

[← Mon parcours](#) **PROJET VALIDÉ**

Analysez des données de systèmes éducatifs

[Mission](#) [Cours](#) [Ressources](#) [Évaluation](#) 60 heures

Mis à jour le lundi 21 novembre 2022

Le nommage des livrables à déposer sur la plateforme a été changé et des indications sur les temps de soutenance ont été ajoutées le 11/05/2022.

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 devant le jury, déposez sur la plateforme, dans un dossier zip nommé "**Titre_du_projet_nom_prénom**", votre livrable nommé comme suit :

Nom_Prénom_n° du livrable_nom du livrable_date de démarrage du projet. Cela donnera :

- *Nom_Prénom_1_notebook_mmaaaa*
- *Nom_Prénom_2_presentation_mmaaaa*

Par exemple, votre premier livrable peut être nommé comme suit :

Dupont_Jean_1_notebook_12022.

Soutenance

Votre soutenance aura lieu avec un mentor évaluateur et durera environ 30 minutes.

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

Votre présentation devrait durer 15 minutes (+/- 5 minutes). Puisque le respect des durées des présentations est important en milieu professionnel, les présentations en dessous de 10 minutes ou au-dessus de 20 minutes peuvent être refusées.

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



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



Manipuler des données avec des librairies Python spécialisées



Maîtriser les opérations fondamentales du langage Python pour la Data Science



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

OPENCLASSROOMS



OPPORTUNITÉS



AIDE



POUR LES ENTREPRISES



EN PLUS



Français



Télécharger dans
l'App Store

