Team 174 Final Report

Sam Wagner, Conner Nations, Braxtyn DeGolyer, Brian Yuhas, Davud Davudov, Edwin Pimentel

Introduction

Gaining insight into the root causes and correlations that lead to success or failure for the teams and drivers in Formula 1 (F1) racing has long been difficult to achieve. Many previous regression, classification, and machine learning methods have struggled to tease out significant answers that go beyond random variance. Cited reasons for this are often the complexity of the data, the changing nature of the sport, heavy bias and noise in the data, and predominantly, the amount of sheer randomness that occurs in an F1 race. Our analysis will attempt to break down these flaws and utilize a unique dataset to provide new insight into constructor and driver performance in F1 racing, potentially providing F1 teams a leg up on their competition.

Problem Definition

Our team will build an F1 Racing dashboard to highlight key factors that may contribute to team or driver success (or failure) in an F1 race by utilizing the FastF1 dataset (via API connection).

Literature Survey

Two papers [1, 2] outline various methods of visualizing Formula 1 data and aerodynamics in a web-based framework. This can guide us towards a web-based visualization. We could potentially enhance this interactivity and include connections between all the visualizations in conjunction with real-world physics. Research [3] attempts to separate the driver's performance from their cars. While the author clearly focused his efforts on building dashboards via Tableau, it seems that a web-based and interactive visualization could represent the data better and provide more insight. Another author [17] concludes that both drivers and cars contribute around 15% and 20% respectively to the outcome of the race. This gives us a good backing on the driver and car relationship. These works suggest that a web-based dashboard is certainly a feasible goal.

Several papers discuss race simulation and strategy. One work [7] developed a methodology for race simulation to determine time to finish the race. This work can help us to identify and determine several key factors needed to improve prediction model accuracy. Another paper [8] has a developed neural network-based prediction model to determine whether a driver should make a pit stop and make any tire change. This gives us a basis for neural networks in Formula 1. An article [9] discusses use of machine learning models to predict impact of tire-change decisions on rank positions. Data cleaning, feature selection methodology, and challenges discussed in this paper can give direction for our project. Two studies [10, 11] analyze the tires of Formula 1 cars under different race conditions. We can complement this information to better understand the effects of tires in a race.

The work outlined in the previous paragraph gives us direction with respect to prediction-based visualization, however we have determined that it would not be prudent for us to pursue a prediction model for this analysis. Nevertheless, it is beneficial to heed the advice of these works in order to determine which independent variables we should consider in our dataset and which visualizations we should produce on our dashboard to be able to tell the best story about the causes and correlations of F1 success.

Proposed Method

Overtake Analysis

Overtaking is a key part of F1 racing and can mark the difference between winning and losing. Although the driver's skill is the predominant determinant of overtaking ability, there are several factors outside of the control of the driver that can make or break an overtake. Some of these factors are the track shape, speed/position of the car in front, DRS performance, wake turbulence, and positioning of other cars (including a potential safety car).

In our dashboard, we will analyze and display the overtaking potential of a driver using several methods centering around differences in car speeds and track sector overtaking "friendliness". Our analysis goal is to determine when a driver has a good or bad chance of overtaking. Overtaking data is not natively available in our data source, instead it must be calculated using time gaps between cars and positional coordinate data. This results in a set of potential overtakes, which must be then classified based on the certain criteria including; track status, DRS position, and pit lane status. In addition to the calculated number of overtakes, we also seeked to quantify the efficiency of an overtake via calculating the time required to set up the overtake. A more efficient overtake is desired since time spent in close proximity to a car ahead will degrade the performance of a car from turbulent air, reducing the effectiveness of aerodynamic surfaces and decreasing the efficiency of cooling systems.

Tire Selection

Tire selection is an integral part of race strategy in Formula 1 and is known to contribute highly to success or failure in a given race. F1's regulating agency, the Fédération Internationale de l'Automobile (FIA), mandates the tires themselves are not to be altered by any teams and are provided via a contract between the FIA and Pirelli, the tire manufacturer. However, Pirelli manufacturers seven different "compositions" of tires (five dry-track tires, one intermediate tire, and one wet-track tire) from which the various F1 teams can choose three to have in their possession during a race. Each composition of tire has various benefits and drawbacks that the teams must weigh (e.g. wet-track tires provide significantly more grip than dry-track tires do, but also reduce top speed of the car).

Our F1 dashboard will attempt to determine the validity of the common F1 assumption that tire selection is integral to race success. We will allow the user to drill down on lap times by year, race, and driver in order to assess the distributions of lap times across the various tire compositions. We will also be sure to include weather conditions in our analysis (and include some weather variables as independent variables in our dashboard) as rainy or cold weather can negatively affect lap times across all drivers regardless of the tire composition.

Driver Trend

One of the several questions we are trying to answer in this project is to determine reasons for bad races, especially for the drivers with an upward trend. From multiple factors, we will focus on drivers past performances, weather conditions, type and direction of the circuit, number of the laps and duration of the race. More specifically, we would like to analyze and investigate the impact of these causes that can result in an unsuccessful race for drivers with winning trends.

In our visualization platform, we will include capability to select and demonstrate race trend change over time for any selected diver. Moreover, we will build a model which can rank these factors for selected drivers for any particular race. This model can help to prevent unsuccessful races and improve chances of winning.

Visualizations

One area F1 analysis is lacking is in visualizations. We want to put together a cohesive and interactable visual based webpage. The consideration was between using D3, streamlit, and Tableau. D3 has more customization than Streamlit and Tableau but would take too long to develop. Tableau is very quick and easy to develop in but has much less customization and intractability. Streamlit is a python package which allows for customizable and easy web page development. It gives the perfect medium between Tableau and D3. Also, since it is a python package it will have seamless intractability with the rest of our code. As discussed above, we have several analyses that we want to explore. In order to visualize these, we will be using plotly and matplotlib graphs. Plotly graphs allow for interactivity like hover over metrics. Our data set also allows for interesting racetrack visuals made in matplotlib.

<u>Experiments/Evaluation</u>

DRS and Overtakes

The drag reduction system (DRS) is a driver-controlled device aimed at reducing aerodynamic drag in order to increase top speed and promote overtaking in specific DRS zones. Since the introduction of DRS in 2011, the number of overtakes in F1 has increased and drivers perform overtakes mostly with the help of this device.

We performed experiments to figure out how much the introduction of DRS impacted overtakes in F1, meaning, how many overtakes were performed with and without utilizing DRS. In our experiments we focused on the Grand Prixs of 2021 and observed position changes in drivers and status of the DRS while this happened. With the data collected we analyzed the number of overtakes per Driver and per Grand Prix (track). The number of overtakes per driver would let us know which drivers are better at overtaking and with what approach (with or without DRS). On the other hand, the overtakes per track give information on which tracks are more overtake friendly and which ones are easier to perform overtakes without the use of DRS.

Observations:

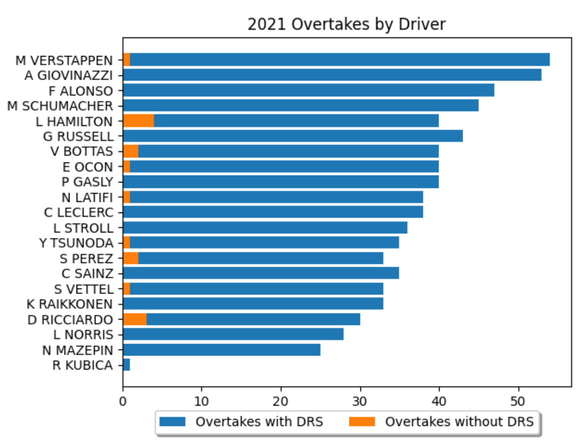

Fig. 1

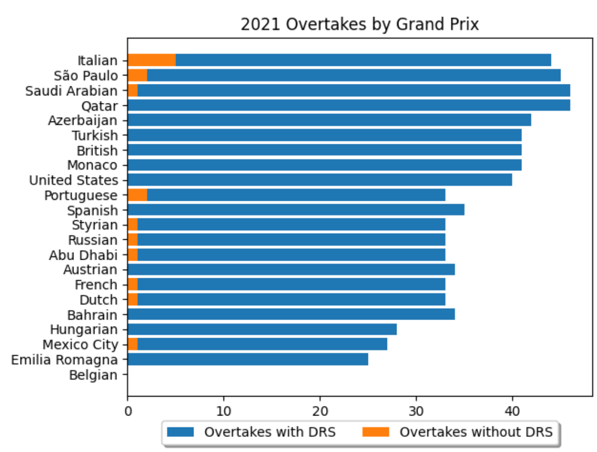

Fig. 2

Fig. 7

During our experiments we were able to observe that most of the overtakes nowadays are performed using DRS, to the point where some drivers perform overtakes exclusively using DRS in most tracks. Additionally, we observed that the track with more overtakes without DRS was the Italian and the hardest to perform overtakes was Emilia Romangna (Belgian GP was canceled). On the other hand, the driver with the most overtakes in the year was Verstappen, and the driver that was able to perform the most overtakes without DRS was Hamilton.

Driver Trend and Overtakes

Often only finishing positions are used to rank the skill of a driver, but in reality a more useful metric is the ease in which they can both gain and hold position. A driver that is able to both keep a car behind for long periods of time while also not wasting time when overtaking will on average be more successful. The following bar charts, Fig.3 and Fig.4, were created to show attacking time (within 1 second of the car ahead) and defending time (the car behind is within 1 second) and build an F1 analytical foundation.

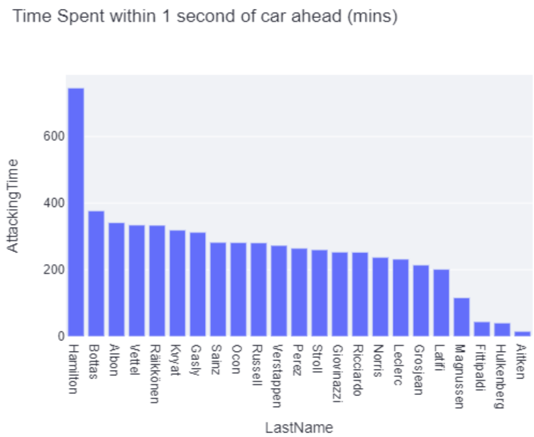

Fig. 3

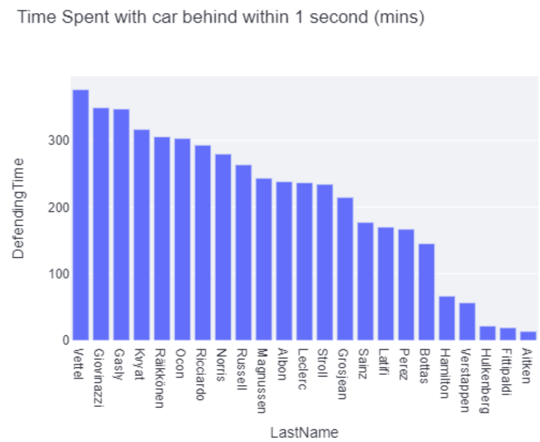

Fig. 4

Following this step, we wanted a way to group the drivers together based on their overtake ability. There are several options in unsupervised learning to do this. We considered spectral clustering, Gaussian Mixture Models, and K-means. After experimenting with each graph they made almost identical results. Thus, we went with the simplest and easiest to compute model in K-means. We also tested with several number of clusters. Three had the best balance between not having too small of clusters and max distance of points from their centroid. In order to better visualize our K-means algorithm we display it on a changeable 2D axis. Also, the same observations are made on a seasonal basis with similar clustering results. When we look at different races and seasons there are a lot of differences between drivers on different tracks. This shows us that most drivers perform at a different level on different tracks, likely a latent variable of how well a particular car is suited to each track layout. We also have to consider that the best driver in qualifying has less chances to overtake as they start in first.

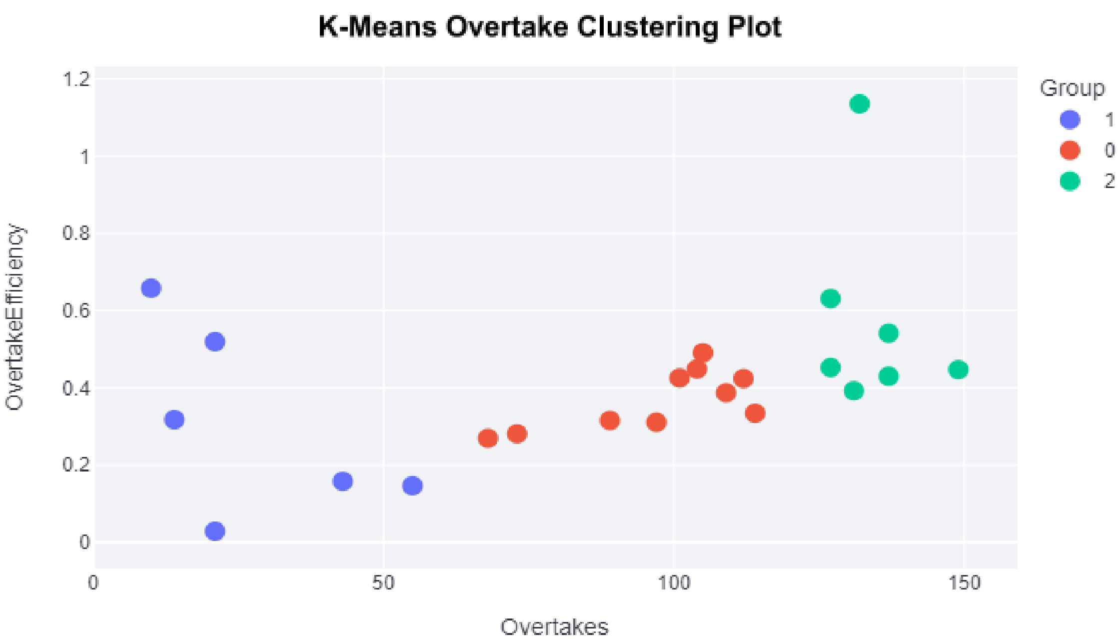

Fig. 5

Tire Analysis

Our team experimented with five years of tire composition and weather data from 2018-2022 to analyze the variation in lap times across F1 races, drivers, and weather conditions in an attempt to determine optimal tire compositions for different situations and different drivers. This analysis was embedded in the larger F1 Dashboard that includes the other analysis topics described in this report.

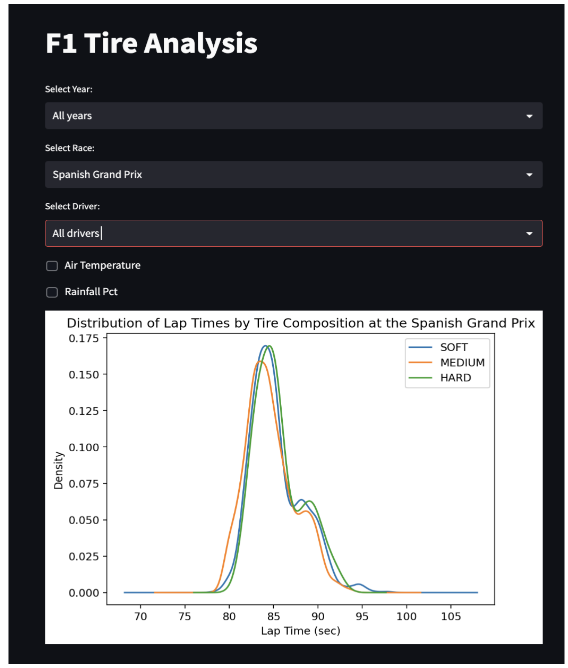

Fig. 6

Our analysis was accomplished by first isolating the necessary variables (lap times, tire compositions, races, drivers, and weather conditions) from the vast dataset available through the FastF1 API connection. Once this was done, all the variables were joined together (some by simple merge, others by aggregation) into a final table which we were able to draw lap time distribution visualizations from on the Streamlit dashboard. Outlier lap times that represented fairly common race events (pit stops, safety car laps, crashes) had to be accounted for which was done by filtering out lap times with a z-score above one. Dropdown menus were then added on the dashboard to allow the user to select the parameters that they desired (race, year, driver, weather conditions).

Common F1 wisdom, which is briefly described in the Proposed Method section above, is that tire composition has a massive effect on lap time and that, because of this, choosing the correct tire composition at the correct points in the race is crucial to success. Vast amounts of resources are invested by F1 teams to make these decisions correctly. However, what we have found in this analysis is that while this is true for some F1 races, it is not necessarily true for others.

Looking at the two figures, Fig.6 and Fig.7, some F1 races appear to foster very little variation across the different tire compositions, while others are particularly volatile. For instance, at the Spanish Grand Prix (left), lap time distributions are almost identical while at the Singapore Grand Prix (right) lap times can be chaotic. What this may suggest for F1 teams' race strategies is that the tire composition matters very little at the Spanish Grand Prix, and therefore the they can afford to be flexible with tire decisions, focus more on the age (wear down) of the tire rather than the composition, and make more strategic decisions in other areas of race strategy. On the other hand, at the Singapore Grand Prix, teams must choose wisely which tire composition they want to use during which periods of the race. Choosing the wrong tire at the wrong time could lead to a significant lap time disadvantage (as is the common F1 wisdom).

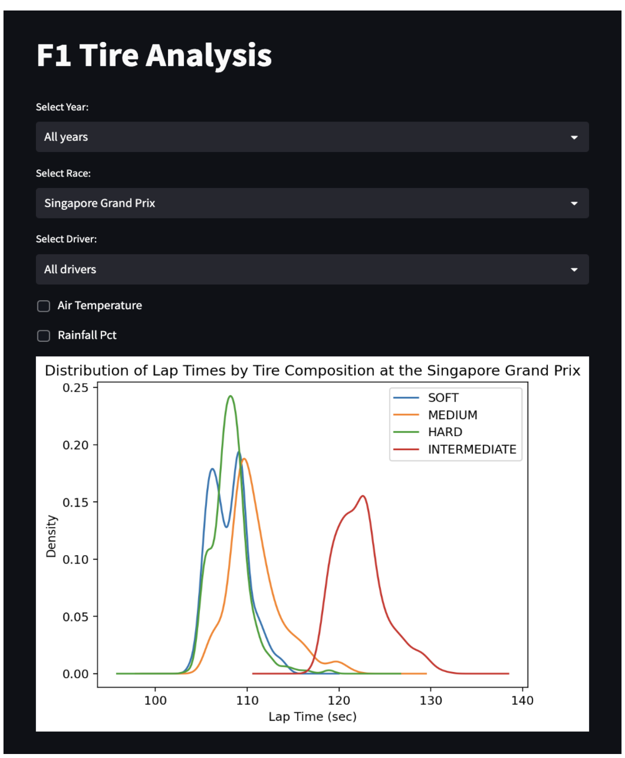

Fig. 7

We were also able to isolate driver performance by comparing their lap times to the average and control for weather changes to determine which drivers perform the better/worse than average on the different tire compsights, but is nevertheless good information for F1 teams to have on their drivers.

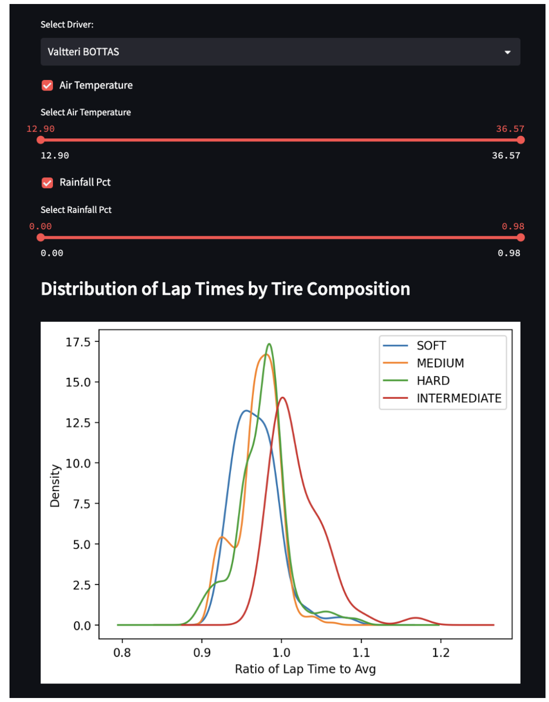

Fig. 8

Dense Geospatial Track Visualization

The abundant and extremely dense GPS positioning data (over 9GB) we discovered in the dataset led us to investigate a dense geospatial visualization. Our team was unable to find others doing this *anywhere* with real data, and our overall information delivery was enhanced with the product of this work. Our method consisted of creating a dense [longitude,latitude] vector matrix in ArcMap of all of our race tracks, and then multiplying the fastF1 gps *relative* vectors [X,Y] matrix by our new ranges for each track. The resulting geospatial charts were astoundingly accurate and we quickly found an opensource solution for sharing this information and plotting these new points onto our solution was a custom MapBox tileset that is contextually accessed through menu boxes and drop down lists in our stream lit dashboard. Our team is excited about plotting even more information onto these tracks, and we do provide some options on our dashboard, but the processing is intensive and beyond the scope of this semester. An example of one of our fully interactive F1 Geospatial tracks is pictured in Fig.9, and is something no other F1 dashboard or analytical tool presents.

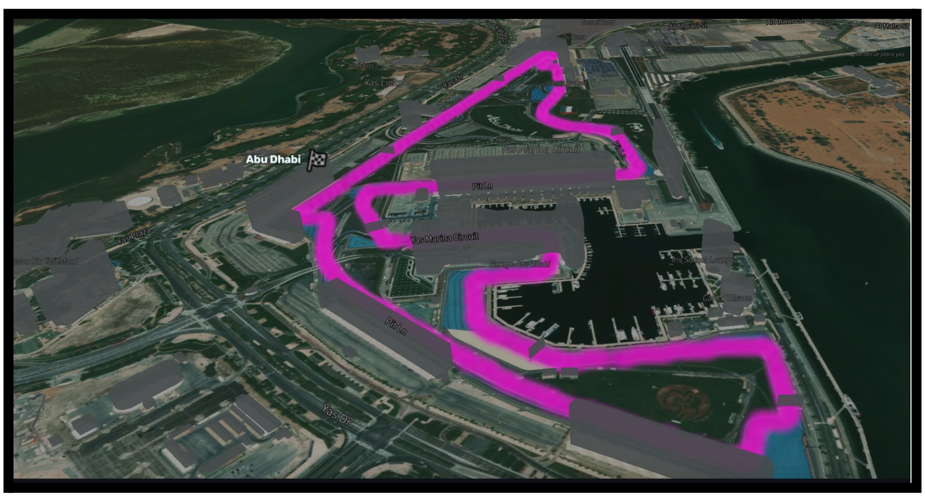

Fig. 9

Conclusions

In this analysis, we were able to make several interesting observations and draw significant conclusions from them. Firstly, in our overtake analysis, we were able to confirm the recent F1 trend of drivers being increasingly reliant on DRS to be able to overtake their opponents. We were also able to determine that the Italian Grand Prix was the easiest track for drivers to perform overtakes without the help of DRS. Additionally, we were able to better isolate driver performance by looking at attacking and defending times, calculating an “overtake efficiency” metric, and clustering drivers into three tiers. Lastly, in our tire analysis, we were able to conclude that the common conception of tire composition heavily affecting lap time might only be applicable on certain tracks. Not all tracks have shown recent evidence of this philosophy holding true.

Additionally, we found that Streamlit was an effective platform to relay our results and explore the data we were experimenting on. We would recommend this platform for future web-based dashboarding projects.

Project effort & contributions were distributed evenly among team members.

Works Cited

1. T. Lampprecht, D. Salb, M. Mauser, H. Van De Wetering, M. Burch and U. Kloos, "Visual Analysis of Formula One Races," 2019 23rd International Conference Information Visualisation (IV), 2019, pp. 94-99, doi: 10.1109/IV.2019.00025.
2. Filipe F. Buscariolo, Julien Hoessler, David Moxey, Ayad Jassim, Kevin Gouder, Jeremy Basley, Yushi Murai, Gustavo R.S. Assi, Spencer J. Sherwin, Spectral/hp element simulation of flow past a Formula One front wing: Validation against experiments, Journal of Wind Engineering and Industrial Aerodynamics, Volume 221, 2022, 104832, ISSN 0167-6105
3. 05-02-2022 Moghaddar Antoine BA EEMCS enhancement of Formula One driver ... (n.d.). Retrieved October 12, 2022, from https://essay.utwente.nl/89793/1/Moghaddar_BA_EEMCS.p df
4. Piquero, A. R., Piquero, N. L., & Han, S. (2021). Identifying the most successful Formula 1 drivers in the turbo era. The Open Sports Sciences Journal, 14(1), 151– 157. https://doi.org/10.2174/1875399x02114010151
5. Bell, A., Smith, J., Sabel, C. E., & Jones, K. (2016). Formula for success: Multilevel modelling of Formula One driver and constructor performance, 1950–2014. Journal of Quantitative Analysis in Sports, 12(2). https://doi.org/10.1515/jqas-2015-0050
6. Bayesian analysis of Formula One race results: Disentangling driver ... (n.d.). Retrieved October 12, 2022, from https://www.researchgate.net/publication/359277496_Baye sian_Analysis_of_Formula_One_Race_Results_Disentangling _Driver_Skill_and_Constructor_Advantage
7. Heilmeier, A., Graf, M., & Lienkamp, M. (2018, November). A race simulation for strategy decisions in circuit motorsports. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 2986-2993). IEEE.
8. Heilmeier, A., Thomaser, A., Graf, M., & Betz, J. (2020). Virtual Strategy Engineer: Using Artificial Neural Networks for Making Race Strategy Decisions in Circuit Motorsport. Applied Sciences, 10(21), 7805.
9. Tulabandhula, T., & Rudin, C. (2014). Tire changes, fresh air, and yellow flags: challenges in predictive analytics for professional racing. Big data, 2(2), 97-112.
10. Önen, Deniz & Rodrig, İlker. (2021). A HEAT TRANSFER ANALYSIS ON FORMULA 1 BRAKE DISCS.
11. Groote, Jesper. (2021). Overtaking in Formula 1 during the Pirelli era: A driver-level analysis. Journal of Sports Analytics. 7. 1-19. 10.3233/JSA-200466.
12. Vrajitoru, Dana. (2019). Trajectory optimization for car races using genetic algorithms. 85-86. 10.1145/3319619.3326792.
13. Düring, B., Fischer, M., & Wolfram, M.-T. (2022). An Elotype rating model for players and teams of variable strength. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 380(2224). doi:10.1098/rsta.2021.0155
14. Bertrand, Q., Czarnecki, W. M., & Gidel, G. (2022). On the Limitations of Elo: Real-World Games, are Transitive, not Additive. doi:10.48550/ARXIV.2206.12301
15. Eichenberger, R., & Stadelmann, D. (2009). Who is the best formula 1 driver? an economic approach to evaluating talent. Economic Analysis and Policy, 39(3), 389– 406. https://doi.org/10.1016/s0313-5926(09)50035-5
16. Bekker, J., & Lotz, W. (2009). Planning Formula One race strategies using discrete-event simulation. Journal of the Operational Research Society, 60(7), 952-961.
17. Rockerbie, D. W., & Easton, S. T. (2022). Race to the podium: Separating and conjoining the car and driver in F1 racing. Applied Economics, 1-14.

18. Gutiérrez, E., & Lozano, S. (2020). Benchmarking Formula One auto racing circuits: a two stage DEA approach. Operational Research, 20(4), 2059-2083.