

1. **Le nombre de musées en France est suffisant pour couvrir toute la population ?**
2. **Choix des sources de données**

Parmi les source mise en disposition nous avons trouvés judicieux pour notre analyse de se basé sur les fichiers :

* **Fréquentations-des-musees-de-france.json**
* **Listes des communes de France Wikip‚dia.html**

En plus de cela nous avons recherche une autre source donnée nous mettant à disposition l'ensemble des villes de France et leurs populations pour une analyse le plus proches possible de la réalité. La source provient de :

* **Populations communales de 1876 à 2021 :** <https://www.insee.fr/fr/statistiques/3698339>

1. **Identification des tables**

Après sélection des données pertinentes pour notre analyse, nous avons identifié les différentes tables.

* **Table de fait :** Fréquentations de musée
* **Table de dimension :** Musées, Ville, Département, Région, Date

1. **Principales anomalies rencontrées**

Après identification des sources de données a utilisées et des tables qui en ressortait, nous avons constatés plusieurs anomalies importantes, notamment :

* **Absences d'éléments de jointure directe entre les tables Fact\_Frequentation\_Musee et Dim\_Musee**: Une remarque directe nous permettait de constater qu'on aurait pu se servir de la référence musée qui était présente dans les 2 tables mais la qualité de cette colonne dans la table **Fact\_Frequentation\_Musee** nous montrait la présence de 86% de valeurs null pour ce champ, l'utiliser revenais à perdre 86% de données. Pour resoudre ce problème, nous avons choisi de basé directement sur la table **Fact\_Frequentation\_Musee** pour extraire les données de la table **Dim\_Musee** qui a donné des résultats plus cohérents. Pour ressortir la clé propre à **Dim\_Musee,** nous avons fait une fusion entre le nom de la ville du musée et le nom du musée car plusieurs musées possédant un même nom mais de ville différente.
* **Incohérence dû au nombre d'habitant total en France de depuis la source web** "**Liste des communes de France les plus peupl‚es  Wikip‚dia.html"** (23 millions) par rapport aux estimations réelles. Raison qui nous a pousser à recherche une source de donnée externe.
* **Absence d'élément de jointure entre la table Dim\_Ville et Dim\_Musee:** Nous avons utilisé le nom de la ville comme clé de jointure.

1. **Traitement des données**
2. **Fact\_Frequentation\_Musee**

Une image contenant texte, capture d’écran, nombre, menu

Description générée automatiquementNous avons exporté les données du fichier **frequentation-des-musees-de-france.json,** puis depuis ceux-ci nous avons effectué les modifications suivantes :

* **Colonnes Supprimées**
* "datasetid"
* "recordid"
* D'autres colonnes supprimées incluent "fields.note", "fields.date\_appellation", "fields.observations", "record\_timestamp", et "fields.ref\_musee".
* **Filtrage des Lignes**
* Suppression des lignes avec une fréquentation nulle ou zéro.
* **Renommage de Colonnes**
* "fields.ville" renommée en "Ville"
* "fields.nomdep" en "Departement"
* "fields.total" en "Total frequentation"
* "fields.regions" en "Region"
* "fields.annee" en "Annee"
* "fields.payant" en "frequentation\_payantes"
* "fields.nom\_du\_musee" en "Nom"
* "fields.gratuit" en "frequentation\_gratuite"
* **Duplication et Modification de Colonnes pour Créer une Clé Étrangère**
* Duplication des colonnes "Ville" et "Nom" en "Ville - Copier" et "Nom - Copier" et transformation en minuscules et les espaces superflus supprimés.
* Fusion de "Ville - Copier" et "Nom - Copier" en une nouvelle colonne "Fk\_musee".

1. **Dim\_Musée**

Une image contenant texte, capture d’écran, nombre, menu

Description générée automatiquement

Nous avons utilisé les données provenant du fichier **frequentation-des-musees-de-france.json,** puis depuis ceux-ci nous avons effectué les modifications suivantes :

* **Colonnes Supprimées**
* "datasetid"
* "recordid"
* "fields.note"
* "fields.ref\_musee"
* "fields.id\_museofile"
* "fields.date\_appellation"
* "fields.observations"
* "record\_timestamp"
* "payant"
* "gratuit"
* **Renommage de Colonnes**
* "fields.ville" en "Ville"
* "fields.nomdep" en "Departement"
* "fields.total" en "total\_frequentation"
* "fields.regions" en "Region"
* "fields.annee" en "annee"
* "fields.nom\_du\_musee" en "Nom"
* **Duplication et Transformation de Colonnes pour Créer des Clés**
* Duplication de "Ville" et "Nom" et transformation en minuscules et suppression des espaces pour créer une clé primaire "Pk\_musee".
* Duplication de "Ville" pour créer une clé étrangère "Fk\_ville" en minuscules et sans espaces.
* **Requêtes Fusionnées**
* Fusion de la table actuelle avec "Dim\_Ville sur la base de la clé étrangère "Fk\_ville". (Les données de villes entre 'Dim\_Ville' et 'Dim\_Musee' avaient des discordances, avec 400 villes non correspondantes initialement. Une jointure externe gauche sur 'Dim\_Musee' a réduit ce nombre à 29, améliorant ainsi l'alignement des données.)
* **Colonnes Supprimées et Renommées Post-Fusion**
* Suppression de la colonne "Fk\_ville" originale après la fusion.
* Renommage de "Dim\_Ville.Pk\_ville" en "Fk\_ville" pour refléter la clé étrangère.
* **Filtrage des Lignes**
* Filtrage des lignes pour ne conserver que les données de l'année 2021. (Avoir la liste de tous les musées en 2021.)

1. **Dim\_Ville**

Une image contenant texte, capture d’écran, nombre, Police

Description générée automatiquement

Nous avons utilisé les données provenant de la source externe **Populations communales de 1876 à 2021,** puis depuis ceux-ci nous avons effectué les modifications suivantes :

* **Colonnes Supprimées**
* Plusieurs colonnes relatives aux populations des années précédentes, telles que "PMUN2020", "PMUN2019", ..., jusqu'à "PTOT1876".
* **Renommage de Colonnes**
* "REG" renommée en "Fk\_region" (indiquant une clé étrangère vers une table de régions).
* "DEP" en "Fk\_dep" (clé étrangère vers une table de départements).
* "LIBGEO" en "Ville".
* "PMUN2021" en "Population".
* **Duplication et Modification de Colonne pour Créer une Clé Primaire**
* Duplication de la colonne "Ville" en "Ville - Copier".
* Renommage de "Ville - Copier" en "Pk\_ville" (indiquant une clé primaire pour la ville).
* La colonne "Pk\_ville" est transformée en minuscules et les espaces superflus sont supprimés pour assurer l'unicité et la cohérence.
* **Modification du Type de Données**
* Le type de la colonne "Population" est converti en entier 64 bits pour une précision numérique appropriée.
* **Filtrage des Lignes**
* Suppression des doublons basée sur la colonne "Pk\_ville" pour assurer l'unicité de chaque ville.

1. **Dim\_Region**

Une image contenant texte, capture d’écran, nombre, logiciel

Description générée automatiquement

Nous avons utilisé les données provenant de la source html **Listes des communes de France Wikip‚dia.html,** puis depuis ceux-ci nous avons effectué les modifications suivantes :

* **Filtrage des Lignes**
* Suppression des lignes où le numéro de département est nul ou égal à "Total". (Ligne en trop on avait le total de communes par dep et de dep par region)
* **Suppression de Doublons**
* Elimination des doublons basés sur la colonne "no Région".
* **Colonnes Supprimées**
* "no Dpt" (numéro de département)
* "Département"
* "Nombre de communes"
* **Modification du Type de Données**
* Conversion du type de la colonne "no Région" en texte.
* La colonne "no Région" modifiée pour traiter correctement les numéros de région.
* **Remplacement de Valeurs**
* Remplacement du texte "Département et région d'outre-mer (DROM)" par "DROM" dans la colonne "Région".
* **Renommage de Colonnes**
* "no Région" renommée en "Pk\_region" (indiquant une clé primaire pour la région).
* "Région" renommée en "Region".

1. **Dim\_Departement**

Une image contenant texte, capture d’écran, nombre, logiciel

Description générée automatiquement

Nous avons utilisé les données provenant de la source html **Listes des communes de France Wikip‚dia.html,** puis depuis ceux-ci nous avons effectué les modifications suivantes :

* **Filtrage des Lignes**
* Suppression des lignes où le numéro de département est nul ou égal à "Total". (Ligne en trop on avait le total de communes par dep et de dep par region)
* Suppression de colonne "Nombre de communes"
* **Renommage de Colonnes**
* "no Région" renommée en "Fk\_region" (indiquant une clé étrangère vers une table de régions).
* "no Dpt" en "Pk\_dep" (indiquant une clé primaire pour le département).
* "Département" en "Departement".
* **Modification du Type de Données**
* Le type de la colonne "Pk\_dep" est converti en texte pour assurer une manipulation correcte des numéros de département.

1. **Indicateurs clés**

Comme indicateurs clés pour notre analyse, nous avons calculer en utilisant le DAX les valeurs de :

1. **Nombre d'habitant**



Cette mesure permet d'avoir le nombre d'habitant total en France. Ce qui nous permettra de faire des observations sur la population sur des axes tels que Région, Ville ou encore département en France.

1. **Nombre de musé**



Cette mesure permet d'avoir le nombre musée total en France. Ce qui nous permettra de faire des observations sur les musées sur des axes tels que Région, Ville ou encore Département en France.

1. **Ratio musée par habitant**





Cette mesure est un indicateur qui mesure la densité relative de musées par rapport à la population d'une région donnée. Il est généralement exprimé sous la forme d'un nombre qui représente le nombre de musées par unité de population, souvent par millier ou par cent milliers d'habitants.

1. **Axes d'analyse**

Pour répondre à la question de savoir si le nombre de musées en France est suffisant pour la population, nous nous service principalement des indicateurs **Nombre d'habitant, Nombre de musées et Ratio musée par habitant.** Il en ressort que :

* Si on observe une région avec une population importante mais un ratio musé par habitant relativement faible, cela peut indiquer une disponibilité relativement faible de musées par habitant dans cette région en prenant en compte l'échelle national
* Une région densément peuplée avec un faible ratio musées par habitant pourrait indiquer un besoin potentiel en termes d'infrastructures culturelles.

Nous les graphiques suivants permettent de montrés notre analyse d'un point de vue visuel.

Une image contenant texte, capture d’écran, Caractère coloré, diagramme

Description générée automatiquement*Figure : Population et musées en France sur une disposition régionales*

Une image contenant texte, capture d’écran, diagramme, Tracé

Description générée automatiquement*Figure : Population et accessibilité Culturelle en France sur les axes régionaux et urbain*

1. **Vision pour les grandes villes et régions et résultat général**

A l'aide de ces graphique (Figure 2), la couche en bleu représente la population et en orange le ratio musé par habitant.

* **Par rapport aux régions :** Allant de Île-de-France à Grand-Est (Les grandes régions), on remarque que le ratio est en dessous de la population ce qui démontre qu'il faut de musée dans ces régions par rapport aux autres régions du pays moins denses.
* **Par rapport aux villes :** Allant de Paris à Rennes (Les grandes villes), on remarque que le ratio est en dessous de la population ce qui démontre qu'il faut plus de musée dans ces régions par rapport aux autres villes du pays moins dense en population.

**Pour les grandes villes et régions il est nécessaire d'augmenter le nombre de musées.**

On peut conclure que **le nombre de musée en France n'est pas suffisant pour la population**. 

1. **Solution BI**
2. **Deux architecture BI**

1. **Modélisation**

Une image contenant texte, capture d’écran, Rectangle, diagramme

Description générée automatiquement*Figure : Modélisation*

Nous avons opté pour une modélisation en **flocon** ayant pour table de fait la table Fact\_Frequentation\_Musee.