Executive Summary - Formula 1 Championship Analysis

This study analyzes the factors influencing final race positions in Formula 1 championships, focusing on data from 2018 to 2023. Key variables such as starting grid positions, constructor performance, fastest lap times, constructor points, laps completed, fastest lap speeds, and pit stops were examined to uncover their impact on race outcomes. The findings reveal that starting grid positions strongly correlate with final positions, with a correlation coefficient of 0.64, emphasizing their significant role in determining race results. Constructor points also play a critical role, showing a moderate negative correlation of -0.53 with final positions, underscoring the competitive advantage of top teams. While drivers starting closer to the front generally achieve better results, variability persists due to factors like driver skill, team strategies, and race dynamics.

The analysis highlights that pit stops have a minimal direct correlation with final positions, indicating their context-dependent impact. Constructor performance analysis reveals that dominant teams such as Mercedes, Red Bull, and Ferrari consistently finish in higher positions with less variability, highlighting their superiority in the sport. These insights provide actionable strategies for teams to optimize race outcomes, enhance fan engagement, and contribute to motorsport analytics research.

Predictive models were developed to forecast final race positions, starting with a simple linear regression that relied solely on grid positions. Enhancements to the model, incorporating fastest lap, constructor ID, and constructor points, significantly improved its performance. An ordinal regression model further refined predictions by treating final positions as ordinal, capturing the ranking nature of the outcome. The Random Forest model demonstrated the best performance by effectively capturing non-linear relationships and providing reliable predictions. The introduction of an interaction term between constructor ID and constructor points added depth to the analysis, highlighting team-specific dynamics but with minimal improvement in overall predictive accuracy. A summary of the model performances is provided in the table below for reference.

MODEL	**MAE**	**RMSE**
Simple Linear Model (~grid)	3.68	4.69
Enhanced Linear Model (~grid + fastestLap + constructorId + constructor points)	3.16	4.31
Ordinal Regression Model (~grid + fastestLap + constructorId + constructor points)	3.26	4.54
Random Forest Model (~grid + fastestLap + constructor points + laps + fastestLapTime)	2.86	3.85
RF with Interaction Term (~grid + fastestLap + constructor points + laps + fastestLapTime + constructor points * constructor_id)	2.94	3.92

The study acknowledges several limitations, including the exclusion of external factors such as weather, track conditions, safety cars, and mid-season changes like driver substitutions and car upgrades. Additionally, the lack of granular data on in-race dynamics, such as lap-by-lap position changes and tire wear, limits the analysis. Expanding the dataset to include historical data and these external variables would enhance the robustness and predictive accuracy of the models. Despite these limitations, this analysis provides valuable insights into Formula 1 race dynamics and delivers reliable predictive tools, benefiting teams, fans, and researchers alike.