NFL ETL report

Extraction:

The data for this ETL was found from a very complete NFL data set, which luckily contained all the data desired. It was read with pandas ‘read\_excel’ function.

<http://www.aussportsbetting.com/historical_data/nfl.xlsx>

Transformation:

It was decided that the bet data should be trimmed down to opening odds for a couple of reasons:

1. They were the only bet data point that was included for every NFL game in the dataset.
2. Win/Loss odds are more directly analogous to those form other sports where bet lines or total score over/under wont have meaning.

Game score columns were replaced with a ‘Home Win’ Boolean column

In the end ‘Date’, ‘Home Team’, ’Away Team’, ‘Home Win’, ‘Home Odds Open’, ’Away Odds Open’

Dropped any NaN columns

Loading

Created database connection to postgres using SQLAlchemy ‘create\_engine’

Loaded final pandas dataframes using ‘to\_sql’ method and connection created by SQLAlchemy dependency