

IBM Cognos Analytics
Version 11.1.0

Explorations - Guide d'utilisation



Table des matières

Chapitre 1. Initiation aux explorations.....	1
Explorations.....	1
Téléchargement de données.....	1
Lancement d'une exploration à partir d'une histoire ou d'un tableau de bord existant	1
Lancement d'une nouvelle exploration à partir du menu Nouveau	1
Lancement d'une nouvelle exploration à partir d'un actif de données sur la page d'accueil.....	2
Ajout d'une source de données.....	2
Chapitre 2. Exploration de relations.....	3
Explorer les relations dans vos données.....	3
Ouverture du diagramme de relation.....	7
Chapitre 3. Visualisations.....	9
Visualisations.....	9
Affichage des cartes dans le panneau de navigation	9
Création d'une visualisation unique.....	10
Création d'une visualisation à l'aide de la fonction de recherche dans des zones de données.....	10
Comparaison de deux visualisations.....	11
Comparaison de deux points de données sur une visualisation.....	12
Analyse avancée des données.....	12
Choix d'un type de visualisation différent.....	13
Connaissances dans les visualisations.....	44
Choix des connaissances corrélées.....	45
Choix des visualisations recommandées.....	45
Choix des visualisations liées.....	45
Chapitre 4. Prévisions.....	47
Prévision.....	47
Fonctions de prévision.....	47
Options de prévision.....	49
Types de visualisation prenant en charge les prévisions.....	51
Données de prévision.....	51
Détails statistiques de prévision.....	54
Modèles de prévision.....	57
Chapitre 5. Principes de l'analyse avancée des données.....	61
Principes de l'analyse avancée des données.....	61
Préparation des données.....	62
Préparation des données pour les champs numériques.....	62
Préparation des données pour les champs catégoriels.....	63
Préparation des données pour les champs cible.....	63
Facteurs clés unidirectionnels.....	64
Facteurs clés bidirectionnels.....	65
Arbre de décisions.....	65
Connaissances dans les visualisations.....	67
Détails en langage naturel.....	76
Relations.....	80
Différences entre Cognos Analytics et Watson Analytics en matière d'analyse avancée des données.....	81

Chapitre 6. Assistant.....	87
Assistant.....	87
Fonctions de l'Assistant.....	88

Chapitre 1. Initiation aux explorations

Explorations

Explorer est un espace de travail flexible permettant de découvrir et d'analyser les données. L'exploration d'une visualisation existante peut également s'effectuer à partir d'un tableau de bord ou d'une histoire. Découvrez les relations masquées et identifiez les tendances qui transforment vos données en connaissances. Les informations corrélées sont représentées par une icône verte avec un nombre sur l'axe des X, l'axe des Y ou le titre d'un graphique.

Lancement des explorations

Utilisez l'une des méthodes disponibles pour lancer les explorations.

Téléchargement de données

Téléchargez un actif de données dans votre dossier **Mon contenu** pour l'utiliser dans votre exploration.

Procédure

1. Cliquez sur l'icône **Nouveau**, puis sur **Télécharger les fichiers**.



2. Accédez à l'emplacement où vous avez sauvegardé votre actif de données et sélectionnez ce dernier. L'actif de données apparaît dans le dossier **Mon contenu**.

Lancement d'une exploration à partir d'une histoire ou d'un tableau de bord existant

Lorsque vous utilisez un tableau de bord ou une histoire, vous pouvez créer ou éditer une exploration directement à partir d'une visualisation.

Pourquoi et quand exécuter cette tâche

Pour ouvrir une visualisation dans une nouvelle exploration ou l'ajouter à une exploration existante, procédez comme suit :

Procédure

1. Ouvrez une histoire ou un tableau de bord existant.
2. Sélectionnez une visualisation.
3. Cliquez sur l'icône **Explorations**  dans la barre d'outils.
4. Sélectionnez **Nouvelle exploration** ou **Ajouter à l'existant**.

Lancement d'une nouvelle exploration à partir du menu Nouveau

Sur la page **d'accueil**, vous pouvez lancer une nouvelle exploration à partir du menu **Nouveau**.

Procédure

1. Cliquez sur **Nouveau** , puis sur **Exploration**.
2. Sélectionnez une source de données, puis cliquez sur **Ajouter**.

La page Points de départ est générée à partir de la source de données que vous avez sélectionnée.

Lancement d'une nouvelle exploration à partir d'un actif de données sur la page d'accueil

Vous pouvez sélectionner le **menu Action** pour un actif de données utilisé récemment dans la page **d'accueil**.

Procédure

1. Sur la page **d'accueil**, si l'actif de données à utiliser s'affiche sous forme de vignette dans la zone

Récent, cliquez sur l'icône **Menu Action** .

2. Sélectionnez **Créer l'exploration**.

Ajout d'une source de données

Ajoutez une source de données à votre exploration pour explorer ses données.

Procédure

1. Dans le panneau **Sources sélectionnées**, cliquez sur **Ajouter une source** .

2. Accédez au dossier **Mon contenu** ou **Contenu de l'équipe** et sélectionnez la source de données à ajouter. Cliquez sur **Ajouter**.

3. Développez la source de données dans la sous-fenêtre **Sources sélectionnées** pour voir les éléments disponibles.

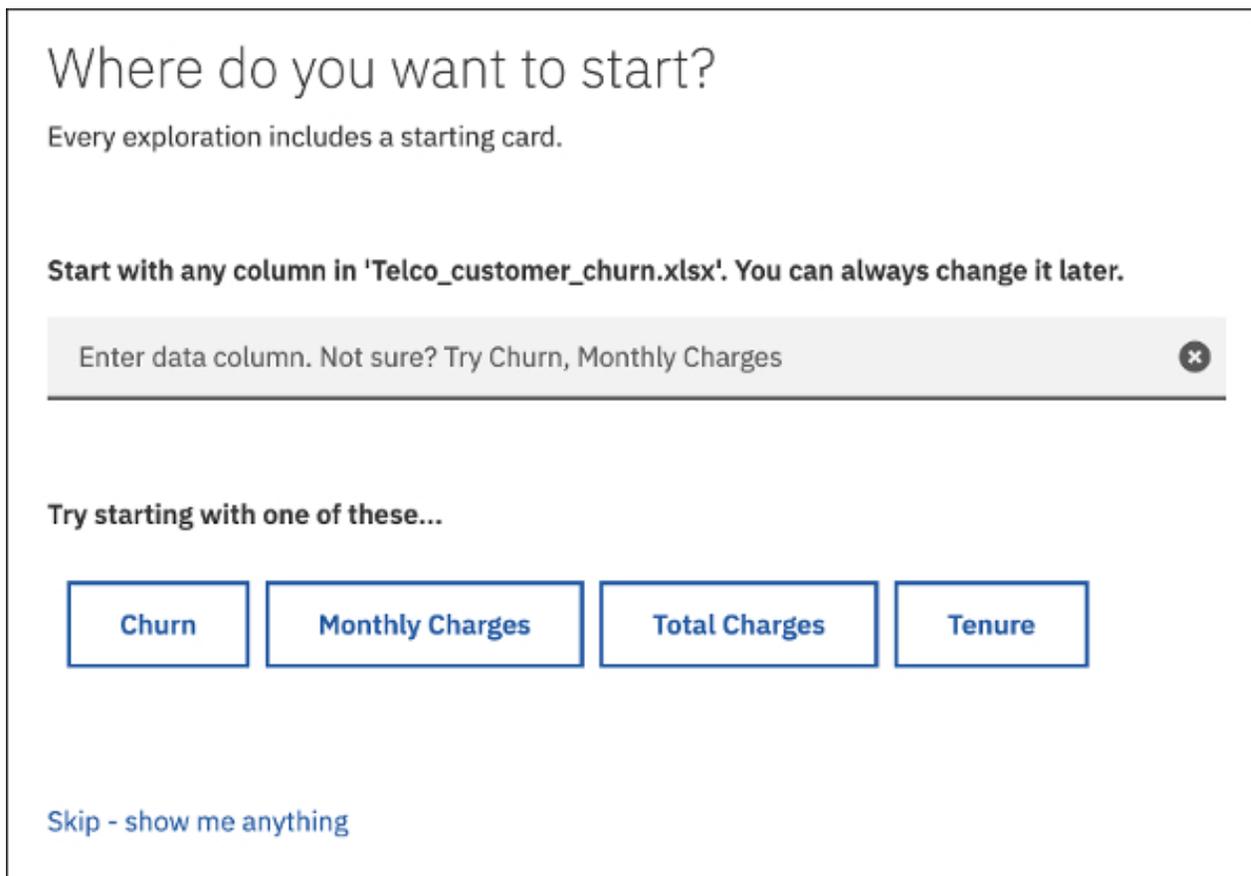
4. Utilisez la page Points de départ pour générer un diagramme de relation à partir de vos données.

Chapitre 2. Exploration de relations

Explorer les relations dans vos données

Lorsque vous créez une exploration, vous pouvez démarrer à partir d'une source de données. La page Points de départ s'affiche et propose des conseils pour vous aider à démarrer.

Vous pouvez saisir un nom de colonne qui apparaît dans votre source de données. Vous pouvez également cliquer sur l'une des colonnes suggérées que le système identifie comme étant intéressante. Si vous n'êtes pas certain de la colonne par laquelle commencer, cliquez sur **Ignorer - me montrer quelque chose** pour afficher un diagramme de relation comportant des visualisations de point de départ suggérées.



Where do you want to start?

Every exploration includes a starting card.

Start with any column in 'Telco_customer_churn.xlsx'. You can always change it later.

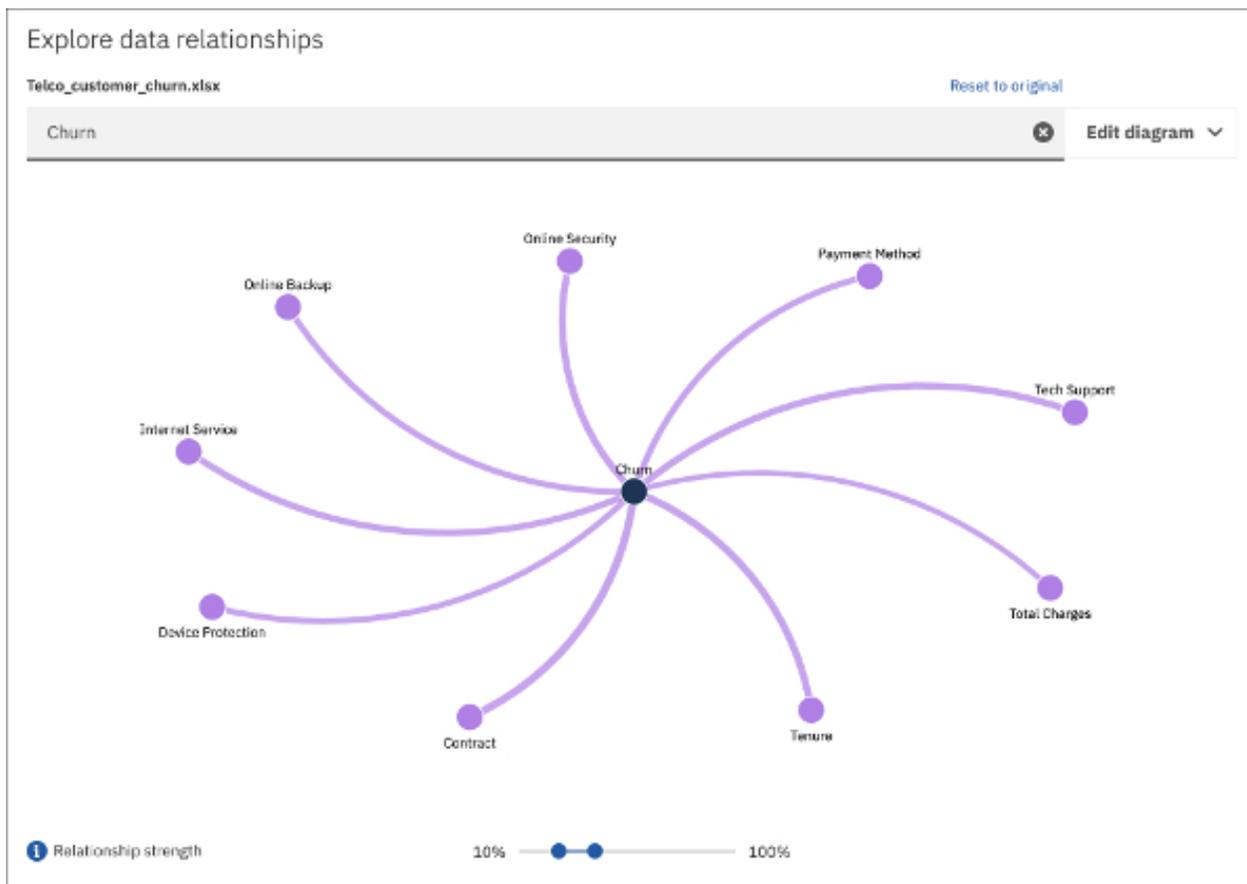
Enter data column. Not sure? Try Churn, Monthly Charges ✕

Try starting with one of these...

[Churn](#) [Monthly Charges](#) [Total Charges](#) [Tenure](#)

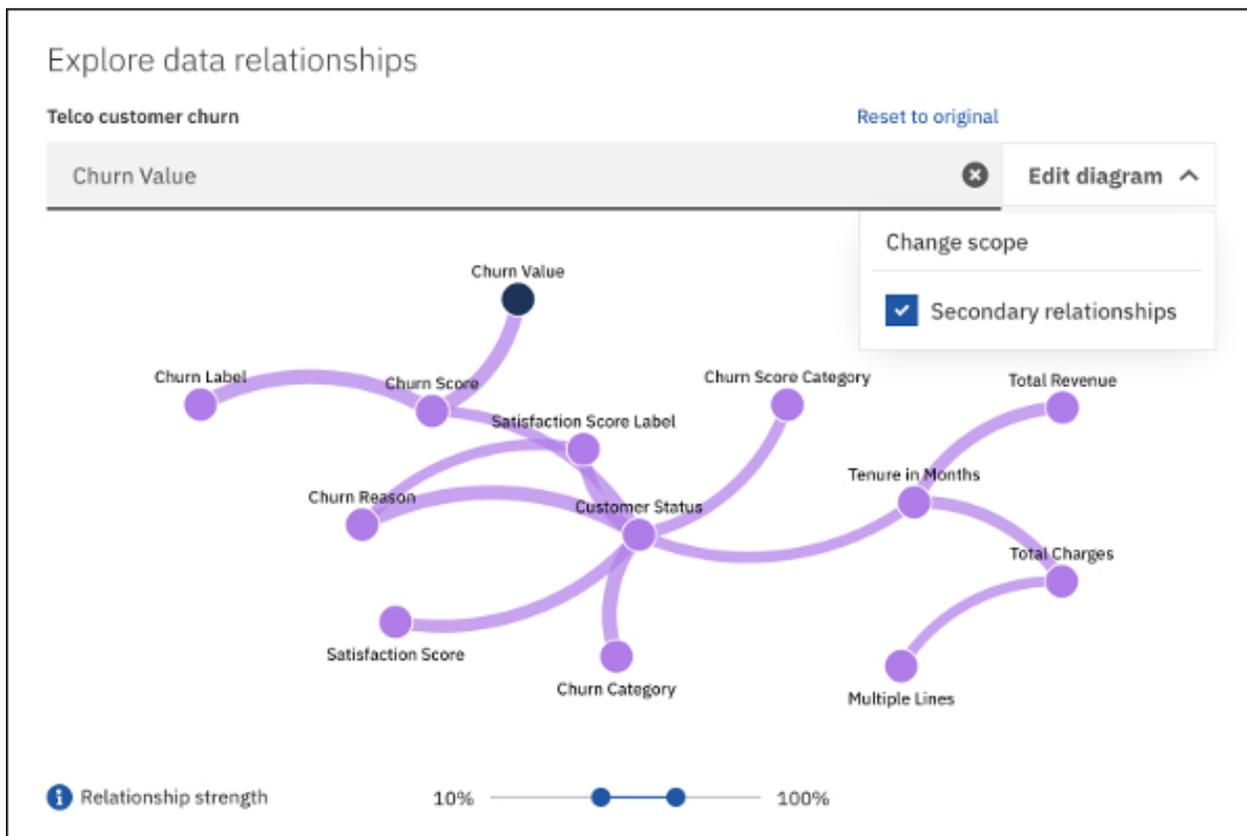
[Skip - show me anything](#)

Dans le diagramme de relation, la colonne avec laquelle vous démarrez est la cible principale : elle est représentée par un noeud bleu foncé. Les zones connexes sont représentées par des noeuds violets. Les lignes connectent les noeuds et représentent les relations. Plus le trait est épais, plus la relation est étroite.



Les relations primaires les plus fortes sont affichées par défaut. Il s'agit des relations directes entre la cible principale et les zones connexes. Les relations secondaires sont les relations entre d'autres zones directement ou indirectement liés à la cible.

Pour afficher les relations primaires et les relations secondaires, cochez la case **Relations secondaires** sous **Editer le diagramme**.



Le diagramme de relation représente ces zones en se basant sur une évaluation statistique des éléments liés. Il ne constitue pas une illustration du modèle de données. Ce dernier peut cependant avoir une incidence sur l'analyse. Lorsque la source de données comporte un grand nombre de lignes, l'analyse est basée sur un échantillon représentatif de l'ensemble des données afin d'améliorer les performances.

Vous pouvez interagir avec le diagramme de relation en sélectionnant un noeud qui vous intéresse. Lorsque vous effectuez cette opération, la liste des visualisations de point de départ suggérées à droite du diagramme est mise à jour pour inclure les noeuds sélectionnés. Vous pouvez également appuyer sur Ctrl tout en cliquant pour sélectionner plusieurs noeuds.

Cliquez sur **Rétablir l'original** pour réinitialiser la portée et la vue de toutes les zones du diagramme de relation au paramètre par défaut.

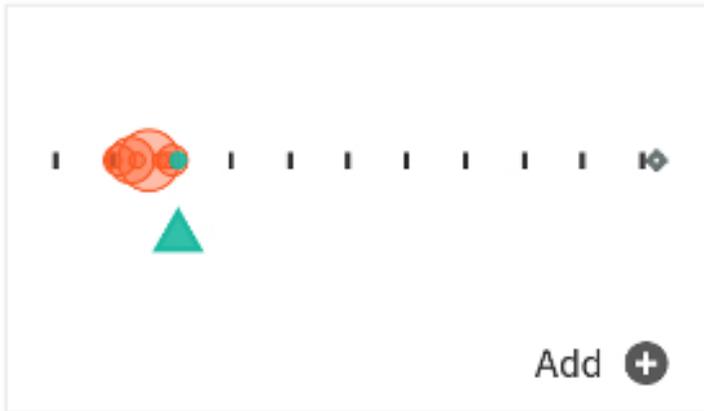
Points de départ suggérés

Les visualisations de point de départ suggérées sont affichées sous forme de miniatures en regard du diagramme de relation. Sélectionnez des noeuds uniques ou plusieurs noeuds dans le diagramme de relation pour générer ces visualisations.

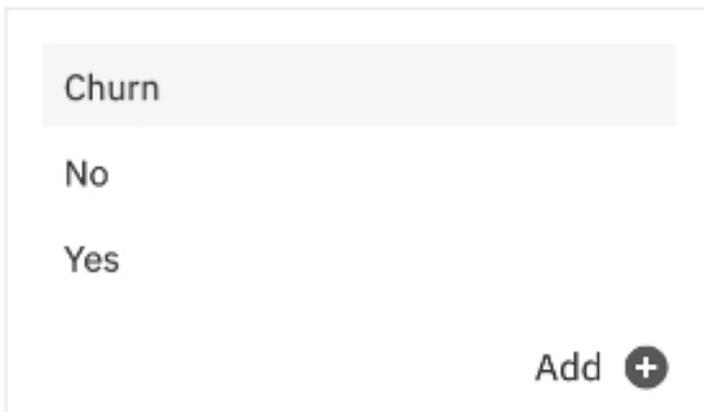
Cliquez sur une visualisation si vous souhaitez l'ajouter à votre exploration et la visualiser en même

temps. Cliquez sur l'icône Plus  dans la visualisation du point de départ pour l'ajouter à votre exploration et gérer la vue en cours.

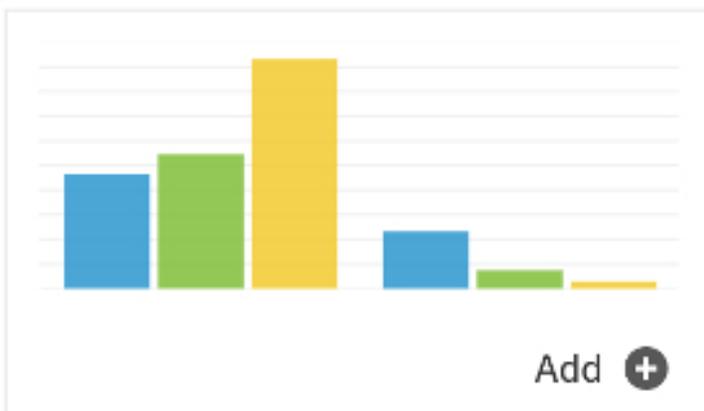
Explore visualizations related to 'Churn'



Churn



Churn



Tenure by Churn colored by Contract

Ouverture du diagramme de relation

Lorsque vous affichez une visualisation et souhaitez revenir au diagramme de relation, utilisez la carte **Relations de données** pour revenir à la vue des points de départ.

Pourquoi et quand exécuter cette tâche

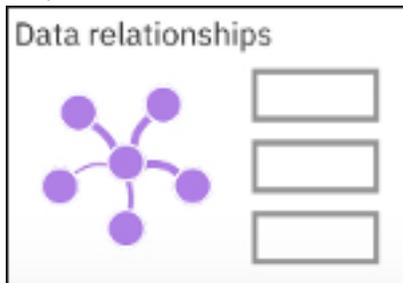
Pour revenir à la vue des points de départ afin d'afficher un diagramme de relation et les points de départ suggérés, procédez comme suit :

Procédure

1. Cliquez sur l'icône **Explorations** dans la sous-fenêtre latérale.



2. Cliquez sur la carte **Relations de données**.



Conseil : La carte **Relations de données** est également accessible à partir du menu **Nouvelle carte** dans la barre d'outils.

Chapitre 3. Visualisations

Visualisations

Vous pouvez modifier le type de visualisation ou les colonnes utilisées dans la visualisation.

Affichage des cartes dans le panneau de navigation

Affichez des miniatures de vos visualisations, appelées cartes, dans le panneau de navigation situé à gauche de la vue principale.

Pourquoi et quand exécuter cette tâche

Les cartes sont une collection de visualisations dans votre Exploration. Elles permettent d'ouvrir vos visualisations pour afficher les détails et les modifier à l'aide des emplacements de données.

Procédure

1. Cliquez sur l'icône **Explorations** dans la sous-fenêtre latérale.



Vos cartes sont répertoriées ici.

2. Cliquez sur la carte qui affiche une miniature de la visualisation que vous souhaitez ouvrir. La visualisation s'ouvre dans la fenêtre principale.
3. Affichez le texte généré dans l'onglet **Détails** ou ajoutez plus de données élémentaires dans l'onglet **Zones**.

Si les onglets **Détails** et **Zones** ne sont pas visibles, cliquez sur **Afficher les détails et les zones** .

Affichage des détails de visualisation

Lorsque vous ouvrez une visualisation, celle-ci s'affiche dans la zone d'exploration principale. Le panneau d'exploration contient l'onglet **Détails**, **Zones** et l'onglet **Propriétés**.

La sous-fenêtre d'exploration n'est pas visible, cliquez sur l'icône **Afficher les détails et les zones** .

Détails de la visualisation

L'onglet **Détails** affiche le texte généré pour décrire les aspects des données représentées dans les visualisations. Ces détails ne sont pas faciles à appréhender dans la visualisation. Ils peuvent par exemple indiquer une moyenne des valeurs au cours d'une période donnée.

Champs

L'onglet **Zones** permet d'ajouter des colonnes pour générer et modifier des visualisations. Ajoutez une colonne à chaque champ obligatoire.

Propriétés

L'onglet **Propriétés** permet de modifier les propriétés qui s'appliquent à vos visualisations.

Création d'une visualisation unique

Il se peut que vous ayez besoin d'une autre visualisation dans votre exploration.

Pourquoi et quand exécuter cette tâche

Pour créer une visualisation unique, procédez comme suit :

Procédure

1. Dans la barre d'outils, sélectionnez **Nouvelle carte**.
2. Sélectionnez la carte vide **Unique**.
3. Dans la fenêtre **Sources**, développez l'actif de données à utiliser.

Si un autre actif de données est ouvert, cliquez sur **Précédent** en regard de son nom.



4. Pour créer une visualisation, effectuez l'une des actions suivantes :

- Faites glisser les éléments de données dans la zone **Créer une visualisation**.

IBM® Cognos Analytics crée une visualisation correspondant aux éléments de données. Par exemple, si vous ajoutez Année ou Département, un tableau est créé. Si vous faites glisser une mesure, telle que Revenus, une visualisation à barres est créée.

- Cliquez sur **Choisir un type** et sélectionnez un type de visualisation. Ajoutez ensuite des données élémentaires à chaque zone.

Création d'une visualisation à l'aide de la fonction de recherche dans des zones de données

Vous pouvez créer rapidement une visualisation qui utilise la fonction de recherche dans des zones de données.

Procédure

1. Cliquez sur l'icône **Explorations** dans la sous-fenêtre latérale.



Vos cartes sont répertoriées ici.

2. Cliquez sur la carte qui affiche une miniature de la visualisation que vous souhaitez ouvrir. La visualisation s'ouvre dans la fenêtre principale.
3. Cliquez sur l'onglet **Zones**.

Si l'onglet **Zones** n'est pas visible, cliquez sur **Afficher les détails et les zones** .

4. Recherchez les données dont vous avez besoin. Cliquez sur les données pour remplir la zone de données.

Résultats

Details **Fields** Properties

 Bars

City|

 California Zip Website Visits Xlsx

 City

Length* Required field

⋮ Blog Visits ⋮

⋮ Unique Visitors ⋮

Click or drag data here

y-start

Click or drag data here

 Target

Click or drag data here

 Color

⋮ Measures group (2)

Click or drag data here

Comparaison de deux visualisations

Vous pouvez créer votre propre comparaison pour analyser les données entre deux visualisations. Vous pouvez également commencer par une comparaison recommandée. Dans les deux cas, un récapitulatif des principales informations et des différences entre les deux visualisations est généré.

Pourquoi et quand exécuter cette tâche

Pour créer une comparaison entre deux visualisations, procédez comme suit :

Remarque : Lorsque vous créez une visualisation, vous pouvez sélectionner une carte de comparaison vide avec deux emplacements pour les visualisations.

Procédure

1. Cliquez sur l'icône **Explorations**  dans le panneau latéral.
La sous-fenêtre **Cartes** s'ouvre.
2. Sélectionnez une carte pour créer une comparaison.
Une visualisation s'affiche.
3. Dans la barre d'outils, cliquez sur **Comparer**.
La page **Comment voulez-vous comparer ?** s'affiche. Elle contient des conseils sur la création de votre propre comparaison ou sur le démarrage avec une comparaison recommandée.
4. Cliquez sur l'icône **Ajouter une carte**  sur une miniature de carte pour l'ajouter à la liste des cartes dans le panneau de navigation. Vous pouvez également cliquer sur la miniature de la carte pour ajouter la nouvelle carte et l'afficher immédiatement.
5. Modifiez éventuellement les données d'une visualisation pour la comparer à l'autre visualisation.
 - a) Sélectionnez l'une des deux visualisations.
 - b) Dans l'onglet **Zones**, modifiez la visualisation de l'une des manières suivantes, par exemple :
 - Supprimer des filtres.
 - Afficher le nombre supérieur ou inférieur.
 - Supprimer les éléments de données.
 - Dans la sous-fenêtre **Sources**, ajoutez de nouvelles données élémentaires ou de nouveaux filtres. Vous pouvez également utiliser la recherche dans les zones de données. Pour plus d'informations, voir [«Création d'une visualisation à l'aide de la fonction de recherche dans des zones de données»](#), à la page 10.

Comparaison de deux points de données sur une visualisation

Vous pouvez sélectionner deux points de données sur une visualisation existante et comparer les données.

Procédure

1. Cliquez sur l'icône **Explorations**  dans le panneau latéral.
La sous-fenêtre **Cartes** s'ouvre.
2. Dans la sous-fenêtre **Cartes**, sélectionnez la carte qui affiche une miniature de la visualisation que vous souhaitez ouvrir.
La visualisation s'ouvre dans la fenêtre principale.
3. Sélectionnez deux points de données sur la visualisation.
4. Cliquez avec le bouton droit de la souris, puis sélectionnez **Comparer par**.
5. Indiquez une colonne pour comparer les deux points de données.
Un tableau affiche des informations sur cette comparaison.

Analyse avancée des données

IBM Cognos Analytics est un outil d'aide à la décision qui permet de gérer et d'analyser les données. Il inclut diverses fonctionnalités en libre-service qui permettent aux utilisateurs de préparer, d'explorer et de partager des données. Dans le cadre de cette offre, Cognos Analytics inclut un certain nombre de techniques prévisionnelles, descriptives et exploratoires, également connues sous le nom d'intelligence numérique. Cognos Analytics utilise de nombreux tests statistiques pour analyser vos données. Il est indispensable de comprendre les définitions de ces tests qui s'appliquent à Cognos Analytics.

Pour plus d'informations, voir le manuel *IBM Cognos Analytics - Guide des tableaux de bord et histoires*.

Choix d'un type de visualisation différent

Les visualisations communiquent les comparaisons, les relations et les tendances. Elles mettent en évidence et clarifient des valeurs numériques. Pour choisir un type de visualisation, prenez en compte ce que la visualisation doit illustrer et ce qui attire les utilisateurs concernés.

Avant de commencer

Pour plus d'informations sur les types de visualisation, voir la documentation relative aux visualisations dans le document *IBM Cognos Analytics - Guide d'utilisation des tableaux de bords et des histoires*.

Procédure

1. Dans la sous-fenêtre **Cartes**, sélectionnez la carte qui représente la visualisation que vous souhaitez ouvrir.
2. Cliquez sur l'icône **Choisir le type de visualisation**  dans la barre d'outils.
3. Cliquez sur le type de visualisation à utiliser.

Vous verrez que chaque type de visualisation communique les données de manière différente. Par exemple, utilisez une visualisation à barres, à colonnes ou à courbes pour comparer un ensemble de valeurs. Utilisez une visualisation à courbes ou en aires pour effectuer le suivi des relations. Utilisez une carte d'arborescence ou une visualisation de graphique circulaire pour examiner les parties d'un tout.

Aires

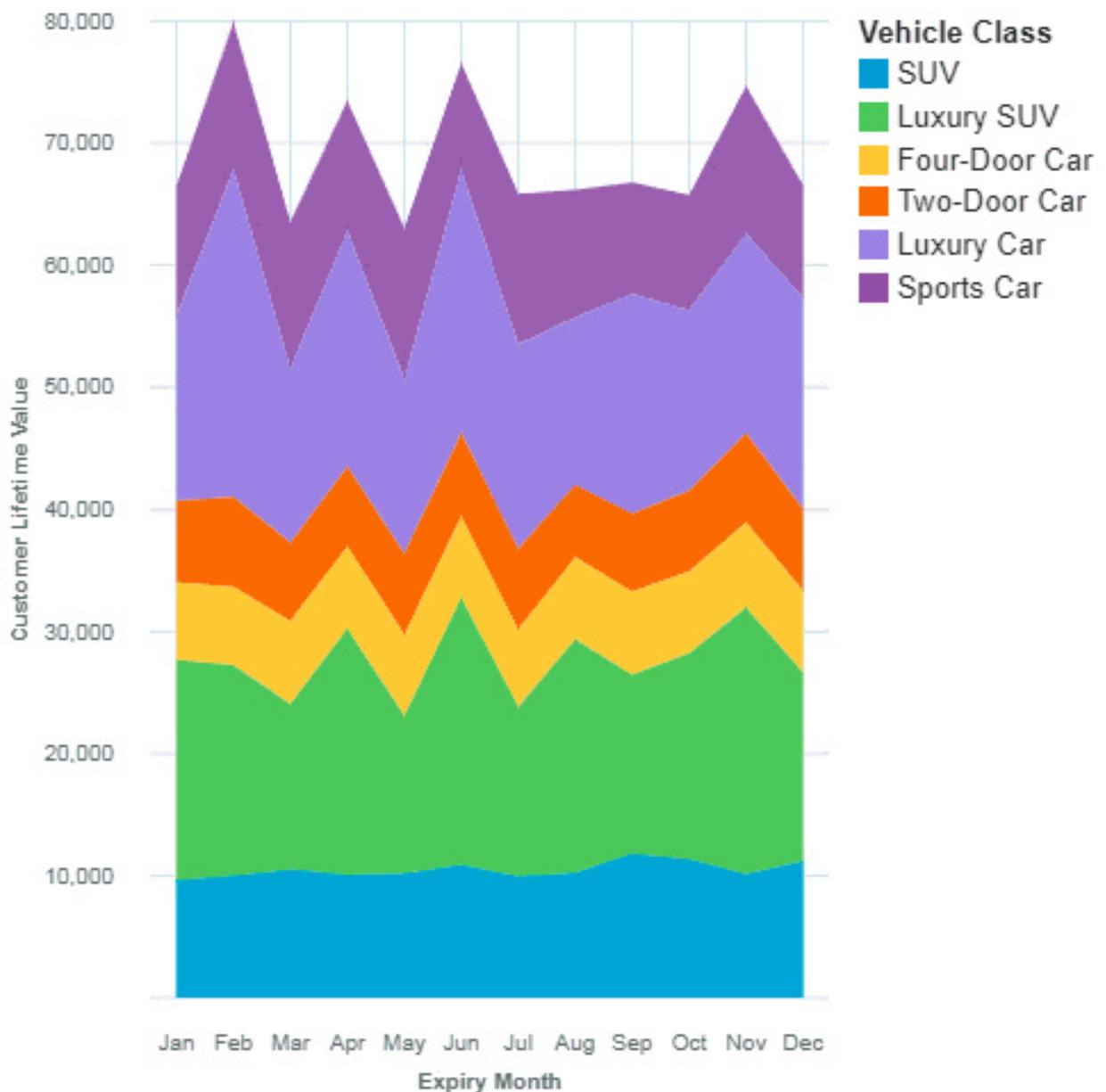
Utilisez une visualisation en aires pour souligner l'ampleur des changements dans le temps.

Les graphiques en aires sont comparables aux graphiques à courbes, mais les aires situées sous les courbes sont remplies d'une couleur ou d'un motif. Les graphiques empilés proportionnels sont utiles pour comparer des contributions proportionnelles dans une catégorie. Elles illustrent la part relative représentée par chaque série de données dans le total.

Etant donné qu'une visualisation en aires superpose les résultats pour chaque colonne ou élément, le total de tous les résultats est facilement visible.

Par exemple, une visualisation en aires est idéale afin d'étudier les revenus au fil du temps pour plusieurs produits.

Par exemple, cette visualisation en aires représente la valeur à vie pour le client par classe de véhicule et par mois. La visualisation en aires superposant les résultats, vous pouvez voir les totaux de chaque mois.



La visualisation en aires a été créée en faisant glisser les données élémentaires suivantes à partir du panneau Sources :

- Faites glisser le type **Mois d'expiration** dans le champ **Axe des X**.
- Faites glisser **Classe de véhicule** dans le champ **Couleur**.
- Faites glisser **Valeur à vie pour le client** dans le champ **Axe des Y**.

Exemples

Vous pouvez voir un exemple de visualisation en nuage de mots dans l'exemple de rapport **Analyse de la valeur de durée de vie du client**. Vous trouverez cet exemple ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

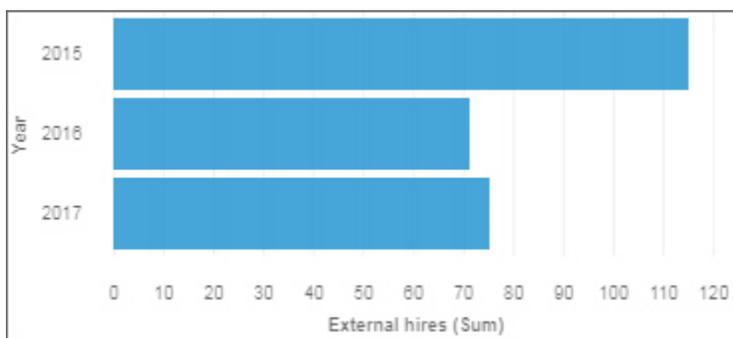
Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Barre

Utilisez une visualisation à barres pour comparer les valeurs entre une ou plusieurs colonnes, telles que les ventes de produits ou les ventes de produits mensuelles.

Les visualisations à barres utilisent des repères de données horizontaux organisés en groupes pour comparer des valeurs individuelles. Vous pouvez utiliser des visualisations à barres pour comparer des données discrètes ou pour afficher des tendances dans le temps.

Une visualisation à barres peut représenter les changements au cours d'une période spécifique ou comparer et opposer plusieurs colonnes sur une période ou dans le temps. Si le nombre de barres est si important que les libellés sont illisibles, filtrez les données pour vous concentrer sur un sous-ensemble de ces données ou utilisez une carte d'arborescence.



Utilisez le champ **Cible** pour représenter les mesures à comparer à une valeur cible.

Utilisez le champ **y-start** pour définir où la mesure doit commencer.

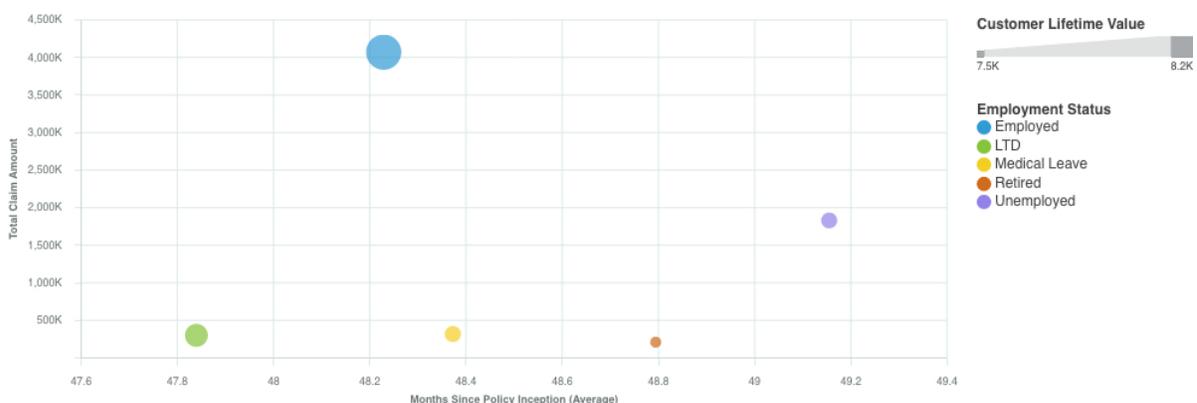
Bulles

Utilisez une visualisation à bulles pour représenter les relations entre des colonnes qui contiennent des valeurs numériques, telles que les revenus et le bénéfice.

Une visualisation à bulles utilise des points de données et des bulles pour représenter des mesures n'importe où le long d'une échelle. Une mesure est tracée sur chaque axe. La taille de la bulle représente une troisième mesure. Utilisez les visualisations à bulles pour représenter des données financières ou n'importe quelle donnée où des valeurs de mesure sont en relation.

Les bulles sont de différentes tailles et de différentes couleurs. L'axe des X représente une seule mesure. L'axe des Y représente une autre mesure et la taille des bulles représente la troisième mesure. Dans l'exemple ci-dessous, la couleur est représentée par un identificateur.

L'exemple illustré représente les mois depuis le début de la règle.



Créez la visualisation à bulles en faisant glisser les données élémentaires suivantes à partir de la section



Analyse des clients de la sous-fenêtre **Sources** :

- Faites glisser **Mois depuis la création de la règle** dans le champ **Axe des X**.
- Faites glisser **Montant total du sinistre** dans le champ **Axe des Y**.
- Faites glisser **Valeur à vie pour le client** dans le champ **Taille**.
- Faites glisser **Situation professionnelle** dans le champ **Couleur**.

Vous pouvez personnaliser la visualisation à bulles. Par exemple, pour créer l'axe des X de la visualisation à bulles comme dans l'exemple, procédez comme suit :

1. Cliquez sur la visualisation, puis dans la sous-fenêtre **Données**, cliquez sur la donnée élémentaire **<Total Claim Amount>**.
2. Cliquez sur 
3. En regard de **Format des données**, cliquez sur  et définissez les options suivantes :
 - **Type de format** : Devise
 - **Symbole de la devise** : K
 - **Position du symbole de la devise** : Fin
 - **Nombre de décimales** : 0
 - **Echelle** : -3 (affiche les valeurs en milliers).
4. Cliquez sur **OK**.

Pour modifier la taille de la visualisation, cliquez sur la visualisation, puis définissez l'option suivante dans la sous-fenêtre Propriétés :

- **Taille - Largeur** : 700 px, **Hauteur** : 300 px

Cliquez sur  pour fermer la sous-fenêtre **Propriétés**.

Exemples

Vous pouvez voir des exemples de visualisation dans l'exemple de rapport **Analyse des valeurs à vie pour le client**. Vous trouverez ces exemples ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Puce

Utilisez des graphiques à puces pour afficher des mesures qui doivent être comparées à une valeur cible.

Dans un centre d'appels, un graphique à puces peut être utilisé pour mesurer des indicateurs, comme le volume d'appels, la rapidité de traitement d'un appel et le pourcentage d'appels abandonnés.

Dans un processus de fabrication, un graphique à puces peut être utilisé pour suivre des indicateurs, comme le nombre d'anomalies et de commandes expédiées

Dans un contexte de remise en forme, un graphique à puces peut être utilisé pour mesurer des indicateurs, comme le nombre de pas effectués et de calories brûlées

Les visualisations à puces comparent une mesure réelle (la puce) à la mesure ciblée (la cible). Les visualisations à puces mettent également en relation les mesures comparées et des zones de couleurs situées en arrière-plan qui fournissent des mesures plus qualitatives, par exemple bon, satisfaisant et médiocre. Elles peuvent être affichées dans un petit format et transmettre efficacement des informations.

Une visualisation à puces comporte une mesure principale unique, par exemple, le cumul annuel à ce jour des revenus, et la compare à une ou plusieurs autres mesures pour enrichir sa signification, par exemple lors d'une comparaison avec une cible. La mesure principale est affichée dans le contexte d'une plage qualitative de performances, par exemple médiocre, satisfaisant et bon.

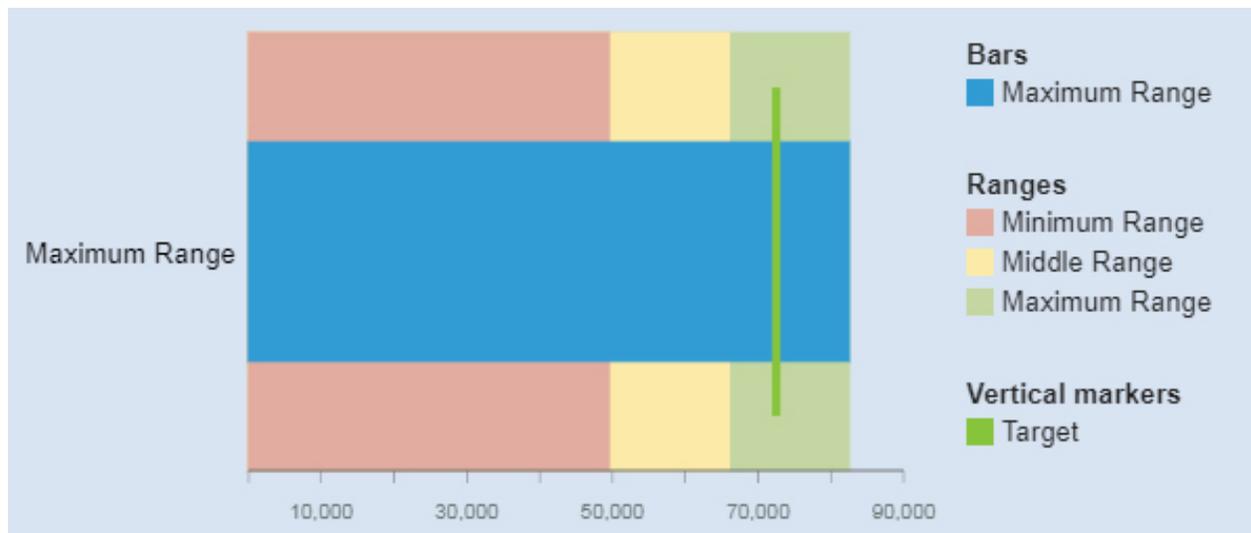
Si vous sélectionnez une visualisation à puces, définissez les zones suivantes :

- Le champ **Barre réelle** indique la mesure réelle.
- Le champ **Cible** indique la mesure cible.

- Le champ **Plage minimale** indique la plage qualitative minimale.
- Le champ **Plage intermédiaire** indique la plage qualitative intermédiaire.
- Le champ **Plage maximale** indique la plage qualitative supérieure.

Remarque : L'accès au détail n'est pas disponible pour la visualisation à puces.

Vérifiez que les plages minimale, intermédiaire et maximale s'appliquent à la mesure réelle et à la mesure cible.



La visualisation à puces a été créée en faisant glisser les données élémentaires suivantes à partir du panneau Sources :

- Faites glisser **Plage minimale** dans le champ **Plage minimale**.
- Faites glisser **Plage maximale** dans le champ **Plage maximale**.
- Faites glisser **Plage maximale** dans le champ **Barre réelle**.
- Faites glisser **Cible** dans le champ **Cible**.
- Faites glisser **Classe de véhicule** dans le champ **Données supplémentaires**.

Exemples

Vous pouvez voir un exemple de visualisation à puces dans l'exemple de rapport **Analyse des valeurs à vie pour le client**. Vous trouverez cet exemple ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

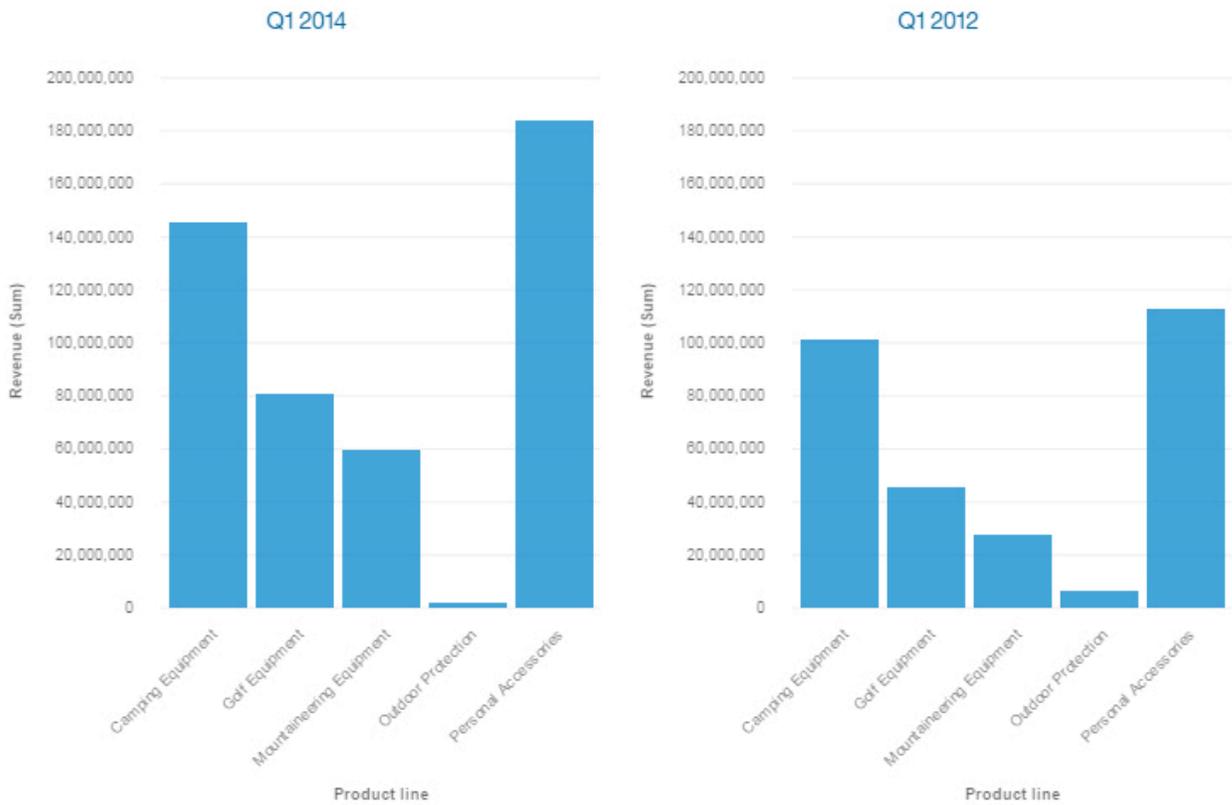
Colonnes

Utilisez une visualisation à colonnes pour comparer les valeurs entre une ou plusieurs colonnes, telles que les ventes de produits ou les ventes de produits mensuelles.

Les visualisations à colonnes utilisent des repères de données verticaux organisés en groupes pour comparer des valeurs individuelles. Utilisez des visualisations à colonnes pour comparer des données discrètes ou pour afficher des tendances dans le temps.

Une visualisation à colonnes représente les changements au cours d'une période spécifique ou comparer et opposer plusieurs colonnes sur une période ou dans le temps. Si le nombre de barres est si important que les libellés sont illisibles, filtrez les données pour vous concentrer sur un sous-ensemble de ces données ou utilisez une carte d'arborescence.

Par exemple, les revenus de chaque ligne de produits sont regroupés par trimestre, ce qui met en avant les performances au cours de chaque trimestre.



Utilisez le champ **Cible** pour représenter les mesures à comparer à une valeur cible.

Utilisez le champ **y-start** pour définir où la mesure doit commencer.

Tableau croisé

Utilisez un tableau croisé si vous souhaitez afficher les relations entre trois colonnes ou plus. Les tableaux croisés affichent les données dans des lignes et des colonnes, les informations étant récapitulées aux points d'intersection.

Par exemple, ce tableau croisé représente les coûts des cours pour chaque département, par organisation.

Course cost	GO Asia Pacific corporate						
	Executive Offices	Finance	Sales (Corporate)	Procurement	Human Resources	Marketing	Summary
2014	10,500	29,750	10,500	18,250	11,000	15,500	95,500
2015	11,500	46,250	14,750	7,750	11,500	30,750	122,500
2016	12,500	43,000	33,250	5,000	14,500	22,500	130,750
2017	13,500	24,000	36,000	7,000	15,750	21,750	118,000
Summary	48,000	143,000	94,500	38,000	52,750	90,500	466,750

A partir de Cognos Analytics version 11.1.4, vous pouvez faire glisser les données du panneau **Sources sélectionnées** et insérer les données dans une colonne/ligne ou faire glisser les données sur les données existantes pour les remplacer.

Lecteur de données

Utilisez un lecteur de données pour visionner une animation présentant l'impact d'une colonne sur les autres visualisations.

Arbre de décisions

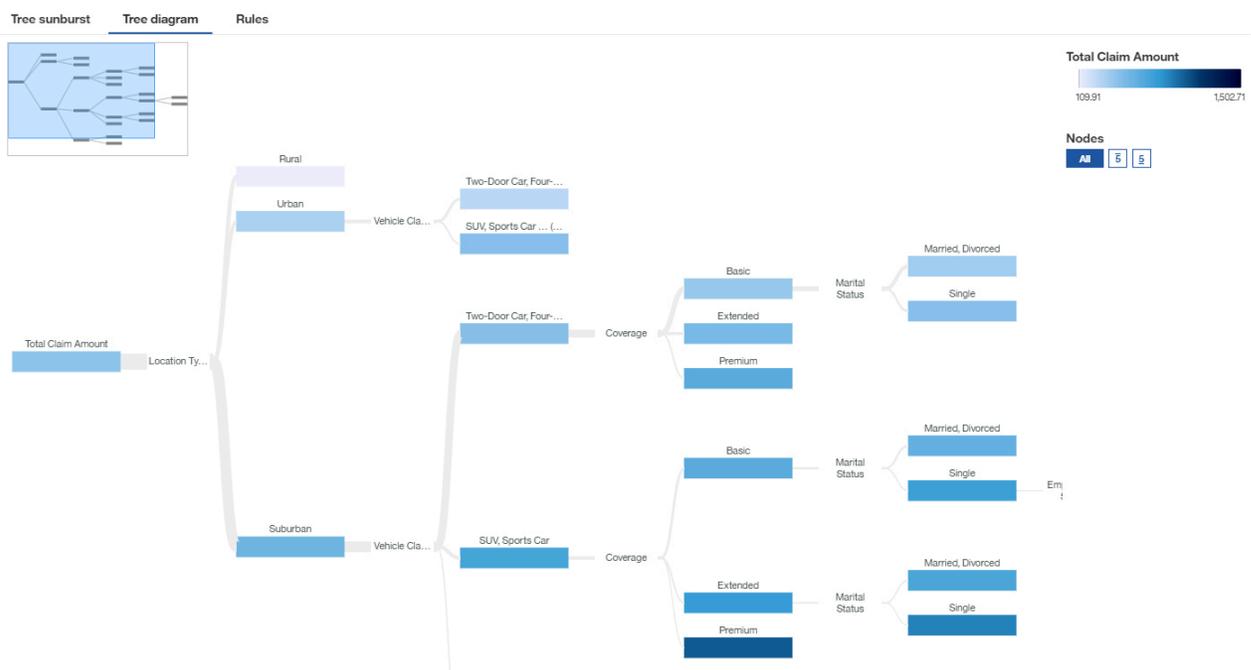
Un arbre de décisions affiche une hiérarchie connectée de cadres qui représentent les valeurs des enregistrements.

Les enregistrements sont répartis dans des groupes appelés "noeuds". Chaque noeud contient des enregistrements statistiquement similaires pour la zone cible. Par exemple, un noeud peut contenir les enregistrements représentant les hommes dont la durée de scolarité est supérieure à 18 ans. Les noeuds peuvent ensuite être utilisés pour prévoir la valeur de la zone d'une cible. Par exemple, le noeud représentant les hommes et la durée de scolarité peut être utilisé pour prévoir le salaire.

Chaque branche d'un arbre de décisions correspond à une règle de décision. Pour plus d'informations sur les règles de décision, voir [«Affichage des règles de décision»](#), à la page 22

Compte tenu du nombre de lignes dans la source de données, l'analyse est basée sur un échantillon représentatif de la totalité des données pour améliorer les performances.

Par exemple, une visualisation de type Arbre de décisions peut ressembler à la figure suivante :



Remarque : Les filtres ne sont pas pris en charge dans les visualisations de type Arbre de décisions.

Pour plus d'informations, voir [«Exploration d'une visualisation incluant un arbre de décisions»](#), à la page 19.

Exploration d'une visualisation incluant un arbre de décisions

Un arbre de décisions est utilisé pour indiquer comment des données sous-jacentes prévoient une cible choisie et met en évidence des informations clés sur l'arbre de décisions.

Pourquoi et quand exécuter cette tâche

La force prédictive d'un arbre de décisions détermine dans quelle mesure les décisions représentées par chaque branche de l'arbre prévoient la valeur de la cible.

Les arbres de décisions ont une seule cible. Si la zone cible de l'arbre de décisions est une zone continue, les indicateurs d'informations clés mettent en évidence les groupes anormalement élevés ou faibles. Si la zone cible de l'arbre de décisions est une zone catégorielle, les informations clés correspondent au mode du noeud. Le mode du noeud représente la ou les catégories de la zone cible qui apparaissent le plus fréquemment au sein du groupe.

Compte tenu du nombre de lignes dans la source de données, l'analyse est basée sur un échantillon représentatif de la totalité des données pour améliorer les performances.

Lorsque vous examinez un arbre de décisions :

- Si vous souhaitez visualiser tous les pilotes, utilisez l'onglet **Diagramme d'arbre** ou **Règles**.
- Si vous souhaitez vous concentrer sur des pilotes clés, utilisez l'onglet **Rayons de l'arbre**.

Pour modifier ou ajouter des facteurs clés, cliquez sur l'icône  sur le champ cible.

Les informations d'analyse sont différentes en fonction du type de cible. Si vous prévoyez une mesure continue, comme le revenu, l'âge ou les bénéfiques, l'arbre de décisions affiche au sein du noeud la valeur moyenne de la cible selon les conditions appliquées jusqu'ici dans le groupe représenté par le noeud. Par exemple, vous avez un arbre qui prévoit les revenus et vous avez une branche représentant le genre, puis la ville. Si vous souhaitez suivre le chemin entre Homme et Chicago, la valeur du noeud Chicago représente les revenus moyens des hommes à Chicago.

Procédure

1. Si vous avez une mesure continue, l'exemple suivant décrit l'arbre de décisions :

La couleur indique si la valeur du noeud est associée à une valeur élevée, moyenne ou basse de la cible. La couleur du noeud est basée sur la moyenne de la cible pour la mesure. Plus la valeur moyenne de la cible est élevée pour un noeud, plus la couleur est foncée.

Par exemple, la figure ci-après présente une visualisation détaillée des dépenses de restauration. Le terminal international est un prédicteur fort de dépenses de restauration élevées pour les professionnels qui voyagent. Le manque de propreté de l'aéroport est un prédicteur de dépenses faibles pour les voyageurs qui se rendent à des conférences ou à des conventions.

Le miniplan vous aide à vous déplacer dans les différentes zones de l'arbre. Il est particulièrement utile si les noeuds sont nombreux.

Dans cet exemple, les cinq valeurs cible les plus élevées sont indiquées avec un nombre. Vous pouvez sélectionner l'une des options suivantes :

- Arborescence complète. Les valeurs les plus élevées ou les plus basses ne sont pas spécifiquement indiquées.



- Cinq valeurs cible les plus élevées. Les cinq valeurs cible les plus élevées sont affichées.

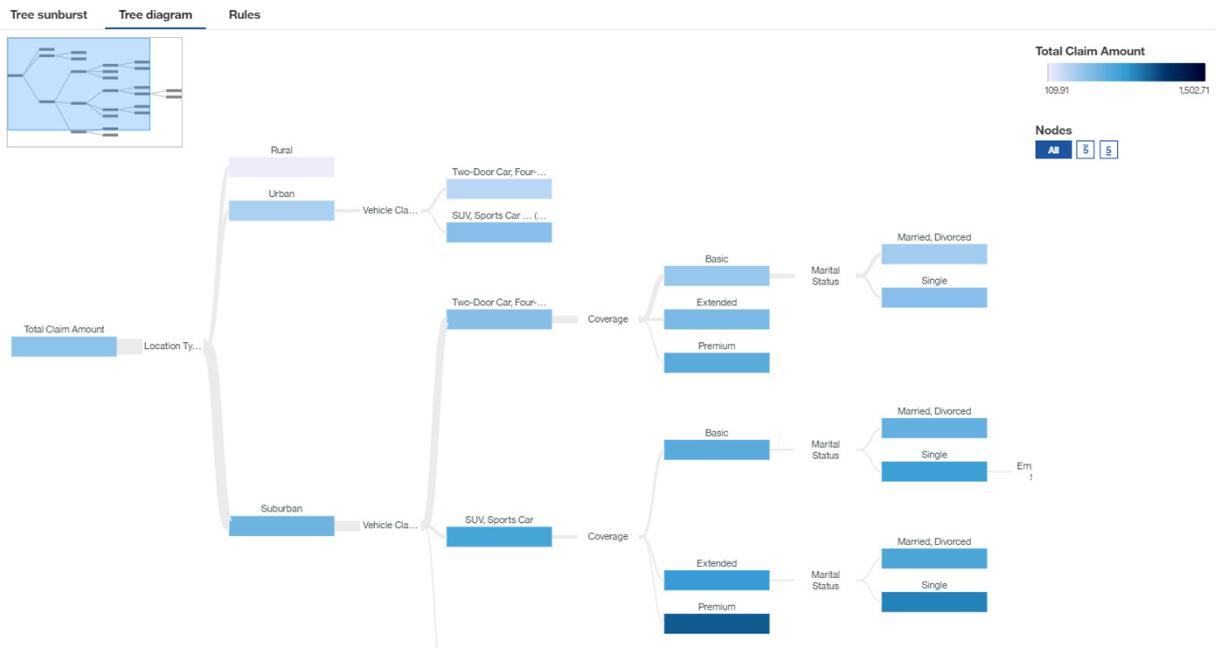


- Cinq valeurs cible les plus basses. Les cinq valeurs cible les plus basses sont affichées.



Si vous avez une mesure catégorielle, sélectionnez la catégorie dont vous souhaitez visualiser les cinq cibles les plus élevées ou les plus basses dans le menu **5 premiers noeuds pour :** ou **5 derniers noeuds pour :**

Les cinq premiers ou les cinq derniers noeuds ne sont pas visibles si vous avez appliqué un zoom avant trop élevé.

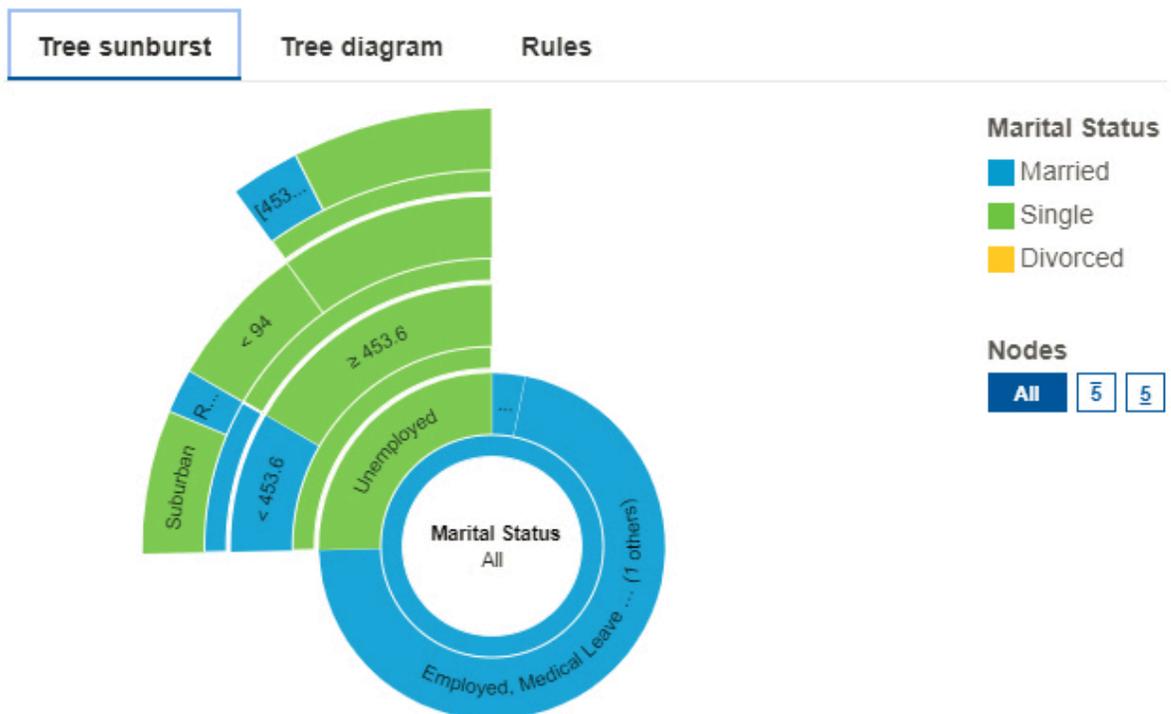


2. Si vous avez une mesure catégorielle, l'exemple suivant présente l'arbre de décisions :

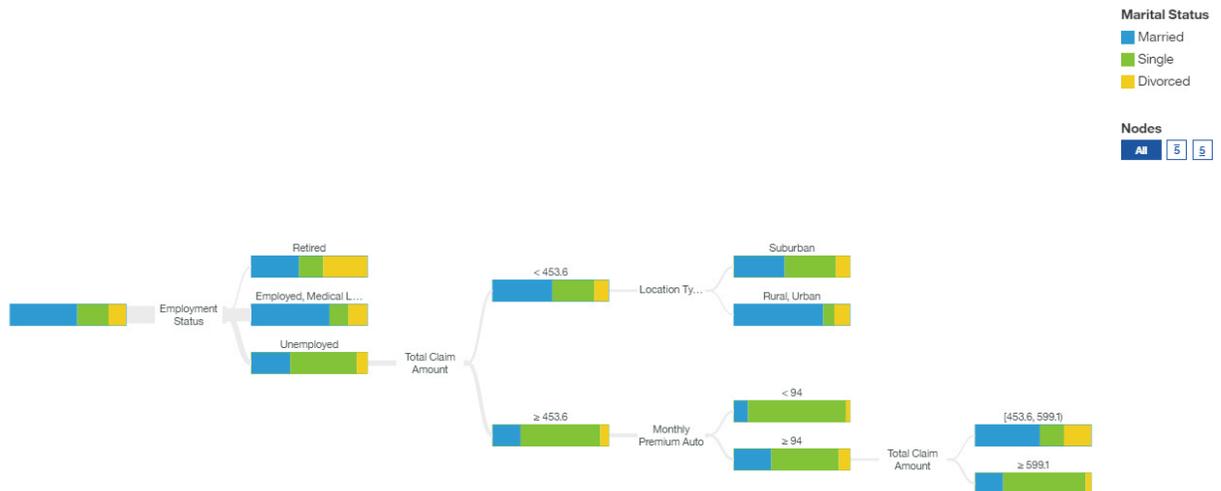
La couleur indique quelles valeurs de zone sont le plus représentées.

Dans l'onglet **Rayons de l'arbre**, vous pouvez constater que si les mesures de l'arbre de décisions sont des prédicteurs forts pour une ou plusieurs valeurs cible, les couleurs sont prédominantes sur ce noeud. Les valeurs non significatives sont omises.

Par exemple, la figure ci-après affiche une visualisation détaillée du statut marital dans l'onglet **Rayons de l'arbre**. Elle indique qu'avoir un emploi est un prédicteur fort pour être marié.



Dans l'onglet **Diagramme d'arbre**, les noeuds représentent visuellement la distribution des personnes en fonction de leur statut marital.



Affichage des règles de décision

Une règle de décision prévoit un résultat dans une zone cible. L'affichage des règles de décision permet de déterminer quelles conditions sont susceptibles d'aboutir à un résultat donné.

Par exemple, examinons un certain nombre de règles de décision hypothétiques qui pourraient prévoir une attrition. Ces règles peuvent identifier des classifications basées sur les tranches d'âge des clients et le nombre de sinistres précédents. A partir de ces règles, on peut observer que les clients qui n'ont aucun ou qu'un seul sinistre et qui ont plus de cinquante dans risquent de résilier leur contrat.

Chaque branche d'un arbre de décisions correspond à une règle de décision.

Procédure

1. Dans un arbre de décisions, appuyez sur **Règles**.
2. Examinez les règles de décision.
3. Pour revenir à la visualisation, appuyez sur **Diagramme d'arbre**.

Analyse de pilotes

Une visualisation de pilotes affiche les pilotes clés, ou prédicteurs, d'une cible. Plus le facteur est à droite, plus le facteur est fort.

IBM Cognos Analytics utilise des algorithmes extrêmement pointus pour obtenir des informations interprétables basées sur une modélisation complexe. Il est inutile de savoir quels tests statistiques doivent être exécutés sur vos données. Cognos Analytics sélectionne les tests appropriés pour les données.

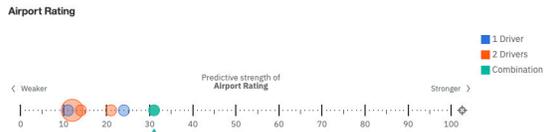
Les pilotes clés pour les cibles continues et catégorielles sont disponibles dans l'analyse des pilotes dans les tableaux de bord et les explorations.

Pour plus d'informations, voir la section *Tests statistiques* du document *IBM Cognos Analytics Tableaux de bord et histoires - Guide d'utilisation*.

Par exemple, cette visualisation incluant une analyse de pilote indique que la combinaison de la satisfaction globale, de l'évaluation de la signalisation, de l'évaluation de la sécurité et de l'évaluation artistique sont les pilotes les plus importants de l'évaluation de l'aéroport cible.

Pour modifier ou ajouter des pilotes clés, cliquez sur l'icône  sur l'emplacement de données cible.

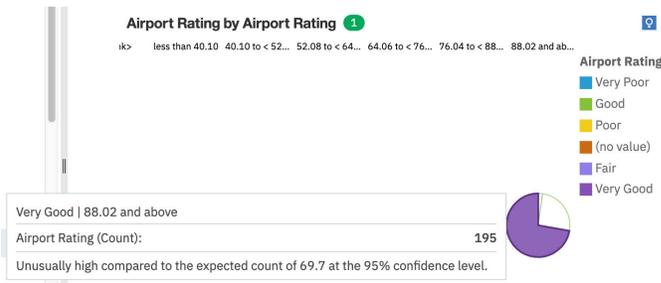
Compte tenu du nombre de lignes dans la source de données, l'analyse est basée sur un échantillon représentatif de la totalité des données pour améliorer les performances.



Si vous placez le pointeur de la souris sur un point de données, la visualisation incluant une analyse de pilotes indique les pilotes qui déterminent l'évaluation de l'aéroport.



Si vous cliquez sur un point de données dans l'arborescence, d'autres visualisations recommandées sont affichées.



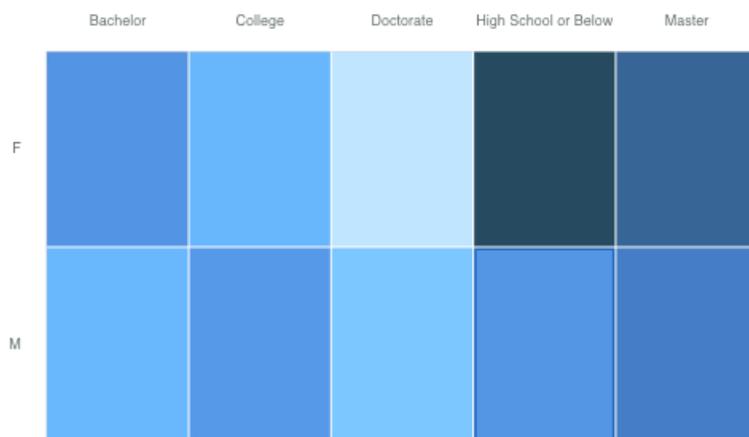
Remarque : Les filtres ne sont pas pris en charge dans les visualisations incluant une analyse de pilotes.

Carte de densité

Utilisez une visualisation de type Carte de densité pour visualiser les relations entre les colonnes, représentées dans une vue matricielle.

Une visualisation de type Carte de densité utilise les couleurs et l'intensité des couleurs pour représenter les relations entre deux colonnes.

Par exemple, cette visualisation de type Carte de densité représente la valeur moyenne à vie pour le client par sexe et formation.



Créez la visualisation de type Carte de densité en faisant glisser les données élémentaires suivantes à

partir du panneau **Sources**  :

- Faites glisser **Genre** dans le champ **Lignes**.
- Faites glisser **Formation** dans le champ **Colonnes**.
- Faites glisser **Valeur à vie pour le client** dans le champ **Densité**.

Exemples

Vous pouvez voir des exemples de visualisation dans l'exemple de rapport **Analyse des valeurs à vie pour le client**. Vous trouverez ces exemples ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Hiérarchie

Utilisez une hiérarchie si vous souhaitez afficher les données dans des lignes et des colonnes.

Par exemple, cette hiérarchie affiche les types de produit.

Binoculars
Climbing Accessories
Cooking Gear
Eyewear
First Aid
Golf Accessories
Insect Repellents
Irons
Knives
Lanterns
Navigation
Packs
Putters

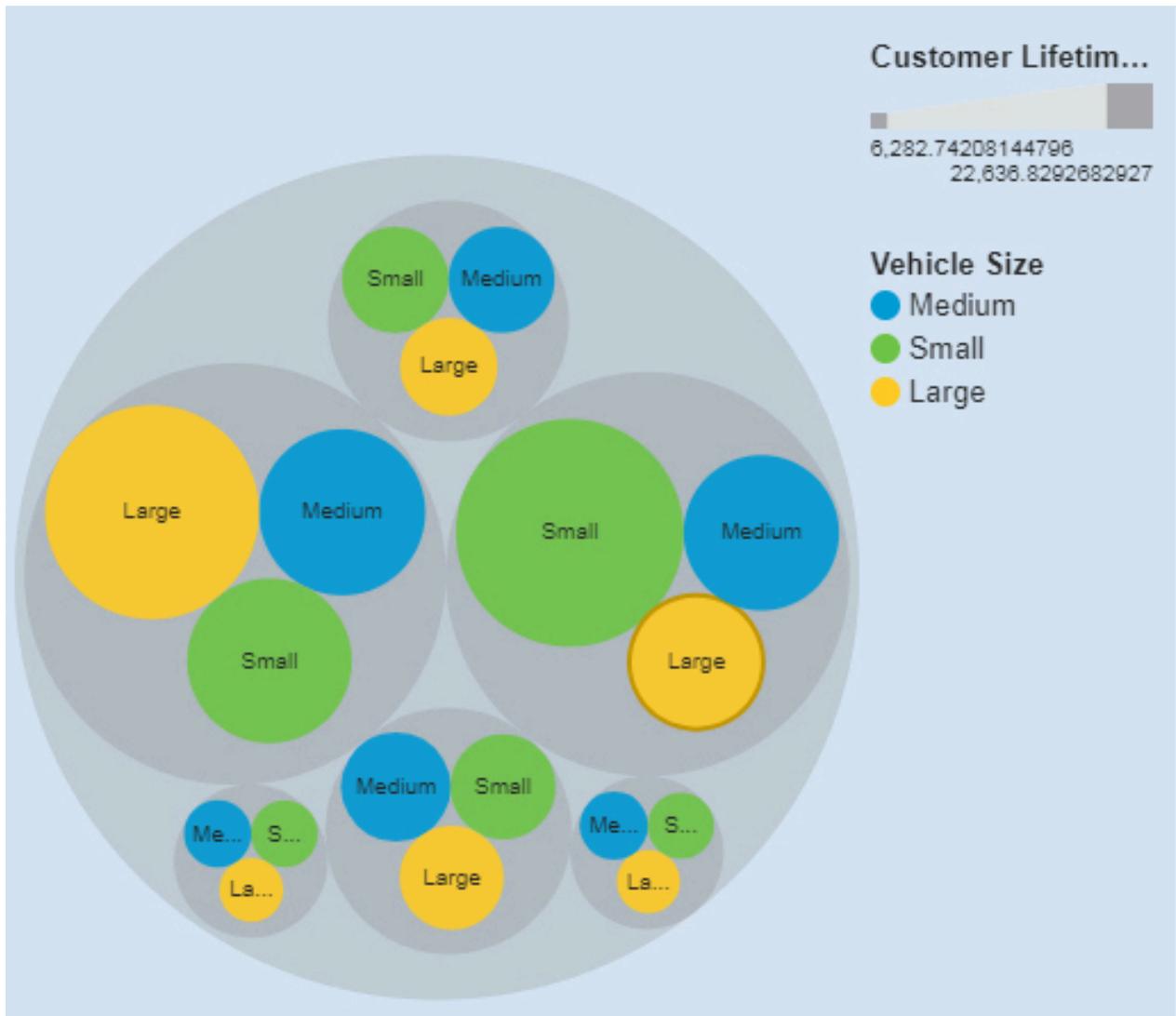
Bulle hiérarchique

Utilisez une visualisation à bulles hiérarchique pour représenter les relations entre des colonnes qui contiennent des valeurs, telles que les pertes nettes. Cette visualisation est comparable à la visualisation à bulles, mais les bulles sont étroitement rapprochées au lieu d'être réparties sur une grille. Les bulles utilisent l'imbrication pour représenter la hiérarchie. Une visualisation à bulles hiérarchique présente une grande quantité de données dans peu d'espace.

La taille de chaque bulle affiche une dimension quantitative de chaque point de données. Elle affiche de nombreux niveaux d'une hiérarchie, ainsi que les relations entre les groupes en fonction d'attributs affectés. Elle utilise la taille et la couleur des bulles pour fournir des informations comparatives sur des catégories.

Les bulles sont de différentes tailles et de différentes couleurs.

Par exemple, cette visualisation à paquet de bulles hiérarchique affiche la valeur à vie pour le client par classe de véhicule et taille de véhicule. Chaque bulle représente une classe de véhicule différente dans l'une des trois tailles de véhicule. La taille de chaque bulle est déterminée par la valeur à vie pour le client de cette classe de véhicule. Les couleurs des bulles sont déterminées par la taille de véhicule.



La visualisation à paquet de bulles hiérarchique a été créée en faisant glisser les données élémentaires suivantes à partir du panneau Sources :

- Faites glisser **Classe de véhicule** et **Classe de véhicule** dans le champ **Bulles**.
- Faites glisser **Valeur à vie pour le client** dans le champ **Taille**.
- Faites glisser **Taille de véhicule** dans le champ **Couleur**.

Exemples

Vous pouvez voir un exemple de visualisation en nuage de mots dans l'exemple de rapport **Analyse de la valeur de durée de vie du client**. Vous trouverez cet exemple ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Indicateur clé de performance

Utilisez une visualisation de type Indicateur clé de performance pour afficher un indicateur clé de performance qui contient deux mesures connexes, telles que les revenus réels et les revenus prévus. Vous pouvez éventuellement afficher un graphique sparkline et une forme significative dans les visualisations de type Indicateur clé de performance.

Une visualisation de type Indicateur clé de performance compare une valeur de base à une valeur cible et affiche la variance entre les deux mesures.

Par exemple, cette visualisation de type Indicateur clé de performance affiche les revenus réels en vert avec une flèche vers le haut pour indiquer que les revenus sont en hausse par rapport aux revenus cible. Dans ce cas, la valeur cible représente les revenus prévisionnels. Un graphique sparkline affiche la forme de la variation au fil du temps et sa couleur est la même que celle de la valeur de base.

TARGET_LINE compared to Revenue for
MonthsAsMember

\$228,762,440.00 ↑

Revenue

\$176,454,488.14 (+29.64%)

Planned Revenue



Créez une visualisation de type Indicateur clé de performance similaire en faisant glisser des mesures de votre propre source de données vers les champs d'une visualisation d'indicateur clé de performance vide :

1. Dans la barre d'outils, sélectionnez **Nouvelle carte**.
2. Sélectionnez la carte vide **Unique**.
3. Cliquez sur **Choisir un type** et sélectionnez le type de visualisation **Indicateur clé de performance**. Ajoutez ensuite des données élémentaires à chaque zone.
4. Faites glisser une mesure sur la zone **Valeur de base**. Cette valeur correspond à la cible réelle.
5. Faites glisser une mesure vers le champ **Valeur cible**.
6. Faites glisser une autre mesure vers le champ **Heure**. Cette valeur crée un graphique sparkline pour votre visualisation Indicateur clé de performance. Vous pouvez ajouter plusieurs mesures, par exemple Années et Mois, au champ **Heure**.

Utilisez les propriétés pour personnaliser une visualisation de type Indicateur clé de performance. Par exemple, les propriétés sont définies par défaut pour afficher la couleur conditionnelle verte lorsque la cible est satisfaite et la couleur conditionnelle rouge lorsque la cible n'est pas satisfaite. Pour afficher la cible réelle dans une autre couleur, développez la règle sous les propriétés, puis sélectionnez une autre **couleur de texte**.

Pour modifier une règle de couleur conditionnelle et sélectionner des couleurs personnalisées, procédez comme suit :

1. Sélectionnez la visualisation de type Indicateur clé de performance dans le panneau d'exploration.
2. Cliquez sur l'onglet **Propriétés**.
3. Dans la section **Règles**, développez la règle que vous souhaitez modifier.
4. Dans la section **Couleur du texte**, sélectionnez une couleur.

Les informations suivantes décrivent les propriétés de l'indicateur clé de performance dans la section **Style de règle** :

- **Couleur du texte**

Définissez la couleur de la valeur, du graphique sparkline et de la forme de l'indicateur.

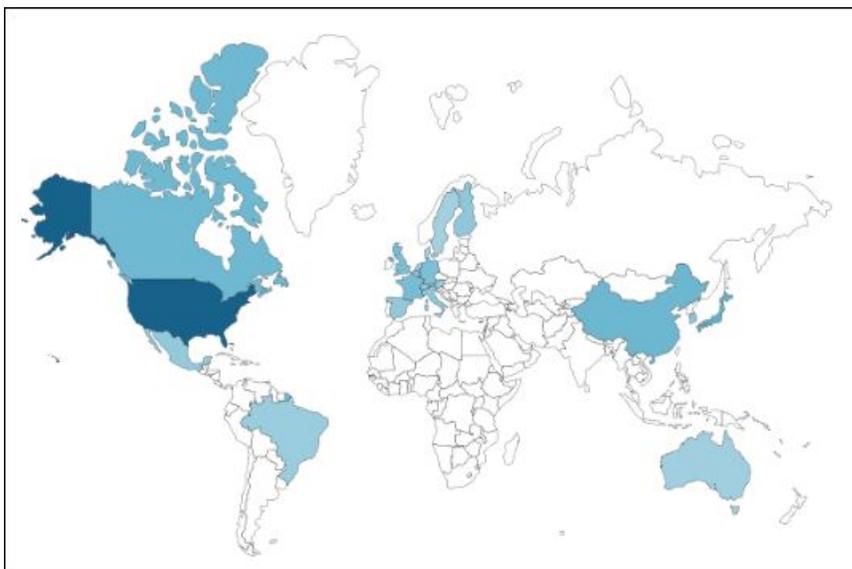
• Indicateur

Sélectionnez une forme à afficher dans la visualisation de type Indicateur clé de performance lorsque la règle est respectée. Par exemple, vous pouvez afficher une flèche vers le bas lorsque votre valeur de base est inférieure à un certain seuil par rapport à la valeur cible.

Carte existante

Utilisez une carte lorsque vous souhaitez afficher des motifs dans vos données par zone géographique. Vous pouvez utiliser une carte existante lorsque vous n'êtes pas connecté à Internet.

Par exemple, cette visualisation de carte existante affiche les revenus par pays avec une couleur plus foncée pour les revenus plus élevés.



Pour plus d'informations, consultez le document https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_ca_legacymaps.doc/ug_ca_legacymaps.pdf.

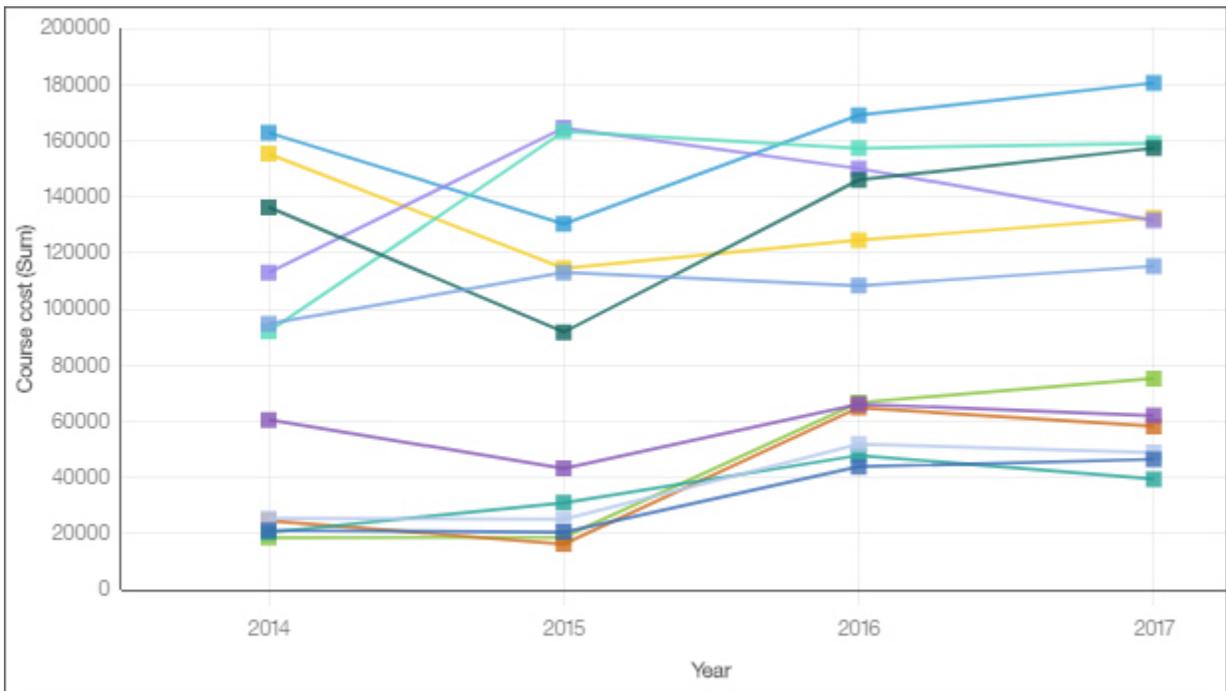
Courbes

Utilisez une visualisation à courbes pour montrer les tendances dans le temps.

Une visualisation à courbes peut comparer des tendances et des cycles, induire des relations entre des variables, ou représenter l'évolution d'une variable unique au fil du temps.

Pour une visualisation à courbes efficace, utilisez une colonne de temps (par exemple, années, trimestres ou jours) pour l'axe des X. Si l'axe des X contient autre chose (par exemple, Canada, Pays-Bas, Royaume-Uni et Etats-Unis), utilisez une visualisation à barres ou à colonnes.

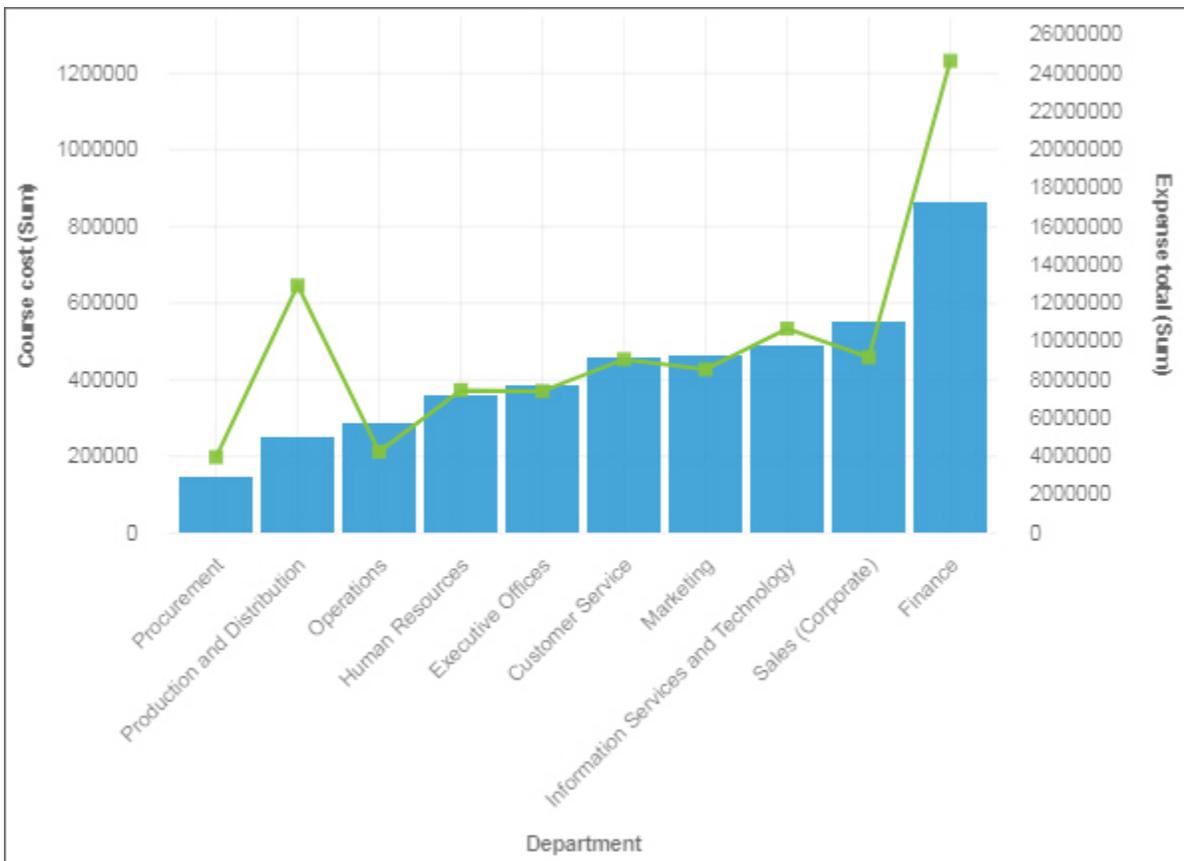
Par exemple, cette visualisation à courbes représente la tendance des coûts des cours par département sur l'année.



Courbes et colonnes

Utilisez une visualisation à courbes et à colonnes pour mettre en évidence les relations entre plusieurs séries de données en combinant des barres et des courbes dans une même visualisation.

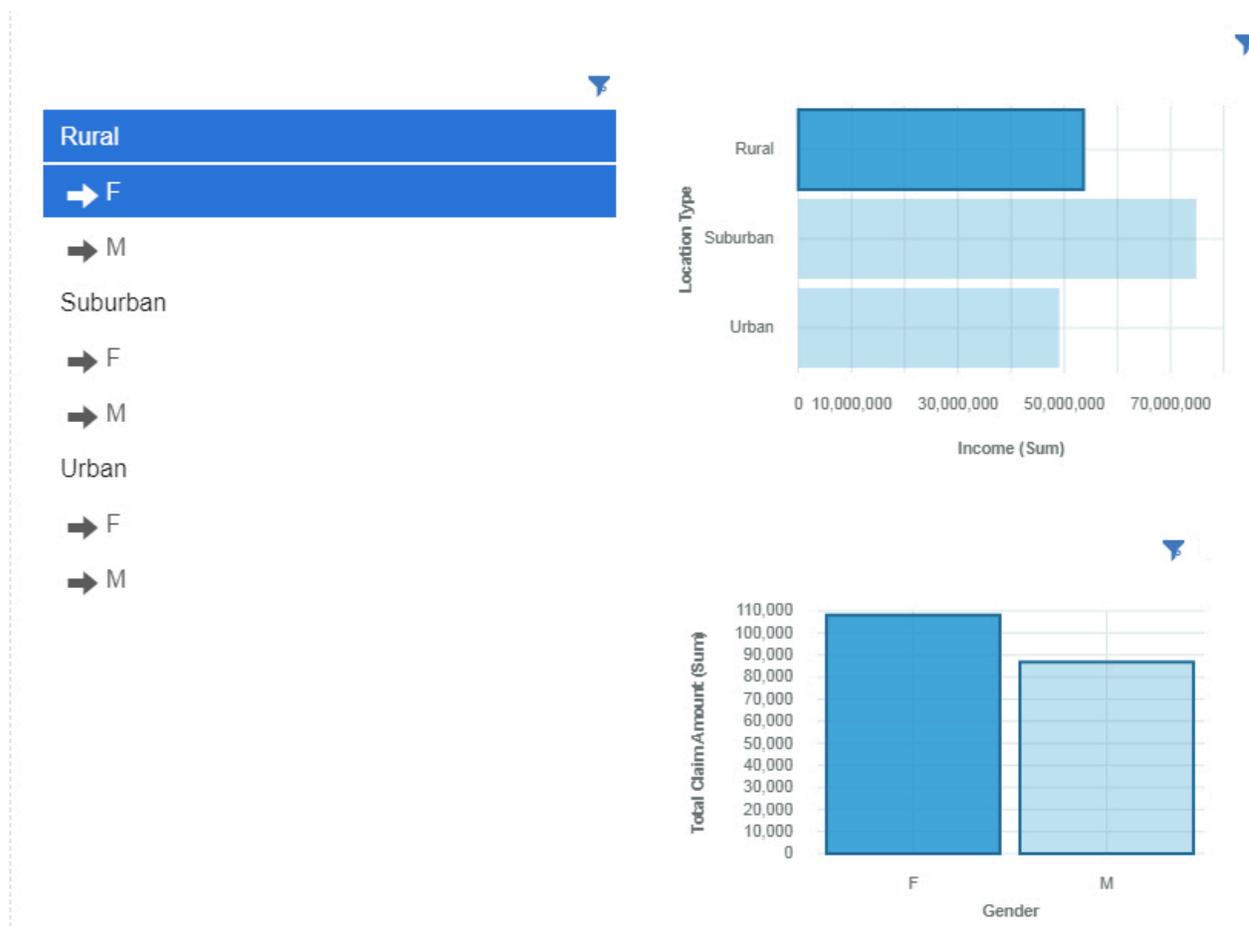
Par exemple, cette visualisation à courbes et à colonnes représente la relation entre le coût des cours et les dépenses totales par département.



Liste

Utilisez une visualisation de type liste pour créer une vue globale des données de manière hiérarchisée.

Une autre utilisation de la visualisation de type liste est la création d'un widget de filtre. L'exemple ci-dessous explique comment vous pouvez utiliser la visualisation à liste en tant que widget de filtre.



Carte

Utilisez une carte lorsque vous souhaitez afficher des motifs dans vos données par zone géographique.

Votre actif de données doit contenir des données géographiques, comme des pays, des états, des provinces ou des continents.

Remarque : Les cartes n'affichent pas les animations si vous réglez les paramètres système d'accès simplifié à cet effet.

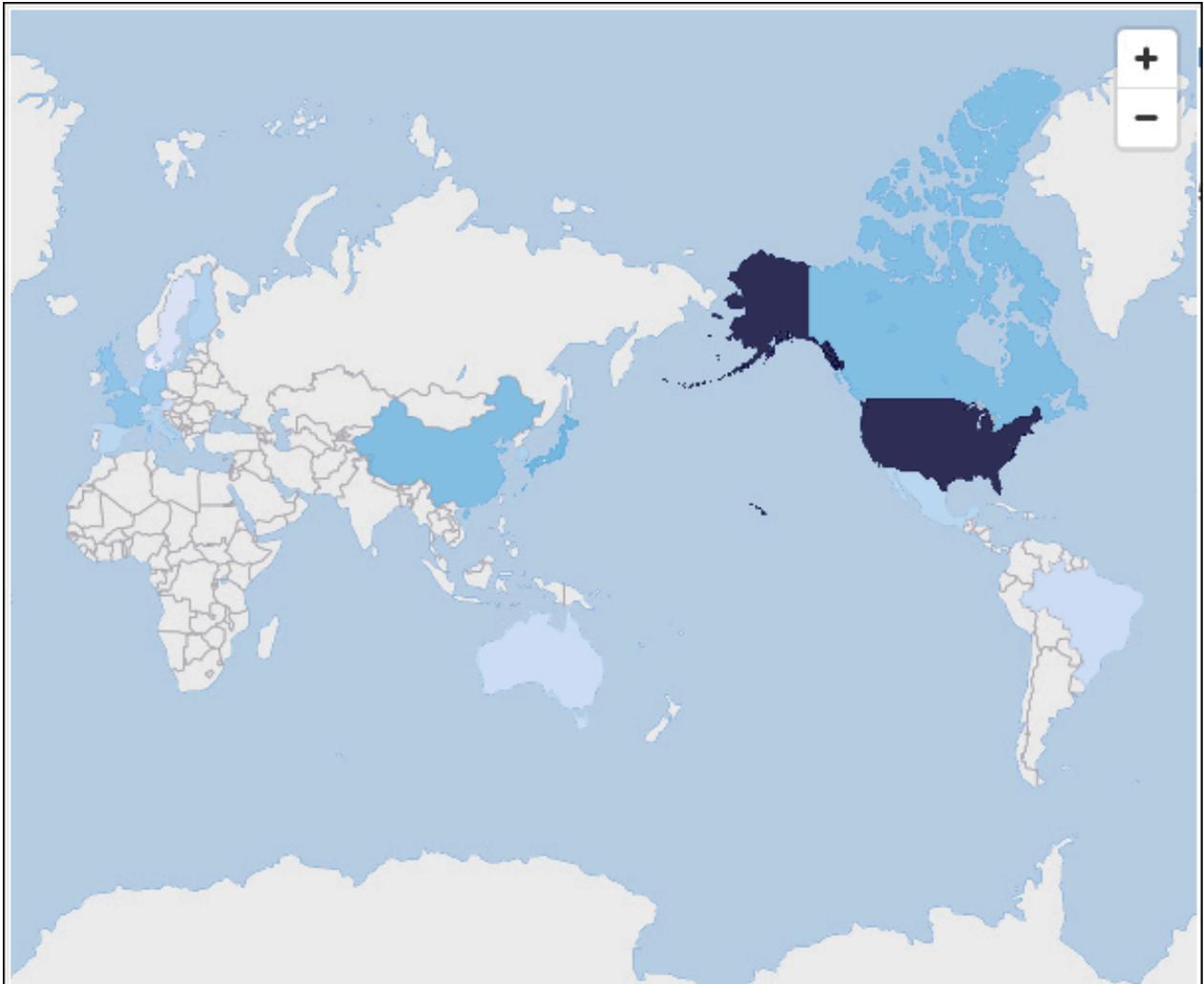
Dans Cognos Analytics, les cartes prennent en charge les continents suivants :

- Amérique du Nord
- Amérique du Sud
- Afrique
- Asie
- Europe
- Antarctique
- Océanie

Pour déterminer si une colonne peut être représentée sous forme de carte, Cognos Analytics analyse un échantillon de 2000 valeurs dans la colonne d'emplacement, à la recherche de noms de lieu reconnaissables. Si au moins 80 % des valeurs sont reconnues comme valeurs de carte, Cognos Analytics génère une carte.

Par exemple, votre colonne d'emplacement comporte quatre pays : Brésil, Chine, Inde et Russie. En raison de l'erreur typographique dans Inde, seules 75 % des valeurs correspondent à des noms de lieu reconnaissables ; vous ne pourrez donc pas afficher de carte comme point de départ. En revanche, si vous avez cinq pays et qu'un seul d'entre eux comporte une erreur typographique, une carte s'affiche.

Par exemple, cette visualisation de carte affiche les revenus par pays avec une couleur plus foncée pour les revenus plus élevés.



Marimekko

Une visualisation Marimekko est comparable à une visualisation à colonnes empilées. Elle affiche des données en utilisant différentes hauteurs et inclut une dimension de données supplémentaire en utilisant différentes largeurs de colonne. La largeur des colonnes est basée sur la valeur affectée au champ de largeur. La hauteur d'un segment spécifique est un pourcentage de la valeur totale de la colonne correspondante .

Vous pouvez identifier rapidement les segments de grande taille, comme une barre verticale spécifique qui représente une grande partie d'une région. Vous pouvez également identifier les blancs, comme une barre verticale sous-représentée dans une région spécifique.

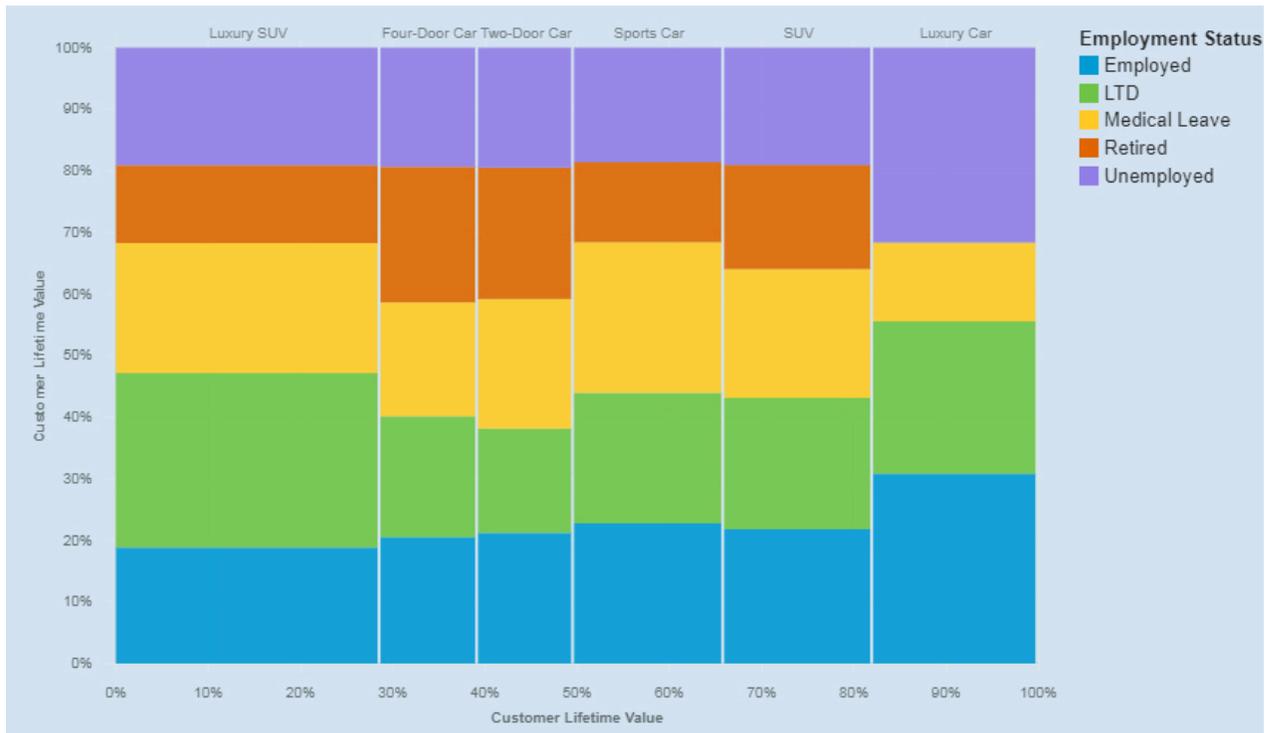
La visualisation Marimekko est utile pour comparer une partie à une totalité lorsque vous devez afficher une mesure/variable supplémentaire.

La visualisation Marimekko permet de représenter simultanément des données dans deux dimensions. Par exemple, des segments de marché sont souvent placés le long de l'axe des X et la largeur de chaque colonne correspond à la valeur financière d'un segment. Vous pouvez utiliser des visualisations

Marimekko pour représenter la contribution de chaque ligne de produits aux revenus ou le produit intérieur brut par pays.

La visualisation Marimekko peut afficher une valeur totale ou partielle. Si vous souhaitez utiliser des pourcentages empilés au lieu du nombre, utilisez l'option **Afficher sous forme de graphique empilé proportionnel avec pourcentages**.

L'exemple ci-dessous affiche la contribution à la valeur à vie pour le client et la situation professionnelle dans différentes classes de véhicules avec l'option **Afficher sous forme de graphique empilé proportionnel avec pourcentages** activée.



La visualisation Marimekko a été créée en faisant glisser les données élémentaires suivantes à partir du panneau Sources :

- Faites glisser **Classe de véhicule** dans le champ **Barres**.
- Faites glisser **Valeur à vie pour le client** dans le champ **Longueur**.
- Faites glisser **Situation professionnelle** dans le champ **Couleur**.

Exemples

Vous pouvez voir un exemple de visualisation en nuage de mots dans l'exemple de rapport **Analyse de la valeur de durée de vie du client**. Vous trouverez cet exemple ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

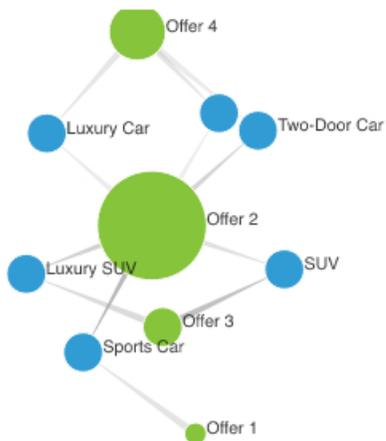
Réseau

Utilisez une visualisation en réseau si vous souhaitez voir les connexions entre les colonnes de votre actif de données. Une visualisation en réseau convient parfaitement pour afficher les connexions, les réseaux et les points d'intersection.

Les visualisations en réseau affichent un ensemble de noeuds représentés par des symboles, et des liens représentés par des chemins pour afficher la relation entre des entités ou des éléments.

Utilisez les champs **De** et **A** pour définir la relation que vous souhaitez étudier.

Par exemple, une visualisation en réseau peut représenter les acceptations d'offre par classe de véhicule.



Créez la visualisation en réseau en faisant glisser les données élémentaires suivantes à partir de la

section **Offres** de la sous-fenêtre **Sources**  :

- Faites glisser le champ **Offre** dans le champ **De**.
- Faites glisser **Classe de véhicule** dans le champ **A**.
- Faites glisser **Accepté** dans le champ **Épaisseur de ligne**.

Définissez ensuite les propriétés de taille et de noeud.

1. Cliquez sur la visualisation, puis sur . Définissez les options suivantes dans la sous-fenêtre **Propriétés** :

- **Taille - Largeur** : 500 px, **Hauteur** : 300 px
- **Taille minimale des noeuds** : 20
- **Taille maximale des noeuds** : 100

2. Cliquez sur  pour fermer la sous-fenêtre **Propriétés**.

Exemples

Vous pouvez voir des exemples de visualisation dans l'exemple de rapport **Analyse des valeurs à vie pour le client**. Vous trouverez ces exemples ici : **Contenu de l'équipe** > **Exemples** > **Rapports** > **Analyse des valeurs à vie pour le client**.

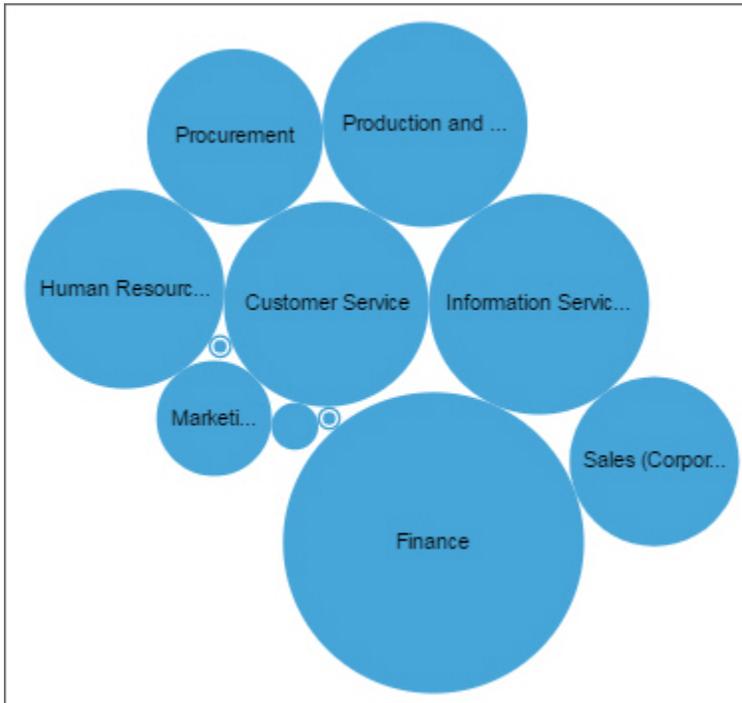
Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Paquets de bulles

Utilisez une visualisation en paquets de bulles pour représenter les relations entre des colonnes qui contiennent des valeurs numériques, telles que les revenus. Cette visualisation est comparable à la visualisation à bulles, mais les bulles sont étroitement rapprochées au lieu d'être réparties sur une grille. Une visualisation en paquets de bulles représente une grande quantité de données dans peu d'espace.

Les bulles sont de différentes tailles et de différentes couleurs.

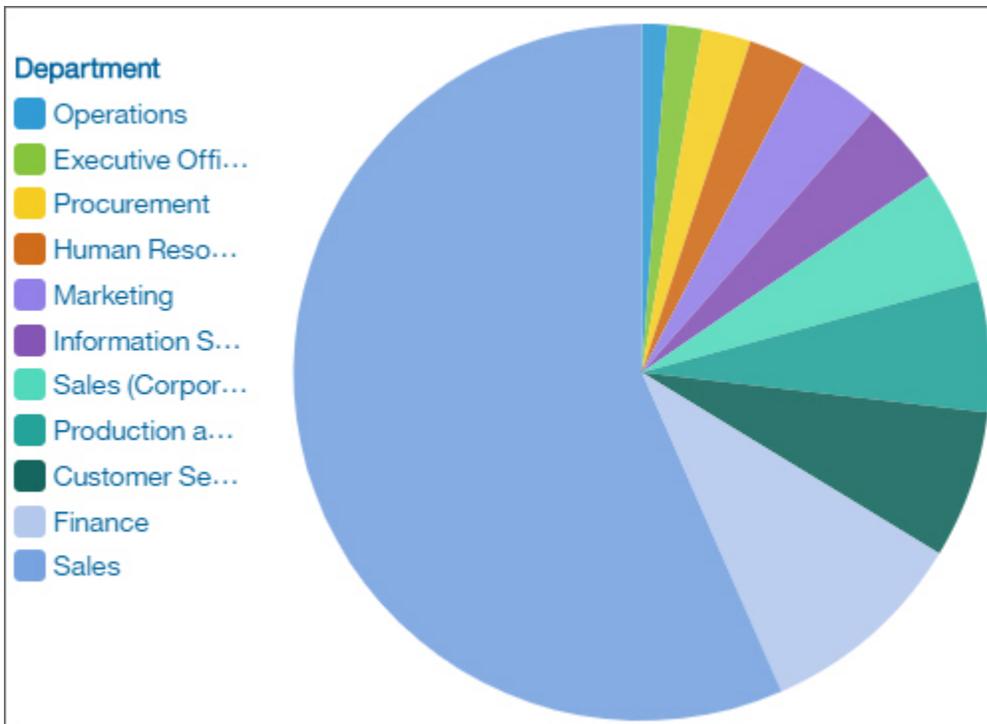
Par exemple, cette visualisation en paquets de bulles représente les employés externes par département. Chaque bulle représente un département différent. La taille de chaque bulle est déterminée par le nombre d'employés externes de ce département.



Graphique circulaire

Utilisez une visualisation de graphique circulaire pour mettre en évidence les proportions. Chaque secteur représente la relation relative de chaque partie par rapport au tout.

Par exemple, cette visualisation de graphique circulaire représente le nombre de jours de cours de chaque département.



Points

Utilisez une visualisation à l'aide de points pour afficher les tendances dans le temps.

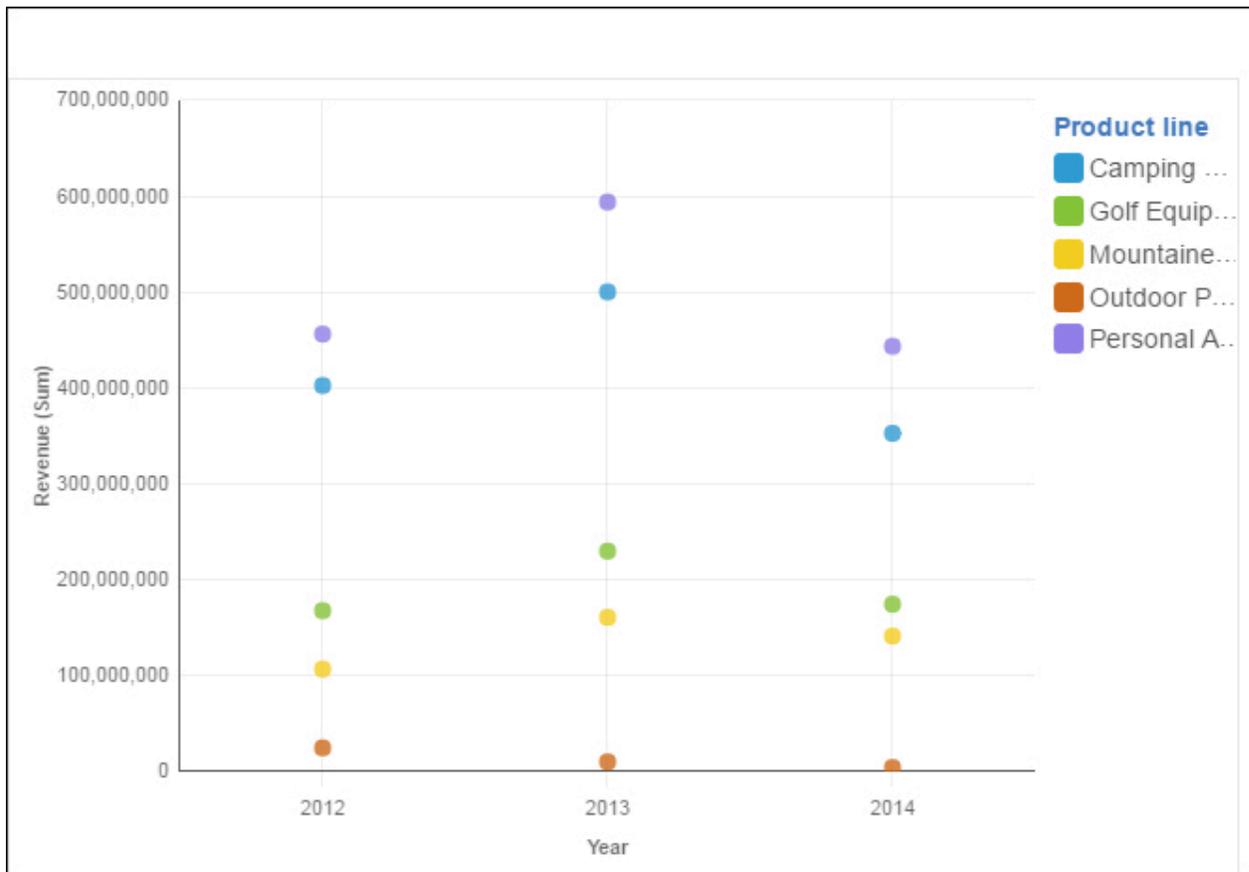
Une visualisation à l'aide de points peut comparer des tendances et des cycles, induire des relations entre des variables ou représenter l'évolution d'une variable unique au fil du temps.

Une visualisation à l'aide de points est comparable à un graphique à courbes sans lignes de connexion.

Pour une visualisation à courbes efficace, l'axe des X doit représenter le temps, par exemple des années, des trimestres, des mois ou des jours. Si l'axe des X contient autre chose (par exemple, Canada, Pays-Bas, Royaume-Uni et Etats-Unis), utilisez une visualisation à barres.

Les valeurs de données sont tracées verticalement.

Par exemple, cette visualisation à courbes représente les revenus trimestriels par type de mode de commande. Les commandes Web ont augmenté de manière spectaculaire au cours de cette période.

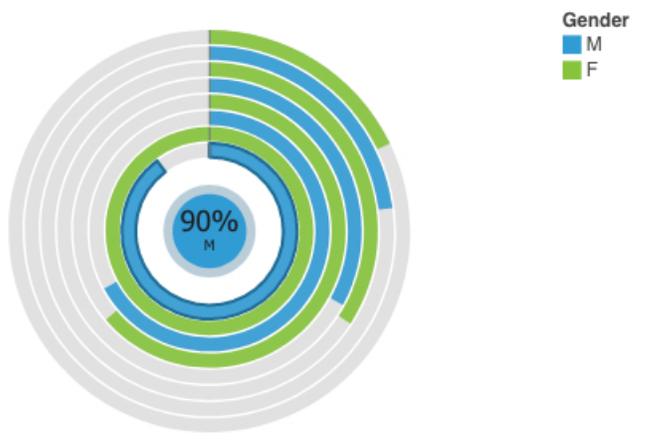


Radiale

Dans une visualisation radiale, chaque barre apparaît dans un cercle, les barres les plus longues représentant les valeurs les plus élevées. Survolez une barre pour en afficher les détails, tels que la valeur exacte représentée par la barre. Chaque barre commence à midi et s'étend dans le sens des aiguilles d'une montre pour les valeurs positives et dans le sens inverse des aiguilles d'une montre pour les valeurs négatives.

Les visualisations radiales, aussi appelées graphiques de type horloge ou compteur de vitesse, affichent des informations à l'aide d'aiguilles sur un cadran. La visualisation radiale n'est valide qu'avec une seule catégorie.

Par exemple, cette visualisation représente les renouvellements par type d'offre et sexe.



Créez la visualisation radiale en faisant glisser les données élémentaires suivantes à partir de la section

Analyse des clients de la sous-fenêtre **Sources**  :

- Faites glisser **Renouveler le type d'offre** dans le champ **Barres**.
- Faites glisser **Nombre de polices** dans le champ **Longueur**.
- Faites glisser **Genre** dans le champ **Couleur**.

L'étape suivante consiste à définir les propriétés de tri des données élémentaires **Renouveler le type d'offre** et **Genre**.

1. Cliquez sur la visualisation, puis dans la sous-fenêtre **Données**, cliquez sur la donnée élémentaire **<Renouveler le type d'offre>**.
2. Cliquez sur .
3. Dans la sous-fenêtre **Propriétés**, pour **Ordre de tri**, sélectionnez **Croissant**.
4. Dans la sous-fenêtre **Données**, cliquez sur la donnée élémentaire **<Genre>**.
5. Dans la sous-fenêtre **Propriétés**, pour **Ordre de tri**, sélectionnez **Décroissant**.
6. Cliquez sur  pour fermer la sous-fenêtre **Propriétés**.

Exemples

Vous pouvez voir des exemples de visualisation dans l'exemple de rapport **Analyse des valeurs à vie pour le client**. Vous trouverez ces exemples ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Nuage de points

Les visualisations en nuage de points utilisent des points de données pour tracer deux mesures n'importe où sur une échelle, et pas seulement aux marques de graduations régulières.

Les visualisations en nuage de points sont utiles pour explorer des corrélations entre différents ensembles de données.

L'exemple ci-dessous présente la corrélation entre les revenus et le bénéfice brut pour chaque type de produit.



Spirale

Une visualisation en spirale affiche les pilotes clés, ou prédicteurs, d'une cible donnée. Plus le facteur est proche du centre, plus il est fort.

IBM Cognos Analytics utilise des algorithmes extrêmement pointus pour obtenir des informations interprétables basées sur une modélisation complexe. Il est inutile de savoir quels tests statistiques doivent être exécutés sur vos données. Cognos Analytics sélectionne les tests appropriés pour les données.

Les pilotes clés pour les cibles continues et catégorielles sont disponibles dans l'analyse des pilotes et dans la visualisation en spirale dans les tableaux de bord et les explorations.

Pour plus d'informations, voir la section *Tests statistiques* dans le manuel *IBM Cognos Analytics - Guide d'utilisation des tableaux de bord et des histoires*.

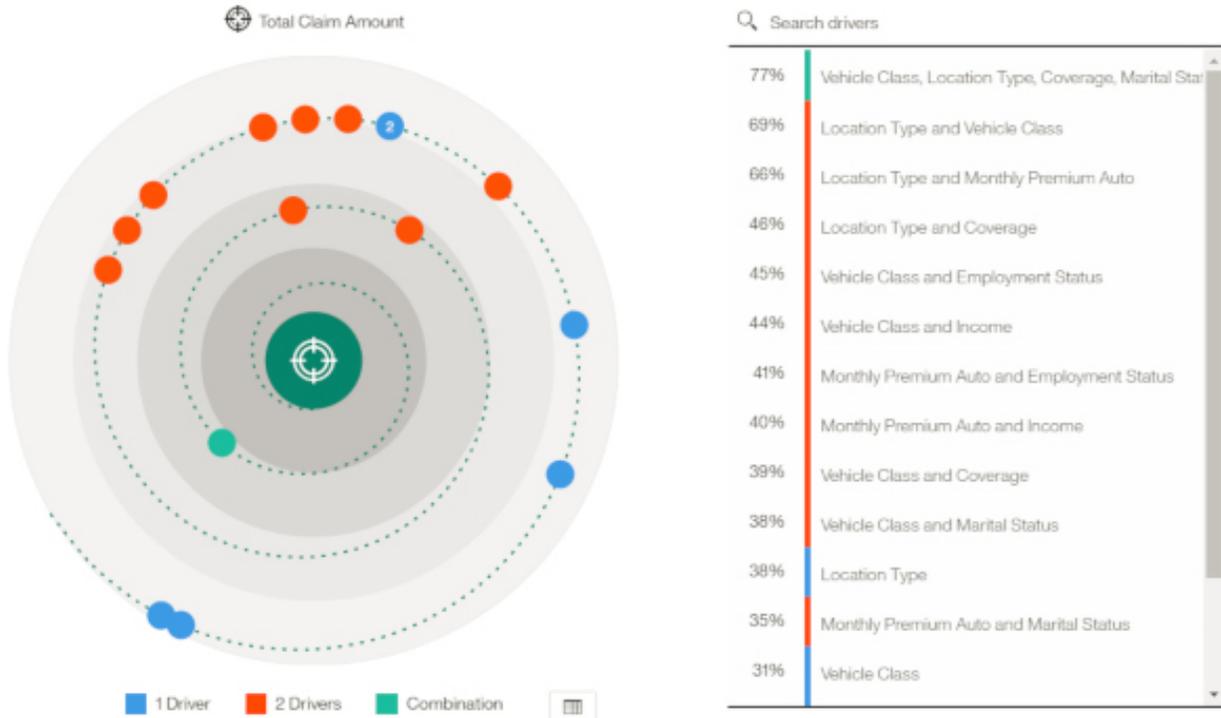
Par exemple, cette visualisation en spirale indique que la combinaison de la classe de véhicule, du type de site, du taux de couverture, du statut marital et de la situation professionnelle sont les pilotes les plus forts de la cible, à savoir le montant total du sinistre.

Vous pouvez exclure des pilotes de l'analyse. Cliquez sur un pilote à l'aide du bouton droit de la souris et

cliquez sur l'icône **Editer les pilotes** . Sélectionnez les pilotes que vous souhaitez inclure dans l'analyse.

Pour modifier ou ajouter des pilotes clés, cliquez sur l'icône  sur l'emplacement de données cible.

Compte tenu du nombre de lignes dans la source de données, l'analyse est basée sur un échantillon représentatif de la totalité des données pour améliorer les performances.

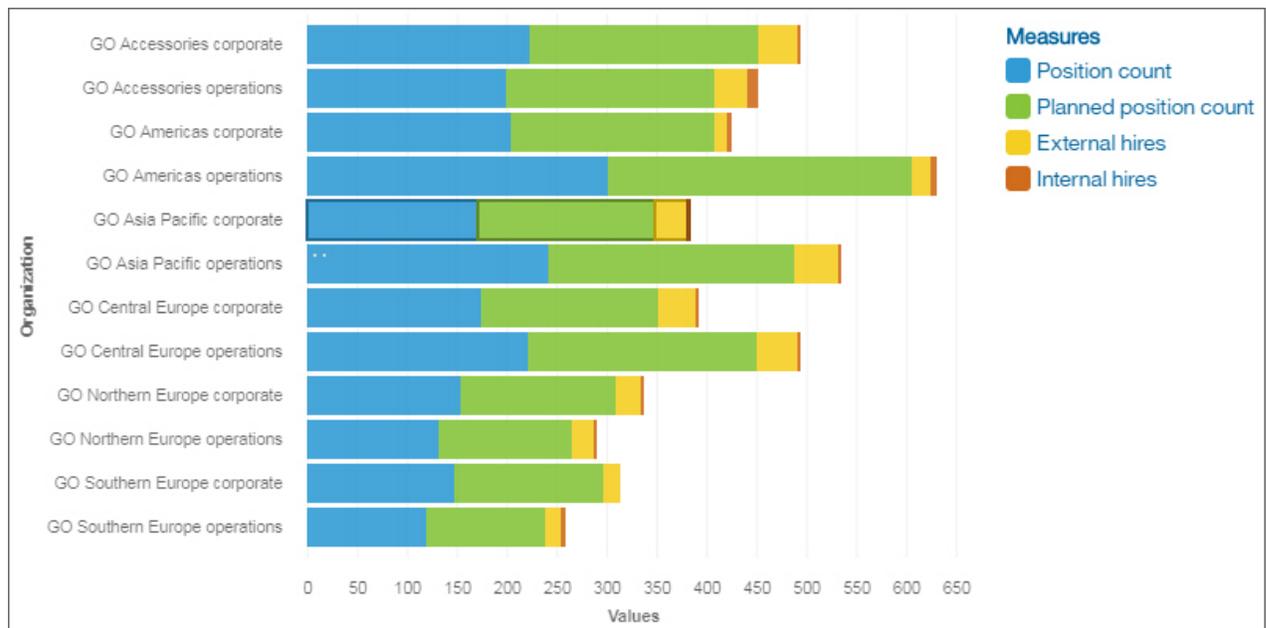


Remarque : Les filtres ne sont pas pris en charge pour les visualisations en spirale.

Barres superposées

Utilisez une visualisation à barres superposées pour comparer les contributions proportionnelles de chaque élément au total, telles que les ventes de produits et les ventes mensuelles de produits.

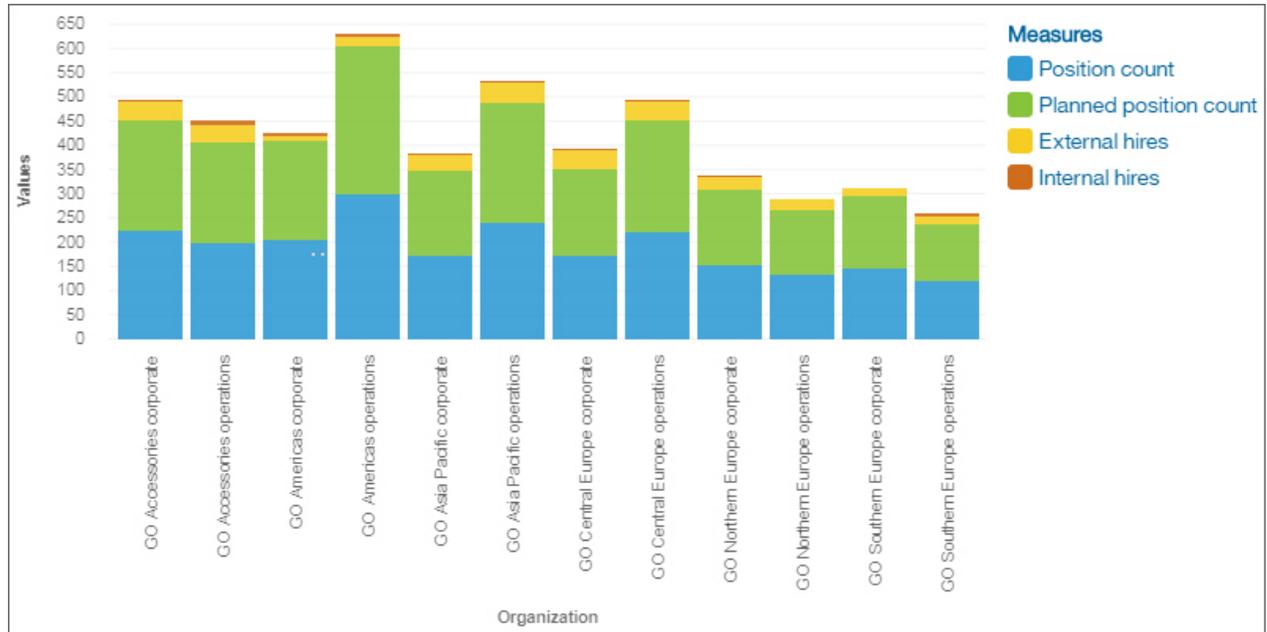
Une visualisation à barres superposées peut représenter les changements au cours d'une période spécifique ou comparer les contributions proportionnelles de chaque élément au total. Si le nombre de barres est si important que les libellés sont illisibles, filtrez les données pour vous concentrer sur un sous-ensemble de ces données ou utilisez une carte d'arborescence.



Colonnes empilées

Utilisez une visualisation à colonnes empilées pour comparer les contributions proportionnelles de chaque élément au total, telles que les ventes de produits et les ventes mensuelles de produits.

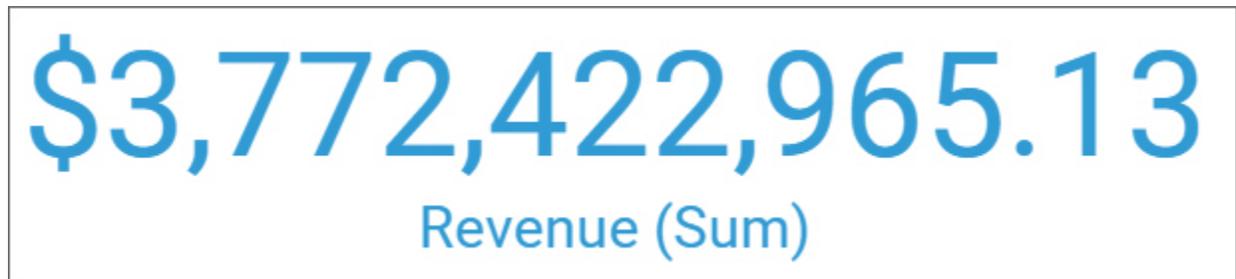
Une visualisation à colonnes empilées peut représenter les changements au cours d'une période spécifique ou comparer les contributions proportionnelles de chaque élément au total. Si le nombre de barres est si important que les libellés sont illisibles, filtrez les données pour vous concentrer sur un sous-ensemble de ces données ou utilisez une carte d'arborescence.



Récapitulatif

Utilisez une visualisation récapitulative si vous souhaitez afficher le total d'une mesure ou le nombre de résultats pour une colonne catégorique.

Par exemple, cette visualisation récapitulative représente les revenus totaux de tous les types de produit.



Par exemple, cette visualisation récapitulative représente le nombre de départements dans votre organisation.

11

Department (Count distinct)

Rayons

Une visualisation en rayons est utilisée pour expliquer comment des données sous-jacentes prévoient une cible choisie et mettent en évidence des informations d'analyse clés.

Pour plus d'informations sur la visualisation en rayons, voir [«Exploration d'une visualisation incluant un arbre de décisions»](#), à la page 19.

Tableau

Utilisez un tableau pour récapituler des informations détaillées de votre base de données, telles que des listes de produits et des listes de clients. Un tableau représente les données en lignes et en colonnes. Chaque colonne affiche toutes les valeurs associées à une donnée élémentaire de la base de données ou un calcul basé sur plusieurs de ces données élémentaires.

Par exemple, ce tableau récapitule le coût des cours de chaque département.

Department	Course cost
Customer Service	459,250
Executive Offices	384,000
Finance	863,750
Human Resources	361,250
Information Services and Technology	491,750
Marketing	465,750
Operations	288,000
Procurement	146,250
Production and Distribution	249,450
Sales	4,056,000
Sales (Corporate)	549,550
Summary	8,315,000

Ajout de colonnes supplémentaires à un tableau

Vous pouvez vous concentrer sur les points qui vous intéressent en ajoutant des données supplémentaires à la visualisation.

1. Faites glisser une autre colonne dans la zone où vous souhaitez placer des données supplémentaires.
2. Insérez-la à côté de la colonne existante.

A partir de Cognos Analytics version 11.1.4, vous pouvez faire glisser les données du panneau **Sources sélectionnées** et insérer les données dans une colonne/ligne ou faire glisser les données sur les données existantes pour les remplacer.

Carte d'arborescence

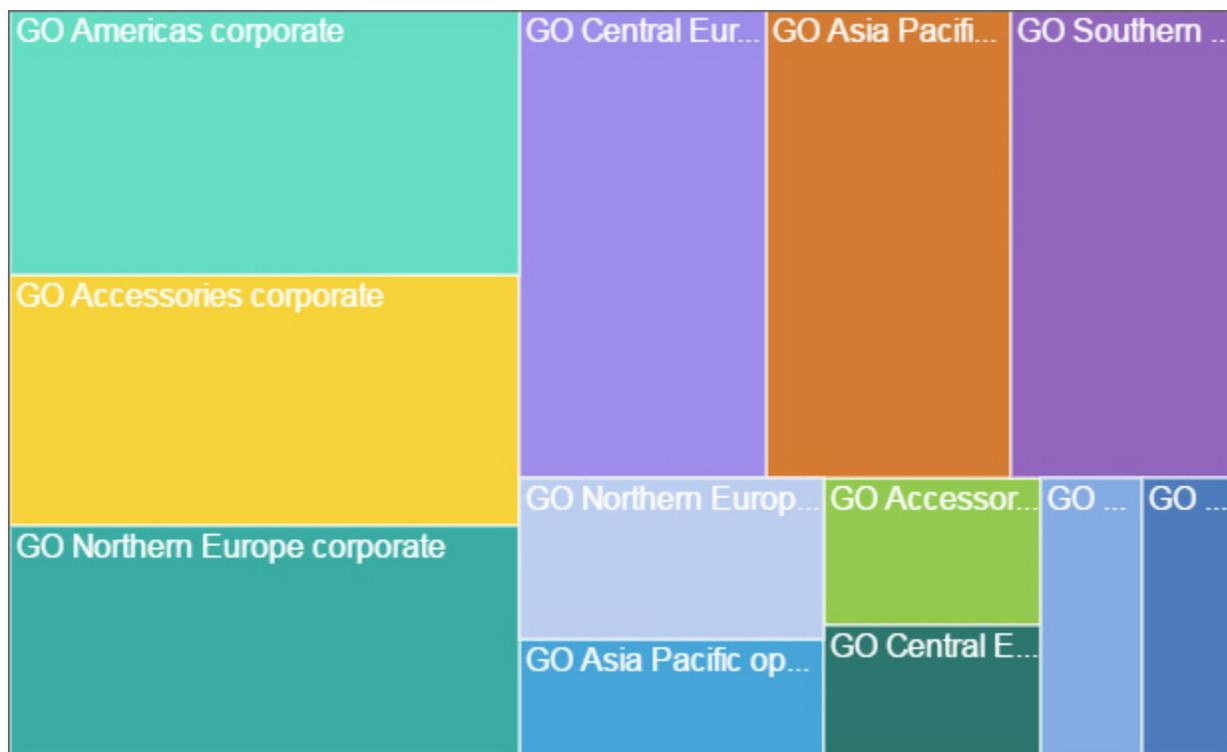
Utilisez une visualisation de carte d'arborescence pour identifier les tendances et les exceptions dans un actif de données volumineux et complexe.

Les cartes d'arborescence représentent les relations entre un grand nombre de composants à l'aide d'une codification utilisant la taille et la couleur dans un ensemble de rectangles imbriqués.

Une carte d'arborescence colorée par catégorie identifie la catégorie de niveau 1 par la couleur. Les tailles des rectangles représentent les valeurs. Dans une carte arborescente colorée par valeur, les tailles des rectangles représentent l'une des valeurs et la couleur représente un second ensemble de valeurs. N'utilisez pas de données incluant des nombres négatifs. Une carte d'arborescence ignore les nombres négatifs.

De nombreux actifs de données possèdent une structure hiérarchique. Vous disposez par exemple de données sur la marge bénéficiaire de produits alimentaires dans une épicerie. Sous la catégorie générale des fruits se trouve la catégorie agrume. Divers agrumes sont répertoriés, tels que le pamplemousse, l'orange et le citron. Une carte d'arborescence vous indique les performances de chaque agrume par rapport aux autres agrumes et aux autres types de produits alimentaires.

Par exemple, cette visualisation de carte d'arborescence représente le coût des cours par organisation.



Pour décocher une case que vous avez cochée, utilisez Ctrl+clic sur cette case.

Graphique en cascade

Utilisez une visualisation en cascade pour comprendre l'effet cumulatif d'une série de valeurs positives et négatives sur une valeur initiale. Les barres d'une visualisation en cascade ne sont pas des totaux.

Une visualisation en cascade montre la manière dont une valeur initiale est augmentée ou diminuée par une série de valeurs intermédiaires, pour obtenir une valeur cumulée finale indiquée dans la colonne tout à fait à droite. Les valeurs intermédiaires peuvent être basées sur le temps ou la catégorie.

Voici quelques exemples de visualisations en cascade :

- Consultation du résultat net après ajout des augmentations et diminutions de revenus et des coûts d'une entreprise au cours d'un trimestre.
- Ventes de produits cumulées sur un an avec total annuel.

Cette visualisation en cascade montre le delta du titulaire de police par mois.



Création d'une visualisation en cascade

1. Créez une exploration. Pour plus d'informations, consultez le document [«Lancement d'une nouvelle exploration à partir du menu Nouveau»](#), à la page 1.
2. Ouvrez l'exemple de module de données : **Sélectionner une source > Contenu de l'équipe > Exemples > Données > Analyse des clients.**
3. Cliquez sur **Visualisations**, puis sur **Cascade** pour ajouter la visualisation en cascade à l'exploration.
4. Cliquez sur **Sources**



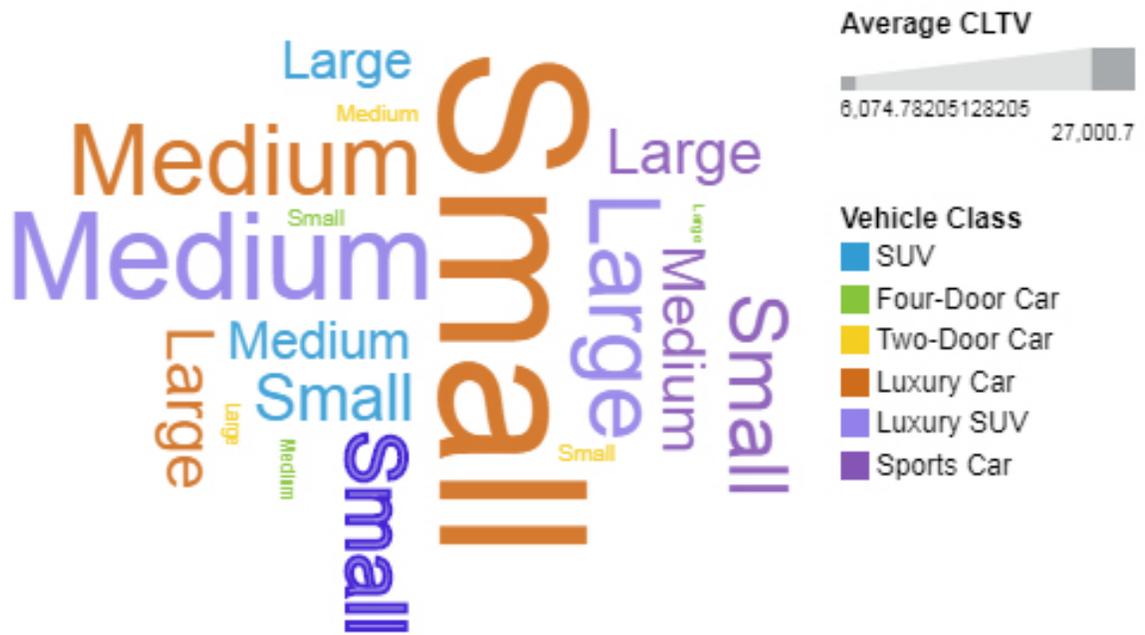
5. Faites glisser les données élémentaires suivantes depuis la section des titulaires de police :
 - Faites glisser **Mois** dans l'emplacement de données **Axe des X**.
 - Faites glisser **Delta** dans l'emplacement de données **Axe des Y**.

Nuage de mots

Utilisez une visualisation en nuage de mots si vous souhaitez afficher une visualisation textuelle d'une colonne. La hauteur du texte représente l'échelle. Le nom lui-même correspond aux différents membres de la colonne.

Conseil : L'actif de données doit contenir au moins 15 colonnes et 100 lignes pour qu'un nuage de mots efficace puisse être créé.

Par exemple, cette visualisation en nuage de mots représente la valeur vie client par taille et classe de véhicule.



Le nuage de mots a été créé en faisant glisser les données élémentaires suivantes à partir du panneau Sources :

- Faites glisser le type **Taille de véhicule** dans le champ **Mots**.
- Faites glisser **CLTV moyenne** dans le champ **Taille**.
- Faites glisser **Classe de véhicule** dans le champ **Couleur**.

Exemples

Vous pouvez voir un exemple de visualisation en nuage de mots dans l'exemple de rapport **Analyse de la valeur de durée de vie du client**. Vous trouverez cet exemple ici : **Contenu de l'équipe > Exemples > Rapports > Analyse des valeurs à vie pour le client**.

Si l'un des exemples d'objets est manquant, contactez votre administrateur.

Connaissances dans les visualisations

IBM Cognos Analytics fournit des informations analytiques qui vous aident à détecter et à valider des relations importantes et des différences pertinentes à partir des données présentées par la visualisation.

Ces connaissances sont disponibles en cliquant sur l'icône **Insights** ⓘ dans les visualisations éligibles. Lorsque vous activez la fonction Insights, le récapitulatif apparaît dans la section **Insights**, les éléments de visualisation connexes sont mis en évidence et des détails sont fournis dans le message d'infobulle correspondant. Vous pouvez contrôler chaque information disponible séparément.

Procédure

1. Dans une visualisation qui prend en charge la fonction Insights, cliquez sur l'icône **Insights** ⓘ.
2. En fonction de la visualisation, les connaissances suivantes sont affichées :
 - **Moyenne** : Indique la moyenne de la valeur cible affichée.
 - **Force prédictive** : Affiche la force prédictive de la relation entre la cible et les zones explicatives.
 - **Ligne d'ajustement** : Indique si une relation linéaire ou quadratique existe entre la cible et les champs explicatifs.

- **Différences significatives** : Affiche les valeurs qui sont nettement supérieures ou inférieures à la moyenne ou à la tendance.
- **Le plus fréquent** : Affiche les valeurs qui sont le plus fréquemment signalées.

Choix des connaissances corrélées

Sur la base de la visualisation, des connaissances corrélées, fondées sur des statistiques, vous sont présentées.

Pourquoi et quand exécuter cette tâche

Si des connaissances corrélées liées à la visualisation principale sont disponibles, une icône verte

associée à un nombre  apparaît sur l'axe des x ou l'axe des y. Ce nombre indique combien de connaissances corrélées sont disponibles.

Pour accéder aux visualisations corrélées, procédez comme suit :

Procédure

1. A partir de la visualisation, cliquez sur l'icône verte .
2. Cliquez sur l'une des connaissances fondées sur des statistiques qui sont présentées dans le menu. Une nouvelle carte est créée.

Choix des visualisations recommandées

Les visualisations recommandées sont des miniatures qui affichent les visualisations susceptibles d'être pertinentes en fonction de vos données.

Procédure

1. Dans la sous-fenêtre **Cartes**, sélectionnez la carte qui représente la visualisation que vous souhaitez ouvrir.
2. Cliquez sur **Choisir le type de visualisation**  dans la barre d'outils. Les visualisations recommandées s'affichent.
3. Cliquez sur la miniature correspondant à la visualisation recommandée que vous souhaitez utiliser.

Choix des visualisations liées

Lorsqu'une visualisation a le focus dans votre exploration, le système recommande des visualisations liées qui ne correspondent pas exactement à ce que vous avez demandé. L'analyse des données a cependant déterminé qu'elles pouvaient présenter de l'intérêt pour vous.

Pourquoi et quand exécuter cette tâche

Les visualisations liées remplacent l'un des éléments de données de la visualisation ou en ajoutent une autre pour créer une nouvelle visualisation. Elles utilisent une combinaison d'interactions utilisateur connues, de statistiques et de potentiel d'intérêt pour suggérer des étapes ultérieures utiles.

Pour accéder aux visualisations liées, procédez comme suit :

Procédure

1. Dans la sous-fenêtre **Cartes**, sélectionnez la carte qui représente la visualisation que vous souhaitez ouvrir.
2. Cliquez sur **Lié**  dans la barre d'outils.

Chapitre 4. Prévisions

Prévision

Utilisez la prévision dans la section IBM Cognos Analytics pour identifier et modéliser la tendance, la saisonnalité et la dépendance temporelle dans les données.

Pour générer des prévisions dans IBM Cognos Analytics, utilisez des outils automatisés qui modélisent les données dépendantes du temps. La sélection et l'optimisation d'un modèle automatisé facilitent l'utilisation des prévisions, même si vous ne connaissez pas bien la modélisation de séries temporelles.

Les prévisions et leurs bornes de confiance sont affichées dans les visualisations dans la continuité des données historiques. Vous pouvez également afficher les détails statistiques des modèles générés si vous souhaitez en savoir plus sur le contexte technique.

La spécification de séries temporelles dans les prévisions requiert souvent la manipulation des données. Cognos Analytics prend en charge un large éventail de séries temporelles sans qu'il soit nécessaire de manipuler les données. Ces séries temporelles incluent entre autres les types de date/heure standard et les zones de temps périodique et cyclique imbriquées. Lorsque les données sont reconnues en tant que série temporelle, elles sont préparées automatiquement. La tendance et les périodes saisonnières appropriées sont détectées et les modèles sont sélectionnés à partir d'un ensemble de neuf types de modèle différents.

Vous pouvez générer des prévisions dans des visualisations à courbes, barres et colonnes. Une prévision permet d'analyser des centaines de séries temporelles par visualisation. Les prévisions et les bornes de confiance sont calculées pour chaque série temporelle et affichées dans la visualisation en tant qu'extensions des données en cours. Vous pouvez inspecter chaque série temporelle séparément et adapter la prévision et les résultats à vos propres données et exigences.

Si les modèles de prévision vous sont familiers, vous pouvez afficher le type de modèle sélectionné, les paramètres de modèle estimés, les mesures de précision standard et les informations récapitulatives de traitement.

Fonctions de prévision

La fonction de prévision permet de modéliser les données de séries temporelles et fournit des prévisions basées sur les données des visualisations.

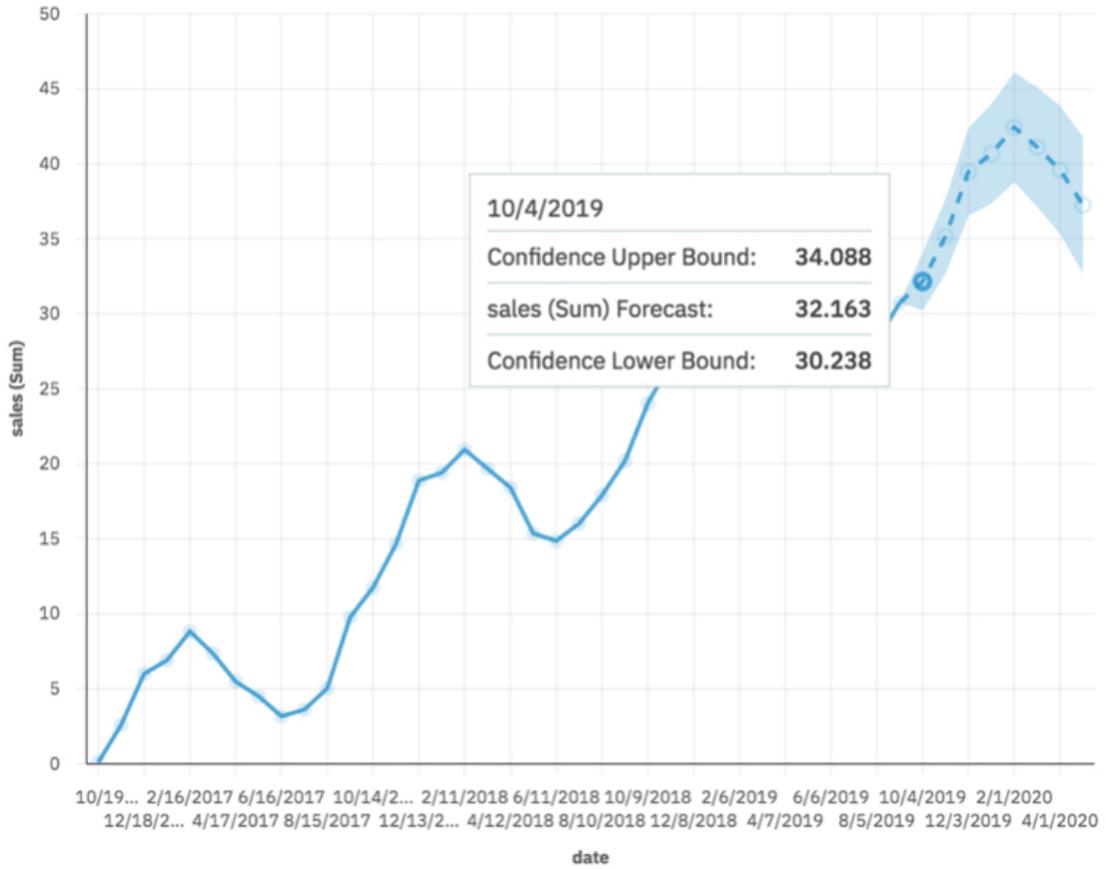
Pour utiliser la fonction de prévision, la visualisation doit être une visualisation à courbes, barres ou colonnes, les données doivent être prises en charge pour la prévision et la fonction de prévision doit être

activée. Lorsque la fonction de prévision est activée, la boîte de dialogue **Prévision**  s'affiche dans l'angle supérieur droit d'une visualisation pour vous permettre de modifier les paramètres de modèle et de prévision, ainsi que les bornes de confiance. Les modèles de série temporelle appropriés pour la visualisation sont estimés et les prévisions s'affichent dans la visualisation. Vous pouvez également visualiser la spécification du modèle de série temporelle et le récapitulatif du traitement des données dans le plateau de données.

L'exemple suivant présente des valeurs de prévision et les bornes de confiance dans une visualisation :

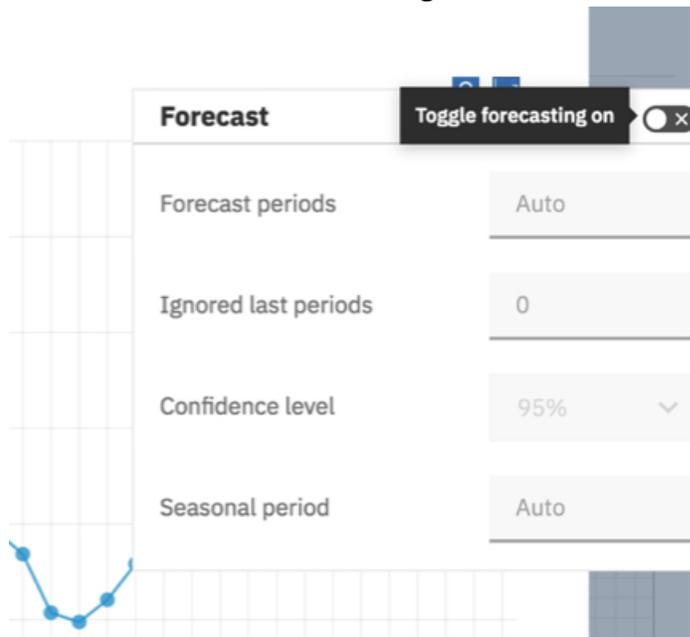


Forecast



Options de prévision

Vous pouvez modifier les prévisions en définissant un certain nombre d'options de période et de niveau de confiance dans la boîte de dialogue **Prévision** .



Une période correspond au plus petit intervalle de temps entre des points voisins dans les données.

Les options suivantes sont disponibles :

Périodes de prévision

Nombre d'étapes à prévoir.

La valeur par défaut est **Auto**, qui représente 20 % de la longueur des données historiques. Toute valeur manquante à la fin d'une série spécifique sera également prévue, mais elle ne sera pas comptabilisée dans le nombre spécifié de périodes de prévision.

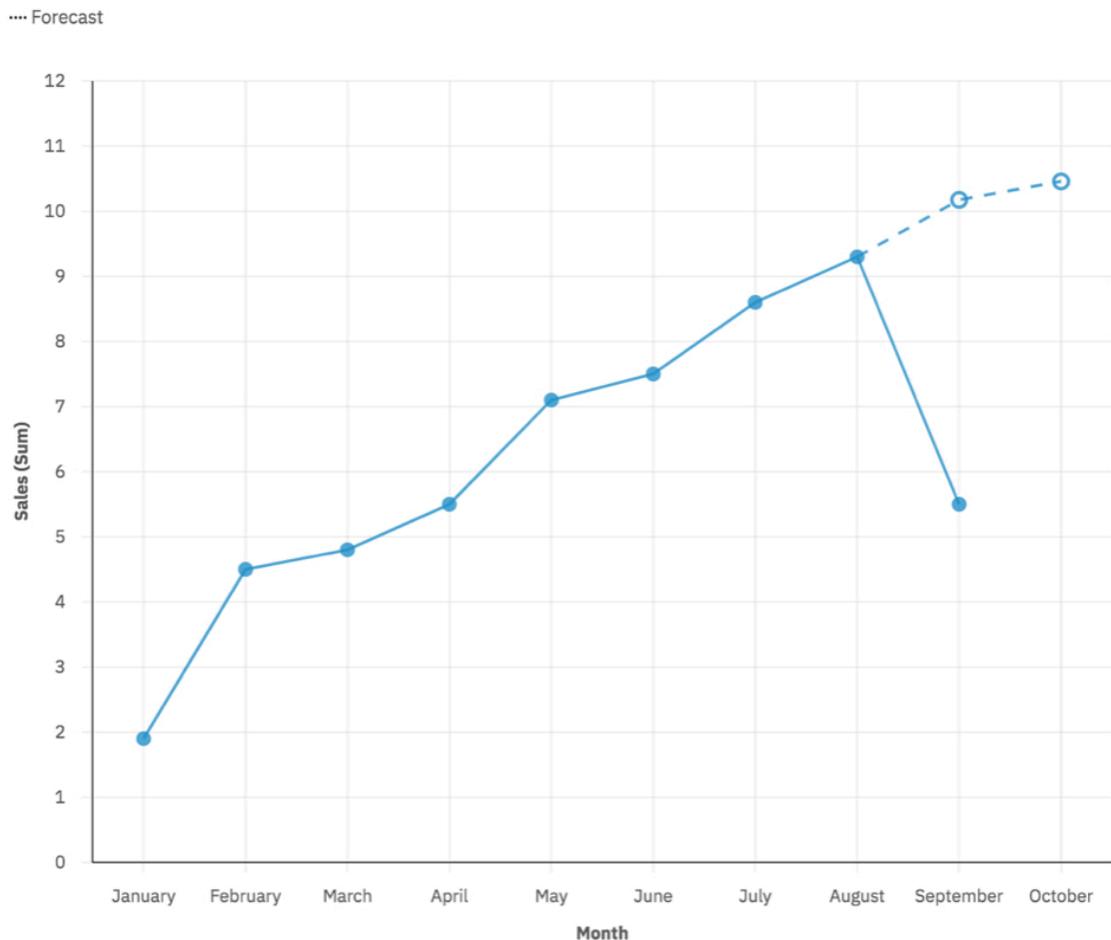
Dernière période ignorée

Ignore un nombre spécifié de points de données à la fin d'une série temporelle lors de la génération du modèle et du calcul des prévisions. Toute valeur manquante à la fin d'une partie non ignorée d'une série sera également prévue. La valeur de la dernière période ignorée doit être spécifiée sous la forme d'un entier non négatif, par exemple : 0, 1, 2, 3.

La valeur par défaut est 0. S'il n'y a pas de valeurs manquantes, toutes les données historiques sont utilisées dans la génération du modèle et le premier point de prévision se trouve après le dernier point de données historiques. Vous pouvez ignorer jusqu'à 100 points de données.

Ignorer la dernière période de données peut être utile lorsque les données sont incomplètes. Par exemple, vous pouvez effectuer une prévision à mi-chemin d'un mois. Excluez ce mois de la prévision en définissant **Dernières périodes ignorées** sur 1.

La visualisation suivante montre une prévision qui ignore les résultats du mois de septembre en définissant **Dernières périodes ignorées** sur 1.



Niveau de confiance

Certitude avec laquelle la valeur réelle doit se trouver dans la plage donnée. Vous pouvez afficher l'intervalle de confiance correspondant dans une infobulle en plaçant le pointeur de la souris sur une valeur de prévision. L'intervalle de confiance s'affiche sous la forme de bornes supérieure et inférieure.

Vous pouvez sélectionner trois niveaux de confiance différents : 90 %, 95 % et 99 %. La valeur par défaut est 95 % et les bornes inférieure et supérieure définissent la plage dans laquelle vous pouvez être sûr à 95 % que la valeur réelle se situe dans cette plage.

Période saisonnière

Saisonnalité avec laquelle générer le modèle. La saisonnalité correspond au moment où la série temporelle présente une variation cyclique prévisible. Par exemple, pendant une période de vacances chaque année.

La valeur par défaut est **Auto**. La valeur **Auto** détecte automatiquement la saisonnalité en générant plusieurs modèles avec des périodes saisonnières différentes et en choisissant la meilleure.

Vous pouvez spécifier une saisonnalité en saisissant un entier non négatif, par exemple 0, 1, 2, 3 comme période saisonnière.

Pour spécifier un modèle non saisonnier, associez **Période saisonnière** à 0 ou 1. Un modèle avec une saisonnalité définie par l'utilisateur s'affiche uniquement si le modèle saisonnier est plus précis que tous les modèles non saisonniers.

Connaissances

Lorsque les fonctions de connaissances et de prévisions sont toutes les deux activées et disponibles pour les visualisations, la boîte de dialogue Connaissances s'affiche en regard de la boîte de dialogue Prévision.

Chaque fonction fournit un ensemble indépendant de résultats d'analyse. Pour plus d'informations, voir «Connaissances dans les visualisations», à la page 67.

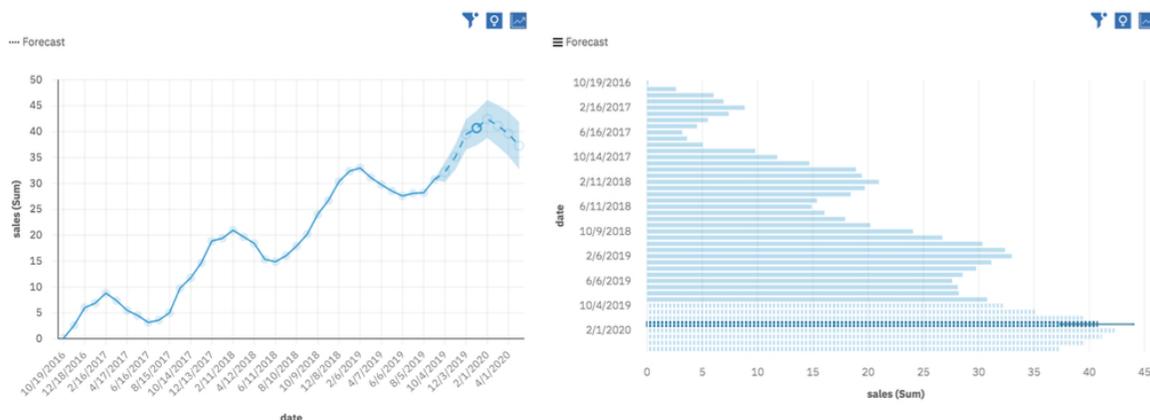
Types de visualisation prenant en charge les prévisions

Les prévisions sont prises en charge dans les visualisations à courbes, barres et colonnes.

Le tableau ci-dessous compare les fonctions d'affichage des prévisions pour chaque visualisation.

Fonctions de prévision	Graphique à courbes	Graphique à barres	Graphique à colonnes
Points de prévision	Cercle ouvert	Barre hachurée	Colonne hachurée
Affichage de l'intervalle de confiance	Région ombrée	Ligne pleine	Ligne pleine
Activer l'intervalle de confiance	Clic sur un point	Clic sur une barre de prévision	Clic sur une colonne de prévision
Nombre d'intervalles de confiance affichés	Tous	1	1

L'image suivante présente une prévision sur des visualisations à courbes et à barres avec les intervalles à barres activés :



Connaissances et prévisions

Les connaissances dans les visualisations fournissent des informations analytiques qui permettent aux utilisateurs d'identifier et de valider les relations importantes et les différences significatives à partir des données présentées dans la visualisation. Les connaissances viennent compléter les prévisions dans les visualisations prises en charge. Les connaissances fournissent un ensemble distinct de résultats d'analyse, et les résultats s'appliquent aux valeurs historiques uniquement. Pour plus d'informations, voir «Connaissances dans les visualisations», à la page 67.

Données de prévision

Les données acceptées pour la prévision incluent des valeurs de mesure correspondant à des points temporels disposés à intervalles réguliers. Vous spécifiez le temps et les mesures dans des visualisations en faisant glisser les champs de temps et les champs de mesure dans les emplacements de visualisation. Vous pouvez éventuellement spécifier des champs de groupe qui catégorisent les valeurs de mesure.

Le tableau suivant récapitule les types de champ et les emplacements de visualisation correspondants pris en charge dans la prévision :

Emplacement	Champs de temps (obligatoire)	Champs de mesure (obligatoire)	Champs de groupe (facultatif)
Emplacement du graphique à courbes	Axe des x	Axe des y	Couleur
Emplacement du graphique à courbes	Barres	Longueur	Couleur
Emplacement du graphique à colonnes	Barres	Longueur	Couleur

Aucune autre visualisation ni aucun autre emplacement de visualisation n'est pris en charge, à l'exception de l'emplacement **Filtres locaux**.

Champs de temps dans les données de prévision

Un champ de temps est identifié par une icône de temps placée devant le libellé du champ dans la sous-fenêtre **Données**.

Vous pouvez spécifier les propriétés d'un champ de temps en utilisant les propriétés **Type de données** ou **Représente Heure**.

Type de données

Un champ est reconnu comme un champ de temps s'il possède l'un des types de données suivants : date, heure ou horodatage. Le type de données est hérité de la source de données et ne peut pas être modifié.

Les types de données date, heure et horodatage sont conçus pour prendre en charge la gamme complète des formats de date et d'heure qui sont couverts par les formats de base et étendus de la norme ISO 8601. Le tableau suivant présente les types de données pris en charge et propose un exemple de format et un exemple de données pour chacun d'entre eux :

Type de données	Exemple de format	Exemple de données
Date	aaaa-mm-jj	2019-07-01
Heure	hh:mm:ss	12:34:56
Horodatage	aaaa-mm-jj'T'hh:mm:ss	2019-07-01T12:34:56

Représente Heure

Un champ est reconnu comme un champ de temps si la propriété de données **Représente** est définie sur **Heure**. Les champs Texte et Entier qui contiennent des données de temps sont également reconnus comme des champs de temps. Les champs de temps sont définis automatiquement lors de l'importation ou de l'enrichissement des données. Les définitions possibles sont Date, Année, Trimestre, Saison, Mois, Semaine, Jour, Heure, Minute ou Seconde.

Si les champs de temps ne sont pas reconnus automatiquement, vous pouvez les spécifier comme des champs de temps. Vérifiez que le format des valeurs de champ est pris en charge ; dans le cas contraire, vous risquez de recevoir une erreur liée à un format de date non pris en charge.

Champs de temps imbriqués

Vous pouvez faire glisser plusieurs champs de temps dans le même emplacement de visualisation pour spécifier un champ de temps imbriqué. Par exemple, un champ représentant une semaine peut être déplacé vers l'emplacement avec un champ représentant le jour pour créer une prévision par jours de la semaine.



Les champs imbriqués dans l'emplacement doivent être classés par ordre hiérarchique temporel. Par exemple, la semaine doit être placée au-dessus de la journée.

Les champs imbriqués ne peuvent pas ignorer les niveaux de la hiérarchie temporelle car ce serait source d'ambiguïté. Le tableau suivant décrit les hiérarchies acceptables :

Champ de temps	Champs inférieurs acceptables
Année	Trimestre, Mois, Semaine, Jour
Trimestre	Mois
Mois	Jour
Semaine	Jour
Jour (de l'année, du mois ou de la semaine)	Heure, Temps
Heure	Minute
Minute	Seconde

Si l'Année ne figure pas dans la hiérarchie temporelle, le système utilise par défaut l'année en cours. Cela peut générer des problèmes liés aux différences entre des années bissextiles et non bissextiles. Dans de tels cas, envisagez de fournir l'année.

Ordre chronologique des données

Les champs de temps spécifiés définissent l'ordre chronologique des points temporels dans la visualisation. Ils sont utilisés pour trier par ordre chronologique lorsque la prévision est activée. L'ordre chronologique inclut les points historiques, ainsi que les nouveaux points prévus. Tous les autres critères de tri spécifiés pour la visualisation sont ignorés lors de l'activation de la prévision.

Tous les libellés de temps non valides sont déplacés vers le début de la séquence et exclus de la génération du modèle et du calcul de la prévision.

Détection de l'intervalle de temps

Il est possible de détecter l'intervalle de temps lorsque les données sont classées par ordre chronologique. L'intervalle de temps correspond à la taille du plus petit intervalle entre deux points temporels adjacents, par exemple "2 semaines". Si des intervalles de temps variables sont détectés, ils doivent tous être des multiples d'entiers du plus petit intervalle. Sinon, les données sont considérées comme irrégulières et ne peuvent pas être prévues. Les points temporels manquants qui se présentent sous la forme d'un résultat de plusieurs intervalles sont remplis pour l'intervalle détecté. Les valeurs de mesure correspondantes sont définies comme manquantes. Si le nombre de valeurs manquantes est supérieur à 33 % de la longueur de la série, une erreur indiquant que trop de valeurs sont manquantes est signalée.

Champs de mesure

Un ou plusieurs champs de n'importe quel type peuvent être spécifiés en tant que champs de mesure pour l'analyse de prévision en les ajoutant à un emplacement de visualisation correspondant. Chaque champ de mesure est analysé séparément. Vous pouvez également spécifier plusieurs séries temporelles en ajoutant un champ à l'emplacement **Couleur**, ce qui fractionne les valeurs de mesure par les catégories du champ spécifié.

Toutes les valeurs de champ de mesure correspondant au même point temporel sont récapitulées à l'aide de l'un des niveaux de récapitulation suivants : **Somme**, **Minimum**, **Maximum**, **Moyenne**, **Nombre** et **Nombre (éléments distincts)**. Le champ doit être numérique pour prendre en charge le récapitulatif **Somme**, **Minimum**, **Maximum** ou **Moyenne**. Tous les types de données et tous les niveaux de récapitulation possibles sont pris en charge pour la prévision. Tenez toutefois compte des points suivants :

- Un petit nombre de valeurs de mesure différentes peut générer des prévisions inattendues ou sans intérêt. Par exemple, lorsque le récapitulatif **Nombre (éléments distincts)** est utilisé.
- Les valeurs de mesure nulles peuvent excessivement influencer les résultats, en particulier lorsqu'elles représentent des mesures manquantes.

Interpolation des valeurs manquantes

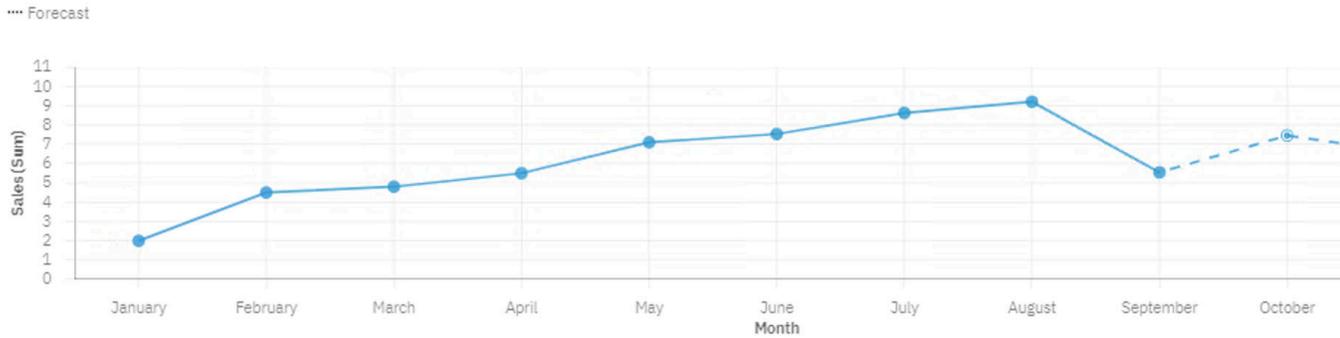
Les valeurs manquantes sont calculées et remplies par l'algorithme **Interpolation linéaire**. Le calcul se base sur les voisins les plus proches dans une série temporelle ordonnée chronologiquement, avec un intervalle de temps détecté. La nouvelle valeur est (valeur précédente + valeur suivante)/2. Par exemple, avec les valeurs [3, 6, valeur manquante, 12], la valeur interpolée pour remplacer la valeur manquante est $(6 + 12) / 2$ ou 9. L'algorithme d'interpolation peut également traiter les valeurs manquantes contiguës.

Les points de données dont les valeurs sont manquantes aux premiers ou aux derniers points temporels historiques sont exclus de la série avant la génération d'un modèle. Les valeurs manquantes aux derniers points temporels historiques sont également prévues.

Détails statistiques de prévision

Une exécution de prévision génère des prévisions et des détails statistiques de prévision. Les détails statistiques de prévision se trouvent dans le plateau de données au bas de chaque visualisation. Chaque série temporelle dans la visualisation contient une seule ligne de détails statistiques. Les détails de prévision sont générés tant que les points temporels sont espacés de manière égale.

Les informations de prévision contiennent des **Statuts** de prévision concernant la série temporelle donnée. Lorsque le statut est **Succès**, les autres zones fournissent des détails sur le modèle et les données utilisés pour la prévision. Lorsque le statut est **Echec**, certains des autres champs, y compris les **Remarques**, fournissent des détails sur la cause de l'échec. Les récapitulatifs des échecs sont toujours fournis dans les avertissements de visualisation.



Data tray

Visualization data incomplete_last_point.csv **Forecasting statistical details**

Model

Sales	Trend	Seasonality	Accuracy %
Sales	Additive	None	31.1%

Les informations de **Modèle** spécifient le type de **Tendance** et de **Saisonnalité** sélectionné pour l'estimation des données de séries temporelles en cas de succès. Le tableau suivant répertorie les différents types disponibles.

Composante de tendance	Composante saisonnière		
	N NONE	A ADDITIVE	M MULTIPLICATIVE
N NONE	(N, N)	(N, A)	(N, M)
A ADDITIVE	(A, N)	(A, A)	(A, M)
Ad ADDITIVE_DAMPED	(Ad, N)	(Ad, A)	(Ad, M)

Mesures de précision

La précision du modèle mesure l'erreur absolue moyenne (MAE, Mean Absolute Error), l'erreur moyenne absolue échelonnée (MASE, Mean Absolute Scaled Error), le pourcentage de précision, la racine de l'erreur quadratique moyenne (RMSE, Root Mean Squared Error), le pourcentage d'erreur absolu moyen (MAPE, Mean Absolute Percent Error) et le critère d'information d'Akaike (AIC, Akaike Information Criterion), qui se basent sur les données de séries temporelles utilisées pour générer le modèle. Toutes les mesures de précision se basent sur les données historiques. Les mesures de précision peuvent également servir d'indicateur de l'exactitude des prévisions, mais elles ne portent pas sur les valeurs futures.

Erreur absolue moyenne (MAE)

Correspond à la différence absolue moyenne entre les valeurs ajustées par le modèle (prévision dans le modèle avec une étape d'avance) et les données historiques observées.

Erreur moyenne absolue échelonnée (MASE, Mean Absolute Scaled Error)

Mesure d'erreur utilisée pour la précision du modèle. Il s'agit de la MAE divisée par la MAE du modèle naïf. Le modèle naïf prévoit la valeur à un point temporel t comme la valeur historique précédente. La mise à l'échelle selon cette erreur signifie que vous pouvez évaluer la qualité du modèle par rapport au modèle naïf. Si la MASE est supérieure à 1, le modèle est pire que le modèle naïf. Plus la MASE est faible, plus le modèle est qualitatif par rapport au modèle naïf.

Pourcentage de précision (% précision)

Indicateur principal de la précision du modèle basé sur les valeurs ajustées. Cet indicateur est spécifié comme le pourcentage de réduction de l'erreur absolue moyenne par rapport au modèle naïf. Il est calculé en soustrayant MASE de 1 et en l'exprimant sous forme de pourcentage. Si la MASE est supérieure ou égale à 1, la précision est définie à 0 % car le modèle ne s'améliore pas par rapport au modèle naïf. Une précision plus élevée indique une erreur de modèle plus faible par rapport au modèle naïf.

Erreur quadratique moyenne (MSE)

Somme du carré des différences entre les valeurs ajustées par le modèle et les valeurs observées qui sont divisées par le nombre de points historiques, moins le nombre de paramètres dans le modèle. Le nombre de paramètres dans le modèle est soustrait du nombre de points historiques pour être cohérent avec une estimation non biaisée de la variance du modèle.

Racine de l'erreur quadratique moyenne (RMSE, Root Mean Squared Error)

Racine carrée de l'erreur quadratique moyenne (MSE). Elle se trouve sur la même échelle que les valeurs des données observées.

Pourcentage d'erreur absolu moyen (MAPE, Mean Absolute Percent Error)

Pourcentage absolu moyen de différence entre les valeurs ajustées par le modèle et les valeurs des données observées.

Critère d'information d'Akaike (AIC) :

Mesure de sélection de modèle. L'AIC pénalise les modèles comportant de nombreux paramètres et tente ainsi de choisir le meilleur modèle en privilégiant les modèles plus simples. L'AIC correspond à la somme du logarithme de la MSE non ajustée, multiplié par le nombre de points historiques, le nombre de paramètres de modèle et les états de lissage initiaux multipliés par 2.

Paramètres

Période saisonnière détectée et estimations des autres paramètres utilisés dans le modèle de lissage exponentiel sélectionné.

Période saisonnière

Nombre d'étapes temporelles dans une période saisonnière utilisée dans le modèle de lissage exponentiel.

Alpha

Facteur de lissage pour les états de niveau du modèle de lissage exponentiel. Les valeurs faibles du paramètre alpha augmentent le volume de lissage. En d'autres termes, une quantité d'historique plus importante est prise en compte lorsque la valeur de alpha est faible. Les valeurs élevées du paramètre alpha réduisent le volume de lissage, ce qui signifie que davantage de pondération est appliquée aux observations plus récentes. Lorsque la valeur est 1, toute la pondération est appliquée à l'observation en cours.

Beta

Facteur de lissage pour les états de tendance du modèle de lissage exponentiel. Ce paramètre se comporte de la même manière que le paramètre alpha, à la différence près qu'il s'applique à la tendance et non aux états de niveau.

Gamma

Facteur de lissage pour les états de saisonnalité du modèle de lissage exponentiel. Son rôle est similaire au paramètre alpha, à la différence près qu'il s'applique à la composante saisonnière du modèle.

Phi

Coefficient d'amortissement du modèle de lissage exponentiel. Les prévisions sur le long terme pouvant générer des résultats non réalistes, il s'avère utile d'utiliser un facteur d'amortissement pour freiner la tendance au fil du temps et générer des prévisions plus prudentes.

Diagnostic

Ces informations incluent le Nombre manquant, la Longueur de série, les Périodes ignorées, le Niveau de tendance, le Niveau de saisonnalité et l'Intervalle de date/heure.

Nombre manquant

Indique le nombre de lignes de données qui comportent des valeurs manquantes ou des points temporels manquants et qui sont positionnées entre la première et la dernière valeur de série valide. Les points temporels non valides, ainsi que les points dont les valeurs sont manquantes aux premiers ou aux derniers points temporels historiques ne sont pas inclus.

Longueur de série

Indique le nombre de points de données utilisés pour la modélisation des séries temporelles. Seuls les points entre la première et la dernière valeur de série valide sont inclus.

Périodes ignorées

Entier, m , qui ignore les m derniers points de données de la série lors de la génération du modèle de lissage exponentiel et du calcul des prévisions. Toute valeur manquante à la fin d'une partie non ignorée d'une série sera également prévue. La valeur par défaut de ce paramètre est 0, ce qui signifie que toutes les données historiques sont utilisées dans la génération du modèle lorsqu'il n'y a pas de valeur manquante. Vous pouvez ignorer jusqu'à 100 points de données potentiels. Les périodes ignorées excluent les points de données lors de la génération d'un modèle, de sorte que la prévision risque d'échouer en raison de facteurs tels que les exigences minimales de longueur des données et la proportion de valeurs manquantes qui dépasse 33 %.

Niveau de tendance

Compare le modèle d'origine, M , au même modèle mais en retirant la composante de tendance. Le niveau de tendance de M correspond à la différence de précision entre le modèle M et le modèle M sans composante de tendance.

Niveau de saisonnalité

Compare le modèle d'origine, M , au même modèle en retirant la composante saisonnière. Le niveau de saisonnalité de M correspond à la différence de précision entre le modèle M et le modèle M sans composante saisonnière.

Intervalle de date/heure

L'intervalle de date/heure représente l'intervalle de temps détecté des données triées par ordre chronologique. Il est identifié comme la plus petite différence entre les points voisins dans les données lorsqu'elles sont triées dans l'ordre chronologique.

Modèles de prévision

Les modèles de lissage exponentiel représentent une classe de modèles de série temporelle couramment utilisée.

Les modèles de lissage exponentiel sont applicables à un ensemble unique de valeurs qui sont enregistrées sur des incréments de temps égaux uniquement. Ils prennent toutefois en charge les propriétés de données qui sont fréquemment utilisées dans les applications métier, telles que la tendance, la saisonnalité et la dépendance temporelle. Toutes les caractéristiques de modèle spécifiées sont estimées en fonction des données observées disponibles. Un modèle estimé peut ensuite être utilisé pour prévoir les futures valeurs et fournir des bornes de confiance supérieures et inférieures pour les valeurs de prévision.

Chaque type de modèle convient à la modélisation d'une combinaison différente de propriétés trouvées dans les données. Le type de modèle qui peut fournir la meilleure correspondance avec les données observées est sélectionné pour modéliser les données observées et est utilisé pour prévoir les futures valeurs.

Algorithmes d'estimation de modèle

Les modèles sont spécifiés par les équations de lissage qui incluent les paramètres de modèle et les états de lissage initiaux. Les paramètres du modèle sont estimés avec des valeurs qui réduisent au minimum l'erreur du modèle.

Equations de lissage

Les modèles de lissage exponentiel tirent leur nom des équations de lissage qui spécifient le modèle. Ils fournissent des formules permettant de calculer les états de lissage pour chaque point observé en utilisant la valeur observée actuelle et les états de lissage précédents. Les équations de lissage fournissent des moyennes pondérées de la valeur actuelle et les états précédents de la série temporelle. La pondération de la valeur ou de l'état en cours est donnée par un paramètre de modèle compris entre 0 et 1, alors que les pondérations des valeurs précédentes diminuent de façon exponentielle.

Equations de lissage de niveau

Tous les types de modèle calculent un état de niveau pour chaque point de série temporelle en utilisant l'équation de lissage de niveau correspondante. Les états de niveau pour le modèle sans tendance ni composante saisonnière sont calculés comme la moyenne pondérée de la valeur de série temporelle au point actuel et de l'état de niveau au point précédent. La pondération associée à la valeur actuelle est un paramètre (**alpha**) dont la valeur est comprise entre 0 et 1. Pour les autres modèles, les tendances et les états saisonniers précédents sont également inclus dans l'équation de lissage de niveau.

Equations de lissage de tendance

Les types de modèle associés à une tendance additive ou une tendance additive amortie calculent un état de tendance pour chaque point de série temporelle en utilisant l'équation de lissage de tendance correspondante. L'état de tendance pour le point actuel est basé sur la différence entre les états de niveau au point actuel et au point précédent, et sur l'état de tendance au point précédent. La pondération associée à la différence entre les états de niveau au point actuel et au point précédent est un paramètre nommé **beta** dont la valeur est comprise entre 0 et 1. Un paramètre supplémentaire (**phi**) est ajouté aux équations de lissage de tendance amortie. **Phi** multiplie la contribution de l'état de tendance à partir du point précédent et sa valeur est également comprise entre 0 et 1. La fonction de ce paramètre est d'estimer le degré d'amortissement des tendances d'un point à l'autre.

Equations de lissage saisonnier

Les types de modèle qui prennent en charge la saisonnalité additive ou multiplicative calculent un état saisonnier pour chaque point de série temporelle. Les états saisonniers sont calculés en utilisant des équations de lissage saisonnier. L'état saisonnier du point actuel inclut la différence entre la valeur de série temporelle et l'état de niveau actuel pour la saisonnalité additive ou le rapport de ces deux mêmes valeurs pour la saisonnalité multiplicative. La pondération associée à ce terme est un paramètre (**gamma**) dont la valeur est comprise entre 0 et 1. Le reste de la contribution provient de l'état saisonnier correspondant dans la période saisonnière précédente. Notez que la période saisonnière a une longueur fixe, et si l'état saisonnier peut changer pour chaque point, seuls les index saisonniers correspondants issus de différentes périodes sont pris en compte ensemble dans les équations de lissage saisonnier.

Etats de lissage initiaux

Les valeurs doivent être spécifiées pour les états de niveau, de tendance et de saisonnalité pour les points qui précèdent la série temporelle. Les valeurs sont nécessaires pour les équations de lissage. Pour calculer les différents états au premier point de la série temporelle, des valeurs d'état aux points précédents correspondants sont requises.

Paramètres du modèle

Chaque équation de lissage utilise des paramètres de modèle correspondants :

alpha

Contrôle les états de niveau.

beta

Contrôle les états de tendance.

gamma

Contrôle les index saisonniers sur les périodes saisonnières.

phi

Paramètre supplémentaire utilisé pour spécifier la tendance amortie.

Ces quatre paramètres possèdent des valeurs comprises entre 0 et 1. Les valeurs plus élevées de **alpha**, **beta** et **gamma** indiquent que les observations plus récentes ont une pondération plus élevée, tandis que les valeurs plus faibles de ces paramètres indiquent des pondérations plus élevées pour les observations plus anciennes. Une valeur plus élevée de **phi** correspond à un degré plus élevé d'amortissement de la tendance des prévisions.

Estimation de modèle

Les paramètres de modèle des équations de lissage sont estimés en se basant sur les données de série temporelle. Les paramètres ne peuvent pas être estimés directement à l'aide d'une formule. Ils sont estimés par un processus itératif qui recherche des valeurs de paramètre qui réduisent au minimum l'erreur de modèle. L'erreur de modèle est calculée comme l'erreur moyenne absolue échelonnée (MASE, Mean Absolute Scaled Error). Les itérations s'arrêtent lorsqu'aucune autre réduction de l'erreur de modèle ne peut être appliquée. Les valeurs de paramètre correspondantes ainsi que les états de lissage initiaux spécifient intégralement le modèle estimé. Ils sont utilisés pour calculer les états de modèle pour tous les autres points de données et pour générer les prévisions de modèle à l'aide d'une équation de prévision correspondante.

Algorithmes de prévision

Un certain nombre d'algorithmes est utilisé pour la prévision.

Avec une étape d'avance

Chaque modèle prend en charge les prévisions avec une étape d'avance en fonction de l'équation de prévision correspondante. Ces prévisions sont nécessaires pour calculer les erreurs de modèle lors du processus d'estimation de modèle.

Les prévisions avec une étape d'avance sont calculées de manière séquentielle pour chaque point de données en utilisant les états de niveau et de tendance calculés pour le point en cours, ainsi que les états saisonniers pour la dernière période saisonnière.

L'erreur de prévision est calculée en soustrayant la valeur de prévision au point précédent de la valeur observée au point en cours. L'erreur de modèle global, qui est utilisée pour estimer le modèle, correspond à la valeur moyenne des erreurs de prévision absolues. Les erreurs plus faibles correspondent à une meilleure qualité d'ajustement. Les mesures de précision affichées dans la section **Détails statistiques de la prévision** fournissent plusieurs récapitulatifs de modèle des erreurs de prévision avec une étape d'avance.

Avec k étapes d'avance

Les prévisions avec k étapes d'avance sont utilisées pour effectuer des prévisions pour n'importe quel nombre de valeurs futures suivant les données de série temporelle observées. Elles sont basées sur les mêmes équations de prévision que les prévisions avec une étape d'avance pour le modèle spécifié.

Le nombre de valeurs de prévision générées correspond à 20 % de la longueur de la série de données historiques par défaut. Vous pouvez spécifier un nombre exact de valeurs à prévoir dans la boîte de

dialogue **Prévision**. Toute valeur manquante à la fin d'une série spécifique sera également prévue, mais elle ne comptera pas dans le nombre spécifié de périodes de prévision.

Bornes de confiance

Les bornes de confiance indiquent le niveau d'incertitude associé à chaque valeur de prévision. Etant donné que les prévisions plus distantes sont moins fiables, ces bornes s'étendent généralement au fil du temps. Les bornes de confiance fournissent des informations pertinentes sur le comportement futur de la série temporelle observée.

Le calcul des bornes de confiance est basé sur la variance globale des erreurs de prévision estimées sur les données observées et d'un facteur qui dépend du modèle spécifié et sur le nombre d'étapes à partir du dernier point observé.

Sélection d'un modèle automatisé dans les prévisions

Plusieurs types de modèle sont utilisés pour créer des modèles candidats pour chaque série temporelle d'une prévision. Les neuf types de modèle disponibles sont normalement utilisés, sauf lorsqu'une composante saisonnière est absente. Seuls trois types de modèle ne tiennent pas compte de la saisonnalité des données.

La valeur par défaut (**Auto**) de l'option "Période saisonnière" détecte la période saisonnière en comparant plusieurs modèles, chacun ayant une période saisonnière candidate différente.

Plusieurs modèles sont comparés à l'aide d'une erreur de modèle et du nombre de paramètres de modèle. Par exemple, lorsque les erreurs de modèle sont les mêmes pour deux modèles, le modèle ayant le moins de paramètres est privilégié. Ce modèle fournit en effet une représentation plus condensée des données observées et tend également à générer des prévisions plus fiables.

Chapitre 5. Principes de l'analyse avancée des données

Principes de l'analyse avancée des données

IBM Cognos Analytics est un outil d'aide à la décision qui permet de gérer et d'analyser les données. Il inclut des fonctionnalités en libre-service qui permettent aux utilisateurs de préparer, d'explorer et de partager des données. Cognos Analytics inclut des techniques prévisionnelles, descriptives et exploratoires, également connues sous le nom d'intelligence numérique. Cognos Analytics utilise de nombreux tests statistiques pour analyser vos données.

Il est indispensable de comprendre les définitions de ces tests qui s'appliquent à Cognos Analytics.

Des algorithmes numériques sont utilisés dans le cadre du flux de travaux pour fournir à l'utilisateur des fonctionnalités qui communiquent les informations sur les propriétés numériques et les relations dans leurs données.

Orientation métier

Contrairement aux logiciels statistiques traditionnels dont le public cible est un analyste de données expérimenté, les algorithmes de Cognos Analytics sont destinés aux utilisateurs qui connaissent l'analyse de données sans toutefois en être des spécialistes. Cela signifie que lorsque Cognos Analytics examine des compromis, il privilégie l'utilité à la complexité.

Fiabilité

Les données métier sont bien plus compliquées que les cas d'école utilisés dans les cours de statistiques ou dans les recherches et les exemples sur le Web. Cognos Analytics utilise des algorithmes puissants qui peuvent prendre en charge toute une gamme de données inhabituelles. Cognos Analytics procède ainsi car les algorithmes plus fragiles peuvent générer des résultats légèrement meilleurs que les algorithmes puissants, mais il est nécessaire de s'assurer qu'ils sont applicables et qu'ils génèrent des transformations de données correctes pour que les résultats soient significatifs. Une légère baisse de précision est le faible prix à payer pour la sécurité qu'offre un algorithme qui ne génère pas de résultats incorrects lorsque les données ne correspondent pas aux attentes.

Intelligence

Pratiquement tous les algorithmes requièrent de prendre des décisions à leur sujet, notamment en matière de niveaux de fiabilité, de combinaisons de champs à explorer ou de transformations de données. Les détails de ces décisions apparaissent dans les descriptions.

Cognos Analytics choisit automatiquement les valeurs appropriées en examinant les propriétés des données. En tant qu'utilisateur, il se peut que vous ne soyez pas conscient de toutes les décisions prises.

Récapitulatif

Dans Cognos Analytics, les procédures et les algorithmes numériques sont conçus pour générer automatiquement des résultats fiables. Pour obtenir la meilleure prévision, classification ou analyse possible, un statisticien professionnel analyse les données à l'aide d'IBM SPSS Statistics ou IBM SPSS Modeler. Cognos Analytics a pour objectif de fournir des connaissances de qualité qui vous aideront à comprendre vos données et leurs relations et ce, automatiquement pour de nombreux types de données différents. Cognos Analytics cherche à fournir des résultats similaires à ceux obtenus par un statisticien professionnel sans déranger l'utilisateur professionnel.

Préparation des données

La préparation des données est une étape de pré-analyse suivie par la plupart des algorithmes d'analyse de données pour s'assurer que les données peuvent être utilisées à des fins d'analyse.

Présentation

La préparation des données est essentielle dans IBM Cognos Analytics. Seules les données préparées sont prises en compte dans l'analyse des pilotes clés, des arbres de décisions et des relations qui sont affichés dans les visualisations d'analyse avancées : Spirale, Analyse des pilotes, Arbre de décisions, Rayon et Explorer les relations. Les données ne sont pas automatiquement préparées pour les autres visualisations et les connaissances correspondantes.

Algorithmes

Tous les algorithmes appliqués reposent sur les valeurs d'un seul champ à la fois. Les valeurs manquantes sont supprimées ou traitées pour chaque champ et tous les champs de facteur de prédicteur numériques sont regroupés. Tous les champs catégoriels sont ajustés pour un grand nombre de catégories et les aberrations sont traitées dans le champ cible. La préparation des données influence les résultats d'analyse, mais les récapitulatifs correspondants de cette préparation ne vous sont pas communiqués.

Détails

La préparation des données et les relations, arbres de décisions et facteurs clés ci-dessous reposent sur un échantillon de données d'environ 10000 lignes lorsque les données d'origine sont plus volumineuses. L'échantillonnage aléatoire de Bernoulli, l'échantillonnage aléatoire à probabilités égales sans remplacement, est appliqué aux données téléchargées et à toutes les sources de données connectées qui prennent en charge l'échantillonnage aléatoire. Sinon, l'échantillonnage systématique est utilisé.

Préparation des données pour les champs numériques

Un champ est traité comme un champ numérique dès lors qu'il contient des informations numériques et que sa propriété d'utilisation correspond à une mesure.

Présentation

Les données numériques pouvant être variées dans leur distribution, IBM Cognos Analytics transforme les champs numériques non cible en catégories ordinales, ce qui réduit la dépendance des algorithmes d'analyse sur le format des données numériques.

Algorithmes

L'algorithme de base utilisé est le regroupement en catégories de fréquences égales. Les données numériques sont divisées en un nombre fixe de catégories qui tentent de placer un nombre égal de lignes de données dans chaque catégorie. Les valeurs manquantes sont placées dans leur propre catégorie. Cognos Analytics tente d'utiliser ses connaissances sur les valeurs manquantes dans les champs prédicteurs pour générer un meilleur modèle. Par exemple, si un champ de données représente le moment auquel un élément a été testé, Cognos Analytics utilise des valeurs manquantes (qui peuvent représenter l'absence de test pour un élément) pour prévoir plus facilement les valeurs des autres champs.

Détails

Certains critères d'exclusion de champ s'appliquent aux champs numériques. Un champ numérique est exclu de toute autre analyse s'il ne contient qu'une seule valeur, y compris la valeur manquante. Sinon, le champ numérique est placé dans une catégorie et le nombre de catégories par défaut est cinq. Si un champ contient plus de 10 valeurs numériques uniques, aucune tentative de regroupement en catégories n'est effectuée et chaque valeur unique est placée dans sa propre catégorie. Si plus de 40 % des lignes contiennent la valeur zéro, cette valeur est toujours placée dans une catégorie distincte. Les valeurs

manquantes sont placées dans leur propre catégorie et n'affectent pas la procédure de regroupement en catégories.

Préparation des données pour les champs catégoriels

Un champ est traité comme un champ catégoriel dès lors que sa propriété d'utilisation correspond à un attribut ou à un identificateur.

Présentation

Les informations principales extraites des champs catégoriels sont la fréquence observée pour chaque valeur de catégorie unique. Des méthodes d'analyse appropriées sont appliquées aux champs catégoriels, mais leur précision et leurs performances peuvent diminuer si le nombre de catégories différentes devient important. La principale étape de préparation des données consiste à commencer à fusionner les catégories lorsqu'elles deviennent nombreuses.

Algorithmes

L'algorithme de base utilisé est le fusionnement des catégories. Les catégories sont triées par fréquence dans l'ordre décroissant et celles qui se trouvent au-delà du nombre par défaut sont fusionnées dans une même catégorie. Les valeurs manquantes sont traitées comme une catégorie distincte unique. En d'autres termes, IBM Cognos Analytics utilise les valeurs manquantes comme pour les champs numériques. Les champs catégoriels sont traités comme des champs nominaux. L'ordre intrinsèque n'est pas supposé parmi les catégories.

Détails

Certains critères d'exclusion de champ s'appliquent aux champs catégoriels. Un champ catégoriel est exclu de toute autre analyse s'il ne contient qu'une seule valeur ou que le nombre de catégories uniques, non fusionnées dépasse 50 % du nombre de lignes de données valides.

Sinon, le champ catégoriel est fusionné et le nombre par défaut de catégories non fusionnées est de 49. Les autres catégories sont fusionnées dans une même catégorie supplémentaire. Toutes les catégories dont le nombre de lignes est inférieur à trois sont fusionnées. Un champ catégoriel est également exclu si le pourcentage de lignes de données valides correspondant à la catégorie fusionnée dépasse 25 %.

Les valeurs manquantes sont traitées comme une catégorie distincte et prises en compte en tant que telles dans l'étape de fusion.

Préparation des données pour les champs cible

La spécification du champ cible est requise pour les visualisations des facteurs clés et des arbres de décisions.

Présentation

Spécifiez toujours le champ cible et au moins un champ supplémentaire. Les modèles sont formés à l'aide des valeurs cible fournies et sont utilisés pour détecter les relations prédictives, puis prédire les valeurs cible en fonction des valeurs des champs en entrée. La préparation des données pour le champ cible diffère de la préparation des données pour les autres champs. Les valeurs manquantes dans la cible ne sont pas utilisées pour générer des modèles, mais les autres informations sont conservées et parfois ajustées pour obtenir des modèles non biaisés.

Algorithmes

L'étape principale de préparation des données relative aux champs cible consiste à supprimer toutes les lignes de données dont la valeur cible est manquante. Elle a lieu avant toutes les autres étapes de préparation des données. Cette étape permet de garantir que seules des informations fiables sont utilisées pour générer des modèles, mais le nombre de lignes supprimées peut s'avérer considérable. Dans de tels cas, le modèle obtenu peut avoir une portée limitée. Les champs cible numériques ne sont pas regroupés, mais les aberrations extrêmes sont traitées pour ne pas avoir un impact négatif sur les

modèles créés par la suite. Les champs cible catégoriels sont traités d'une manière très proche de celle des autres champs catégoriels, mais les valeurs manquantes ont été supprimées pour les cibles catégorielles.

Détails

Les aberrations extrêmes sont détectées en fonction des limites inférieure et supérieure. La limite supérieure est déterminée à l'aide d'un percentile supérieur défini afin que seuls 2,5 % des valeurs cible possèdent une valeur supérieure. La différence entre le percentile supérieur et la médiane est multipliée par 2,5 et ajoutée à la médiane pour obtenir la limite supérieure. Des étapes similaires sont appliquées pour obtenir la limite inférieure. Les valeurs cible qui se trouvent au-delà des limites calculées sont remplacées par la valeur limite correspondante dans toutes les analyses ultérieures.

Facteurs clés unidirectionnels

Les facteurs clés unidirectionnels sont un outil explicatif basé sur un modèle.

Présentation

A partir d'un champ cible, l'outil utilise un modèle statistique pour analyser tout autre champ de données disponible et estime sa force de prévision des valeurs cible. Ces champs de données sont appelés facteurs ou prédicteurs cible. Chaque champ de données potentiellement pertinent est analysé et seuls les principaux facteurs relatifs à leur force prédictive sont affichés. Vous obtenez des informations sur les facteurs disponibles et leur classement en fonction de leur force prédictive pour la cible spécifiée dans les données. Les résultats d'analyse des facteurs unidirectionnels sont disponibles à la fois dans l'analyse des facteurs et dans les visualisations en spirale. L'exploration visuelle de chaque facteur n'est activée que pour la visualisation d'analyse de facteur dans Explorer.

Algorithmes

L'analyse de chaque facteur unidirectionnel est basée sur un modèle statistique qui inclut la cible et un prédicteur catégoriel unique. Le modèle est appliqué après l'étape de préparation des données pour le champ cible et tous les champs prédicteurs potentiels. Par exemple, tous les champs de prédicteurs numériques sont regroupés lors de l'étape de préparation des données et traités comme des champs catégoriels dans l'analyse. L'analyse de variance unidirectionnelle est appliquée pour les cibles numériques et le test d'indépendance du khi-carré est appliqué pour les cibles catégorielles avec l'ajustement du khi-carré pour les données fragmentées.

Pour chaque champ de la liste des facteurs potentiels, un test d'hypothèse est effectué pour déterminer si le champ a un impact important sur la cible. Seuls les champs qui réussissent le test et possèdent une force prédictive suffisamment élevée sont sélectionnés comme facteurs clés unidirectionnels possibles.

Détails

Une analyse préliminaire basée sur des fonctionnalités intelligentes réduit le nombre de facteurs potentiels dans certains cas. L'objectif est de supprimer les champs non pertinents ou redondants. La liste des facteurs utilisés est disponible dans l'interface utilisateur et vous pouvez ajouter des facteurs initialement exclus à l'analyse. Les 20 premiers facteurs obtenus dont la force prédictive est supérieure à 10% peuvent être affichés.

Certaines restrictions sont appliquées à la taille des données pour améliorer les performances et la vitesse. Si les données contiennent plus de 250 champs, les champs les moins pertinents sont exclus avant l'analyse des facteurs. Vous pouvez rajouter les champs exclus dans l'analyse via l'interface utilisateur, comme décrit précédemment. Si les données spécifiées contiennent plus de 10000 lignes, un échantillon d'environ 10000 lignes peut être utilisé pour l'analyse des facteurs. Un avertissement s'affiche en pareil cas : *Pour améliorer les performances, en raison du nombre de lignes dans la source de données, l'analyse est basée sur un échantillon représentatif de l'ensemble des données.* Les résultats doivent correspondre approximativement à ceux qui auraient été obtenus si toutes les lignes des données d'origine avaient été utilisées.

Facteurs clés bidirectionnels

Les facteurs bidirectionnels s'appuient sur la modélisation et le classement de paires de prédicteurs indépendants à un moment donné.

Présentation

A partir d'un champ cible, IBM Cognos Analytics utilise un modèle statistique pour analyser une paire d'autres champs de données et estime sa force de prévision des valeurs cible. Une recherche sur différentes paires de prédicteurs n'est généralement pas exhaustive et certaines paires bien classées peuvent être exclues des résultats finaux. L'objectif est de fournir un aperçu et une variété de paires de prédicteurs dont la force prédictive est supérieure à celle des modèles à un prédicteur affichés comme facteurs unidirectionnels. Par conséquent, les connaissances obtenues des facteurs unidirectionnels sont étendues et l'utilisateur reçoit des informations pertinentes sur les paires de champs dans les données. Les résultats d'analyse des facteurs unidirectionnels et des facteurs bidirectionnels sont tous deux disponibles dans l'analyse des facteurs et les graphiques en spirale. Ils peuvent être affichés séparément en sélectionnant une option d'affichage de graphique correspondante. Chaque facteur unidirectionnel ou bidirectionnel affiché peut être étendu dans une nouvelle visualisation directement à partir de la visualisation Analyse du facteur dans Explorer.

Algorithmes

L'analyse de chaque facteur bidirectionnel repose sur un modèle statistique qui inclut la cible et une paire de prédicteurs indépendants. Le modèle est appliqué après la préparation des données et la génération de tous les facteurs unidirectionnels. Le premier prédicteur de la paire est sélectionné parmi les 50 premiers facteurs unidirectionnels et le second, parmi les 25 premiers facteurs unidirectionnels. Cette stratégie de recherche garantit que la plupart des principales paires de prédicteurs sont prises en compte pour la modélisation. L'analyse de variance bidirectionnelle est appliquée pour les cibles numériques et le test d'indépendance du khi-carré est appliqué pour les cibles catégorielles avec l'ajustement du khi-carré pour les données fragmentées.

Pour chaque paire de champs envisagée, un test d'hypothèse est effectué pour déterminer si la paire a un impact important sur la cible. Seules les paires qui réussissent le test et possèdent une force prédictive suffisamment élevée sont sélectionnées comme facteurs bidirectionnels possibles.

Détails

La restriction de la sélection des champs de données et des lignes de données pour les facteurs unidirectionnels s'applique également aux facteurs bidirectionnels. Cela est normal car les champs prédicteurs potentiels des facteurs bidirectionnels sont sélectionnés à partir des principaux facteurs unidirectionnels en fonction de leur force prédictive respective. Toutefois, la pertinence du modèle d'un facteur unidirectionnel et la force prédictive minimale ne sont pas requises pour leur entrée dans un modèle bidirectionnel. La force prédictive du facteur bidirectionnel obtenu doit être supérieure d'au moins 10 % et offrir une amélioration relative d'au moins 10 % par rapport à la force prédictive de chacun des facteurs unidirectionnels inclus. L'amélioration relative est calculée sous la forme du pourcentage de la différence entre 100 % et la force prédictive du facteur unidirectionnel imbriqué. Les facteurs bidirectionnels obtenus qui remplissent ces critères sont classés en fonction de leur force prédictive et les 20 premiers peuvent être affichés.

Arbre de décisions

Les arbres de décisions sont des modèles plus complexes que les facteurs unidirectionnels et bidirectionnels. Ils étendent la séquence sous forme de modèles de combinaison. La principale différence est la suivante : les arbres de décisions prennent en charge la détection d'interaction entre plusieurs prédicteurs et permettent donc d'obtenir des connaissances plus approfondies que les facteurs.

Présentation

A partir du champ cible, l'algorithme effectue des recherches dans tous les autres champs de données et les ajoute au modèle pour en améliorer la fiabilité de prévision des valeurs cible. La recherche dans les

différents prédicteurs est itérative ; une fois que la recherche a ajouté un prédicteur, elle continue d'ajouter le prochain prédicteur qui améliore le plus le modèle. Son objectif est de rechercher le meilleur ensemble de prédicteurs et un moyen optimal de les combiner afin de générer un modèle optimal. Les connaissances provenant des arbres de décisions sont présentées sous la forme de règles de décision dans lesquelles la combinaison de prédicteurs et les valeurs correspondantes fournissent une prévision unique pour la valeur cible. Les règles de décision sont classées par force pour que vous puissiez facilement rechercher les règles les plus pertinentes et intéressantes. Les règles de décision générées par l'arbre de décisions s'excluent mutuellement. Les règles de décision fournissent également un ensemble complet de règles afin qu'une règle correspondante existe pour toute combinaison de valeurs de prédicteur dans les données. La force prédictive globale de l'arbre de décisions, améliorée par rapport à celle du modèle de base, est également disponible. Les résultats sont disponibles dans trois visualisations différentes : rayons, arborescence et règles de décision. Chacune d'elles offre certains avantages en affichant la structure de l'arbre de décisions et le contenu correspondant des règles de décision. La force prédictive globale de l'arbre de décisions est également disponible dans la visualisation d'analyse de facteur.

Algorithmes

Le modèle d'arbre de décisions est généré après la préparation des données et la génération de tous les facteurs unidirectionnels. Le premier prédicteur d'arborescence est sélectionné comme principal facteur unidirectionnel. Les catégories du prédicteur sont fusionnées lorsque l'impact négatif sur la force prédictive est inférieur à un certain seuil. L'étape suivante consiste à rechercher le meilleur prédicteur pour fractionner chaque noeud d'arborescence constitué des catégories fusionnées. Le processus se poursuit jusqu'à ce qu'une règle d'arrêt s'applique à un noeud d'arborescence. Deux options d'arrêt sont possibles : toutes les catégories de chacun des prédicteurs candidats sont fusionnées dans un même noeud ou le nombre de noeuds dépasse le nombre maximal de noeuds admis. Les catégories comportant un nombre de lignes inférieur au minimum spécifié sont toujours fusionnées avec une autre catégorie. Cela signifie qu'aucun des noeuds de l'arborescence ne peut contenir moins de lignes que le nombre minimal de lignes. La même procédure est utilisée pour les cibles continues et catégorielles ; seule la fonction d'impureté est différente.

Détails

Fonctions d'impureté

Les valeurs des fonctions d'impureté sont utilisées comme critère principal pour le fractionnement et la fusion des noeuds d'arborescence potentiels. Le total de la fonction d'impureté pour les arborescences continues correspond à la somme des carrés par noeud, tandis que la mesure d'impureté Gini est utilisée pour les cibles catégorielles. Le total d'impureté Gini correspond à la somme des carrés des proportions des nombres dans toutes les catégories cible par noeud qui est soustraite de la valeur un, le tout étant multiplié par le nombre de lignes. L'amélioration de la valeur de la fonction d'impureté permet un gain d'informations.

Lors du fractionnement de chaque noeud, IBM Cognos Analytics recherche le champ prédicteur dont le gain d'information est le plus élevé ; ce gain correspond à l'impureté du noeud parent de laquelle est soustraite l'impureté totale de tous les noeuds enfant potentiels. Avant que Cognos Analytics sélectionne le prédicteur, Cognos Analytics tente de fusionner certains noeuds enfant potentiels qui correspondent initialement à chaque catégorie de prédicteur. La perte d'informations est calculée en soustrayant l'impureté des noeuds non fusionnés de l'impureté des noeuds fusionnés. Tant que la perte d'informations est inférieure à un seuil donné, les noeuds sont fusionnés. Ce processus permet de créer des arborescences relativement petites, faciles à visualiser et à comprendre, tout en conservant la force globale de l'arborescence.

Règles d'arrêt

Les noeuds candidats sont toujours fusionnés s'ils sont basés sur moins de 25 lignes. Si toutes les catégories d'un prédicteur sont fusionnées, ce dernier ne peut pas être utilisé pour le fractionnement d'un certain noeud. Si aucun des prédicteurs ne peut fractionner le noeud spécifique, le processus s'arrête pour ce noeud. Le processus global de génération de l'arborescence s'arrête lorsqu'aucun des noeuds ne peut être fractionné ou si le nombre de noeuds générés dépasse 36.

Importance de variable

L'importance de variable correspond à une réduction relative des erreurs de l'arborescence lorsque le prédicteur correspondant est inclus dans l'arborescence. Elle est calculée en comparant les erreurs d'une arborescence initiale et une arborescence restreinte générée par le reste des prédicteurs de l'arborescence initiale. L'erreur de l'arborescence initiale est soustraite de l'erreur de l'arborescence restreinte et le résultat est divisé par l'erreur de l'arborescence restreinte. Les variables dont l'importance est nulle ou négative sont supprimées de l'arborescence. L'erreur d'arborescence correspond à la somme des carrés pour les cibles continues et à une erreur de classification pour les cibles catégorielles.

Force prédictive

La force prédictive de l'arborescence avec la cible continue est calculée comme pour les facteurs clés. Le contenu des noeuds feuille est pris en compte. La contribution de la variance pour chaque noeud feuille est ajoutée et divisée par la variance globale des données. Il s'agit d'une erreur relative pour l'arborescence. Elle est soustraite de la valeur un pour obtenir une force prédictive compatible avec la mesure R carré utilisée pour les facteurs clés.

Pour les cibles catégorielles, Cognos Analytics calcule la précision de la classification en fonction de l'erreur de classification ajoutée à partir de tous les noeuds feuille. L'amélioration de la précision de la classification par rapport à celle du modèle de base, également connue sous le nom de R carré ajusté des nombres, est renvoyée comme force prédictive de l'arborescence. Elle est calculée en soustrayant l'erreur d'arborescence de l'erreur du modèle de base et en divisant le résultat par l'erreur du modèle de base. Par exemple, la précision de la classification du modèle peut atteindre 95 %, mais si la classe majoritaire apparaît pour 90 % des lignes des données, la force prédictive renvoyée pour l'arborescence n'est que de 50 %. Cela correspond au cas de la cible continue où le modèle de base est représenté par la valeur moyenne globale. La force prédictive mesurée par R carré repose sur l'amélioration relative de l'arborescence dans le cadre de la réduction de la variance globale.

Cognos Analytics n'affiche que les arborescences dont la force prédictive est supérieure à 10 %. Une arborescence de cible continue est affichée dans une analyse de facteur ou une visualisation en spirale si sa force prédictive est supérieure à celle du facteur clé le plus fort. Sinon, elle n'est pas affichée dans ces graphiques car les facteurs clés fournissent déjà toutes les connaissances pertinentes.

La force prédictive d'un arbre de décisions qui utilise les données utilisées pour générer l'arbre de décisions est calculée. Elle est connue pour être biaisée et fournir des estimations optimistes sur les performances de l'arbre de décisions pour des données similaires de la même source de données. Cognos Analytics réduit l'écart en optimisant l'algorithme afin que le surajustement des données d'apprentissage soit minimisé.

Connaissances dans les visualisations

Les connaissances dans les visualisations fournissent des informations analytiques qui permettent aux utilisateurs d'identifier et de valider les relations importantes et les différences significatives à partir des données présentées dans la visualisation.

Présentation

Les connaissances sont gérées et récapitulées dans la section **Connaissances** qui est disponible dans chacune des visualisations éligibles. Lorsque vous activez la fonction Connaissances, le récapitulatif apparaît dans la section **Connaissances** et les éléments de visualisation connexes sont mis en évidence. Des détails sont fournis dans le message d'infobulle correspondant. Vous pouvez contrôler chaque information disponible séparément.

Algorithmes

Le type de connaissances dépend des données affichées par la visualisation. Les types de connaissances disponibles sont **Moyenne**, **Force prédictive**, **Différences significatives**, **Ligne d'ajustement** et **Les plus fréquents**. Le type "Moyenne" fournit la moyenne des récapitulatifs affichés, tandis que "Les plus

fréquents" fournit la catégorie qui apparaît le plus souvent dans les données. Les autres connaissances sont liées à des tests d'analyse et des tests statistiques plus avancés. L'objectif est de fournir des informations fiables qui permettent une meilleure description des données affichées et la détection des relations attendues dans la population représentée par ces données.

Détails

L'analyse des connaissances s'appuie toujours sur les lignes de données qui sont utilisées pour créer les récapitulatifs affichés dans la visualisation. Cela signifie que l'intégralité des données sont utilisées pour les connaissances à moins qu'un filtrage ne soit appliqué aux données d'origine.

Certains tests statistiques et certaines analyses utilisés dans les connaissances requièrent non seulement les récapitulatifs de données affichés dans la visualisation, mais aussi des récapitulatifs supplémentaires. Par exemple, le test des différences significatives sur plusieurs catégories d'un champ explicatif requiert les nombres et variances de chaque catégorie en plus des données affichées. Ces récapitulatifs supplémentaires sont extraits d'une base de données avec les récapitulatifs requis pour la visualisation. Tous les récapitulatifs sont traités par les connaissances, mais seuls les récapitulatifs requis sont disponibles dans la visualisation. L'analyse des connaissances est toujours basée sur les lignes de données utilisées pour créer les récapitulatifs affichés dans la visualisation.

Restrictions

Si des connaissances ne sont pas disponibles immédiatement dans une visualisation, cela peut être dû à l'un des éléments suivants :

- Le type de visualisation lui-même ne prend pas en charge les connaissances.
- Les données de la visualisation ont peut-être été tronquées.
- La combinaison du niveau de récapitulation, du type de champ et du rôle d'un champ sélectionné ne correspond pas aux exigences de l'une des connaissances disponibles.

Types de visualisation pris en charge pour les connaissances

Les types de visualisation suivants prennent en charge les connaissances :

- Aires
- Barres
- Bulles
- Colonnes
- Carte de densité
- Bulle de hiérarchie
- Courbes
- Courbes et colonnes
- Carte
- Paquets de bulles
- Graphique circulaire
- Points
- Radial
- Nuage de points
- Barres superposées
- Colonnes empilées
- Carte d'arborescence
- Nuage de mots

Les petites extensions multiples sont prises en charge pour certaines connaissances, notamment Les plus fréquents et Différences significatives.

Niveaux de récapitulation

Les niveaux de récapitulation pris en charge sont **Nombre**, **Moyenne**, **Somme**, **Minimum** et **Maximum**. Toutes les autres valeurs, telles que **Nombre (éléments distincts)**, peuvent empêcher la recommandation de connaissances. Certains algorithmes ne prennent en charge que des niveaux de récapitulation spécifiques. Remplacer le niveau de récapitulation par défaut par l'une des valeurs prises en charge peut éventuellement permettre d'activer les connaissances.

Types de champ

Les types de champ peuvent être conçus en interne comme continus ou catégoriels en fonction des valeurs du champ sélectionné.

Type de champ	Description
Catégoriel	Variable qui peut accepter une valeur parmi un nombre limité et généralement fixe de valeurs possibles. Une variable catégorielle affecte chaque individu ou autre unité d'observation à un groupe particulier ou à une catégorie nominale en fonction d'une propriété qualitative. Par exemple, le pays dans lequel réside une personne.
Continu	Variable qui est utilisée pour décrire des valeurs numériques, telles qu'une plage de 0 à 100 ou de 0,75 à 1,25. Une valeur continue peut être un entier, un nombre réel ou une date et une heure.

Rôles des champs

IBM Cognos Analytics affecte un rôle à chacun des emplacements de champ dans une visualisation prise en charge. Un rôle de champ peut être désigné comme l'un des rôles suivants, en fonction de l'emplacement de la visualisation.

Rôle du champ	Description
Réponse	Variable qui est prédite ; également appelée cible ou variable dépendante. Elle se trouve généralement sur l'axe des Y.
Explicatif	Variable qui aide à expliquer les modifications dans la réponse ; également appelée prédicteur ou variable indépendante. Elle se trouve généralement sur l'axe des X.
Groupe	Variable qui est traitée comme un facteur de regroupement facultatif ou explicatif, utilisé pour déterminer le nombre de modèles générés dans l'algorithme. Par exemple, peut correspondre à l'emplacement Couleur d'une visualisation à colonnes.
Pondération	Variable qui définit les pondérations de régression facultatives, utilisées pour calculer le modèle de régression. Par exemple, ceci peut correspondre à l'emplacement Taille d'une visualisation à bulles.
Répétition	Variable qui crée de petits multiples, la visualisation étant répétée une fois pour chaque valeur distincte de la variable. Par exemple, peut correspondre à l'emplacement Répéter (lignes) d'une visualisation de graphique circulaire.
Points	Variable qui définit la forme des données et des points de données, utilisés pour le calcul du modèle. Par exemple, ceci peut correspondre à l'emplacement Points d'une visualisation en nuage de points.

Comme exemple général, dans une visualisation à barres avec les emplacements suivants, les mappages de rôle dans cette visualisation sont définis comme suit :

- Barres (axe des Y), explicatif
- Longueur (axe des X), réponse

- Couleur, groupe

Connaissances dans les visualisations pour les nombres

Des connaissances sur les nombres sont disponibles lorsqu'un nombre est affiché pour chaque catégorie d'un champ catégoriel unique.

Ces connaissances sont également disponibles lorsque le nombre s'affiche pour chaque combinaison de catégories d'une paire de champs catégoriels dans la visualisation. Dans ce dernier cas, la paire peut correspondre à deux champs explicatifs, comme dans les lignes et les colonnes d'une carte de densité, ou à un champ explicatif et un champ de répétition, comme dans les barres et les répétitions (colonnes) d'un graphique à barres.

Les connaissances des nombres de catégories combinées de trois champs catégoriels sont prises en charge pour un champ explicatif et la combinaison de deux champs de répétition, comme dans les segments, les répétitions (colonnes) et les répétitions (lignes) d'une visualisation de graphique circulaire.

Présentation

Utilisez ce type de visualisation si vous cherchez à comparer le nombre d'éléments dans différentes catégories ou dans une combinaison de catégories.

Algorithmes

IBM Cognos Analytics renvoie le nombre moyen de toutes les catégories du champ de réponse spécifié et applique des tests statistiques pour détecter les catégories dont les nombres sont les plus différents de la moyenne d'un point de vue statistique.

Les visualisations comportant deux ou trois champs catégoriels et des nombres pour chaque combinaison de catégories sont traitées de manière différente. Cognos Analytics compare non seulement les nombres dans les catégories, mais il détecte également les relations entre les champs catégoriels. Cognos Analytics traite le premier champ comme le champ de réponse et les autres champs comme le champ explicatif.

Cognos Analytics signale la catégorie la plus fréquente dans les visualisations par un champ explicatif catégoriel, un ou deux champs de répétition catégoriels et un champ de réponse numérique.

Détails

Champ catégoriel unique

Le premier test appliqué est le test du khi-carré des fréquences égales pour déterminer si des nombres disponibles sont nettement différents de la moyenne. Si le résultat du test est pertinent, Cognos Analytics applique le test d'influence du khi-carré séparément pour chaque catégorie. Cognos Analytics calcule les tailles d'effet des catégories pour lesquelles le test d'influence est pertinent d'un point de vue statistique et renvoie les catégories dont les tailles d'effet sont les plus élevées parmi les différences significatives.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissances
Exactement 1 Niveau de récapitulation = Nombre	N/A	N/A	N/A	N/A	Moyenne Différences significatives

Deux champs catégoriels

Cognos Analytics traite le premier champ catégoriel comme le champ de réponse et le second comme le champ explicatif. Le champ numérique initial est utilisé comme une entrée dans les algorithmes.

Le test d'indépendance du khi-carré avec ajustement des données fragmentées est utilisé pour déterminer s'il existe une relation entre le champ de réponse et le champ explicatif. Si le résultat du test est pertinent, Cognos Analytics calcule la force prédictive de ce modèle comme le R carré ajusté des nombres, en filtrant les catégories dont la fréquence est faible. La relation est déclarée fiable et la force prédictive est renvoyée si elle est supérieure à 10 %.

Si le résultat du test ci-dessus est significatif, toutes les combinaisons de catégories explicatives et de réponse sont analysées de manière plus approfondie par l'application du test d'influence du khi-carré à chaque combinaison. Les combinaisons de catégories explicatives et de réponse dont le test d'influence est significatif sont considérées comme influentes. La taille d'effet est calculée pour chacune des combinaisons de catégories influentes et celles dont les tailles d'effet sont les plus élevées sont signalées comme différences pertinentes.

Si les rôles des deux champs catégoriels sont explicatifs et répétitifs, l'algorithme le plus fréquent est appliqué. Les nombres sont additionnés pour chaque catégorie distincte du champ explicatif. La somme la plus élevée est déclarée, ainsi que le nombre de catégories ayant cette somme. Notez que le champ de répétition n'est pas utilisé par cet algorithme. Il se déclenche uniquement lorsque l'algorithme est appliqué.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Répétition	Connaissance
Exactement 1 Niveau de récapitulation = Nombre	Exactement 2 Catégoriel	N/A	N/A	N/A	N/A	Force prédictive Différences significatives
Exactement 1 Niveau de récapitulation = Nombre	Exactement 1 Catégoriel	N/A	N/A	N/A	Exactement 1 Catégoriel	Force prédictive Différences significatives Le plus fréquent

Trois champs catégoriels

Ces algorithmes ne sont appliqués que lorsqu'il existe un champ explicatif et deux champs de répétition. La combinaison des deux champs de répétition est traitée comme s'il s'agissait d'un champ catégoriel unique, où les catégories sont les paires de catégories des deux champs de répétition.

La force prédictive est calculée exactement comme dans le cas des deux champs catégoriels, en utilisant les champs de répétition appariés comme prédicteur du champ explicatif. Le test d'indépendance du khi-carré avec ajustement des données fragmentées est utilisé pour tester la signification de la relation, et la mesure R carré ajusté des nombres avec les catégories à faible fréquence filtrées est utilisée pour trouver la force prédictive.

Les différences significatives sont calculées exactement comme dans le cas des deux champs catégoriels, en identifiant les combinaisons du champ explicatif et des champs de répétition appariés pour lesquels le nombre est inhabituel. Le test d'influence du khi-carré permet de tester la signification de chaque combinaison, et les combinaisons dont la taille d'effet est la plus élevée sont signalées.

L'algorithme le plus fréquent est appliqué exactement comme dans le cas des deux champs catégoriels, en additionnant les nombres pour chaque catégorie distincte du champ explicatif. La somme la plus élevée est déclarée, ainsi que le nombre de catégories ayant cette somme. Notez que les champs de répétition ne sont pas utilisés par cet algorithme. Ils se déclenchent uniquement lorsque l'algorithme est appliqué.

Réponse	Explicatif	Groupe	Pondération	Points	Répétition	Connaissance
Exactement 1 Niveau de récapitulation = Nombre	Exactement 1 Catégoriel	N/A	N/A	N/A	Exactement 2 Catégoriel	Force prédictive Différences significatives Le plus fréquent

Différences entre Cognos Analytics version 11.1 R2 et R3

Pour les visualisations dotées de deux champs catégoriels, lorsque le champ de réponse possède une catégorie représentant des données manquantes (la catégorie "(aucune valeur)"): :

- Dans Cognos Analytics version 11.1 R2, le calcul du R carré ajusté des nombres pour la force prédictive a omis des valeurs de données pour la catégorie des données manquantes. Dans Cognos Analytics version 11.1 R3, les valeurs sont incluses. Cela peut affecter la force prédictive indiquée pour les cartes de densité où l'emplacement des données de densité comporte des données de nombres agrégés et où les lignes et les colonnes correspondent à deux champs catégoriels.
- Dans Cognos Analytics version 11.1 R2, les différences significatives ne signalent aucune cellule inhabituelle associée à la catégorie des données manquantes. Dans Cognos Analytics version 11.1 R3, les cellules sont indiquées. Cela peut affecter les différences significatives affichées pour les cartes de densité où l'emplacement des données de densité comporte des données de nombres agrégés et où les lignes et les colonnes correspondent à deux champs catégoriels.

Connaissances dans les visualisations pour les récapitulatifs par un ou plusieurs champs explicatifs

Des connaissances pour les récapitulatifs sont disponibles lorsque le niveau de récapitulation est Moyenne, Somme, Minimum ou Maximum pour un champ de réponse continu. Ces connaissances sont calculées et affichées au niveau de chaque catégorie d'un champ explicatif catégoriel unique ou de chaque combinaison de catégories d'une paire de champs explicatifs catégoriels dans la visualisation.

Présentation

Utilisez de telles visualisations lorsque vous cherchez à comparer des valeurs d'un champ de réponse dans des catégories différentes ou des combinaisons de catégories de champs explicatifs.

Algorithmes

Si le niveau de récapitulation est Moyenne, IBM Cognos Analytics détecte toute relation entre le champ de réponse et les champs explicatifs et calcule la force prédictive du modèle correspondant. Si les différences entre les valeurs moyennes des catégories explicatives sont significatives du point de vue statistique, Cognos Analytics identifie les catégories explicatives les plus différentes ou la combinaison de catégories la plus différente comme différences significatives.

Lorsque le niveau de récapitulation de la réponse est Somme, Cognos Analytics calcule la somme moyenne entre les catégories explicatives ou les combinaisons de catégories. Si les différences des sommes des catégories sont significatives du point de vue statistique, Cognos Analytics identifie les catégories explicatives ou les combinaisons de catégories les plus différentes comme différences significatives.

Pour l'ensemble des graphiques applicables, les connaissances sur la moyenne indiquent la valeur moyenne récapitulée des réponses dans toutes les catégories explicatives. Si le niveau de récapitulation du champ de réponse est Moyenne, la moyenne pondérée est calculée à l'aide de la valeur affichée et du nombre pour chaque catégorie explicative.

Détails

Moyenne par champ explicatif unique

Si le niveau de récapitulation du champ de réponse est Moyenne et qu'un seul champ explicatif catégoriel est disponible, Cognos Analytics applique l'analyse de variance unidirectionnelle. Cognos

Analytics utilise la statistique F pour vérifier si les valeurs moyennes des catégories explicatives sont égales. Si des différences sont significatives, Cognos Analytics calcule le R carré ajusté comme force prédictive de la relation entre le champ de réponse et le champ explicatif. Une relation fiable et sa force prédictive sont renvoyées à l'utilisateur si la force prédictive dépasse 10 %.

Si les différences entre les moyennes sont importantes, Cognos Analytics effectue un test d'influence t pour détecter les catégories les plus différentes de la moyenne globale. Cela implique de calculer l'erreur standard de chaque moyenne de catégorie et de comparer la moyenne à la moyenne globale à l'aide de la statistique de test t. Pour les catégories présentant des différences significatives, Cognos Analytics calcule également la taille d'effet correspondante et renvoie les catégories dont la taille d'effet est la plus élevée parmi les différences significatives.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1 Niveau de récapitulation = Moyenne Continu	Exactement 1 Catégoriel	N/A	Facultatif Tout	N/A	Moyenne Force prédictive Différences significatives

Moyenne par deux champs explicatifs

Pour les graphiques dans lesquels le niveau de récapitulation du champ de réponse est Moyenne et où deux champs explicatifs catégoriels sont disponibles, Cognos Analytics applique l'analyse de variance bidirectionnelle. Cognos Analytics utilise la statistique F pour vérifier si les valeurs moyennes des combinaisons de catégories explicatives sont égales. Si des différences sont significatives, Cognos Analytics calcule le R carré ajusté comme force prédictive de la relation entre le champ de réponse et les deux champs explicatifs. Cognos Analytics calcule également le R carré ajusté pour les modèles unidirectionnels qui incluent chacun un seul champ explicatif. Si la force prédictive d'un modèle bidirectionnel est supérieure à 10 % et qu'elle est également supérieure à 10 % par rapport aux modèles unidirectionnels correspondants, Cognos Analytics affiche la force prédictive du modèle bidirectionnel et renvoie une relation fiable entre le champ de réponse et les deux champs explicatifs. Sinon, si la force prédictive maximale des modèles unidirectionnels dépasse 10 %, Cognos Analytics renvoie une relation fiable entre le champ de réponse et le champ explicatif unique correspondant avec sa force prédictive. Si la force prédictive maximale des modèles unidirectionnels ne dépasse pas 10 %, Cognos Analytics renvoie l'absence de relation entre le champ de réponse et les champs explicatifs.

Si les différences entre les moyennes de toutes les combinaisons de catégories sont importantes, Cognos Analytics effectue également un test d'influence t pour détecter les combinaisons de catégories les plus différentes de la moyenne globale. Ce test est similaire à celui utilisé pour un champ explicatif simple. La principale différence est la suivante : au lieu de prendre en compte les catégories d'un champ explicatif unique, Cognos Analytics prend en compte les combinaisons de catégories des deux champs explicatifs. Les combinaisons de catégories dont les tailles d'effet sont les plus élevées sont signalées parmi les différences significatives.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1	Exactement 2 Catégoriel	Facultatif (traiter comme Explicatif)	Facultatif	N/A	Moyenne Force prédictive

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Niveau de récapitulation = Moyenne Continu		Catégoriel			Différences significatives

Somme par un ou deux champs explicatifs

Pour les graphiques dans lesquels le niveau de récapitulation du champ de réponse est Somme et où un ou deux champs explicatifs catégoriels sont disponibles, Cognos Analytics applique le test de comparaison des sommes. Ce test détecte si l'une des sommes est différente de la somme moyenne de toutes les catégories explicatives ou combinaisons de catégories. Si ce test est significatif, Cognos Analytics procède au test d'influence des sommes, qui compare la somme de chaque catégorie ou combinaison de catégories à la somme moyenne. Pour chacun des tests pertinents, Cognos Analytics calcule également la taille d'effet correspondante. Les catégories ou combinaisons de catégories dont les tailles d'effet sont les plus élevées sont renvoyées comme différences significatives.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1 Niveau de récapitulation = Somme Continu	1 ou 2 Catégoriel	Facultatif (traiter comme Explicatif) Catégoriel	N/A	N/A	Moyenne Différences significatives

Minimum ou Maximum par un ou deux champs explicatifs

Pour les niveaux de récapitulation Minimum ou Maximum, seule la connaissance Moyenne est disponible. Elle est calculée comme la valeur moyenne de la valeur minimale ou maximale de la réponse dans toutes les catégories explicatives ou combinaisons de catégories.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1 Niveau de récapitulation = Min ou Max Continu	1 ou 2 Catégoriel	N/A	N/A	N/A	Moyenne

Connaissances dans les visualisations pour deux champs continus

Des connaissances pour deux champs continus sont disponibles lorsqu'une visualisation implique deux champs continus et un champ de points ou de groupe catégoriel facultatif.

Présentation

Utilisez des visualisations comme un nuage de points pour les deux champs continus, éventuellement divisés selon des catégories du champ de groupe. L'objectif principal est de détecter les relations entre les champs continus et d'inclure également le champ de groupe catégoriel. Les résultats contiennent la force prédictive de la relation détectée, la description de la relation fournie par les courbes d'ajustement

et les points qui s'éloignent considérablement des courbes d'ajustement comme différences significatives.

Algorithmes

IBM Cognos Analytics calcule plusieurs modèles de régression qui impliquent l'un des champs continus comme la réponse et les autres champs continus comme un champ explicatif. Le champ de groupe catégoriel facultatif est utilisé comme facteur de modèle. En plus des contributions des modèles additifs correspondant au champ explicatif, Cognos Analytics prend en compte le carré du champ explicatif et tous les termes d'interaction incluant un facteur. Un modèle de régression qui fournit un ajustement optimal des données est sélectionné parmi un certain nombre de modèles possibles. La courbe d'ajustement correspondante est dérivée d'un modèle linéaire ou quadratique. Si un champ de groupe catégoriel facultatif a été fourni, la courbe d'ajustement peut produire une ligne ou une courbe quadratique différente pour chaque catégorie du facteur. Un facteur contenant jusqu'à trois catégories est actuellement pris en compte afin de ne pas surcharger la visualisation.

Chaque point d'une visualisation représente un nombre de lignes dans les données et est défini dans le champ **Points**. Les nombres de lignes correspondants qui sont basés sur le champ de réponse définissent les pondérations de fréquence utilisées pour générer les modèles de régression. Les pondérations de régression sont utilisées indépendamment de la pondération de fréquence lorsque Cognos Analytics génère les modèles de régression.

Détails

Deux zones continues

Lorsque Cognos Analytics applique la régression linéaire multiple pour deux champs continus, l'un de ces champs est choisi comme la réponse et l'autre comme un champ explicatif du modèle. Cognos Analytics prend en compte les termes du modèle linéaire et du modèle quadratique. Si le modèle quadratique est pertinent d'après le test F et que sa force prédictive relative s'est améliorée de plus de 10 % par rapport au modèle linéaire, Cognos Analytics renvoie sa force prédictive et affiche la courbe quadratique à partir du modèle généré. Cette courbe affiche les valeurs prédites de la réponse en fonction des valeurs correspondantes du champ explicatif. Sinon, le modèle de prédicteur linéaire est pris en compte. S'il est pertinent et que sa force prédictive dépasse 10 %, Cognos Analytics renvoie sa force prédictive et affiche une ligne représentant les valeurs prédites du champ de réponse en fonction des valeurs du champ explicatif correspondantes. Si le modèle linéaire n'est pas admissible, la moyenne est renvoyée comme la courbe d'ajustement et aucune relation n'est signalée entre les deux champs continus.

Si une relation linéaire ou quadratique est détectée, Cognos Analytics vérifie également les différences entre les valeurs prédites et observées du champ de réponse. Ces différences sont appelées résidus et Cognos Analytics effectue un test de résidus de Student pour détecter les aberrations. Les points qui s'écartent grandement des relations détectées sont affichés comme différences significatives dans le graphique correspondant.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1	Exactement 1	N/A	Facultatif	Facultatif	Force prédictive
Niveau de récapitulation = tout	Continu		Continu	Tout	Courbe d'ajustement
Continues					Différences significatives

Champ du groupe de champs catégoriels

Si un champ de groupe catégoriel est spécifié en plus des deux champs continus, il est utilisé comme facteur dans la régression linéaire multiple où l'un des deux champs continus est choisi comme le champ de réponse et l'autre comme un champ explicatif. Cognos Analytics prend en compte les termes du modèle linéaire et quadratique pour les champs explicatifs continus associés aux contributions du facteur. Si le modèle quadratique ou le modèle linéaire qui inclut le facteur est significatif d'après le test F et que sa force prédictive relative s'est améliorée de plus de 10 % par rapport au modèle linéaire avec un champ explicatif continu seul, Cognos Analytics génère quatre modèles supplémentaires. Ces modèles incluent toutes les interactions possibles de champ continu explicatif et de facteur. Un modèle avec la valeur maximale du R carré ajusté, également pertinent, est sélectionné comme modèle final. Il est utilisé pour créer une courbe d'ajustement pour chaque catégorie du prédicteur indépendant. Sinon, le modèle linéaire avec un champ explicatif continu est testé pour évaluer sa pertinence et signalé si sa force prédictive est supérieure à 10 %. Si ce n'est pas le cas, aucune relation fiable n'est établie entre les champs et la moyenne globale est renvoyée comme courbe d'ajustement.

Si une relation fiable est détectée, Cognos Analytics recherche également les différences entre les valeurs prédites et observées du champ de réponse. Cognos Analytics effectue un test de résidus de Student pour détecter les aberrations et les afficher comme écarts significatifs dans le graphique correspondant.

Restrictions

Le tableau ci-après décrit les conditions qui déterminent si des connaissances sont suggérées pour cet algorithme.

Réponse	Explicatif	Groupe	Pondération	Points	Connaissance
Exactement 1	Exactement 1	Exactement 1	Facultatif	Facultatif	Force prédictive
Niveau de récapitulation = tout	Continu	Catégoriel	Tout	Tout	Courbe d'ajustement
Continu					Différences significatives

Champ des pondérations de régression

Il est possible d'utiliser un champ continu facultatif afin de spécifier des pondérations de régression pour le modèle. La pondération de régression d'une valeur disponible correspond à l'influence de l'observation sur les paramètres du modèle généré.

Détails en langage naturel

La fonction Détails en langage naturel est une fonction texte qui étend les visualisations affichées avec des récapitulatifs supplémentaires. Cette fonction fournit des connaissances qui sont obtenues à partir d'une analyse de données significative pour vous.

Présentation

La fonction Détails en langage naturel est disponible pour les visualisations qui sont créées dans une exploration. Les détails sont alors affichés sous forme de texte dans le panneau de détails correspondant. Vous pouvez ainsi obtenir les informations les plus pertinentes pour les données consultées dans un format en langage naturel. Des récapitulatifs et des informations supplémentaires non disponibles dans la visualisation sont également présentés.

Algorithmes

Les détails de base fournissent des récapitulatifs simples des données qui ne sont pas immédiatement visibles dans la visualisation affichée. Vous pouvez obtenir ces détails en spécifiant d'autres visualisations connexes, mais de telles étapes d'exploration deviennent dès lors inutiles car les récapitulatifs associés sont disponibles sous forme de texte.

Les détails basés sur les connaissances dans les visualisations décrivent les informations que vous pouvez obtenir dans la boîte de dialogue Connaissances de la visualisation affichée ou des visualisations connexes. Elles clarifient les informations affichées et permettent également d'ajouter des insights supplémentaires qui ne sont pas disponibles dans la visualisation affichée.

Détails

Les détails sont basés sur les champs qui sont affichés dans la visualisation correspondante. L'analyse associée peut s'appuyer sur des récapitulatifs supplémentaires, mais elle n'inclut pas de champs non spécifiés dans la visualisation. Les récapitulatifs et les détails sont convertis en texte traduisible à l'aide de modèles et non d'une fonction complète de génération en langage naturel. Le texte présenté contient donc des répétitions, mais cela ne réduit ni la quantité ni la qualité des informations affichées.

Détails de base en langage naturel

Les détails de base en langage naturel offrent des récapitulatifs supplémentaires des données affichées dans la visualisation ou mettent en évidence les informations disponibles avec des détails supplémentaires.

Présentation

Ces informations fournissent une vue plus globale des données tout en restant pertinentes pour les récapitulatifs affichés que vous étudiez. Si vous modifiez le niveau de récapitulation spécifié pour le champ de réponse, une partie ou l'ensemble des détails de base sont également modifiés car la portée de la visualisation a changé. Des informations supplémentaires sont fournies dans le contexte de la tâche principale.

Algorithmes

Les détails de base calculés dépendent du niveau de récapitulation spécifié pour le champ de réponse dans la visualisation correspondante. Les niveaux de récapitulation possibles pour les réponses sont Nombre (éléments distincts) et Nombre pour tout champ et Somme, Moyenne, Minimum et Maximum pour les champs numériques. Les détails basés sur un nombre sont utilisés pour la plupart des niveaux de récapitulation des réponses, mais des récapitulatifs de correspondance supplémentaires sont fournis avec les niveaux de récapitulation des réponses Somme, Moyenne, Minimum et Maximum. Les récapitulatifs basés sur un même nombre sont générés à la fois pour les niveaux de récapitulation des réponses Nombre et Nombre (éléments distincts).

Détails

Total

Le nombre total est affiché pour la réponse et tout champ explicatif de la visualisation. Ce nombre n'inclut pas les valeurs manquantes de la réponse et il est généré, sauf si le niveau de récapitulation de la réponse est Somme ou Moyenne. Tous les champs catégoriels spécifiés dans la visualisation sont traités comme des champs explicatifs à des fins d'analyse détaillée.

Nombre pour les champs explicatifs

Le nombre est également calculé pour chaque catégorie du champ explicatif affiché, sauf si le niveau de récapitulation de la réponse est Somme. L'algorithme sélectionne et renvoie les principales catégories, les nombres correspondants et les pourcentages des nombres par rapport au nombre total du champ explicatif. Cette procédure est appliquée à chaque champ explicatif du graphique.

Somme

Si le niveau de récapitulation affiché est Somme pour un champ de réponse numérique, IBM Cognos Analytics récapitule la somme totale correspondante pour le champ de réponse. Si la visualisation contient plusieurs champs catégoriels explicatifs, la somme est calculée pour chaque catégorie et chaque champ explicatif. L'algorithme sélectionne et renvoie les principales catégories, les sommes correspondantes et les pourcentages des sommes par rapport à la somme totale de chaque champ explicatif.

Moyenne

Si le niveau de récapitulation affiché est Moyenne pour un champ de réponse numérique, Cognos Analytics récapitule la moyenne générale correspondante pour le champ de réponse.

Plage

Les récapitulatifs des plages sont donnés par le minimum et le maximum lorsque la somme ou la moyenne du niveau de récapitulation est calculée pour un champ de réponse numérique dans les combinaisons de catégories des champs explicatifs spécifiés. Si un seul champ explicatif est spécifié, les catégories qui contiennent le minimum et le maximum sont également affichées.

Minimum et maximum

Si le niveau de récapitulation affiché est Minimum ou Maximum pour un champ de réponse numérique, Cognos Analytics affiche la valeur minimale ou maximale correspondante dans toutes les catégories d'un champ catégoriel explicatif. Si plusieurs champs catégoriels explicatifs sont spécifiés, le minimum ou le maximum est calculé dans toutes les combinaisons de catégories possibles.

Détails de base en langage naturel basés sur les connaissances

Les détails basés sur les connaissances fournissent une description textuelle des connaissances que vous pouvez obtenir à l'aide de la boîte de dialogue Connaissances dans la visualisation.

Présentation

Les détails basés sur les connaissances fournissent également des informations provenant de visualisations connexes, à caractère informatif et faciles à appréhender dans le contexte du graphique en cours. Vous disposez ainsi d'informations plus complètes liées à la visualisation en cours.

Algorithmes

Les détails basés sur les connaissances exploitent directement la gamme complète des calculs et tests statistiques pris en charge par les connaissances. Ils récupèrent également les résultats des visualisations connexes et les compilent dans un message significatif. Par ailleurs, ils fournissent des données d'analyse approfondies basées sur les résultats des connaissances en générant des informations supplémentaires lisibles sous forme de texte.

Détails

Champ explicatif unique

A partir d'un champ de réponse et d'un champ explicatif catégoriel, les détails utilisent les connaissances pour détecter la relation entre le champ de réponse et le champ explicatif. Une analyse est appliquée selon que le niveau de récapitulation de la réponse est Moyenne ou Nombre et la force prédictive est renvoyée si une relation pertinente est détectée.

Si le champ explicatif est numérique et que la fonction Insights génère une courbe d'ajustement, IBM Cognos Analytics renvoie une pente positive ou négative pour la courbe d'ajustement et indique si une relation quadratique est détectée. Si la relation est quadratique, Cognos Analytics renvoie également le point extremum. Cognos Analytics calcule la valeur minimale ou maximale de la réponse et la valeur explicative correspondant à l'extremum de la réponse.

Deux champs explicatifs

Si deux champs explicatifs sont disponibles, Cognos Analytics détecte la relation entre la réponse et les deux champs explicatifs et la relation entre la réponse et chaque champ explicatif séparément. Si la force prédictive de la relation avec les deux champs explicatifs apporte une amélioration relative de plus de 10 % par rapport à chaque relation distincte, Cognos Analytics indique que la réponse est affectée par les deux champs explicatifs. Sinon, la réponse est affectée par un seul champ explicatif ou par chaque champ explicatif séparément, mais pas ensemble.

Différences significatives

Les détails renvoient toutes les différences significatives détectées par les connaissances pour les données utilisées si le niveau de récapitulation de la réponse est Nombre, Moyenne ou Somme. Ces différences sont signalées pour les catégories de chaque champ explicatif séparément, ainsi que pour les combinaisons de catégories de deux champs explicatifs. Des indications sont également fournies pour préciser si des valeurs sont anormalement élevées ou basses.

Arbres de décision

Les détails renvoient la force prédictive d'un arbre de décisions et une liste des champs prédicteurs utilisés pour fractionner les noeuds d'arborescence. Un champ prédicteur dont l'importance de variable est la plus élevée et son pourcentage d'amélioration de l'importance de variable par rapport aux autres champs des arbre de décisions sont renvoyés si le ratio est supérieur à deux. Des détails similaires peuvent être affichés pour un champ dont l'importance de variable est la plus basse.

Détails en langage naturel pour des séries temporelles

Les détails des séries temporelles fournissent des informations de type texte basées sur l'analyse des données de série temporelle et des modèles de prévision correspondants.

Présentation

IBM Cognos Analytics renvoie les détails des séries temporelles pour une visualisation créée dans une exploration chaque fois que les données de la visualisation contiennent une série temporelle unique et qu'un modèle de prévision est généré. Si les données sont appropriées, les informations des séries temporelles sont générées même si la boîte de dialogue **Prévision** ne figure pas dans la visualisation. Lorsque la boîte de dialogue **Prévision** est présente, elle produit le même modèle par défaut que les informations des séries temporelles lors de l'activation. Les points de série temporelle sont triés automatiquement dans l'ordre chronologique à des fins de détection des informations, mais contrairement à la fonction de prévision, les points temporels affichés dans la visualisation ne sont pas triés.

Algorithmes

Les détails des séries temporelles sont basés sur un modèle de lissage exponentiel pour les séries temporelles observées. Les valeurs des séries temporelles observées et les composantes du modèle généré sont utilisés pour créer des informations pour la série temporelle : valeurs inhabituelles, effets saisonniers et informations de tendance. Chaque type d'information dépend d'une combinaison de données différente et de composantes de modèle de lissage exponentiel correspondantes.

Détails

Valeurs inhabituelles

Un modèle de lissage exponentiel fournit une valeur prédite pour chaque point temporel observé. Une valeur prédite à un point temporel représente la prévision avec une étape à l'avance au point temporel précédent. Un intervalle de confiance pour chaque valeur prédite est calculé et utilise la variance de valeur prévue correspondante qui dépend du modèle. Une valeur de série temporelle observée située en dehors de l'intervalle de confiance pour la valeur prévue correspondante basée sur le modèle est considérée comme une *valeur inhabituelle*.

Les valeurs inhabituelles sont détectées en fonction du modèle de lissage exponentiel sélectionné pour la série temporelle. Le niveau de confiance utilisé pour calculer les intervalles de confiance des prévisions est de 99,74 %. Jusqu'à cinq valeurs inhabituelles sont indiquées en répertoriant les points temporels correspondants. Cognos Analytics ne répertorie pas les points par ordre chronologique, mais par ordre décroissant de la distance à partir de l'intervalle de confiance. Les points plus inhabituels sont répertoriés en premier. Les valeurs inhabituelles sont signalées comme anormalement élevées ou anormalement basses, dans la mesure du possible.

Une valeur inhabituelle détectée au dernier point temporel est indiquée séparément. Cela peut signaler que les données sont incomplètes. Par exemple, une valeur récapitulée pour le dernier mois peut refléter les données quotidiennes sur une partie du mois uniquement.

Effets saisonniers

Les informations des *effets saisonniers* indiquent la durée saisonnière d'une série temporelle identifiée par le modèle. La durée saisonnière correspond à une durée fixe d'un schéma saisonnier établi dans la série temporelle. Par exemple, la variation moyenne de la température sur 12 mois établit un schéma annuel. Ces informations fournissent également la force des effets saisonniers et signalent les périodes comportant les valeurs saisonnières les plus élevées et les plus basses.

La durée saisonnière est obtenue à partir du modèle sélectionné. Elle est dérivée de la période saisonnière et de l'intervalle de date ou d'heure qui est indiqué dans les détails statistiques des prévisions. Un modèle saisonnier n'est sélectionné que s'il fournit un ajustement supérieur à tous les modèles non saisonniers. La période saisonnière du modèle saisonnier sélectionné est obtenue en comparant des modèles à plusieurs périodes saisonnières candidates.

Les effets saisonniers apparaissent comme faibles, modérés ou forts selon la valeur de la force calculée. La force des effets saisonniers est calculée sous la forme d'une réduction de l'erreur du modèle par le modèle saisonnier par rapport au modèle non saisonnier correspondant et divisé par l'erreur du modèle non saisonnier. Elle diffère de la force de la saisonnalité renvoyée dans les détails statistiques de prévision où l'écart de précision entre les deux modèles est indiqué.

Les valeurs saisonnières les plus élevées et les plus basses sont calculées en fonction des moyennes des composantes saisonnières sous-jacentes de tous les modèles saisonniers de la série temporelle. Les périodes correspondantes sont indiquées si les valeurs moyennes sont toujours les plus élevées, ou les plus basses, sur la majorité des modèles saisonniers.

Tendance

Les informations de *tendance* indiquent l'orientation positive ou négative globale des valeurs de série temporelle, le cas échéant. Elles indiquent également la force de la tendance.

Les composantes de niveau et de tendance sont toutes les deux extraites du modèle de lissage exponentiel correspondant. Seule la composante de niveau est utilisée si le modèle n'a pas de composante de tendance. Elle définit une courbe de tendance pour les données de série temporelle. La mesure d'association de tau de Kendall et le test statistique correspondant sont ensuite calculés pour la courbe de tendance. Ils détectent une orientation globale positive ou négative des valeurs de série temporelle. Les différentes plages de valeurs de tau définissent le degré de la force indiquée pour la tendance : faible, modéré ou fort.

Pour plus d'informations sur les modèles de lissage exponentiel, voir [«Modèles de prévision»](#), à la page 57.

Relations

Les visualisations des relations dans une exploration sont affichées initialement lorsque vous spécifiez des données à explorer.

Présentation

IBM Cognos Analytics offre un bref aperçu des relations parmi les paires de champs qui s'intéressent à un seul domaine d'intérêt. La visualisation est composée de plusieurs onglets, chacun couvrant un domaine d'intérêt différent. Ces informations s'avèrent très utiles pour vous orienter dans une multitude de relations pertinentes disponibles dans des données à explorer davantage, si nécessaire.

Algorithmes

Le domaine d'intérêt initial est déterminé à partir de l'analyse sémantique des données, mais vous pouvez spécifier un autre domaine d'intérêt. Chaque onglet fournit un diagramme réseau avec des champs comme noeuds et des liens entre les paires de noeuds qui représente la force relative de la relation entre les noeuds. Les liens du domaine d'intérêt dominant le graphique, mais les autres paires de champs

associées dont les relations sont fortes sont également affichées. Vous pouvez ajuster un curseur pour afficher un nombre plus ou moins grand de noeuds dans le réseau.

Détails

Données à analyser

Les relations utilisent des données non récapitulées pour calculer la force de la relation parmi toutes les paires de champs prises en compte. Pour normaliser la mesure de la force de la relation et la rendre comparable entre toutes les paires de champs, tous les champs numériques sont regroupés dans la première étape. Tous les champs des données sont traités comme catégoriels. Le regroupement appliqué est un regroupement par fréquences égales qui génère cinq catégories. Vous trouverez des détails supplémentaires dans la section sur la préparation des données pour les champs numériques.

Force de la relation

Les données de chaque paire de champs catégoriels sont d'abord mises de type Tableau pour toutes les combinaisons de catégories de champs qui se trouvent dans les données. A partir des données de type Tableau, IBM Cognos Analytics applique le test d'indépendance du khi-carré pour déterminer si les champs sont indépendants. Si l'écart d'indépendance est important, Cognos Analytics calcule la taille d'effet à partir de la statistiques du khi-carré. Il s'agit du V de Cramer, largement utilisé comme mesure d'association entre deux champs catégoriels. Les valeurs de cette mesure sont comprises entre 0 et 1 et Cognos Analytics renvoie la valeur de la force de la relation, exprimée sous forme de pourcentage. Les relations dont la force est inférieure à 10 % ne sont pas signalées car elles sont considérées comme trop faibles pour présenter un intérêt pratique.

Limitations de performances

Le calcul de la force de la relation entre toutes les paires de champs du jeu de données est prohibitif pour les données volumineuses. Cognos Analytics limite à 100 le nombre de champs traités pour pouvoir fournir une réponse rapide. Toutefois, ces champs sont sélectionnés par un autre processus et la perte possible de relations pertinentes est minimisée. Si les données contiennent plus de 10000 lignes, Cognos Analytics obtient un échantillon aléatoire de cette taille pour des raisons de performances. Cette taille de données garantit une perte minimale de précision de l'estimation de la force de la relation.

Différences entre Cognos Analytics et Watson Analytics en matière d'analyse avancée des données

Les méthodes d'analyse avancée de Watson Analytics ont été adaptées et améliorées dans IBM Cognos Analytics. Par conséquent, un certain nombre de différences peuvent être observées.

Préparation des données

La préparation des données est essentielle dans Cognos Analytics et IBM Watson Analytics car seules des données préparées sont entrées dans les analyses pour les facteurs clés et les arbres de décisions. La préparation des données diffère entre les deux offres sur plusieurs points importants.

Différence	Cognos Analytics	Watson Analytics
Valeurs manquantes	Les valeurs manquantes dans le champ cible sont traitées de la même manière dans les deux produits ; les lignes pour lesquelles la valeur est manquante dans le champ cible sont exclues de toutes les analyses ultérieures. En revanche, les valeurs manquantes sont traitées différemment pour les champs prédicteurs.	Attribution en fonction de la relation entre la cible et un prédicteur. Watson Analytics exclut les champs prédicteurs qui comportent plus de 25 % de valeurs manquantes.

Différence	Cognos Analytics	Watson Analytics
	<p>Une catégorie distincte pour les prédicteurs indépendants.</p> <p>Les prédicteurs numériques sont toujours regroupés et les valeurs manquantes sont traitées comme une catégorie distincte.</p>	
Fusion de catégories	<p>Cognos Analytics classe les catégories par fréquence d'occurrence dans les données et regroupe toutes les catégories commençant par 50^{ème} dans une même catégorie. Toutes les catégories dont la fréquence est inférieure à trois sont également fusionnées. Les catégories fusionnées doivent représenter moins de 25 % des lignes de données, faute de quoi, le champ n'est pas utilisé pour l'analyse. Les catégories cible sont également fusionnées.</p>	<p>La fusion des catégories pour les prédicteurs dépend des valeurs cible. Seules les catégories contenant un nombre suffisant de valeurs de champ cible sont fusionnées.</p>
Traitement des aberrations	<p>Les aberrations dans les champs cible continus sont traitées en définissant des valeurs limites extrêmes et en remplaçant les valeurs au-delà de ces limites par les valeurs limites correspondantes à des fins d'analyse. L'analyse s'en trouve améliorée à la fois pour les facteurs clés et les arbres de décisions car cela facilite la détection des relations avec la cible. Les aberrations traitées lors de la préparation des données prédictives ne sont pas signalées dans la visualisation.</p>	<p>Le traitement automatisé des aberrations n'est pas disponible.</p>
Echantillonnages	<p>Cognos Analytics analyse des échantillons de données qui contiennent environ 10000 lignes si les données d'origine dépassent cette taille. Un avertissement est affiché chaque fois que l'analyse correspondante est effectuée sur un échantillon de données, mais que la force prédictive des facteurs clés et des arbres de décision générés est censée être proche de la force prédictive des modèles obtenus sur les données complètes correspondantes.</p>	<p>Watson Analytics utilise l'ensemble complet des lignes des données pour générer les facteurs clés et les arbres de décisions. Cela peut avoir un impact négatif sur les performances.</p>

Facteurs clés

Différence	Cognos Analytics	Watson Analytics
Facteurs clés	<p>Cognos Analytics offre des facteurs clés pour les cibles continues et catégorielles dans la visualisation en spirale et la visualisation d'analyse de facteur.</p> <p>Les étapes de préparation des données affectent à la fois les facteurs clés unidirectionnels et bidirectionnels, mais l'analyse de Cognos Analytics pour les facteurs clés bidirectionnels diffère de celle de Watson Analytics en ce qui concerne les critères utilisés pour sélectionner les facteurs à présenter à l'utilisateur.</p>	<p>Les facteurs clés des cibles continues et catégorielles sont disponibles dans la visualisation en spirale.</p>
Tests statistiques	<p>Pour les cibles catégorielles, Cognos Analytics utilise le test d'indépendance du khi-carré pour déterminer si deux champs catégoriels sont indépendants.</p>	<p>Watson Analytics utilise le test de rapport de vraisemblance pour les cibles catégorielles.</p>
Critère de sélection	<p>Cognos Analytics requiert que la force prédictive du facteur clé bidirectionnel apporte une amélioration relative de plus de 10 % par rapport à celle des facteurs unidirectionnels correspondants. L'amélioration relative de la force prédictive est mesurée par le pourcentage de réduction de l'erreur du facteur clé bidirectionnel par rapport à l'erreur minimale des deux facteurs clés unidirectionnels correspondants. Ce critère garantit que seuls les facteurs bidirectionnels pertinents apparaissent.</p>	<p>Watson Analytics utilise l'interaction bidirectionnelle comme critère de sélection pour les facteurs clés bidirectionnels. Un facteur clé bidirectionnel doit posséder une taille d'effet d'interaction supérieure à un certain seuil pour être sélectionné.</p>
Force prédictive	<p>Pour les graphiques en spirale et les graphiques d'analyse des facteurs avec une cible continue, la force prédictive correspond au R carré ajusté.</p> <p>La force prédictive de la cible catégorielle correspond au R carré ajusté des nombres. Cette valeur est obtenue en calculant l'amélioration de la précision de la classification du modèle par rapport au modèle de constante et en la divisant par l'erreur de classification du modèle de constante. Le modèle de constante prévoit toujours le mode cible et son exactitude de classification est évaluée par la fréquence du mode. Une relation prévisible fiable est indiquée lorsque la force prédictive du modèle est supérieure au seuil par défaut de 10 %.</p>	<p>La force prédictive d'une cible continue correspond au R carré.</p> <p>La force prédictive de la cible catégorielle est calculée comme la précision de classification.</p>

Différence	Cognos Analytics	Watson Analytics
	En conséquence, la force prédictive d'une visualisation en spirale avec une cible catégorielle peut être différente dans Watson Analytics et Cognos Analytics, même pour une cible et des entrées comparables.	

Arbre de décisions

Différence	Cognos Analytics	Watson Analytics
Arbre de décisions	Outre les visualisations Arborescence et Règles de décision, Cognos Analytics fournit la nouvelle visualisation Rayon. Les différences dans la génération d'un arbre de décisions dans Cognos Analytics ou Watson Analytics dépendent du fait que la cible soit continue ou catégorielle. La force prédictive des arbres de décisions est affichée dans les visualisations en spirale et les visualisations d'analyse des facteurs dans Cognos Analytics, mais uniquement si elle est supérieure à la force prédictive du premier facteur clé.	Les visualisations Arbres de décisions et Règles de décision sont disponibles. La force prédictive de l'arbre est toujours disponible dans la visualisation en spirale.
Cible continue	Les arbres de décisions d'une cible continue sont générés à partir de la somme des carrés dans Watson Analytics et Cognos Analytics. Cognos Analytics utilise directement le gain d'informations associé à la taille d'effet pour le test F dans Watson Analytics, lorsqu'il fusionne et fractionne des noeuds d'arborescence. Cela fournit des fractionnements multi-directionnels pour les noeuds des arbres de décisions.	Watson Analytics utilise l'algorithme CHAID et les tests F correspondants lorsqu'il fusionne et fractionne des noeuds d'arborescence. Cela fournit des fractionnements multi-directionnels pour les noeuds des arbres de décisions.
Règles d'arrêt	Cognos Analytics limite le nombre de noeuds d'arborescence à 36. Le nombre minimal de lignes correspondant aux noeuds d'arborescence est 25. Les champs dont la force prédictive ne contribue pas sont supprimés des arborescences Cognos Analytics. Cognos Analytics génère des arborescences plus petites qui conviennent mieux à la visualisation et aux connaissances tandis que des arborescences d'une précision	Watson Analytics limite la taille de l'arborescence en limitant le nombre de niveaux d'arborescence à cinq. Le nombre minimal de lignes correspondant aux noeuds d'arborescence est 50. Les arborescences dans Watson Analytics n'ont tendance à être plus précises que si elles contiennent un nombre bien plus important de noeuds.

Différence	Cognos Analytics	Watson Analytics
	comparable, voire supérieure, à celle de Watson Analytics sont générées.	
Importance de variable	L'importance de variable correspond à l'amélioration relative de la force prédictive lorsque la variable est ajoutée à l'arbre de décisions. Les prédicteurs dont l'importance de variable est inférieure ou égale à un seuil défini sont exclus de l'arbre de décisions. Ce seuil est fixé à zéro pour la plupart des arborescences. Ce seuil est légèrement augmenté pour les arborescences basées sur un nombre plus petit de lignes de données et dont la force prédictive est plus faible pour réduire les effets de surajustement possibles.	L'importance de variable est calculée en fonction de la sensibilité des résultats lorsque la variable accepte des valeurs différentes.
Force prédictive	<p>La force prédictive des arbres de décisions avec une cible continue correspond au R carré.</p> <p>La force prédictive de la cible catégorielle correspond au R carré ajusté des nombres.</p> <p>Le R carré ajusté des nombres soustrait l'erreur de classification d'arborescence de l'erreur de classification de base obtenue en sélectionnant toujours la catégorie cible la plus fréquente et en divisant le résultat par l'erreur de classification de base. Il représente l'amélioration relative obtenue par l'arbre de décisions. En conséquence, la force prédictive des arbres de décisions avec une cible catégorielle peut être différente dans Watson Analytics et Cognos Analytics, même pour des arbres de décisions comparables.</p>	<p>La force prédictive des arbres de décisions avec une cible continue correspond au R carré.</p> <p>La force prédictive pour la cible catégorielle correspond à la précision de classification.</p>

Commandes utilisateur

Différence	Cognos Analytics	Watson Analytics
Commandes utilisateur	<p>Vous spécifiez un champ comme continu en associant les propriétés de données Usage à Mesure ou comme catégoriel (nominal) en les associant à Identificateur ou Attribut.</p> <p>Permet d'exclure et d'inclure des prédicteurs directement dans les visualisations correspondantes par l'intermédiaire de la boîte de dialogue d'édition des facteurs.</p>	<p>Vous spécifiez les champs à utiliser comme continus, nominaux ou ordinaux en définissant les propriétés de données appropriées dans Affiner.</p> <p>Permet d'exclure et d'inclure des prédicteurs dans les modèles par l'intermédiaire d'une autre propriété de données Optimiser.</p>

Chapitre 6. Assistant

Assistant

IBM Cognos Analytics inclut un assistant intégré qui prend en charge les entrées textuelles pour vous aider à analyser rapidement vos données et à simplifier les données analytiques. Quelques étapes suffisent pour accéder à des sources de données clés, créer des visualisations et les faire glisser dans le canevas Exploration ou Tableau de bord. La saisie de texte est prise en charge uniquement en anglais.

Vous pouvez utiliser deux interfaces :

1. Panneau complet de l'assistant
2. Panneau compact de l'assistant

Panneau complet de l'assistant

Le panneau complet de l'assistant comprend trois sections : barre d'outils, sortie et zone d'entrée. Vous pouvez exécuter des actions dans la barre d'outils, telles que la réinitialisation de votre historique. La sortie est affichée dans une zone de défilement verticale qui inclut toutes des réponses antérieures. La zone d'entrée appelée par défaut **Poser une question** permet d'entrer un texte sous la forme d'une conversation.

Ouvrez la vue complète du panneau en cliquant sur l'icône **Assistant** . Si vous utilisez le panneau compact dans **Exploration** uniquement, vous pouvez accéder au panneau complet en cliquant sur **Ouvrir en vue complète**.

Panneau compact de l'assistant

Le panneau compact est disponible uniquement dans l'interface **Exploration**. Vous pouvez ouvrir le panneau compact en cliquant sur **Poser une question** dans la zone de ce panneau.

Le panneau compact affiche uniquement le dernier texte entré et la réponse associée. Pour afficher l'ensemble de l'historique, cliquez sur le lien **Ouvrir en vue complète** dans le panneau ou sur l'icône

Assistant .

Dans la zone **Poser une question** vous pouvez cliquer sur la flèche du clavier vers le haut ou vers le bas pour afficher une liste des commandes recommandées. Lorsque vous entrez du texte, une fonction de saisie anticipée propose des suggestions. Cette fonction est une excellente façon de formuler des questions appropriées et cela évite de taper au clavier.

Pour répéter votre dernière question, cliquez sur la bulle de réponse ou utilisez le raccourci clavier (Maj + flèche vers le haut), puis appuyez sur Entrée. Vous pouvez parcourir vos questions précédentes en appuyant sur Maj + flèche vers le haut/bas.

Les visualisations qui apparaissent dans le panneau **Assistant** sont condensées et peuvent exclure certaines informations. Vous pouvez afficher les détails de la visualisation dans **Explorer** ou **Tableau de bord**. A partir de l'**Assistant**, vous pouvez faire glisser des visualisations sur un canevas **Exploration** ou **Tableau de bord**.

The screenshot shows the IBM Cognos Analytics interface. On the left, a sidebar contains navigation icons and a list of fields: 'reet Name', 'undred Block', 'ocal Area', 'ase Type', and 'inute'. A button labeled 'show Hour and Local Area' is visible. Below this, a visualization titled 'Hour and Local Area' is shown, featuring a word cloud and a legend for 'Local Area' with categories like Marpole, Grandvi..., Kensing..., Sunset, Renfrew..., Kerrisdale, Shaugh..., Kitsilano, Fairview, Riley Park, Dunbar..., and Arbutus... On the right, a table displays data for 'Hour' and 'Local Area'.

Hour	Local Area
0	Grandview-Woodland
	Kensington-Cedar Cottage
	Mount Pleasant
1	Arbutus Ridge
	Hastings-Sunrise

Pour effacer votre historique de conversations et le restaurer dans son état initial, cliquez sur l'icône

Réinitialiser  dans la barre d'outils. Si vous entrez reset ou clear, le résultat est le même. L'icône **Réinitialiser** apparaît dans le panneau complet uniquement ; elle n'est pas disponible dans le panneau compact.

Important : L'actualisation de l'onglet du navigateur Cognos Analytics réinitialise également l'historique.

Fonctions de l'Assistant

L'Assistant est une fonction puissante qui vous permet de poser des questions en langage naturel. La présente rubrique décrit certaines des commandes que vous pouvez utiliser avec l'assistant.

Posez des questions en utilisant le langage naturel. Par exemple, montrer le Bénéfice est synonyme de répertorier le Bénéfice, qu'est-ce que le Bénéfice ?, plus d'informations sur le Bénéfice et Bénéfice.

Voici quelques questions fréquemment posées pour démarrer :

help

Affiche une présentation et des informations d'utilisation générales.

show data

Répertorie toutes les sources de données prises en charge auxquelles vous avez accès dans **Mon contenu** et **Contenu de l'équipe**. Les sources de données prises en charge sont les suivantes :

- Fichiers téléchargés (csv, tsv, xls,xlsx et zip)
- Les modules de données
- Ensemble de données
- Des packs Framework Manager

Remarque : Reportez-vous à la rubrique relative à *l'enrichissement des packs*, dans le document *Gestion*.

Lorsque les résultats dépassent le nombre d'éléments affichés, cliquez sur **Plus** pour afficher jusqu'à 100 sources de données. Faites défiler la liste vers le bas et cliquez sur **Moins** pour réduire la liste.

show source <nom_source>

Affiche une liste de zones et de détails pertinents pour la source de données indiquée. La barre d'outils affiche la source <nom_source> utilisée dans le contexte. Par défaut, Cognos Analytics utilise la source active dans votre panneau de tableau de bord ou d'exploration. A partir des résultats, vous pouvez cliquer sur les noms de zone pour obtenir plus d'informations sur ces zones. Cliquer sur des noms de zone revient à exécuter la commande `show column <nom_colonne>`.

Lorsque les résultats dépassent le nombre d'éléments affichés, cliquez sur **Plus** pour afficher jusqu'à 100 sources de données. Faites défiler la liste vers le bas et cliquez sur **Moins** pour réduire la liste.

show column <nom_colonne>

Pour la colonne spécifiée, les informations et les zones associées s'affichent. Cliquer sur les zones associées équivaut à demander `show chart <colonne1> et <colonne2>`. Si la colonne spécifiée est destinée à avoir des influenceurs, vous pouvez entrer `what influences <nom_colonne>` pour afficher sa liste d'influenceurs.

what influences column <nom_colonne>

Affiche une liste de zones qui influence les résultats dans la colonne spécifiée.

show chart <colonne1> and <colonne2>

Affiche les visualisations qui représentent la relation entre <colonne1> et <colonne2>. Faites défiler les visualisations en cliquant sur les flèches gauche et droite. Chaque visualisation inclut une icône d'information dans le coin supérieur droit. Survolez l'icône pour afficher les descriptions des données sous-jacentes. Vous pouvez éventuellement entrer d'autres colonnes, mais un nombre de colonnes excessif peut rendre les visualisations moins efficaces.

Si vous cliquez sur `Show related visualizations`, les visualisations sont renvoyées en fonction des champs d'influence et des champs connexes.

Si vous cliquez sur `Create dashboard from the charts`, un tableau de bord basé sur les graphiques les plus récemment générés est créé. Le même tableau de bord est créé en saisissant `Create related dashboard`. Si les graphiques contiennent des agrégations supérieures ou inférieures, ces modificateurs sont appliqués au tableau de bord généré.

L'application d'agrégations et de filtres peut améliorer la précision et permettre de créer des visualisations plus convaincantes. Les agrégations courantes comprennent le total, la moyenne, le nombre, le maximum/minimum, le haut/bas, le meilleur/pire, etc. Voici quelques exemples d'agrégations :

show top <num> <colonne1> par <colonne2>

Affiche les valeurs supérieures à partir de <colonne1> en fonction du contexte de <colonne2>. Par exemple, afficher les 5 meilleures ventes par région. Si <num> n'est pas spécifié, la valeur par défaut est 10.

<colonne1> est une mesure agrégée ou non agrégée, alors que <colonne2> est une colonne catégorielle.

show average <nom_colonne>

Affiche la moyenne de toutes les valeurs trouvées dans <nom_colonne>.

how many <nom_colonne>

Si <nom_colonne> est une catégorie, le nombre d'éléments distincts est renvoyé. Si <nom_colonne> est une mesure, la somme totale est renvoyée.

show maximum <nom_colonne>

Affiche la valeur la plus élevée trouvée dans <nom_colonne>.

show minimum <nom_colonne>

Affiche la valeur la plus faible trouvée dans <nom_colonne>.

show total <nom_colonne>

Affiche le total de la somme pour toutes les valeurs trouvées dans <nom_colonne>.

Vous pouvez ajouter des filtres pour des chaînes géographiques (telles que Pays ou Etat) ou des chaînes temporelles (telles que Mois ou Année). Le filtrage par date n'est pas pris en charge pour le moment. Les agrégats et les filtres peuvent être combinés pour produire des résultats plus granulaires. Voici quelques exemples, basés sur des échantillons de données :

- show Education by Income where Income is less than 1000
- show Education by Income where Income > 100K

Vous pouvez utiliser K (pour dénoter des milliers) ou M (pour dénoter des millions).

- show Revenue in 2017 and 2018
- show Income by Month for New York City
- what are the top 5 States by average Inventory, excluding California

Les visualisations filtrées comprennent une icône de filtre (), située dans le coin supérieur droit du graphique. Survolez l'icône pour afficher le filtre appliqué.

create dashboard

Génère un nouveau tableau de bord basé sur la source de données sélectionnée. Vous pouvez modifier les visualisations, les onglets, les commandes, etc. et enregistrer votre nouveau tableau de bord. Par défaut, le tableau de bord généré inclura des analyses avancées et des graphiques prédictifs.

La génération automatique de tableaux de bord pour des sources de données plus importantes peut entraîner des problèmes de performances. Pour contourner ce problème, vous pouvez entrer `create simple dashboard` pour générer un tableau de bord de base. Vous pouvez ensuite modifier le tableau de bord en remplaçant les graphiques par des visualisations plus sophistiquées, telles que l'analyse des pilotes ou les diagrammes en spirale.

