

FORMULA 1 ANALYTICS

BIG Data



Ayush Praveen - PES1UG20CS900

Yash Vardhan Soni - PES1UG20CS518

Anirudh Chandrashekhar - PES1UG20CS056

Pracheth Thamankar - PES1UG20CS469

PROBLEM STATEMENT

- Using statistical analysis methods to gain insights on the various factors that affect lap time and driver performance.
- Analyzing the best strategies for various race circuits.



DATASET

Kaggle Dataset - <https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>

- 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
- sprint_results.csv
- status.csv

The dataset consists of all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships from 1950 till the latest 2022 season.

PRE-PROCESSING

Feature Engineering done on the whole dataset to take care of missing values and convert the data to the required format.

```
sprints_df['fastestLapTime'] = pd.to_timedelta(  
    '00:' +sprints_df['fastestLapTime'], errors='coerce')  
sprints_df.fastestLapTime.fillna(pd.Timedelta(0), inplace=True)  
  
sprints_df['time'] = pd.to_timedelta(sprints_df['milliseconds'], unit = 'ms')  
sprints_df.time.fillna(pd.Timedelta(0), inplace=True)
```

	time	milliseconds	fastestLap	fastestLapTime
0 days 00:25:38.426000	1538426.0	14.0	0 days 00:01:30.013000	
0 days 00:25:39.856000	1539856.0	17.0	0 days 00:01:29.937000	
0 days 00:25:45.928000	1545928.0	17.0	0 days 00:01:29.958000	
0 days 00:25:49.704000	1549704.0	16.0	0 days 00:01:30.163000	
0 days 00:26:02.537000	1562537.0	16.0	0 days 00:01:30.566000	

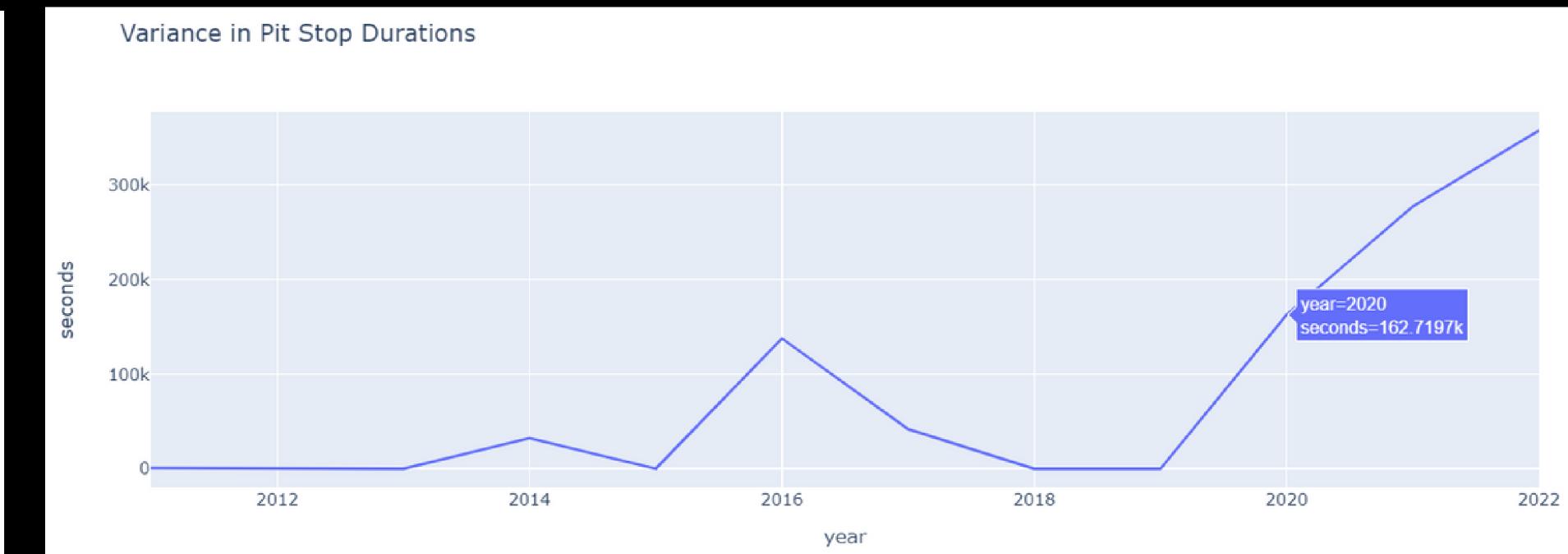
```
drivers_df['name'] = drivers_df['forename'] + ' ' + drivers_df['surname']  
drivers_df.drop(labels=['forename', 'surname'], axis=1, inplace=True)  
drivers_df['dob'] = pd.to_datetime(drivers_df['dob'])
```

	driverId	driverRef	dob	nationality	name
0	1	hamilton	1985-01-07	British	Lewis Hamilton
1	2	heidfeld	1977-05-10	German	Nick Heidfeld
2	3	rosberg	1985-06-27	German	Nico Rosberg
3	4	alonso	1981-07-29	Spanish	Fernando Alonso
4	5	kovalainen	1981-10-19	Finnish	Heikki Kovalainen

PIT STOP ANALYSIS

How pit stop durations have varied over time?

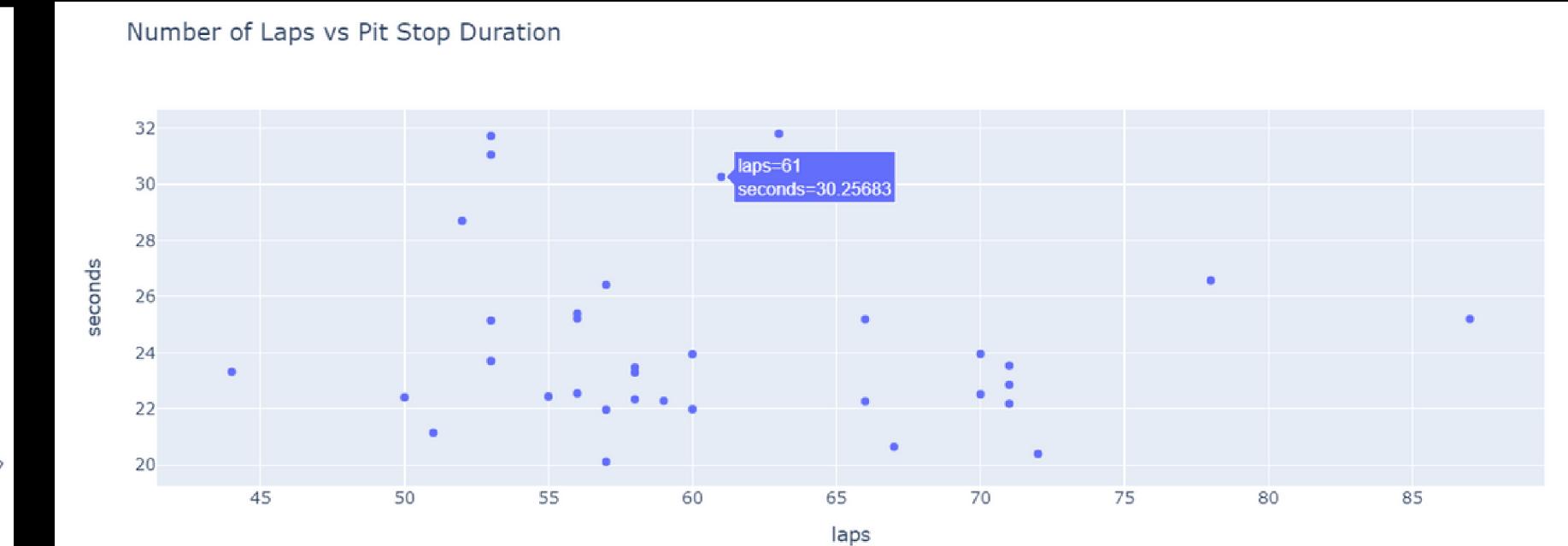
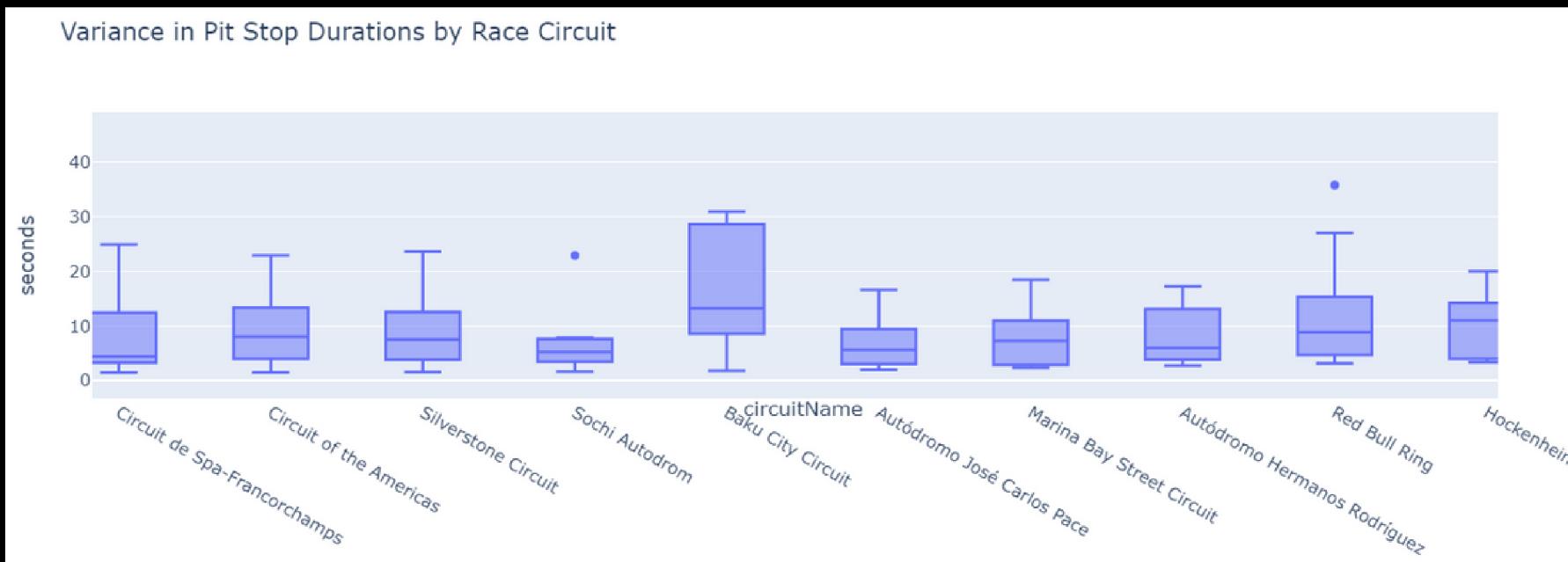
1. Pit stop durations saw a significant increase from 2012 to 2014, more specifically from 2013 to 2014 with the beginning of the Turbo-Hybrid era with new regulations being introduced.
2. Pit stop durations have remained fairly stable from 2014 onwards.
3. Majority of the durations are centered around 20-30s.
4. Variance drastically increases from 2019.



PIT STOP ANALYSIS

Relationship between Pit Stop duration and race circuit.

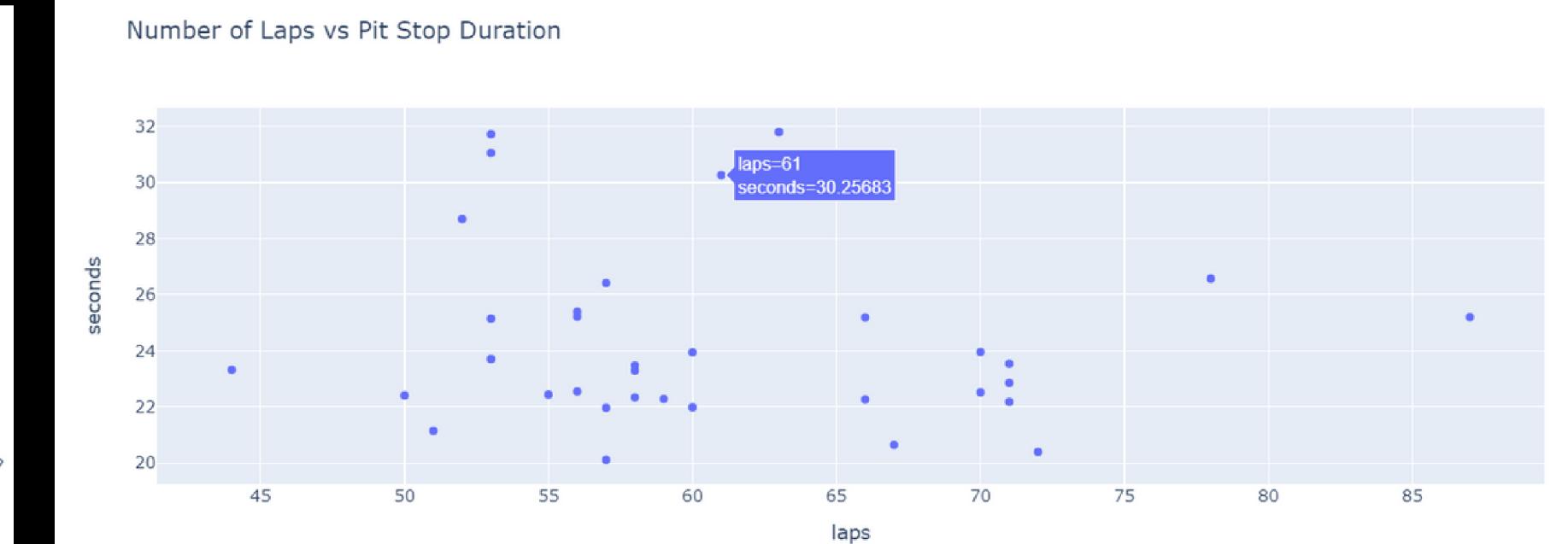
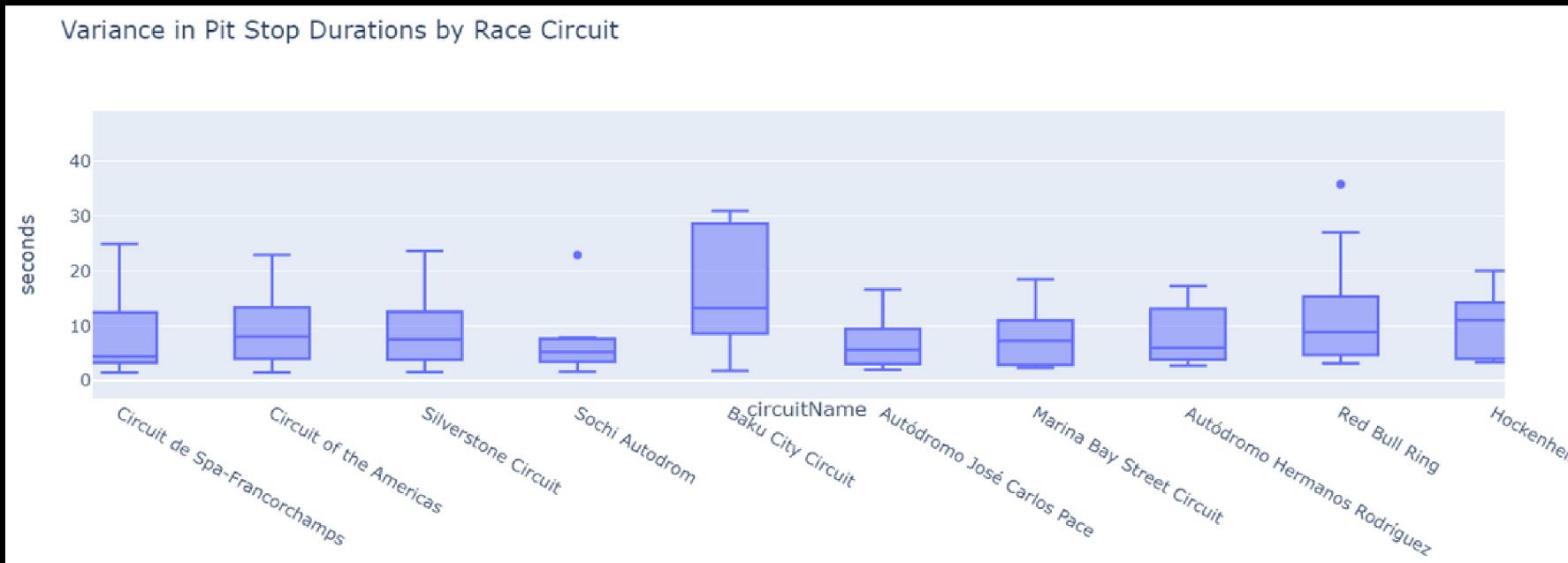
1. Race circuits do have an impact on pit stop durations.
2. Number of laps (and hence length of the circuit) doesn't seem to have any correlation with pit stop durations.
3. Some circuits have larger variances than others, but overall the variances appear to be fairly consistent.



PIT STOP ANALYSIS

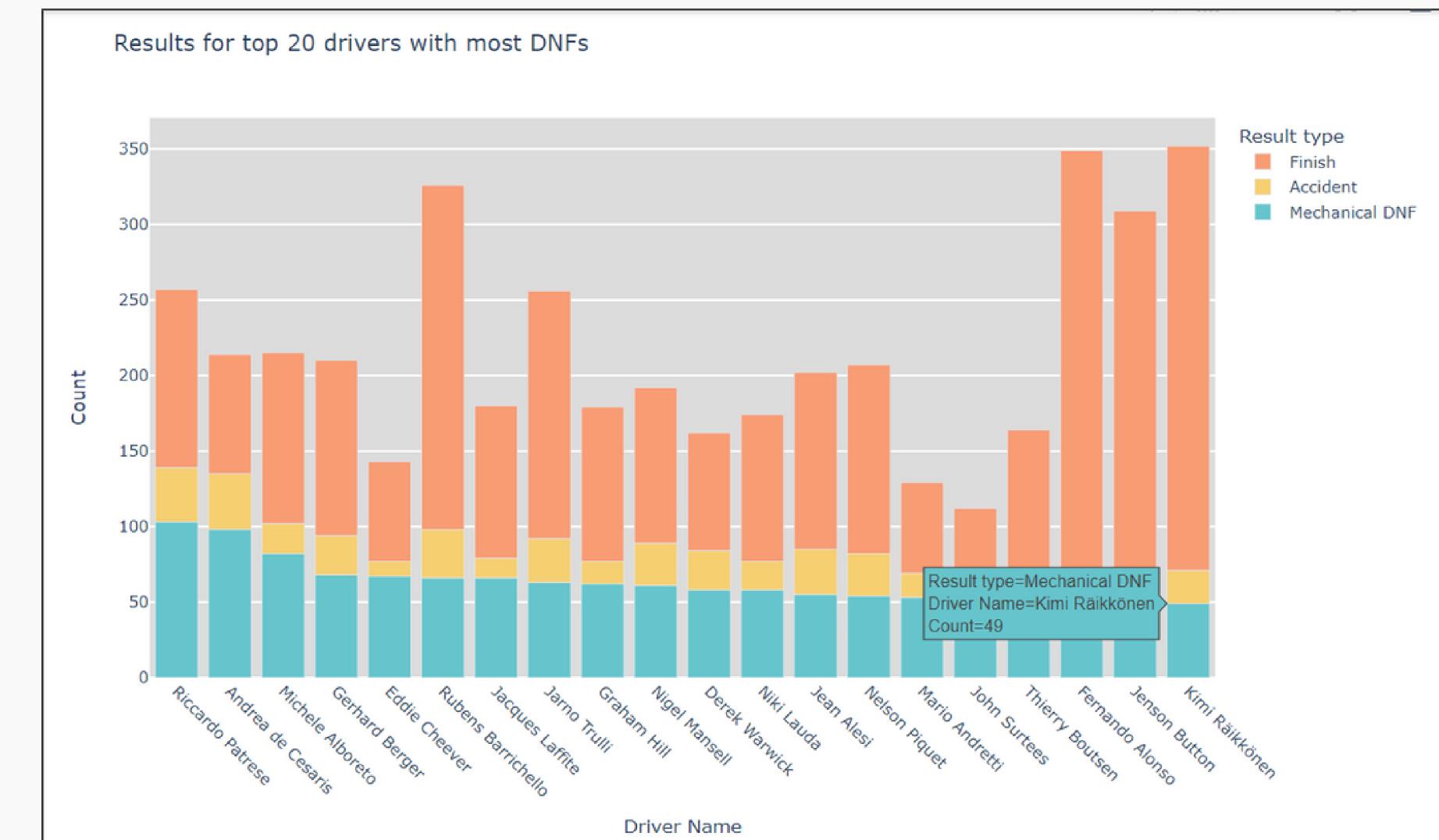
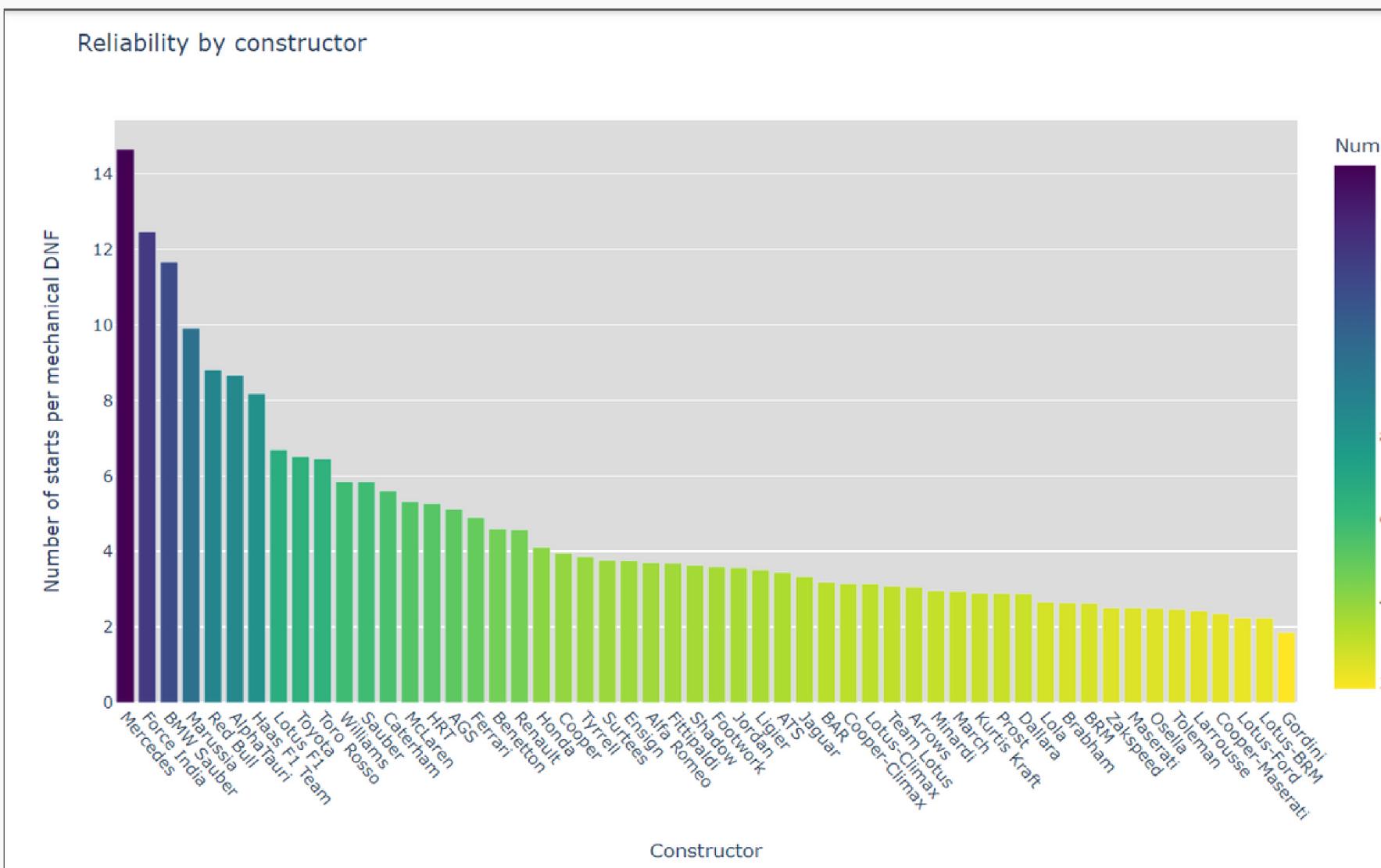
Relationship between Pit Stop count and race circuit.

1. Race circuits have a significant impact on pit stop count. We see that Mugello has seen by far the highest number of pit stops on average.
2. Number of laps (and hence, length of circuit) do not seem to have any correlation with pit stop count.
3. Variance also looks to be dependent on race circuit. This means that circuits with higher variance do not have a definite pit stop strategy that every team counts on.



Mechanical Failure Analysis

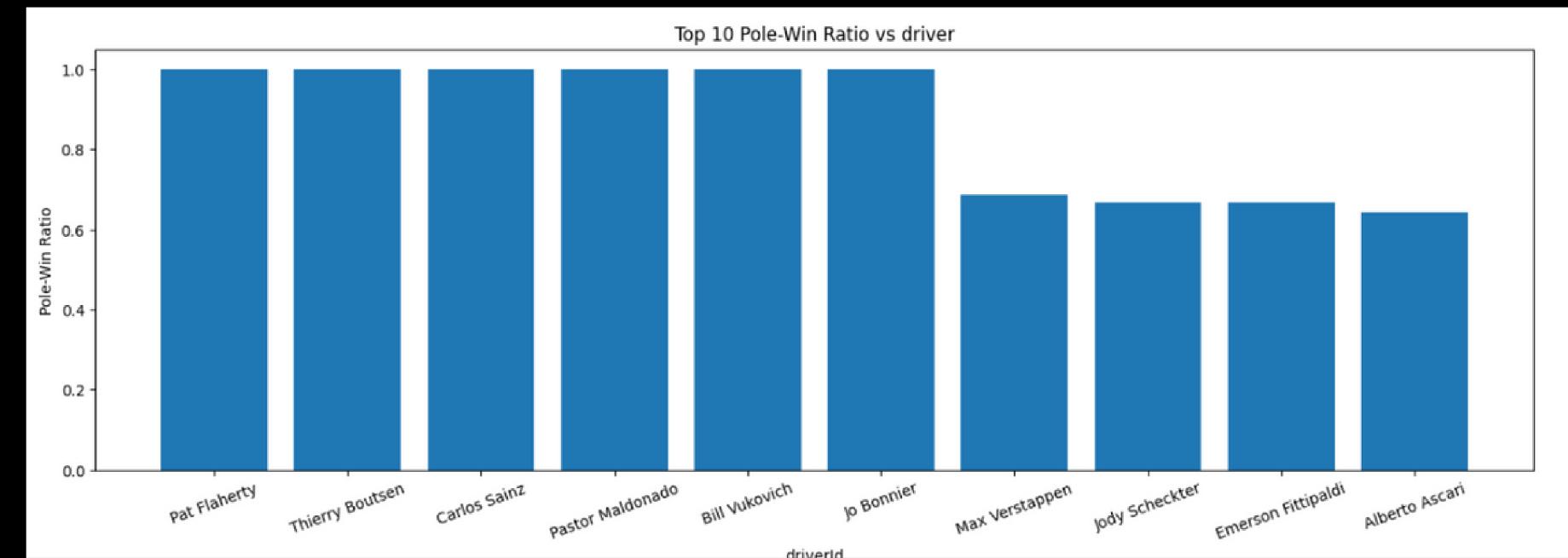
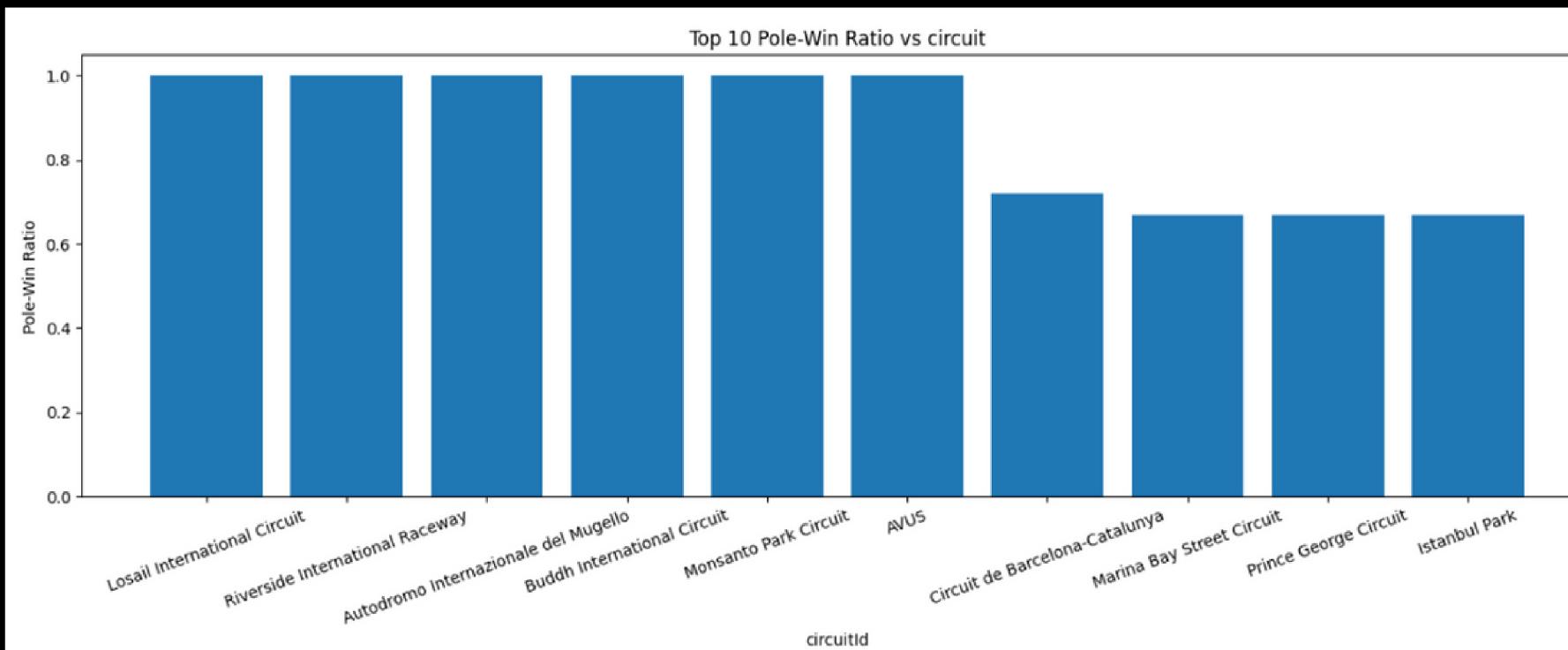
- Reliability by constructor is plots the number of MDNFs* of a constructor by the number of race starts.
 - Top 20 drivers with the most DNFs looks at drivers with the highest number of MDNFs and other accidents which did not allow him to complete the race.



*MDNF - Mechanical Did Not Finish

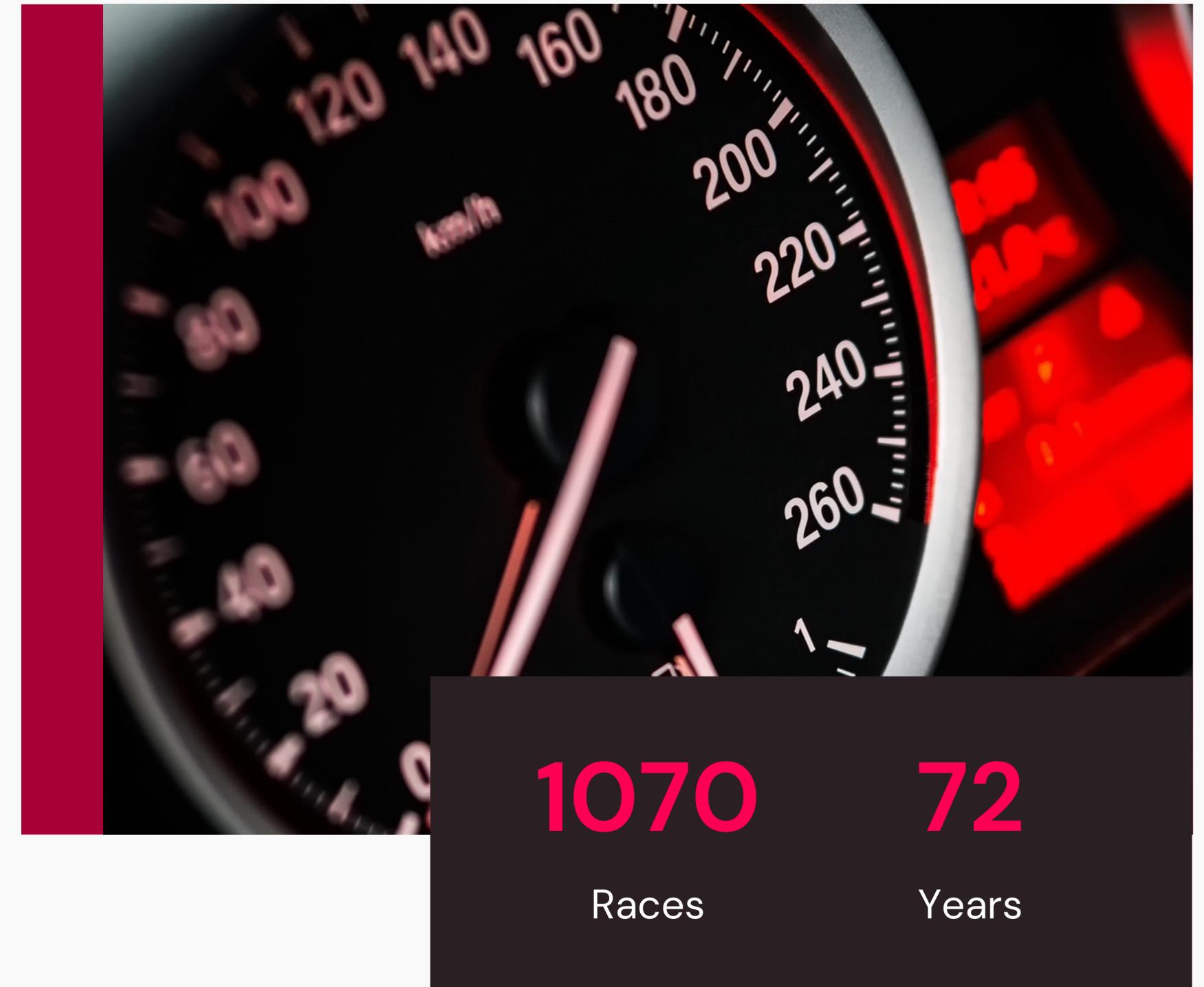
POLE TO WIN CONVERSION RATES

- Pole to win conversion rate is the percentage of drivers who are successful in converting pole position into a win.
- The first plot shows the circuits which favour drivers who start on pole.
- The second plot shows the drivers who are most successful in converting pole position into a win.



FUTURE SCOPE

- Finding the best pit stop windows for different race circuits.
- Finding the Best drivers of all time.
- Finding the best constructors of all time.
- Predicting the best race strategy for different circuits.



Thank
you!