**STAT 613 Project Proposal**

\*\* Please have an outline in RMarkdown for your proejct with the topic, data source, and an overview for what you plan to do for your project as well as a timeline. Identify who (if anyone) you are working with and how you plan to divide the work. What question do you plan to answer?\*\*

**Team Members:**

Devika Noir, Julia Smadja, and Dominic Thomas

**Topic/Question:**

D.C. United has hired DJD to analyze their futbol match data so that they may learn which environmental factors most influence game outcomes: wins, losses, and ties. DJD plans to ingest data on travel, stadium, and weather conditions specific to match days, in addition to gameplay data such as goals scored or shots taken.

DJD will identify relevant game and environment features by developing a model that can assist D.C. United in predicting the outcome of a game given certain conditions and test the models validity. Finally, DJD plans to develop an interactive and visualisation app so that D.C. United can continue to draw insights from the data at the conclusion of this project.

**Data Source:**

* Provided by DC United:
  + OPTA: MLS League Match Data by Team
  + DC United Travel Schedule
* Manually Assembled:
  + WUnderground: Weather History
  + Wikipedia: Stadium

**Timeline:**

* Data Ingestion/Cleaning - Feb 1 - Mar 1 DONE
* Exploratory Visualization and Analysis - Mar 1 - Mar 15
* Feature Selection/Extraction - Mar 22 - Mar 29
* Model Selection & Evaluation - Mar 29 - Apr 6
* Shiny App Development & Presentation - Apr 6 - End of Semester