

ESILV - Python for data analysis
Modalité de notation

Date de rendu : 10 janvier minuit

Il vous est attribué un jeu de donnée à analyser, sur la page d'[attribution des datasets par élève : https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEFkF4/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEFkF4/edit?usp=sharing).

On vous demande d'analyser le jeu de donnée qui vous est attribué et de faire à partir de là :

1. Un powerpoint expliquant les tenants et aboutissant du problème, vos réflexions sur la question posée, les différentes variables que vous avez créées, comment se situe le problème dans le contexte de l'étude, etc : **25 %**
2. Un code en python :
 - a. Data-visualisation (utilisez matplotlib, seaborn, bokeh...) – montrez le lien entre les variables et la cible : **25%**
 - b. Modélisation – prenez scikit-learn essayez plusieurs algorithmes, changez les hyper paramètres, faites une grille de recherche, comparez les résultats de vos modèles dans des graphiques : **25%**
3. Transformation du modèle en API Django : **25%**

Un mail + un github + le remplissage d'une feuille google drive sont attendus :

Pour répondre, vous devrez :

1. Mettre votre travail sur votre compte github :
 - a. Un Readme résumant la tâche à effectuer et vos conclusions
 - b. La version pdf de votre powerpoint
 - c. Le code au format jupyter notebook
 - d. L'api django ou Flask
2. Mettre le lien vers votre github sur la page suivante :
<https://docs.google.com/spreadsheets/d/1CHVkt2Bk7RBycCZypeMLjxcM6U2MynypbGFFWEFkF4/edit?usp=sharing>
3. Envoyer un mail à romain.jouin@gmail.com + le professeur que vous avez eu en cours, avec :
 - a. Sujet : "ESILV - Python for data analysis - devoir 2021"
 - b. Contenu :
 - Nom / prénom
 - Lien github
 - c. Pièce jointe
 - Ppt (le même que sur le github)

Les professeurs restent à votre disposition pour toute question éventuelle, et essayez aussi de contacter les chercheurs responsables des data sets : il y a souvent leurs mails dans les descriptifs, et ils seront sans doute heureux de voir que leur travail est réutilisé, et on verra s'ils répondent.

Le travail principal du data-scientist est un travail de compréhension des données, et oser poser les questions est une qualité indispensable dans ce métier.