**Wil Noble**

**CIDM 4310**

**Dr. Babb**

**6/28/21**

**Proposal:**

Initially for my case I chose the banking industry as a point of interest. I have since decided to do a complete 180 and change my case completely. Now that I know more about the capabilities of BI and DSS, and the sources of data needed to create a BI/DSS, I have decided to change my case.

With the issue of data availability made apparent during my search for it, I had to rethink what I could do. While browsing for various APIs, I came across <http://ergast.com/mrd/>. Ergast is an API that allows developers access to a database of continually updated Formula 1 statistics from the start of the championship in 1950 to the current championship. I chose this dataset because I really enjoy Formula 1 and because it has sources of both dynamic and static data.

For my static data, I plan to pull and upload all the qualifying lap time data from Ergast and upload it to my VPS for use in answering questions within Jupyter Notebook. For my dynamic data, I plan to utilize python to setup a scheduled script to pull the most current race results and utilize the data with Jupyter Notebook to answer the questions I have. The race results data is not the most dynamic data ever, at most it would be updated once a week. However, I believe the concepts will still apply.

**Questions to be addressed:**

* Standings of the driver’s championship for the current championship
* Standings of the constructor’s championship for the current championship
* Lap times for the most current race compared by lap for each driver
* Comparison of qualifying lap times throughout the years

**Tools and Techniques:**

For the visualization of the driver and constructor standings, I believe utilizing Pandas to create bar graphs would be the best course of action. As for the lap time comparisons, I believe using pandas to create a line graph would be the best approach. I think having a different color for the line based on the constructor and having the driver’s name abbreviation under each dot could be the most readable. For the comparison of the qualifying times throughout the years, I believe this one will be the most difficult to quantify and visualize. Throughout the years, the format of qualifying has changed, circuits have changed, and weather could factor in. For these reasons, I believe it would be best to take and average of the fastest times at circuits that have remained on the calendar since the advent of qualifying. I believe the best way to visualize it would be a dot plot using Pandas.