**ETL Project Proposal**

**Extraction**

Our primary data source was Kaggle. From this website we found two different CSV files that contained data on NFL players.

One of our CSV files (nfl\_draft.csv) contained draft data for all players from 1985 to 2015. The other CSV file (QBStats\_all.csv) contained various performance statistics for all players from 1986 to 2016.

**Transform**

Using Pandas we read the nfl\_draft CSV and converted it to a Data Frame. We then cleaned it up to only contain Quarterbacks drafted from 2004 on. The cleaned up Data Frame included the year drafted, the round, the overall position in the draft and the players position.

We next used Pandas to read in the QBStats\_all CSV and converted it to a Data Frame. We cleaned it up to only contain quarterback stats from 2004 on. This clean-up also required significant clean-up of the player name in order to retrieve a name capable of being joined with the draft data. A count of the number of games played by each of the players was calculated for use later in our analysis. The statistics data was also grouped by the player and saved to a new Data Frame.

We then combined the new statistics Data Frame with the draft Data Frame and the previously calculated game count was added to give another field that could be used to further filter the data.

**Load**

We then used the .to\_sql function in Pandas to convert the Data Frame to a sqlite table. We also needed to use sqlite3 and the cursor function to duplicate our table and create a sqlite table with a primary key in order to use it in a Flask app.

We then used SQLAlchemy in order to map our new table and create our app routes as well as to return data from our table into a JSON list of dictionaries. Each dictionary contains statistics data and draft data for each player drafted since 2004.

Lastly we used HTML, Bootstrap and CSS to create a website interface for visualizing the functions for our app.