



ÉTUDIANT EN SCIENCE DES DONNÉES

Dan SEBAG

Formation

2022 - 2025

2-ème année en cours BUT SD
(Science des données, ex STID)
BAC + 3 - IUT Paris Rives de seine

2022

Baccalauréat Technologique
(STI2D)

Coordonnées



07 81 37 41 09



20 Rue du Parc,
92300 Levallois-Perret



dan.sebag06@gmail.com



<https://www.linkedin.com/in/dan-sebag-33a379273/>

Informations complémentaires

Sport

Fort Intérêt pour le sport de tout type, pratique de manière amateur une grande variété de sports en semaine et le week-end

Animateur Scout

Animateur dans un mouvement scout durant l'année (dimanches), et les camps d'été (juillet)

Option théâtre de la seconde à la terminale

Permis B

Compétences

Compétences en statistiques :

- **Test** et Méthodes d'**échantillonnage**.
- Visualisation des données. (création de graphiques)
- Interprétation des résultats

Compétences en Informatique :

- **SQL**
- Programmation en **R**, **Python** et **SAS**
- Analyse statistique **R**, **Excel**
- **Tableau** et **PowerBI**
- Suite **Office**

Autres compétences :

- **Anglais**.
- Esprit d'analyse et de synthèse.
- Communication et esprit d'équipe.

Projets réalisés durant la formation

Collecte automatisée des données du site Action avec Python.

Interrogation de l'API du site Action pour récupérer tous les produits de la catégorie habitat.

- Identification et interrogation de l'API avec Python.
- Stockage des données dans des dictionnaires et nettoyage.
- Transformation des données en fichiers texte individuels par produit.
- Création d'un fichier Excel regroupant tous les produits sur une seule feuille.
- Correction automatique des détails restants, comme l'affichage correct des prix en euros.
- Automatisation du code à l'aide de fonctions pour une utilisation sur différentes catégories de produits.

Analyse statistique d'une base donnée avec R.

Analyse statistique d'une base de données sur la consommation de drogue pendant une partie de jeu vidéo afin de déterminer une potentielle baisse de performances ou des résultats différents.

- Réception de la base de données, traitement et nettoyage avec R.
- Conduction d'un test de corrélation entre la variable de consommation de drogue et les autres variables.
- Élaboration de graphiques mettant en avant les variables les plus corrélées.
- Création de catégories d'individus en vue de les comparer à la variable de consommation.
- Examen approfondi de l'ensemble des résultats et rédaction du rapport final.

L'analyse des données statistiques n'a pas permis de mettre en évidence des différences significatives entre les joueurs ayant consommé de la drogue et ceux qui n'en avaient pas. La taille réduite de l'échantillon et les écarts de performances relativement faibles entre les joueurs rendent les conclusions imprécises