

DATA VISUALIZATION PROJECT FORMULA 1 DATA

By Group 8 :

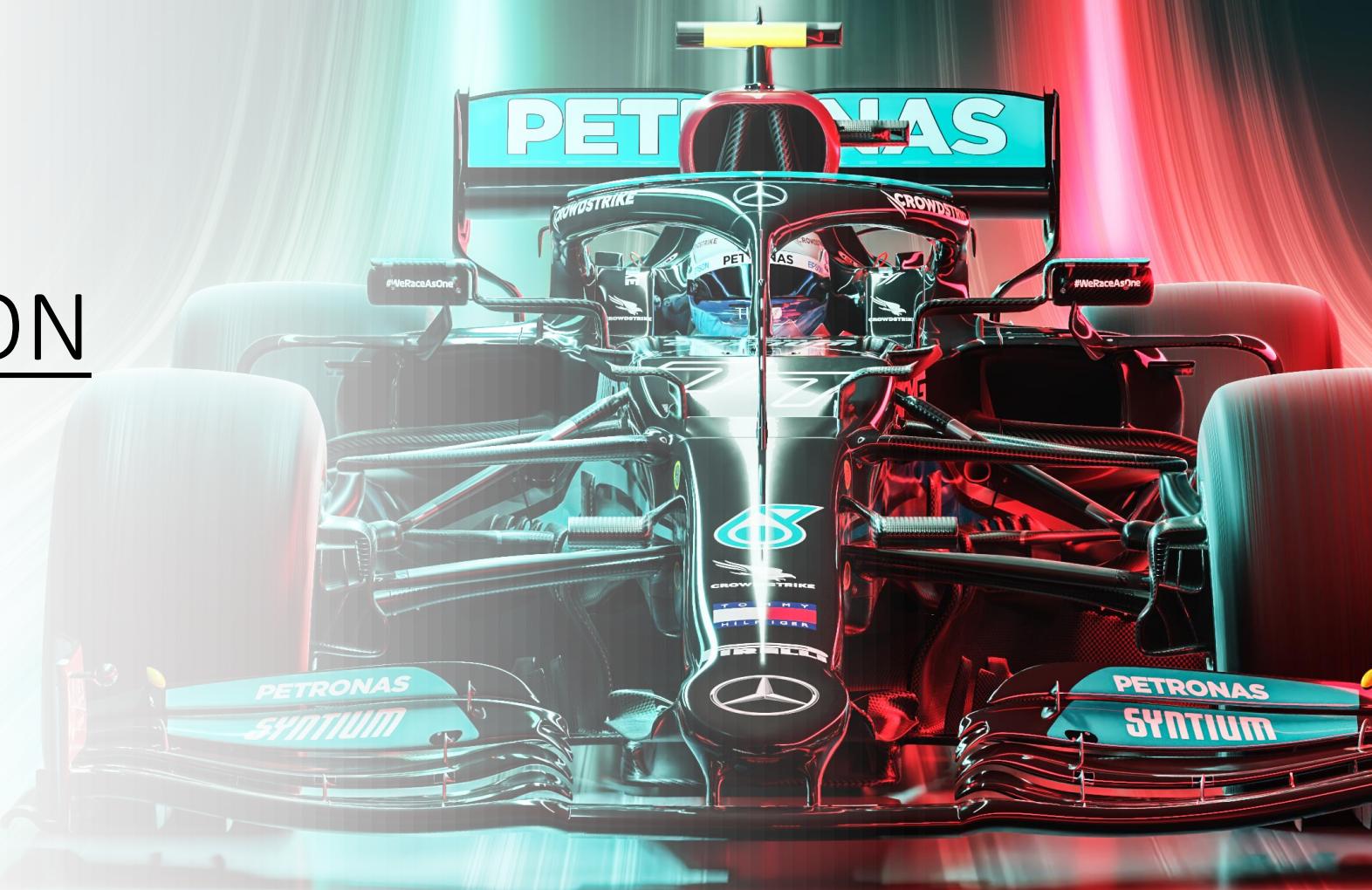
Bhavik Patel

Namratha Sampath

Smeet Sheth

Sourab Saklecha

Yuti Khamker



AGENDA

1. Introduction
2. Motivation
3. Objective
4. Development Life-Cycle
5. Project Timeline
6. Literature Review
7. Data sources
8. Data Process Diagram
9. EDA
9. Linear Regression
10. Visualization
11. Data Model Review
12. Conclusion



INTRODUCTION

What is F1?

- Pinnacle of motor-sport. Top 10 teams of the World. 20 Top Drivers
 - Complex and data driven
 - Extreme precision and accuracy



A photograph of a man in a racing team pit garage. He is wearing a blue zip-up hoodie with 'ALPHATAURI' and a Formula 1 logo on the sleeve, and a black headset with a microphone. He is looking intently at a large array of computer monitors displaying various race-related data, including track maps and performance metrics. The background shows other team members and equipment, creating a busy and focused atmosphere.

Motivation

- 14,500 Components in a car.
- 300+ Sensors on every car
- 1.5 terabytes of data generated-every weekend!
- 11.8 Billion data points generated by a team every year.

PROJECT OBJECTIVES

Visualize and Analyse historic performance
of Teams
using Power BI/Tableau



Trend Analysis for new data (2023
Season) vs historic data
Using Python



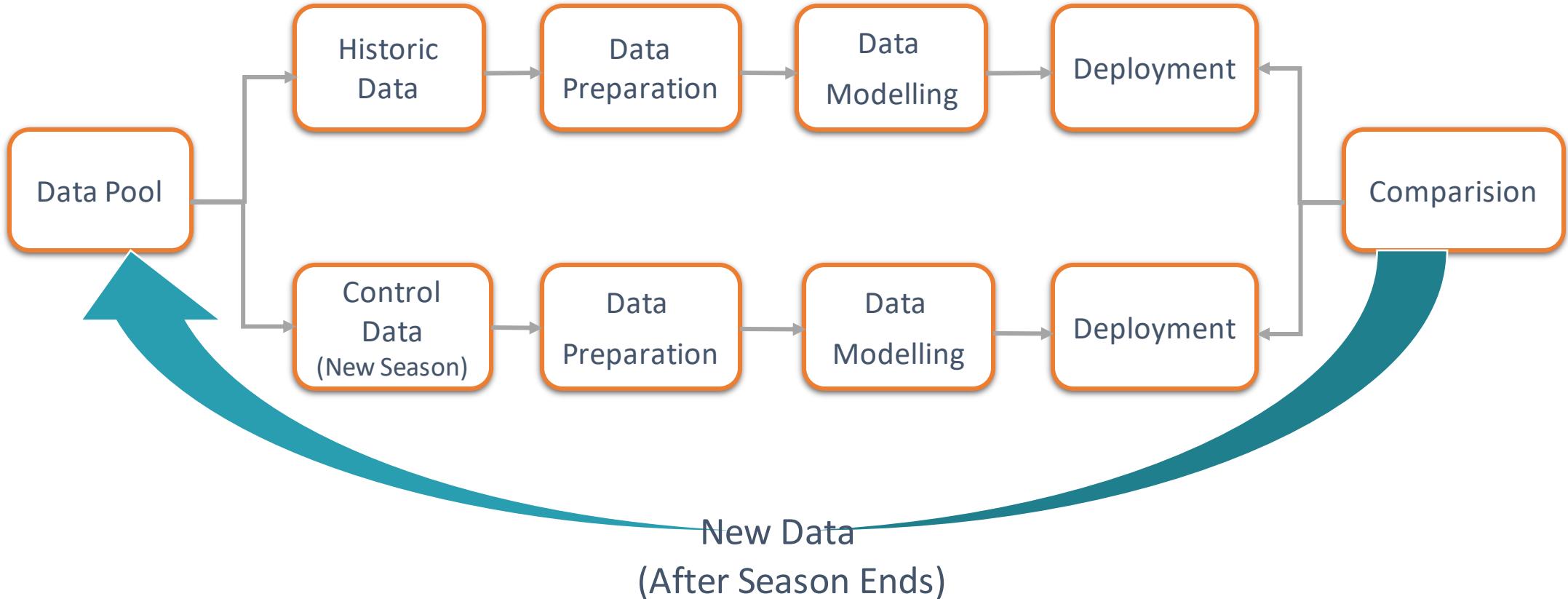
Visualize and Analyse historic performance
of Drivers
using Power BI/Tableau



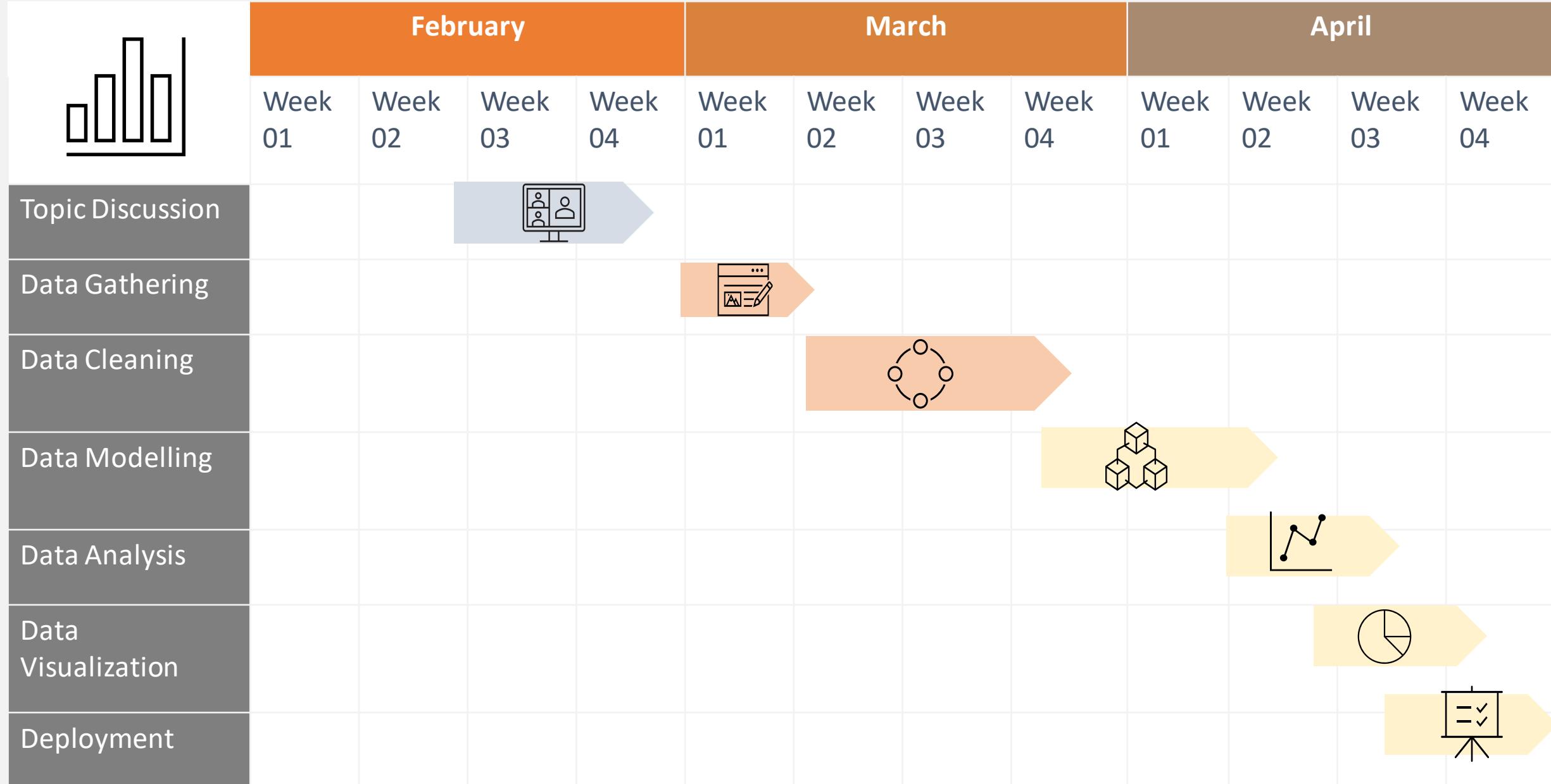
Analyse the impact of changes to technical
regulations on the performance of the car.
using Python



Hybrid Waterfall and incremental Model



Project Timeline



Literature Review

Name of the Article	Model Used	Results
Vishnu Singh Khatuwal, Digvijay Puri. 'Business Intelligence Tools for Dashboard Development'	Data Analysis Methods, Dashboard Design Principles, Self-Service BI, Connectivity and Data Integration, Data Visualization Techniques, User Interaction and Exploration	The paper provides the applications of BI dashboard technologies in real time scenarios.
Marco Capo ; Aritz Pérez ; Jose Lozano. 'An efficient K-means clustering algorithm for massive data'	Clustering, parallelization, unsupervised learning, K-means, K-means++, Mini-batch. Boundary Weighted K-means algorithm.	The Boundary Weighted K-means method (BWKM), a clustering algorithm, is introduced that offers both theoretical assurances and real-world competitiveness in its attempt to tackle challenges with large datasets.
Prescott Delzell, Patrick McCabe, Adam Mourad. 'Automation Of Data Analysis In Formula 1'	Process of post-race analysis	Companies may improve their racing efficiency and perhaps climb the Formula 1 standings by using these dashboards.
Ankur Patil, Nishtha Jain , Rahul Agrahari. 'A Data-Driven Analysis of Formula 1 Car Races Outcome'	Principal Components Analysis	Using a set of statistical techniques, they concluded that most of the variables are strongly correlated with each other. The original feature space can be significantly reduced to a lower-dimensional subspace without a significant loss of information.

DATA SOURCES

Dataset from Kaggle

Fast API

Ergast API

Wikipedia

FIA website

Snapshot of Survey

Formula 1 Survey

smeetpiyush.sheth@sjsu.edu Switch account  Draft saved 

Name: Smeet Sheth 

Age: 22

Gender: Male Female Other: 

E-mail: Your answer

Please select your preferred way of watching F1 races:

At venue
 TV
 Online
 Mobile
 Other: _____

Preferred Team: Mercedes

Preferred Driver: Lewis Hamilton

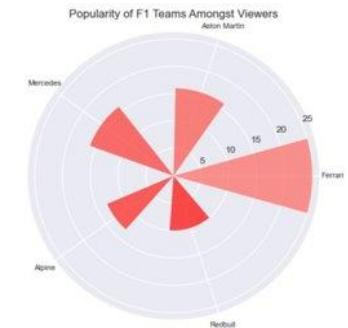
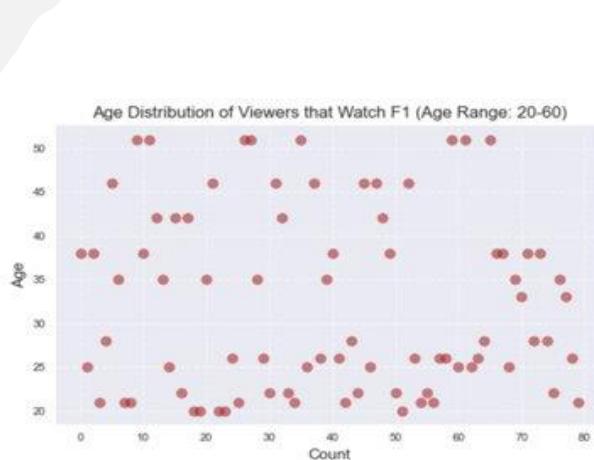
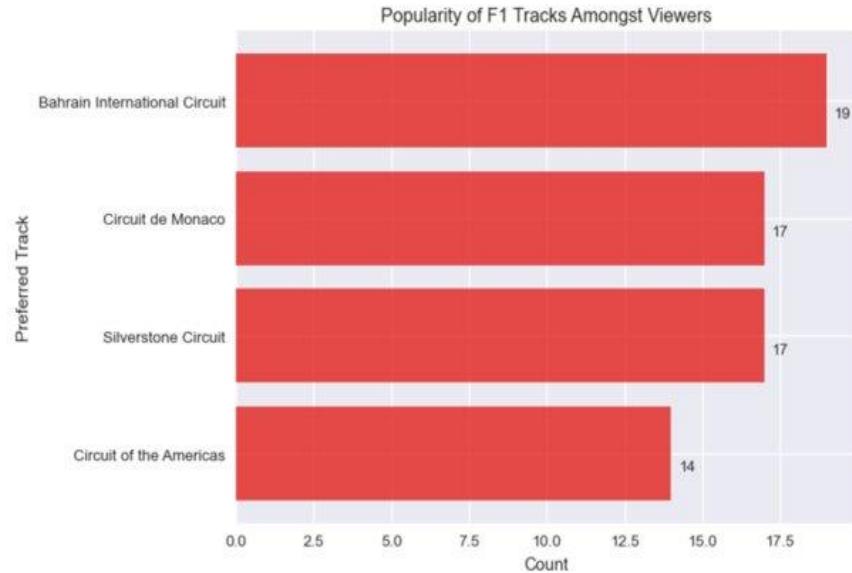
Preferred Circuit: Abu Dhabi

Please select the platform you use to receive updates about F1 races and news:

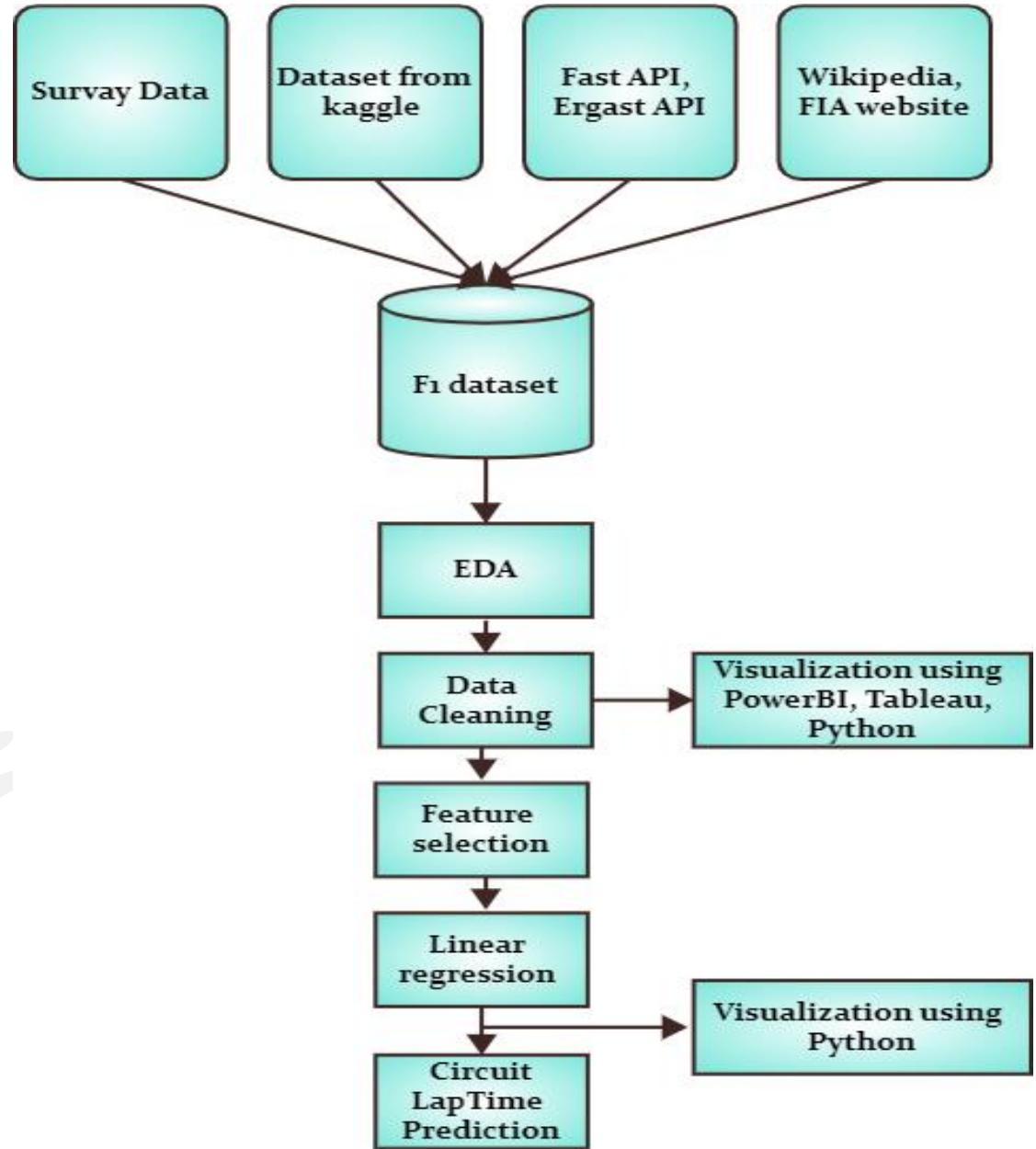
Official Website
 Fan forums
 TV
 Social Media
 Other: _____

Real Time Insights

- The polar chart gives information on how Ferrari is the most popular F1 team amongst the participants.
- Most of the participants that watch F1 from our survey belong to the age group 20-30.
- Bahrain International Circuit is the most popular track amongst our respondents.



Data Process Diagram



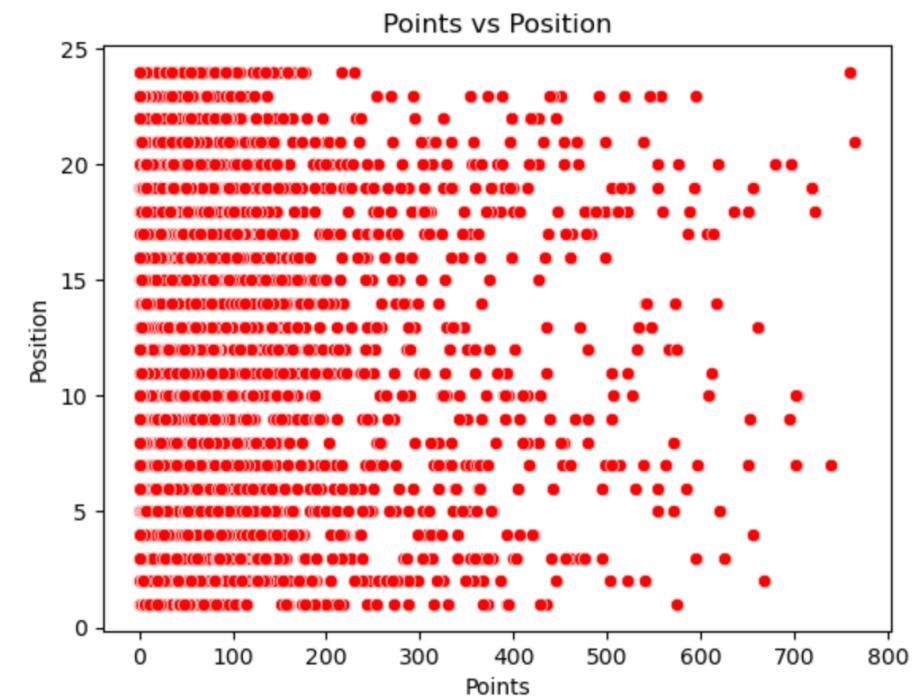
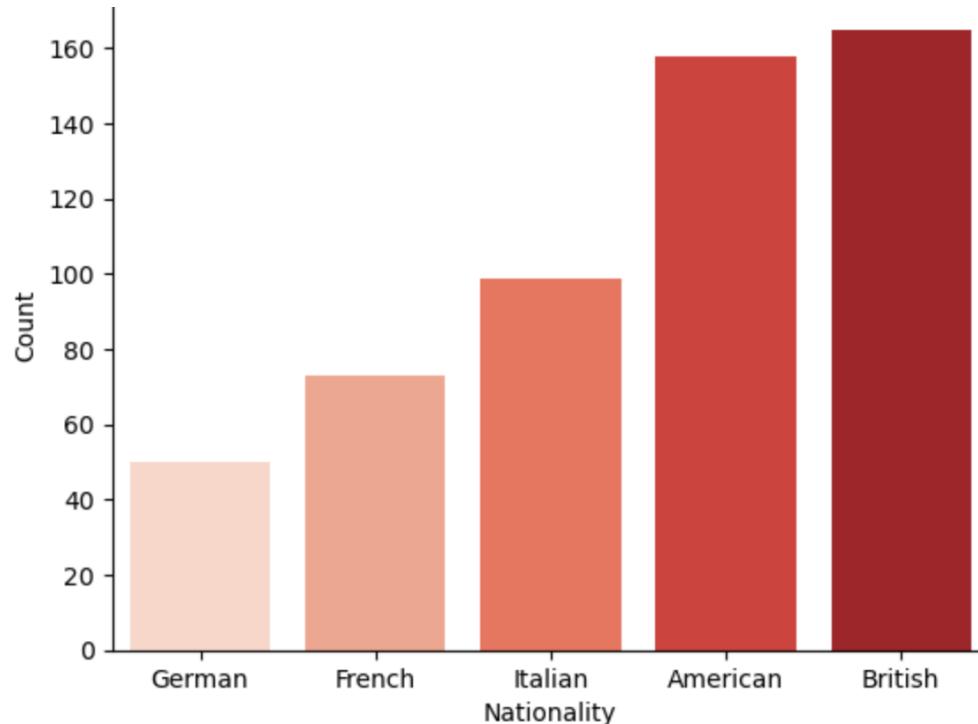
Exploratory Data Analysis

- The exploratory data analysis was conducted concurrently with the data cleaning.
- An analysis of the correlation between the columns was conducted using a correlogram.

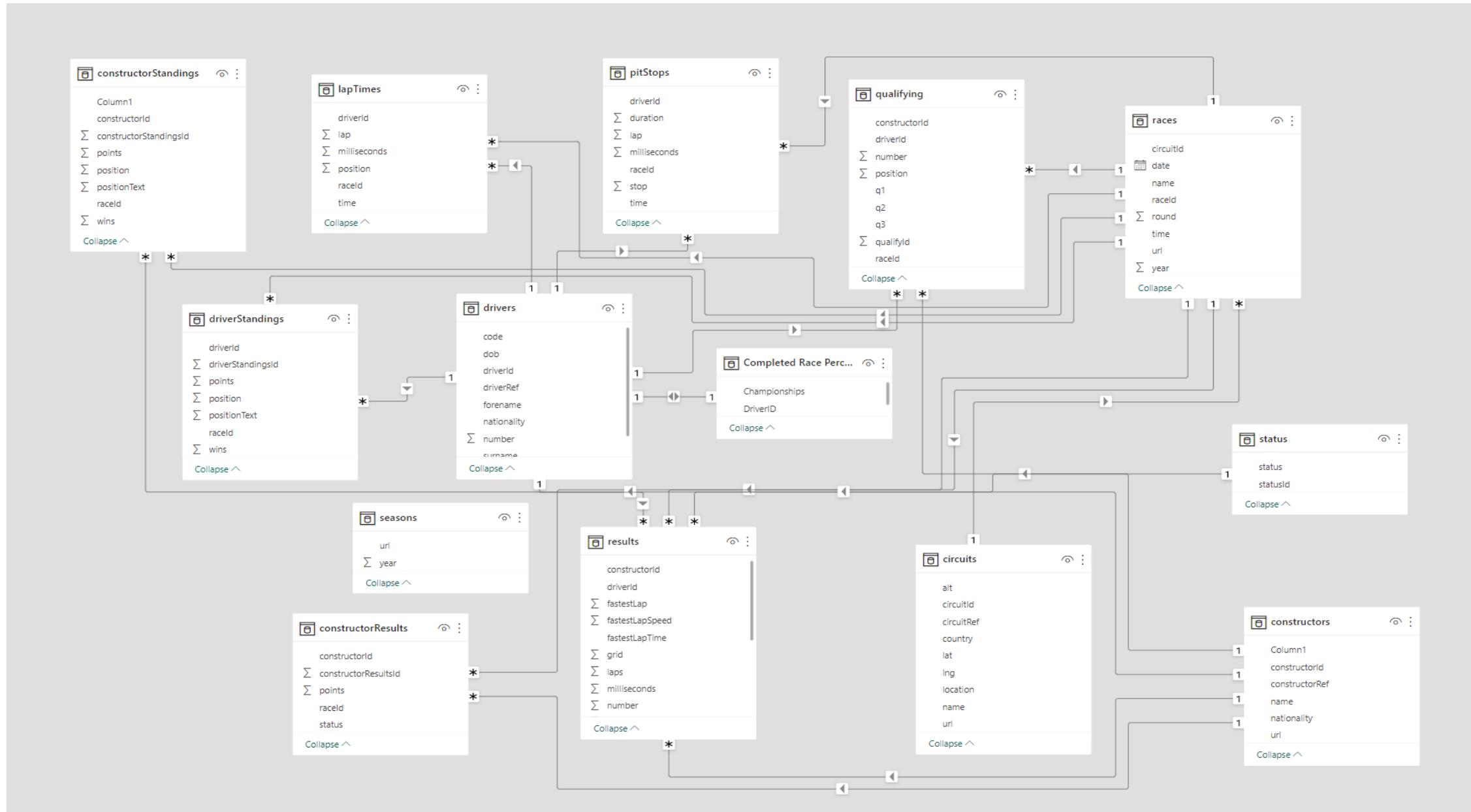
Correlation Matrix														
driverId	1.00	0.17	-0.21	-0.00	0.88	0.20	0.08	1.00	0.08	0.17	1.00	0.31	0.85	-0.11
lat	0.17	1.00	-0.19	0.09	-0.03	-0.16	-0.14	0.17	0.09	0.05	0.17	0.17	0.17	-0.07
lng	-0.21	-0.19	1.00	-0.20	0.11	0.05	-0.11	-0.21	0.01	-0.02	-0.21	-0.21	-0.21	0.24
lap	-0.00	0.09	-0.20	1.00	-0.13	-0.07	-0.04	0.09	0.01	0.09	0.15	0.04	0.02	-0.76
position	0.88	-0.03	0.11	-0.13	1.00	0.03	0.01	0.75	0.07	0.16	0.09	0.06	-0.05	0.27
milliseconds	0.20	-0.16	0.05	-0.07	0.03	1.00	-0.01	0.06	-0.07	0.21	0.07	0.02	-0.03	0.20
points	0.08	-0.14	-0.11	-0.04	0.01	-0.01	1.00	0.23	-0.04	0.10	-0.21	0.24	0.09	0.10
constructorId	1.00	0.17	-0.21	0.09	0.75	0.06	0.23	1.00	0.08	0.17	1.00	0.50	1.00	-0.12
statusId	0.08	0.09	0.01	0.01	0.07	-0.07	-0.04	0.08	1.00	0.26	0.08	0.08	0.08	0.02
positionOrder	0.17	0.05	-0.02	0.09	0.16	0.21	0.10	0.17	0.26	1.00	0.17	0.17	0.17	-0.30
qualifyId	1.00	0.17	-0.21	0.15	0.09	0.07	-0.21	1.00	0.08	0.17	1.00	0.59	0.02	-0.11
standingsId	0.31	0.17	-0.21	0.04	0.06	0.02	0.24	0.50	0.08	0.17	0.59	1.00	0.07	-0.11
driverStandingsId	0.85	0.17	-0.21	0.02	-0.05	-0.03	0.09	1.00	0.08	0.17	0.02	0.07	1.00	-0.11
year	-0.11	-0.07	0.24	-0.76	0.27	0.20	0.10	-0.12	0.02	-0.30	-0.11	-0.11	-0.11	1.00

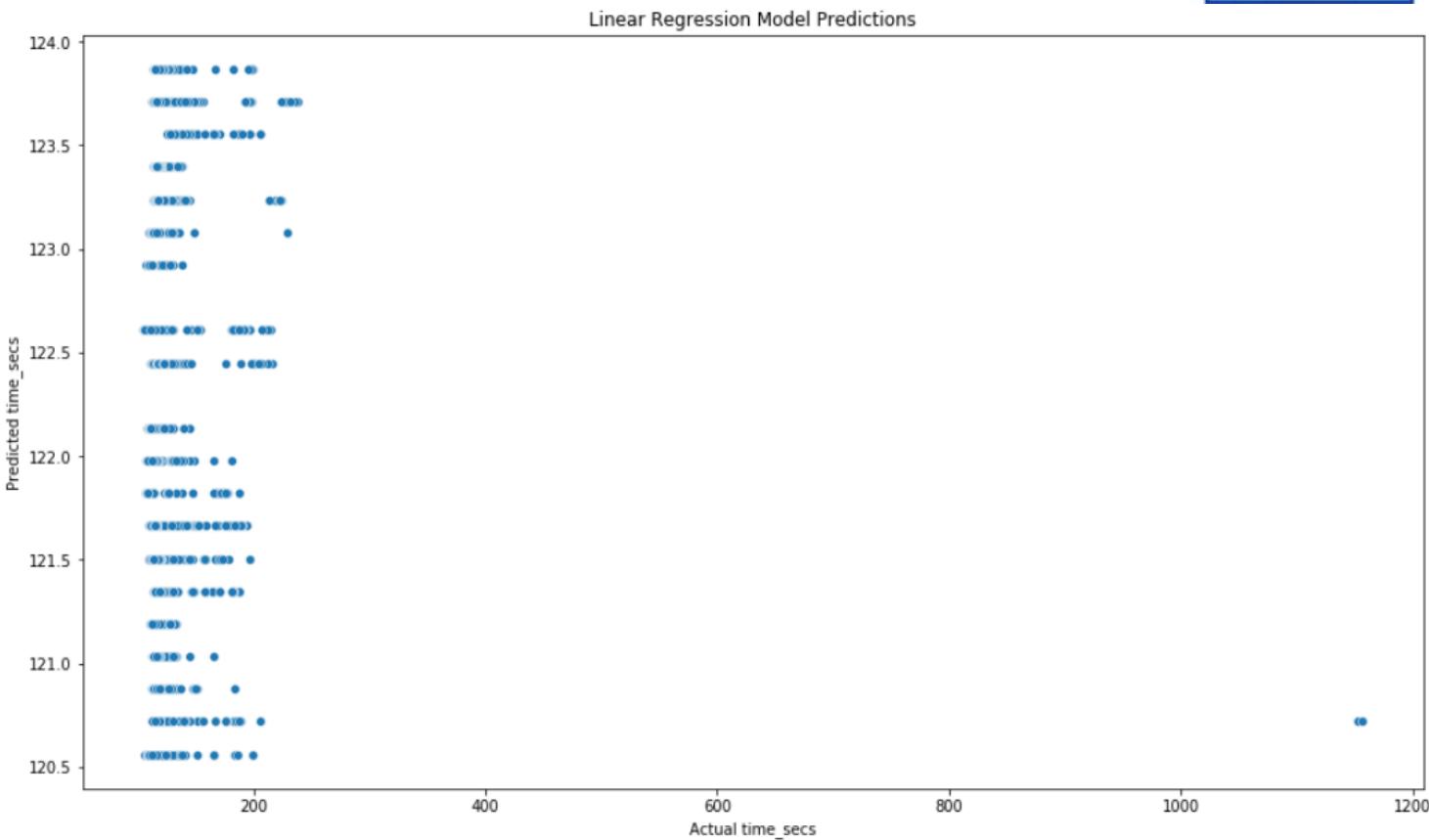
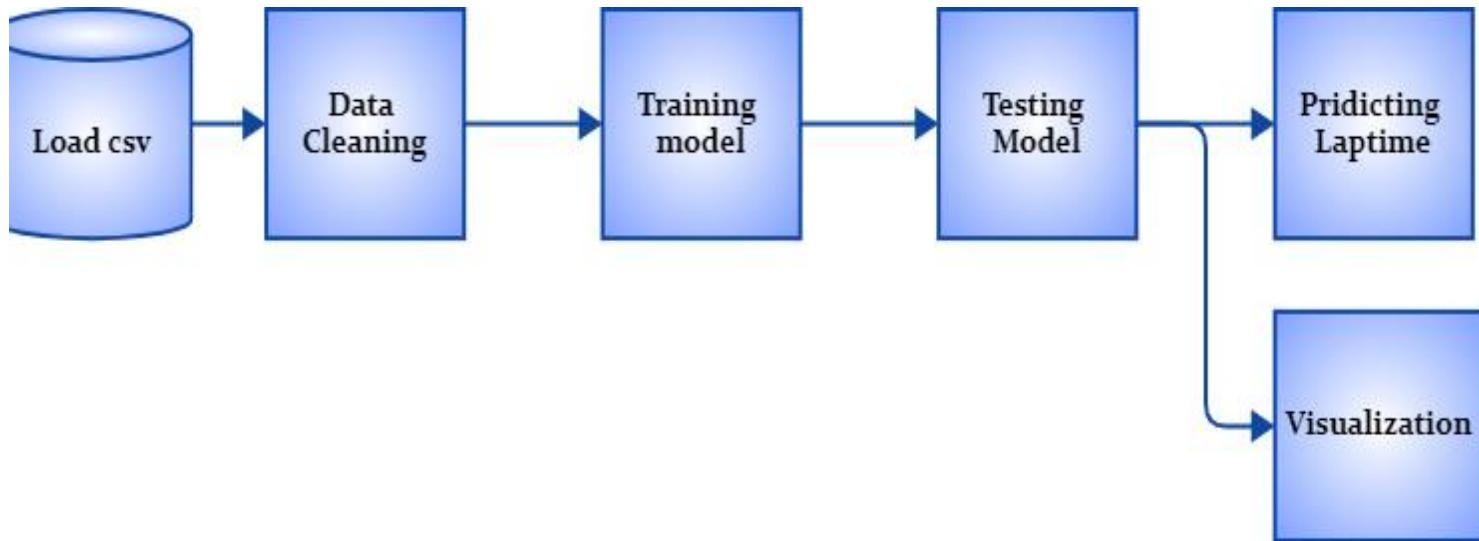
Exploratory Data Analysis

- The bar plot shows a significant number of British drivers in Formula 1, with German drivers being noticeably the least represented of all the countries.
- It implies that the points system is well balanced. It indicates that drivers in similar positions consistently earn similar point totals, reflecting fairness and competitiveness of the championship.



DATAMODEL REVIEW :



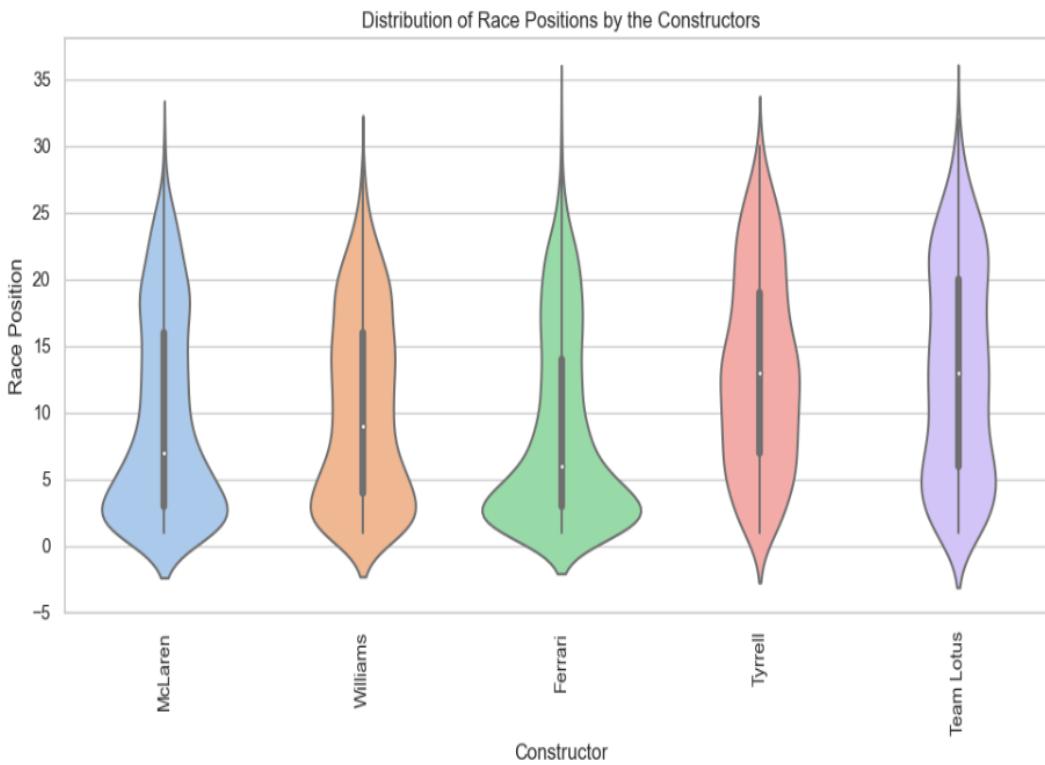


Predicting Using Linear Regression

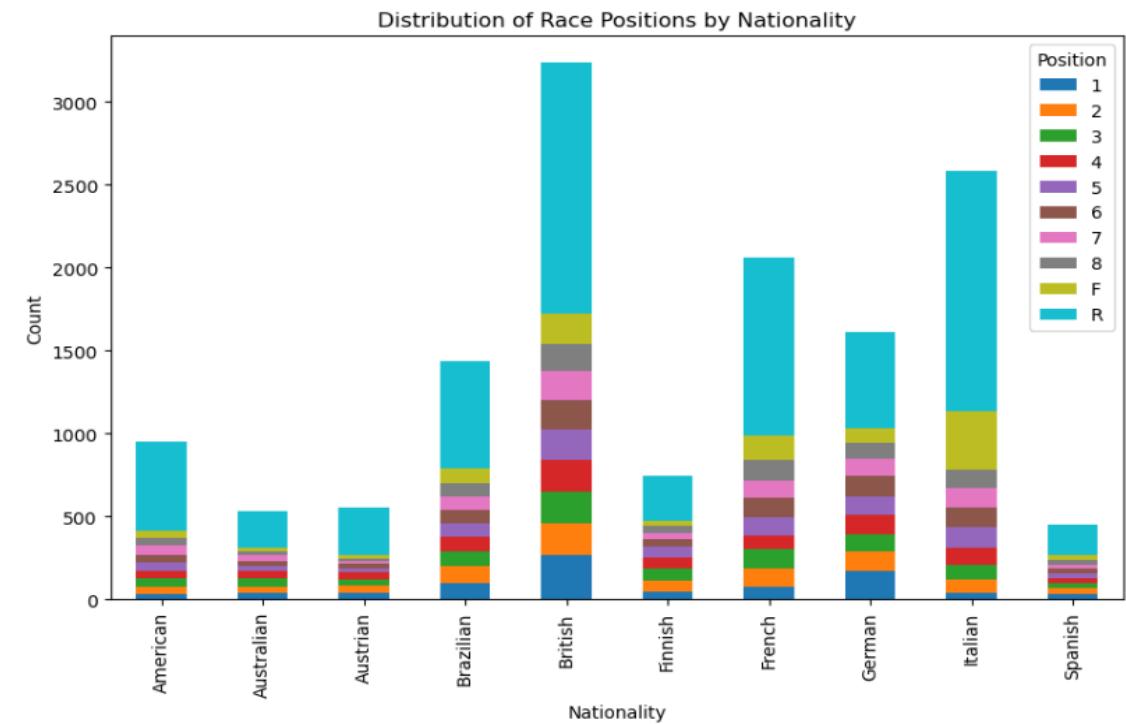
- Load the data
- check for null values and duplicate values
- Reshape the values
- Data is split into 80%-20% for training and testing the model using Linear Regression.
- Once the model is ready, we can predict average Lap_Time for any specific year and CircuitID

Visualizations

Constructors with their respective positions.

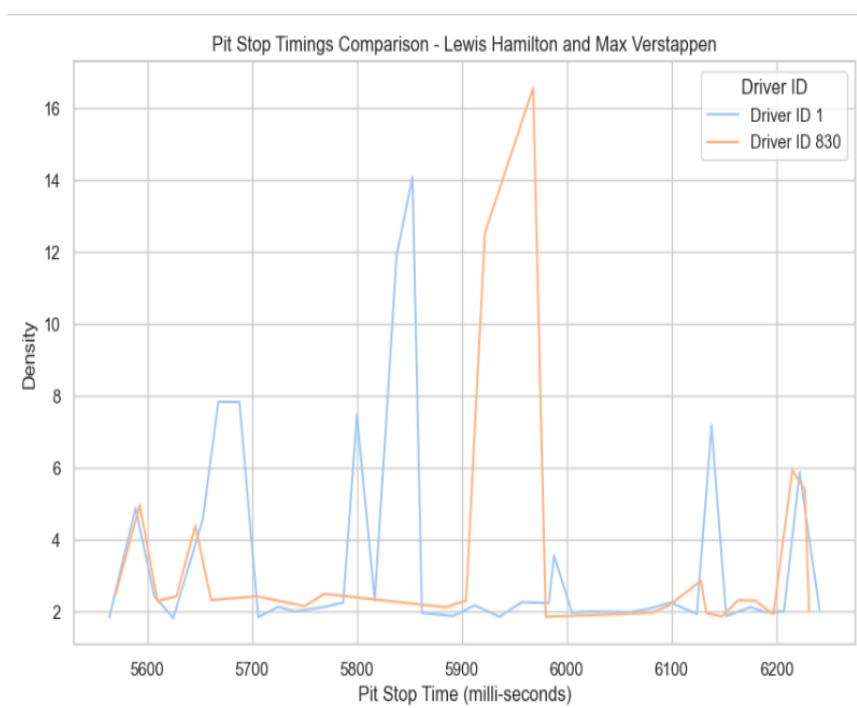


Drivers' positions based on their nations



Cont.

Pit-stop duration: Lewis vs Max

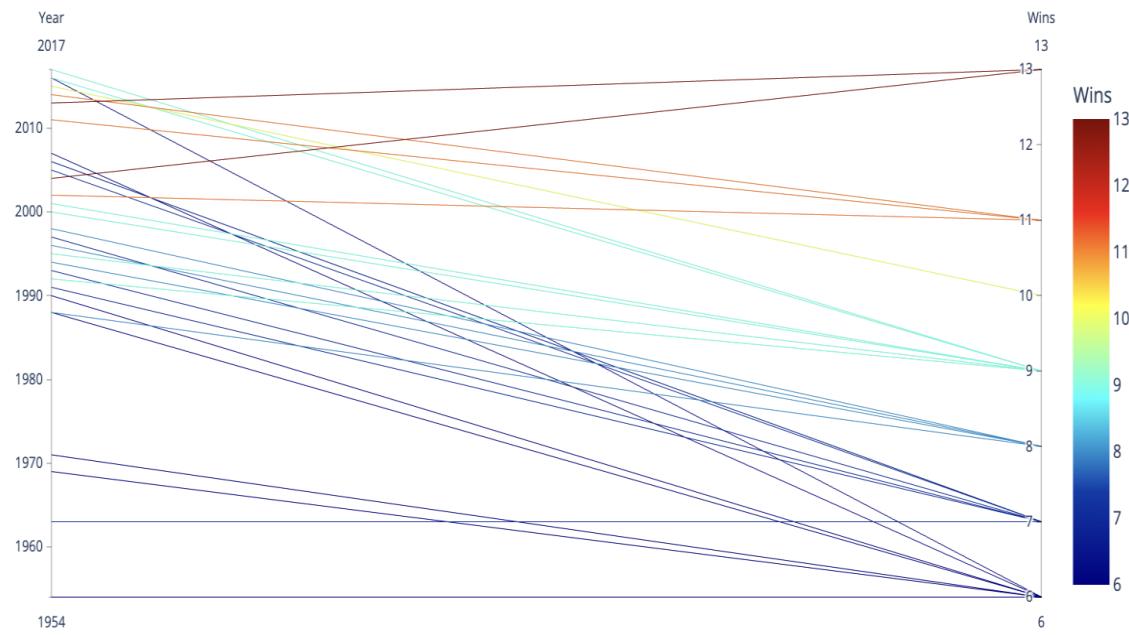


Spread of races across the globe

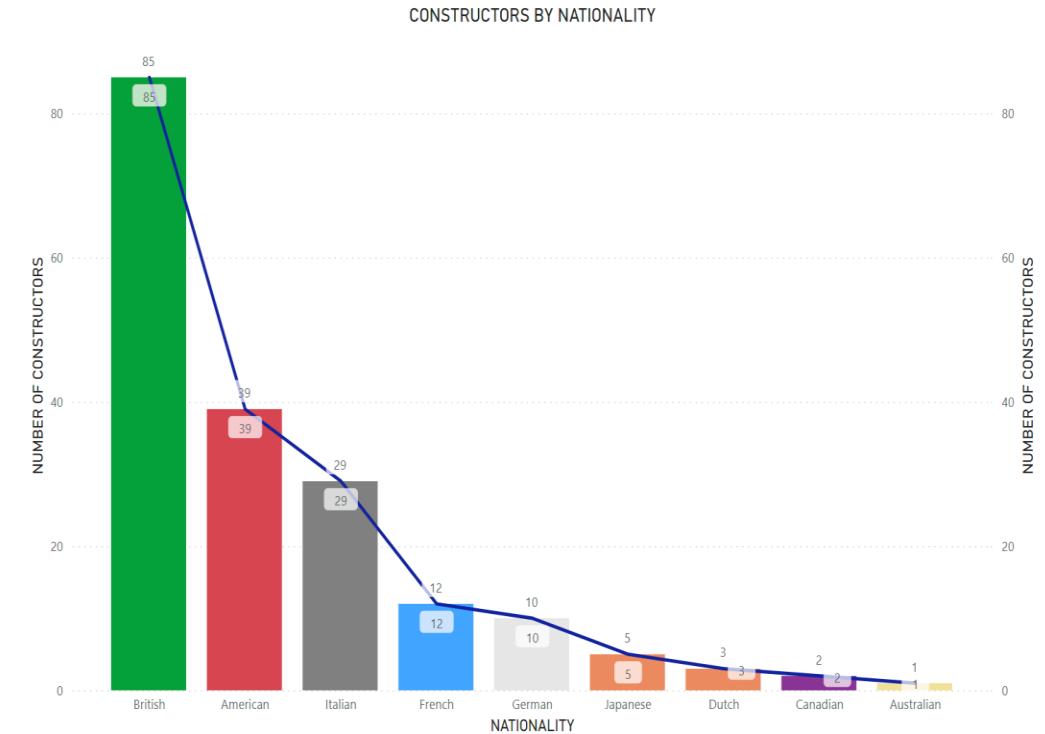
Number of races in different countries from 1950 to 2023



Most wins by a driver in a single season (Parallel Coordinate Plot)



CONSTRUCTORS BY NATIONALITY



DEMO





FORMULA 1 ANALYTICS DASHBOARD

HOME

DRIVER DASHBOARD

TEAM DASHBOARD

FUNNEL REPORT

CONSTRUCTORS BY NATIONALITY

PITSTOP DURATION BY LAP

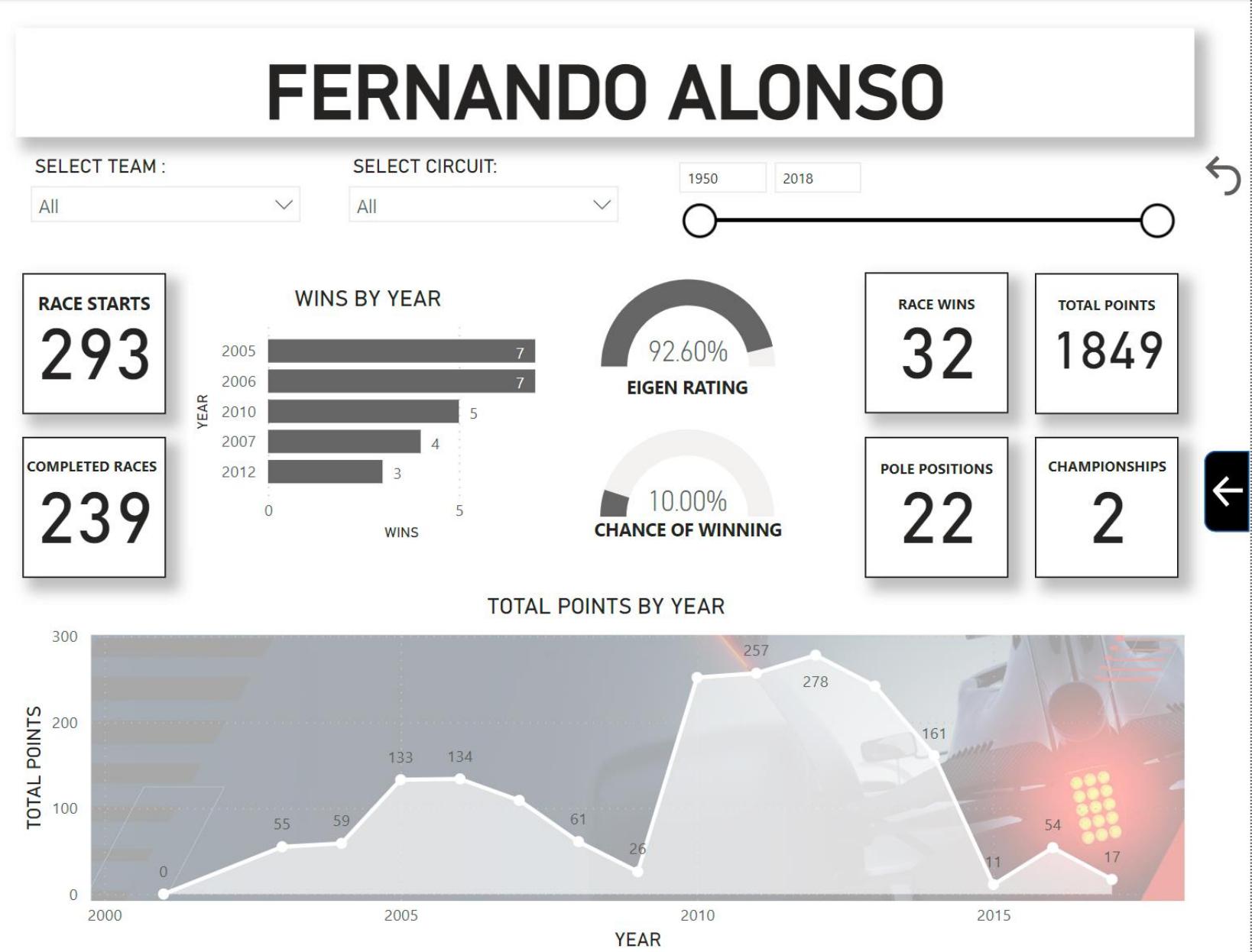
F1

DRIVER DASHBOARD

SELECT DRIVER :
FERNANDO ALONSO



AGE : 42
CAR NUMBER : 14





DRIVER DASHBOARD

SELECT DRIVER :

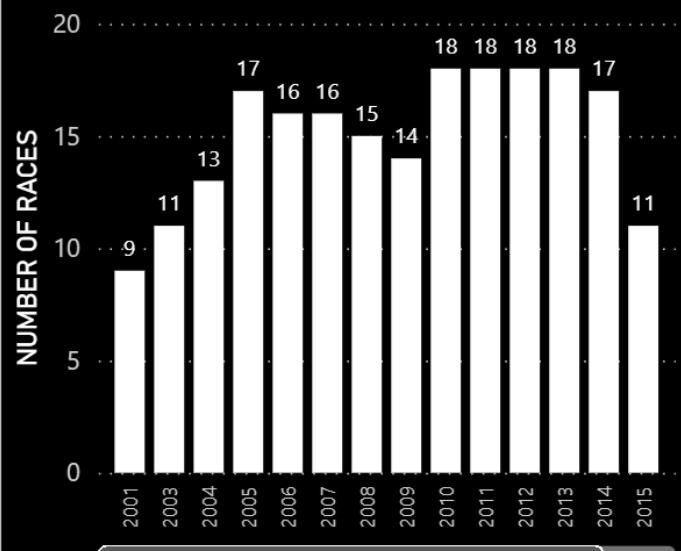
FERNANDO ALONSO



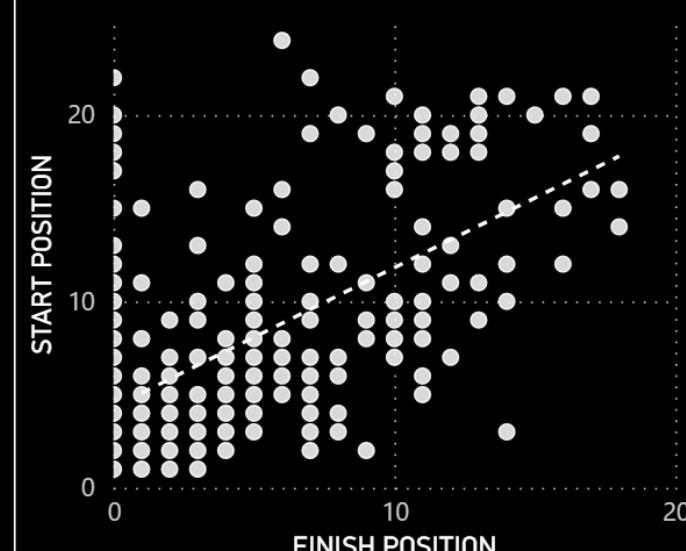
AGE : 42

CAR NUMBER : 14

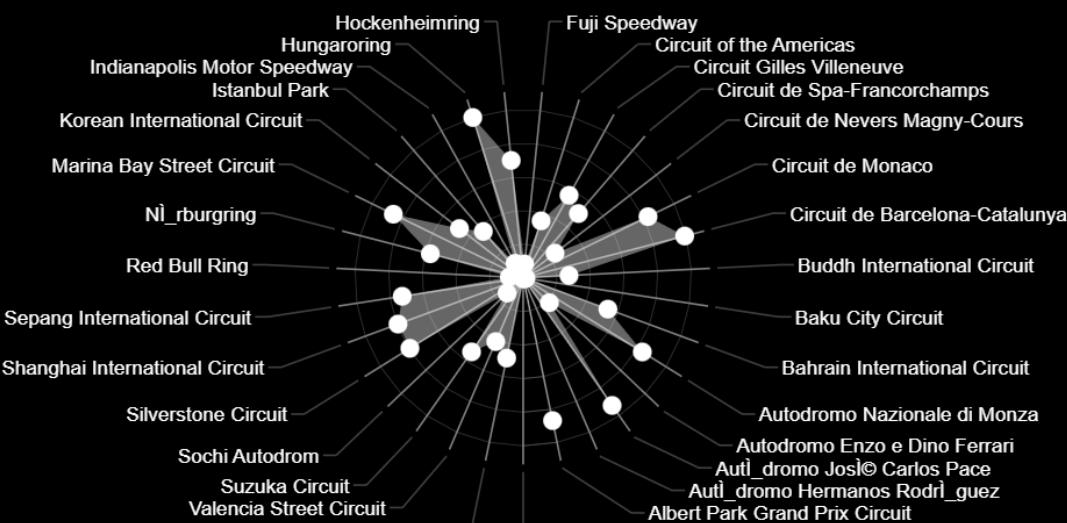
NUMBER OF RACES BY YEAR



START POSITION VS END POSITION



POINTS SCORED BY CIRCUIT





TEAM DASHBOARD

SELECT TEAM:

MCLAREN

McLaren



MCLAREN

SELECT DRIVER:

All

SELECT CIRCUIT:

All

1950

2018



RACE STARTS

1625

RACE WINS

178

POLE POSITIONS

155

COMPLETED RACES

1121

TOTAL POINTS

5314

CHAMPIONSHIPS

8

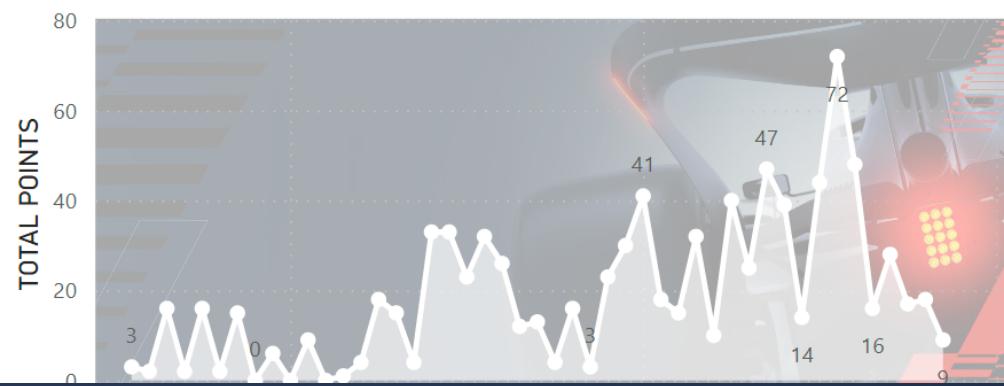
81.00%

EIGEN RATING

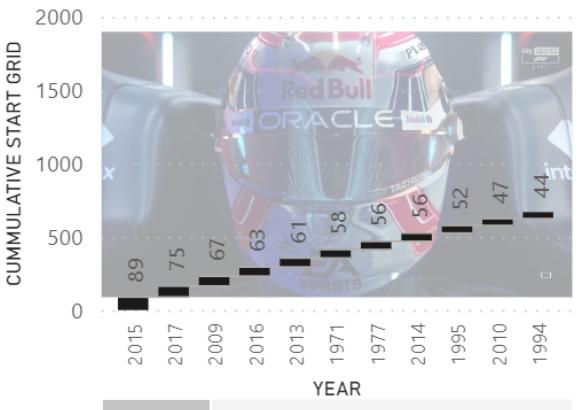
12.00%

CHANCE OF WINNING

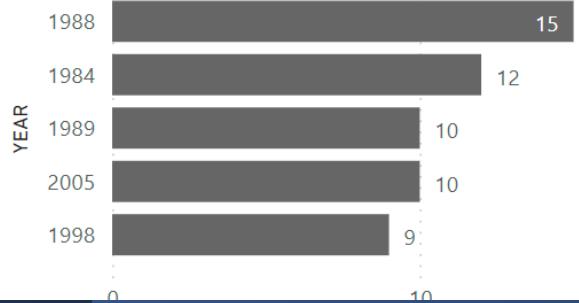
TOTAL POINTS BY YEAR

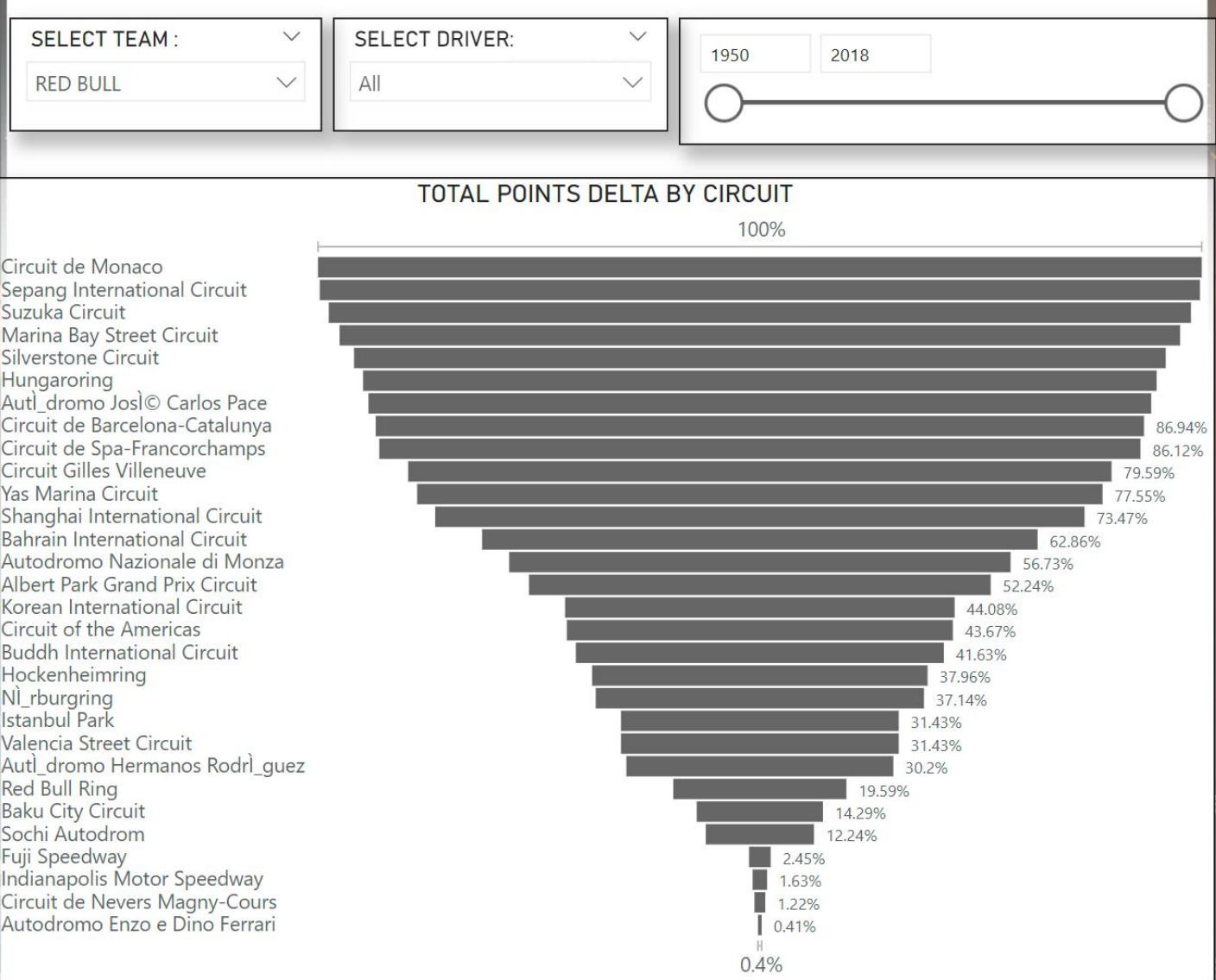


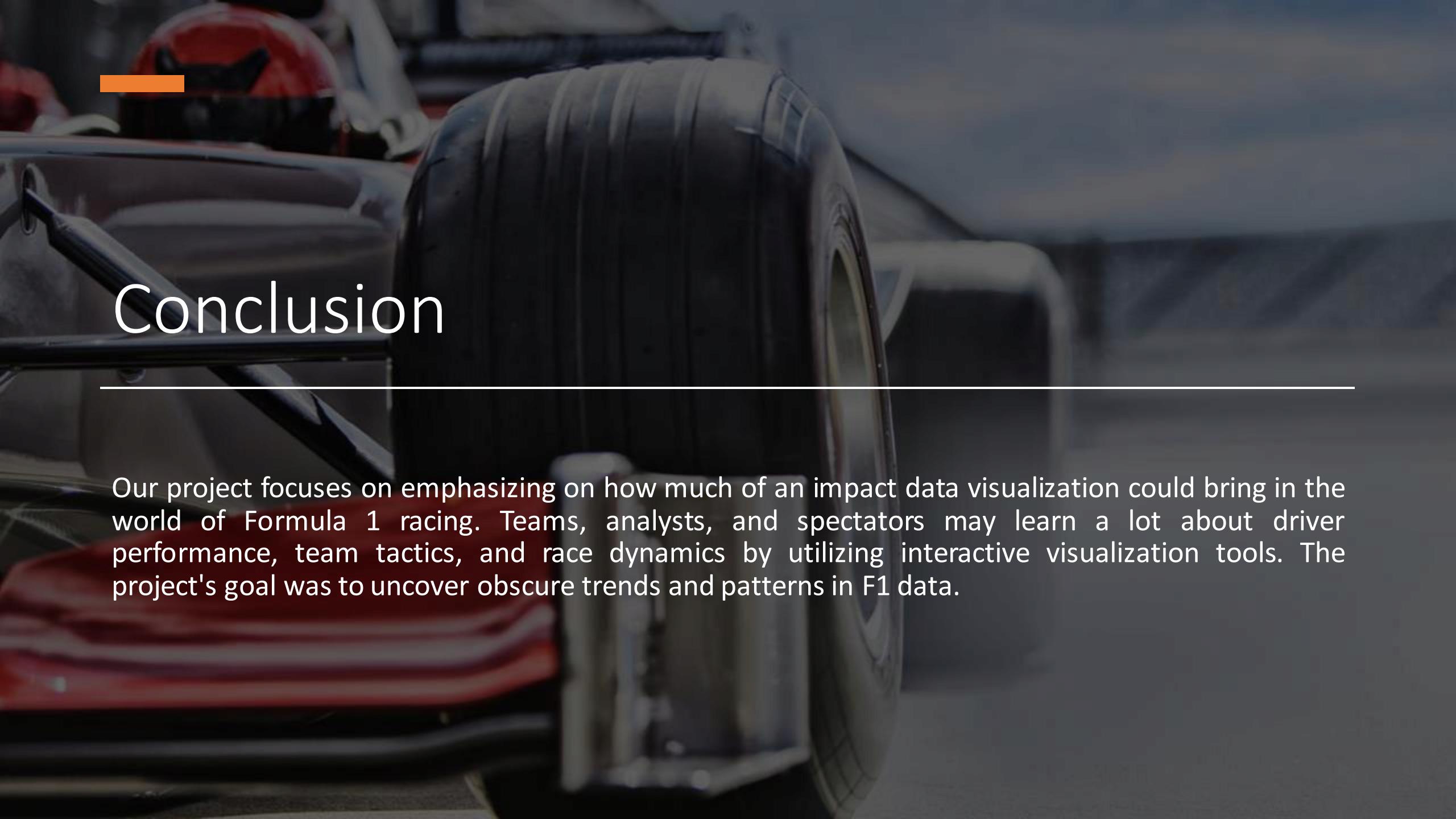
CUMMULATIVE START GRID BY YEAR



WINS BY YEAR







Conclusion

Our project focuses on emphasizing on how much of an impact data visualization could bring in the world of Formula 1 racing. Teams, analysts, and spectators may learn a lot about driver performance, team tactics, and race dynamics by utilizing interactive visualization tools. The project's goal was to uncover obscure trends and patterns in F1 data.



THANK YOU!