All about the 49ers.

I was given a specific task to correlate game data from the 49ers. The concentration was to focus on the relationship between where the location of the games were, an how it was affected by weather and if that factored on the over under of the games.

Extract.

I was able to find data on two specific sites that gave a robust amount of data in regards to scores, locations, outcomes of the games. The data was in the form of a downloadable csv file, which was compiled by specific parameter needed. I chose for this subset the last 3 years, analyze this data. These 3 years were important since it was the 3 years coached by the current head coach. Each head coach has a specific emphasis on offensive and defensive calling.

https://www.pro-football-reference.com/teams/sfo/2016.htm

https://www.pro-football-reference.com/teams/sfo/2017.htm

https://www.pro-football-reference.com/teams/sfo/2018.htm

https://pypi.org/project/sportsreference/

http://www.nflweather.com/en/searches/100725

One of the main concerns were to join the csv files into one database. Each had different fields. I chose to keep them separate but provide a primary key of opponents as the primary key.

The database created was a relational database.

A screenshot of a social media post

Description automatically generatedA close up of a piece of paper

Description automatically generated

Some of the key columns were the dates of the games, the outcome score, and weather. An important column is the covered/not covered column which will help determine if an upcoming game played against has a high probability of that outcome, in comparison to past data.

Transform

Using python and a pandas extension, I was able to delete columns that did not have an effect on this analysis. Having multiple years of data, was easy to cull and munge thru with either joining of dataframes, or a simpler process with excel manipulation.

A screenshot of a social media post

Description automatically generated

Transfering the data to postgres was difficult. I was able to install the psycopg2 into python and appear to have connection to the database. It’s the actual transferring of the data into the database I was having trouble with.

A screenshot of a cell phone

Description automatically generated

When loading data into the tables in my ETLproject\_DB, I was receiving errors of the following.

A screenshot of a cell phone

Description automatically generated

Scouring the stack overflow and python help sites, most were remedied by changing the type into text/integer with 255 characters limitation. With this being the date, It wasn’t so easy.