The goal

The main research goal is to set a program that will predict the race time for each driver. This program will be useful for the constructors and for the gamblers. It will depend on the following features:

- Weather
- Driver
- Track
- PitStop Time

Datamining & Machine learning algorithm

For this project, we will implement a multiple regression on these different variables:

Variables	Example
Y= Race Time	*1:30:00*
X1=Weather (Boolean)	*Sunny*
X2=Driver (Hot Encoding)	*Hamilton*
X3=Track (Hot Encoding)	*Mexico*
X4=PitStop Time	*2.55"*

The Dataset

We found our dataset on Kaggle, he is related to all statistics about the Formula 1. This dataset has 13 different files. We will focus on different files such as the "result" one. Its dimensionality is 18 for a cardinality of around 23770. For the weather, we will have to find another dataset that we will merge to the actual one.