

BUT STID - Semestre 3

COMPETENCE 4 : Développer un outil décisionnel - parcours VCOD

Collecte de données web

F.GARNIER



Plan de la Saé

Chapitre 1 : Web Scraping

Chapitre 2 : API

Sujet noté non alternant

Sujet noté – non alternant

1) Contexte Web scraping

Pour ce sujet, vous serez libre de choisir le site à scraper. Pensez à analyser les données à traiter, à évaluer la difficulté de récupération de ces données par la technique de web scraping et enfin à imaginer des indicateurs qui aient du sens pour l'application.

TRAVAIL A FAIRE

Etapes à suivre pour réaliser l'application Python :

- Recherche des données et du site à scaper en lien avec notre domaine de compétences et contenant des données géographiques : réfléchir à l'intérêt des données et aux indicateurs potentiels à produire.
- Validation par votre enseignant
- Scraping des données du site web dans un dataframe.
- Tableau de bord à partir du dataframe en Python : au moins une carte, un graphique, un tableau de données.

Attention à bien choisir des indicateurs pertinents !

2) Contexte API : Taux de réussite des auto-écoles à l'examen du permis de conduire.

Le jeu de données concerne le taux de réussite des auto-écoles en 2018 à l'examen du permis de conduire. Ce jeu de données dispose de données de localisation (rue, ville, code postal des auto-écoles) mais pas de données au bon format géographique permettant une géolocalisation sur une carte (latitude, longitude).

Le jeu de données est au format Xls et est récupérable sur Updago : TxReussiteAutoEcoles.xls

Nous utiliserons donc une API pour traduire les informations Rue+CP+Ville en coordonnées géographiques (latitude + longitude). Il s'agit de l'**API Adresse** (Base Adresse Nationale) <https://api.gouv.fr/les-api/base-adresse-nationale>

TRAVAIL A FAIRE

L'objectif de ce contexte est de faire une application python permettant :

- de charger le jeu de données initial dans un dataframe ;
- de rajouter deux colonnes latitude et longitude dans ce dataframe en utilisant l'API Adresse. Observer la documentation de l'API en cliquant sur le lien donné page précédente ;
- d'afficher une carte des **auto-écoles d'une ville avec leur taux de réussite au permis B en affichant des marqueurs de couleurs différentes par taux de réussite. La ville sera saisie par l'utilisateur de l'application.**

Ex de carte pour Niort :



- un graphique permettant de comparer les taux de réussite dans les villes de l'ex région Poitou-Charentes : La Rochelle, Niort, Poitiers et Angoulême.
- une carte nationale de la moyenne des taux de réussite par commune (dataframe à construire).
- un autre indicateur que vous jugerez utile à l'application.

Remarques :

- Pensez à nettoyer vos données en ôtant toutes les lignes non nécessaires ou contenant des valeurs manquantes ou aberrantes pour les variables nécessaires à l'établissement de ces indicateurs.
- Ne chargez vos dataframes qu'avec les données utiles (pénalisation sinon)

3) Evaluation

Les deux livrables seront à restituer sur Updago pour le **16 décembre** au plus tard. Les éléments d'évaluation seront les suivants :

- Opérationnalité (*réponses aux besoins exprimés et respect du cahier des charges*)
- Automatisation de l'outil et optimisation.
- Ergonomie,
- Bonnes pratiques de développement : Commentaires, Indentation, **Découpage modulaire**, Gestion des erreurs avec try catch si besoin...