**Formula 1 history through data analysis**

The selected topic for our project is "Formula 1 history throughugh data abalysis", we seleceted this topic because of all the available and verifyable information available, the diverse variables that determine the outcome of this sport and twy to solve through data, the main questions all F1 fans have asqued themselves at some point. What provides the best leverage in the sport, the man, the machine or the strategy?

**Dataset**

We looked for information that can be usable for this purposes, the characteristics of the data that we could use for the purposes of our project are the following:

* Should be accurate.
* Should be verifyable.
* Should be diverse enough to perform our analysis.

We found a group of datasets that could be useful for our pusposes on the following link:

<https://www.kaggle.com/williamhaworth/formula-1-kaggle-learning/data>

In the link mentioned above, we found datasets related to Formula 1 racing from the year 1950 to 2021, the datasets that we will use are formatted in csv and contain information about:

circuits.csv constructor\_results.csv constructor\_standings.csv constructors.csv driver\_standings.csv drivers.csv lap\_times.csv pit\_stops.csv qualifying.csv races.csv results.csv seasons.csv status.csv

**Questions to be answered**

Ww will try to answer through data, questions that fans and analysts have asked about this sport since its very beginning, which is more important? The man, the machine or the strategy? we will leave budget ans sponsors out of the analysis this time.

Some of the questions we have initially designed with the available datasets are the following:

* How relevant is the driver vs the car to determine a championship output?
* How relevant is the pit stop average time to a championshio output?
* Which are the most relevant mechanical failures to determine a constructor, championship output?
* How relevant is the number of pole positions to the championship output?

**Machine learning models**

After cleaning and grouping the relevant datasets, we will use a non supervised machine learning model to expose relevant correlations that are not evident, select the most meaningulf ones and then predict possibles outcomes through supervised machine learning.

1.- Quién puede ganar el proximo F1?

Posiciones Eliminar puntos

Corredores

1. Es probable que el que sale en primer lugar termina en primer lugar
2. Tiempos de pits afectan el resultado final
3. Fallas mas cruciales en equipos de los últimos lugares
4. Constructores históricos por malos resultados
5. A los cuantos resultados negativos el constructor deja de participar
6. Tiempos de pits stops proporcionales a la posición en la carrera
7. Variables pits stops, tiempos de calificación
8. Factor mas determinante
9. Modelos de años anteriores, de años atrás
10. Factores que han cambiado en la F1
11. Standing de los corredores en la f2, en charting, indi

Si siempre estas en el numero uno en la f2 en la f1

**Que factor es mas importante en la carrera, el piloto o la constructora.**

Cuales son los factores que tienen mas peso en el resultado de la carrera

Cuales son los factores que afectan el resultado de una carrera?

* Tiempos de pits
* Ganar calificación
* Cantidad de campeonatos de constructora
* Cantidad de fallas de constructora
* An;os de experiencia del piloto
* Mejores tiempos /lap times
* Pista que se corre
* Resultados previos de constructoras y pilotos

Cluster de factores de la escuderia y del piloto main

sergio\_gaytan - este es mi prueba de commi