

POWER BI – INITIATION

Dans ce TD 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 le TD, veuillez télécharger le jeu de données « police-department-incidents.csv », qui se situe dans le lien drive ci-dessous :

<http://bit.ly/SISE-POWERBI>

Exercice 0

• NETTOYAGE ET MANIPULATION DES DONNEES

1. Charger les données issues du fichier « police-department-incidents.csv » [ACCUEIL / OBTENIR LES DONNEES]
2. 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. [ACCUEIL / MODIFIER LES REQUETES]
3. La colonne « Date » indique une heure « 00 :00 :00 ». Modifier la colonne Date pour ne garder que la date sans l'heure. [TRANSFORMER / TYPE DE DONNEES / DATE / AJOUTER UNE NOUVELLE ETAPE]
4. Les données après 2015 étant incomplètes, conserver uniquement les données avant le 31/12/2014. [FILTRE CHRONOLOGIQUE / AVANT LE...]
5. Compter les valeurs en sélectionnant la colonne d'identifiant (IncidntNum). Puis supprimer l'étape en cliquant sur la croix à droite de l'étape. [TRANSFORMER / STATISTIQUES / COMPTER LES VALEURS]
6. 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.
7. Ajouter une nouvelle colonne correspondant à la colonne « Address » en lettres minuscules (« adresse_min »). [AJOUTER UNE COLONNE / FORMAT / minuscules]
Puis supprimer la colonne « Address » initiale

8. On souhaite rajouter une colonne « adresse_complete » correspondant à la concaténation de « adresse_min » et « PdDistrict ». [AJOUTER UNE COLONNE / COLONNE PERSONNALISE] & permet la concaténation entre 2 champs
9. Supprimer la colonne « PdDistrict » puis remonter cette étape vers le haut dans l'enchaînement des étapes appliquées. [CLIC DROIT SUR L'ETAPE / DEPLACER VERS LE HAUT] Il est possible de descendre ou de remonter des étapes dans cet enchaînement. De plus à chaque fois que l'on clique sur une étape on visualisera l'état de la mise en forme à cette étape de la requête.
Cliquez sur la dernière étape (« personnalisée ajoutée ») et comprendre le message d'erreur obtenu pour la colonne « adresse_complete ».
Supprimer ensuite cette étape de suppression de colonne « PdDistrict »
10. Modifier l'étape de concaténation des colonnes « Address » et « PdDistrict » en double-cliquant sur l'étape.
Essayer de concaténer la colonne « adresse_complete » avec la colonne « IncidntNum » : on obtient une erreur.
Modifier le type de données (Texte) de la colonne « IncidntNum » pour pouvoir concaténer des variables de même type.
Faire attention à l'enchaînement chronologique des étapes.
11. Associer à chaque ligne le salaire moyen du District correspondant. [AJOUTER UNE COLONNE / COLONNE CONDITIONNELLE]
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

12. On souhaite, sur notre rapport, associer des valeurs issues d'une autre source de données. Cela va créer une autre table.
Utiliser une autre source de données : site web [ACCUEIL / NOUVELLE SOURCE / WEB]
<http://sanfranciscopolice.org/police-district-maps>

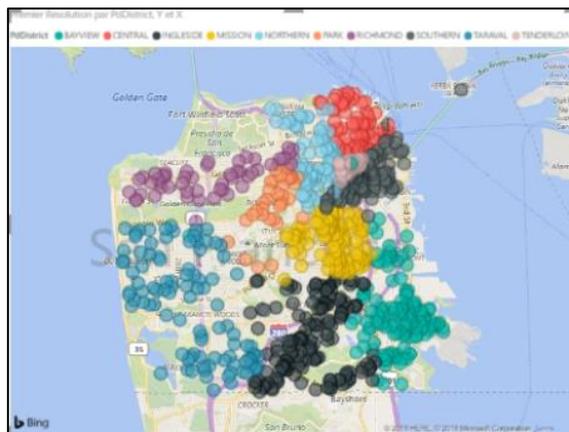
Power BI va repérer qu'il y a un tableur sur la page.

Exercice 1

• VISUALISATION DES DONNEES

Vous travaillez pour le San Francisco Police Department. Vous êtes en charge 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 Visualization pour vous en sortir auprès de votre chef.

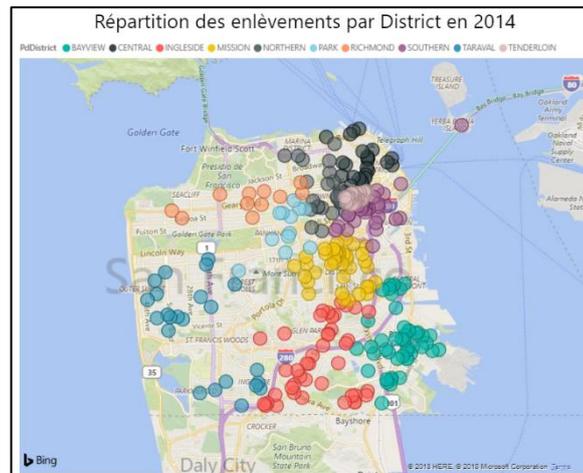
1. Dans l'éditeur de requête, cliquer sur Fermer et appliquer : cela aura pour but d'appliquer les requêtes sur les données chargées
2. Localiser sur une carte les incidents de kidnapping reportés
Attention X est la longitude tandis que Y est la latitude
Ajouter « Resolution » en info-bulle et filtrer au niveau de l'élément avec « Category ».
On peut différencier des filtres sur l'élément visuel, sur la page ou sur le rapport (l'ensemble des pages du rapport).



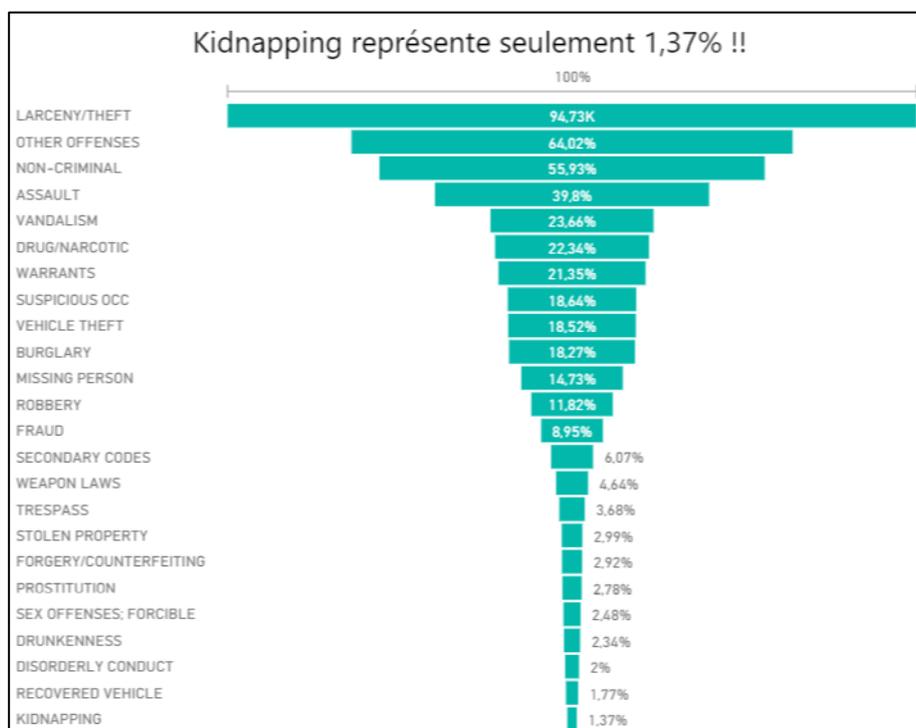
Répartition des enlèvements par District

3. Nous souhaitons filtrer que pour l'année 2014. Pour cela nous allons retourner dans l'éditeur de requêtes.
Il est possible de naviguer entre Power Query et Power BI Desktop pour modifier notre requête ensuite notre rapport de visualisation s'actualisera avec les modifications réalisées.
Dans l'éditeur de requêtes, dupliquer la colonne de date et ne garder que l'année [CLIC DROIT / DUPLICATION DE LA COLONNE / TRANSFORMER / DATE / ANNEE]. Renommer cette colonne « Année ».
Retourner dans Power BI Desktop en faisant « fermer et appliquer ».
Ajouter « Année » dans le filtre sur l'élément. [TYPE DE FILTRE / FILTRE DE BASE]

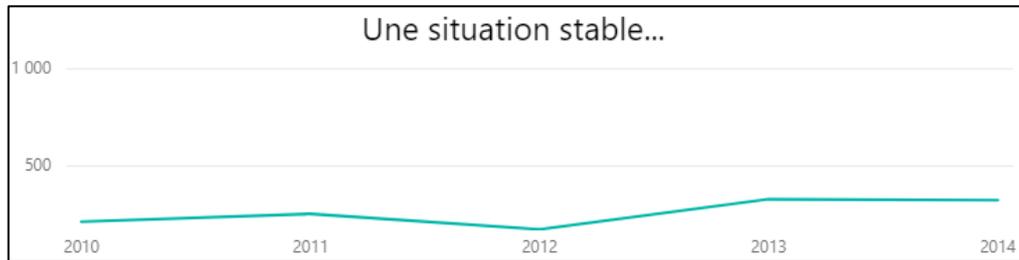
- Changer le titre de la cartographie en « Répartition des enlèvements par District en 2014 »
[SYMBOLE ROULEAU PEINTURE (FORMAT) / TITRE]



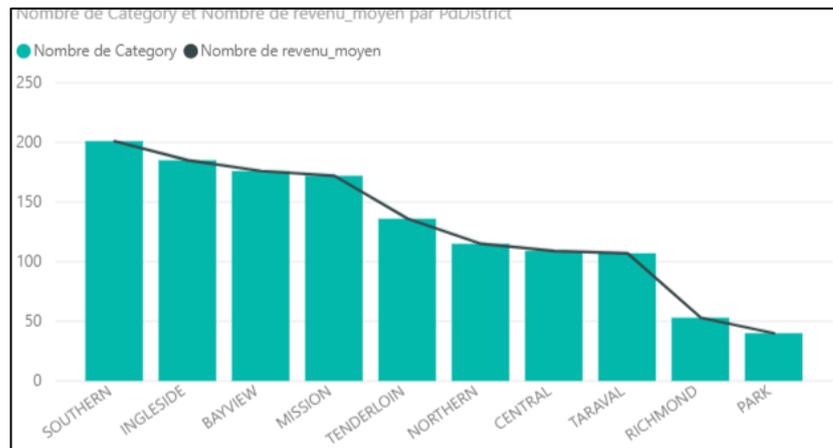
- On veut mettre l'accent sur le nombre de kidnapping en comparaison avec d'autres incidents. Pour minimiser le nombre de kidnapping et donner l'impression que ce sont parmi les incidents les moins nombreux nous décidons de faire un **graphique en entonnoir** avec un filtre judicieux pour avoir la représentation suivante :



6. Insérer un graphique pour suivre les évolutions des kidnappings selon les années. Vos chiffres semblent bien supérieurs aux années précédentes, modifions la visualisation de ce graphique pour que cela ne soit pas trop pénalisant vis-à-vis de votre chef ! Jouer sur les échelles pour « aplanir » la hausse des enlèvements.



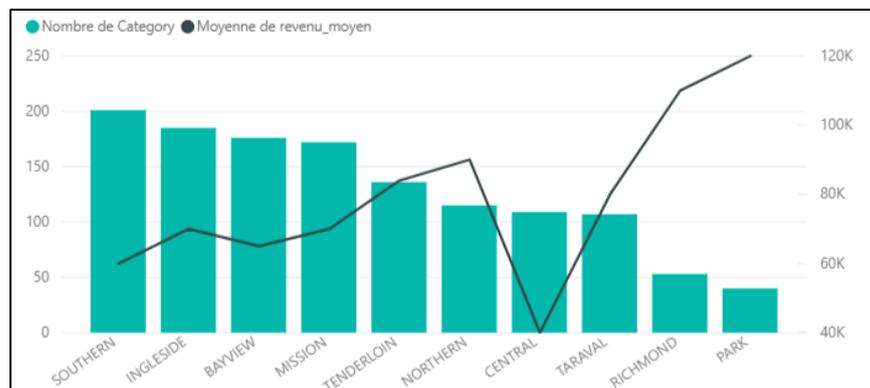
7. Réaliser un « graphique en courbe et histogramme groupés » pour visualiser le nombre d'enlèvements par district comparé avec le revenu moyen d'un habitant du quartier (« Category » en Valeurs de colonnes et « revenu_moyen » en Valeurs de lignes) :



On n'obtient pas ce que l'on souhaite avec en valeur de ligne « nombre de revenu_moyen ». Cela est dû au fait que la colonne « revenu_moyen » que l'on a ajouté est en type de données « n'importe lequel ».

Il faut donc retourner dans l'éditeur de requêtes pour détecter le type de données (nombre entier) et appliquer la modification.

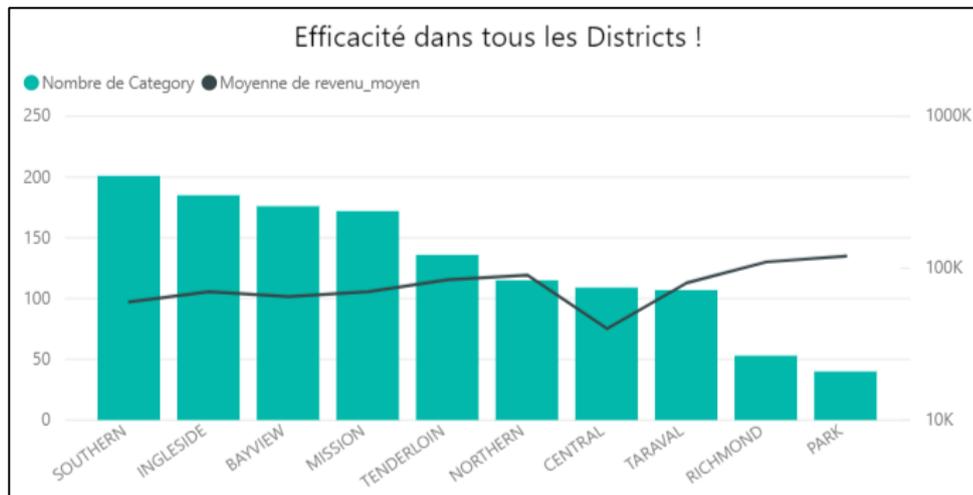
Afficher la moyenne et non le nombre de « revenu_moyen ».



8. On peut déduire que plus le revenu moyen est élevé, moins les enlèvements sont nombreux !

Cela n'est pas bon à communiquer car il ne faudrait pas donner l'impression que votre service est plus actif dans les districts aisés.

Pour atténuer cette impression, choisir une échelle logarithmique pour l'axe Y des lignes afin de « lisser » cette tendance.



9. Cliquer sur « Poser une question »



« Poser une question » vous permet de poser une question en anglais et avoir la réponse chiffrée.

Essayer « How many Kidnapping »

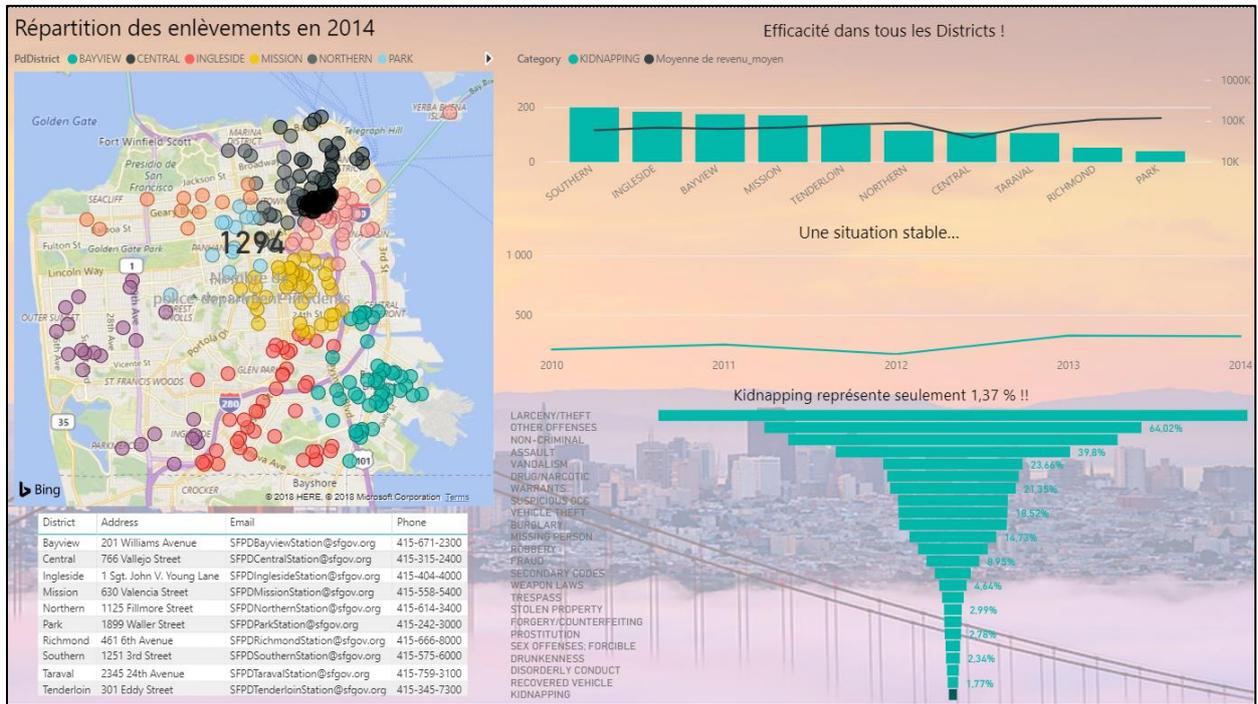
Sur le graphique « Efficacité dans tous les Districts ! » cliquer sur les différentes barres correspondant aux différents quartiers pour voir l'évolution du chiffre.

10. Incorporer à votre page les données issues du site internet

Pour cela, ajouter un tableau et sélectionner les 4 champs du jeu de données

11. Ajouter le fond d'écran de votre choix !

Bravo vos talents en Data Visualization vous ont permis de vous en sortir face à votre chef ! Vous avez même reçu des félicitations sur votre travail de cette année !



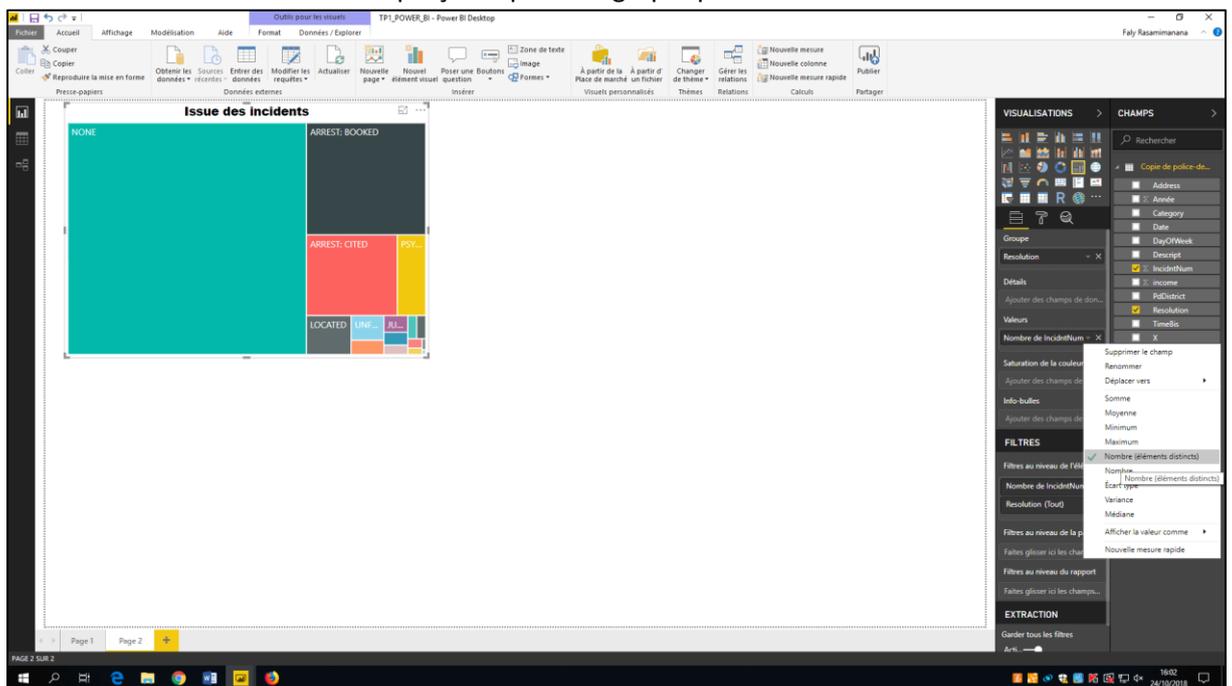
Exercice 2

Supposons maintenant que vous êtes maire de San Francisco. Vous avez les moyens financiers d'allouer d'avantages de ressources à la lutte contre les violences. Votre objectif est d'avoir moins de 80 000 actes de violences par an et d'augmenter la part de personnes arrêtées et envoyées en prison à 20 %. Vous décidez de construire un tableau de bord pour l'année 2014 pour visualiser si les objectifs ont été atteints et, si ce n'est pas le cas, visualiser vers quel district et à quelle période de la semaine devront être alloué le plus de ressources.

1. Créer un graphique **Tree map** en haut à gauche du rapport afin de pouvoir visualiser les résolutions d'incidents. Vous lui donnerez le titre suivant : "Issue des incidents". L'alignement devra être centré et le texte de taille 12. Ajouter un filtre sur la page selon l'année et sélectionner uniquement l'année 2014.

Attention : quand vous préciserez l'onglet valeur, vous veillerez à ce que la mesure ne soit pas agrégée et qu'elle comptabilise le Nombre d'éléments distincts.

Aperçu du premier graphique :



2. Créer un **Histogramme empilé 100 %** en haut à droite du rapport afin de pouvoir visualiser l'issue des incidents (« Resolution ») en fonction du district dans lequel l'incident a lieu. Vous ajouterez une ligne de constante de couleur rouge sur le graphique au seuil de 20%. Cela constitue le niveau d'incidents pour lequel on souhaite que le coupable soit arrêté et condamné (ARREST ; BOOKED). Vous lui donnerez le titre suivant : "Résolution des incidents par district". L'alignement devra être centré et le texte de taille 12, Police : Arial Black.

- Créer un **Histogramme empilé** en bas à gauche du rapport indiquant le nombre d'incident par jour de la semaine. Vous lui donnerez le titre suivant : "Incidents par jour de la semaine". L'alignement devra être centré et le texte de taille 12, Police : Arial Black. Vous veillerez à ce que l'échelle s'étende de 0 à 15 000. Vous ferez varier la couleur des barres en spécifiant l'option « saturation de la couleur ». Vous ferez en sorte que les valeurs les plus élevées soient en rouge et les valeurs les plus faibles en vert.
- Créer une **Carte à plusieurs lignes** qui indique le nombre d'incidents qui ont mené à l'arrestation et la condamnation du coupable pour chaque année. Pour cela vous utiliserez un filtre sur l'élément qui ne retiendra que les résolutions « ARREST ; BOOKED » et « NONE ». Vous lui donnerez le titre suivant : "Chiffres clés". L'alignement devra être centré et le texte de taille 12, Police : Arial Black. Vous ferez apparaître les chiffres du graphique en rouge. Vous ajouterez une bordure à ce graphique.
- Créer une **jauge** qui indiquera le nombre d'incidents. Vous veillerez à ce que la jauge s'étende de 0 à 100 000. Vous fixerez une valeur cible de 80000. Vous lui donnerez le titre suivant : "Nombre d'incidents par rapport à l'objectif annuel". L'alignement devra être centré et le texte de taille 12, Police : Arial Black.

Répondez aux questions suivantes :

- L'objectif de ne pas dépasser 80 000 incidents est-il atteint ?
- Quelles sont les régions pour lequel l'objectif de 20% n'est pas atteint et donc pour lesquels nous devrions allouer plus de moyens ?
- Quelles sont les jours de la semaine pour lesquels nous devrions allouer le plus de moyen ?

Résultat final :

