Proposal: Extract data from Kaggle regarding nfl salaries and statistics, and create a relational database for these data.

EXTRACT: The data we extracted was from Kaggle. The following data sources were used:

<https://www.kaggle.com/kendallgillies/nflstatistics> (Basic\_Stats.csv)

<https://www.kaggle.com/trolukovich/football-players-salaries> (football\_salaries.csv)

TRANFORM: We imported the data into Python and created dataframes from the CSV files. From the dataframes, we dropped data that we were not looking to import into PostgreSQL.

We also created dataframes for the teams and players tables so that we could create tables with primary and foreign keys and that were 4th normal form.

In the end we had four different csv files outputted:

Salary\_Table.csv

Players\_table.csv

Team\_table.csv

Basic\_Stats1.csv

LOAD:

We created the following tables:

Players, teams, salaries, stats

We loaded these files into PostgreSQL manually using the import files function. After loading, we did need to clean up the data a bit more. We wanted to create relations between the teams and players tables so we needed to use an alter table statement to add a column for team\_id in the players table, and we utilized an update statement to update the foreign key value.