

Jointures entre bases de données

Bienvenue dans cette vidéo sur les jointures entre bases de données.

Vous pouvez télécharger les ensembles de données pour suivre les étapes dans votre propre version de Tableau.

Des données provenant de différentes sources

Pour bon nombre d'entreprises, les données dont elles disposent sont réparties dans différents systèmes. Par exemple, des données financières dans une base de données SQL Server et des données sur les produits dans Amazon Redshift. Les données peuvent être hébergées dans plusieurs environnements. Mais nous souhaitons effectuer des analyses sur l'ensemble des données.

Pour cette vidéo, nous disposons de deux sources de données basées sur fichier : une au format Excel et une autre au format CSV. Nous utilisons uniquement des fichiers plats afin de faciliter la compréhension.

Dans le cas présent, nos données de ventes disposent d'un ID produit mais aucune autre information sur le produit vendu. Quant à nos données de produit, elles se présentent sous forme de tableau comportant l'ID produit, la catégorie, la sous-catégorie et le nom du produit.

Il s'agit d'une configuration idéale pour une jointure entre bases de données. Si nous effectuons une jointure gauche des données de produit vers les données de vente en fonction de l'ID produit, nous ajouterons ces informations à chaque transaction de vente pertinente. Pour plus d'informations sur les jointures, reportez-vous à la vidéo sur les types de jointures.

Jointures entre bases de données

Pour créer cette source de données intégrée à l'aide d'une jointure entre bases de données,

- Commençons par nos ventes : cliquez sur Excel, puis ouvrez le fichier Sales 2016.
 - Afin de nous y retrouver, double-cliquons sur cette table et renommons-la « Ventes »
- Nous allons maintenant ajouter une *connexion* de données vers cette source.
 - Dans la zone Connexions, cliquez sur le bouton Ajouter, mais PAS sur le bouton Nouvelle source de données dans la barre d'outils.
 - Choisissez Ajouter, puis Fichier texte
 - Et ouvrez le fichier Products 2016.
- La feuille a déjà été ajoutée à l'espace de travail.
 - Renommons-la simplement « Produits ».
- Donnons maintenant un nom plus précis à la source de données en indiquant « Ventes et produits 2016 ».
- Tout comme une jointure au sein d'une même source de données, nous pouvons modifier cette jointure : cliquez sur l'icône, puis sélectionnez une jointure gauche.
- Nous constatons que nos données ont été jointes au niveau des lignes avec ID produit.

Dans la grille, nous pouvons voir que les informations sur le produit (en jaune) ont été ajoutées à chaque ligne à partir des données de ventes (en bleu).

Utilisation de sources de données intégrées

Maintenant que nous avons une nouvelle source de données unique, essayons-la.

- Nous constatons que nous avons une seule source de données dans le volet Données, mais que nos champs sont décomposés par table, tout comme avec une jointure au sein d'une même source de données.
 - Si besoin, nous pouvons désactiver cette option, en cliquant avec le bouton droit et en choisissant, par exemple, Regrouper par dossier au lieu de Regrouper par table de sources de données

- Et créer une hiérarchie pour les sous-catégories et les catégories.
 - Ajoutez ID produit et Nom du produit.
 - Nous pouvons masquer l'ID produit de la table CSV, car il ne sert que pour la clause JOIN.
- Déplaçons Catégorie vers Lignes, et développons.
- Déplaçons Ventes vers Colonnes.
- Et décomposons par date de commande.

Une source de données intégrée fonctionne de la même manière que toute autre source de données à source unique : elle peut même être publiée et enregistrée, contrairement aux fusions.

Conclusion

Merci d'avoir suivi cette vidéo sur les jointures entre bases de données. Nous vous invitons à découvrir les autres vidéos de formation gratuite pour en apprendre davantage sur l'utilisation des produits Tableau.