

1. Introduction

In the electrifying realm of Formula One, the path to victory is not solely carved by the roar of engines or the sheen of cutting-edge machinery. It is a multifaceted odyssey, where the intricate ballet of team dynamics, driver prowess, and strategic acumen plays out against the backdrop of relentless technological innovation and intense financial stakes. Group A11 embarks on an analytical expedition to dissect and comprehend the myriad components that converge on the tarmac, influencing the outcomes of each high-speed contest. Our investigation seeks to demystify the symbiosis between the steely resolve of drivers and the tactical ingenuity of their teams, deciphering the formula that propels some to bask in the glory of the checkered flag while others merely chase it.

The objective of our study is ambitious yet precise: to unearth and delineate the key success factors in the Formula 1 industry, wielding statistical methodologies and advanced data visualization as our tools of discovery. With a dataset that chronicles the saga of F1 from 1950 to 2023, we stand poised to construct a social network analysis that reveals the collaborative pulse of this sport, mapping the constellations of relationships that form the beating heart of the racing community. From the raw data's lineage of races, drivers, and teams, to the victories and standings that shape the sport's history, our analysis will illuminate the complex web that underlies the surface of this global phenomenon.

This project is not merely a probe into the statistical undercurrents of Formula One; it is a narrative of how leadership, vision, and communal spirit intertwine to forge champions. We aim to capture the essence of F1 success, framing a perspective that encapsulates both the quantifiable and the intangible, the analytical and the emotional.

2. Data And EDA

Our data source is from the “[Formula 1 World Championship (1950 - 2023)](https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?select=races.csv)” dataset from Kaggle. The dataset comprises all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships spanning from the inception of the championship in 1950 till the early half of 2023 season. However, since 2010 was the year when the points scoring system became synonymous with the one we recognize today, we narrowed our focus to data from 2010 onwards, which resulted in 5757 instances in our dataset. Also, we extracted key features that we believe are beneficial to our analysis. They include year, race, constructor, driver, nationality, position order, championship standing, points, grid, laps, pit stop time, and qualifying time. By concentrating on the period and key features, we ensure consistency and relevance in our analysis, allowing us to draw meaningful insights.

We then conducted exploratory analysis to uncover some very interesting big picture insights.



Figure 1:Map of the most popular cities

Formula One is a premier international racing event, therefore we wanted to gain an overview of the popular cities that have hosted races worldwide from 2010 to 2023. Observing the geographical graph above, certain regions and cities emerge as prominent destinations for these events. Europe is a popular region to host Formula 1 races due to its rich history in motorsport, with passionate fan bases and many iconic circuits like Circuit de Spa-Francorchamps, Silverstone Circuit, Circuit de Monaco, and Autodromo Nazionale Monza. The continent's proximity to major teams and manufacturers further enhances its appeal as a logistical hub for hosting races. Additionally, notable cities worldwide, such as Montreal, Austin, Sao Paulo, Sakhir, Abu Dhabi, Suzuka, Singapore, and Melbourne, play significant roles in the Formula 1 calendar.

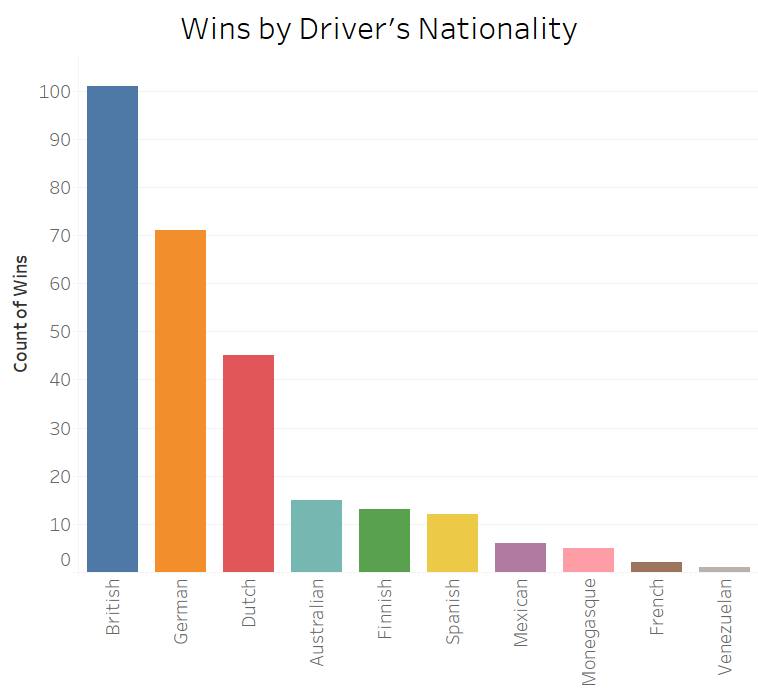


Figure 2: Number of wins by nationality of driver

Following on from the cities the races were held, we wanted to explore the diverse spread of nationalities among winning drivers. Interestingly, this distribution aligns with the prevalence of races held in the European region. Notably, Hamilton's dominance in the 2010s established British drivers at the forefront, while German drivers like Vettel and Rosberg also maintained their presence.

Intrigued by a potential correlation between a driver's nationality and their success, we conducted a chi-squared test, revealing a significant association between the two variables. On the other hand, a similar test revealed no significant association between the country of the race and their success. Thus, we intend to delve deeper into this relationship between nationality and performance in our subsequent network analysis.

Next, we looked at the more dominant forces in F1 of recent years.

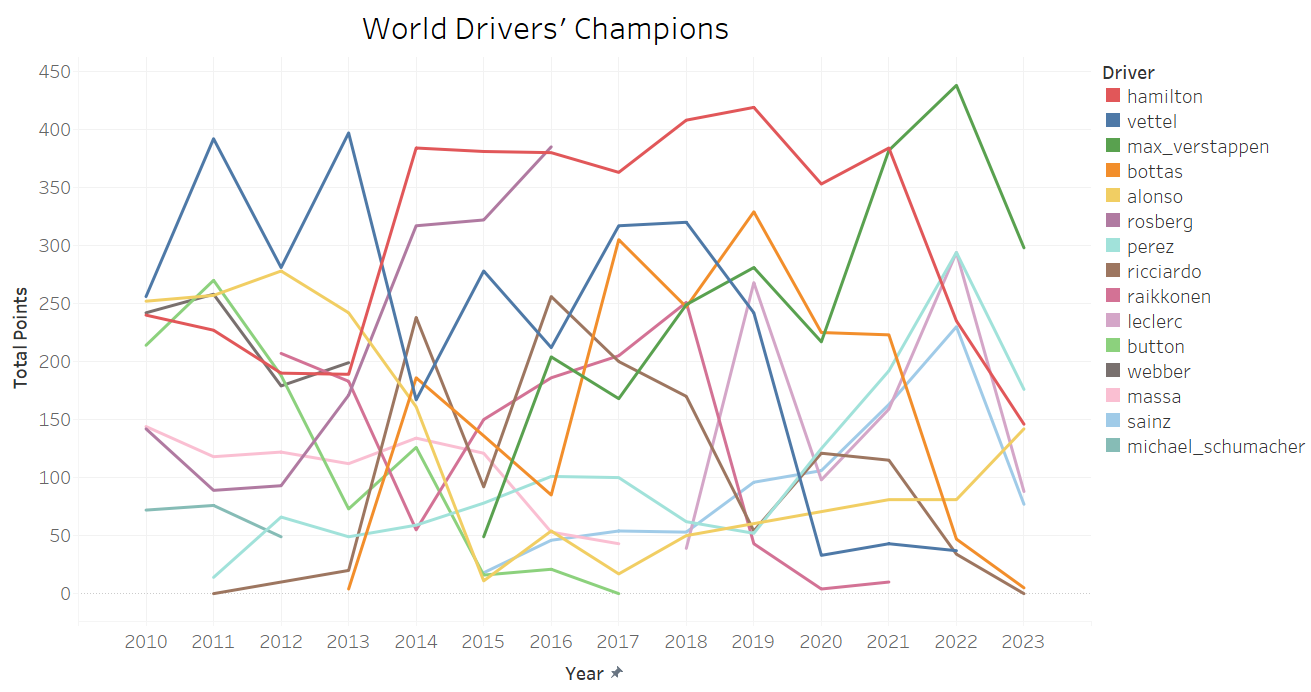


Figure 3: Drivers vs Points since 2010

Lewis Hamilton's formidable presence with Mercedes has redefined dominance in Formula 1, while Max Verstappen's recent emergence has stirred considerable excitement within the F1 community.

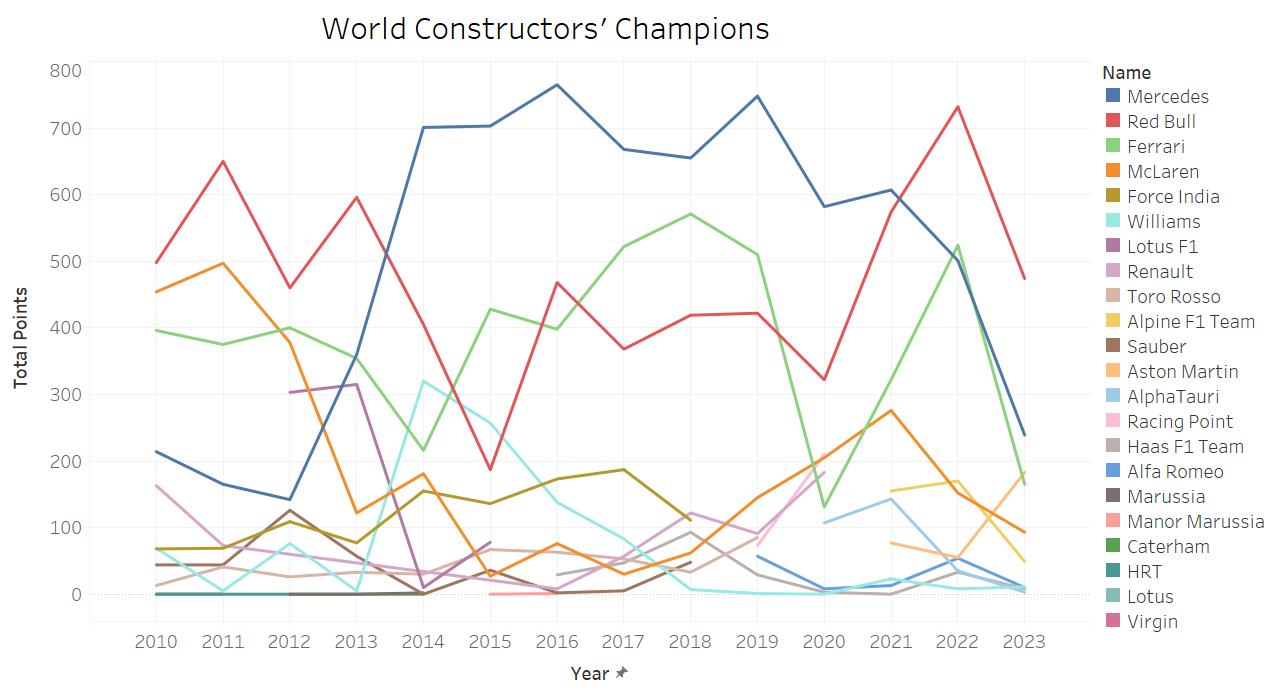


Figure 4: Constructors vs Points since 2010

Notably, the landscape of recent years has been shaped by the unparalleled success of two major constructors based in the United Kingdom: Mercedes and Red Bull. These powerhouse teams stand as the clear frontrunners in the sport, with Ferrari closely contesting for prominence in third place.

3. Statistical Analysis

Upon completion of our EDA, we were curious as to what drives these drivers and constructors to championships and wins. We conducted two different regression analyses to determine the crucial factors in race performance. The logistic regression was modeled with the dependent variable being a binary outcome of win or lose. The linear regression was modeled with the dependent variable being their finishing position.

These were the variables we looked at in our models:

#grid : position they started the race

#totalPoints: points they accrued in that year's championship

#laps: laps in that race

#year: year of the race

#round: which round (i.e which week of racing was it)

#ageAtRace: age at race

#month: which month (whether time of year has an effect)

#totalwins: totalwins accrued in that year

From the logistic regression analysis, we can infer that factors such as grid position, number of laps completed in the race, total wins, championship standing, and total points are significant predictors of race wins in Formula 1. This suggests that drivers who start with better grid positions, have a history of winning races, hold higher championship standings, and accumulate more points throughout the season are more likely to win races. On the other hand, the linear regression analysis reveals that grid position, number of laps in the race, round of the race, month of the race, championship standing, and total points are significant predictors of finishing position. This indicates that similar factors influence both winning races and achieving higher finishing positions, emphasizing their importance in overall race performance.

The insights from the logistic regression and linear regression can guide teams and sponsors in identifying key drivers to support and invest in, as well as inform race strategy decisions, such as optimizing grid position and focusing on consistent performance throughout the season. The round or gameweek in which the race is held can indeed be surprisingly important. In Formula 1, the season consists of multiple races held at different circuits around the world, and the timing of each race within the season can have significant implications for team strategies and driver performance. For example, races held later in the season may carry more weight in terms of championship standings, leading teams to adopt more aggressive or conservative race strategies depending on their position in the standings. Additionally, the conditions at different circuits can vary based on the time of year, with races held in certain months experiencing different weather patterns or track temperatures, which can significantly impact car performance and tire wear. Certain months may coincide with specific weather conditions or regional climates, affecting track conditions and driver visibility. For instance, races held in the middle of summer may experience higher temperatures, leading to greater tire degradation and requiring teams to adjust their setups accordingly.

Overall, the significance of the round and month variables underscores the multifaceted nature of Formula 1 racing, where factors such as circuit characteristics, weather conditions, and strategic considerations can all influence race outcomes. Understanding these nuances allows teams to better prepare for each race and optimize their chances of success throughout the season.

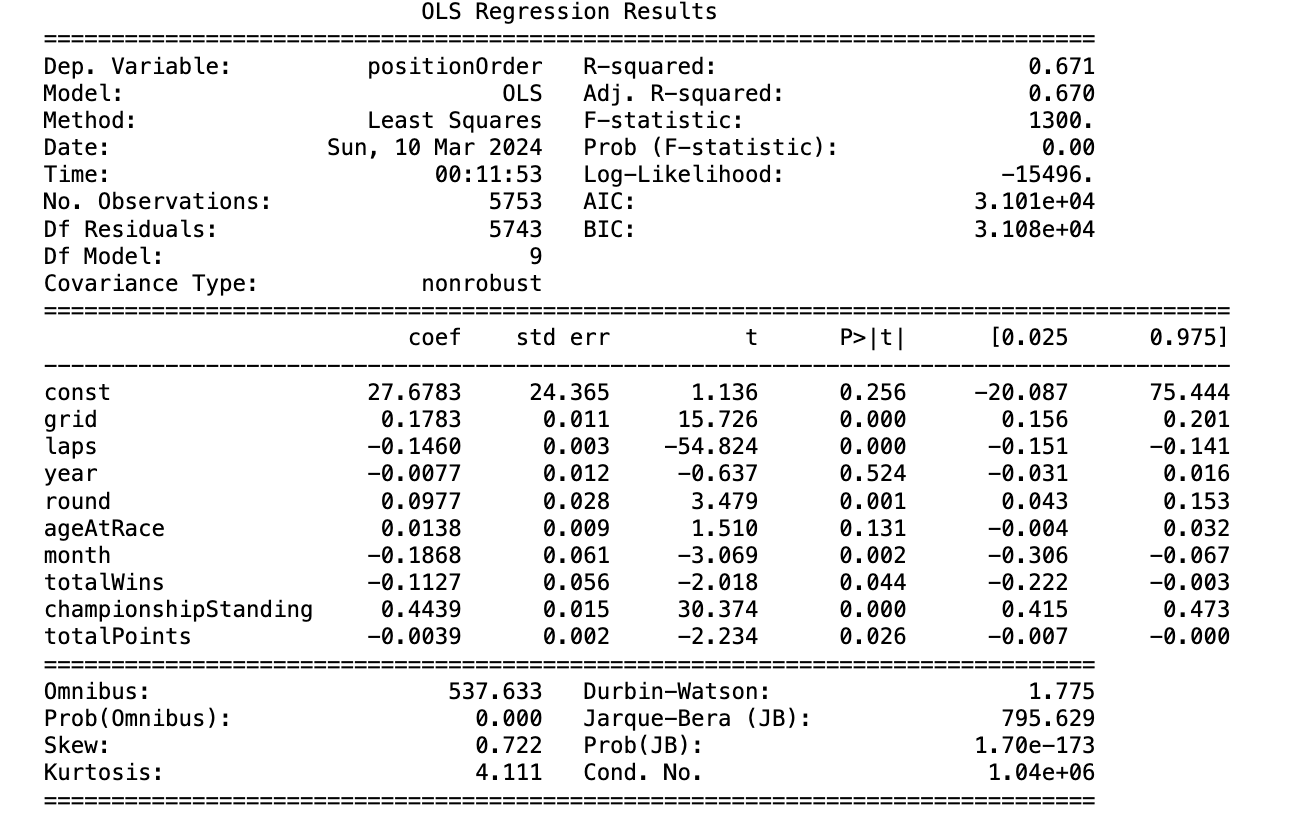
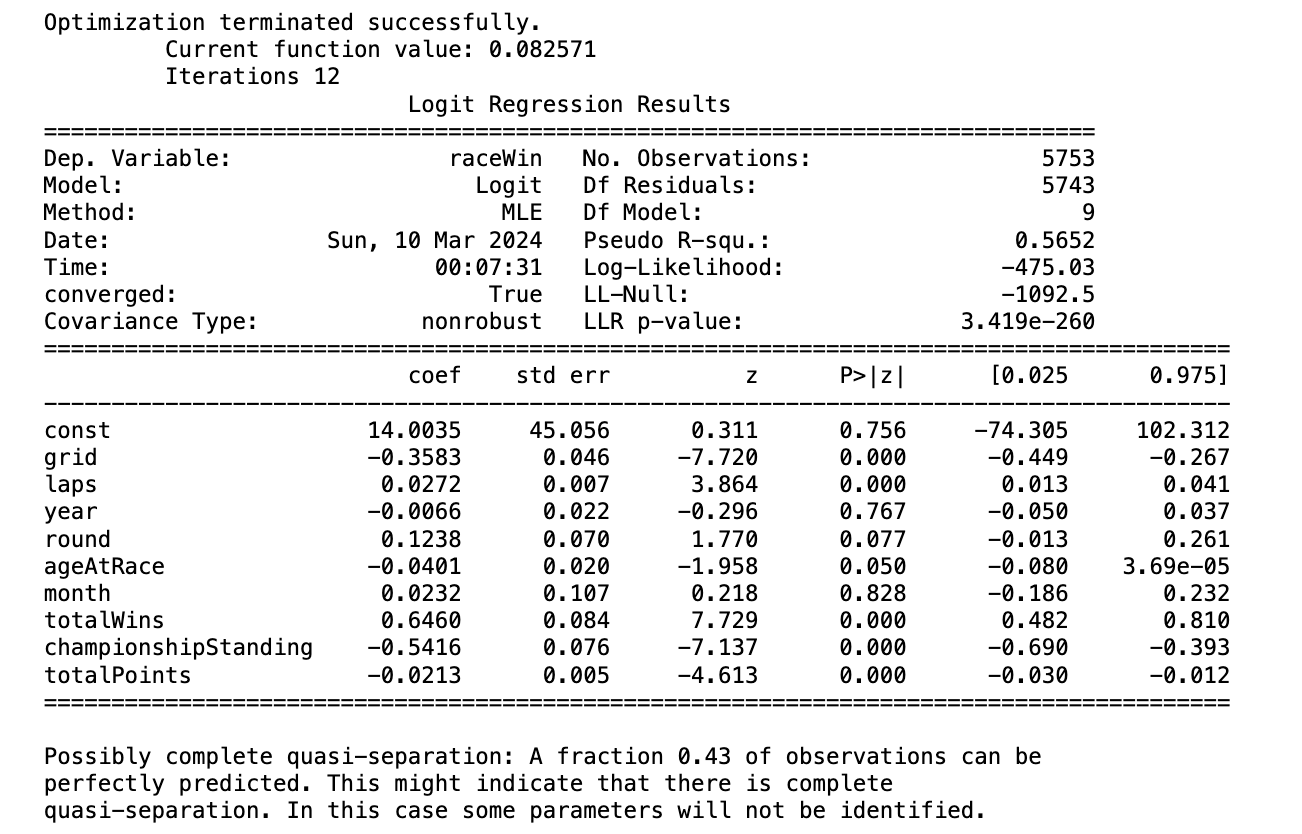


Figure 5: Side by side comparison of the two models (Logistic regression on the left and linear on the right)

4. Network Analysis

1. Average pitstop speed network:

This network analysis sheds light on the relationships between drivers and teams based on average pitstop durations. The visualization highlights drivers as blue nodes and teams as green nodes, with the thickness of edges indicating the average time spent during pitstops. Notably, Williams emerges as a standout performer, evidenced by the numerous thin edges connected to it, suggesting that they boast the fastest average pitstop times among all teams.

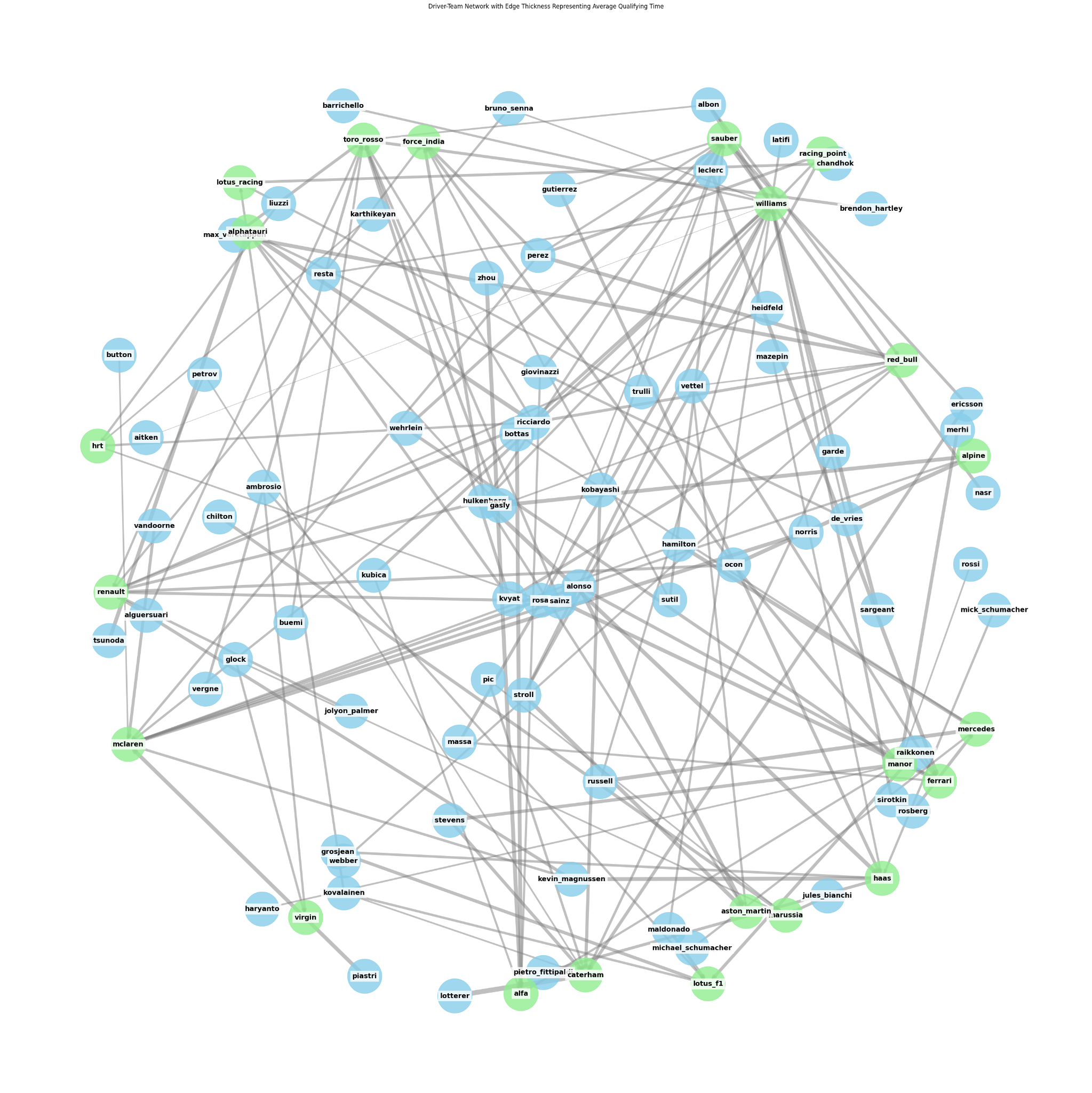


Figure 6: Pitstop speed network

Delving deeper, the network analysis underscores a strategic imperative for lower-order teams. With limited resources and often less competitive cars, these teams prioritize optimizing pitstop efficiency as a means to gain a competitive edge. For them, enhancing pitstop times becomes a crucial aspect of their race strategy, offering a tangible opportunity to make up for deficits in raw race pace.

Conversely, higher-order teams, equipped with more advanced technology and greater resources, may afford a degree of flexibility when it comes to pitstop times. While they also strive for efficiency, their focus may be more balanced between pitstops and overall car performance. Nevertheless, even for these teams, minimizing pitstop times remains imperative for maximizing their chances of success.

1. Qualifying times network

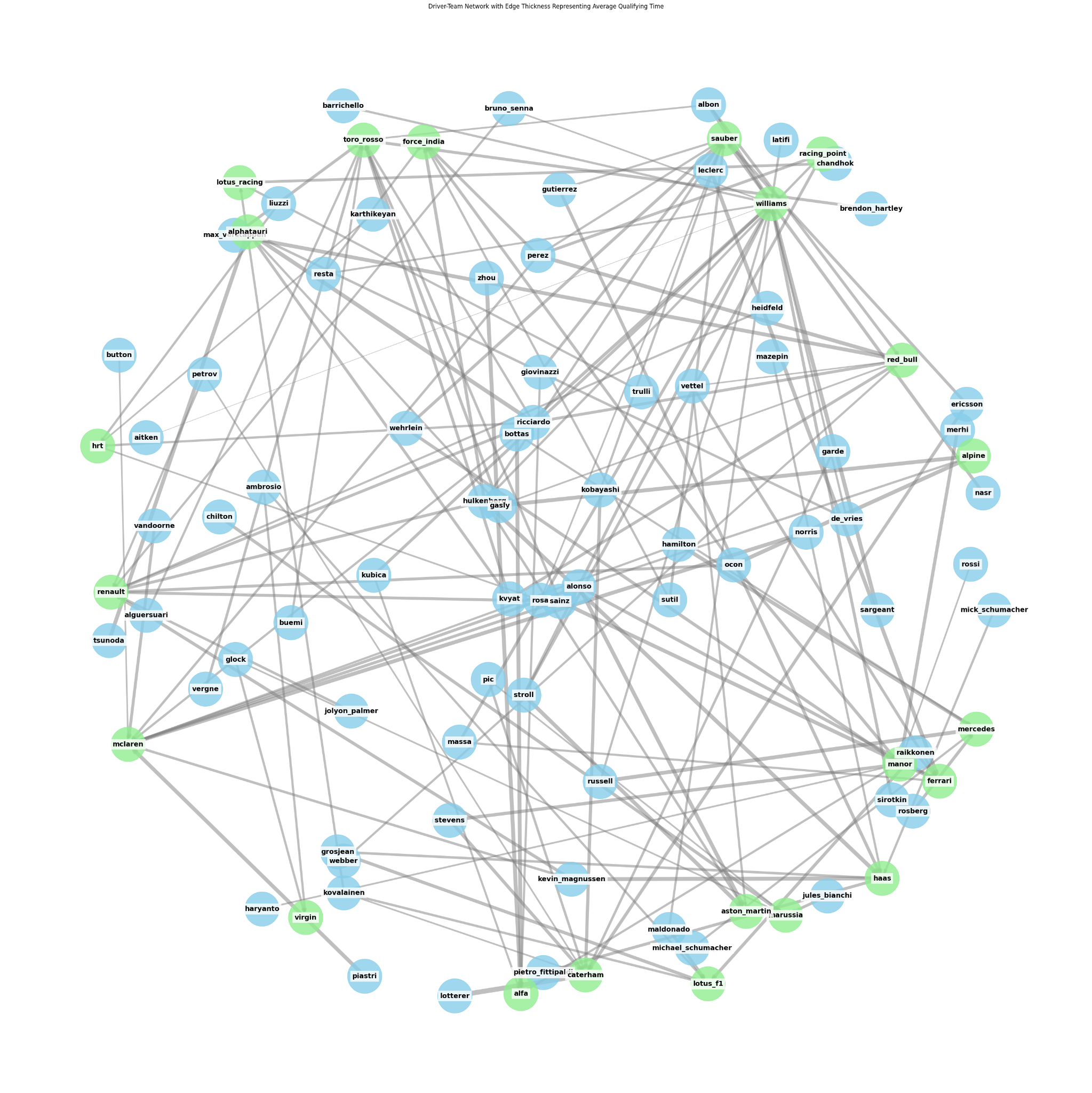


Figure 7: Qualifying time network

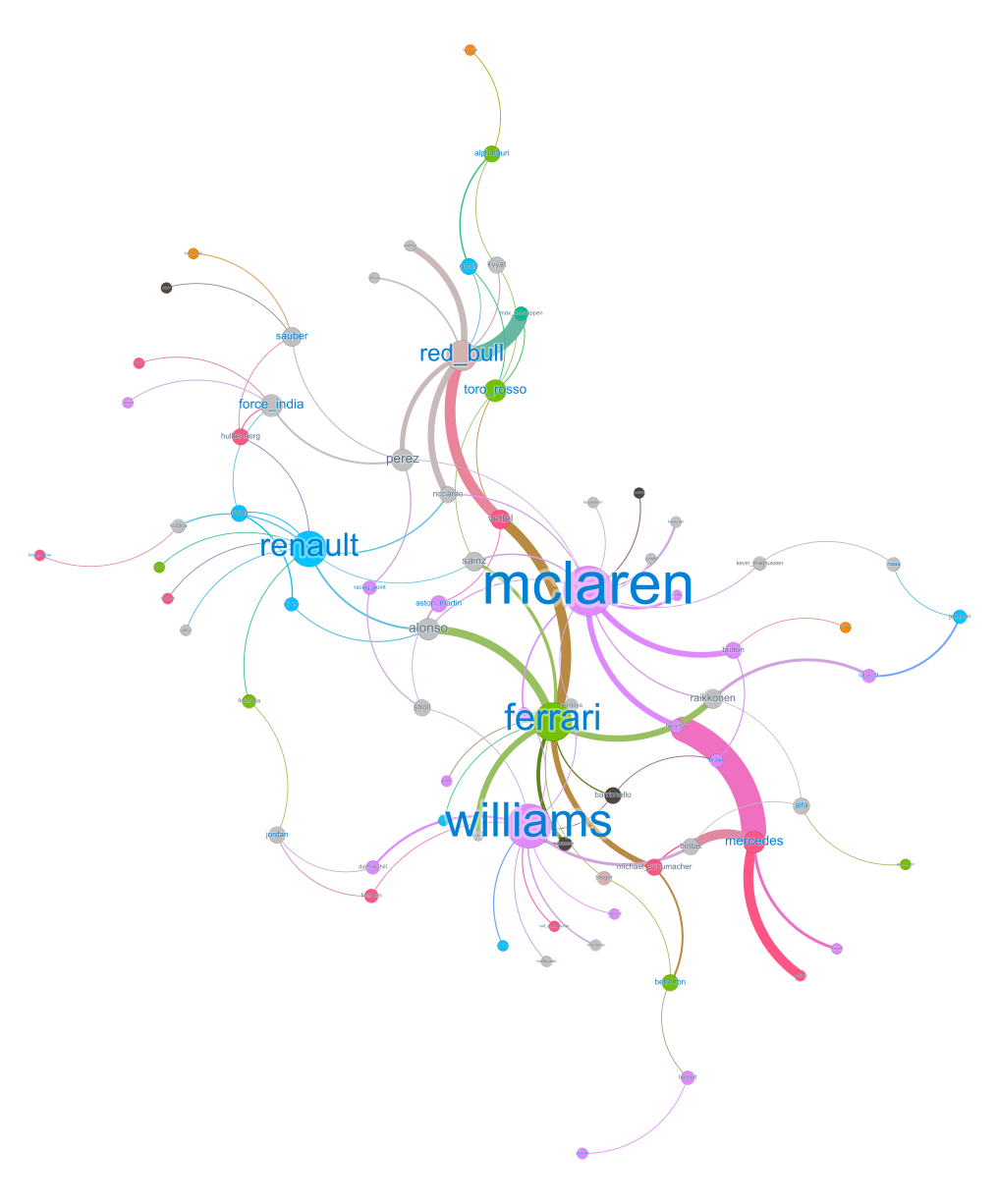
This network analysis focuses on the interplay between drivers and teams through the lens of average qualifying times. The visualization portrays drivers as blue nodes and teams as green nodes, with edge thickness symbolizing the average time spent during qualifying sessions. Thicker edges indicate slower qualifying times, a crucial factor influencing race start positions and subsequent race outcomes.

Lower-tier teams such as Williams and Toro Rosso are frequently challenged by slower qualifying times, which inevitably impact their starting positions on the grid. These teams often grapple with technical limitations or budget constraints, hindering their ability to compete with the top-tier contenders. Conversely, powerhouse teams like Red Bull Racing and Mercedes-AMG Petronas Formula One Team tend to dominate qualifying sessions, leveraging their substantial resources and technical expertise to secure favorable grid positions. However, even these elite teams occasionally encounter setbacks, such as mechanical failures or strategic missteps, which can result in suboptimal qualifying performances.

The qualifying session in Formula 1 serves as a critical juncture for teams to fine-tune their cars and strategies based on real-time feedback from drivers and engineers. For lower-tier teams, this period represents an opportunity to identify and address performance gaps, with a focus on optimizing car setups and maximizing potential. Conversely, top-tier teams often approach qualifying with a different mindset, aiming to secure favorable grid positions early on before experimenting with alternative strategies and setups. Despite their divergent approaches, all teams recognize the importance of qualifying in shaping their race-day prospects, underscoring the strategic depth and competitive nature of Formula 1.

1. Performance and nationality network

This network provides an insightful analysis of the relationships between drivers and teams within the Formula 1 racing circuit, leveraging a network plot to illustrate their collaborations and successes. This innovative approach not only showcases the competitive dynamics but also highlights the sport's international diversity and strategic alliances, offering a unique perspective on F1's intricate ecosystem. The nodes, distinguished by colors corresponding to nationalities, symbolize the diverse array of talent and expertise present within the F1 ecosystem. The edges, varying in thickness based on the points earned, serve as visual indicators of the strength and efficacy of collaborations.

Figure 8: Performance and nationality network

Within the network analysis, degree centrality emerges as a pivotal metric, offering insights into the significance of nodes within the Formula 1 (F1) ecosystem. Degree centrality, as reflected in the provided tables, quantifies the influence of both teams and drivers, with larger nodes indicating entities of greater importance. Notably, prominent teams like McLaren and standout drivers such as Perez, Alonso, and Sainz command substantial centrality values, underscoring their pivotal roles in shaping F1 dynamics and outcomes.

Degree Centrality Tables (Teams and Drivers)

|  |  |
| --- | --- |
| Team | Centrality |
| McLaren | 0.093 |
| Red\_bull | 0.082 |
| Renault | 0.072 |
| Ferrari | 0.061 |
| Mercedes | 0.051 |
| Force\_india | 0.052 |
| Williams | 0.052 |

|  |  |
| --- | --- |
| Driver | Centrality |
| Perez | 0.052 |
| Alonso | 0.041 |
| Sainz | 0.041 |

#### 

#### 

#### 

Furthermore, the thickness of edges between nodes serves as a visual cue to the strength of collaborations within the F1 network. Thicker edges signify robust partnerships between drivers and teams, indicative of successful alliances that yield commendable results on the racetrack. Noteworthy collaborations, exemplified by teams like Ferrari, Mercedes, and Red Bull, often translate into higher point accumulations, highlighting the strategic prowess and effectiveness of these alliances in navigating the competitive landscape of F1.

Additionally, the network analysis unveils intriguing trends regarding nationality clusters, shedding light on potential patterns of dominance or prevalence within the sport. These insights provide teams with valuable strategic cues, guiding their decisions and approaches as they seek to optimize performance and capitalize on emerging opportunities in the ever-evolving world of F1 racing. By leveraging these analytical findings, stakeholders can gain a deeper understanding of the intricate dynamics at play within the F1 network, empowering them to make informed decisions and strategic investments that drive success on and off the track.

This network analysis offers a powerful tool for visualizing the competitive landscape, highlighting the importance of centrality, strategic collaborations, and the international diversity within the sport. This approach provides both fans and professionals with a new perspective on the dynamics of F1 racing, although it also underscores the necessity of comprehensive data for a fully accurate representation.

5. Conclusion

The factors that contribute to the success of Formula 1 have been studied extensively. It has been found that several elements play a significant role in determining drivers' performance on the track. Among these factors, the ones that stand out are drivers who begin races from better grid positions, have a history of winning races, and hold higher championship standings while accumulating more points throughout the season. These findings highlight the importance of having a strong starting position and maintaining consistent performance throughout the season to continue winning races.

In addition, the analysis indicates that while pitstop strategies are crucial for all teams, they are particularly vital for lower or mid-level constructors, who can leverage improvements in pitstop efficiency and qualifying times to gain a competitive edge. Williams was highlighted for having the fastest pitstop times, suggesting a focused effort on this aspect can result in considerable gains in race standings.

The social network analysis of top five finishes since 2010 shows that certain teams, specifically Ferrari, Mercedes, and Red Bull, frequently achieve high-quality victories, which may be attributed to effective collaboration with drivers and the successful execution of team strategies. With many races won, McLaren and Renault's prominence points to their potential for future success in the sport.

In conclusion, the Formula 1 industry is characterized by a combination of statistical performance metrics and complex social dynamics, where strategic decisions, team collaboration, and driver skills all play critical roles in achieving victory. The findings of this report provide valuable insights into the multifaceted nature of success in Formula 1 racing and suggest areas of focus for teams looking to improve their performance in future seasons.