

**Formula One Race Data**

Maxime THOOR  
Louis TINEL  
Jules ROUSSEAU  
Nils CHARRIER  
Edouard DANSETTE  
Pierre-Maxime COSTA

1. **Présentation de Formula One Data Race**

* L’entreprise

Formula One Data Race (FORD) est une société d’études et d’analyses d’écuries et de pilotes de formule 1. En se basant sur l’étude des statistiques des Grand Prix de formule 1, elle délivre des pronostics et conseille les investisseurs et parieurs, particuliers comme professionnels.

Pour pouvoir fournir des pronostics les plus fiables possibles, FORD s’appuie sur un set de données complet et varié et analyse chaque aspect des courses de Formule 1 afin d’anticiper quel pilote ou quelle écurie a le plus de chance de remporter le prochain Grand Prix. Les investisseurs peuvent ainsi investir sur les constructeurs automobiles et les parieurs sur les pilotes.

* Les membres de l’entreprise

Maxime THOOR, chef de projet

Louis TINEL, s’occupe de l’extraction des données et du nettoyage

Jules ROUSSEAU, s’occupe du Data Warehouse

Nils CHARRIER, s’occupe de la partie Datamart

Pierre-Maxime COSTA et Edouard DANSETTE, s’occupe du reporting et de l’analyse des données

1. **Présentation des données**

Les différents sets de données sont les suivants :

*circuit.csv drivers.csv results.csv  
 constructor\_results.csv lap\_times.csv seasons.csv  
 constructor\_standing.csv pit\_stops.csv status.csv  
 constructor.csv qualifying.csv  
 driver\_standing.csv races.csv*

Ces jeux de données contiennent toutes les informations nécessaires sur les statistiques des courses des Grand Prix. Les statistiques concernent chaque aspect des courses : les temps réalisés par chaque pilote en course et en qualification, leurs positions, leurs points gagnés. Ces statistiques sont également valables pour les constructeurs.

Avant d’effectuer un éventuel tri, l’ensemble des membres de l’équipe se sont répartis sur différentes études afin d’en tirer des conclusions évidentes pour la suite de notre projet. Grâce à cela, nous avons choisi d’en éliminer 3 d’entre eux qui sont :

*constructor\_standing.csv*

*driver\_standing.csv*

*seasons.csv*

En effet, chacun de ces datasets ne possédaient pas d’éléments majeurs pour délivrer des statistiques importantes pour notre entreprise.

Eléments qu’on ne peut pas prendre en compte (le fait que plusieurs constructeurs collaborent sur une même voiture)