


 PROJET VALIDÉ

Analysez des données de systèmes éducatifs

MISSION COURS RESSOURCES ÉVALUATION

 60 heures

Mis à jour le mercredi 2 décembre 2020

Vous êtes Data Scientist dans une **start-up de la EdTech**, nommée **academy**, qui propose des contenus de formation en ligne pour un public de niveau lycée et université.



Votre entreprise EdTech

Mark, votre manager, vous a convié à une réunion pour vous présenter le projet d'**expansion à l'international** de l'entreprise. Il vous confie **une première mission d'analyse exploratoire**, pour déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion.

Voici les différentes questions que Mark aimerait explorer, que vous avez notées durant la réunion :

- Quels sont les pays avec un fort potentiel de clients pour nos services ?
- Pour chacun de ces pays, quelle sera l'évolution de ce potentiel de clients ?
- Dans quels pays l'entreprise doit-elle opérer en priorité ?

Votre mission

Mark vous a donc demandé de réaliser une analyse pré-exploratoire de ce jeu de données. Il vous a transmis cet email à la suite de la réunion :

Hello,

Les données de la Banque mondiale sont disponibles à l'adresse suivante :

<https://datacatalog.worldbank.org/dataset/education-statistics>

Ou en téléchargement direct à ce [lien](#).

Je te laisse regarder la page d'accueil qui décrit le jeu de données. En résumé, l'organisme "EdStats All Indicator Query" de la Banque mondiale répertorie 4000 indicateurs internationaux décrivant l'accès à l'éducation, l'obtention de diplômes et des informations relatives aux professeurs, aux dépenses liées à l'éducation... Tu trouveras plus d'info sur ce site :

<http://datatopics.worldbank.org/education/>

Pour la pré-analyse, pourrais-tu :

- Valider la qualité de ce jeu de données (comporte-t-il beaucoup de données manquantes, dupliquées ?)
- Décrire les informations contenues dans le jeu de données (nombre de colonnes ? nombre de lignes ?)
- Sélectionner les informations qui semblent pertinentes pour répondre à la problématique (quelles sont les colonnes contenant des informations qui peuvent être utiles pour répondre à la problématique de l'entreprise ?)
- Déterminer des ordres de grandeurs des indicateurs statistiques classiques pour les différentes zones géographiques et pays du monde (moyenne/médiane/écart-type par pays et par continent ou bloc géographique)

Ton travail va nous permettre de déterminer si ce jeu de données peut informer les décisions d'ouverture vers de nouveaux pays. On va partager ton analyse avec le board, alors merci de soigner la présentation et de l'illustrer avec des graphiques pertinents et lisibles !

Livrables attendus

- Un **notebook** comportant les analyses pré-exploratoires réalisées (non cleané, pour comprendre votre démarche).
- Un **support** de présentation pour la soutenance.

Pour faciliter votre passage au jury, déposez sur la plateforme, dans un dossier nommé "*P2_nom_prenom*", tous les livrables du projet. Chaque livrable doit être nommé avec le numéro du projet et selon l'ordre dans lequel il apparaît, par exemple "*P2_01_notebook*", "*P2_02_support*", et ainsi de suite.

Modalités de la soutenance

5 min - Rappel de la problématique et présentation du jeu de données

15 min - Présentation de l'analyse pré-exploratoire du jeu de données et vos conclusions sur la pertinence de l'usage du jeu de données pour répondre aux questions stratégiques que se pose l'entreprise

5 à 10 minutes de questions-réponses

Démarrage du projet

Vous vous demandez par où commencer ? Suivez nos conseils dans cette video pour vous organiser et faire le point avant votre première session de mentorat !

05:32

Compétences évaluées



Mettre en place un environnement Python



Manipuler des données avec des librairies Python spécialisées



Utiliser un notebook Jupyter pour faciliter la rédaction du code et la collaboration



Effectuer une représentation graphique à l'aide d'une librairie Python adaptée



Maîtriser les opérations fondamentales du langage Python pour la Data Science

OPENCCLASSROOMS



ENTREPRISES



CONTACT



EN PLUS



 Français 

