

FORMULA 1 UNVEILED

Data-Driven Insights and Visual Narratives

TEAM MEMBERS:

- Chaithra Lal Nair
- Shraddha Ramprakash Gupta
- Utkarsh Vinay Bhosekar

Website Link: <https://formula-1-7jab.onrender.com/>

Github Link: <https://github.com/shraddhagupta27/Formula-1>

ABSTRACT:

Our project aims to demystify F1 for newcomers by developing a user-friendly website elucidating the sport's fundamentals through interactive visualizations. By presenting data visually engagingly, we intend to make F1's rich history and complex statistics more approachable, thereby broadening its appeal and fostering a more informed fanbase.

Existing platforms often cater to seasoned enthusiasts, offering detailed analyses that may overwhelm beginners. Our approach focuses on simplicity and engagement, utilizing visual tools to bridge the knowledge gap and invite a wider audience into the world of F1 racing.

INTRODUCTION:

Formula 1 (F1) stands as the pinnacle of motorsport, captivating audiences worldwide with its blend of cutting-edge technology, exceptional driving skills, and strategic depth. The sport's intricate dynamics generate vast amounts of data, encompassing race results, driver statistics, team performances, and circuit information. Effectively visualizing this data is crucial for uncovering patterns, trends, and insights that might otherwise remain obscured.

Data visualization serves as a bridge between complex datasets and comprehensible insights, enabling enthusiasts and analysts to explore and understand multifaceted information intuitively. In the context of F1, visualizations can reveal the performance trajectories of drivers and teams, the evolution of race strategies, and the global expansion of the sport over time. For newcomers to F1, these visual tools demystify the sport's complexities, making it more accessible and engaging.

The existing visualizations, while informative, may lack interactive elements that facilitate deeper exploration, and the use of complex dashboards can be overwhelming for beginners.

Recognizing these gaps, our project aims to create an interactive platform tailored specifically for newcomers to Formula 1. Our objectives include:

- **Simplifying Complex Data:** Presenting F1 data in an accessible manner that allows users to grasp fundamental aspects of the sport without prior extensive knowledge.
- **Interactive Learning:** Incorporating interactive visualizations that enable users to engage with the data actively, fostering a more profound understanding through exploration.
- **Highlighting Trends and Facts:** Showcasing historical trends, interesting facts, and key milestones in F1 to provide context and enrich the learning experience.

By focusing on these objectives, our goal is to create an engaging website that offers clear explanations of F1 fundamentals, showcases trends over time, shares intriguing facts, and incorporates interactive visualizations. This approach aims to make the sport more accessible and enjoyable for new fans, fostering a deeper understanding and appreciation of F1.

PROCESS:

Data Collection and Preparation: Our project utilizes a comprehensive dataset sourced from Kaggle, which compiles historical Formula 1 data from 1950 to 2023. This dataset encompasses various aspects of the sport, including races, drivers, constructors, circuits, and results. To ensure the most current insights, we supplemented this dataset with 2024 data extracted directly from the official Formula 1 website.

The raw data comprised 14 CSV files, each detailing different facets of Formula 1. For our analysis, we focused on specific files:

- **circuits.csv:** Information about race circuits.
- **races.csv:** Details of each race event.
- **drivers.csv:** Profiles of drivers.
- **constructors.csv:** Information on constructor teams.
- **results.csv:** Outcomes of races.

Before analysis, we performed several preprocessing steps to ensure data quality and relevance:

1. **Data Cleaning:** Addressed missing values, corrected inconsistencies, and standardized data formats across the datasets.
2. **Data Integration:** Merged related datasets to create a unified view, facilitating comprehensive analysis. For instance, combining race results with driver and constructor information provided a holistic perspective on race outcomes.
3. **Feature Engineering:** Created new variables to enhance analysis. Examples include calculating cumulative race counts for circuits and determining win rates for drivers and constructors.

4. **Data Transformation:** Converted data types and formats as necessary to support specific analyses and visualizations.

These preprocessing steps were crucial in preparing the data for effective visualization and analysis, ensuring that our insights into Formula 1 are both accurate and meaningful.

VISUALIZATION METHODS:

In our project, we employed a variety of visualization techniques to effectively convey the rich and multifaceted data associated with Formula 1 racing. These methods were selected to highlight different aspects of the sport, from historical trends to individual performances, ensuring a comprehensive and engaging presentation for users.

1. Animated Choropleth Map: The animated choropleth map serves as a compelling visualization of Formula 1's global expansion from 1950 to 2020. By shading countries based on the cumulative number of races hosted within specified year ranges, the map provides a clear depiction of the sport's geographical growth. The animation feature allows users to observe the progression over time, illustrating how Formula 1 has evolved from its European origins to become a worldwide phenomenon. As the animation progresses, viewers can see the sport's expansion across different continents, with new countries joining the roster of race hosts. The color-coded shading offers an immediate understanding of each country's involvement, with darker shades indicating a higher number of races hosted. This dynamic visualization not only highlights the historical growth of Formula 1 but also emphasizes its increasing global reach and popularity over the decades.

2. Lollipop Chart: To identify circuits that have hosted the most races, we created a lollipop chart. This chart combines dots and lines to represent the number of races held at each circuit, offering a clear comparison among the top venues. The lollipop chart's design emphasizes individual data points, making it easier to discern differences in race counts across circuits.

3. Treemap: A treemap is a data visualization technique that represents hierarchical data through nested rectangles, where each rectangle's size corresponds to a specific quantitative value, and its color can represent an additional dimension. This method allows for efficient comparison of proportions within the hierarchy, making it particularly useful for displaying large datasets in a compact form. In our project, we utilized a treemap to analyze the contributions of various drivers to the success of their respective constructors in Formula 1. Each constructor is depicted as a larger rectangle, with nested rectangles inside representing individual drivers and their respective win counts. The area of each nested rectangle is proportional to the number of wins achieved by the driver for that constructor, providing a clear visual representation of each driver's impact within the team.

To enhance interactivity, we implemented a ***drill-down feature*** in the treemap. Users can click on a constructor's rectangle to delve deeper into the data, revealing a more detailed treemap that displays the top drivers for that specific constructor. In this drill-down view, the size of each rectangle corresponds to the number of wins the driver secured for the constructor, allowing users to easily compare drivers' contributions within the team. This interactive approach facilitates a comprehensive understanding of team dynamics and individual performances in Formula 1

4. Interactive Line Chart: To analyze the performance of top drivers over recent seasons, we developed an interactive line chart depicting the number of wins per season for the top 10 drivers from 2015 onwards. This time-series visualization allows users to track performance trends, compare drivers across seasons, and identify patterns such as dominance periods or performance improvements.

5. Grouped Bar Chart with Dropdown: A grouped bar chart is an advanced visualization technique used to display comparisons across multiple categories within different groups. Unlike a simple bar chart, it groups bars together based on a shared category, making it ideal for showing relationships within and across groups. In our work, this visualization effectively highlights the performance of Formula 1 constructors by grouping their drivers' contributions side by side. This allows viewers to analyze not only the total performance of constructors but also the distribution of points between their two drivers.

Each of these visualization methods was chosen to align with specific data characteristics and storytelling objectives. By employing a diverse set of visual tools, we aimed to create an engaging and informative platform that caters to both newcomers and seasoned fans of Formula

WEBSITE HOSTING:

The website for our project was hosted on Render, a modern cloud platform that simplifies deployment and scaling of web applications. Render was chosen for its ease of use, seamless integration with GitHub for continuous deployment, and its ability to host both static and dynamic websites efficiently. The platform's support for Flask applications allowed us to deploy our visualization-focused website with minimal configuration effort. Additionally, Render's scalable infrastructure ensures that the website remains responsive and performant, even as we plan to add more features or handle larger datasets in the future.

CHALLENGES AND ITERATIONS:

Dynamic Interactivity in Flask: Unlike Dash, Flask does not natively support dynamic interactivity (e.g., live callbacks). Adding these features would require heavy reliance on JavaScript (e.g., Plotly.js) or frameworks like Flask-SocketIO, increasing complexity. Moving from Dash to Flask meant transitioning from a Python-first framework to one that relies on templates and embedded JavaScript. This required rethinking the implementation of certain interactive features, such as the treemap's drill-down.

Template Integration Challenges: Embedding multiple Plotly visuals (e.g., choropleth map, treemap, and bar chart) into a single HTML template required careful management of layout and JavaScript integration, which limited additional functionality.

Combining Multiple Drill-Down Layers: Adding multiple layers of drill-downs (e.g., country → circuit → constructors) required designing and managing a hierarchical data pipeline, which was complex to implement within the constraints of Flask and Plotly.

Incorporating 2024 Data: One of the challenges faced was the lack of 2024 season data in the dataset, which was essential to make the website more current and relevant. To address this, the 2024 data had to be manually generated and integrated with the existing dataset. However, this process introduced issues such as mismatched column names, missing IDs, and inconsistent data formats, leading to errors during merging and aggregation. Significant effort was required to clean, normalize, and validate the data, ensuring proper alignment across fields like race IDs and constructor details.

Website issue: We have hosted the site on Render using its free plan, which can result in slower loading times. Occasionally, the site may display errors like "Page is not working" or "Bad Gateway" initially. However, after refreshing or reloading the page, the website typically becomes accessible within a few seconds. This issue can be resolved by upgrading to a better plan in the future, which would provide improved performance and reliability.

RESULTS AND INSIGHTS:

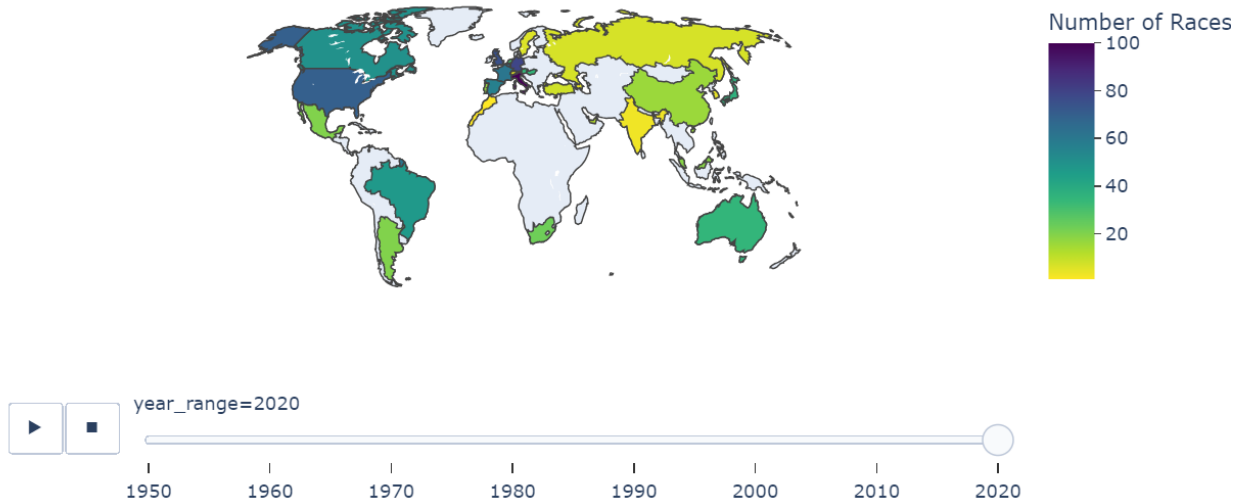
This section highlights the key takeaways and observations derived from the visualizations and analysis performed during the project. The insights provide a deeper understanding of Formula 1's historical trends, circuit preferences, constructor dominance, and driver performance.

- 1) **Expansion of Formula 1 Circuits Over Time:** The animated choropleth map revealed the geographical spread of Formula 1 circuits from the 1950s to 2020. A noticeable trend was the shift from European-dominated circuits in the early years to a more global

presence, with significant expansions in Asia and the Middle East after the 2000s. This reflects F1's strategy to grow its global fan base.

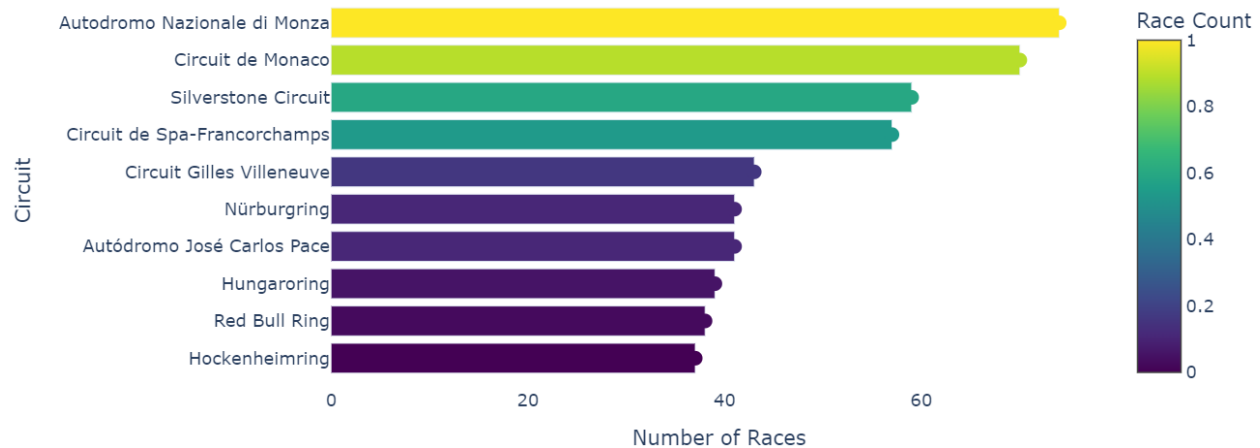


Expansion of Formula 1 Circuits Over Time



- 2) **Top Circuits by Race Count:** The lollipop chart showed the most popular circuits in terms of races hosted. Iconic circuits like "Monza" and "Silverstone" continue to lead in race count, emphasizing their enduring legacy in the sport. This trend also underscores the strong connection between certain circuits and F1's heritage.

Top Circuits by Number of Races (Colorblind-Friendly Lollipop Chart)



- 3) **Top 10 Constructors by Number of Race Wins:** The treemap visualization of the top 10 constructors by race wins reveals notable insights into driver contributions and career trajectories. A significant observation is the career of Lewis Hamilton, who began his Formula 1 journey with McLaren in 2007, securing his first World Championship with them in 2008. In 2013, he transitioned to Mercedes, where he achieved six additional World Championships and numerous race victories.

This career progression is evident in the treemap, where Hamilton's contributions are represented across both McLaren and Mercedes, with the size of the rectangles indicating his substantial impact on each team's success. Hamilton's movement between teams is a prime example of drivers switching allegiances over the years, influencing the distribution of race wins among constructors. The treemap effectively highlights such dynamics, showcasing how individual drivers have bolstered multiple teams' performances throughout their careers. This visualization not only underscores the dominance of certain constructors but also provides a nuanced understanding of how driver movements and contributions shape the competitive landscape of Formula 1.

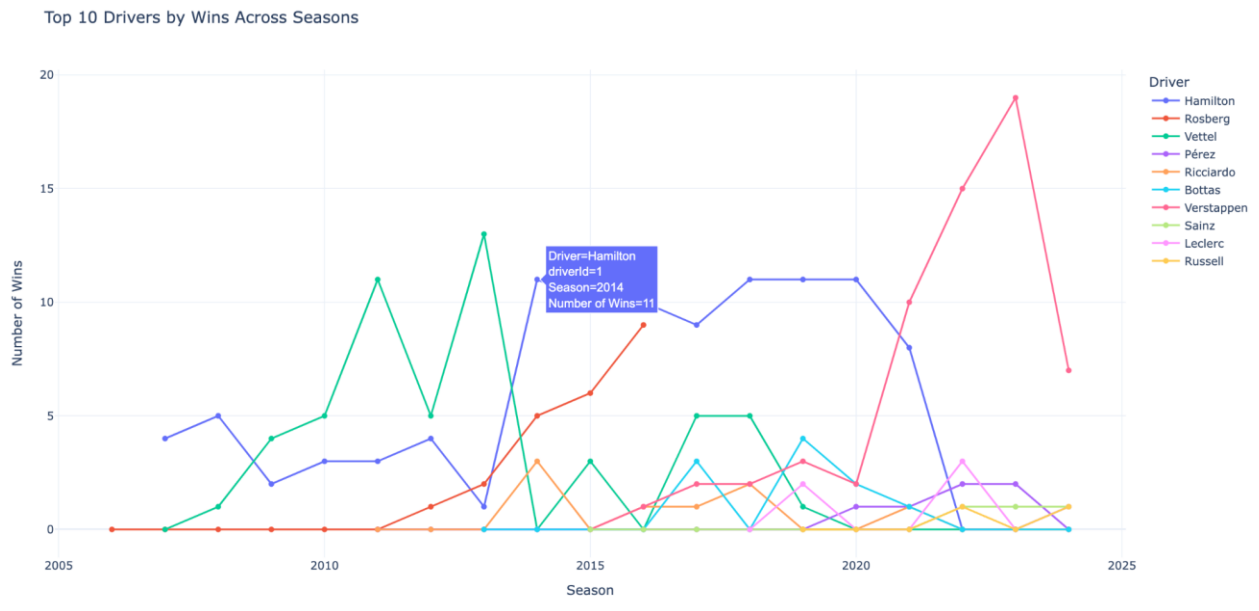
Top 10 Constructors by Wins



4) **Line Graphs:** The line chart depicting the "Top 10 Drivers by Wins Across Seasons" provides a rich narrative of driver dominance and shifting competitive dynamics in Formula 1 over the years. One of the most striking trends is Lewis Hamilton's meteoric rise and sustained dominance, particularly between 2014 and 2020, coinciding with Mercedes' supremacy during the turbo-hybrid era. This visualization effectively showcases how Hamilton consistently outperformed his peers during this period, achieving double-digit wins in multiple seasons.

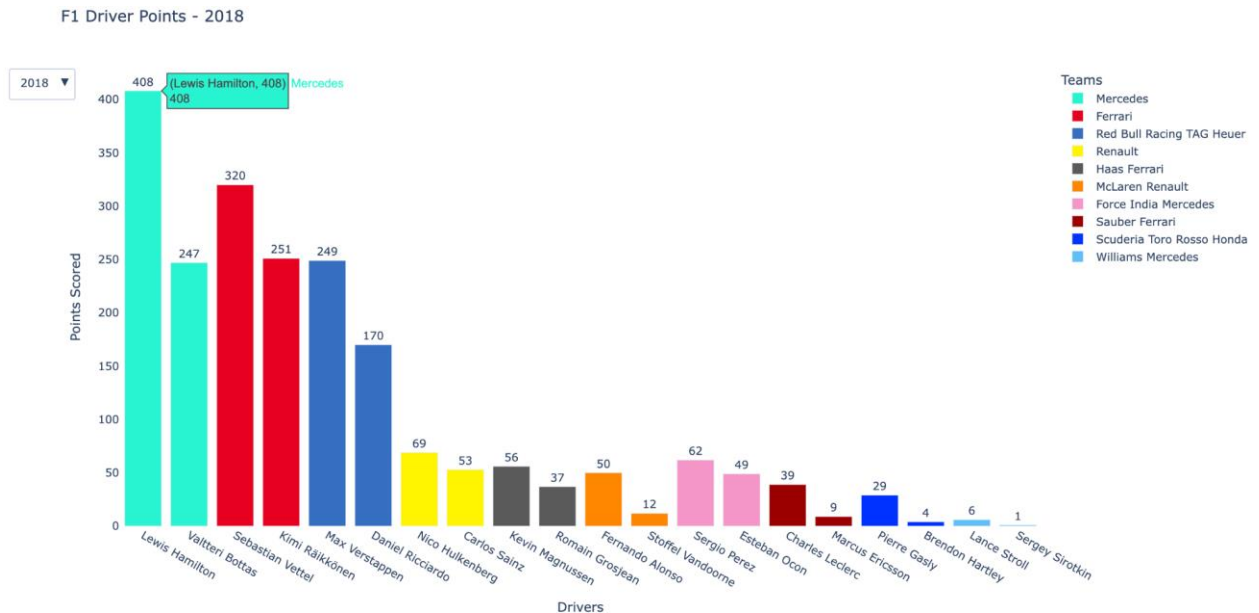
Another interesting insight is the decline of Sebastian Vettel's winning streak after his peak years with Red Bull from 2010 to 2013, where he achieved four consecutive World Championships. The chart also highlights Max Verstappen's rise to prominence starting in 2020, reflecting Red Bull's resurgence as a top constructor. Additionally, drivers like Nico Rosberg, who had a standout season in 2016, and newer talents like Charles Leclerc and George Russell, demonstrate how the competitive landscape has evolved, with emerging stars carving out their place in Formula 1 history. This visualization provides a

comprehensive overview of individual performances, offering a temporal perspective on the ebb and flow of driver dominance in the sport.



- 5) **Dropdown Bar Charts:** In our project, the grouped bar chart is used to showcase the performance of Formula 1 constructors across seasons. Each group represents a constructor, and within each group, the individual bars represent the points scored by the constructor's two drivers. This approach allows viewers to visually compare how constructors relied on their drivers in different seasons. For example, in the 2018 season, Mercedes stands out with both Lewis Hamilton and Valtteri Bottas contributing significantly to the team's total points. The chart provides a clear breakdown of driver-specific contributions while maintaining a grouped structure for constructors.

To enhance interactivity, a dropdown feature was added, allowing users to select a specific season between 2015 and 2024. Upon selecting a season, the grouped bar chart dynamically updates, enabling focused analysis of that year's constructor and driver performance. Team-specific colors were employed to maintain consistency and make it easy for viewers to track constructors across seasons. This visualization not only highlights individual driver performance but also provides insights into team dynamics and strategy, making it a valuable tool for exploring Formula 1 data.



LEARNINGS AND CONCLUSIONS:

This project demonstrates the transformative potential of combining modern web technologies and advanced data visualization techniques to create a highly engaging and educational user experience. Formula 1, as a sport, generates vast amounts of data, and while this complexity can intimidate newcomers, our project bridges the gap by presenting the information in an intuitive, visually appealing, and interactive format. The visualizations go beyond raw numbers—they tell stories that resonate with both new fans seeking to understand the sport and seasoned enthusiasts looking to delve deeper into its intricacies.

The use of tools such as Flask, Dash, and Plotly enabled us to seamlessly integrate interactive charts and dynamic features into a cohesive website. Each visualization was thoughtfully designed to simplify the sport's complexity while preserving the richness of the data. For instance, the treemap reveals the dominance of top constructors while allowing users to drill down into individual drivers' contributions, offering a layered perspective. Similarly, the line graph captures the rise and fall of legendary drivers, enabling users to appreciate their performances across seasons.

Through this project, we learned the importance of balancing aesthetics, interactivity, and performance optimization. While aesthetics enhance user engagement, interactivity empowers users to explore the data on their own terms, and performance optimization ensures a smooth and responsive experience, even for complex visualizations like treemaps. The iterative process of experimentation, analysis, and refinement helped us identify visualization methods that effectively

communicate key insights, such as the dominance of Ferrari, the global expansion of Formula 1 circuits, and record-breaking performances by drivers like Lewis Hamilton and Max Verstappen.

The website serves as more than just an educational tool; it acts as a gateway for newcomers to explore and appreciate Formula 1's rich history and legacy. It fosters a deeper understanding of the sport by making its data accessible, engaging, and informative. By providing dynamic filtering options, interactive elements, and compelling visual narratives, the website opens up the world of Formula 1 to a broader audience, helping to grow the fanbase.

In conclusion, this project exemplifies how data visualization and web technologies can make even the most complex datasets approachable and engaging. By focusing on storytelling and interactivity, we have created a tool that not only educates but also inspires curiosity and passion for Formula 1, bringing its vibrant history and exciting future to life for a global audience.

FUTURE WORK:

Network Graphs for Driver-Team Relationships: Develop network graphs where nodes represent drivers and teams, and edges represent collaborations. This would highlight the interconnected history of F1 and the legacy of certain teams.

Performance Radial Charts: Use radial bar charts to compare driver performances in categories like overtakes, pole positions, fastest laps, and podium finishes across seasons.

Driver vs. Driver Head-to-Head Tool: Implement an interactive comparison feature allowing users to compare two drivers across metrics like wins, podiums, lap times, and career trajectories on selected circuits.

Race Outcomes by Start Position: Introduce a visual that correlates starting grid positions with final race outcomes, using violin plots or swarm plots to show distributions.

Dynamic Storytelling with Data: Add narrative visualizations, using scrolly-telling techniques, to walk users through iconic races, historic rivalries, or the evolution of F1 teams.

REFERENCES:

- Formula 1 Official Website: <https://www.formula1.com>
- Plotly Documentation: <https://plotly.com/python/>
- Flask Framework: <https://flask.palletsprojects.com>
- Dash Framework: <https://dash.plotly.com/>
- Render: <https://www.programonaut.com/host-your-web-application-for-free-with-render-step-by-step/>
- <https://www.pathfinder-analytics.net/post/introduction-to-treemaps-in-plotly-express-python>
- Inspirations:
https://www.reddit.com/r/dataisbeautiful/comments/16c8m1i/oc_total_points_accumulated_by_each_f1_driver_in/
- [\[OC\] Formula 1 Driver Ranking 2022 Season : r/dataisbeautiful](#)