

CS 6630: Process Book - Formula 1 Stats

Team Members: Harshini Keerthi Vasan, Vinu Sreenivasan, Sanjana Aravindan

Basic Info

Project Title

Formula 1 Stats

Team Members

1. Sanjana Aravindan, u1135274, u1135274@utah.edu
2. Harshini Keerthi Vasan, u1136693, u1136693@utah.edu
3. Vinu Sreenivasan, u1138972, vinu.sreenivasan@utah.edu

Project Repository

<https://github.com/HarshiniVasu/Formula-1-Stats>

Overview

Formula 1 Stats is a data visualization project that would allow the user to visualize the most successful teams (car brands) over a period of time and top circuits in the world. It would also show a comparison between various drivers who were active over a period of time. This project would include three main visualizations:

1. Showing the most successful teams over a period of time and the respective drivers who represented those teams in those years.
2. A comparative visualization between driver statistics who were active over a period of years.
3. Showing a map visualization of all the regions in the world that have formula 1 circuits (race tracks). This part would also show the top circuits in the world on map.

Background and Motivation

We, as a team are ardent fans of Formula 1 racing. This racing is a perfect blend of technology and the automobile industry. The mere adrenaline rush that the viewers get watching this sport, is something that gets us excited about working on this project. We are inspired enough to visualize various interesting correlations between all the drivers of the league and eminent teams like Ferrari, Mercedes and McLaren.

Though there are many visualizations on this data online, we couldn't find a place which shows the most successful teams over a selected period of years. Also, most of the visualizations did not have a comparative analysis of drivers over a period of time (for example from 2010-2015). This idea motivated us to take up the project.

Related Work

Formula 1 racing is a very well renowned sport. While we see a lot of visualizations on FIFA world cup and Olympics, we don't get to see that much for Formula 1 racing. That was one of the reasons that we chose to work on an unconventional sport data and build visualizations on it.

Project Objectives

We aim to show the highly sought after mainstream Formula 1 teams that are/were in the league over a period of time that the user wishes to visualize. The main questions that we are trying to answer are,

1. What are the top most circuits in the world? This would be on the basis of the number of races that took place on those circuits between 1970 and 2018.
2. What are the teams that scored well on those circuits respectively?
3. What are the most successful teams in a sorted order over a specific period of time?
4. Who are the drivers who drove for those successful teams in that specific period of time?
5. What are the basic details and stats of the driver such as points earned by the driver on a specific period of time?
6. Comparison of two or more driver's performance in terms of points and average lap time over a specific period of team.

We would like to use the knowledge we have in d3 to visualize the formula 1 data set. We will use concepts covered in class, assignments and also a few other d3 based visualizations to build an effective user interface with interesting visualizations of formula 1's records.

Data

We did a significant amount of research and found a set of multiple csv files related to formula 1 racing from Ergast Developer API. These csv files provide mapping of various sets of data using primary keys. There is a circuit (race track) data file which comprises of a unique key id, the track location of the races with its latitude and longitude. The constructors file (f1 teams) has fields like id, nationality of the team and their standings game-wise and year-wise (1970-2018). The driver's data file includes the history of their games (wins), the team they raced for, their standings in global level year wise as well as match wise. There is another interesting csv file which provides the status of the drivers with respect to every game denoting the number of laps, damages, collisions, etc.

Data Processing

We have done a bit of data cleansing and re-organized the data according to our visualization specification. We have multiple csv files like circuits (race tracks), constructors (teams), drivers, results, status, races. Using the primary keys (ids) present in every csv file, we combined the data that were only necessary from all of these files like driver statistics (nationality, DOB, points, position, laps), team statistics (nationality, results, laps), status (the issues or status tracking during the racing - oil leak, collision, engine, accident, axle, count of laps successfully). We also split the raw data year-wise so that retrieval of the data year-wise for team statistics and driver statistics visualization becomes easier.

```
In [5]: races_results_circuit_drivers = pd.merge(races_results_circuit, drivers_data[['driverId','driverRef','number','forename','surname','dob','nationality']], on='driverId')

In [6]: races_results_circuit_drivers.rename(columns={'year':'season',
                                                 'name_x':'raceName',
                                                 'circuitRef':'circuitId_x','name_y':'circuitName'},
                                             inplace=True)

In [7]: races_results_circuit_drivers.rename(columns={'dob':'dateOfBirth','forename':'givenName','surname':'familyName','grid':'results.grid','laps':'results.laps'},
                                             inplace=True)

In [8]: races_results_circuit_drivers_cons = pd.merge(races_results_circuit_drivers, constructors_data[['constructorId','constructorRef','name','nationality']], on="constructorId")

In [9]: races_results_circuit_drivers_cons.rename(columns={'nationality_x':'nationality','nationality_y':'nationality.1','name':'name_x'},
                                                inplace=True)

In [10]: races_results_circuit_drivers_cons_status = pd.merge(races_results_circuit_drivers_cons, status_data, on='statusId')

In [11]: races_results_circuit_drivers_cons_status.rename(columns={'status':'results.status'}, inplace=True)

In [12]: races_results_circuit_drivers_cons_status.columns

Out[12]: Index([u'raceId', u'season', u'round', u'circuitId', u'raceName', u'date',
                u'driverId', u'constructorId', u'number_x', u'results.grid',
                u'position', u'positionText', u'results.laps', u'points', u'statusId',
                u'circuitId_x', u'circuitName', u'location', u'country', u'lat', u'lng',
                u'driverRef', u'number_y', u'givenName', u'familyName', u'dateOfBirth',
                u'nationality', u'constructorRef', u'name_x', u'nationality.1',
                u'results.status'],
                dtype='object')

In [13]: races_results_circuit_drivers_cons_status = races_results_circuit_drivers_cons_status.drop(['number_y','circuitId','constructorId', 'driverId'], axis=1)

In [14]: races_results_circuit_drivers_cons_status.rename(columns={'number_x':'number'}, inplace=True)

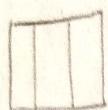
In [15]: races_results_circuit_drivers_cons_status = races_results_circuit_drivers_cons_status[['season','round','raceName','date','circuitId_x','circuitName','number','position','positionText','points','driverRef','givenName','familyName','dateOfBirth','nationality','constructorRef','name_x','nationality.1','results.grid','results.laps','results.status','location','country','lat','lng']]
races_results_circuit_drivers_cons_status.columns

Out[15]: Index([u'season', u'round', u'raceName', u'date', u'circuitId_x',
                u'circuitName', u'number', u'position', u'positionText', u'points',
                u'driverRef', u'givenName', u'familyName', u'dateOfBirth',
                u'nationality', u'constructorRef', u'name_x', u'nationality.1',
                u'results.grid', u'results.laps', u'results.status', u'location',
                u'country', u'lat', u'lng'],
```

Exploratory Data Analysis and Design Evolution

The images below are the initial set of visualizations that we considered doing as a part of the project during our project proposal phase.

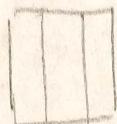
Driver Champions:



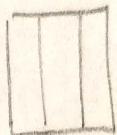
1990



1991



1992



1993



1994



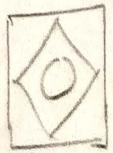
1995



1996



1997



1998



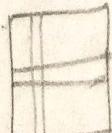
1999



2000



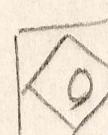
2001



2002



2003

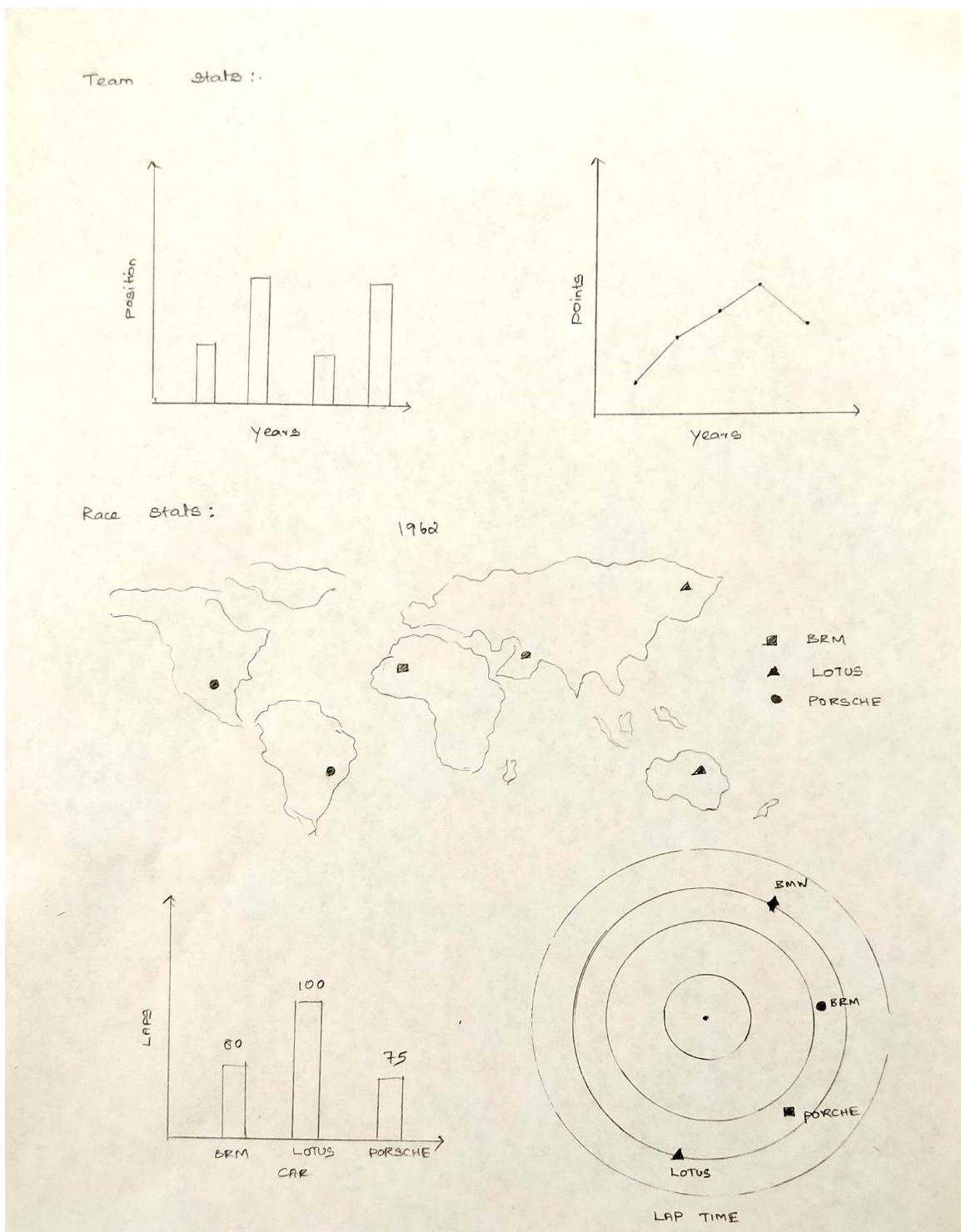


2004

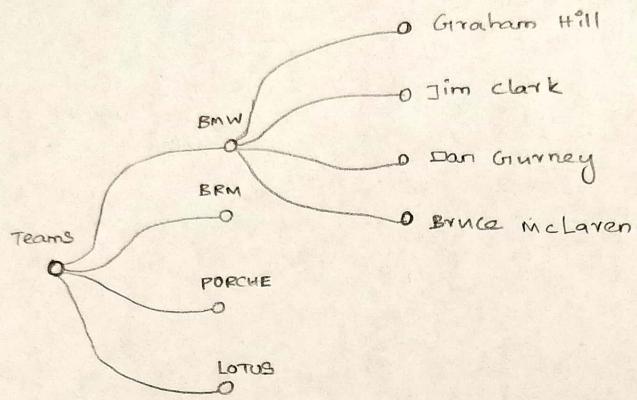


2005

Team Statistics Images



Team info :

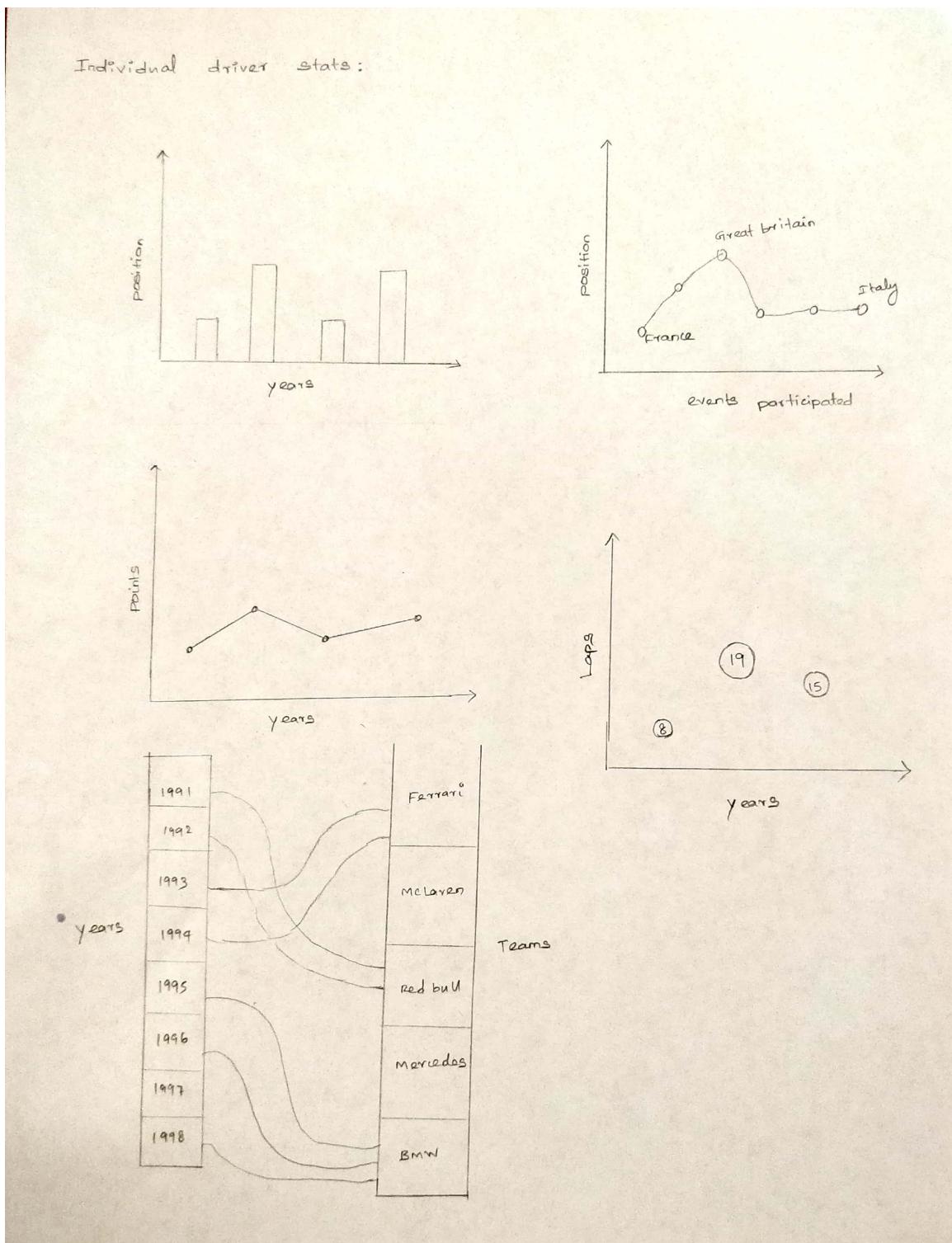


Treemap for teams

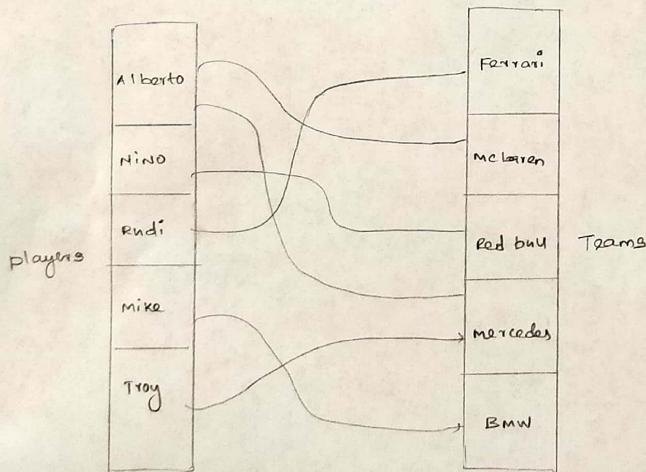
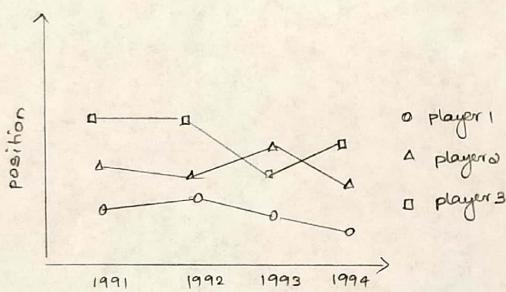
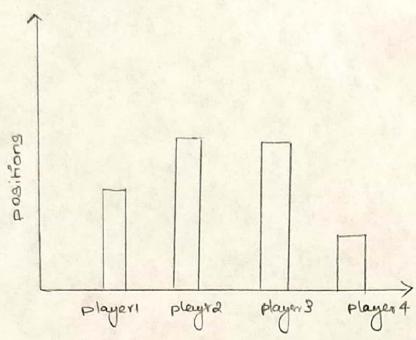
Team Stats : Races/Wins

< Races Wins >		
945	Ferrari	228
818	McLaren	182
720	BMW	79
740	BRM	55
620	PORCHE	35
750	LOTUS	27
710	Renault	32
720	TOYATA	26
680	PROST	36
600	Lola	17

Individual Driver Statistics Images



comparison of drivers per year



While the visualizations above depict a lot of interesting correlations, we could figure out that these visualizations are very generic and does not answer any specific question or give any distinct insight to the data. After a feedback session with a peer team, and with the TA, we came up with a modified version of our visualizations. The visualizations shown below are well organized and provides a directional flow to the project.

Modified Design and Implementation

Design Changes

One of the images below shows the home page of our project. This would contain three tabs - Teams, Drivers, Circuits.

The first tab - "Teams" would have a year chart from 1970 - 2018. Selecting a particular year would give us a stacked horizontal bar chart with races/wins of the top most car teams in that year. There would also be a year slider functionality over the year chart that would help the user visualize the top teams over a specific period of time (for example from 1970 -1985). In order to make it interactive, we have decided to display the names and other details such as nationality and points of the drivers who represented a particular team on clicking the bar corresponding to the team.

The second tab - "Drivers" will have a drop down box of years with to and from labels. On selecting a range of years, we would get a list of drivers who drove during those years. Clicking on a particular driver would give his statistics. Below the names of the drivers and the description of the selected driver (teams he represented, nationality, date of birth), we would give the user an option to compare the drivers. The user can select one or more drivers to compare (using a drop-down box) with the default selected driver. This would display two line charts with the average lap time and the position between drivers compared.

The third tab - "Circuits" will represent a map that highlights all the regions in the world that have formula 1 circuits. We would further have a button "Top Circuits". Clicking on that would modify the map and highlight all the regions of the world that have the top most circuits. There would also be a table below showing the top circuits of the world with its statistics. We would rank the circuits based on the number of races that happened in that circuit. The more the races, the higher the rank. To make the map interactive, we have decided to provide an info box. Clicking on a region would give us the name of the circuit, the number of races that happened on that circuit and the name of the team that was the most successful on that circuit with respect to wins and the number of wins that the successful team had on that circuit.

Implementation

This section would highlight the designs that were finalized for the project after several brainstorming sections and the corresponding implementation phases of the corresponding design.

We decided to have a well structured and an attractive home page for our project. We believe, this would enhance the user experience. We have used HTML and Bootstrap to design our Home page. Below are the screen shots of the same.

We have also created About and Contact tabs to help the users understand about our project and contact us for any further details about it. We would also be happy to hear any feedback that would help us improve our design.



As ardent fans of Formula-1 Racing, the three of us, Vinu Sreenivasan, Harshini Keerthi Vasan and Sanjana Aravindan decided to visualize and find out interesting facts about the best teams and the best players in the Formula-1 racing history. Formula-1 fandom has increased tremendously over the years, and the three of us have been lucky to witness the growth of players and teams from different parts of the world. We hope our project gives a good insight about the drivers, circuits and the best teams of Formula-1 racing.

To launch our visualization, simply click the home button on the navigation bar and click the "Get Started" button. The main page will show up with four tabs linking to Teams, Drivers, Circuits and Demo respectively. The Teams tab will have interesting visualizations about various teams in the Formula-1 history. The Drivers tab would highlight the specifics of a particular driver selected. The Circuits tab would give us an overview about all the F1 race circuits in the world. The Demo tab provides a link to the youtube screencast of our project. So what are you guys waiting for, go ahead and enjoy the visualization! :)

Our Process book gives a detailed analysis about the design choices that we had to make for this project. It also highlights the various problems that we faced during the course of this project and the methods that helped us in fixing them. You can find the link to the process book below:

Process Book : <https://github.com/HarshiniVasu/Formula-1-Stats/blob/master/Process%20Book.pdf>

Harshini Keerthi Vasan

[LinkedIn](#)

[Email](#)

Sanjana Aravindan

[LinkedIn](#)

[Email](#)

Vinu Sreenivasan

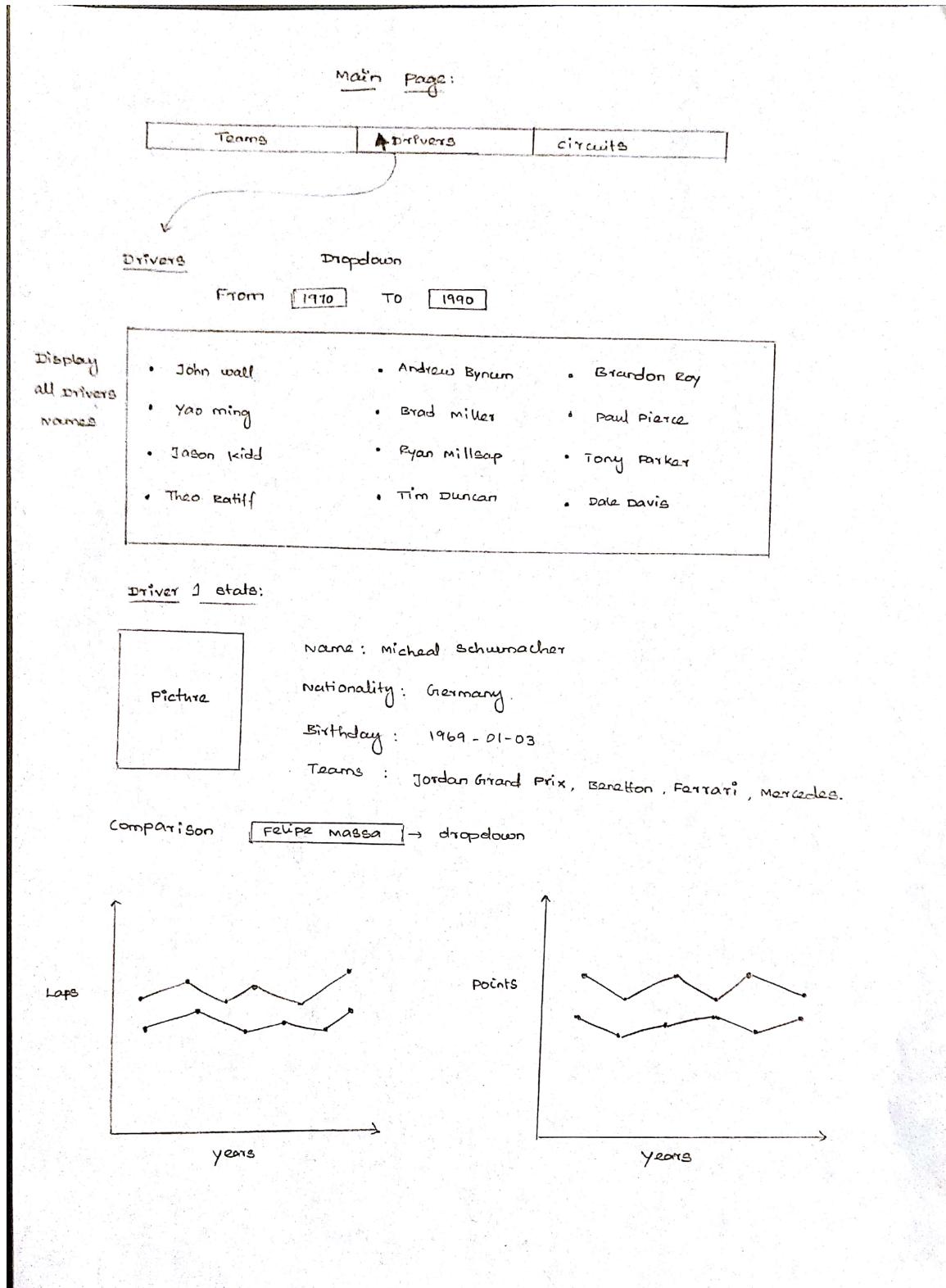
[LinkedIn](#)

[Email](#)

Drivers Tab Implementation

The screen shot below is our main page that comes up after clicking the "Get Started" button in the Home Page. It contains four tabs - Teams, Drivers, Circuits and Demo.

We have our implementation screen shots below the design snap shots to help the user have an idea about our transition from a paper-based design to a developed web page.



Home About Contact

Year From **1970** • Year To **1981**

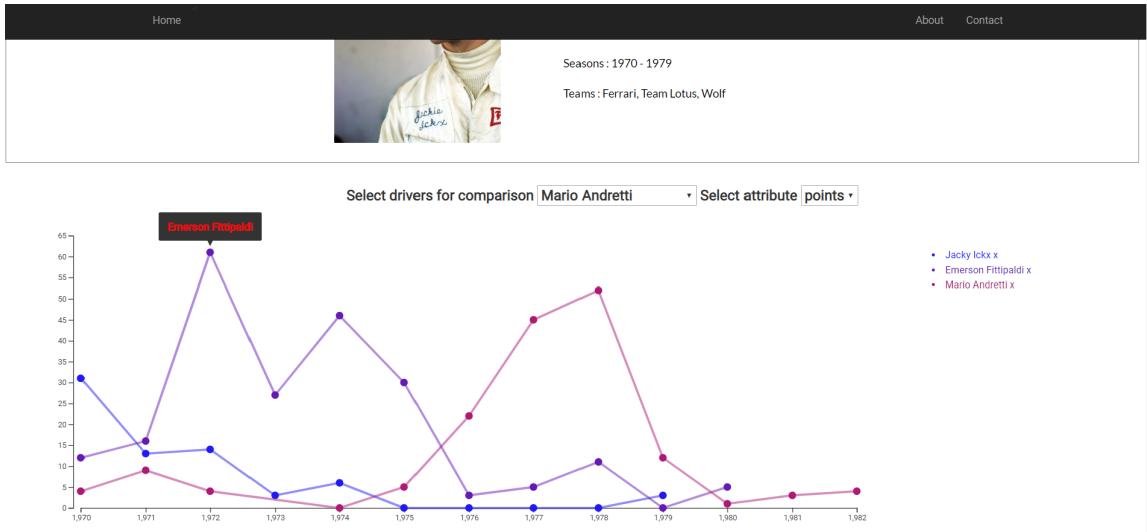
Select another driver

• Jacky Ickx	• Denny Hulme	• Emerson Fittipaldi	• Rolf Stommelen	• Henri Pescarolo
• Francois Cevert	• Graham Hill	• Chris Amon	• Clay Regazzoni	• Jackie Stewart
• Mario Andretti	• Ronnie Peterson	• Jackie Oliver	• Jean-Pierre Beltoise	• Peter Gethin
• Brian Redman	• Andrea de Adamich	• Tim Schenken	• Nanni Galli	• Dave Charlton
• Derek Bell	• Reine Wisell	• Howden Ganley	• Mike Hailwood	• Jean-Pierre Jarier
• Gijs van Lennep	• Niki Lauda	• Carlos Pace	• Carlos Reutemann	• Arturo Merzario
• Patrick Depailler	• Jody Scheckter	• James Hunt	• John Watson	• Jochen Mass
• Hans-Joachim Stuck	• Vittorio Brambilla	• Tom Pryce	• Jacques Laffite	• Ian Scheckter
• Jean-Pierre Jabouille	• Alan Jones	• Brian Henton	• Harald Ertl	• Brett Lunger
• Tony Trimmer	• Emilio de Villota	• Rupert Keegan	• Riccardo Patrese	• Hector Rebaque
• Patrick Tambay	• Gilles Villeneuve	• Bruno Giacomelli	• Didier Pironi	• Rene Arnoux
• Derek Daly	• Keke Rosberg	• Nelson Piquet	• Eddie Cheever	• Geoff Lees



Jacky Ickx
DOB: 1/1/1945
Nationality: Belgian
Seasons: 1970 - 1979
Teams: Ferrari, Team Lotus, Wolf

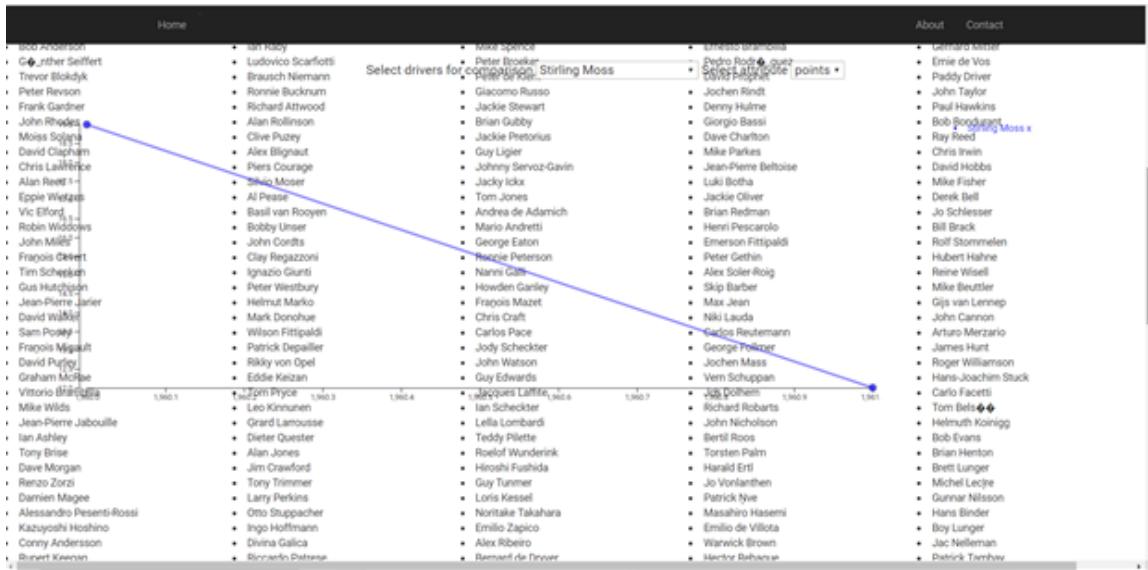
The Drivers tab would highlight the specific details of a particular driver selected within a certain year range from the drop down by clicking the "Select Another Driver" button. We also have included a different kind of visualization below the driver details that would allow us to see a comparison between two or more drivers based on their points and laps. This comparison is also a part of our paper-based design. The screen shot for the same is below.



Design Issues

We faced a few problems while designing this page. Firstly, the driver names had a lot of ASCII characters that we had to resolve by cleaning the dataset. We also had alignment issues while including both the driver details and the line chart comparison

in the same page. We had to re-scale the svg to make them both of them fit in the same page without issues. Below is a screen shot of the problem we faced.



Teams Tab Implementation

The Teams tab is the default page of our main screen. This tab would provide us with a bar chart of various teams based on the number of races they participated over a particular period of time (range of years can be selected with the help of the year slider) and the number of wins they had during those years. This tab would also give us the details of drivers who participated for those teams over those years. Earlier, while designing contents for this tab, we thought we would just include a table of driver details. Later we decided to make it more attractive by using a bubble chart that would contain the flags of the country that the respective drivers represent. The bubble chart would be visible by clicking on a particular team name in the bar chart. Hovering on those flags in the bubble chart would highlight the details of the drivers such as the driver's name, nationality and the number of races he participated for the particular team. The radius of the circle in the bubble chart is scaled by the number of races the driver participated in. Below are the paper based design snapshots and screen shots of our implemented design.

Team - (constructors)

year slider

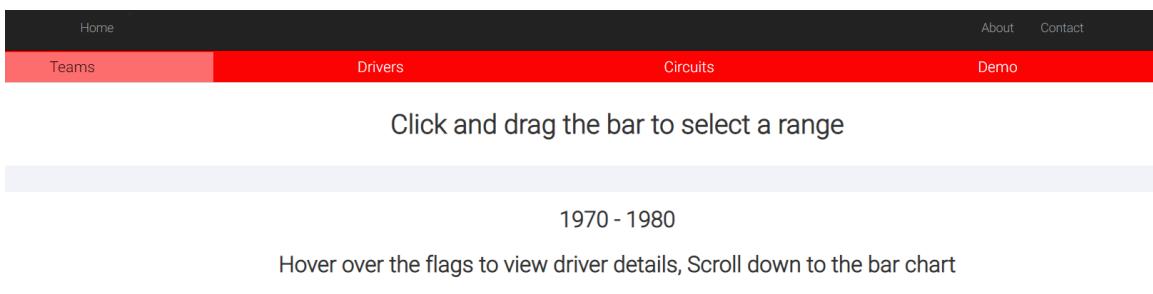
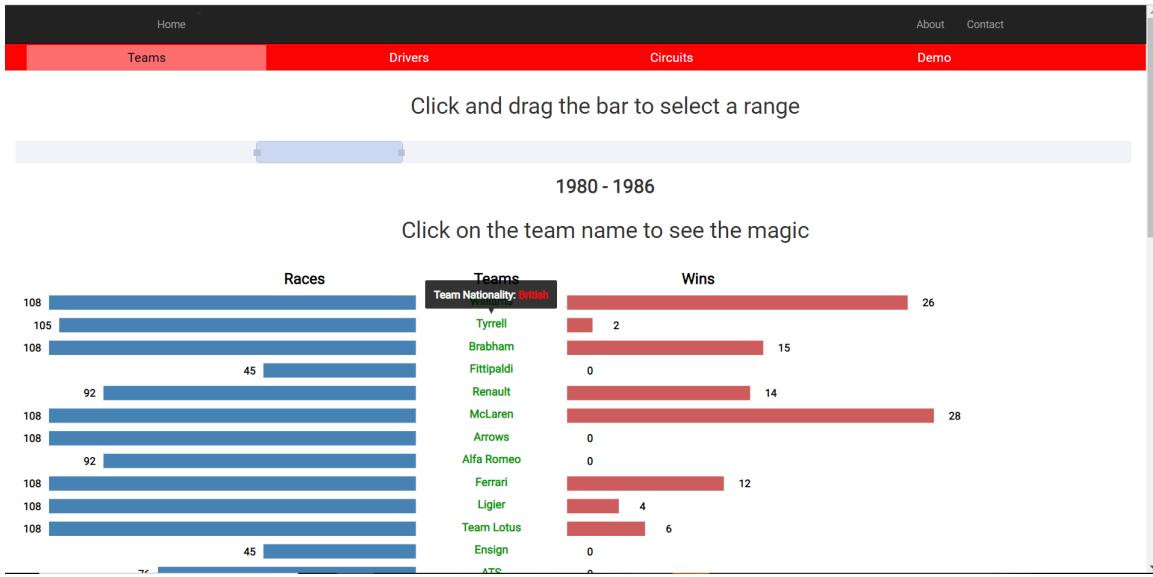


	Races	wins
945	Ferrari	208
818	McLaren	182
671	LOTUS	114
401	BRM	79
403	BMW	72
394	Red Bull	53
337	Ford	35
250	Minardi	27
219	Toro Rosso	23
208	EMD	17

on click

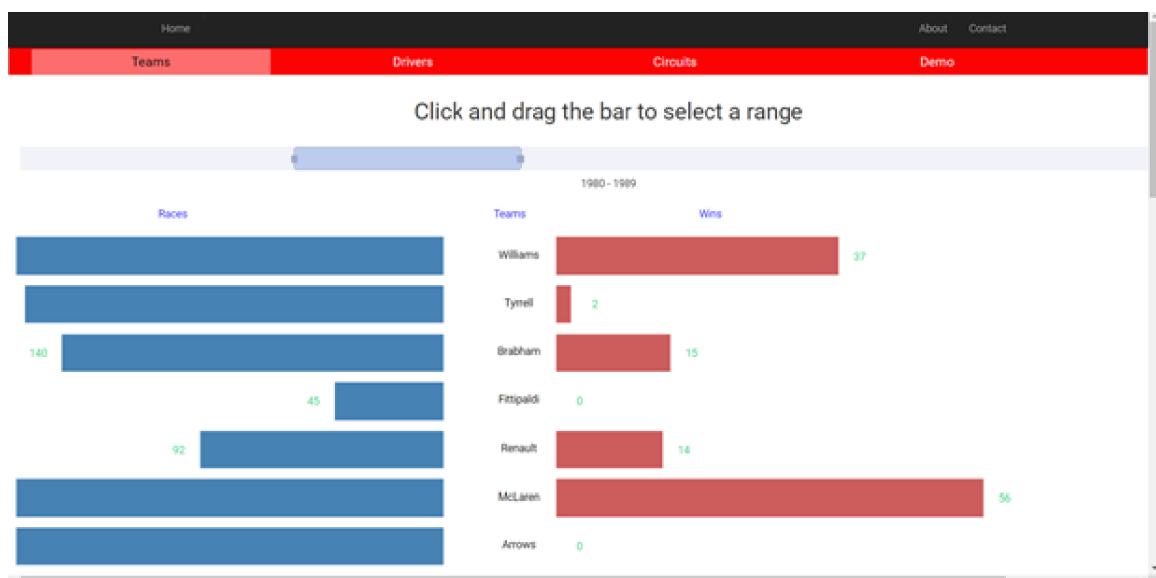
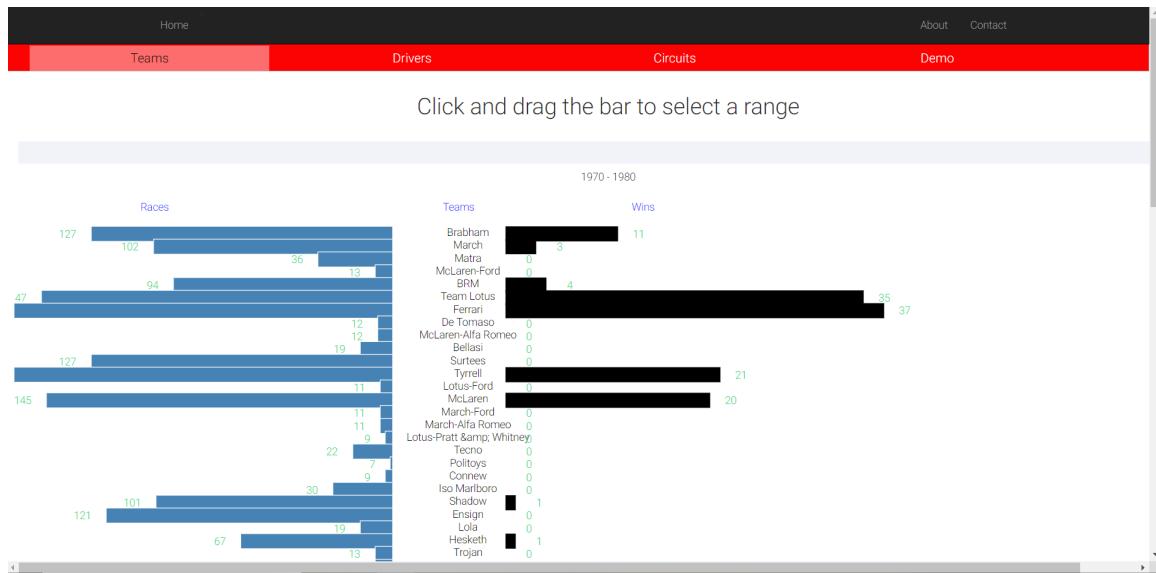
Drivers (Ferrari)

	Laps	Nationality	points
Sebastian Vettel	58	canada	90
Kimi Raikkonen	56	France	89
Lewis Hamilton	43	Germany	88
Fernando Alonso	40	Hungary	84
Felipe Massa	40	Spain	82
Michael Schumacher	39	malaysia	80

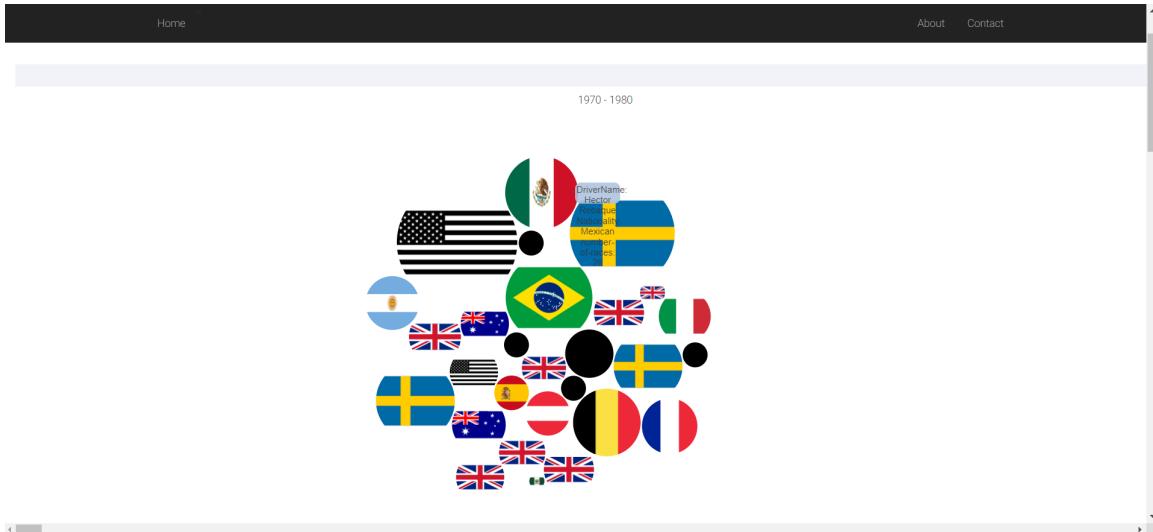


Design Issues

We had a lot of scaling issues with the bar chart design. There was a specific problem that took us a lot of time to figure out. The height of the bars in the bar chart were dynamically increasing with every year interval change. We were using bandwidth that is used with scaleBand for scaling the height. Later, we solved this issue by fixing the height to a particular value. Below is a screen-shot of our problem.



We also had a few black circles popping up in the bubble chart . Those black circles were of the countries that had spaces in their names like "South Africa", "New Zealand". We later re-modified our code to consider only the first word of the countries name that included space and named the flag accordingly in our images dataset. Below is a screen-shot of the problem.



Circuits Tab Implementation

The Circuits tab contains two buttons, "All Circuits" and "Top Circuits". Clicking the "All Circuits" button would highlight the location of all the race circuits in the world. The race circuits are represented as circles on the map. Hovering on the circle would give us the details of the circuit such as the name of the race circuit, the location and the country where the circuit is present, the team that has won the most on that circuit and the number of wins that the most successful team has had on that circuit. Clicking on a circle (circuit) on the map would show an image of the car of the most successful team and the corresponding details such as circuit name, the most successful team and number of wins. The actual implementation of this tab is same as the paper-based design implementation that we had. We just included an extra functionality of showing up the most successful team car images and their details as a part of this view. The snapshot and the screen-shots below highlight our design implementation.

circuit (map)

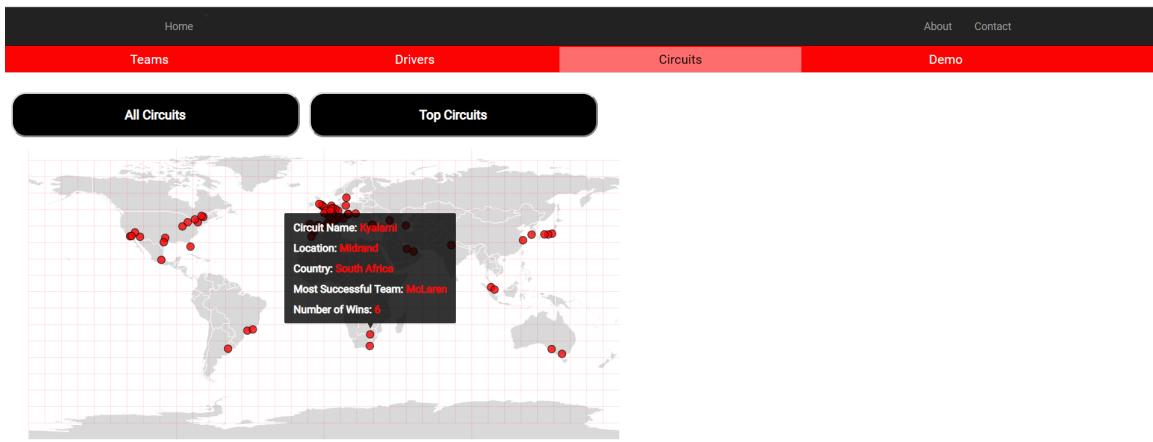
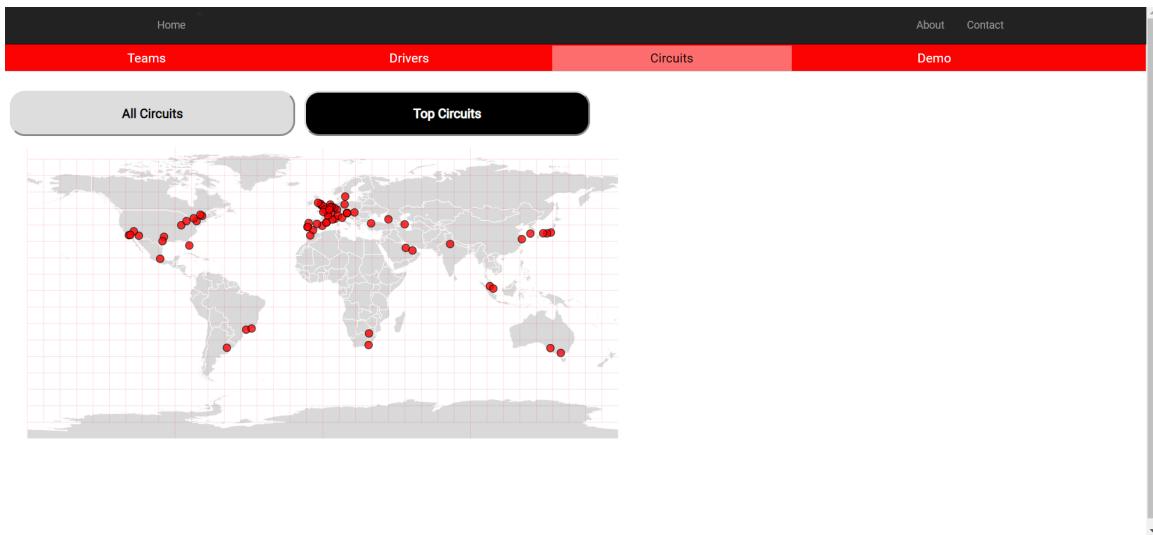
All

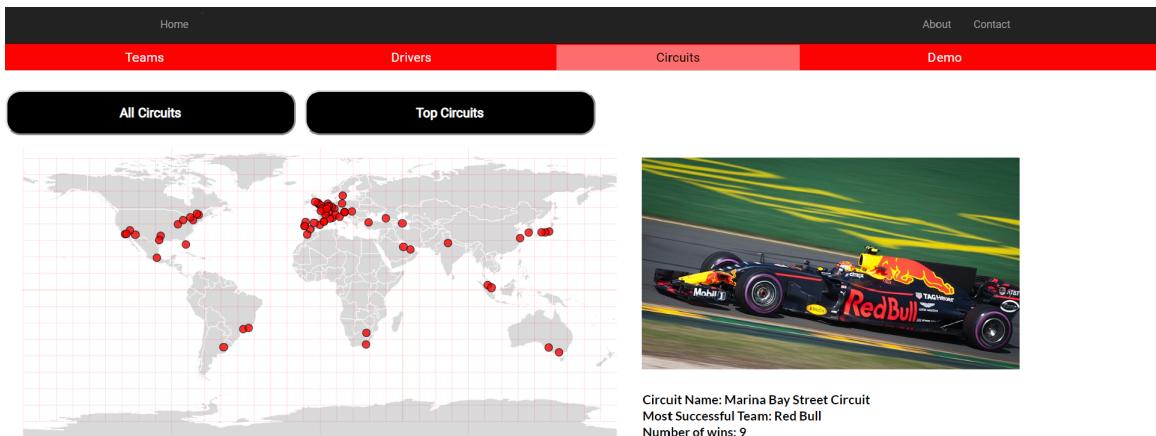
Top circuits



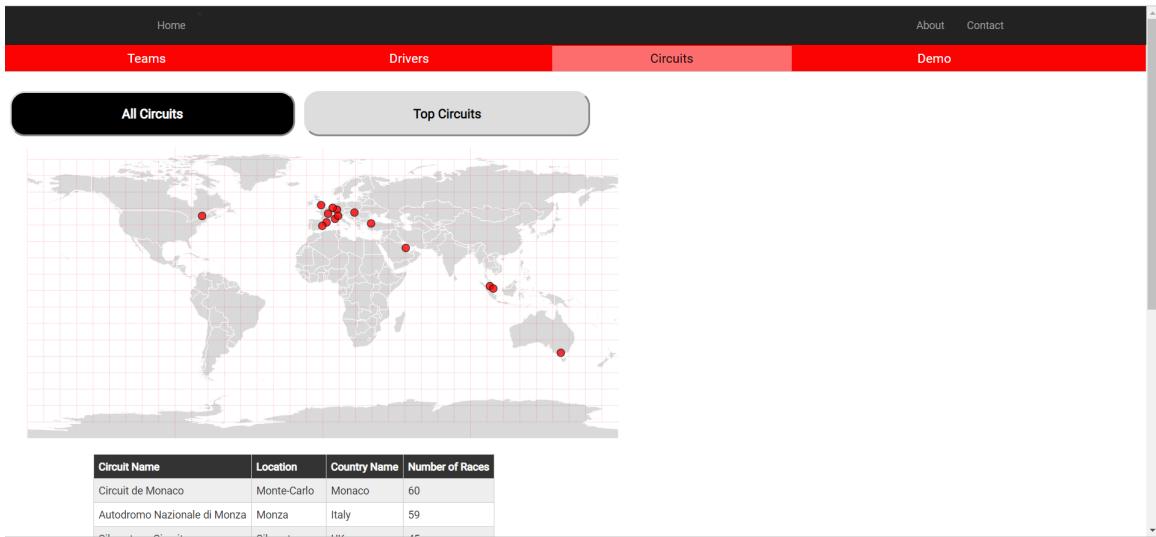
Top circuits:

Las Vegas	Nevada	USA	65
Gilles Villeneuve	Montreal	Canada	64
Hungaroring	Budapest	Hungary	51
Sepang	Kuala Lumpur	Malaysia	48
Bahrain	Sakhir	Bahrain	40
Hockenheimring	Hockenheim	Germany	38
Shanghai	Shanghai	China	35
Albert Park	Melbourne	Australia	34





Clicking on the other button named "Top Circuits" would highlight the top 15 circuits in the world based on the total number of races on those circuits so far. The circles highlighted as a part of the "Top Circuit" button functionality would also exhibit the same behavior as the "All circuits" button view. This view would also have a table with specific details (Circuit Name, Location, Country, Total number of races on that circuit) about all the top 15 circuits.



The screenshot shows a dark-themed web application interface. At the top, there's a navigation bar with links for "Home", "About", and "Contact". Below the navigation bar is a table titled "Circuits" with the following data:

Circuit Name	Location	Country Name	Number of Races
Circuit de Monaco	Monte-Carlo	Monaco	60
Autodromo Nazionale di Monza	Monza	Italy	59
Silverstone Circuit	Silverstone	UK	45
Circuit de Spa-Francorchamps	Spa	Belgium	43
Circuit Gilles Villeneuve	Montreal	Canada	38
Hockenheimring	Hockenheim	Germany	35
Hungaroring	Budapest	Hungary	32
Circuit de Barcelona-Catalunya	Montmelo	Spain	27
Albert Park Grand Prix Circuit	Melbourne	Australia	22
Sepang International Circuit	Kuala Lumpur	Malaysia	19
Circuit de Nevers Magny-Cours	Magny Cours	France	18
Bahrain International Circuit	Sakhir	Bahrain	13
Marina Bay Street Circuit	Marina Bay	Singapore	10
Istanbul Park	Istanbul	Turkey	7
Valencia Street Circuit	Valencia	Spain	5

Design Issues

There were not many design issues with this tab. A few locations and circuit names had ASCII characters that we removed completely. The figure of the most successful team's car was not fitting at the right place in the svg. We had to scale the image and the margin to achieve the right fit of the image.

Demo Tab

The Demo tab contains a link to our screen cast video.

Evaluation

Our motive of the project was always to build a simple and clean design to view interesting insights about Formula-1 racing. Initially, when we started with the project, we thought that we had all the necessary data to build our visualization. Gradually, when we started working on those designs we realized that we do not have the necessary information yet. Therefore, we had to do a lot of data manipulation to achieve the kind of visualization we have now. We wanted our visualization to answer a few specific questions while we built the prototype of this project. The content below will highlight the questions and the corresponding tab/visualization that answers it.

1. What are the top most circuits in the world? This would be on the basis of the number of races that took place on those circuits between 1970 and 2018.

This question is answered by the **Circuits** tab in our visualization which highlights the top most circuits in the world on a "Top Circuit" button click. On clicking the button, top most circuits will be displayed in a table.

2. Which are the teams that scored well on those circuits respectively?

This question is also answered by the **Circuits** tab which highlights the most successful team on that particular circuit on a mouse hover over the circle. On clicking the circle, the most successful team's details will be displayed.

3. What are the most successful teams over a specific period of time?

This question is answered by the **Teams** tab which provides a comparison of the races and wins of the teams in a bar chart representation over a specific period of time.

4. Who are the drivers who drove for those teams in that specific period of time?

This question is also answered by the **Teams** tab which provides a bubble chart representation of the players who played for those teams in that specific period of time.

5. What are the basic details and stats of a particular driver who was playing in a specific period of time?

This question is answered by the **Drivers** tab which provides an overview of the driver details.

6. Comparison of two or more driver's performance in terms of points and laps over a specific period of time.

This part is also answered by the **Drivers** tab which provides a comparison of drivers with a line chart based on points and laps.

A future enhancement to this project module could be to visualize the shift in trend of different drivers based on lap timings.

References

- [1] Ergast Developer API : <https://ergast.com/mrd/db/#csv>
- [2] Formula 1 : <https://www.formula1.com/>