

# Midterm – Project Proposal Outline

**Name:** *Matthew Godbold*

**Project Title:** *Data Fueled Insights: Analyzing Formula 1 Through Interactive Visualizations*

## **I. Introduction / Background**

*Formula 1 has been renowned as the pinnacle of motorsports for the past 70 years. Its single seat, open-wheel, open cockpit, high-tech race cars are a marvel of engineering, both mechanically and aerodynamically. However, while many fans appreciate the speed and skill of the drivers, few are aware of the incredible amount of data collected from the cars, tracks, drivers, and pits.*

*In fact, the sheer amount of data collected is a masterful conduction of data collection and correlation. By communicating this data to its audience, Formula 1 can offer fans a deeper understanding of the sport, allowing them to explore their favorite drivers, constructors, and tracks in a way incomparable to any other sport. With so much data available, it's no surprise that fans are eager to get their hands on it. By creating a data visualization dashboard, fans can explore and analyze the data, gaining valuable insights into the performance of their favorite drivers and constructors. From lap times and tire wear to fuel consumption and engine performance, the possibilities are endless. By providing fans with this level of insight, Formula 1 is not only enhancing the fan experience, but also providing valuable data that can be used to improve the performance of the cars and drivers themselves.*

## **II. Audience**

*Formula 1 has a global fanbase of nearly 1.55 billion and has the highest YoY increase of any sport at nearly 30%. It also crowns the sports entertainment industry for social media engagement with an active following of 50 million users. This is made possible by the social media teams, utilizing data in an easily comprehensible way. To give fans centralized data visualizations, I am building a dashboard for interactive visualizations.*

## **III. Dataset(s)**

**Source:** <https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?select=status.csv>

**Description:**

*The data set consists of thirteen CSV files (20.14 MB) containing the following features:*

Circuits	
<i>ID</i>	<i>77</i>
<i>Circuit Reference</i>	<i>77</i>
<i>Name</i>	<i>77</i>
<i>Location</i>	<i>77</i>
<i>Country</i>	<i>77</i>
<i>Latitude</i>	<i>77</i>

Constructor Results	
<i>Constructor</i>	<i>12.2K</i>
<i>Results ID</i>	<i>12.2K</i>
<i>Race ID</i>	<i>12.2K</i>
<i>Constructor ID</i>	<i>12.2K</i>
<i>Number of Points</i>	<i>12.2K</i>

Constructor Standings	
<i>Constructor</i>	<i>12.9K</i>
<i>Standings ID</i>	<i>12.9K</i>
<i>Race ID</i>	<i>12.9K</i>
<i>Constructor ID</i>	<i>12.9K</i>
<i>Number of Points</i>	<i>12.9K</i>
<i>Position Number</i>	<i>12.9K</i>
<i>Position Text</i>	<i>12.9K</i>
<i>Number of Wins</i>	<i>12.9K</i>

Constructors	
<i>Constructor ID</i>	<i>211</i>
<i>Constructor</i>	<i>211</i>
<i>Reference</i>	<i>211</i>
<i>Name</i>	<i>211</i>
<i>Name</i>	<i>211</i>
<i>Nationality</i>	<i>211</i>
<i>URL</i>	<i>211</i>

Driver Standings	
<i>Driver Standings</i>	<i>33.9K</i>
<i>ID</i>	<i>33.9K</i>
<i>Race ID</i>	<i>33.9K</i>
<i>Driver ID</i>	<i>33.9K</i>
<i>Number of Points</i>	<i>33.9K</i>
<i>Position Number</i>	<i>33.9K</i>

Drivers	
<i>Driver ID</i>	<i>857</i>
<i>Driver Reference</i>	<i>857</i>
<i>Number</i>	<i>857</i>
<i>Driver Number</i>	<i>857</i>
<i>Driver Code</i>	<i>857</i>
<i>First Name</i>	<i>857</i>
<i>Last Name</i>	<i>857</i>
<i>Date of Birth</i>	<i>857</i>
<i>Nationality</i>	<i>857</i>
<i>URL</i>	<i>857</i>

Lap Times	
<i>Race ID</i>	<i>538K</i>
<i>Driver ID</i>	<i>538K</i>
<i>Lap Number</i>	<i>538K</i>
<i>Position Number</i>	<i>538K</i>
<i>Time</i>	<i>538K</i>
<i>Time in ms</i>	<i>538K</i>

Pit Stops	
<i>Race ID</i>	<i>9634</i>
<i>Driver ID</i>	<i>9634</i>
<i>Stop Number</i>	<i>9634</i>
<i>Lap Number</i>	<i>9634</i>
<i>Time of Pit Stop</i>	<i>9634</i>
<i>Duration</i>	<i>9634</i>
<i>Duration in ms</i>	<i>9634</i>

Qualifying	
<i>Qualify ID</i>	<i>9575</i>
<i>Race ID</i>	<i>9575</i>
<i>Driver ID</i>	<i>9575</i>
<i>Constructor ID</i>	<i>9575</i>
<i>Car Number</i>	<i>9575</i>
<i>Position Number</i>	<i>9575</i>
<i>Q1 Time</i>	<i>9575</i>
<i>Q2 Time</i>	<i>9575</i>
<i>Q3 Time</i>	<i>9575</i>

Races		Results	
Race ID	1102	Result ID	25.8K
Year	1102	Race ID	25.8K
Round	1102	Driver ID	25.8K
Circuit ID	1102	Constructor ID	25.8K
Name	1102	Car Number	25.8K
Date	1102	Grid Position	25.8K
Time	1102	Final Position	25.8K
URL	1102	Position Text	25.8K
		Position Order	25.8K
		Number of Points	25.8K

#### **Foreseen Issues:**

*At this point in the project, there are no foreseen issues. All data appears to be cleaned and prepared for visualizations.*

#### **IV. Proposed Visualizations**

*description of visualizations you intend to create*

*describe any interactivity that you intend to include*

*include example images for the **types** of visualizations*

*example images can come from unrelated datasets and sources*

*cite any images you include from external sources (if it's from a website, the URL is fine)*

*Formula 1 is a unique sport in that the competition is multifaceted. There are two main competition levels: the driver and the constructor. On the driver level, there are twenty drivers that compete for points (earned based on finishing position) that go towards winning the Driver's Championship. On the other hand, the ten constructors (comprised of two drivers each) compete for team points which are based on the points earned by your two drivers. For this reason, I plan to create a dashboard that is separated into two main pages, one for drivers and one for constructors.*

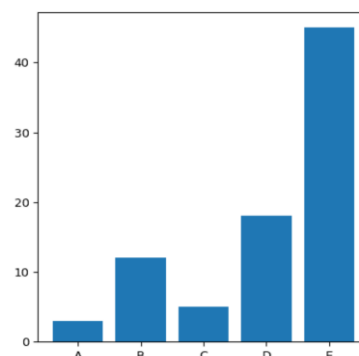
#### **Driver Visualizations:**

##### *Driver Success*

**Description:** A bar chart comparing the driver vs. their wins and points per grand prix.

**Interactivity:** The user will be able to select the driver of their choice.

**Source:** <https://www.python-graph-gallery.com/barplot/>

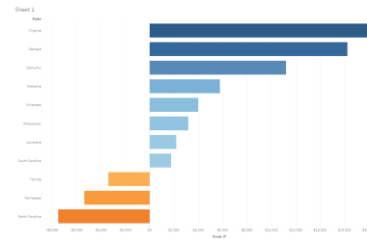


### Driver Conversions

**Description:** A diverging bar chart showing the drivers conversions per grand prix.

**Interactivity:** The user will be able to select the driver.

**Source:** <https://www.thedataschool.co.uk/timothy-manning/make-clean-diverging-bar-chart-tableau-tips-tableautimothy>

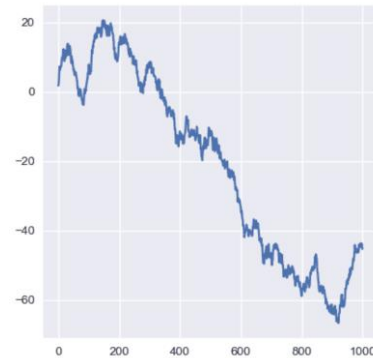


### Driver Position

**Description:** A line chart will be used to show the drivers lap time per lap and position per lap for each grand prix.

**Interactivity:** The user will be able to select the driver, lap time or position, and the grand prix.

**Source:** <https://www.python-graph-gallery.com/line-chart/>

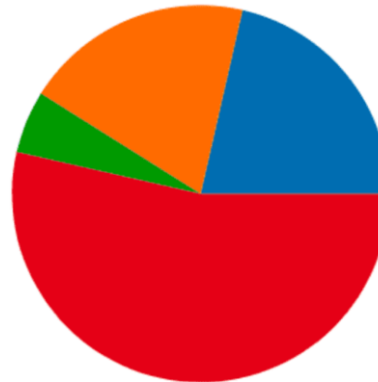


### Pit Duration

**Description:** A pie chart will be used to show the time the driver spent in the pit during the total duration of the race.

**Interactivity:** The user will be able to select the driver and grand prix.

**Source:** <https://www.python-graph-gallery.com/pie-plot/>



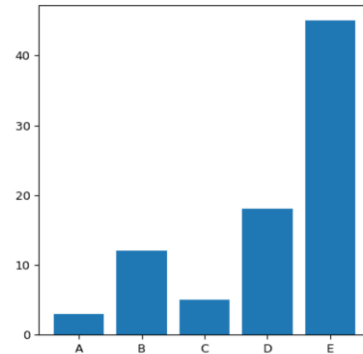
## Constructor Visualizations:

### Constructor Success

**Description:** A bar chart comparing the constructor wins and points vs other constructors per grand prix or all time.

**Interactivity:** The user will be able to select the constructor, grand prix or all time, and if they view wins or points.

**Source:** <https://www.python-graph-gallery.com/barplot/>



### Mapping the Team

**Description:** Map placing the nationality of each of its drivers.

**Interactivity:** The user will be able to select the constructor of their choice.

**Source:** <https://www.python-graph-gallery.com/map/>

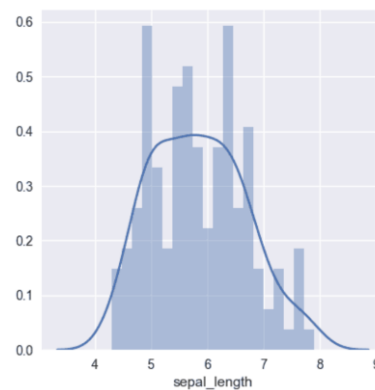


### Points Garnered

**Description:** A histogram showing which races garnered the most points for the team in a season or all time.

**Interactivity:** The user can select the constructor and the year or all time.

**Source:** <https://www.python-graph-gallery.com/histogram/>

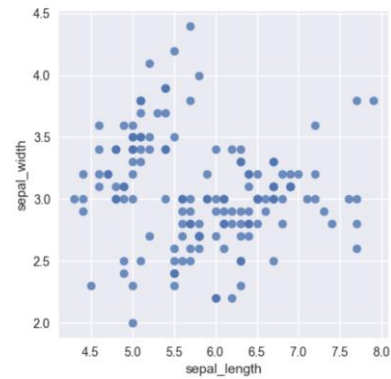


### Driver's Over Time

**Description:** Scatter plot showing the constructor's driver positions overtime.

**Interactivity:** User is able to select the constructor.

**Source:** <https://www.python-graph-gallery.com/scatter-plot/>



## V. Plan

### Plan A:

To build the visualizations, Python will be employed along with several common packages namely: Polars, Matplotlib, Plotly, and Seaborn. Streamlit will be used to build the dashboard and all user capabilities.

### Plan B:

If things don't go to plan, Tableau will be used to create the visualizations and dashboard. This is the backup because Tableau provides an easy to use, all-in-one tool for creating the application.