Project Report

Extract

The raw fast food data was imported as a csv and was taken from:

https://www.kaggle.com/khushishahh/fast-food-restaurants-across-us

The raw obesity data was imported as a csv and was taken from:

<https://nccd.cdc.gov/dnpao_dtm/rdPage.aspx?rdReport=DNPAO_DTM.ExploreByTopic&islClass=OWS&islTopic=&go=GO>

Transform

The raw obesity data set had quite a lot of redundant columns that where removed for the purposes of the project. Out of the 43 columns in the spreadsheet only 7 where relevant. The columns where also renamed and reordered. In order to make a foreign key that would be able to link to the fast food table in the database, I make a column with abbreviations of the states in America.

The raw fast food data had a few redundant column dropped and reordered in a fashion that would make sense to someone using the database. The data set required new rows to separate entries that had two fast food restaurants that shared the same location. I was able to find out that there were only a few entries that had shared locations using “.unique()”. I exported the csv and used the search tab to separate the entries. I also used this time to replace all the entries that were wrongly capitalized or had incorrect punctuation (namely Subway and McDonald’s respectively).

Load

The final database that was chosen was a relational database in SQL. Two tables created, namely: fast\_food\_db and obesity\_df. It was done like this due to the one to many relationship the two tables have with each other. The common column name is “State\_Abbrev” so in the obesity\_db it would act as a primary key and in the fast\_food\_db it would be a foreign key.