Group 3 – Project 3 Write Up

Group Members:

- Marta Baker
- Katie Rebeck
- Kevin Carney

Introduction:

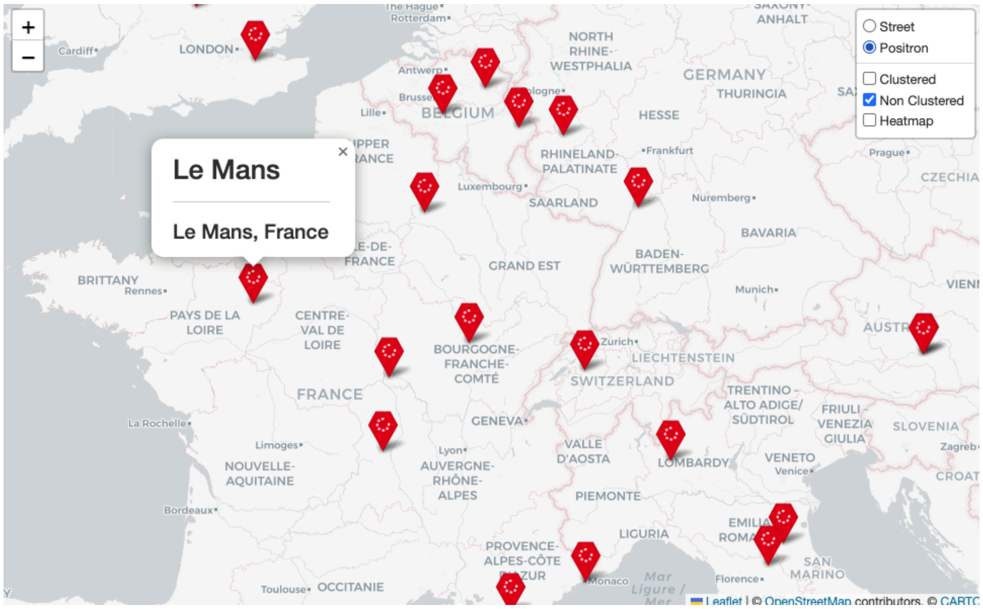
The purpose of Project 3 was to choose a data set found on Kaggle or other similar services. Once found, the directive was to take the data set and build a full stack website to include interactive visualizations and at least one map using the data. For group 3, we opted to use a dataset built around Formula One (hereafter referred to as F1) racing. The dataset was titled "*Formula 1 World Championships (1950 – 2024)"* by a user named Vopani. Per the provenance, the data was based on information found from Ergast.com, which is an API that is dedicated to covering F1 racing.

Definitions:

- **Circuit:** The location where an F1 race occurs.
- **Constructor:** The team that a driver is associated with.
- **CSV:** Comma Separated Value worksheet, similar to an excel worksheet.
- **Kaggle:** Website hosting a database of open source datasets for use in data analysis.

Our objective was to create a high-level overview of what F1 is. Additionally, we sought to analyze not only where the various circuits were located, but also the performance of the drivers grouped by nationality. The dataset included 14 CSVs, of which only six were ultimately used. Data cleaning operations were minimal, with the bulk of the work being the removal of extraneous columns that were not relevant to our research questions. Other operations involved the consolidation of USA and United States into a single grouping for the circuits.csv “country” column, and removal of null values.

When it came to devising a design for our application and visualizations, we reviewed the F1 website and used that overall color scheme and layout to help guide our decisions. For the map element of our application, we wanted to enable the user to quickly see the overall locations of the various circuits on a world map. From here, we wanted to give a user the capability to zoom in and click on the pins to see what the name of the circuit was and what city it was located in.



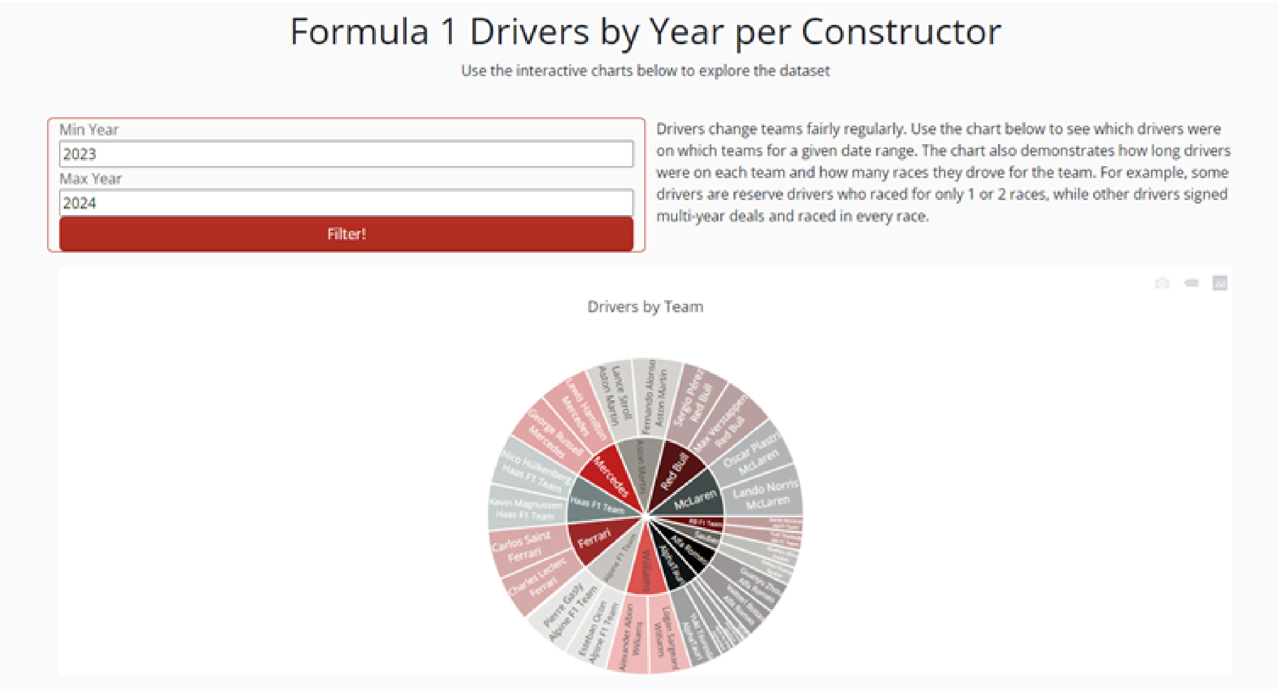
Le Mans

Le Mans, France

○ Street
● Positron

☐ Clustered
☑ Non Clustered
☐ Heatmap

When determining how to answer the question related to the most recent constructor that a driver had been associated with, it was deemed that the most evocative method of conveying that information was through a sunburst chart. This way a user could not only see the list of the various constructors that are currently active, but subsequently click on each constructor’s entry on the visualization and see all the drivers currently active for that team. Additional functionality was added to enable a range of years to be applied to the visualization, resulting in an automatically updating sunburst showing the results.



Formula 1 Drivers by Year per Constructor

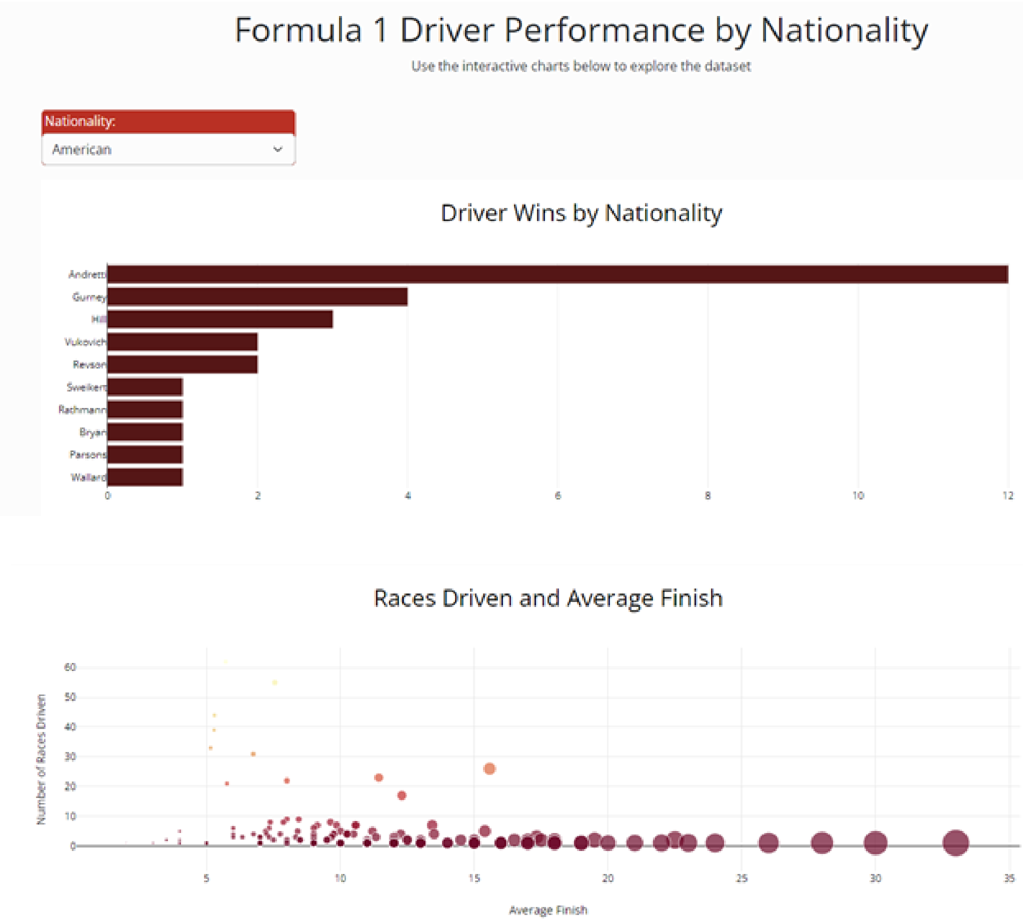
Use the interactive charts below to explore the dataset

Min Year: 2023

Max Year: 2024

Drivers change teams fairly regularly. Use the chart below to see which drivers were on which teams for a given date range. The chart also demonstrates how long drivers were on each team and how many races they drove for the team. For example, some drivers are reserve drivers who raced for only 1 or 2 races, while other drivers signed multi-year deals and raced in every race.

Our final visualizations answered the question pertaining to the overall performance for drivers grouped by nationality. As this data accounted for all races and drivers since 1950, it was felt that restricting it to first place finishes would enable the visualizations to be more effective in conveying the findings. A drop-down box was added to enable a user to select a nationality they wanted to view, and it would return with a bar chart filtering the results based on the user input. This chart would not only give the names of the drivers from that nationality, but also the number of 1st place finishes that they've had. Additionally, a bubble chart will populate that shows the average finish versus number of races that drivers of that nationality have achieved, with each driver having their own bubble whose size was scaled based on overall finish.



Formula 1 Driver Performance by Nationality

Use the interactive charts below to explore the dataset

Nationality: American

Overall, as the dataset was based on recorded performance in sanctioned events, any sort of inherent bias should be minimal. That stated, there was missing data from earlier races, it is possible that were this missing data to be implemented into the results, certain outcomes could be affected. Of note would be the performance based on nationality. But an analysis of those races, as well as independent research to try and find the missing information would require a great deal of time and effort which was outside of the available time for this project.

In general, time was the biggest limitation for this project. There were eight other CSVs that contained a myriad of data that could have lent itself to additional inquiries and research. But due to the deadline for the presentation, those other queries were not able to be pursued.

This project was a significant undertaking for our group, but we learned a lot about full stack development and hosting. Were supplemental investigations to occur, we believe that there could be other relationships within the dataset to analyze. For instance, were there any years where notable developments occurred? Was there a circuit where a historically underperforming constructor went on a streak? If so, how did that influence and change their subsequent circuits? Do “home turf” circuits have any impact on performance for drivers who are originally from that area? These questions and more could be a research project of their own. Were we to look into trying to pursue these additional inquiries, we have thankfully had a strong baseline of experience to build from.