GRANULAR ANALYSIS OF DRIVER PERFORMANCE METRICS BY RACE CHARACTERISTICS

FORMULA 1 DRIVER & RACE RESULTS 2014 to 2019

CU-VIRT-DATA-PT-12-2023-U-LOLC PROJECT3
GROUP 6

Mia, Elsie, Dan, Trevor
4/18/2024

EXECUTIVE OVERVIEW

SIGNIFICANCE

MOTOR RACING IS AN INHERENTLY RISKY ACTIVITY WITH HIGH RATES OF INCIDENTS on TRACK


ALL CRASHES F1 2022 SEASON

PROBLEM

Formula One racing continues to grapple with a **persistently high rate of incidents on track** which can lead to morbidity or mortality among drivers and other stakeholders.

Despite advancements in safety technology and regulations, these incidents remain a significant concern, posing risks to the well-being of drivers and the integrity of the sport.

SOLUTION

SOLUTION

Leverage accessible data and innovative tools to enhance safety awareness and mitigate risks in Formula One racing.

OUR SOLUTION: FORM1

A comprehensive safety hub

FORM1 is an is an interactive dashboard based on Formula One timing and static data.

Its principal aim is to provide an overview of which track characteristics affect driver performance, whilst allowing the user to probe the data and explore the differences in performance across track locations

OUR SOLUTION: FORM1

STORY BEING TOLD:

Story of DRIVER SAFETY IN TOP-CLASS MOTOR SPORTS

RELATIONSHIP BEING ILLUMINATED:

the relationship between DRIVER PERFORMANCE AND TRACK FEATURES

MANNER in which the STORY is told:

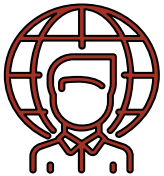
4 MAIN VISUAL ELEMENTS are utilized to illustrate the impact of extraneous variables on race results

UNIQUE VALUE PROPOSITION

How FROM1 competes

01: Democratizing Safety Data

Making data available to *everyone*



02: Proactive RISK MItigation

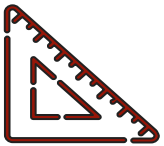
Live data allows us to forecast dangerous situations before they arise

03: COllaborative SAfety Solutions

Level the playing field so everyone stays safe

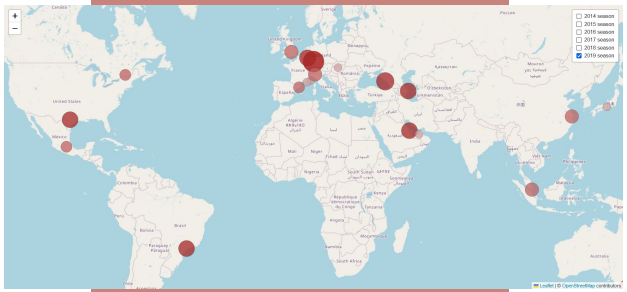
04: EMpowering DECision MAking

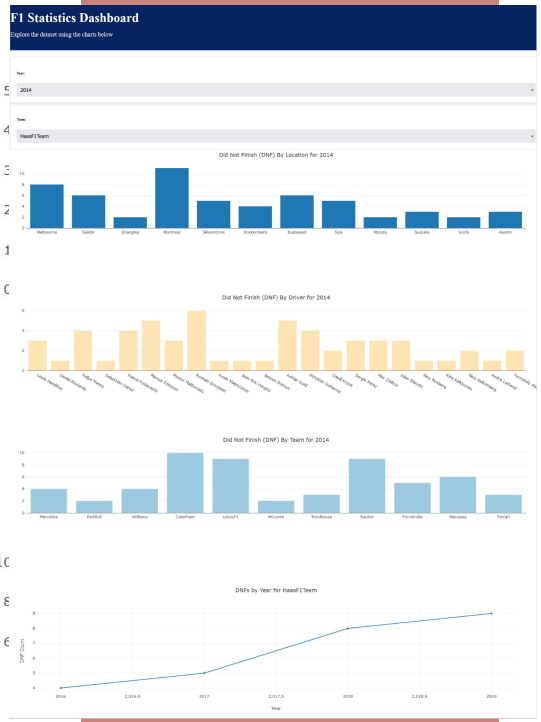
Give stakeholders the tools to prioritize safety without losing out on the competitiveness



THE MODEL

GEOMAP ELEMENT





INTERACTIVE MODELING HTML ELEMENT

PROJECT OVERVIEW

PROJECT 3

STORYTELLING USING DATA VISUALIZATIONS

Project 3
GOAL

Extract, transform, and load data using a database to tell a story by visualizing the data in different forms.

OUR Project 3
GOAL

Analyze Formula 1 driver performance metrics over time across all track locations to illustrate results and determine the impact, if any, of extraneous variables on race results

PROJECT COMPLETION WORKFLOW

STAGE 1

Explore a problem within a scope (risk management and safety in motorsports), identify a need, & determine if project application offers solution to problem

STAGE 3

Write code to present data via multiple visualizations

INTENTION

Apply what's been learned module 1 - module 15 with a refined scope of interest (finance, healthcare, custom) to tell a story via data visualizations

STAGE 2

Find data, fetch it and load into a database, clean the data, and combine the data for use by all team members

STAGE 4

Determine if insights can be made based on presentation of data

DATA EXPLORATION & WORKFLOW DELEGATION

How we split up the work.

MIA METNI

PROJECT ITEMS:

- Find F1 data set to import to Postgresql database
- Build map visual showing track location and race info
- Design presentation

TREVOR

PROJECT ITEMS:

- Find a new library to import Postgresql tables to a Python file for additional data transformation
- Create graph showing total DNFs filtered by team and driver over time

ELSIE

PROJECT ITEMS:

- Create ERD for Postgresql database
- Convert Postgresql tables to JSON files using Python
- Create track location vs. DNF graph

DAN

PROJECT ITEMS:

- Find and import track location and weather data using an APIs
- Create weather vs. track location graph
- Help build map visual

DATA & DELIVERY

DATA SOURCING

While there is a wealth of public F1 race data, most comes from one source: the Ergast API, an unofficial collection of race info giving race and driver metrics.

Our main data set came from the F1 Timing Data GitHub repo, an F1 fan's own collection of data from 2014-2019 using additional accident and failure data not from the Ergast API.

Weather data came from Open Weather Map's One Call API 3.0, which lets users call historical weather data based on timestamps.

IMPLEMENTATION, APPLICATION, & VALIDATION

DATA CLEANING

Our original data were CSVs and came from an API. For the API, we imported the data using Python, cleaned it, and exported it to a CSV.

Using Postgresql, we created a database and created tables for each CSV, then imported the files. We then did additional data cleaning and joined tables to create a main table and individual tables, which we also exported to CSVs.

We connected the database to Python using a new library we found online called psycopg2 (we did this in Jupyter Notebook). In Python, we cleaned data and exported JSON files to use to create our visuals using JavaScript.

DATA USAGE

1. Once data is finalized and new csv is saved.
2. Read data from new CSV file into a Pandas DataFrame.
3. Converts each row of the DataFrame into a dictionary and appends it to a list of dictionaries.
4. Converts the list of dictionaries into a JSON string and writes it to a JSON file.
5. Reads the JSON data from the file and writes it to another JSON file .
6. Loads the JSON data from the original file and pretty-prints it.
7. Writes the pretty-printed JSON data to a new file .
8. These steps were repeated several time to create smaller datasets specific to each members part.

This workflow involves data conversion and storage using Pandas and JSON files.

MAP ELEMENT

We designed a map using JavaScript and Leaflet to mark all track locations across the world by season. Each marker has a pop up that gives location info as well as some information on the race for the selected season.

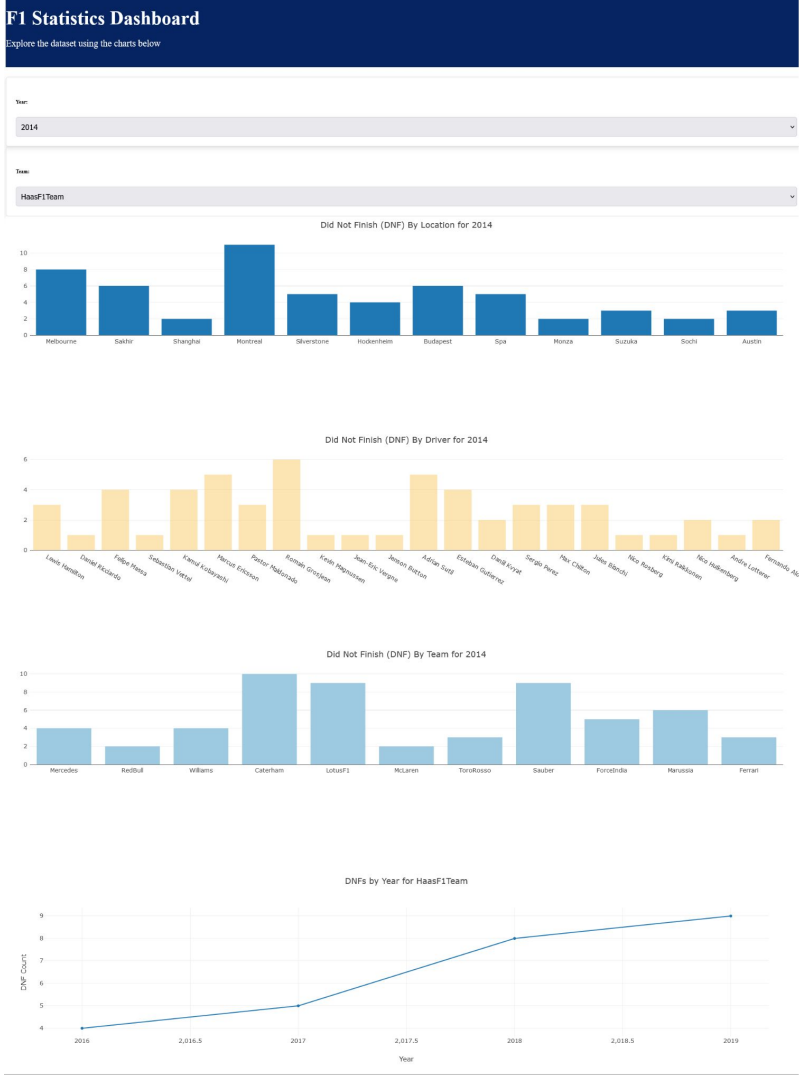
WEB GRAPHICS ELEMENT

We used JavaScript to create interactive graphs that allow the user to see the relationship between F1 track per season and different metrics, such as the Did Not Finish (DNF) status of a driver or the weather of the track on race day.

WEB GRAPHICS ELEMENT

- Historical Data for Each Season to Identify DNFs for each race

- Filter by Location, Team, or Driver



FINDINGS

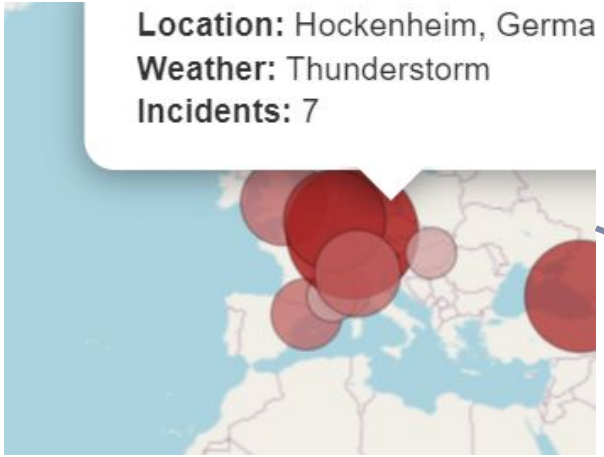
VISUAL ELEMENT: MAP

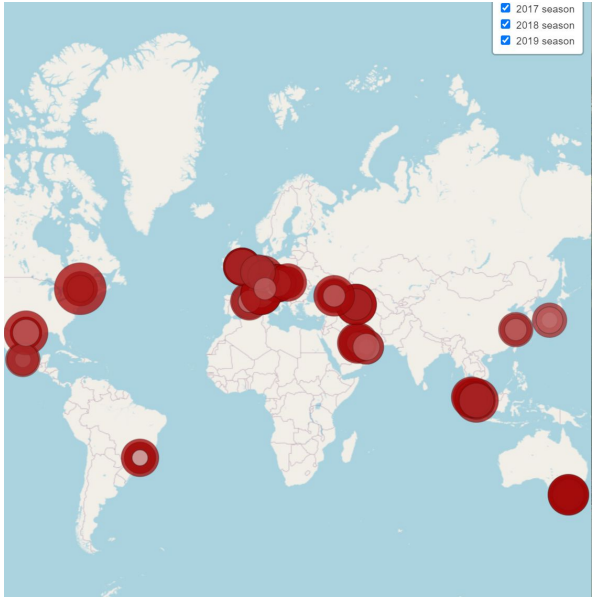
VISUAL ELEMENT: MAP CODE

```
// Function to calculate marker color based on the number of incidents
function calculateColor(incidents) {
    // Define color gradient based on the number of incidents
    // Modify the gradient as needed
    if (incidents > 5) {
        return '#a30505';
    } else if (incidents > 3) {
        return '#a62929';
    } else if (incidents > 1) {
        return '#bf6363';
    } else {
        return '#cfa1a1';
    }
}
```

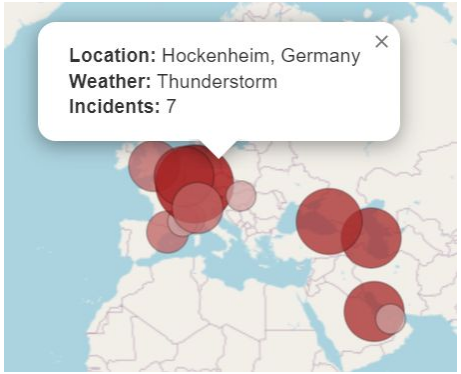
```
// Function to calculate marker radius based on the number of incidents
function calculateRadius(incidents) {
    // Define a scaling factor for the radius
    // Modify the factor as needed
    let scaleFactor = 10;
    return Math.sqrt(incidents) * scaleFactor;
}
```

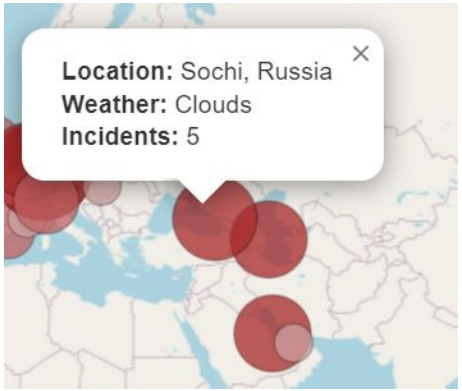
MAPPING: RELATIONSHIP BETWEEN DRIVER PERFORMANCE AND TRACK LOCATION





MAPPING: RELATIONSHIP BETWEEN DRIVER PERFORMANCE AND TRACK LOCATION AND WEATHER CONDITIONS



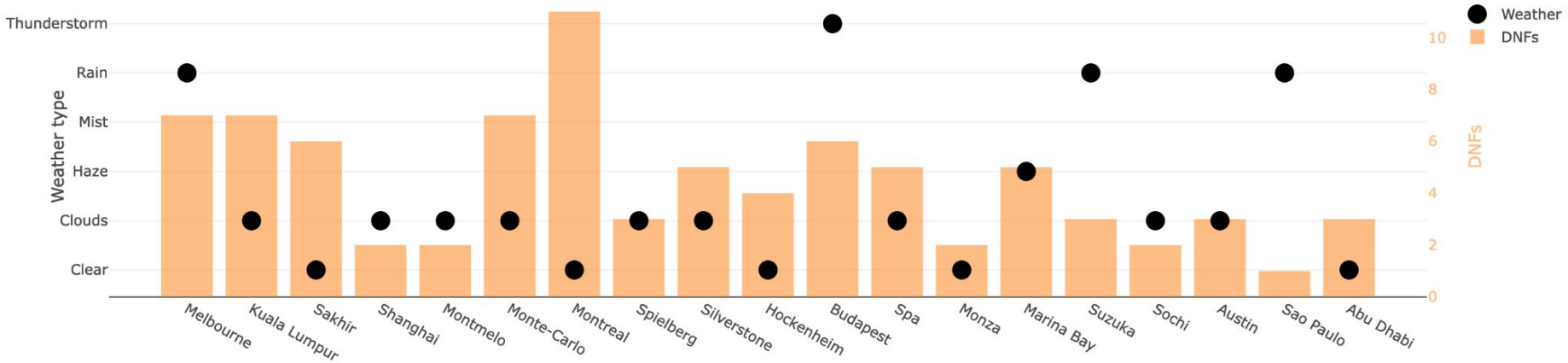


HTML: WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS

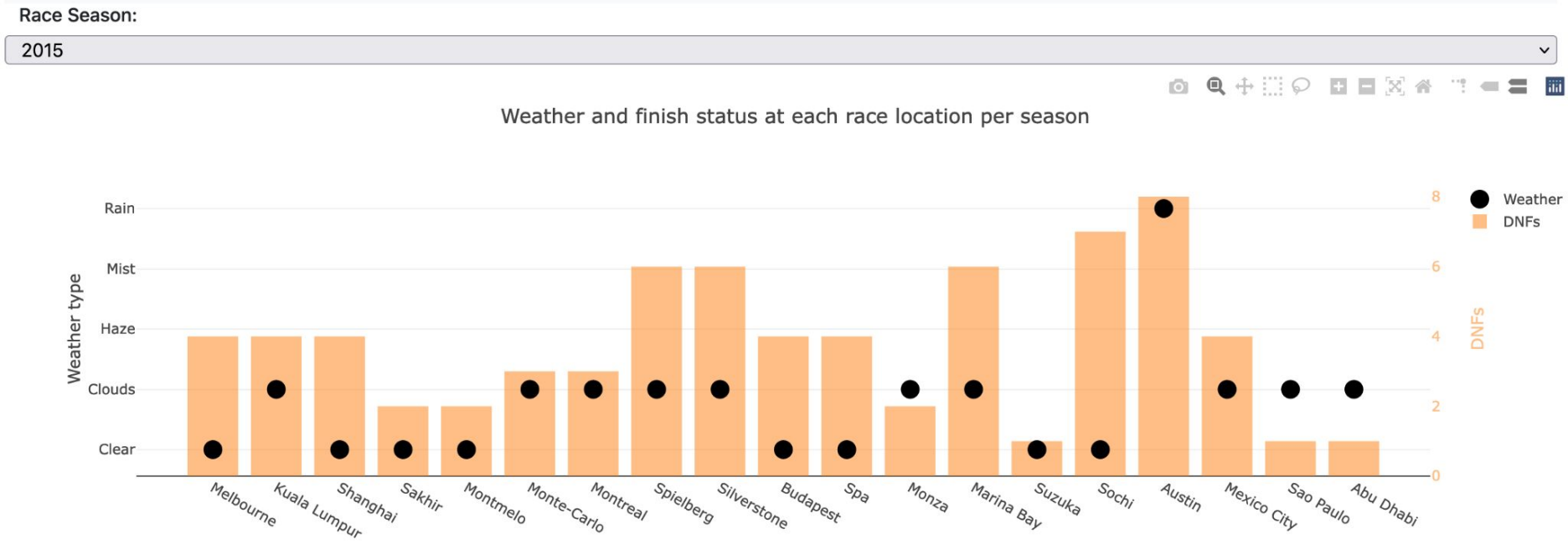
Race Season:

2014



HTML: WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS

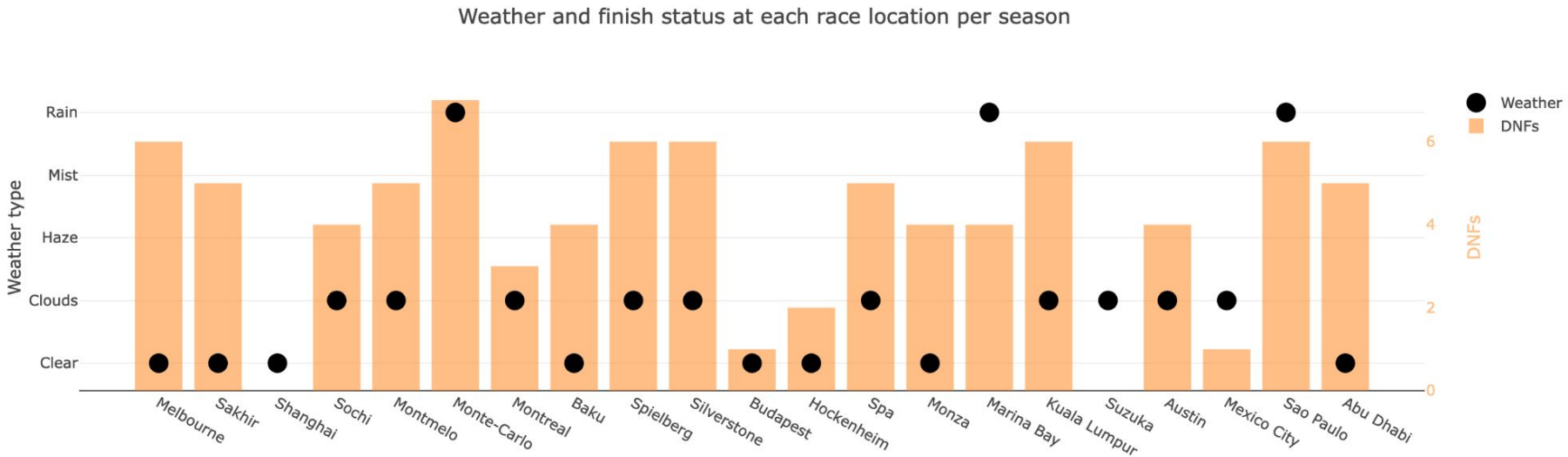


HTML: WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS

Race Season:

2016

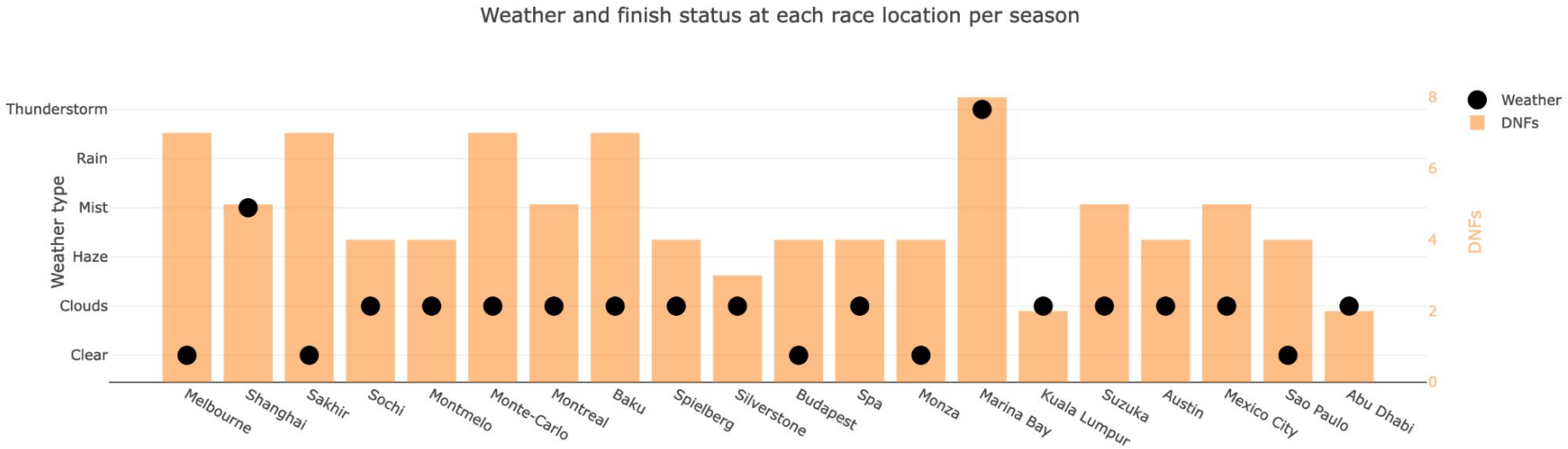


HTML: WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS

Race Season:

2017

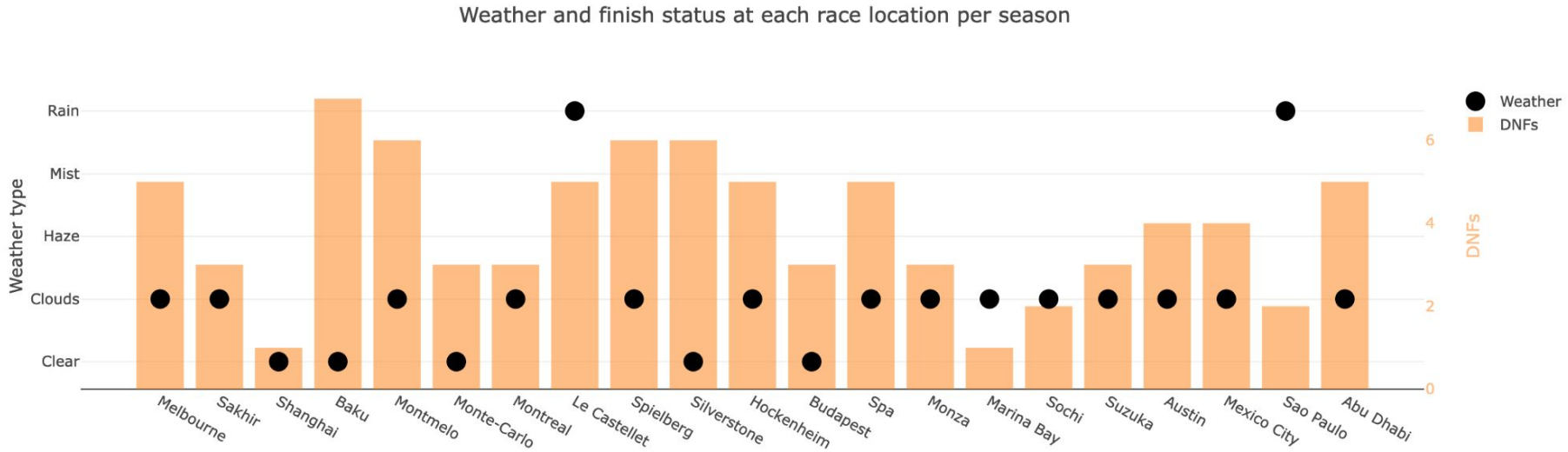


HTML:
WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS

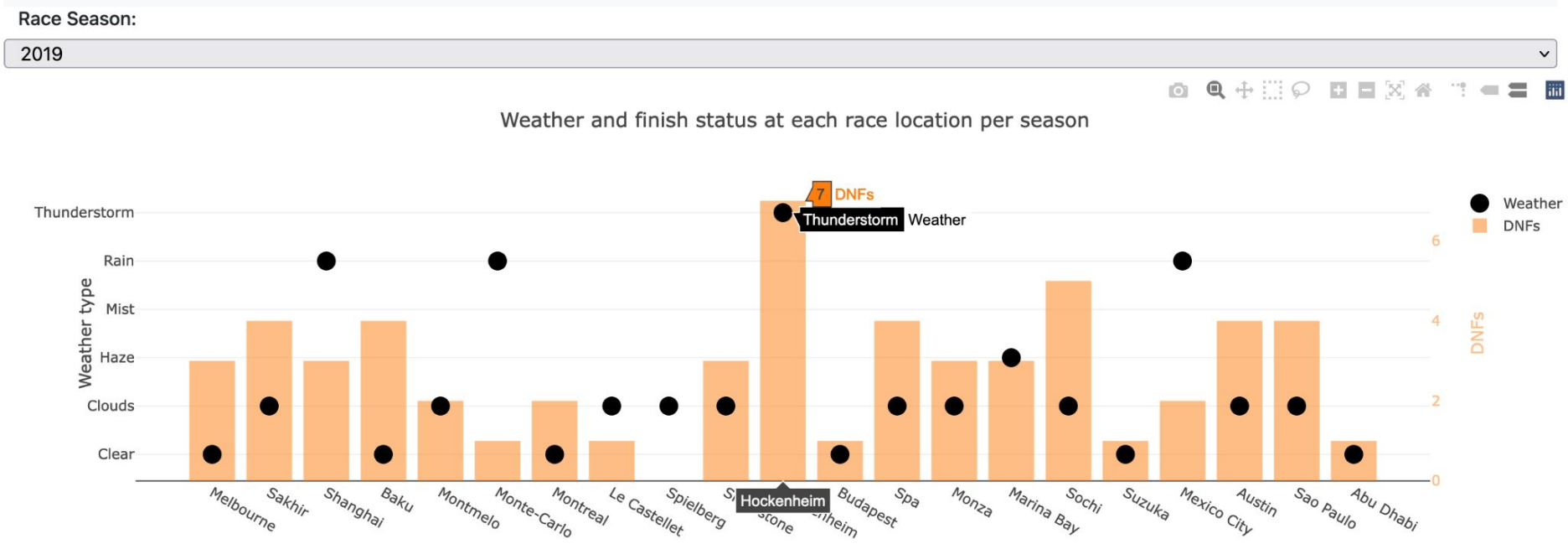
Race Season:

2018



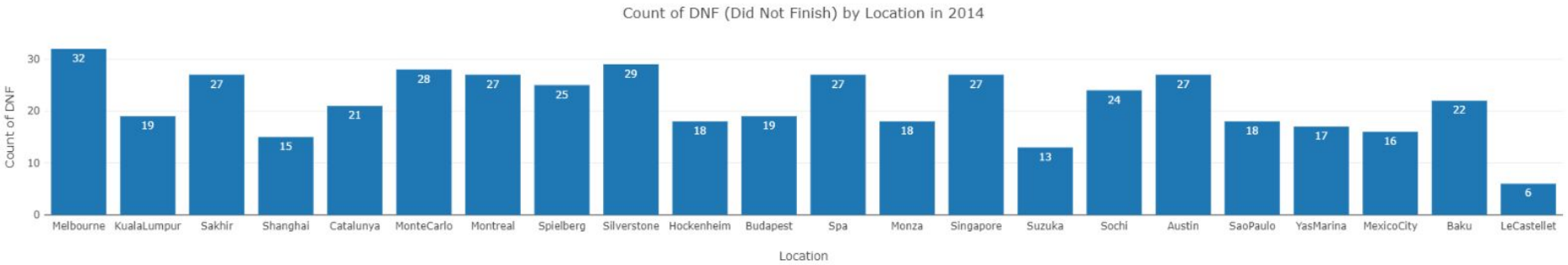
HTML: WEATHER PER TRACK

POSSIBLE RELATIONSHIP BETWEEN RATE OF INCIDENCE OF INJURY AND TENDENCY FOR LOCATION TO HAVE PARTICULAR WEATHER CONDITIONS



DNF by Location

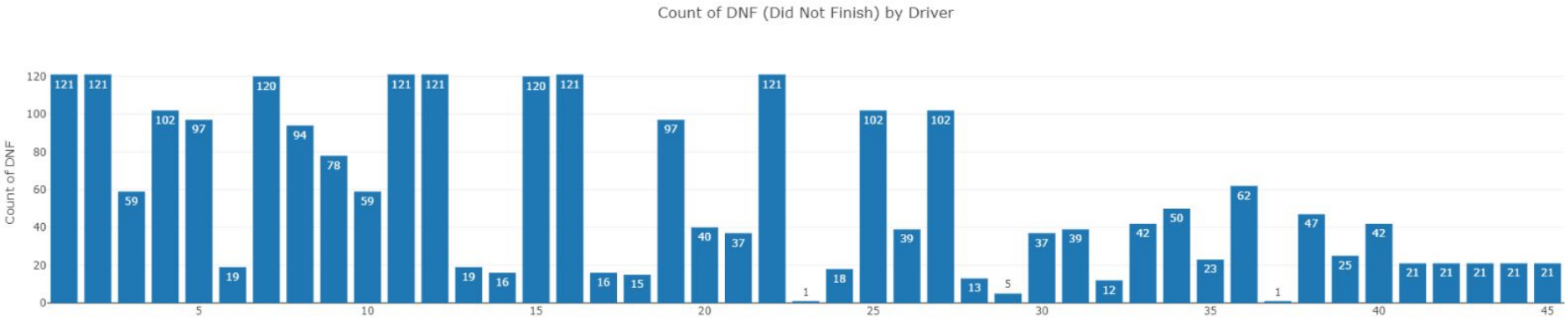
What is driving the DNF?



DNF by Driver

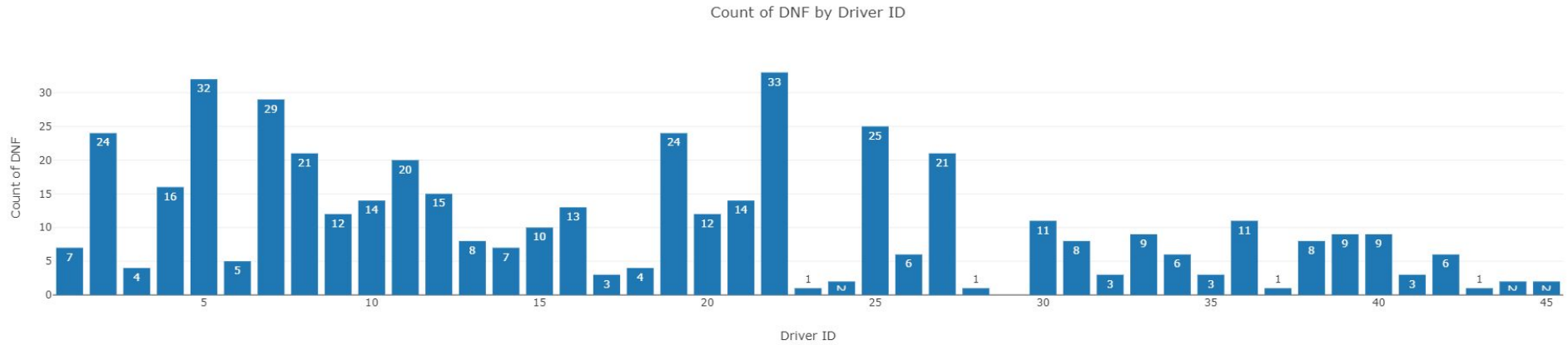
Limitation of data
DNF data provided by seasons
Lack of location Driver_id
Merging tables resulted in incorrect multiplication of data points

Recommendation
Consider looking at the granular data



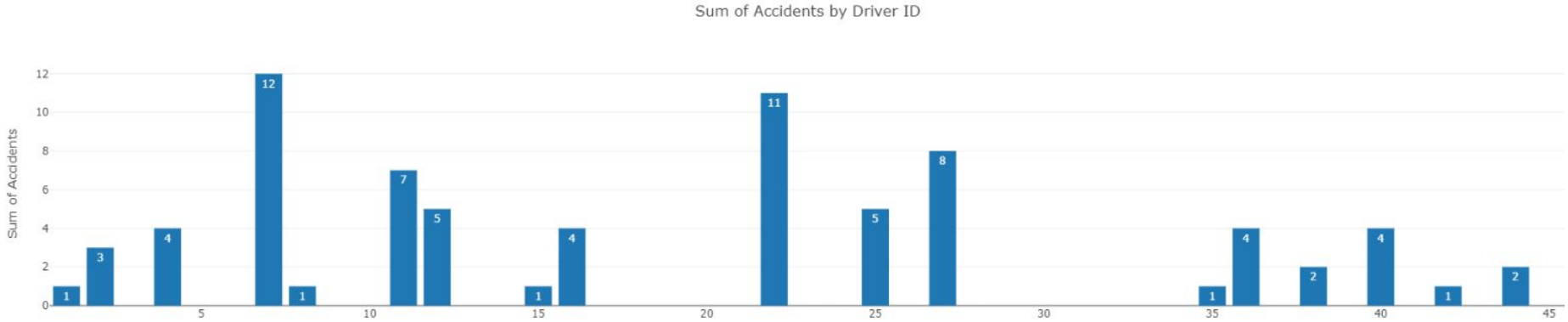
DNF BY INDIVIDUAL DRIVER

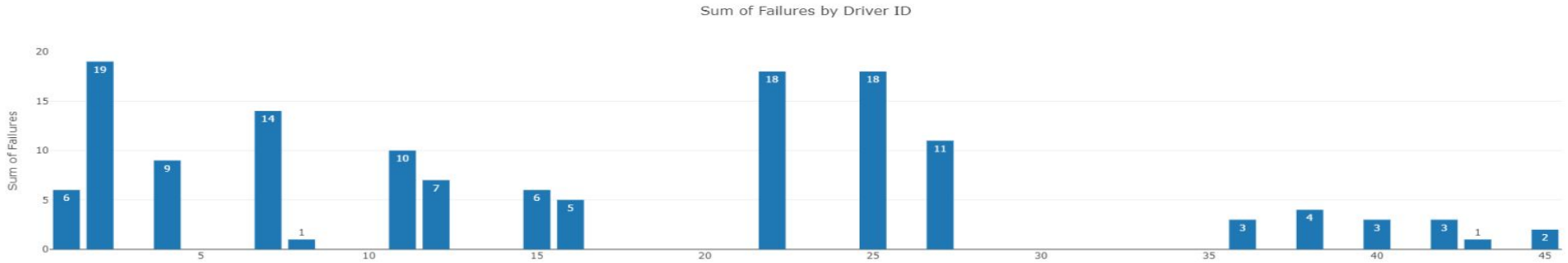
While accidents and failures account for a significant portion of DNF incidents, they don't paint the full story.



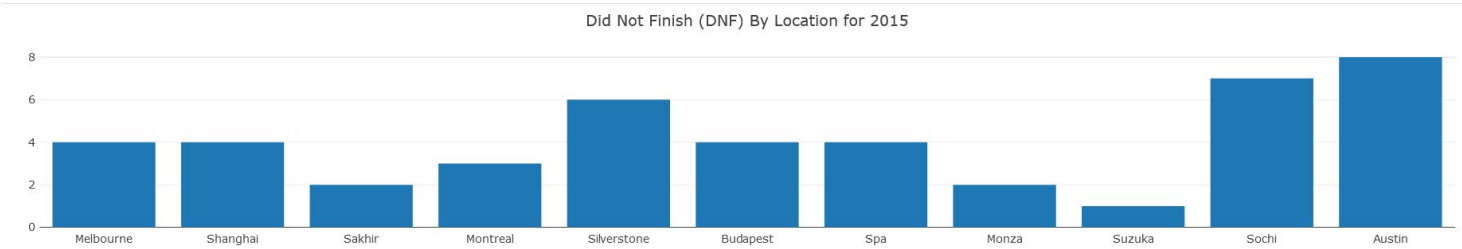
RELATIONSHIP BETWEEN INCIDENCE OF FAILURES AND ACCIDENTS PER DRIVER

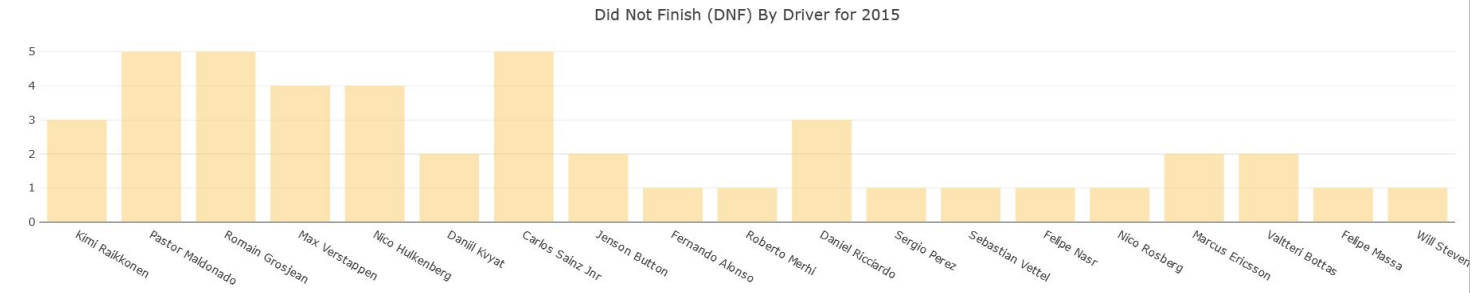
Despite efforts to analyze DNF, we are unable to conclude on the causes of DNF and any correlation.

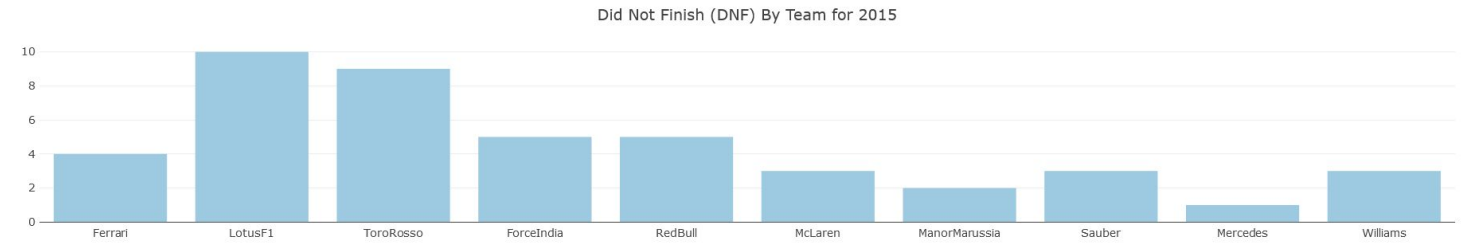




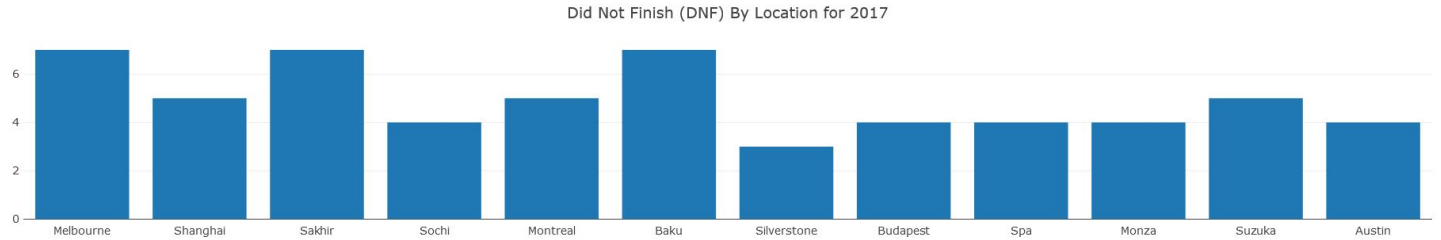
HTML: DNF DATA FOR 2015

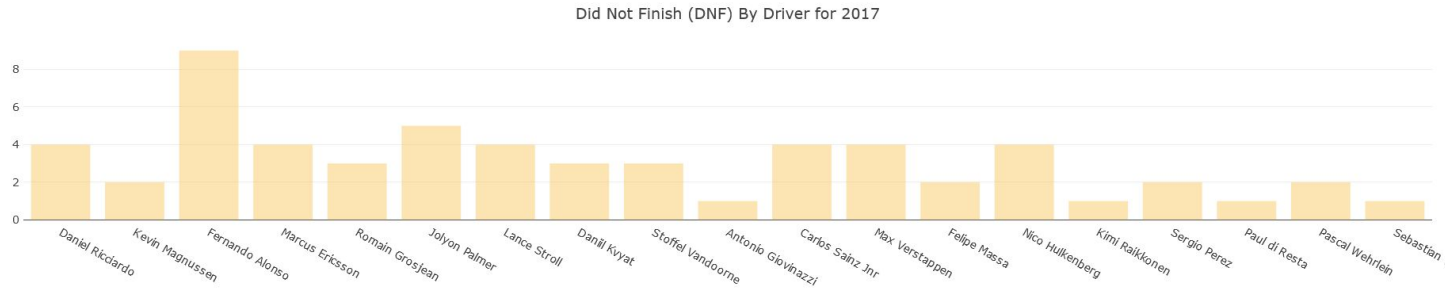


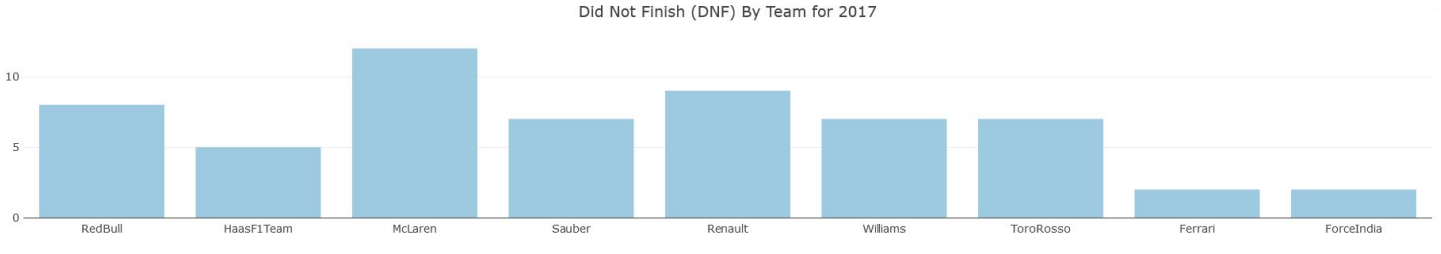




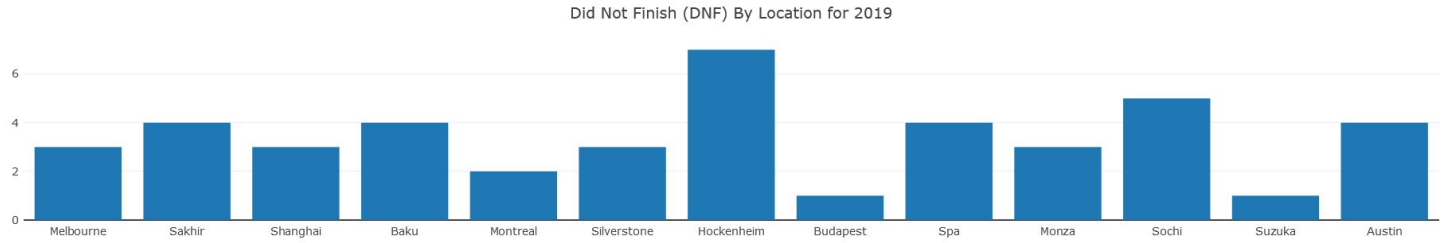
HTML: DNF DATA FOR 2017

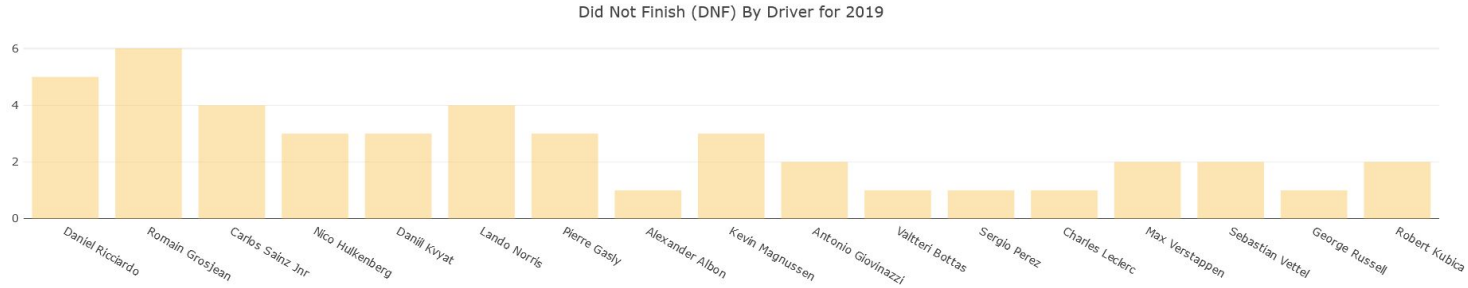


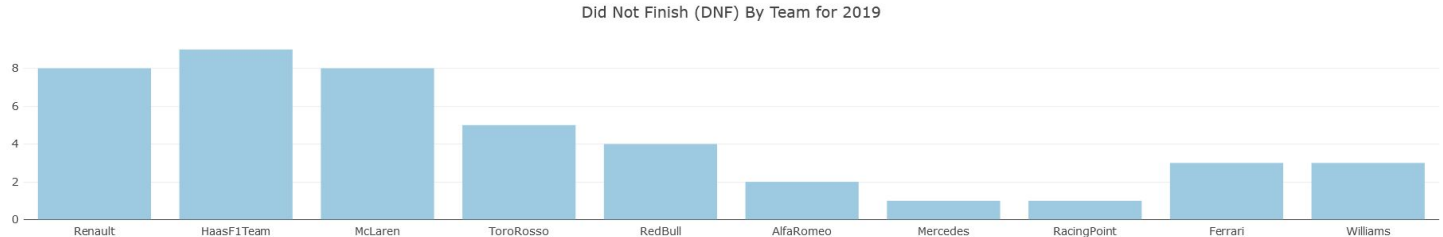




HTML: DNF DATA FOR 2019



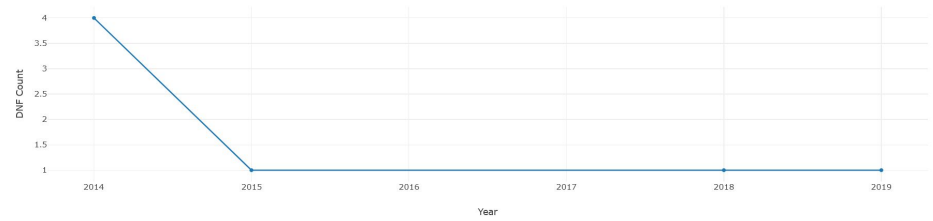




HTML: Yearly DNFs for the Biggest Teams in F1

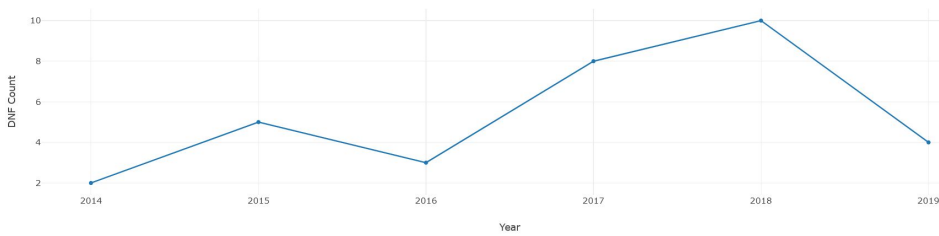
Team:

Mercedes



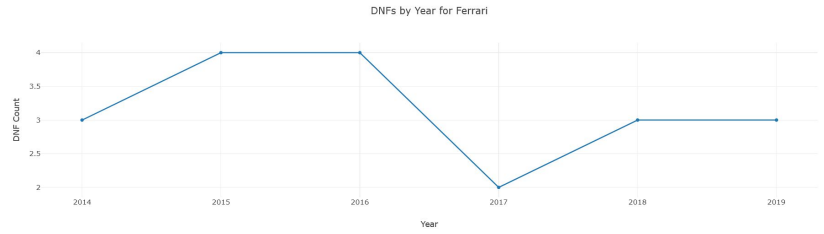
Team:

RedBull



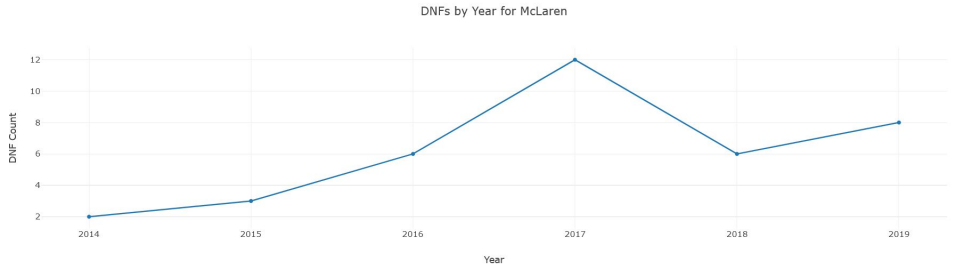
Team:

Ferrari



Team:

McLaren



TAKEAWAYS

FORM1 serves to tell the story of driver safety in top-class motor sports by providing a comprehensive overview of how different track characteristics influence driver performance across various Formula One circuits.

FORM1
SO YOU CAN DRIVE AND SURVIVE

C YA