

Fusion de données

Bienvenue dans cette vidéo consacrée à la fusion de données. Vous pouvez télécharger le classeur d'exercices pour suivre les étapes dans votre propre logiciel Tableau.

Qu'est-ce que la fusion de données ?

Imaginons que vous disposiez de deux chaînes de magasins, Office City et Coffee Chain. Vous stockez vos données correspondant à Office City dans un système et celles correspondant à Coffee Chain dans un autre. La fusion des données permet de combiner, dans une même vue, les données de plusieurs sources de données. Au lieu d'effectuer la jonction des données au niveau des lignes, comme une jointure entre bases de données, la fusion des données envoie des requêtes distinctes aux sources de données et agrège les résultats à un niveau commun dans Tableau.

Les relations de fusion

Deux sources de données sont présentes dans le volet Données : Office City et Coffee Chain. Vous pouvez facilement passer de l'une à l'autre. Notez cependant comme les dimensions et les mesures varient. La fusion des données nécessite au moins un champ commun dans les deux sources de données, par exemple, État. Si les champs n'ont pas le même nom, mais possèdent des membres en commun, vous pouvez définir la relation manuellement.

À titre d'exemple, nous savons que les champs Région d'Office City et Territoire de Coffee Chain contiennent tous deux les valeurs Centre, Est, Sud et Ouest. Nous pouvons donc les définir comme étant équivalents. Accédez au menu Données et sélectionnez Modifier les relations ; Office City peut alors devenir la source principale. Sélectionnez ensuite Personnalisé et ajoutez une relation : nous choisissons Région et Territoire.

Notez que Tableau a établi une relation entre les deux champs et indique aussi la relation automatique : État. Vous pouvez également renommer le champ d'une source de données pour qu'elle corresponde à l'autre. Cliquez une première fois, puis une deuxième, et saisissez Région à la place de Territoire pour Coffee Chain. Si les deux champs portent le même nom, Tableau crée automatiquement une relation entre les deux sources de données.

Maintenant que les champs communs sont définis, nous pouvons commencer la fusion.

Comment réaliser une fusion

Commencez par sélectionner la source de données Office City. Faites glisser Ventes d'Office City vers l'étagère Colonnes et État vers l'étagère Lignes. Dans le volet Données, une coche bleue est apparue en regard d'Office City. Lorsque vous êtes connecté à plusieurs sources de données dans Tableau, la première que vous utilisez dans la vue devient la source de données principale. Elle est alors identifiée par cette coche bleue. Passons à la source de données secondaire. Notez l'icône de liaison orange en regard d'État. Étant donné qu'État est déjà dans la vue, Tableau effectue la fusion automatiquement, ce que confirme le lien orange.

Un lien gris brisé apparaît également en regard du champ Région. Région étant un champ commun aux deux sources, il s'agit d'un autre champ de liaison potentiel qui n'est pas utilisé dans la vue actuelle. Si vous voulez effectuer une fusion en fonction de ce champ, cliquez sur le lien gris pour l'activer et créer la relation. Terminons la fusion en faisant glisser Ventes de Coffee Chain vers l'étagère Colonnes. Une coche orange est désormais visible en regard de la source de données Coffee Chain et sur la pile dans la vue. Cet élément signifie qu'il s'agit de notre source de données secondaire.

Comment cette vue est-elle créée ? Tableau interroge chaque source de données sur la somme des ventes par État, puis affiche les valeurs dans une même vue. Les informations concernant les ventes de chaque État proviennent d'Office City, car Tableau renvoie les informations pour tous les membres de champ (États) présents dans la source de données principale, que la source de données secondaire dispose ou non de données.

Inversion des sources de données principale et secondaire

Il est important de noter que la source principale et la source secondaire sont déterminées feuille par feuille. Autrement dit, cette priorité n'est pas préservée dans l'ensemble du classeur. Lorsque vous vous trouvez sur une nouvelle feuille, les sources de données dans le volet Données n'affichent aucune coche orange ou bleue qui différencie la source principale de la source secondaire. Les relations établies sur la feuille de calcul précédente ne sont pas conservées.

Que se passe-t-il si nous inversons les sources de données et utilisons Coffee Chain comme source principale au lieu d'Office City ? Commencez par déplacer Ventes de Coffee Chain vers la vue, puis utilisez État. À présent, la vue comporte environ la moitié des États seulement, par rapport à l'exemple précédent. La raison est simple : l'ensemble de données de Coffee Chain comporte moins d'États qu'Office City. Tableau affiche tous les États de Coffee Chain, puis extrait uniquement les informations pertinentes d'Office City. Cela signifie que les États d'Office City qui n'existent pas dans Coffee Chain ne seront pas représentés dans la vue.

Conclusion

Merci d'avoir suivi cette vidéo sur la fusion des données. Nous vous invitons à regarder la vidéo présentant des aspects supplémentaires de la fusion de données, ainsi que nos autres vidéos de formation gratuite pour en apprendre davantage sur l'utilisation des produits Tableau.