Data Clean up and Analysis

The raw fast food data set was taken from:

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

The raw obesity data set was taken from:

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

Transformation

It looks like the raw obesity data set has a lot of columns so probably more than half will be removed and reordered. I want analysts, students and food journalists to be able to able to. In order to make a foreign key that would be able to link to the fast food table in the database, I’ll make a column with abbreviations of the states in America.

For the raw fast food data I used “.unique()” and found out that there are a few entries in the “Names” column that share address. For the purposes of the database that's going to be made, I’m going to need to separate them. I also see that there are several ways the creator of the spreadsheet has spelt “subway” so ill need to normalize any inputs that have been spelt differently

Type of database

The final production database was a relational database.

Final tables to be used:

* Fast food data
* Obesity stats