**SAE SCRAPPING FORMULE 1**

**Introduction :**

**Pourquoi avoir choisi la Formule 1 comme sujet de scraping ?**

La Formule 1 est un sport emblématique qui allie histoire, performance et compétitivité. Le choix de collecter et d’analyser le classement des pilotes par année depuis 1950 s’explique par plusieurs raisons :  
La F1, créée en 1950, offre plus de 70 ans de données sur les classements des pilotes. Ces informations retracent l’évolution des performances, des époques de domination (comme celles de Schumacher ou Hamilton) et des rivalités légendaires. Cela permet de mettre en lumière les dynamiques du sport à travers les décennies. Le classement annuel des pilotes est une donnée bien définie et standardisée, rendant son extraction et son analyse plus accessibles. C’est un point d’entrée idéal pour un projet de scraping, car ces données sont souvent présentées sous forme de tableaux ou de listes sur des sites spécialisés. Avec les données de classement, il est possible de mener des analyses comparatives entre différentes époques, en tenant compte des changements dans les règlements, les formats de compétition, et le nombre de courses par saison. Cela offre une perspective unique sur l’évolution de la compétitivité dans ce sport. Bien que le classement annuel des pilotes semble être une donnée simple, le scraping de ces informations peut poser des défis intéressants, notamment si les données proviennent de sites historiques ou dynamiques. C'est une opportunité d’affiner des compétences techniques tout en traitant des informations cohérentes et exploitables. Travailler sur un sujet lié à un sport aussi passionnant que la F1, avec une communauté mondiale de fans, est une source de motivation. Ce projet combine la curiosité pour l’histoire de la F1 avec le plaisir de manipuler des données significatives pour produire des analyses et des visualisations intéressantes. En somme, choisir la F1 et ses classements comme sujet de scraping permet d’explorer un pan riche et fascinant de l’histoire du sport automobile, tout en relevant un défi technique et analytique stimulant.

**Description des rôles :**

Bastian : Analyse, Visualisation

Théo : Collecte, Traitement

**Contribution individuelle :**

Lors de cette SAE, nous avons contribués tous les deux à hauteur 50% chacun sur chaque partie de la SAE.

**La collecte :**

Nous avons découpé la conception du code afin de travailler plus vite. Théo s’est occupé du code de récupération et Bastian des url à indiquer pour récupérer les données.

Contribution : 50% chacun

**Le traitement :**

Bastian s’est occupé de la mise en format en csv avec l’aide de Théo.

Contribution : 50% chacun

**L’Analyse et la Visualisation :**

Lors de l’analyse Bastian et Théo ont réaliser le Power BI.

Contribution : 50% chacun

**Site web choisis et légalité :**

Le site web que nous avons choisis est le site [fomula1](https://www.formula1.com/) qui recense de nombreuses statistiques sur la formule 1

Nous avons pu accéder à toutes les données souhaitées excepté les tags dont nous n’avons aucune utilité comme l’indique la sortie du robots.txt.

Sitemap: https://www.formula1.com/sitemap.xml

User-Agent: \*

Disallow: /en/latest/tags/\*

Allow: /

**Librairies choisies :**

Java.io pour créer le csv.

Jsoup pour le scrapping.