|  |  |  |
| --- | --- | --- |
| Checkpoint I | Checkpoint I: Project Proposal | |
| Group: | <G01> |
| Date: | <2023/09/15> |
|  |  |

# Domain

“Qual o melhor piloto de Fórmula 1 em cada circuito”

# Dataset Description

O dataset a ser utilizado foi re tirado do [kaggle](https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?select=drivers.csv) que consiste em varios ficheiros .csv, como por exemplo lap\_times, circuits e results.

This dataset is valuable for Formula 1 enthusiasts, analysts, and researchers interested in studying the history and performance of Formula 1 drivers, teams, and races over several decades. Researchers and data analysts can use this data to perform various analyses, including driver performance over time, team performance, historical trends, and more. Please refer to the dataset's documentation on Kaggle for specific details about its structure and contents.

# Example Questions

1ª: Como evoluiu o lap time e o pit stop time por circuito ao longo dos anos?

De forma a responder a esta questão vamos utilizar os valores de Lap Time e Pit Stop Time recolhidos ao longos dos anos de um determinado circuito selecionado, que podemos obter através do nosso dataset e apresentá-los de forma a vermos a sua evolução.

2ª: De entre os pilotos em comparação qual deles possui o melhor Average Lap Time ?

Para tal, de entre os pilotos e circuito selecionados vamos obter os seus tempos médios de volta ao circuito e utilizá-los de forma a perceber qual teve a corrida mais rápida e estável.

3ª: Como se comparam as performances gerais dos pilotos tendo em conta um circuito selecionado onde ambos correram?

Nesta questão, usaremos dados como posição inicial / posição final, volta mais rápida, entre outros, de forma a poder fazer uma comparação final e mais conclusiva (por exemplo num radar chart onde os valores máximos são as melhores estatísticas já obtidas no circuito em análise) da performance de dois pilotos que já tenham corrido no circuito selecionado mesmo que em anos diferentes.

4ª: Qual a diferença entre as estratégias utilizadas pelos pilotos em cada circuito quando comparados os Lap Time ao longo da corrida?

Para responder a esta questão, vamos analisar a variação das velocidades de cada volta da corrida de forma a perceber se a estratégia utilizada pelo piloto/equipa foi a melhor quando comparada a outros pilotos no mesmo circuito.

5ª: Num determinado circuito qual dos pilotos em comparação conseguiu perder menos tempo em Pit Stop Time?

Na quinta questão, vamos buscar todos os tempos em que os pilotos se encontraram no pit stop num determinado circuito comparando-os entre pilotos de forma a chegar a uma conclusão sobre o overall dos pilotos quando comparados nas mesmas condições.

# Data Sample

Seleção de circuito em análise:

from: circuits.csv

circuitId,name,country,lat,lng

70,"Red Bull Ring","Austria",47.2197,14.7647

Desenvolvimento do tempo por volta ao longo de um ou mais Grand Prix's:

from: lap\_times.csv

raceId,driverId,lap,position,time

841,808,6,4,"1:33.573"

Comparação do tempo de paragem entre multiplos carro:

from: pit\_stops.csv

raceId,driverId,stop,lap,duration

841,153,1,1,"26.898"

Ligação entre seleção de circuito e os seus resultados:

from: races.csv

raceId,circuitId,name,date

1036,4,"Spanish Grand Prix","2020-08-16"

Evolução dos resultados de um piloto ao longo dos demais Grand Prix's:

from: results.csv

resultId,raceId,driverId,grid,position,fastestLapTime,fastestLapSpeed

26078,1110,855,17,13,"1:50.188","228.830"