

Atelier Power BI

Dans cette manip nous allons effectuer de la manipulation de données et de la réalisation de tableau de bord. Le jeu de données qui sera utilisé concerne les incidents recensés par le **San Francisco Police Department**.

Ce jeu de données comprend la catégorie d'incidents, la date et le lieu de celui-ci et sa résolution.

La source de données est disponible dans le lien ci-dessous :

<https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry>

Pour commencer , veuillez télécharger le jeu de données `police-department-incidents.csv`.

Exercice 1

NETTOYAGE ET MANIPULATION DES DONNÉES

Charger les données issues du fichier `<<police-department-incidents.csv >>`

Aller dans l'éditeur de requête pour commencer la mise en forme des données. Cela va nous permettre de nettoyer nos données avant de créer notre rapport.

La colonne `<< Date >>` indique une heure `<< 00:00:00 >>`.

Modifier la colonne Date pour ne garder que la date sans l'heure.

Les données après 2015 étant incomplètes, conserver uniquement les données avant le **31/12/2014**.

Compter les valeurs en sélectionnant la colonne d'identifiant (**IncidentNum**).

Puis supprimer l'étape en cliquant sur la croix à droite de l'étape.

Supprimer maintenant les doublons : clic droit sur la table entière en haut à gauche puis supprimer les doublons.

Compter de nouveau les valeurs. Puis supprimer cette étape.

Ajouter une nouvelle colonne correspondant à la colonne « Address » en lettres minuscules (`<< adresse_min >>`). Puis supprimer la colonne « Address » initiale

On souhaite rajouter une colonne « **adresse_complete** » correspondant à la concaténation de << **adresse_min**>> et « **PdDistrict** ».

Associer à chaque ligne le salaire moyen du District correspondant.
Nommer la << **revenu_moyen** » selon les informations suivantes :

PARK	120000
RICHMOND	110000
NORTHERN	90000
TENDERLOIN	84000
TARAVAL	80000
MISSION	70000
INGLESIDE	70000
BAYVIEW	65000
SOUTHERN	60000
CENTRAL	40000

Vous travaillez pour le **San Francisco Police Department**. Vous êtes responsable du **service kidnapping** et vous devez réaliser un rapport pour votre hiérarchie concernant le travail effectué en 2014. Ce rapport devra inclure plusieurs graphiques cependant vos chiffres ne sont pas très bons. Vous allez donc miser sur vos talents en Data Visualisation pour vous en sortir auprès de votre chef.

On veut mettre l'accent sur le nombre de **Kidnappings** en comparaison avec d'autres incidents. Comment faire pour minimiser le nombre de kidnappings et donner l'impression que ce sont parmi les incidents moins nombreux ?

Insérer un graphique pour suivre les évolutions des kidnappings selon les années.

On constate que les nouveaux chiffres sont nettement supérieurs aux années précédentes, modifiez la visualisation de ce graphique pour que cela ne soit pas trop pénalisant vis-à-vis de votre chef.

Réaliser un graphique en courbe et histogramme groupés pour visualiser le nombre d'enlèvements par district comparé avec le revenu moyen d'un habitant du quartier (Catégorie en valeur de colonnes et **revenu_moyen** en valeurs de lignes) :

Exploiter la fonctionnalité Q&A de Power BI pour vérifier vos résultats.