# Group F Topic Proposal

**Title of the project: Formula 1 Analytics (Group F)**

**Project Participants**

- Yu Cheng | yc3984@columbia.edu
- Christa Chiao | cjc2297@columbia.edu
- Lauren Deitz | lmd2231@columbia.edu
- Nikki Gerjarusak | sg4011@columbia.edu
- Donald Stephens | ds3478@columbia.edu

**Brief description of the proposed visualizations / analyses**

Our questions are broken out by category below. We will not try to answer *all* of these questions, but this is a sense of what is possible.

**Driver Performance**

- Race Wins over the last 10 years
- Race Podiums (Top 3) over the last 10 years
- Fastest Laps (by season/by circuit/all time)
- Qualifying Positions
- Points earned (by race/by season/all time)
- Clinches (by race/by season/all time)
- Driver Success on purpose-built circuits (ie. Austin) versus closed public roads (ie. Monaco)

**Constructor Statistics**

- Pit stop trends over time (do pit stop times vary by driver? by circuit?)
- Constructor rankings/winningest constructors

**Circuit-Specific Statistics**

- Purpose-built circuits v closed public roads (performance/timings)
- Which circuits have the most incidents or accidents? Specific parts of the circuit?

- Is there a circuit that has a greater impact on timings?
- Is there a circuit that has correlation with tire burnout rate?
- Historic lap times

**Car Element Performance**

- Top 3 rankings/titles by chassis manufacturer
- Top 3 rankings/titles by engine manufacturer
- Top 3 rankings/titles by tire manufacturer

**Financial Analysis**

- Breakout of Financial Costs/Spend against Constructor Performance

**Links to data sources / API etc.**

The full scope of the project is still in discussion, but the primary data source is a Kaggle dataset, which is sourced from a motor sports racing API (Ergast Developer API). There is a chance that we would incorporate data from the Racing Statistics website as well.

**Outline briefly which types of visualizations you plan to use.**

For over time data that show comparison features, we'll include **line graphs**, or other **time series graphs**. We could include an **interactive map feature**, illustrating the various circuits or races that might make up an F1 season and the globe-trotting nature of the sport. If there is data to include a **topographical dynamic map** for race track elevation, we are interested in illustrating this element of the circuit. If the data allows, we'd like to illustrate crashes or contact between drivers by location on the race track or the lap number of the race.

**Are there any significant hurdles that you have doubts about? Would not solving them render the project incomplete?**

There are not any significant hurdles that threaten the completeness of our project. Finding and joining in outside data to our core data set could be challenging (ie. tracking down records of engine manufacturers over time) although this is dependent on the direction that we'd take this analysis.