Data Analyst Assignment

Context

The Data Scientists at your client's organisation would like to build a machine learning model associated with the F1 race results dataset to predict the constructor and driver race winners. The datasets are located in separate csv files. They require an exploratory data analysis to be carried out on the data. More information about the data can be found here.

The requirements are as follow:

results.csv

Parent table - all subsequent table relationships and aggregations use this dataset as a reference

status.csv

The csv file contains the status of a driver's race.

constructors.csv

The csv file containing the constructors' information

constructor_standings.csv

The csv file containing the constructors' race standings information

races.csv

The csv file contains all the race information related to the results.

drivers.csv

The csv file contains all the drivers' information related to the results.

circuits.csv

The csv file contains the race circuit's information.

pitStops.csv

The csv file contains the pit stops that a driver took during a race

lapTimes.csv

The csv file contains the driver's lap times during a race.

driverStandings.csv

The csv file contains the driver standings.

Task

Develop the datasets and analysis to complete the following:

- Carry out an analysis and report on the constructors' success based on championship season wins.
- Carry out an analysis and report on the drivers' success based on race wins.
- Carry out an analysis and report on the risks associated with race tracks based on accidents and collisions.
- Carry out an analysis and report on the fastest race tracks based on lap times.
- Recommend engineered features to the Data Scientists for the ML model they are building. Analyse and explain why you'd think the engineered features would boost ML model performance.

Output

- Source code
- Combined dataset as per the requirements, in a data file format of your choice.
- Analysis and reports including but not limited to tables and charts

Summary of Requirements:

- Complete the task and create artifacts like you would in the context of the job that you have applied for.
- We expect to see code which you would be happy to put in production. (Code that can be run locally as well as in a cloud provider of your choice)
- Choose whatever language you're comfortable with.
- Document your code.
- Explain your design decisions.
- If something is not clear don't hesitate to ask or just make an assumption and go with it.