

Documentation du Projet de Visualisation des Données de Santé Publique



1. Introduction

Le projet a pour objectif d'analyser et de visualiser les données relatives à la grippe en France, en mettant l'accent sur les hospitalisations, les passages aux urgences et la couverture vaccinale. L'objectif était de rendre ces données accessibles à travers une interface interactive afin de permettre aux utilisateurs d'en tirer des insights et de prendre des décisions informées.

Notre rôle au sein du projet a été celui de **Développeurs et Data Analysts**. Nous avons appliqué nos compétences en **Data Science, Visualisation de données et Communication** pour créer une plateforme permettant de visualiser et d'explorer ces données en temps réel. L'interface a été construite avec **Dash**, intégrant des visualisations interactives telles que des graphiques et des cartes choroplèthiques, permettant une exploration détaillée des données de santé publique.

2. Objectifs et Compétences Utilisées

2.1 Objectif du Projet

L'objectif principal de ce projet était de fournir un tableau de bord interactif pour la visualisation des données de santé publique en France, notamment celles relatives à la grippe. Nous avons cherché à permettre aux **utilisateurs cibles, à savoir les autorités sanitaires**, d'analyser les tendances des hospitalisations, des passages aux urgences et de la couverture vaccinale à travers des visualisations interactives, ainsi que d'examiner les quantités d'actes vaccinaux ou de doses de vaccins, tout en tenant compte des variations géographiques et démographiques.

2.2 Compétences Développées

Les compétences suivantes ont été mobilisées tout au long du projet :

- **Data Science** : Traitement des données, analyse statistique, et élaboration de modèles d'analyse.

- **Visualisation de données** : Création de visualisations interactives permettant d'explorer les données en profondeur.
- **Géomatique** : Manipulation des données géospatiales pour permettre une visualisation régionale.
- **Communication** : Création de rapports et de présentations claires pour les parties prenantes, ainsi que la conception d'une interface utilisateur intuitive.

3. Méthodologie et Outils

3.1 Collecte et Préparation des Données

Les données utilisées proviennent de sources publiques, incluant des informations sur les passages aux urgences, les hospitalisations et la couverture vaccinale des départements en France. Nous avons nettoyé et préparé ces données à travers les étapes suivantes :

- **Nettoyage des données** : Gestion des valeurs manquantes et anomalies.
- **Transformation des données** : Conversion des formats, agrégation, et filtrage.
- **Fusion des sources de données** : Pour enrichir les données et faciliter l'analyse.

Les outils utilisés pour ces étapes incluent principalement **Python** avec les bibliothèques **Pandas** et **NumPy**, ainsi que **Scikit-Learn** pour le machine learning (prédiction).

3.2 Analyse des Données

Une fois les données nettoyées et préparées, nous avons effectué des analyses exploratoires pour comprendre les tendances et identifier les éventuelles anomalies.

- **Corrélations** entre les indicateurs clés (ex. : taux de vaccination et hospitalisations).
- **Visualisations préliminaires** (histogrammes, boxplots, etc.) pour mieux comprendre la distribution des données.

Les outils utilisés pour ces analyses étaient :

- **Pandas** pour la manipulation des données.
- **Matplotlib / Seaborn** pour la création de visualisations statiques.

3.3 Modélisation et Prédiction

Nous avons initialement prévu de mettre en place des modèles de prédiction pour anticiper l'évolution de certains indicateurs, tels que le nombre d'hospitalisations ou de passages aux urgences. Cependant, après plusieurs essais et une évaluation

approfondie des données, nous avons décidé de ne pas retenir ces modèles prédictifs en raison des limitations suivantes :

Modèles envisagés :

- **Régression linéaire** : Pour prédire le nombre d'hospitalisations en fonction de la couverture vaccinale et d'autres variables.
- **Modèles de classification** : Tels que les **arbres de décision** ou **forêts aléatoires** pour classer les départements selon leur niveau de risque épidémique.
- **Modèles ARIMA** : Pour prédire les séries temporelles des cas de grippe sur plusieurs mois.

Limitations des données :

- **Données manquantes ou incomplètes** : Plusieurs périodes étaient partiellement couvertes, ce qui a rendu l'entraînement de modèles de prédiction fiables difficile.
- **Granularité insuffisante** : Les données étaient souvent agrégées à un niveau trop global (par exemple, au niveau national) pour permettre une modélisation précise à l'échelle départementale.
- **Fluctuations imprévues** : Des événements imprévus, comme des pics de cas non anticipés, ont rendu difficile la modélisation stable de l'évolution des indicateurs.

En conséquence, après des tests de validation croisée et des évaluations des performances des modèles, nous avons choisi de privilégier une **analyse descriptive** approfondie et des **visualisations interactives**, plutôt que d'implémenter des modèles prédictifs moins fiables.

3.4 Visualisation des Données et Interface Interactive

L'une des principales étapes du projet a été de rendre les données accessibles de manière interactive. Nous avons conçu une interface web interactive avec **Dash**, permettant aux utilisateurs de visualiser et d'explorer les données en temps réel. Les principales fonctionnalités de l'interface incluent :

- **Graphiques interactifs** (barres, courbes, etc.) pour comparer les taux d'hospitalisations, de passages aux urgences et de couverture vaccinale.
- **Cartes choroplèthiques** pour visualiser la répartition géographique des données.
- **Filtres dynamiques** permettant de segmenter les données par départements, classes d'âge ou période.

Les outils utilisés pour cette étape sont :

- **Dash** : Framework Python pour créer des applications web interactives.
- **Plotly** : Pour la création de graphiques interactifs.

- **GeoPandas** : Pour manipuler et afficher des données géospatiales.

3.5 Communication des Résultats

La communication claire des résultats a été un élément essentiel du projet. L'interface interactive que nous avons développée permet aux utilisateurs de visualiser les données en temps réel. Par ailleurs, nous avons produit des **rapports** et des **présentations** pour résumer les résultats de manière accessible, même pour un public non technique.

Les outils utilisés pour la communication étaient :

- **PowerPoint** : Pour la présentation des résultats aux parties prenantes.
- **Jupyter Notebooks** : Pour documenter et partager les analyses.

4. Résultats et Impact

4.1 Résultats Clés

Les résultats de notre analyse ont permis de tirer plusieurs conclusions importantes :

- Une **corrélation positive** a été observée entre une couverture vaccinale plus élevée et un nombre réduit de passages aux urgences.
- Des **disparités géographiques** importantes existent, avec certains départements ayant des taux d'hospitalisations beaucoup plus élevés que d'autres.
- Des **zones sensibles** ont été identifiées, nécessitant des interventions ciblées, en particulier dans les départements où la couverture vaccinale est faible.

4.2 Impact pour les Parties Prenantes

Le tableau de bord interactif a permis aux décideurs de mieux comprendre les tendances de la grippe et d'ajuster leurs stratégies de santé publique en conséquence. Grâce à la visualisation en temps réel des données, les autorités de santé publique peuvent rapidement identifier des zones à risque et prendre des mesures appropriées.

5. Conclusion et Perspectives

5.1 Conclusion

Ce projet a été une excellente occasion de mettre en œuvre et de développer des compétences en **Data Science**, **Visualisation des données**, et **Communication**. Grâce à l'utilisation de **Dash** et **Plotly**, nous avons créé une plateforme intuitive et informative qui permet d'analyser et de visualiser les données de manière dynamique.

5.2 Perspectives

Pour aller plus loin, voici quelques pistes de développement pour de futurs projets :

- **Prédiction et Modélisation** : Bien que nous n'ayons pas pu implémenter de modèles prédictifs fiables cette fois-ci en raison des limitations des données, un futur projet pourrait se concentrer sur l'amélioration de la collecte de données pour permettre une modélisation plus robuste.
- **Données en temps réel** : L'ajout de flux de données en temps réel permettrait de rendre le tableau de bord encore plus utile pour la surveillance continue des tendances, avec des mises à jour instantanées pour un suivi plus réactif.
- **Amélioration de l'interface utilisateur** : L'ergonomie de l'interface pourrait être améliorée pour faciliter encore plus la navigation et l'exploration des données, notamment en ajoutant des filtres interactifs avancés.

De plus, nous prévoyons de développer une **version 2 (V2)** du projet, qui mettra l'accent sur une interactivité améliorée et des performances accrues. L'architecture de la V2 sera la suivante :

- **ETL Python (Pandas / Airflow / API)** : Le traitement des données sera effectué avec **Pandas** pour le nettoyage et la transformation des données. Nous intégrerons également **Airflow** pour la gestion des workflows d'extraction, transformation et chargement (ETL) et exposerons les résultats via une API REST ou GraphQL.
- **Front JavaScript (React + ECharts/Plotly)** : Le front-end sera développé avec **React** pour une interface dynamique et réactive. Les visualisations interactives seront réalisées avec **ECharts** et **Plotly** pour une meilleure performance graphique.
- **Dashboard interactif** : L'objectif est d'avoir une plateforme de visualisation évolutive, où les utilisateurs pourront interagir avec les données en temps réel et appliquer des filtres personnalisés selon leurs besoins.

Ce projet nous a permis de renforcer nos compétences techniques et de mieux comprendre l'importance d'une communication claire des résultats, surtout dans le cadre de la santé publique.