Data Analysis Summary

1. Tool & Framework Specification:

Programming Language: Python

Libraries & Frameworks Used:

1. **Data Handling & Processing:**
    - Pandas – for loading, processing, and aggregating race results data
    - NumPy – for numerical operations
2. **Data Visualization:**
    - matplotlib – for plotting Graphs.
    - seaborn – for enhanced statistical plotting.
3. **Machine Learning:**
    - scikit-learn
    - Models used: Random Forest Regressor (for predicting race outcome)

      Markov chains (To predict the future team of a driver)

2. Feature Engineering:

Newly Created Features:

Driver Consistency: Average finishing position, Average qualifying position.

Team Strength: Average Constructor Points, Win Reliability Score.

Track Complexity: Average positions gained, Position variability, Driver position variability.

3. Model Training and Evaluation (For Predicting Race Outcome):

- **Model used:** Random Forest Regressor (from scikit-learn)
- **Train-Test split ratio:** 80-20
- **Preprocessing Steps:** Label Encoding for Categorical features- Driver ID and Constructor ID

- **Hyperparameter Tuning:**

  Methodology: Grid Search Cross Validation (5-fold CV).

  Parameters: n_estimaters (200), max_depth (20), min_samples_split (10).

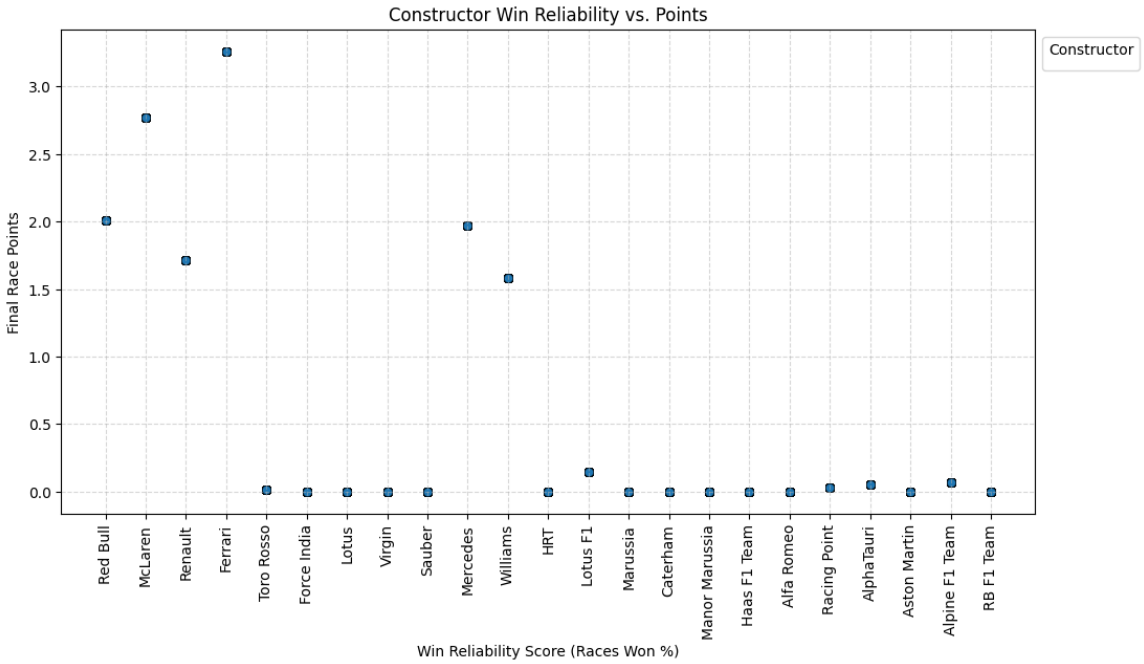
- **Evaluation Results:**

  Mean Absolute Error (MAE): 1.6914428951393188
  Root Mean Squared Error (RMSE): 2.577681315708374
  R^2 score: 0.888097119398918

- **Cross Validation Results:**
  Mean Cross-Validation MAE: 2.233675309221616
  Mean Cross-Validation RMSE: 3.2477314504037316
  Mean Cross-Validation R^2: 0.7984372978015104

4. Key Insights and Visualizations:

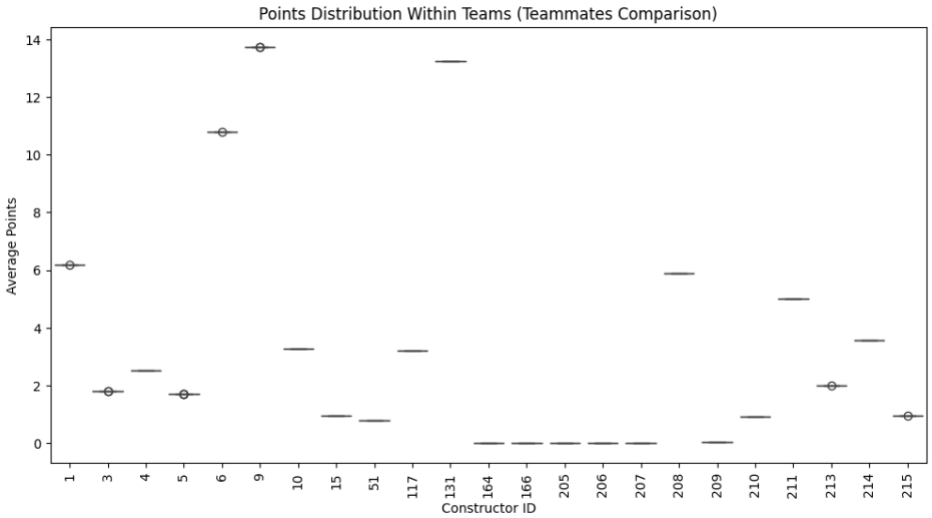
1. Constructor win reliability (Scatter plot):



Inference:

- Ferrari and MC Laren are the most reliable constructors.

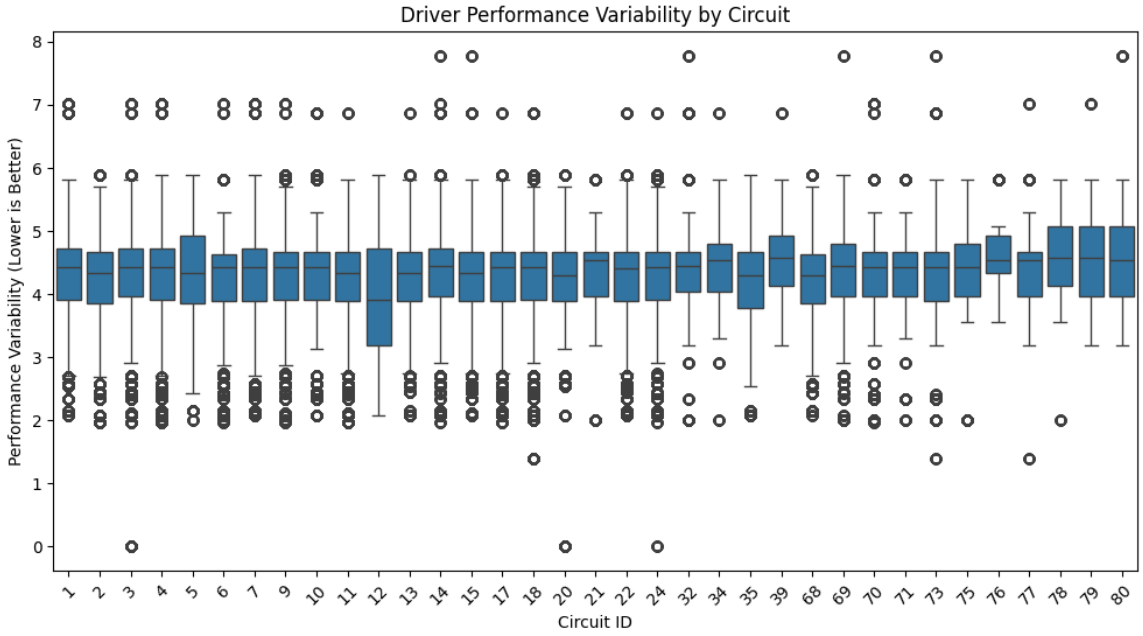
2. Team Dynamics (Box plot):



Inference:

- Most teams do not have a single dominant member.
- There is no single member dominance in most of the teams.

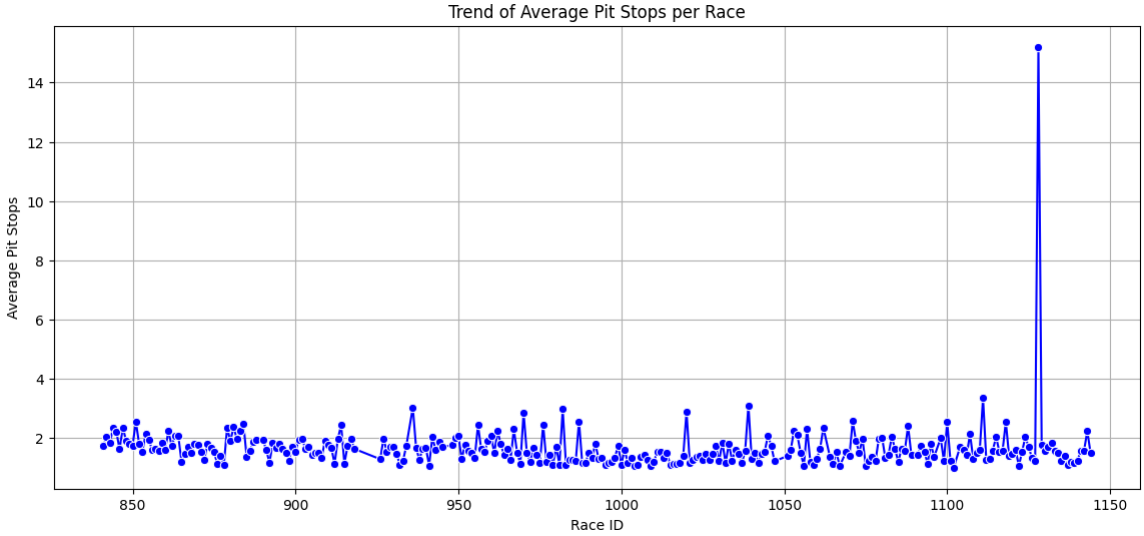
3. Driver Adaptability to Circuits:

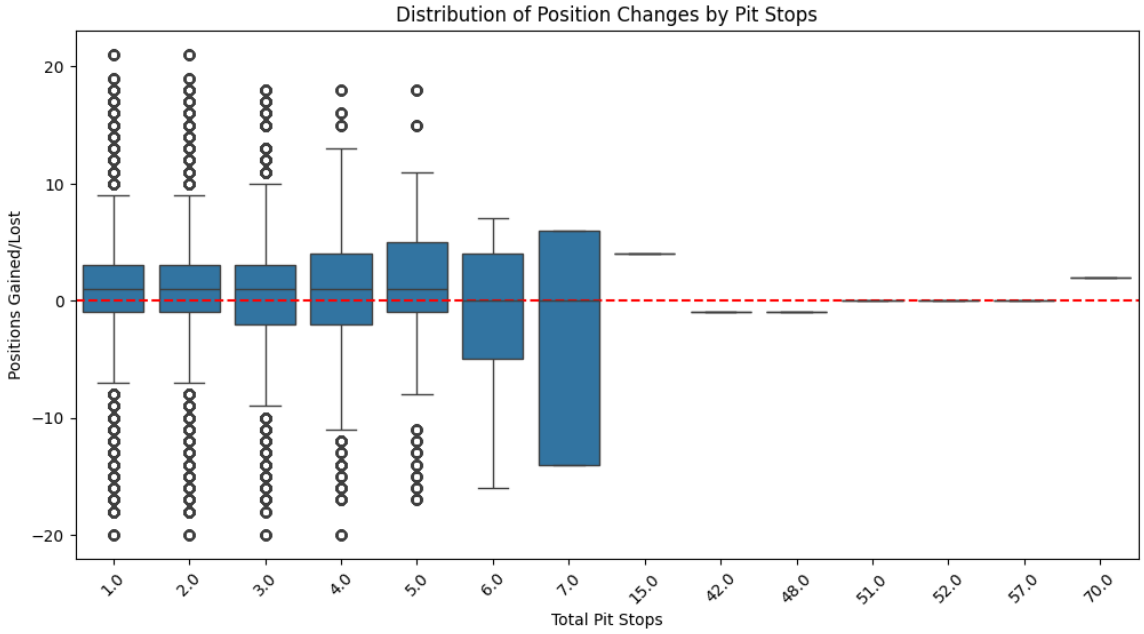


Inference:

- The adaptability score ranges from 4-5 for all circuits.

4. Pit Stop Strategies:

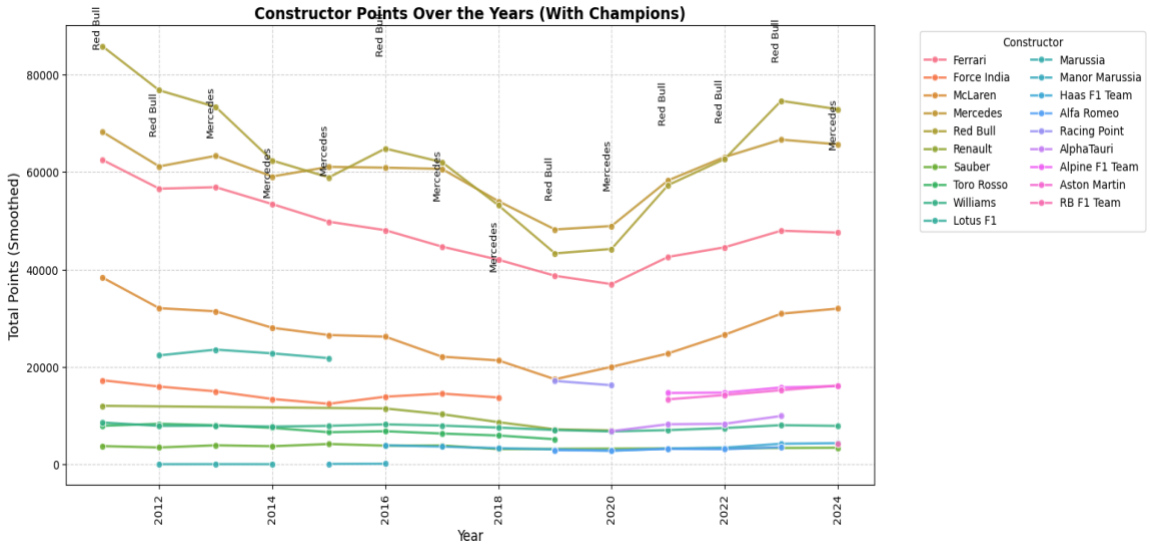




Inference:

- The average number of Pit stops for per race is approximately 2.
- The average number of Pit stops for per team is approximately 2.
- As the number of Pit stops increases the number of positions lost also increases.

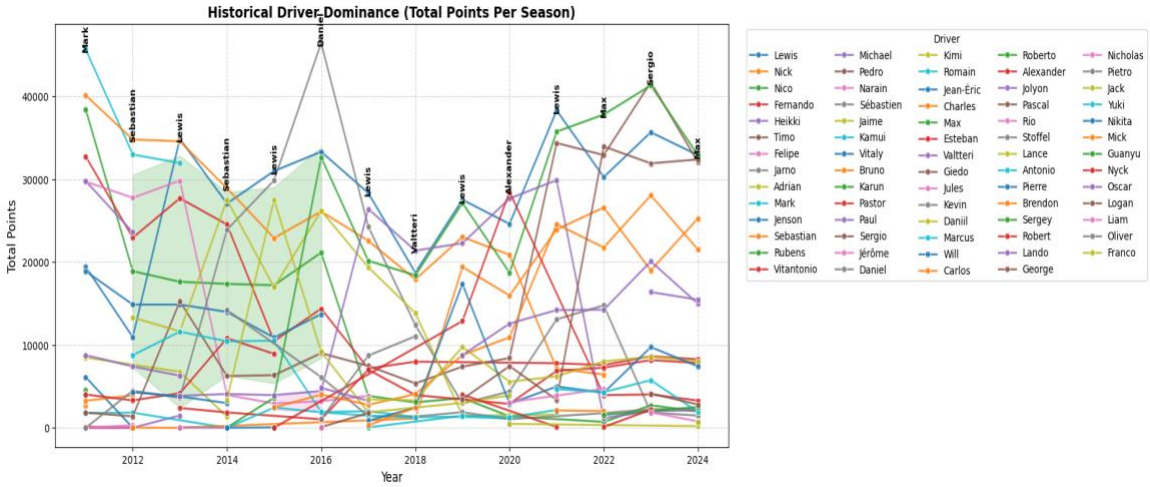
5. Constructor trends over the years:



Inference:

- Over the years Red Bull and Mercedes have remained the top teams.
- Ferrari has consistently ranked as the third-best team.
- Toro Rosso has been the most consistent team, maintaining a steady total points each year.

6. Driver Trends over the years:



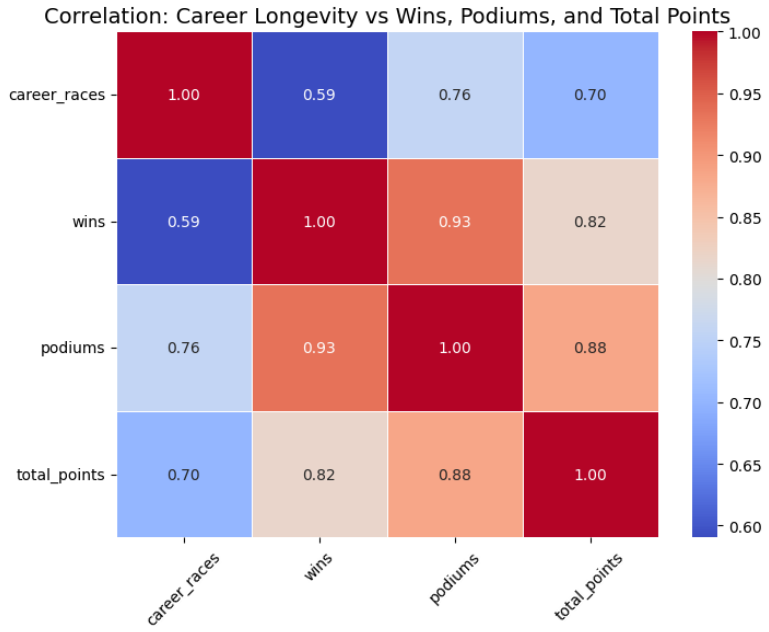
Inference:

- Lewis is the top performer.
- Daniel scored the highest point in 2016.

PROBLEM STATEMENT SOLUTIONS:

1. **Driver & Constructor Performance**
   - Top 3 Dominant Drivers (by Win Ratio)
     Lee Wallard - Win Ratio: 50.00%, Wins: 1, Podiums: 1
     Juan Fangio - Win Ratio: 41.38%, Wins: 24, Podiums: 35
     Bill Vukovich - Win Ratio: 40.00%, Wins: 2, Podiums: 2

   - Top 3 Dominant Constructors (by Win Ratio)
     Brawn - Win Ratio: 23.53%, Wins: 8, Podiums: 15
     Matra-Ford - Win Ratio: 22.50%, Wins: 9, Podiums: 15
     Mercedes - Win Ratio: 19.79%, Wins: 129, Podiums: 298

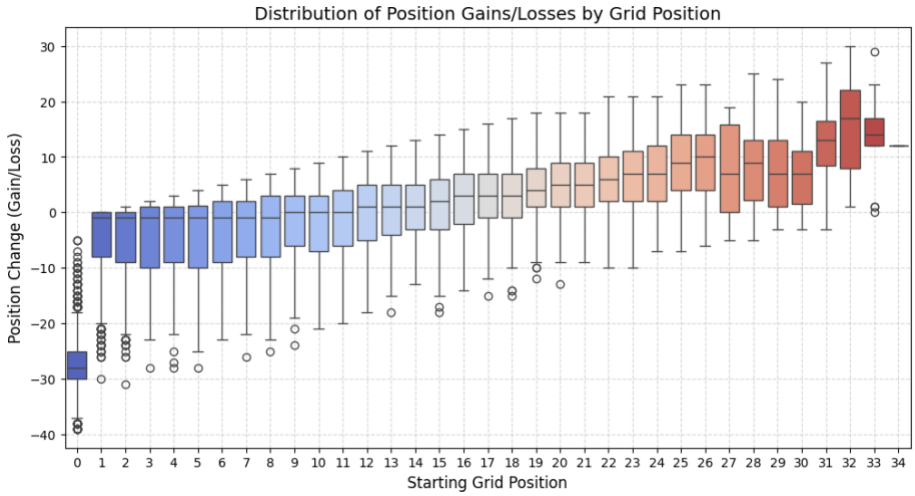


Inference: Career longevity mostly depends on the points scored by the driver.

2. Qualifying vs. Race Performance

Top 3 Drivers Who excel in making up positions:

driver	grid	avg_position_gain	total_races
George Amick	25	23.0	1
Bud Tingelstad	28	19.0	1
Ernst Klodwig	32	17.0	2

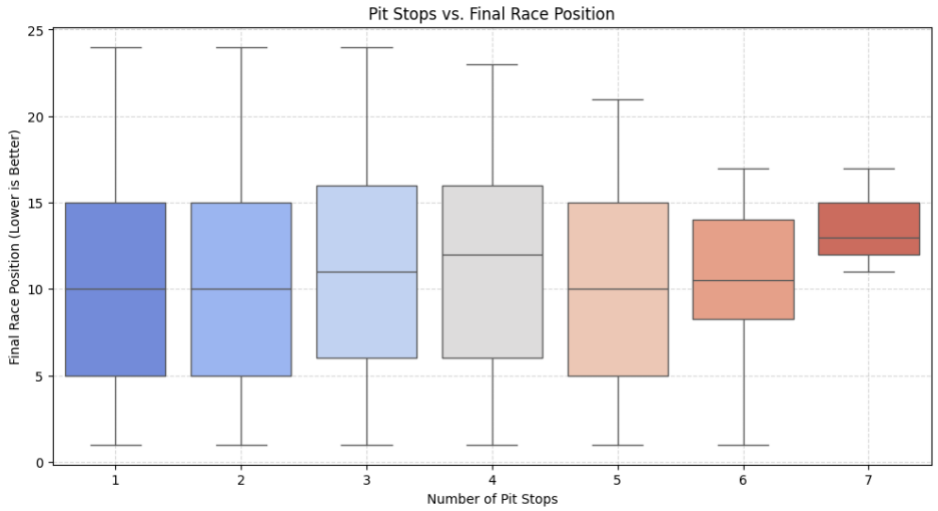


Inference: It is evident that the position gain is high if the starting grid position is 20 or above.

3. Pit Stop Strategies

Top 10 Fastest Pit Stop Drivers:

driver	avg_pit_time	total_pit_stops
Jack Doohan	22.080000	1
Michael Schumacher	22.541644	90
Nick Heidfeld	22.933320	25



Inference: As the number of Pit stops increases the finishing position is lower.

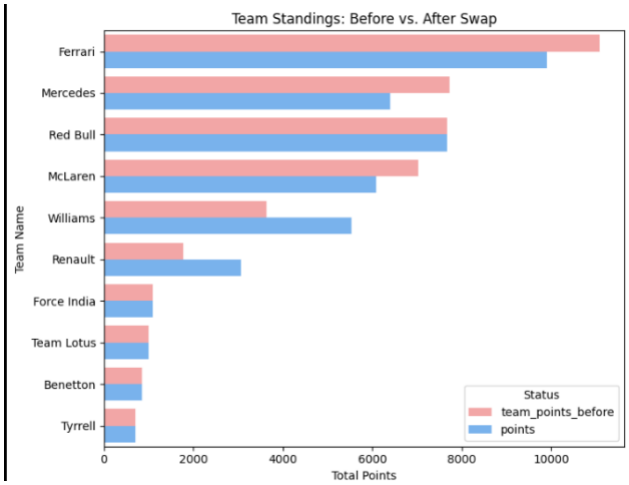
4. **Head-to-Head Driver Analysis:**

Most Competitive F1 Rivalries (Head-to-Head Win Ratios):

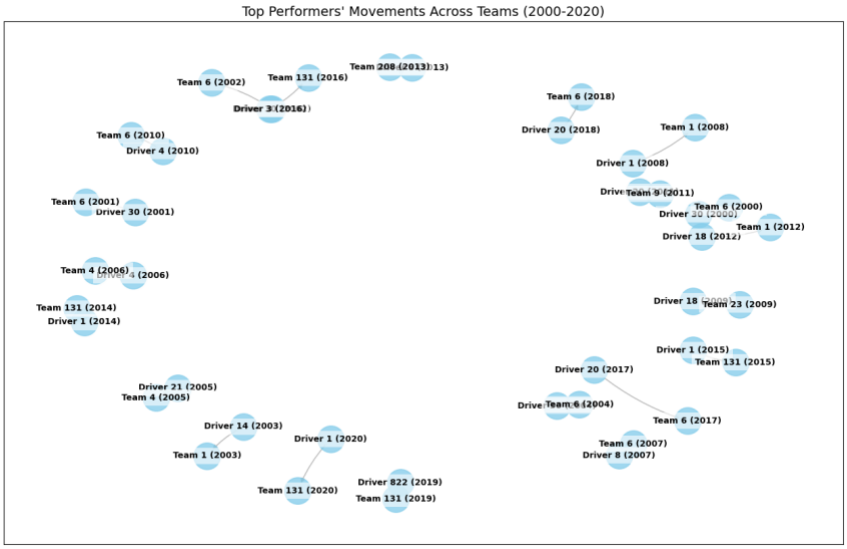
driver_pair	races_competed	driver_A_win_ratio
Felipe Massa vs. Daniel Ricciardo	128	0.5
David Coulthard vs. Mika Häkkinen	122	0.5
Jarno Trulli vs. Mark Webber	176	0.5

5. **Hypothetical Driver Swaps:**

Showing the impact on team standings after swapping the following drivers:
Valtteri with Fernando

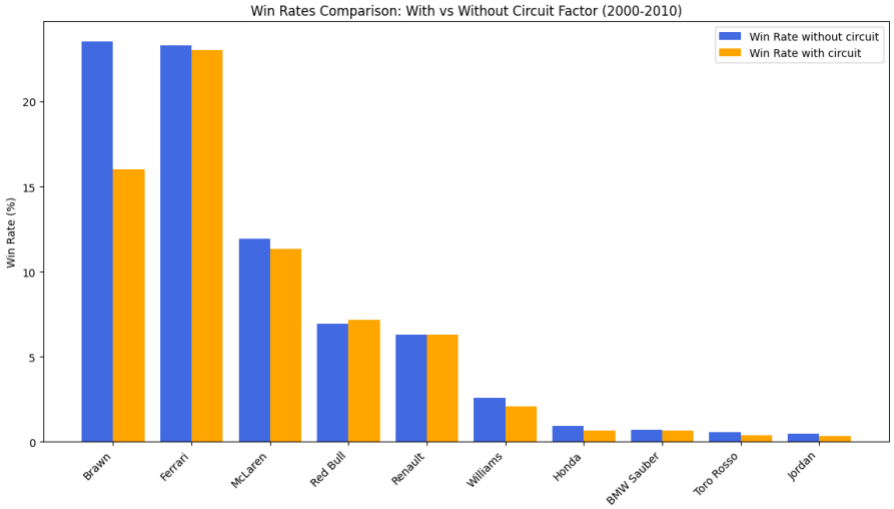


6. **Driver Movements & Team Networks:**



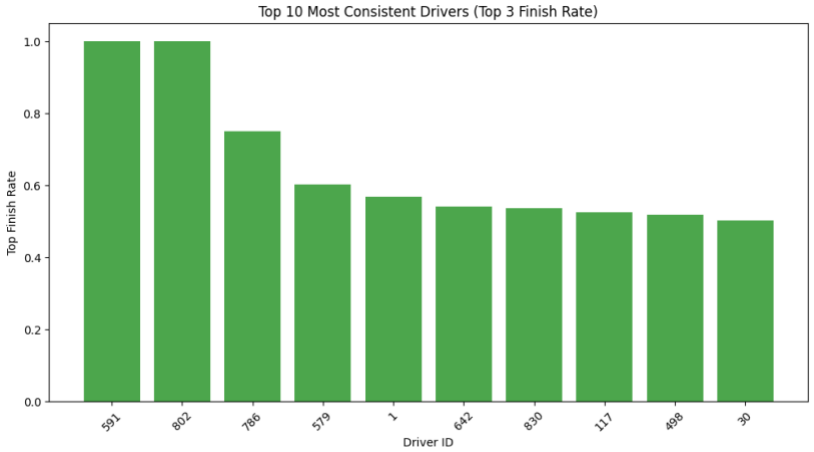
Inference: Team 6 has faced the most number of driver transitions.

7. Team Performance Comparison:

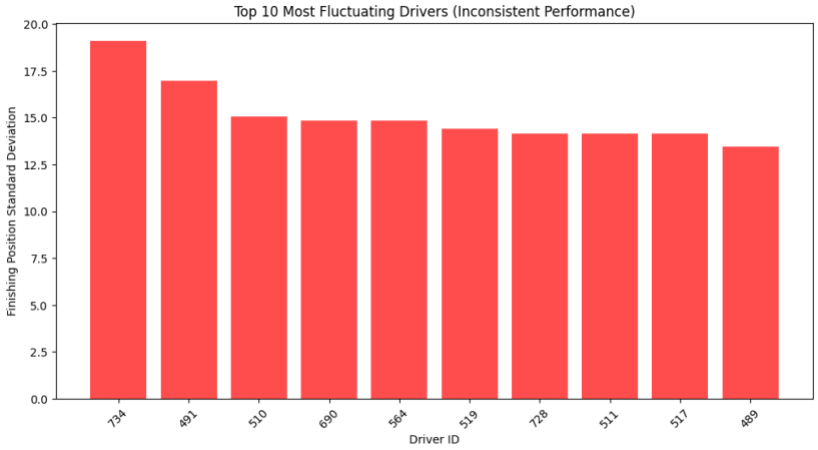


Inference: Ferrari's win rate is similar in both the cases
Brawn's win rate is the most affected by circuit Factor.

8. Driver Consistency in Race Performance:

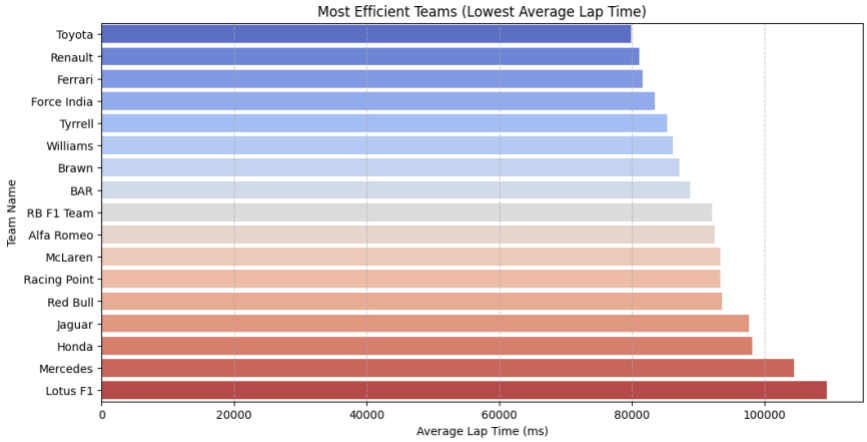


Inference: George Amick and Serafini are the most inconsistent drivers.



Inference: Bobby Ball and Alfonso Thiele are the most consistent drivers.

9. Lap Time Efficiency:



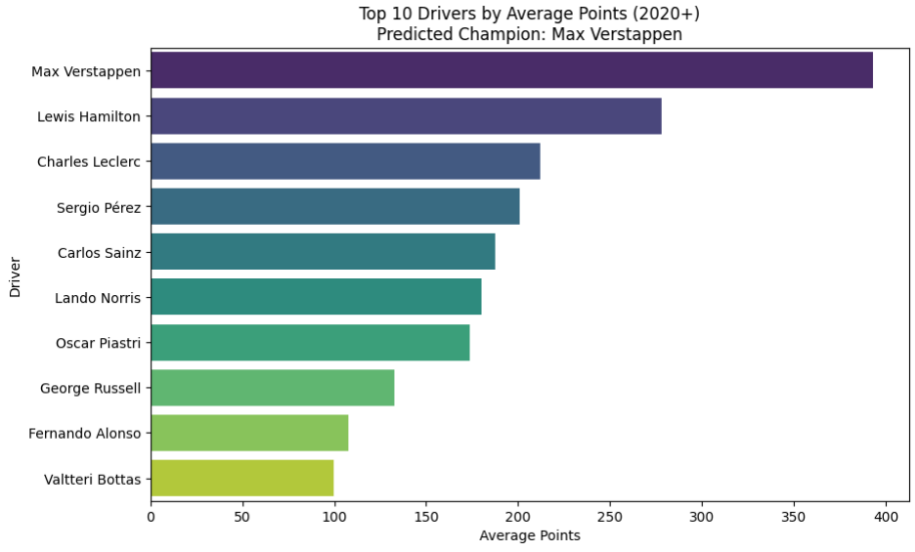
Inference: Toyota has the least average lap time and Lotus has the highest average lap time.

10. Best Team Lineup:

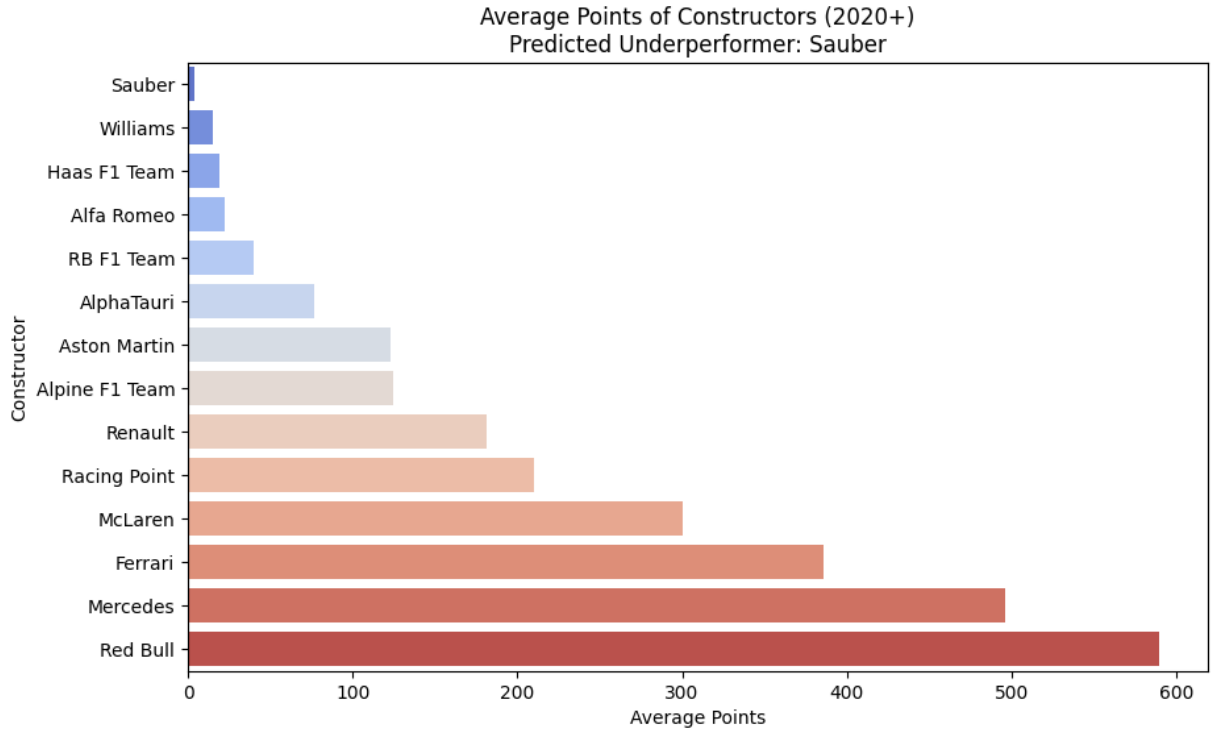
Best Team Lineup (Top 4 Drivers):

driver	total_points	wins	podiums	races	composite_score
Max Verstappen	1964.5	55	81	107	6334.5
Lewis Hamilton	1389.5	21	51	106	3459.5
Charles Leclerc	1060.0	6	33	107	2020.0
Sergio Pérez	1004.0	6	31	105	1924.0

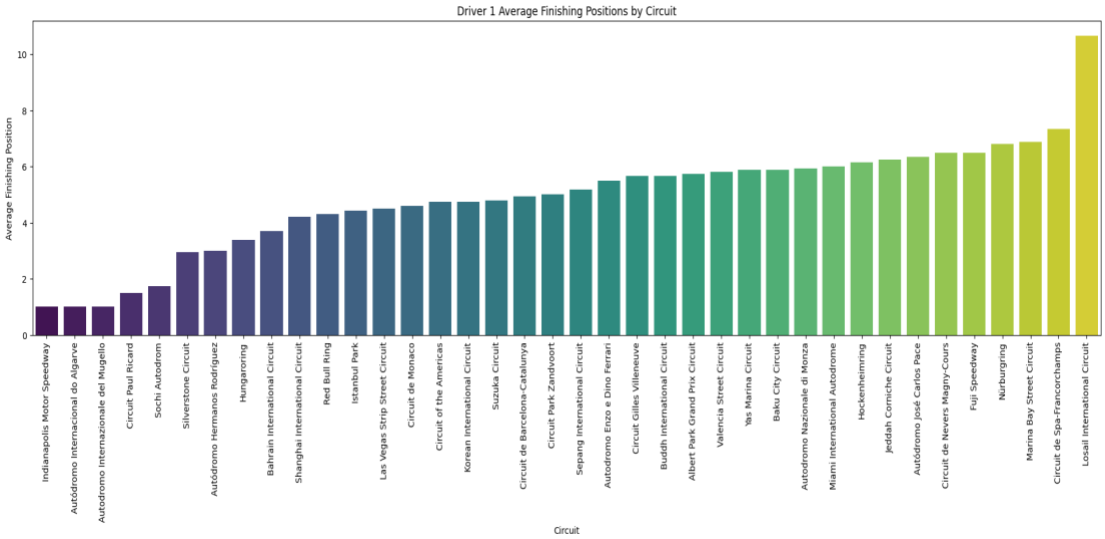
11. Predictions for 2025 Season:



12. Struggling Teams Analysis:



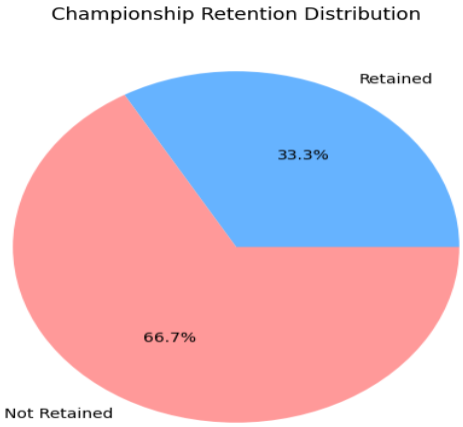
13. Driver-Specific Track Struggles:



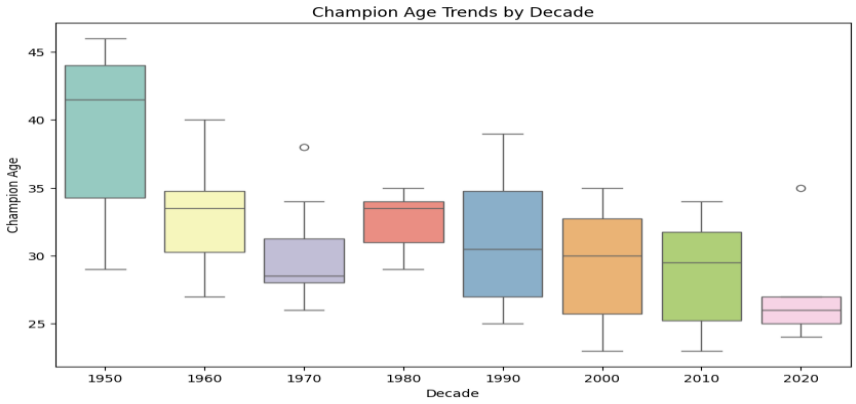
Inference: Losail International Circuit is the most difficult track with an average finish position of 10.

14. Championship Retention Probability:

The Historical Championship Retention Rate is 0.3333333333333333



15. Champion Age Trends:



Inference: The age range of champions has been decreasing over the years. The average age of champions in 1950's is 42 and in 2020's is 27.