Fachhochschule Dortmund
University of Applied Sciences and Arts

Report

Machine Learning With Python

Assignment-1

By

Raghupati Bhetwal

FH Dortmund

1. Retrieve data from different sources as shown below and saved in .csv Format.

Data1	https://github.com/toUpperCase78/formula1-datasets/blob/master/formula1_2019season_drivers.csv
Data2	https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020/code?datasetId=468218&sortBy=voteCount
Data3	https://www.kaggle.com/datasets/deepcontractor/froza-horizon-5-cars-dataset
Data4	https://data.world/ogletree/tour-down-under-2018
Data5	https://github.com/toUpperCase78/formula1-datasets/blob/master/formula1_2021season_drivers.csv

2. Data2 is taken for the Visualization and Data Pre-processing:
    a. Read the data and merged.

```
# merging all seperate dataframe into single dataframe as df
con1 = pd.merge(result_df, races_df, on ='raceId')
con2 = pd.merge(con1, drivers_df, on = 'driverId')
con3 = pd.merge(con2, driver_standings_df, on = 'driverId')
con4 = pd.merge(con3, constructor_df, on ='constructorId')
df = pd.merge(con4, stats_df, on ='statusId')
pd.get_option("display.max_columns",None)
df.head()
```

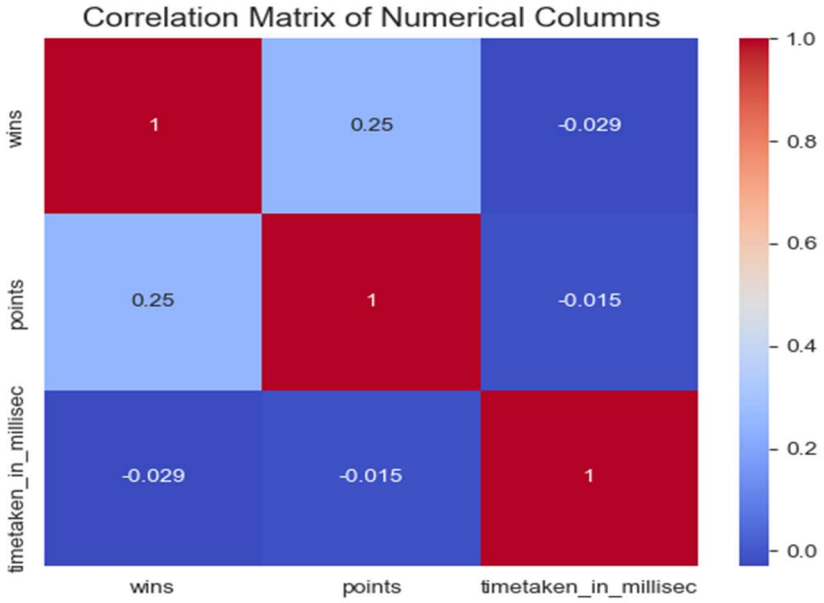
	resultId	raceId_x	driverId	constructorId	number_x	grid	position_x	positionText_x	positionOrder	points_x	...	raceId_y	points_y	position_y	positionTe
0	1	18	1	1	22	1	1	1	1	10.0	...	18	10.0	1	
1	1	18	1	1	22	1	1	1	1	10.0	...	19	14.0	1	
2	1	18	1	1	22	1	1	1	1	10.0	...	20	14.0	3	
3	1	18	1	1	22	1	1	1	1	10.0	...	21	20.0	2	
4	1	18	1	1	22	1	1	1	1	10.0	...	22	28.0	3	

    b. **Checked Missing values:** There was no missing values present.

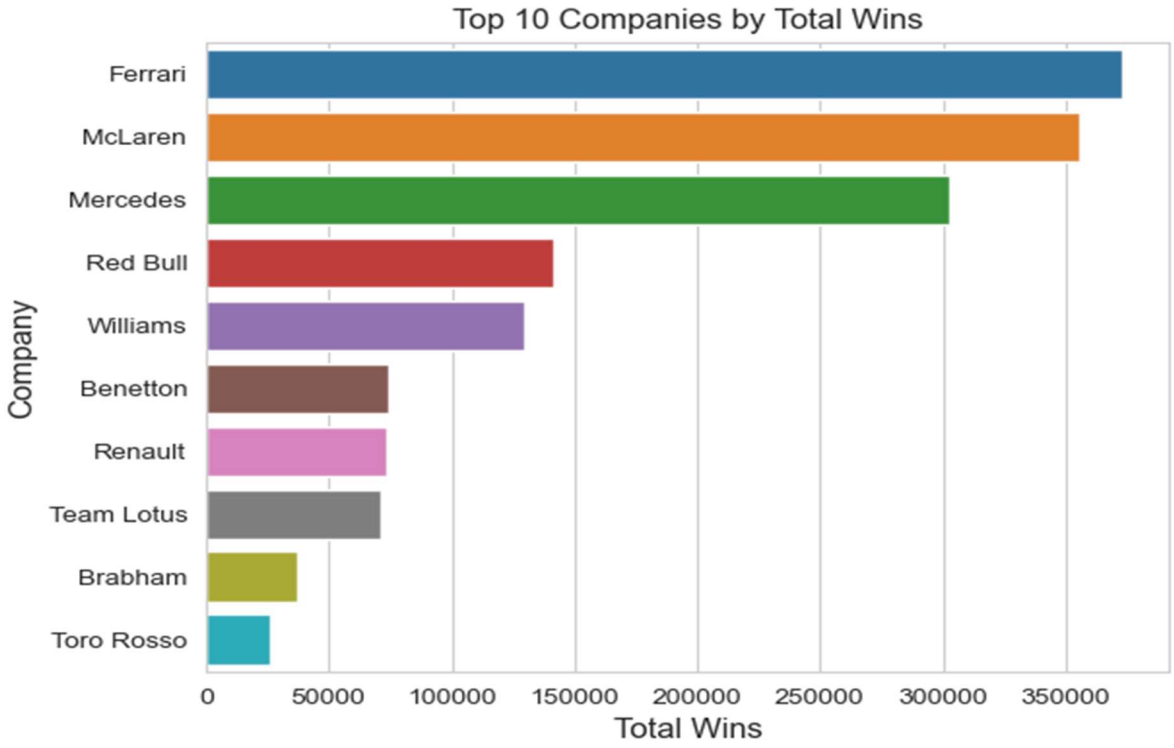
```
In [10]: df.isnull().sum().sum()

Out[10]: 0
```

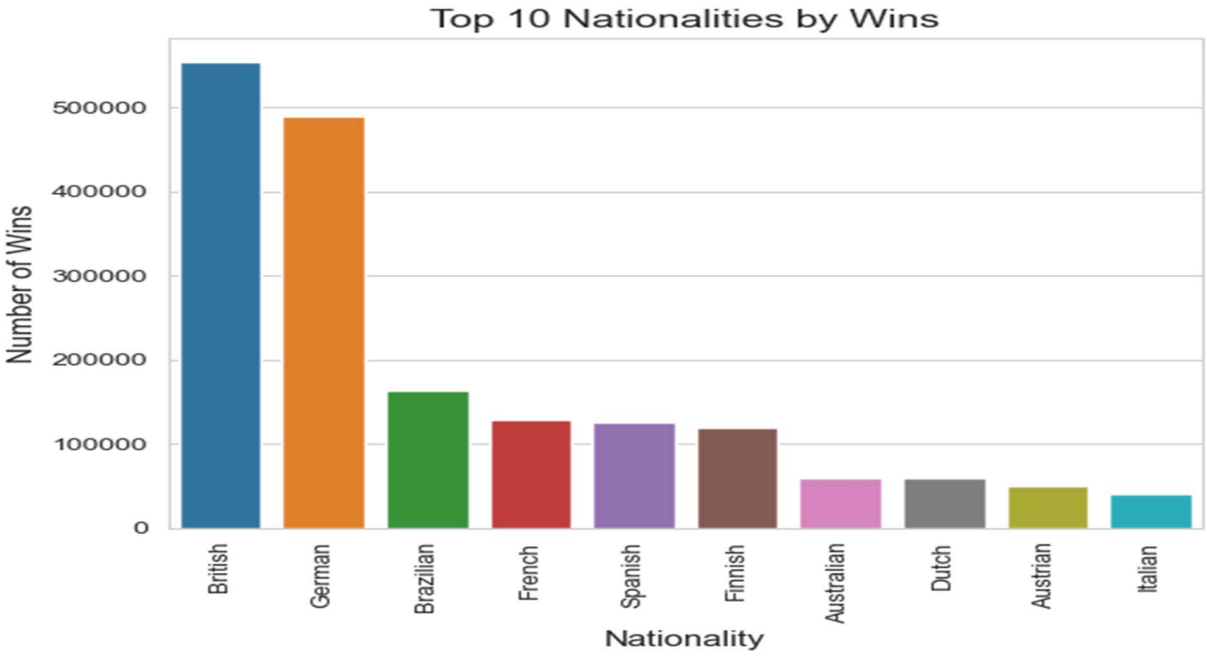
c. Taken some of the columns data Visualization:
Correlation Matrix: The correlation matrix displays the relationship between numerical variables in the data. The heatmap shows ‘timetaken_in_millisec' has a strong negative correlation with 'wins' and 'points' (i.e. inverse relation).



d. **Top 10 Car Companies by Wins:** The bar chart displays the top 10 companies with the highest total wins. It shows that “Ferrari” has the highest number of wins, followed by ‘McLaren’ and 'Mercedes'.



e. **Top 10 Nationalities by Wins:** The bar chart displays the top 10 nationalities with the highest number of wins. It shows that drivers from 'British (Britain)' have the highest number of wins, followed by 'Great Britain' and 'Brazil'.



f. **Wins by Company and Driver Nationality:** The stacked bar chart displays the number of wins by company and driver nationality for the top 10 companies. It shows that 'British' has the highest number of wins across various driver nationalities, followed by 'Mercedes' and 'McLaren'.