

Projet #3

Projet d'Analyse des Prénoms des Bébés aux États-Unis

Développeur Full Stack Big Data - L'École Multimédia



Sommaire

- [Présentation](#)
- [Contraintes](#)
- [Libertés](#)
- [Méthodologie](#)
- [Contraintes](#)
- [Objectifs](#)
- [Cahier des charges](#)
 - [Analyse des Données](#)
 - i. [Acquisition et Préparation des Données](#)
 - ii. [Analyses Spécifiques](#)
 - [Développement de l'Interface Web avec FastAPI](#)
 - i. [Configuration de l'Environnement](#)
 - ii. [Développement de l'Interface Utilisateur](#)
 - iii. [Structure de l'Interface Web](#)
 - iv. [Fonctionnalités Interactives](#)
 - v. [Design et Convivialité](#)
 - vi. [Conclusion](#)
- [Rendu final](#)
- [Conseils](#)
- [Compétences à valider](#)
- [ANNEXES](#)

Contexte

Présentation

L'administration américaine de la sécurité sociale (USSSA ou SSA) a rendu publiques les données sur la fréquence des prénoms attribués aux bébés depuis 1880 jusqu'en 2018. Ce projet vise à analyser ces données pour identifier les tendances et les modèles dans l'attribution des prénoms.

Une interface utilisateur sera développée pour présenter les résultats de vos données.

Charte graphique

L'application respectera la charte graphique du SSA :

- Couleurs : <https://coolors.co/0b4678-1976d2-e1d7cd-a26769-6d2e46>
- Typographie : Sans serif

Contraintes

- Le projet devra être réalisé en groupe d'au moins 3 personnes ;
- Le développement doit être réalisé en Python, Pandas, Matplotlib et FastAPI ou Django

Libertés

- Vous pouvez proposer l'utilisation de modules Python de votre choix, à condition de les justifier ;
- Vous travaillerez à partir du jeu de données suivants : National data que vous trouverez à la page suivante : <https://www.ssa.gov/oact/babynames/limits.html>

Méthodologie

Vous devrez démontrer au client que vous êtes capable de travailler selon des méthodes itératives et des pratiques "Agile". C'est une des conditions posées pour que vous soyez chargés du projet final.

Contraintes

- Utilisation de méthodologies de travail collaboratives et Agile.
- Production d'une documentation progressive
- Utilisation d'outils de modélisation (de type Merise ou UML, par exemple)

Objectifs

À l'issue du module, vous devrez avoir réalisé les éléments suivants :

- Installer Django ou FastAPI
- Modéliser des données ORM
- Architecturer une application avec FastAPI ou Django
- Sécuriser les accès aux données
- Analyser les données avec Pandas & des notions de statistiques descriptives
- Analyser des données en utilisant des graphiques

Cahier des charges

Description des fonctionnalités

1. Analyse des Données

Acquisition et Préparation des Données

- Téléchargement et chargement des données:
 - Les données sont téléchargées depuis un fichier "names.zip" et chargées en utilisant Pandas.
- Exploration initiale : Calcul du nombre de naissances par sexe pour une année donnée et concaténation des données de toutes les années pour créer un ensemble complet.
- Agrégation des données : Création de tableaux pivots pour représenter les naissances par année et par sexe.
- Visualisation des naissances : Représentation graphique du nombre de naissances par année.
- Calcul des proportions** : Ajout d'une colonne pour la proportion des naissances par prénom, sexe et année, et vérification de la somme des proportions pour chaque groupe.
- Extraction des sous-ensembles : Extraction de sous-ensembles de 1000 prénoms les plus populaires par groupe avec une fonction spécifique.

Analyses Spécifiques

- Tableau pivot par prénom et année : Création d'une table pivot des naissances par prénom et par année.
- Étude de tendances : Analyse et représentation graphique des tendances de prénoms spécifiques comme "John" et "Harry".
- Mesure de la diversité des prénoms : Évaluation de la diversité des prénoms au fil du temps.
- Analyse par décennie : Regroupement et analyse des tendances des prénoms par décennie.

- Diversité géographique : Exploration des différences de popularité des prénoms par région.
- Longueur des prénoms : Analyse des tendances liées à la longueur des prénoms.
- Noms composés : Étude de l'évolution de l'utilisation des noms composés.

2. Développement de l'Interface Web avec FastAPI

Configuration de l'Environnement

- Mise en place de FastAPI : Configuration de l'environnement pour créer une API RESTful.

Développement de l'Interface Utilisateur

- Routes pour accès aux données : Création de routes permettant d'accéder aux données et d'afficher des graphiques interactifs avec Plotly.
- Intégration avec PostgreSQL
- Stockage des données : Utilisation de PostgreSQL pour le stockage des données, permettant une récupération rapide et efficace.

Structure de l'Interface Web

- Conception intuitive : L'interface web est conçue pour offrir une expérience utilisateur intuitive et conviviale.
- Diagramme UML : Un diagramme UML simple représente les relations entre les années, les prénoms et les naissances.

Fonctionnalités Interactives

- Sélection de prénoms : Permet aux utilisateurs de sélectionner un ou plusieurs prénoms à visualiser.
- Filtrage par année : Possibilité de filtrer les données par année pour observer les tendances sur des périodes spécifiques.
- Zoom et défilement : Graphiques interactifs avec des options de zoom et de défilement pour une exploration détaillée.

Design et Convivialité

- Attention à l'esthétique : Conception esthétique et conviviale pour améliorer l'expérience utilisateur.

Conclusion

- Synthèse des découvertes : Résumé des principales tendances et découvertes de l'analyse des données.
- Perspectives d'avenir : Discussion sur les perspectives et améliorations futures du projet.

Rendu final

Votre rendu final prendra la forme d'une archive Zip et devra comporter les éléments suivants :

- Code source du projet.
- Spécification technique.
- Documentation complète.
- Suivi détaillé des réunions et plan de travail partagé.
- Présentation orale avec démonstration du projet.
- L'adresse de votre dépôt Github

Le jeu de données ne sera pas intégré à l'archive, par contre spécifiez la source et donnez le lien vers les données dans vos documents.

Cette archive aura comme titre le nom du groupe suivi de votre classe.

Exemple: Projet3_groupe_A_CDA.zip

Votre dossier comprendra au maximum une quinzaine de pages.

Par ailleurs, vous devrez faire une présentation orale de votre travail devant le client, comprenant une démonstration de l'application. Vous aurez 15 minutes pour convaincre le client.

Conseils

- Assurez-vous de bien comprendre les objectifs et les exigences du projet dès le début.
- Adoptez une approche itérative et agile pour le développement du logiciel.
- Utilisez des outils de gestion des tâches et de collaboration pour faciliter le travail d'équipe.
- Effectuez des tests rigoureux pour garantir la qualité du logiciel.
- Documentez soigneusement chaque étape du processus de développement.
- Assurez-vous de respecter les délais et de communiquer efficacement avec les parties prenantes.
- Pensez à mesurer les temps de parole pour la présentation orale

Compétences à valider

RNCP32123BC06 - Architecte BIG DATA et science de données

- 6.5 : Écrire des algorithmes d'analyse de données
- 6.8 : Maîtriser l'analyse et la science de données

RNCP32123BC08 - Développement WEB

- 8.1 Utiliser les principaux outils de développement web
- 8.2 Produire une application Orientée Objet

ANNEXES

- Support Pandas : <https://pandas.pydata.org/>
- Support FastAPI : <https://fastapi.tiangolo.com/>
- Support Python <https://github.com/aahWeb/Python-Expert-/tree/main>