Visual Analytics

Course Project

MSc in Business Administration and Data Science

Student ID	Student name
S143003	Diana Laura Janikowski
S158386	Jenő Tóth
S137041	Rasmus Kongstad Thomsen
S158793	S M Ahasanul Karim

Number of characters: 33,766

Submission date: 12. February 2023

Table of contents

Table of Figures

Abstract

Over the past 72 years, Formula 1 utilized 76 different circuits all over the world. In the 2022 season, there were 22 races, while in the 2023 race there will be 24 organized. Our research looks at all the tracks that have been used since 1996 to determine which of them are the most exciting. We look at 6 variables for each race: top speeds, finish time differences, different finishes compared to seasonal results, pit stops, crashes, and overtakings. Based on these, we generate scores for each race, which are then averaged for each circuit, allowing us to compare different tracks. Our analysis concludes with a recommendation to the Fédération Internationale de l'Automobile about the inclusion of the Korea International Circuit from 2024 on. We also provided an interactive dashboard where users can adjust the variables to observe different scenarios. The dashboard can be accessed at https://public.tableau.com/app/profile/jeno.toth/viz/Formula1CircuitExcitementAnalysis/Overview#1.

Introduction

The focus of our research described in this report is Formula 1 (F1) which is considered to be “the world’s most prestigious motor racing competition” (Formula 1, 2023a). Formula 1 teams comprise a number of professional staff, such as drivers, car engineers, and pit stop crew. (Statathlon, 2018). Engineers compete to build the fastest cars, while drivers race on a number of tracks each year. Formula 1 cars are only used in F1 and the teams build new cars for each season in order to gain an advantage. Formula 1 currently consists of ten teams with two drivers each and each season runs from March until the end of November. The number of races has increased overtime, last year consisted of 22 races whereas 2023 will feature 24 of them (Formula 1, 2023b; Formula 1, 2023c).

The first Formula 1 championships took place in 1950 and the dataset we used in our research features historical Formula 1 racing records gathered over the last 72 years since the first

championships (Ergast, 2023; Formula 1, 2023). The dataset contains information on each race including the venue, the results, lap times, pit stops, and more.

A lot of money is involved in Formula 1 and there are high costs of running a team and therefore they have to retain sponsors to be able to compete (Formula 1, 2022). Part of the reason they are able to obtain the huge amount of money through sponsors can be found in the sport's viewership (Formula 1, 2022). Formula 1 reported accumulated 1,55 billion TV viewers across the 2021 season with 445 million unique viewers tuning in to at least one race (Formula 1, 2022). They also reported that the total live attendance across all the Grands Prix accumulated to 2,69 million (Formula 1, 2022). There are a number of stakeholders whose aim, among others, is to create excitement for viewers. Our research focuses on one aspect of this excitement. In this paper, we will analyze the last 72 years' Formula 1 data to determine which circuits are the most exciting, finishing off with a recommendation for which track should be reintroduced to the competition.

Literature Review

Before creating any kind of visualization, it is important to look for guidelines, principles, and methods to make sure that our dashboards are as effective as possible. Midway (2020) lays down 10 principles for effective data visualization. Some of them are technical, such as which software to use, while others are concerned with the sizes, colors, and shapes of the visuals. These principles are not advised for a specific subject, instead they serve as a guideline for any scientific paper which intends to include visuals. As the target audience of the article is broad, the principles are more foundational rather than in-depth, and not all of them were useful for our specific tasks (for example, Principle 5 concerns uncertainty and confidence intervals, which were not a part of our research). Despite this, a number of the principles were kept in mind during our research, such as the meaning of colors, and asking for outside opinion.

Janes et al. (2013) distinguishes between useful dashboards and dashboards that are designed to fulfill other purposes, such as displaying as much information as possible and demonstrating graphical abilities. The authors distinguish two steps when designing a useful, efficient dashboard: choosing the right data, and choosing the right visualizations. When choosing the right data, they distinguish three steps, all of which should be contemplated. These are the goal of the study; the questions, which delimit the object of the study, and the measures, which look at how to answer the question in a quantitative way. For the second step, the paper argues that there is no single right visualization to choose. Instead, a number of considerations should be taken to make the dashboard useful for the user. These include helping the user understand both the context and the meaning of the data.

Bach et al. (2022) describe in detail the different dashboard design patterns, their uses and shortcomings. They distinguish between eight groups of patterns. These can be categorized as Content Dashboard Design Patterns and Composition Dashboard Design Patterns. The first group includes Data Information (how values are visualized), Meta Data (what metadata is included), and Visual Representation (what charts, graphs are used to visualize). The second group includes Page Layout, Screenspace (for example whether there are multiple pages), Structure (the structure of information across pages), Interaction, and Color. Based on these patterns, the authors give a number of guidelines so that the dashboard will not be overwhelming to the user, while conveying as much information as possible.

In the fifth chapter of his book, Edward Tufte (1990) describes how to use colors in visuals. He describes that the primary goal of using color in information design is to "do no harm". This means avoiding excessive use of colors, using them in a non-concise way, and in ways difficult to interpret, such as using light colors and white next to each other. As such, we used colors sparsely, mainly on the race overview dashboard. Here, colors served two purposes. It helped distinguish between drivers, and it also conveyed further information about which team each driver is in.

Despite our limited use of colors, we still had to determine which color to use throughout our dashboards. Knaflic (2015) described a number of criteria when choosing a color. Due to accessibility and to attract attention on our white backgrounds, we used blue colors to highlight important parts. We used a gradient from light blue to dark blue on our circuit locations map. When making a map with gradients, there are a number of shortcomings we had to avoid. Crameri et al. (2020) describes poor use of color gradients, and gives recommendations on how to use colors so that they are accessible and reduce complexity. The article focuses on two of the most used gradients: rainbow color maps and red-green pairs. It highlights their shortcomings, such as rainbow maps having the most attractive color (yellow) in the middle, and red-green maps being inaccessible. They also establish that multi-color gradients are often difficult to interpret.

We designed our dashboards to be user-interactive. In the overview pages, users can select the year, as well as the race. In our interactive analysis dashboard, they can use 6 sliders to set parameters for later calculations. It is interesting to look at what makes an interactive dashboard like ours more engaging for the users. Boy et al. (2015) tests in a number of web-based experiments whether a narrative displayed before the user interactions increases engagement. Their hypothesis, based on previous research, was that it does. They tested both storytelling, and no-storytelling versions of the same visualizations. They measured the number of interactions a user performs in different scenarios. Their results did not confirm their hypothesis. These results were not significant, however, therefore their conclusion stated that narratives do not increase user engagement, but do not affect results significantly negatively. When creating our dashboards, we decided only to include texts about the general purpose, and brief descriptions about the variables and our assumptions. Perhaps the most influential narrative was the predetermined parameters. These were based on our assumptions described in this report, and while users can freely change them, they may be victims of the anchoring bias (Tversky and Kahneman, 1974).

Data Preparation

We used data generated by the Ergast Developer API (Ergast, 2023). The dataset can also be found on Kaggle (2022). Kaggle being a public space makes the dataset widely available to other users as well (Kaggle, 2023). However, despite the number of available analyses using the dataset we were not able to find any that focused on the same research direction, and therefore we deemed the dataset to be a good foundation for dashboard creation and analysis despite it already being used by other users. The following entity relationship diagram provides an overview of the data:

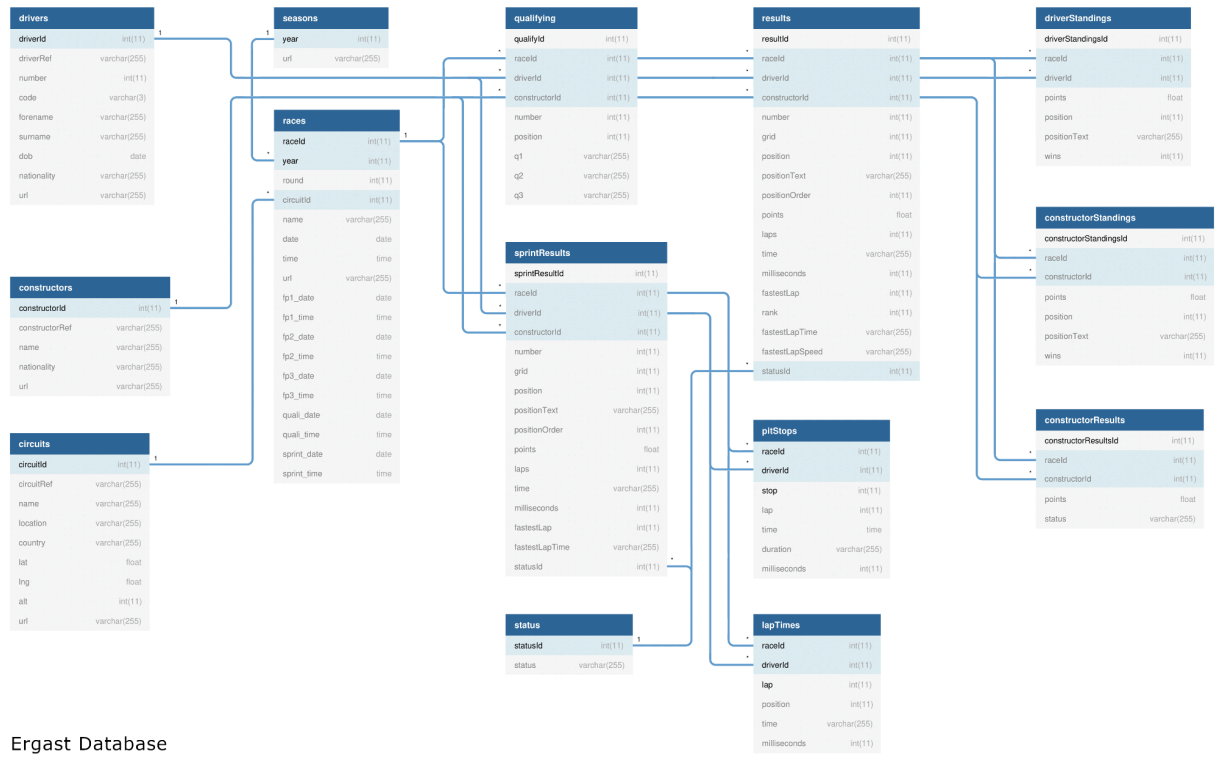

Figure 1. Ergast Database ER diagram (Ergast, 2023)

Our last access to the data was on 30 December 2022, by this time it contained all relevant information for the whole of the 2022 season. The data is built up from 14 datasets, these are described in more detail in the Appendix. While there was no missing data within seasons, there were a number of points which were not available in earlier years. For our purpose, the most important ones from this category were lap times, top speeds, and pit stops. For the creation of

new variables, we used the software Alteryx. The following table contains the most important of these new variables.

Table	New variable	Description
circuits	count	How many times a circuit has been used for an F1 race.
circuits	latest_year	When was the last F1 race held at the circuit.
lap_times	total_time	The total time it took for a driver to complete X laps in a race.
lap_times	takeover_number	If the position of a driver in lap X-1 was M, and in lap X it decreased to N (meaning he overtook at least 1 person), then the takeover_number for lap X is M-N.
lap_times	takeover_weight	Takeovers at the top of the table are more exciting than those happening at the bottom. If the driver took over someone such as in the example above, the takeover weight for that turn is the number of total drivers divided by the square root of N.
lap_times	lap_progress	For every minute it estimates how many laps the driver completed, based on the known number of laps completed and what fraction of the seconds have elapsed in the next lap.
races	diff_1_2	The time difference between the first and second places' finishes.
races	diff_2_3	The time difference between the second and third places' finishes.
races	crashes	Counts how many times the following statuses appeared in each race: "Accident", "Collision", "Spun off", "Broken wing", "Heat shield fire", "Engine fire", "Fire", "Injured", "Injury", "Fatal Accident", "Collision damage", "Damage", "Debris"
results	finish_diff	For each driver, it looks at the driver's finish position in the race, as well as his final position that season. It takes the average of all the absolute values of the drivers in each race.

Table 1. Most Important Variables Calculated for Excitement Estimation

Data Analysis

For our analyses, we determined the six most important factors for excitement to be top speed, overtakings, the number of pit stops, the number of crashes, the difference between the drivers' final positions in a race compared to their seasonal finishes, and the time differences between first, second, and third places. We gathered these for every race, as well as the averages for every season. There were a number of missing values: lap times were only available from 1996, top speeds from 2004, and pit stops from 2011. Because of this, we decided only to include races from 1996 and onwards in our calculations. There were also large fluctuations between years, due to the constant evolution of the sport. As such, we compared the values within seasons.

For every race, we used the following data for our analysis: "Average top speeds of the drivers", "Finish distance between first and second", "Finish distance between second and third", "Overtaking weight", "Finish position difference compared to season", "Number of pit stops", "Average length of pit stops", and "Number of crashes". For each year, for each variable, we normalized these data with the following formula: $\frac{Max-x}{Max-Min}$, where x is the specific value, Max is the maximum of such values in a given year, and Min is the minimum. With this normalization, all of our values were numbers between 0 and 1, with 0 being the lowest in each season, while 1 is the highest. Our hypothesis is that the larger each of these variables are, the more exciting a race is, therefore we summed them up in our next steps.

We added "Finish distance between first and second" and "Finish distance between second and third" together such that 66% of the value came from the first variable. Similarly, we added "Number of pit stops" and "Average length of pit stops" with the first value accounting for 80% of the sum.

We designed our algorithm such that it takes in weights for all 6 of the created variables, and for every race it calculates an excitement score:

$\sum_{i=1}^{6} \frac{W_i * i}{W_t} * 100$, where i are the variables, W_i is the weight of the variable and W_t is the total weight. After this, we grouped all the races by the circuits they were held on, and took average.

With all of this, we had an excitement score for every circuit that has been used since 1996. Due to the normalization process, this score should not be interpreted as a number, instead, it should serve as a measure to compare how exciting a racetrack is. Because of the nature of our calculations, races before 2010 will have lower weights as at least one variable (pit stops, and perhaps top speed) will always be zero there. Initially, we planned on using weights for the years when a race occurred, as we believe newer races should account for more, since they represent the current Formula 1 environment better. Due to these missing observations, we decided to not include the weights, as the numbers for earlier races will be slightly lower anyways.

In our dashboards users can freely adjust the weights to see what are the results of different models. For this analysis, we used estimates from Motorsport Network (2017) to determine what weights are optimal. The weights associated were the following: “Top speed”: 90, “Finish time difference”: 85, “Overtakings”: 82, “Pit stops”: 75, “Different finishes compared to season”: 65, “Crashes”: 60. The chart showing displaying the fan surveys from 2017 depicting the key features of Formula 1, is shown below:

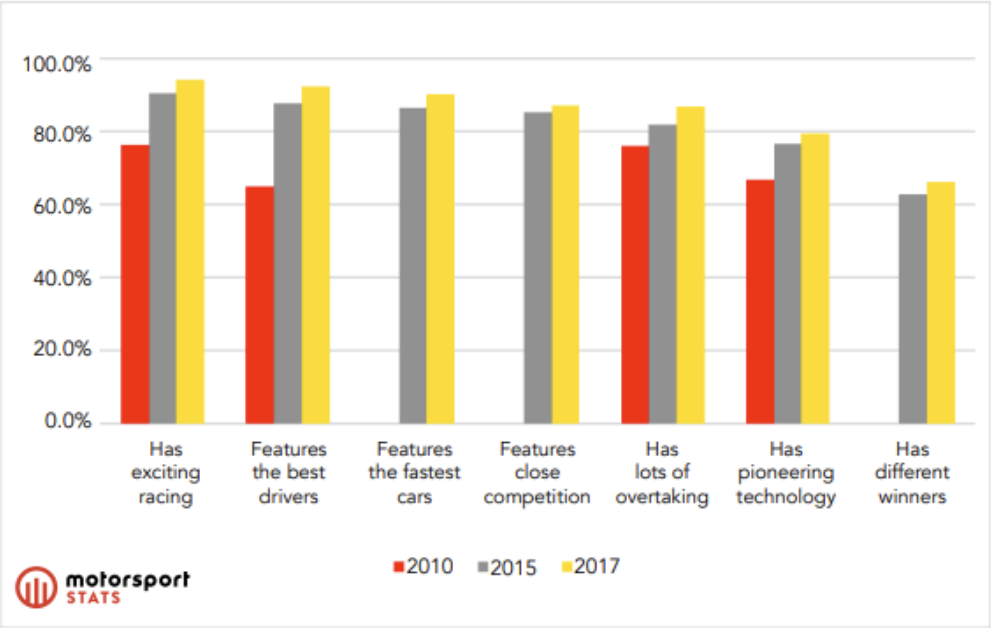

Figure 2. Key features of Formula 1 (Motorsport, 2017)

At this point, we had a dataset of 40 circuits with each of their excitement scores. While our data analysis was good to provide results, we had to look at each of the tracks individually for feasibility. While the most exciting track according to our model was the Losail International Circuit, there was only one F1 race held there, therefore we cannot say that it is not an outlier.

Our pick for the most exciting circuit currently not in use was therefore the Korea International Circuit. The track has been used 4 times in the past, most recently in 2013. It is also a feasible spot to organize a race at (compared to for example another exciting track, the Sochi Autodrom in Russia). The Korea International Circuit is South Korea's first and only international racing track meeting FIA's Grade A standard (Korea Circuit, 2023). It was designed by Hermann Tilke, a German civil engineer who has over the span of 40 years designed many other circuits (Korea Circuit, 2023; Porsche, 2023; Tilke, 2023). It is divided into a 5.615 km long track which includes streets as well as a 3.045 km long permanent track (Korea Circuit, 2023). It is located in the country's tourism, leisure and business development hotspot which further adds value to the attractiveness of our suggestion besides its excitement level (Korea Circuit, 2023).

Investigating the different tracks across the previously mentioned measurements the data show that the Korean Grand Prix scores higher on average across all categories. When looking at the average top speed there are only small differences across the tracks, but when it comes to the number of overtakes and the number of pit stops, the Korean Grand Prix had higher averages than most of the other tracks, and therefore it is not unsurprising that it has one of the highest excitement scores.

We also looked at the least exciting tracks that are currently being used. The least exciting according to our model is the Circuit de Monaco. This makes sense, as the circuit has narrow tracks, not giving space for overtakings (Formula1, 2023d). It also has short runways, resulting in low top speeds. Despite this, it is perhaps the most iconic Formula 1 track, and therefore it makes up in other ways in excitement. Because of these factors which cannot be measured by our model, we decided to propose the inclusion of the Korean Grand Prix not as a substitute for an existing track, but rather to further increase the number of races held. The following table shows the rest of the top 10 results of our analysis:

Rank	Circuit Name	Excitement score	Number of races hosted	Latest race's year
1	Losail International Circuit	50.4	1	2021
2	Baku City Circuit	47.5	6	2022
3	Jeddah Corniche Circuit	46.3	2	2022
4	Circuit de Spa-Francorchamps	46.3	55	2022
5	Korea International Circuit	46.2	4	2013
6	Autodromo Internazionale del Mugello	45.1	1	2020
7	Fuji Speedway	40.9	4	2008
8	Autodromo Nazionale di Monza	40.8	72	2022
9	Shanghai International Circuit	40.4	16	2019
10	Sepang International Circuit	40.3	19	2017

Table 2. The top 10 most exciting circuits according to our model

To test our model, we looked at a qualitative list of tracks by WhichCar (2021). The following table contains their top 10 picks and what our model predicted for the most exciting races. The two ranks do not share much in common. One of the attributes that could explain this is that rain was mentioned multiple times as a factor for excitement, however it was missing from our analysis.

Circuit name	WhichCar place	Our predicted place
Bahrain International Circuit	10	5
Hungaroring	9	18
Monza Circuit	8	3
Algarve International Circuit	7	7
Istanbul Park	6	12
Suzuka International Racing Course	5	8
Baku City Circuit	4	1
Red Bull Ring	3	17
Circuit de Spa-Francorchamps	2	2
Interlagos	1	11

Table 3. Comparison of Predicted Results with WhichCar (2021)

Shortcomings

There are a number of shortcomings that we must acknowledge in our analysis. These come from the limits of the data we used. Perhaps the biggest estimation we took was regarding overtakings. Because we only had data for each lap, but not for within laps, we could not see if there were perhaps two drivers who overtook each other back and forth within a lap. Despite this, our hypothesis is that there is a large correlation between overtakings between and within laps, as the position of the start gate does not influence where these overtakings take place.

While we had a lower number of overtakings this way, it is acceptable, as we were only interested in the ratios between tracks, which should not be significantly influenced by this.

Another error that we could not calculate due to the lack of data was weather. A rainy race is fundamentally different from a dry one, featuring more crashes and perhaps more excitement. This can influence our results, as tracks with more frequent rains could be estimated to be more exciting, and outliers are also harder to spot. Finally, we lacked sufficient information to determine what weights should be given for the parameters. While our estimates resulted in different results from experts' opinions, we believe that with further data collection about the weather, and with more research about weight assignments our model can be optimized.

Dashboards

Target audience

Our target audience is the leadership of FIA (Fédération Internationale de l'Automobile), the organization behind Formula 1 and the team responsible for selecting the venues (FIA, 2023). Though FIA is driving a sport it's also an organization with the goal of earning money. Their earnings come mainly through sponsors and advertisers which in turn comes from the potential audience of F1. There is a lot an organization can do to expand or streamline to increase their earnings, but we will assume that one of the priorities is increasing the audience and that is what we have focused on.

While our analysis above provided a conclusive answer, we wanted the stakeholders to be able to interact with our model and use their knowledge to tweak it how they see fit. Because of this, we created an interactive dashboard where they can both explore the statistics of each season and each race over the past decades, as well as input the weights for our 6 variables themselves.

We have four pages to our dashboard, the front page with facts regarding where the races have been held and the main stats on all time Formula 1. From here, users can navigate to three other pages. The first page is about the individual seasons where the user selects a given year, and can then see stats regarding the drivers and constructors for that season. Then we have a page about the individual races, where the user can select a race, and then see the statistics related to the race and an animation of the progression of the race. Finally, we have the final page where the user can see and track the excitement score of each circuit and use the filters to change the weights of the different things that make racing exciting.

Dashboard walkthrough

When opening the dashboard, the user is met with an introductory front page with general information regarding Formula 1. Here, there is a short introduction to the problem and a description of the analysis intended with the dashboard. Directly beneath, we have placed a table showing all time statistics about Formula 1 such as the unique number of tracks, the unique number of teams and drivers throughout history. Beneath this, the user has the navigation for the report and can switch to the three other pages with navigation buttons. On the right side of the dashboard the user will see an interactive map with locations of circuits that have hosted a Formula 1 Grand Prix. Hovering over the dots the user will see a tooltip showing the name of the circuit and the number of times it has hosted a Grand Prix. This is also illustrated by the gradient of the color as the darker the dot the more times it has hosted. To complement the map, the user will see a table to the right showing the different circuits and the number of times it has been used, sorted from most to least. In the bottom of the page, we have placed a table showing statistics on the used circuits. This includes the average top-speed, the average number of pit-stops, and the average number of overtakes. These statistics are all factors in what makes a race exciting, and they will be used more in depth on the Interactive Excitement Analysis dashboard as part of the analysis.

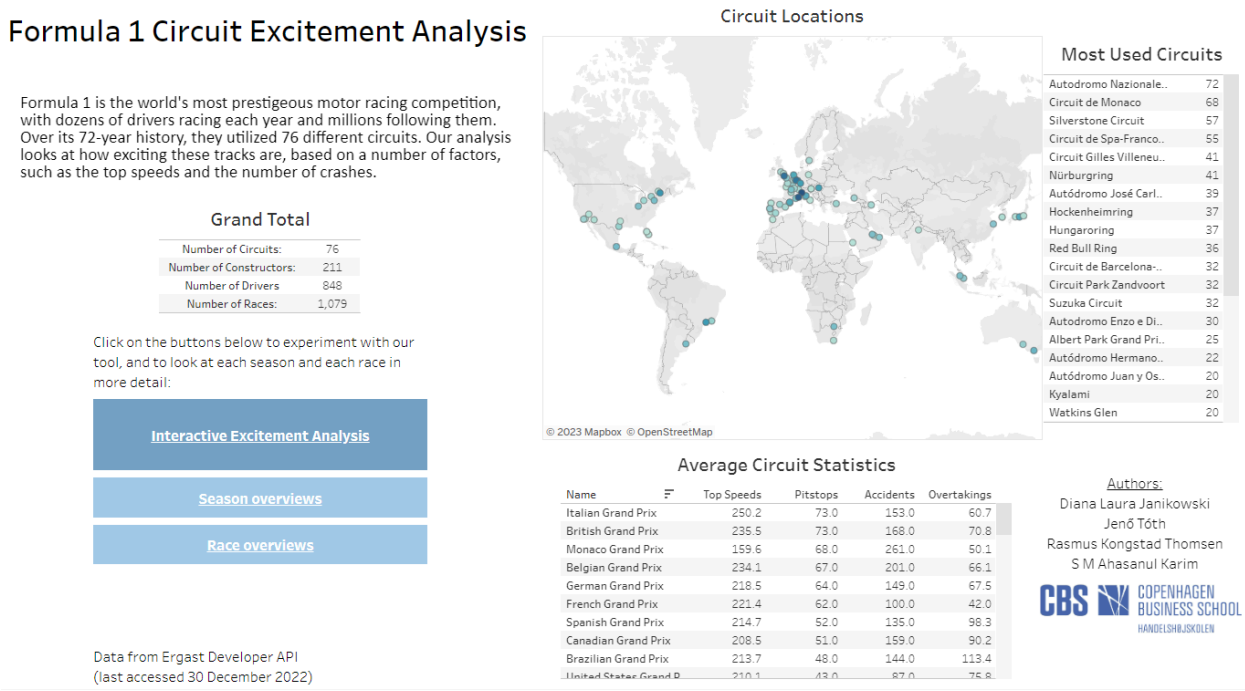

Figure 3. Front page of the dashboard

When the user navigates to the Race overviews page, they will see a table displaying race statistics for the selected race compared to the season averages. This includes the number of laps, the average top-speed, the number of overtakes, number of pit-stops, and the difference in total time finish between the first and second place. To the right of the table, the user will see the outline of the currently selected track. Right next to this, the user has the option to do some filtering to select any race since 1996. The first filtering option that the user has is the Year filter. This is a single select filter as the user can only see one race at a time in the visuals, and it therefore wouldn't make sense to select more than one year. The next filtering option is the Circuit Name filter which like the year filter is a single select filter as the user only can see one race at a time. After picking a race, the user can start an animation which will show all of the drivers' progress throughout the race. The animation can be played and paused by the user, and they can also select any time during the race via a slider.

The bottom part of the dashboard is a visual showing the development of the drivers through the selected race. When the animation is running the visual will display the race from start to finish with the drivers ending up in their finishing positions. In the top right corner, the user has the option to navigate back to the front page with a navigation button.

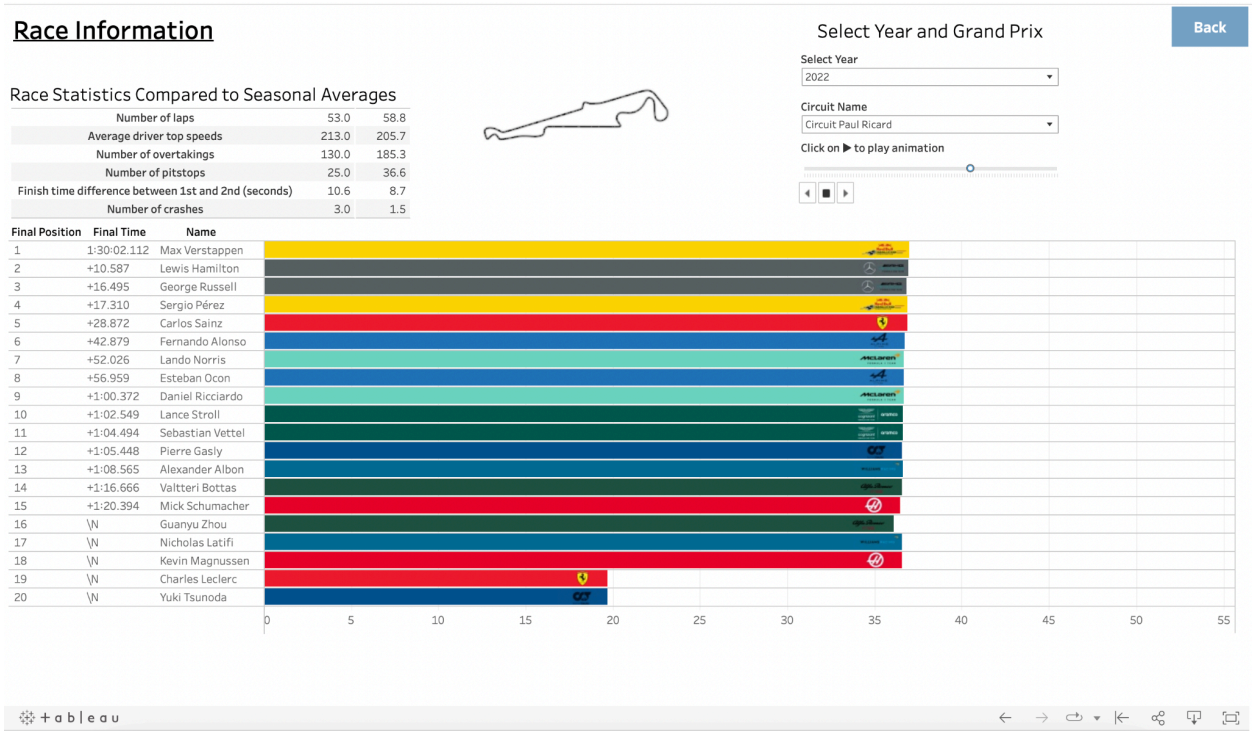

Figure 4. Race Overview page of the dashboard

When the user navigates to the Season overviews page, they will see a table displaying the average race statistics for the given season. This includes the average difference in total time finish between the first and second place, the average number of overtakes and pit-stops, and the average top-speed. In the top of the page, the user has the option to filter the data to get the desired outcome. For this dashboard there is one filter, and this gives the user the option to choose a given year. This is a single select filter as the page displays information per season, and it therefore wouldn't make sense to select more than one year at a time. To the right of the table, the user will see two tables displaying the most up to date standings for the drivers'

championship and the constructors' championship. These are sorted from most to least points gathered over the season. In the bottom of the dashboard, we have a visual showing the development in the driver standings throughout the selected season, with each line representing a different driver. The line color explanation can be found right next to the visual, and in case the user can't differentiate between them, the tooltip for the graph will also give the desired information. In the bottom right corner, the user can find a table showing the selected season's race winners, displaying all the names of the tracks, and the winning driver and their team. As with the other pages, the user has the option to navigate back to the front page with a navigation button in the top right corner.

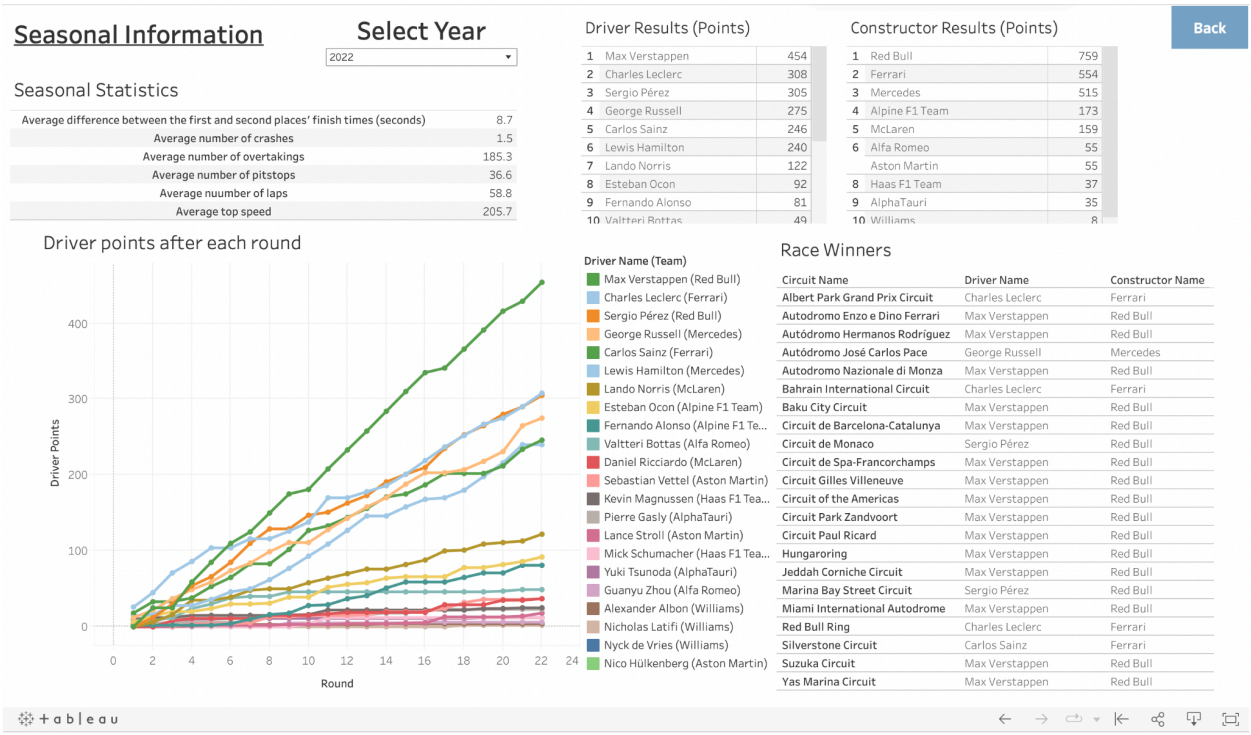

Figure 5. Seasons Overview page of the dashboard

When the user navigates to the Interactive Excitement Analysis page, they will see a filtering selection which is essential to generate any meaning in the tables. The filtering section is prefaced with an explanation as to why the following selections have been included. Then, the user can change the filters based on their beliefs to generate the most and least exciting races.

There are six filters the user has to take into account, which is the top-speed, the finish time difference, the different finish, the overtakes, the pit-stops, and the crashes. All have to be in the interval of 0 – 100 and have a short explanation to make the user experience better. The preset filtering has been determined based on a survey from Motorsport as previously described (Motorsport, 2017).

When the filtering has been set, the user can see a table showing the excitement score for all the tracks sorted from most to least exciting. On the right side of the dashboard the user can see two tables. One table is showing the top five most exciting tracks not used in the 2022 season, which has previously held at least three races. Another table is showing the five least exciting tracks used in the 2022 season.

This is followed by a disclaimer saying that only races after 1995 count towards the excitement score due to missing data. As with the other pages, the user has the option to navigate back to the front page with a navigation button in the top right corner.

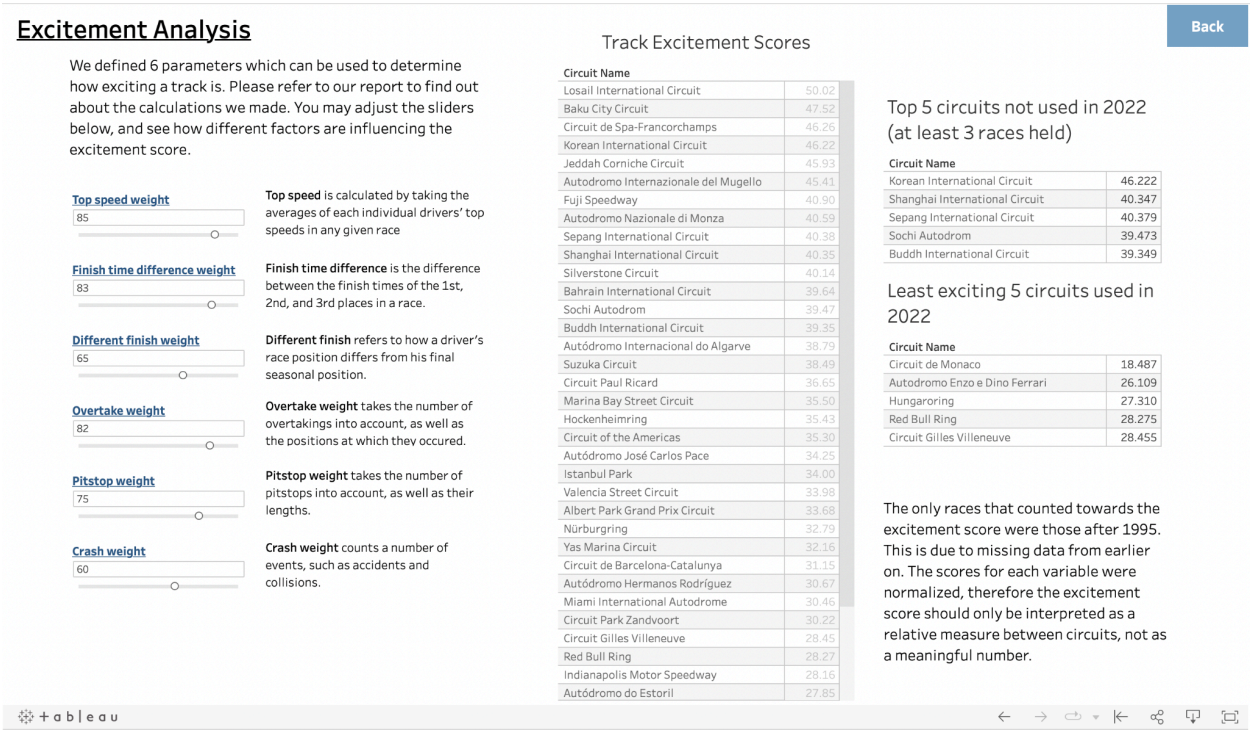

Figure 6. Interactive Excitement Analysis page of the dashboard

Dashboard objects

The focus of the design of the dashboard is to effectively provide insights into Formula 1 for our target group based on the underlying data. The dashboard should support the insights that it provides and make it possible for the user to make valuable data-driven decisions. A best practice rule regarding dashboards is to keep it simple to not disturb the point of the report and therefore we have taken steps to remove unnecessary clutter and noise in the dashboards.

When building dashboards there are three main points one needs to take into consideration to ensure that the dashboard is accessible to all the users regardless of physical or visual impairments. The three main points to ensure can be found in the Web Content Accessibility Guidelines and they are that the data presented are 1. perceivable, 2. operable, and 3. understandable (W3C, 2008). To accommodate some of these points we have for instance chosen the bold font for the KPIs.

We have kept a consistent format across the pages to ensure a comfortable viewing experience and to make the report easy to use. On each page the placements for the tables and visualizations have been chosen in accordance with the zigzag motion that most users tend to follow when looking across the screen (Knaflic, 2015). The zigzag involves the movement of the eyes from the upper left corner to the upper right corner, followed by eye movement to the lower left corner and the lower right corner in the end. In the same regard we have placed all filtering in the same place on the right side across all the pages.

Color choices

While designing our dashboards we used colors sparingly and therefore all pages have a neutral white background with all text including tables in a dark gray color. Across all pages we have used a bold font to highlight all the KPIs. Whenever we used colors, they were used in order to

attract attention (Knaflic, 2015). Additionally, we avoided excessive use of green and red as they belong to the most frequent color blindness spectrum (Knaflic, 2015).

On the front page we have a map with the locations of the circuits and for this we have chosen a map with the countries in a light gray color and the dots to represent the places in blue. We have chosen the color blue as it is the brand color of FIA. According to Knaflic (2015), creating color palettes for branding can be a major undertaking for a company and might be required to be used, and therefore we have decided to incorporate the color in order to align our visualizations with FIA's branding. Besides the map, blue has been also used for buttons in our dashboards.

On the dashboard page related to the seasons we have created a line chart showing the development in points for each driver throughout the season. To be able to distinguish the different drivers lines they have randomly been assigned a color. To ensure all users can utilize the dashboard this has been made up for by adding the relevant information as a tooltip to the lines.

On the dashboard page related to the individual races we have created a bar chart showing the development for the race finishes. To ease the use of the chart the bars are not only in the color of the drivers' respective teams but also showing their team logo to be able to distinguish between them. For the excitement page of the dashboard, we have a bar chart showing the excitement level of each race circuit and for this we have chosen the same blue color as for the map.

Accessibility

To ensure that the users of the dashboard can utilize it properly we need to consider the format of the report. When building a dashboard in Tableau there are the options to build it for desktops, tablets, or smartphones. As all these have very different formats, we have to choose between them and based on our target group, we have chosen to build the dashboard in the desktop format. This covers a large portion of the most used devices in a company setting as it

can be used perfectly for the following three formats, Generic Desktop (1366 x 768), Large Desktop Monitor (2560 x 1440), and iPad Pro 12.9-inch (1366 x 1024). This also means that the dashboard has not been optimized for tablets or smartphones and the user will therefore have to scroll when using the dashboard on these devices. The report was built with the most common reading flow in mind as it should be read from left to right and from top to bottom. To guarantee that all users from our target group can use the report, it has been published to Tableau Public and it can therefore be accessed through a single weblink.

Conclusion

For a number of decades now, Formula 1 has been perhaps the most influential motor racing competition in the world. There are a number of reasons behind this, such as the constant technological advancements and the drama that surrounds each season. One of these factors are the tracks that are constantly changing, bringing new challenges for teams. Our analysis looked at the complete history of F1, with a focus on data since 1996, to analyze which of these circuits are the most exciting. We looked at a number of variables which could make a track exciting, such as how close a race is and how fast drivers go. Based on our analysis, our recommendation to the FIA is to once again include the Korea International Circuit. While our calculations were in depth, we are aware that different stakeholders may hold different views about what makes a track exciting. As such, we provided an interactive dashboard, where users can adjust weights to their likings and look at the outcomes for different scenarios. Besides this, our dashboard also contains overviews of all the past 72 years of Formula 1, as well as all of the individual races since 1996.

References

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., Chen, M. (2020). Dashboard Design Patterns. arXiv.

Boy, J., Detienne, F., & Fekete, J.-D. (2015). Storytelling in Information Visualizations. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15: CHI Conference on Human Factors in Computing Systems. ACM.

Crameri, F., Shephard, G. E., & Heron, P. J. (2020). The misuse of colour in science communication. In Nature Communications (Vol. 11, Issue 1). Springer Science and Business Media LLC.

Edward Tufte, Envisioning Information. Chapter 5: Color and Information. Graphics Press, 1990.

Ergast. (2023). *Ergast Developer API*. Ergast.com. Retrieved January 5, 2023, from http://ergast.com/mrd/

FIA. (2023). *Organisation | Federation Internationale de l'Automobile*. FIA. Retrieved January 5, 2023, from https://www.fia.com/organisation

Formula 1. (2022). *Formula 1 announces TV, Race attendance and digital audience figures for 2021 | Formula 1®*. F1. Retrieved January 5, 2023, from https://www.formula1.com/en/latest/article.formula-1-announces-tv-race-attendance-and-digital-audience-figures-for-2021.1YDpVJIOHGNuok907sWcKW.html

Formula 1. (2023a). *About F1 | Formula One World Championship Limited*. Formula One World Championship Limited. Retrieved January 5, 2023, from https://corp.formula1.com/about-f1/

Formula 1. (2023b). *F1 Schedule 2022 - Official Calendar of Grand Prix Races*. F1. Retrieved January 5, 2023, from https://www.formula1.com/en/racing/2022.html

Formula 1. (2023c). *F1 Schedule 2022 - Official Calendar of Grand Prix Races*. F1. Retrieved January 5, 2023, from https://www.formula1.com/en/racing/2023.html

Formula 1. (2023d). *Monaco Grand Prix - Monte Carlo.* F1. Retrieved January 5, 2023, from https://www.formula1.com/en/information.monaco-circuit-de-monaco-monte-carlo.2ZWRtIcSI6ZzVGX1uGRpkJ.html

Janes, A., Sillitti, A. and Succi, G. (2013). Effective dashboard design. Cutter IT Journal. 26. 17-24.

Kaggle. (2022). *Formula 1 World Championship (1950 - 2022)*. Retrieved January 5, 2023, from https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?select=constructor_standings.csv

Kaggle. (2023). Kaggle: Your Machine Learning and Data Science Community. Retrieved January 5, 2023, from https://www.kaggle.com/

Korea International Circuit. (2023). *About KIC | Overview*. Korea International Circuit. Retrieved January 5, 2023, from https://en.koreacircuit.kr:442/web?site_id=3&menu_id=169&

Knaflic, C. N. (2015). *Storytelling with Data - A Data Visualization Guide for Business Professionals*. Wiley.

Midway, S. R. (2020). Principles of Effective Data Visualization. In Patterns (Vol. 1, Issue 9, p. 100141). Elsevier BV.

Motorsport. (2017). Formula1 in 2017. *Motorsport*. https://cdn-3.motorsport.com/static/documents/Fan_Report_Final.pdf

Porsche. (2023). *What goes into making the perfect racetrack?* Porsche. Retrieved January 5, 2023, from https://www.porsche.com/stories/design/hermann-tilke-how-to-design-a-racetrack

Statathlon. (2018). *A Thorough Analysis of the Pit Stop Strategy in Formula 1*. Statathlon. Retrieved January 5, 2023, from https://statathlon.com/analysis-of-the-pit-stop-strategy-in-f1/

Tilke. (2023). *HERMANN TILKE «*. Tilke Engineers & Architects. Retrieved January 5, 2023, from https://tilke.de/en/hermann-tilke/

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, 185(4157):1124–1131.

WhichCar. (2021). *Every current Formula 1 track, ranked*. WhichCar. Retrieved January 5, 2023, from https://www.whichcar.com.au/opinion/every-current-formula-1-track-ranked

W3C, (2008): Web Content Accessibility Guidelines (WCAG) 2.0. Edited by Wendy Chisholm, John Slatin, Jason White, Ben Caldwell, Michael Cooper, Loretta Reid, Gregg Vanderheiden. Retrieved January 5, 2023, from https://www.w3.org/TR/WCAG20/

Appendix

Circuits dataset:

Column name	Description	Sample value
circuitId	Unique ID for every circuit.	1
circuitRef	Reference name of the circuit.	albert_park
name	Name of the circuit.	Albert Park Grand Prix Circuit
location	Area where circuit is located.	Melbourne
country	Country where circuit is located.	Australia
lat	Latitudinal coordinate of the circuit.	-37.8497
lng	Longitudinal coordinate of the circuit.	144.968
alt	How many meters above sea level the circuit lays.	10
url	Link to the circuit's Wikipedia page.	http://en.wikipedia.org/wiki/Melbourne_Grand_Prix_Circuit

Constructor_results dataset

Column name	Description	Sample value
constructorResultsId	Unique ID for constructor results.	186
raceId	Unique ID for races.	36
constructorId	Unique ID for constructors.	1
points	The number of points a constructor obtained in a race.	14
status	Almost completely empty column (only has values for one constructor in one season).	D

Constructor_standings dataset

Column name	Description	Sample value
constructorStandingsId	Unique ID for constructor standings.	60

raceId	Unique ID for races.	23
constructorId	Unique ID for constructors.	10
points	The number of points a constructor obtained in total in a season after a race.	0
position	Constructor’s position in the season after a race.	10
positionText	String of “Position”.	10
wins	Total number of wins in a season for constructor.	0

Constructors dataset

Column name	Description	Sample value
constructorId	Unique ID for constructors.	5
constructorRef	Reference name for constructors	torro_rosso
name	Name of the constructor.	Torro Rosso
nationality	Country of origin for the constructor.	Italian
url	Wikipedia link to constructor’s page.	http://en.wikipedia.org/wiki/Scuderia_Toro_Rosso

Driver_standings dataset

Column name	Description	Sample value
driverStandingsId	Unique ID for driver standings	70
raceId	Unique ID for races.	22
driverId	Unique ID for drivers.	2
points	The number of points a driver accumulated over the season.	20
position	The driver’s final position in a race.	5
positionText	String of “Position”.	5
wins	How many wins the driver accumulated over the season.	0

Drivers dataset

Column name	Description	Sample value
driverId	Unique ID for drivers.	13
driverRef	Driver's name reference	massa
number	Driver's track number.	19
code	3-letter code for driver.	MAS
forename	Driver's first name.	Felipe
surname	Driver's last name.	Massa
dob	Driver's date of birth.	4/25/1981
nationality	Driver's nationality.	Brazilian
url	Link to the Wikipedia article of driver.	http://en.wikipedia.org/wiki/Felipe_Massa

Lap_times dataset

Column name	Description	Sample value
raceId	Unique ID for races.	841
driverId	Unique ID for drivers.	20
lap	The lap number within a race.	5
position	The position of the driver at the end of the lap.	1
time	The time the lap took for the driver.	01:32.3
milliseconds	Number of milliseconds the lap took for the driver.	92342

Pit_stops dataset

Column name	Description	Sample value
raceId	Unique ID for races.	842
driverId	Unique ID for drivers.	808

stop	The stop number for the driver in the race.	1
lap	The lap number within a race.	15
time	The time at which the pit stop occurred.	16:30:10
duration	How long the pit stop took.	23.182
milliseconds	“Duration” in milliseconds.	23182

Qualifying dataset

Column name	Description	Sample value
qualifyId	Unique ID for qualifiers.	2
raceId	Unique ID for races.	18
driverId	Driver’s track number.	9
constructorId	Driver’s track constructors.	2
number	Driver’s track number.	4
position	Grid position of driver based on the qualifiers..	2
q1	Best time in first round of qualifiers.	01:26:1
q2	Best time in second round of qualifiers.	01:25:3
q3	Best time in third round of qualifiers.	01:26:9

Races dataset

Column name	Description	Sample value
raceId	Unique ID for races	1095
year	Year the race was held in	2022
round	The round of the season	21
circuitId	Unique ID for circuits	18
name	The name of the event	Brazilian Grand Prix
date	The date of the race	11/13/2022
time	The time of the race	18:00:00

url	Wikipedia link to the race	http://en.wikipedia.org/wiki/2022_Brazilian_Grand_Prix
fp1_date	Date of the first free practice sessions	11/11/2022
fp1_time	Time of the first free practice sessions	15:30:00
fp2_date	Date of the second free practice sessions	11/12/2022
fp2_time	Time of the second free practice sessions	15:30:00
fp3_date	Date of the third free practice sessions	\N
fp3_time	Time of the third free practice sessions	\N
quali_date	Date of the quail races	11/11/2022
quali_time	Time of the quail races	19:00:00
sprint_date	Date of the sprint races	11/12/2022
sprint_time	Time of the sprint races	19:30:00

Results dataset

Column name	Description	Sample value
resultId	Unique ID for results	1
raceId	Unique ID for races	18
driverId	Unique ID for drivers	1
constructorId	Unique ID for constructors	1
number	Unique number in more modern seasons for drivers	22
grid	Position where driver started the race	1
position	Position of the driver	1
positionText	String format of position	1
positionOrder	Orders for positions	1
points	Amount of achieved points	10
laps	No. of completed laps	58
time	Time taken to complete the race	34:50.6
milliseconds	Time in milliseconds	5690616

fastestLap	The fastest lap no of the driver	39
rank	Rank of the driver in the races	2
fastestLapTime	Time taken to complete the fastest lap	01:27.5
fastestLapSpeed	The speed of the fastest lap	218.3
statusId	Unique Id for statuses	1

Seasons dataset

Column name	Description	Sample value
year	Year the of the season held in	2022
url	Wikipedia url link to the specific season	http://en.wikipedia.org/wiki/2022_Formula_One_World_Championship

Sprint_results dataset

Column name	Description	Sample value
raceId	Unique ID for races	1061
driverId	Unique ID for drivers	830
constructorId	Unique ID for constructors	9
number	Unique number in more modern seasons for drivers	33
grid	Position where driver started the race	2
position	Position of the driver	1
positionText	String format of position	1
positionOrder	Orders for positions	1
points	Amount of achieved points	3
laps	No. of completed laps	17
time	Time taken to complete the race	25:38.4
milliseconds	Time in milliseconds	1538426
fastestLap	The fastest lap no of the driver	14

fastestLapTime	Time taken to complete the fastest lap	01:30.0
statusId	Unique Id for statuses	1

Status dataset

Column name	Description	Sample value
statusId	Unique Id for statuses	4
status	Description of the status of a race	Collision