F1 Data Analyzer Process Book

Overview and Motivation

Formula 1 is a sport deeply driven by data, where factors like driver performance, race strategies, and car engineering play a crucial role in determining outcomes. The Formula 1 dataset provides a comprehensive collection of race results, lap times, pit stops, and driver standings, making it an ideal resource for data-driven exploration. This project is motivated by a keen interest in sports analytics and the challenge of extracting meaningful insights from complex real-world data. By leveraging this dataset, we aim to analyze performance trends, optimize race strategies, and apply machine learning techniques to predict race outcomes, ultimately bridging the gap between data science and motorsports.

Related Work

- Formula 1 World Championship (1950 - 2024)
- Github Repository

Questions

For the new F1 fans, it is important to see the history and the previous F1 seasons as well.

Primary Question:

- How can we easily visualize and show the history of the F1 sport to the new fans?

Sub-Questions:

- What have been the teams and drivers with the most achievements?
- What trends can be found within the different drivers and teams with the most achievements?

These questions evolved from a broader historical performance analysis to a detailed, strategic insight into race-day decision-making.

Data

Sources:

- Formula 1 World Championship (1950 - 2024)
- Data used

Collection Methods:

- Directly accessed from provided open-source repositories such as Kaggle.

Cleaning Methods:

- Handling missing or inconsistent data:
    - Looking at the data we received from Kaggle files such as the Driver.csv, results.csv, etc, had missing values that we decided to replace with zero. In this case, we understand that the drivers and constructive team points might not be exactly correct due to the missing values.

- Encoding/Decoding driver and team identifiers
    - Since the data was distributed within different files, each row for each observation, such as driver names, constructive team names, race results, etc, has been encoded with IDs, which we had to merge the data throughout different files to achieve the data dimension and specification for each chart.

- Transforming data to structured formats for visualization
    - Since in our proposal we have specified the charts and the specific data that will be shown within the chart, we were able to construct and extract important variables needed for each chart. Operations such as left merge, group by, sum, and size were needed to reach the needed data.

	year	driverId	Driver Name	Total Points	Races Won	Races Qualified First
2	2007	1	Lewis Hamilton	109.0	4	5
3	2008	1	Lewis Hamilton	98.0	5	7
4	2009	1	Lewis Hamilton	49.0	2	4
5	2010	1	Lewis Hamilton	240.0	3	1
6	2011	1	Lewis Hamilton	227.0	3	1
7	2012	1	Lewis Hamilton	190.0	4	8
8	2013	1	Lewis Hamilton	189.0	1	5
9	2014	1	Lewis Hamilton	384.0	11	7
10	2015	1	Lewis Hamilton	381.0	10	11
11	2016	1	Lewis Hamilton	380.0	10	12
12	2017	1	Lewis Hamilton	363.0	9	11
13	2018	1	Lewis Hamilton	408.0	11	11
14	2019	1	Lewis Hamilton	413.0	11	5
15	2020	1	Lewis Hamilton	347.0	11	10
16	2021	1	Lewis Hamilton	385.5	8	8
17	2022	1	Lewis Hamilton	233.0	0	0
18	2023	1	Lewis Hamilton	217.0	0	1

	year	constructorId	Constructor Name	Total Points	Races Won	Races Qualified First
2	1968	1	McLaren	0.0	0	0
3	1971	1	McLaren	10.0	0	0
4	1972	1	McLaren	53.0	1	0
5	1973	1	McLaren	58.0	3	0
6	1974	1	McLaren	75.0	4	0
7	1975	1	McLaren	53.0	3	0
8	1976	1	McLaren	75.0	6	0
9	1977	1	McLaren	60.0	3	0
10	1978	1	McLaren	15.0	0	0
11	1979	1	McLaren	15.0	0	0
12	1980	1	McLaren	11.0	0	0
13	1981	1	McLaren	28.0	1	0
14	1982	1	McLaren	69.0	4	0
15	1983	1	McLaren	34.0	1	0
16	1984	1	McLaren	143.5	12	0

	Driver Name	Championships
2	Michael Schumacher	7
3	Lewis Hamilton	7
4	Juan Fangio	5
5	Max Verstappen	4
6	Alain Prost	4
7	Sebastian Vettel	4
8	Ayrton Senna	3
9	Jack Brabham	3
10	Niki Lauda	3
11	Jackie Stewart	3
12	Nelson Piquet	3
13	Jim Clark	2
14	Mika Häkkinen	2
15	Alberto Ascari	2
16	Graham Hill	2

	Driver Name	Total Points	Races Won	Pole Positions	Avg Lap Time
2	Lewis Hamilton	4820.5	105	107	96752.70828201211
3	Nick Heidfeld	259.0	0	1	92514.57860014432
4	Nico Rosberg	1594.5	23	30	98486.722695099
5	Fernando Alonso	2329.0	32	23	96208.60932704201
6	Heikki Kovalainen	105.0	1	1	100620.95779601407
7	Kazuki Nakajima	9.0	0	0	95103.3509127789
8	Sébastien Bourdais	6.0	0	0	95072.25470219436
9	Kimi Räikkönen	1873.0	21	19	93918.77925146325
10	Robert Kubica	274.0	1	1	95376.17329700272
11	Timo Glock	51.0	0	0	101818.09896820383
12	Takuma Sato	44.0	0	0	91514.21721677075
13	Nelson Piquet Jr.	19.0	0	0	94326.72842438638

Exploratory Data Analysis

The next thing on our to-do list is to replace the Demo data with the data that we have cleaned and decoded to demonstrate the data and explore the trends or insights that exist in our data.

Design Evolution

We employed the Five Design-Sheet Methodology in the charts that we have decided:

- **Bar Chart:** For the given season driver metrics, we will demonstrate total points, races won, and pole positions achieved by the driver.

- **Line Chart:** Team points, races won by the team, and pole positions achieved by the team.

- **Heatmap:** To enable comparative analysis of driver performance across multiple metrics in one view.

- **Radial Graph:** To visualize the distribution of driver and constructor championships across F1 history, showing dominance and historical trends.

The final design will combine clear visual encoding, interactivity for user engagement, and insightful comparative analysis.

Ideate:

F1 visualization

Points | Wins | Lap time

or

Points / Wins / Lap time / ...

Input 1

Pick your F1 season

2022
2023
2024 * click
2025

Filter:

Input 2

Pick your team

AMG Petronas
Haas
Red Bull
McLaren
Ferrari

or

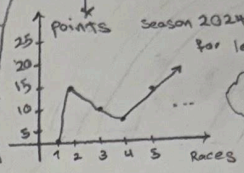
Team

Pick your Driver * click

Input 3

Hamilton
Max Verstappen
Carlos Sainz
...

Points



Season 2024 for Lewis Hamilto

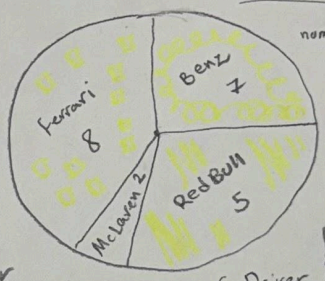
Generating...

Having names would be easier than having logos as an input.

Ferarri 86 wins
Red Bull 71 wins
Benz 54 wins
...

} over 1950 - 2024 years

Categorize:



number of championships won by a Team / this can Also be done for Drivers (Drivers with most championships)

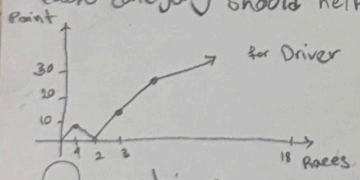
In History

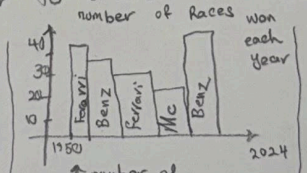
- Driver with most championships
- Teams with the most championships
- the fastest car in each season
- ...

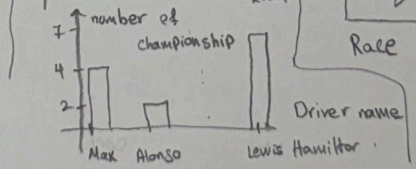
1. How we want to recieve the Input from User
2. Visualizations for a given Team
3. Visualizations for a given Driver
4. Visualization for the best records in history.

Combine & Refine:

Given different tabs or panels for each category should help the user to understan Data Better!





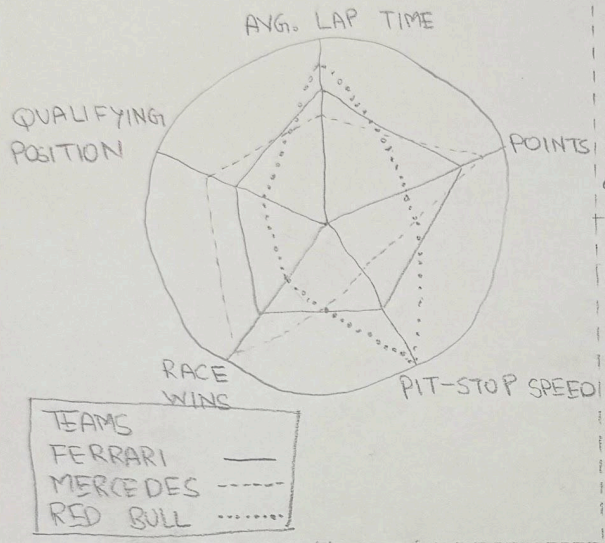


Race circuits around the World

Questions:

1. How do we want to Identify the Patterns between various Data?
2. What colors would make the most sense for each category?
3. How can we seperate each visualization between each category so the user can distinguish the differces?

LAYOUT



INFORMATION

We use a radial graph to show the F1 data here. We have variables like avg lap time, championship points, pit-stop speed, race wins and qualifying position and categorize it via team names using diff color codes.

OPERATIONS

We can have interactive buttons or checkboxes at the top of the chart. Toggling those removes / brings back variables to the radar chart.

Hovering on the chart should show the exact values and also explain the metric.

Clicking on a team, should open another chart showing only more details about that team.

FOCUS / PART I

→ Visualizes 5-6 metrics of the same type without overwhelming users

→ Highlights asymmetries in performance (eg: a team with fast laps but slow pit stops)

→ Radial axes can distort perception (eg: metrics on the right side may appear more prominent)

→ Requires normalization (eg: scaling lap times to a 0-100% range)

ALTERNATIVES CONSIDERED

Parallel Coordinates - too complex for casual users

Bar Charts - limited to 2-3 metrics without cluttering

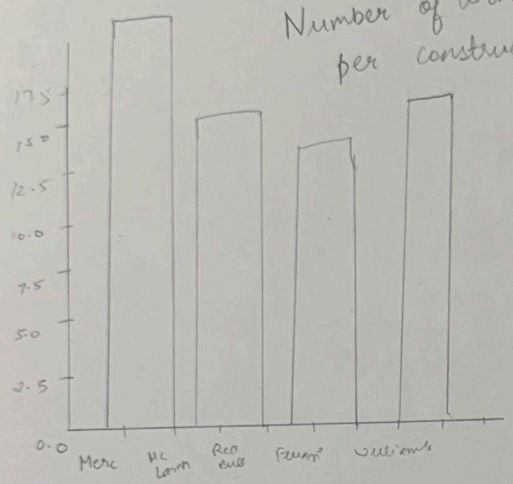
DISCUSSION

PROS:
Quick comparision
Compact design
Aesthetic appeal

CONS:
Cognitive load
Overlap issues
Normalization challenges

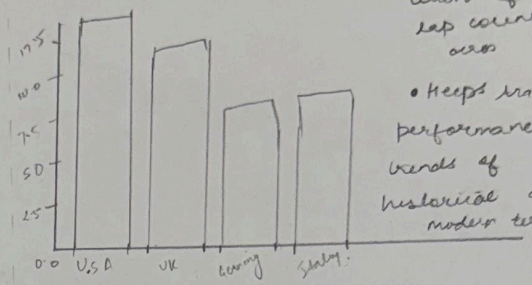
MITIGATIONS:
add a normal data toggle to show raw data

LAYOUT



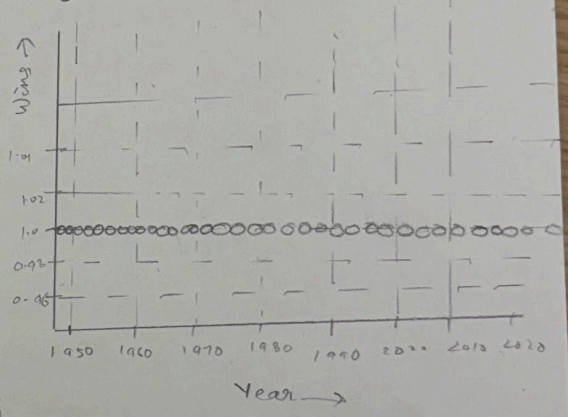
- This section presents the average fastest lap time achieved by different constructors over the years
- A lower lap time indicates a faster car

INFORMATION



- Displays the distribution of lap counts across
- Helps track performance trends of historical and modern teams

OPERATIONS



FOCUS

Understanding Race Pace vs Winning Strategies

- The relationship between fastest lap times and actual race wins.
- Identifying whether consistent performance is more valuable

Focus: Relationship between Pole Position

- The impact of starting position on race result
- Analyzing whether securing pole position leads to high race victories

DISCUSSIONS

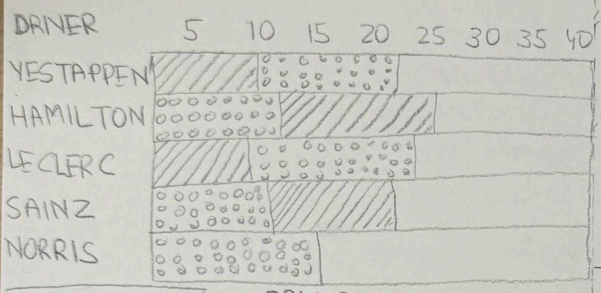
- Speed vs Strategy: Is it better to have the fastest car.
- Impact of Lap Count:

Do races with more laps require different strategies

- Role of Pits Stops: How do pit stop strategies influence race.

LAYOUT

F1 RACE STRATEGY TIMELINE



TIRE TYPES
- SOFT
- MEDIUM
- HARD

DRIVER STATISTICS

VESTAPPEN
Pit Stops - 2, Avg Lap - 1:32:456
HAMILTON
Pit Stops - 2, Avg Lap - 1:32:893
LECLERC
Pit Stops - 2, Avg Lap - 1:33:124
SAINZ
Pit Stops - 2, Avg Lap - 1:33:456

FOCUS / PART I

→ Replace complex charts with a grid based timeline for intuitive, side-by-side driver comparisons.

→ Encodings:

Color: Tire type (categorical) + lap time (sequential)

Symbol: Pit stops with duration

→ Minimalist design (no axes or complex shapes)

→ Combines temporal (lap progression), categorical (tire strategy), and quantitative (lap time) data in one view.

INFORMATION

We are using a heatmap to represent the data here. Each row represents a driver (eg: Verstappen, Hamilton) and each column represents laps. (0 to total laps). The cells are color coded cells like red for soft, yellow medium, black hard. We show all the key metrics in the side panel.

OPERATIONS

HOVER:

Over a cell → Tooltip shows exact lap time, tire type, and pit-stop duration.

Over a pit-stop marker → highlight all laps affected by that pit stop

FILTER:

Click a driver's row → gray out other drivers for focused comparision

DISCUSSION

PROS:

→ Easy to draw

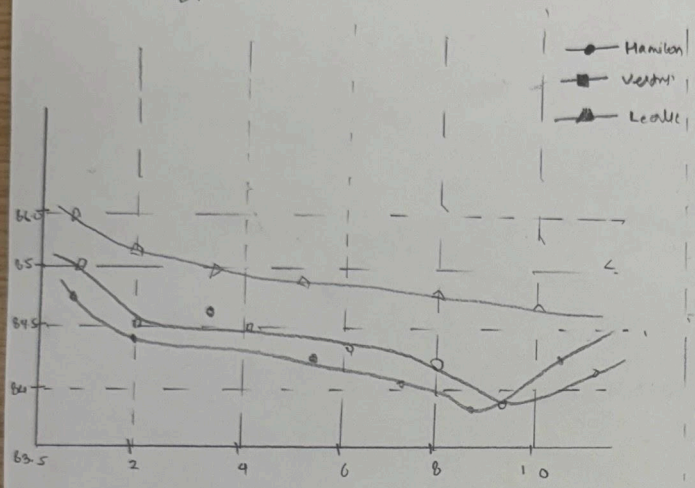
→ Comprehensive

CONS:

→ Limited scalability

→ Manual adjustments

LAYOUT



LAP TIME ACROSS LAPS FOR DIFFERENT DRIVERS

FOCUS / PART I

- Identify which driver has the fastest average lap time
- compare performance b/w drivers
- Trends in consistency and improvement
- Detect Anamolies

WHY THIS DESIGN?

- Allow easy comparison of driver performance over a race
- Helps analyze team strategies and efficiency of pit stops
- Provides an intuitive visual representation of race insights

INFORMATION

- Lap times of multiple drivers over a series of laps
- Performance comparison bet drivers
- Trends in consistency and pace improvement

OPERATIONS

① Click on a lap point in the graph: Expands the lap details to show telemetry data, including speed, acceleration

② Select a race from drop down: Dynamically update the dashboard to

③ Hover on a pit stop in a heat map: Display pop-up detail of the stop, including duration

④ Click on the leaderboard: Explained driver statistics

DETAIL.

Data Sources:

- FIAPI
- live data learning
- live timing data for real time visualization

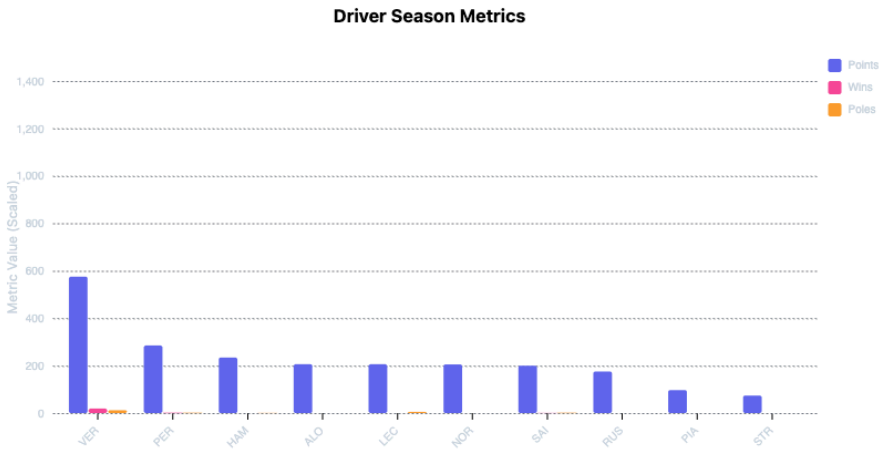
Technologies and Tools

- Frontend: D3J, React.js
- Python: (Django / Flask)
- Database: PostgreSQL

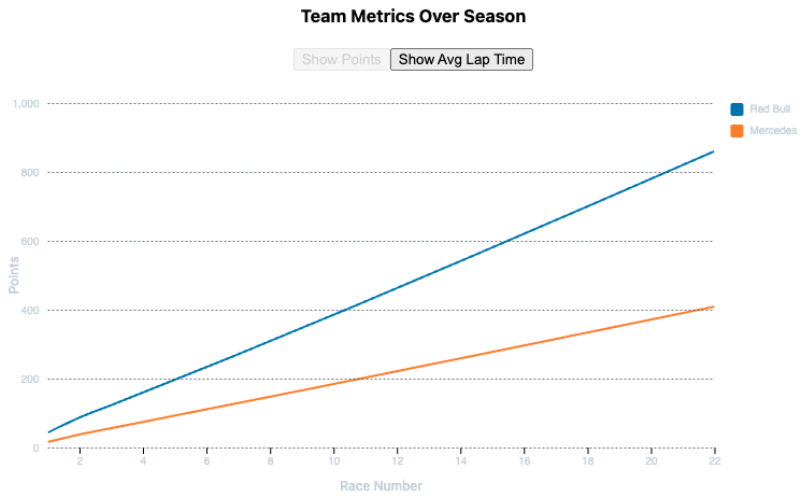
Implementation

Interactive visualizations developed with D3.js and JavaScript:

- **Bar Chart:** Interactive filters and tooltips.

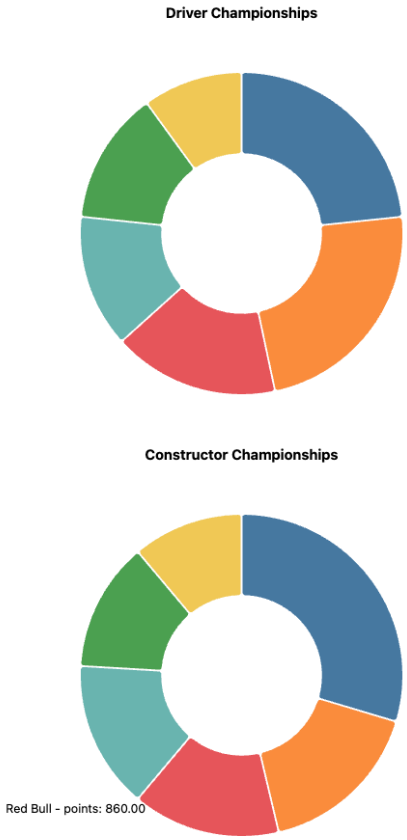


- **Line Chart:** Zoom, pan, toggle metrics.



- **Heatmap:** Hover interactivity and cell highlighting.

- **Radial Graph:** Hover details and historical drill-down.



Evaluation

To be done at the end.