

Explorez et nettoyez un dataset simple et retirez des premières observations

Dans ce webinaire vous apprendrez à exploiter des données recensées aux États-Unis afin d'orienter une entreprise de l'EdTech dans ses choix de développement à l'international. On vous demande pour cela de produire des résultats actionnables et impactants dans l'objectif de convaincre votre client du classement des villes que vous lui proposerez.

Compétences visées :

- Exploration et nettoyage de données
- Produire des observations pertinentes

Énoncé

En tant que Data Scientist vous venez de décrocher une mission avec une plateforme d'éducation en ligne leader sur le marché européen, vous allez bientôt signer le contrat et vous consacrer au sujet pour les deux prochaines semaines.

Ce leader de l'EdTech voit grand, ayant une forte implantation en France et en Europe, les dirigeants souhaitent également se développer sur le territoire américain qui regorge d'opportunités pour la formation en ligne, de par la taille de la population et des besoins liés aux métiers du numérique. L'entreprise souhaite démarrer son implantation aux USA en sélectionnant les territoires géographiques les plus prometteurs afin de lancer des campagnes publicitaires ciblées.

Support de travail

L'exercice est directement présenté sous la forme d'un notebook Jupyter, **téléchargez donc le notebook correspondant et répondez aux questions**. Les sorties des cellules sont en réalité déjà fournies, mais à vous de trouver les lignes de codes qui permettent de les obtenir.

Les données et le notebook sont directement disponibles dans [ce fichier zip](#).

Note

Vous êtes étudiant sur le parcours Data Scientist ? Ce scénario vous rappellera certainement un de vos projets ! Il s'inspire en effet du projet 2 de Data Scientist, où vous devez réaliser un exercice exploratoire similaire, à partir des données de la Banque Mondiale.

Cependant, dans cet exercice, les recommandations doivent être faites à l'échelle des villes et pas des pays, comme c'était le cas dans le projet 2. Si l'objectif est donc similaire, la dataset et le résultat attendu sont bien différents dans cet exercice.

Ressources

[Découvrez les librairies Python pour la data science](#)