

LEVY, DAN
LOUPRET, JEAN-CHRISTOPHE
MARTEL, BENJAMIN
PERRON-ABRAN, PASCAL

Proposition de projet
ÉQUIPE 6

Travail présenté à Christian Gagné
GIF-4101

Faculté des sciences et génie
Université Laval
22 octobre 2025

Analyse de données FastF1

Énoncé du problème

Dans la Formule 1 moderne, les écuries utilisent l'analyse rapide et précise de grandes quantités de données télémétriques pour prendre des décisions. Idéalement, l'analyse de ces données devrait se faire rapidement et simplement pour optimiser les stratégies et les décisions de l'écurie efficacement.

Les données officielles de la Formule 1 sont disponibles via un API, mais l'utilisation directe de celles-ci est complexe. Ces données sont brutes, peu documentées et il est difficile de faire des liens entre plusieurs aspects (télémétrie, météo, événements en piste).

Notre principal enjeu est donc de transformer ces données brutes en informations pertinentes et compréhensibles. Cette situation est problématique, car sans méthode d'analyse claire et reproductible, l'interprétation des données reste limitée.

Proposition d'amélioration

Le projet vise à concevoir une plateforme d'analyse des données de Formule 1 à partir de la bibliothèque FastF1, permettant :

- l'extraction automatisée des données de performances historiques;
- la génération de visuels comparatifs en fonction de paramètres clés (pilote, conditions météorologiques, tours, type de pneu, etc.);
- l'analyse statistique de la performance et de la régularité;
- la création d'un rapport sur les stratégies et les écarts de performance.

L'objectif final est d'identifier les bonnes et les mauvaises stratégies à utiliser durant une course pour maximiser la performance.

Méthodologie proposée

Le projet suivra un pipeline typique d'analyse de données en Python :

1. **Collecte et préparation des données**
 - Utilisation de l'API FastF1 pour télécharger les données;
 - Nettoyage des données aberrantes et la transformation en DataFrames pandas locaux pour réduire le temps d'accès aux données.

2. Analyse exploratoire et statistique

- Étude de la régularité en fonction des paramètres clés;
- Détection d'événements clés (arrêts aux stands, safety car, incidents);
- Utilisation de méthodes statistiques simples pour interpréter les performances.

3. Modélisation et visualisation

- Développement d'un module Python générant automatiquement des graphiques interactifs (matplotlib, seaborn, plotly);
- Création d'un tableau de bord ou d'un rapport automatisé présentant les résultats;
- Implémentation d'un modèle prédictif basique (ex. : régression linéaire) pour estimer les temps au tour attendus selon les paramètres (type de pneu, charge de carburant, etc.).

4. Validation et documentation

- Validation de la cohérence des données et comparaison des résultats avec d'autres sources;
- Rédaction d'une documentation claire.

Jeux de données proposés

Source et fiabilité

Les données seront extraites à partir de la bibliothèque libre d'utilisation FastF1 (<https://docs.fastf1.dev>), qui s'appuie directement sur les flux de télémétrie de la Formule 1 Live Timing. FastF1 est maintenu activement sur GitHub par une communauté d'ingénieurs et d'analystes indépendants, garantissant une fiabilité élevée et une bonne compatibilité avec les Grands Prix récents (à partir de 2018).

Pourquoi ce jeu de données ?

Le choix de la bibliothèque FastF1 pour l'extraction des données s'explique par sa richesse et son accessibilité. Le jeu de données qu'elle fournit est complet, incluant des informations détaillées sur les tours, les secteurs, les pneus, la météo, et d'autres aspects pertinents. De plus, ce jeu de données est librement accessible, sans droits d'auteur restrictifs, ce qui facilite son utilisation dans le cadre du projet. Enfin, l'utilisation de ces données permet de traiter un problème réaliste et concret dans le domaine de l'ingénierie des données, offrant ainsi une application pratique et significative.