

Mini Projet – Analyse Exploratoire des Admissions Hospitalières

Rapport EDA

Réalisé par Ibrahima Gassama

1. Contexte

Les données hospitalières analysées proviennent d'un fichier CSV contenant des informations sur les patients, les admissions, les médecins, les hôpitaux ainsi que des éléments financiers et médicaux.

L'objectif de ce mini-projet est :

- De nettoyer le dataset,
- D'extraire des tendances significatives,
- De visualiser ces tendances,
- Et de déployer une application Streamlit interactive accessible en ligne.

2. Description du dataset

Le fichier comporte les variables suivantes :

- **Patients** : âge, sexe, groupe sanguin
- **Condition médicale** : ex. Asthme, Cancer, Diabète
- **Informations hospitalières** : hôpital, médecin, type d'admission
- **Finances** : montant facturé
- **Résultats médicaux** : test normal, anormal, etc.

Le dataset contient **n** patients (selon le CSV).

3. Nettoyage réalisé

Les principales opérations appliquées :

- Conversion des dates (`Date of Admission`, `Discharge Date`)
- Normalisation des noms (capitalisation, suppression espaces)
- Conversion numérique (`Age`, `Billing Amount`)
- Suppression des doublons
- Calcul de la **durée de séjour** :
$$\text{LengthOfStay} = \text{Discharge Date} - \text{Admission Date}$$

Les données deviennent donc cohérentes, prêtes pour l'analyse.

4. Analyse exploratoire (résultats)

- **Âge des patients** : réparti de façon assez large avec une moyenne autour de X ans.
- **Conditions médicales fréquentes** : Asthme, Cancer, Obésité dominant.
- **Types d'admission** : Urgences et Admissions Électives apparaissent majoritaires.
- **Résultats des tests** : majorité de tests « Normaux » mais présence de cas anormaux.
- **Finances** : le coût facturé varie fortement, certaines admissions sont marquées par des coûts élevés.
- **Durée du séjour** : observée généralement entre X et Y jours.

5. Visualisations principales

1. Histogramme : distribution de l'âge.
2. Countplot : répartition des types d'admission.
3. Scatter Plot (Plotly) : Coût vs durée du séjour.
4. (Optionnel) Heatmap admissions par hôpital.

Ces visualisations sont intégrées au notebook et à l'application Streamlit.

6. Application Streamlit

Une version en ligne interactive est disponible :

<https://hashimi-barry-mini-proj-q09v63iy73-streamlit-app-f1bvvy.streamlit.app/>

Elle permet :

- De consulter les données nettoyées,
- D'interagir avec les visuels,
- D'effectuer une exploration filtrée.

7. Conclusion

L'analyse montre des tendances significatives :

- Certains diagnostics mènent à des séjours plus longs et coûteux.
- Les urgences sont fréquentes.
- Les données nettoyées permettent une exploration fiable.

L'étude pourrait être approfondie en :

- Ajoutant des tests statistiques,
- Évaluant la corrélation entre variables,
- Appliquant du machine Learning (classification des admissions coûteuses).

