

TP Power BI : Analyse des données Airbnb

Objectifs pédagogiques

1. Comprendre comment importer et transformer des données dans Power BI.
2. Apprendre à nettoyer et harmoniser des données (suppression des doublons, gestion des valeurs manquantes, harmonisation des formats).
3. Créer un modèle relationnel entre différentes tables.
4. Concevoir des visualisations pertinentes (cartes, graphiques, KPI).
5. Répondre à des questions métiers à partir des données.
6. Réaliser le même TP sur Tableau Software

Étape 1 : Préparation des données

Fichiers fournis

1. `listings_with_issues.csv` : Informations sur les logements.
2. `hosts_with_issues.csv` : Informations sur les hôtes.
3. `reviews.csv` : Avis des utilisateurs.
4. `bookings.csv` : Réservations.

Étape 2 : Nettoyage et transformation des données

Tâches à réaliser dans Power BI

1. **Importer les fichiers :**
 - Charger les quatre fichiers CSV dans Power BI.
1. **Nettoyer les données :**
 - **Supprimer les doublons** dans `listings.csv` et `hosts_with_issues.csv`.
 - **Gérer les valeurs manquantes :**
 - Dans `listings.csv` : remplacer les valeurs manquantes de `average_rating` par la moyenne des notes non manquantes.
 - Dans `hosts_with_issues.csv` : remplacer les valeurs manquantes de `host_name` par "Unknown Host".
 - **Harmoniser les noms des villes :** Standardiser les noms dans les colonnes `city` en les mettant en majuscules.
1. **Créer des colonnes calculées :**
 - Dans `listings.csv` :
 - Ajouter une colonne "Catégorie" pour classer les logements :
 - **"Économique"** : si `price_per_night < 100`.
 - **"Standard"** : si $100 \leq \text{price_per_night} < 300$.
 - **"Haut de gamme"** : si `price_per_night ≥ 300`.

Étape 3 : Modélisation

Tâches à réaliser dans Power BI

1. **Créer un modèle relationnel :**
 - Joindre les tables en utilisant les relations suivantes :
 - listings.csv ↔ reviews.csv : Relation basée sur listing_id.
 - listings.csv ↔ bookings.csv : Relation basée sur listing_id.
 - hosts.csv ↔ bookings.csv : Relation basée sur host_id.
1. **Configurer les cardinalités :**
 - Identifier les relations en "1 à plusieurs" et les configurer correctement dans Power BI.

Étape 4 : Création de visualisations

Visualisations attendues

1. **Carte géographique :**
 - Répartition des logements par ville.
 - Utiliser une carte pour visualiser le nombre de logements dans chaque ville.
1. **Histogramme :**
 - Distribution des prix par nuit (price_per_night).
 - Ajouter des filtres pour explorer cette distribution par ville et par catégorie (économique, standard, haut de gamme).
1. **Graphique à barres :**
 - Top 10 des logements avec la meilleure note moyenne (average_rating).
1. **Graphique de tendance :**
 - Évolution mensuelle du nombre de réservations (bookings.csv).
1. **Indicateurs clés (KPI) :**
 - Nombre total de réservations.
 - Revenu total généré (somme de total_amount dans bookings.csv).
 - Note moyenne globale des logements.

Étape 5 : Questions métiers

Questions auxquelles répondre avec Power BI

1. Quelle est la répartition des logements par ville ?
2. Quels sont les logements les mieux notés ? Quel est leur prix moyen ?
3. Quel est le revenu total généré par les superhôtes ?
4. Quelle est la tendance des réservations au fil du temps ?
5. Quelle est la distribution des logements selon leur catégorie de prix ?

Bonus : Analyse avancée

1. Analyse des superhôtes :

- Créer un segment de superhôtes et analyser leur contribution aux réservations et aux revenus.

1. Avis des utilisateurs :

- Analyser les notes des avis (`reviews.csv`) et identifier des corrélations avec les prix des logements.