[Date]

[Titre du document]

[Sous-titre du document]

1. Expliquer pourquoi c’est nécessaire et pourquoi ce n’est pas nécessaire
2. Démontrer que les données sont pertinentes pour l’analyse
3. Ex : on est au niveau mondial donc les communes n’ont pas de sens
4. Population totale = somme population par région

**ETUDE DE CAS 1**

1. **IDENTIFIONS LES SOURCES PERTINENTES A L’ANALYSE :**

Notre analyse va permettre de vérifier si le nombre de musée est suffisant pour couvrir la population. Pour cela, nous avons identifiées les sources ci-dessous.

* 1. ***Les sources pertinentes à notre analyse sont :***
* Musée de France : permet d’avoir les informations importantes sur les musées comme la référence du musée, le nom officiel du musée, la région, la ville, le département, la répartition géographique du musée. Elle représente une table de dimension « **DIM\_Musée** » dans notre analyse avec pour clé primaire pk\_id\_museofile.
* Fréquentation des musées : permet d’avoir les informations sur les musées fréquentées (Fk\_id\_museeofile), fréquentation totale des musées (fields.total), la ville et région de chaque musée fréquenté, la fréquentation gratuit, fréquentation payante, le nom du musée. Elle représente une table de fait « **Fact\_frequentaion-des-musees-de-France** » avec pour clé primaire pk\_id\_frequentation.
* Population de 283 villes : permet d’avoir les informations sur villes et les communes de plus de 30.000 habitants avant 2024 et les communes ayant compté 30.000 habitants puisque notre étude est au niveau de toute la population de France, donc on va s’intéresser aux villes. Pour chaque ville, nous avons le département associé, la région et sa population de 1968 à 2021. Elle représente une table de dimension « DIM\_Ville » avec pour clé primaire Pk\_CodeInsee ».
  1. ***Sources complémentaires :***
* Données sur la population de chaque région de France. Cette source de donnée me permettra en plus d’avoir ma population totale. Elle représente une table de dimension « DIM\_Région » avec pour clé primaire pk\_code\_region »
* Table Date : pour avoir les jours, mois, trimestre, semestre et années. Elle représente la table de date de notre analyse avec pour clé primaire « pk\_date\_appellation »
  1. Les sources qui ne sont pas pertinentes à l’analyse
* Liste des communes de France : nous donne les informations sur les communes de France par régions ce qui n’est pas pertinent pour notre analyse car nous nous intéressons à la population totale qui ne peut être obtenue qu’à partir des villes.
* Liste des sites remarquables : cette source n’est pas utile à notre analyse car elle donne les sites patrimoniaux qui ne sont pas les musées et n’entre pas dans notre contexte d’analyse qui traite essentiellement sur les musées et la population de France.

1. **LES TYPES D’ANOMALIES SUR CHAQUE TABLE ET LE TRAITEMENT REALISE**
2. Anomalies et Traitements réalisés sur la table « FACT\_frequentation-des-musées-de-France »

* **Anomalie**

Colonne « datasetid » n’est pas pertinente pour l’analyse

* **Traitement réalisé** :

Suppression de la colonne datasetid.

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

Description générée automatiquement

* **Anomalie**

Colonne « recordid » n’est pas pertinente pour l’analyse.

* **Traitement réalisé** :

Suppression de la colonne recordid

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

Description générée automatiquement

* **Anomalie**

Colonne « record\_timestamp» n’est pas pertinente pour l’analyse.

* **Traitement réalisé** :

Suppression de la colonne record\_timestamp.

Une image contenant texte, nombre, logiciel, Police

Description générée automatiquement

* **Anomalie**

Duplication de la valeur d’une région de France : Ile-De-France et Île-de-France sont présentes à la fois

* **Traitement réalisé** :

Remplacer la valeur Ile-De-France par Île-de-France pour n’avoir qu’une seule.

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

Description générée automatiquement

* **Anomalie**

Duplication de la valeur d’une région de France « ST PIERRE ET MIQUELON », « Saint-Pierre-et-Miquelon » sont présentes à la fois.

* **Traitement réalisé** :

Remplacer la valeur ST PIERRE ET MIQUELON par Saint-Pierre-et-Miquelon pour n’avoir qu’une seule.

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

Description générée automatiquement

* **Anomalie**

La table FACT\_frequentation-de-musee n’a pas de clé primaire

* **Traitement réalisé** :

Création d’une colonne « index » pour faire office de colonne de clé primaire.

Une image contenant texte, capture d’écran, nombre

Description générée automatiquement

1. Anomalies et Traitements réalisés sur la table « DIM\_Musee»

* **Anomalie**

Présence du symbole « | » sur les noms des régions entrainant un mauvais orthographe de celles-ci comme suit : Hauts|||||-de-||||France, Villefranche|||||-de-||||Lonchat, Val|||||-de-||||Moder, Île|||||-de-||||France, Hauts|||||-de-||||France

* **Traitement réalisé**

Remplacer la symbole « | » par le vide

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

Description générée automatiquement

* **Anomalie**

Mauvais orthographe du nom officiel du musée écrit comme suit : musèèèé

* **Traitement réalisé**

Remplacement de la valeur erronée « èèèé » par « ée »

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

Description générée automatiquement

* **Anomalie**

Présence du symbole « | » sur les noms officiels de musée.

* **Traitement réalisé**

Remplacer la symbole « | » par le vide

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

Description générée automatiquement

* **Anomalie**

Mauvaise orthographe du nom des villes écrit comme suit à travers la valeur « èèè ».

* **Traitement réalisé**

Remplacement de la valeur erronée « èèè » par « é.

Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

Description générée automatiquement

* **Anomalie**

Présence des colonnes telles que datasetid, recordid et geolicalisation qui ne nous servent pas dans notre analyse.

* **Traitement réalisé**

Suppression des colonnes datasetid, recordid et geolicalisation.

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

Description générée automatiquement

* **Anomalie**

La table DIM\_Musée n’a pas de clé primaire

* **Traitement réalisé** :

Renommer la colonne fields.ref en pk\_id\_museofile pour faire office de colonne de clé primaire.

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

Description générée automatiquement

* **Anomalie**

Présence des colonnes telles que geometry.type", "geometry.coordinates.0", "geometry.coordinates.1" qui ne nous servent pas dans notre analyse.

* **Traitement réalisé**

Suppression des colonnes geometry.type", "geometry.coordinates.0", "geometry.coordinates.1".

Une image contenant texte, nombre, ligne, logiciel

Description générée automatiquement

1. Anomalies et Traitements réalisés sur la table « DIM\_Ville »

* **Anomalie**

Présence des crochets sur les colonnes des années de la table DIM\_Ville.

* **Traitement réalisé**

Remplacer les crochets par les valeurs des années.

Une image contenant texte, nombre, ligne, Police

Description générée automatiquement

* **Anomalie**

Colonne des années ont été pivotées.

* **Traitement réalisé**

Dépivoter les colonnes des années de telle sorte à avoir une colonne pour toutes les années et une autres pour les populations.

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

Description générée automatiquement

* **Anomalie**

Présence du caractère « [c] dans la colonne commune.

* **Traitement réalisé**

Remplacer la valeur [c] dans la colonne commune par le vide.

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

Description générée automatiquement

* **Anomalie**

Présence du caractère « [e] dans la colonne commune.

* **Traitement réalisé**

Remplacer la valeur [e] dans la colonne commune par le vide.

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

Description générée automatiquement

* **Anomalie**

Présence du caractère « [g] dans la colonne commune.

* **Traitement réalisé**

Remplacer la valeur [g] dans la colonne commune par le vide.

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

Description générée automatiquement

* **Anomalie**

Présence du caractère « [d] dans la colonne commune.

* **Traitement réalisé**

Remplacer la valeur [d] dans la colonne commune par le vide.

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

Description générée automatiquement

* **Anomalie**

Présence du caractère « [f] dans la colonne commune.

* **Traitement réalisé**

Remplacer la valeur [f] dans la colonne commune par le vide.

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

Description générée automatiquement

1. Anomalies et Traitements réalisés sur la table « DIM\_Region »

* **Anomalie**

Absence de la clé primaire dans la table DIM\_Region.

* **Traitement réalisé**

Nomenclature de la colonne code région en « pk\_code\_region »

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

Description générée automatiquement

1. L’onglet qualité de chaque table utilisée

* Onglet qualité de la table « Fact\_frequentation-de-musse »

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

Description générée automatiquement

* Onglet qualité de la table « DIM\_Musee »

Une image contenant texte, nombre, Police, ligne

Description générée automatiquement

* Onglet qualité de la table « DIM\_Ville»

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

Description générée automatiquement

* Onglet qualité de la table « DIM\_Date »

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

Description générée automatiquement

* Onglet qualité de la table « DIM\_Region »

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

Description générée automatiquement

1. **Réalisation des calculs en DAX**

Dans notre analyse nous avons identifiés plusieurs calculs qui nous permettront de répondre à la problématique posée qui est de savoir si le nombre de musée est suffisant pour couvrir la population.

1. Nombre de musées : NombreMusee = COUNT(DIM\_musees[PK\_id\_museofile])

Une image contenant capture d’écran, texte, logiciel

Description générée automatiquement

* **Explication** : Nombre de musée permet d’avoir le nombre de musée total en France pour toutes les régions. Pour se faire nous avons compter les clés primaires des musées dans la table DIM\_Musee.
* **Résultat souhaité**:

Une image contenant texte, Police, logo, Graphique

Description générée automatiquement

1. Nombre de population :

PopulationTotale = CALCULATE (SUM('DIM\_Région'[Population municipale]))

Une image contenant capture d’écran, texte

Description générée automatiquement

* **Explication** : Nombre de population permet de connaître la population totale de France pour toutes les régions de France.
* **Résultat souhaité**

Une image contenant texte, Police, logo, Graphique

Description générée automatiquement

1. Nombre de musée fréquenté par régions

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

Description générée automatiquement

**Explications** : Nous avons représenté par un diagramme en aire le nombre de musée fréquenté par régions. Nous pouvons donc remarquer qu’il y’a beaucoup plus de musées fréquentés en « Ile-de-France » que dans les autres régions. Par-ailleurs certaines régions n’ont quasiment pas de musée signifiant comme les régions de Guadeloupe jusqu’à Saint-Pierre-et-Miquelon.

1. **Nombre de musée fréquenté par ville.**

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

Description générée automatiquement Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

* **Explications :** Ce graphique en barres représente le nombre de musées accessibles aux populations par ville. On remarque que grand nombre de musées sont présents et très fréquentés à Paris par rapport à d’autres villes donc le nombre de visites des musées par ville n’est pas proportionnel. Il y’a des villes où les fréquentations sont quasiment insignifiantes du fait de l’inexistence des musées.

1. **Total de fréquentations des musées**

Une image contenant capture d’écran, texte, logiciel

Description générée automatiquement

* **Explications :**

Le total de fréquentation des musées permet de déterminer la somme des totaux de fréquentations des musées en termes de personnes.

* **Résultat souhaité :**

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

Description générée automatiquement

1. **Nombre de régions de France**

Une image contenant capture d’écran, texte

Description générée automatiquement

* **Explications** : Le nombre de régions de France est obtenu en comptant le nombre de région dans la table ville de manière distincte.
* **Résultat souhaité** :

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

Description générée automatiquement

1. Moyenne de personnes accessibles par musée en France

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

Description générée automatiquement

* **Explication :**

La moyenne de fréquentation est obtenue en faisant la moyenne sur le champ total de la table fréquentation. Ça permet de voir en moyenne le nombre de personnes dans un musée.

* **Résultat obtenu :**

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

Description générée automatiquement

1. Moyenne des populations par régions

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

Description générée automatiquement

* **Explications :**

La moyenne de population est obtenue en faisant la moyenne de la population municipale dans la table DIM\_Région

* **Résultat souhaité :**

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

Description générée automatiquement

1. **Evaluer le nombre de musée au fil des années (si ça augmente ou ça diminue)**

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

Description générée automatiquement

**Explications** : Ce Graphique en courbe représente la variation des musées visité au fil des années. On remarque que le nombre de musée diminue au fil des ans.

**Conclusion** : le nombre de musée ne sera pas suffisant pour couvrir la population

1. Pages des rapports Power BI

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

Description générée automatiquement

Description :

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

Description générée automatiquement

**ETUDE CAS 2**

1. Proposons 02 architectures BI complètes pour la collecte des données jusqu’à la visualisation

L'architecture BI est conçue pour fournir une description complète d’un projet BI, depuis la collecte des données jusqu'à la visualisation, en passant par la transformation et la modélisation des données.

* **PREMIERE ARCHITECTURE**

**Source de données**

**Datawarehoouse : Stocke les données**

**Visualisation : Tableau, Looker, Power BI**

**ETL : nettoyage, transformation, merge, chargement**

**Talend**

**Microsoft SSIS,**

**SAS Data Integration Studio**

Google BigQuery Amazon, Apache Hive

**DataMart**

VISUALISER

VISUALISER

VISUALISER

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

Description générée automatiquement

**Population**

**FrequentationMusée**

**Musées**

CHARGER

Une image contenant vitesse, objets en métal, cercle, roue

Description générée automatiquement

**Données internes** : Fichier sur la population de France

**EXTRACTION**

EXTRACTION

**Données externes** : Fichier sur les musées, fréquentation des musées, les villes de France

**Description** :

L’ETL extrait les sources de données provenant des données internes et externes, les nettoie, les transforme, les fusionne et les charge dans le datawarehouse. Ensuite le datawarehouse stocke les données qui seront ensuite reparties dans les datamart représentant chacun une base de données dont le contenu est en rapport avec une activité précise de l’entreprise. Enfin ces données sont affichées à l’utilisateur sous forme de tableau de bord, de reporting, pour permettre la prise de décision.

* DEUXIEME ARCHITECTURE

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

Description générée automatiquement

**MuleSoft Anypoint Platform**

**WSO2 Enterprise Integrator**

**Description** :

* **Exposition** : le consommateur ne connaît que l’ESB. Il invoque le service que ce dernier lui expose. Cette invocation se fait sur un protocole et avec un format de données qui sont indépendants du fournisseur de service ;
* **Routage** : l’ESB détermine le fournisseur de service à invoquer (éventuellement en s'appuyant sur un registre ou annuaire de services) ;
* **Transformation** : l’ESB réalise une médiation de format vers celui pris en charge par le fournisseur de service ;
* **Invocation** : c'est l’ESB qui invoque le fournisseur de service.

1. Présentation de la modélisation

Une image contenant texte, diagramme, logiciel, conception

Description générée automatiquement