

The goal

The main research goal is to set a program for the constructors' managers and engineers which can predict the maximum accrued *Pitstop* time that a team can afford depending on the:

- Race time
- Weather
- Driver
- Track

So, thank to this program the managers can know what is the eligible maximum accrued *Pitstop* to win the race or to finish in the targeted race time.

Datamining & Machine learning algorithm

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

Variables	Example
• $Y = \text{Max accrued PitStop Time}$	<i>YYY</i>
• $X1 = \text{Race Time}$	<i>Targeted Time 1:30:00</i>
• $X2 = \text{Weather (Boolean)}$	<i>Sunny</i>
• $X3 = \text{Driver (Hot Encoding)}$	<i>Hamilton</i>
• $X4 = \text{Track (Hot Encoding)}$	<i>Mexico</i>

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 "pitStops" one. Its dimensionality is 7 for a cardinality of around 6250. For the weather, we will have to find another dataset that we will merge to the actual one.

