ETL Project

In our project, we decided to do food review and we got our data from Kaggle. Mine wound up being on restaurant reviews and we had the idea that we both worked for an Asian themed foodie review site.

Overall our data was very easy to navigate from the beginning and didn't need much cleaning. I combined the types of restaurants and left out the 'Japanese name' column. Then split the data by stars to category type and stars to money. This was to see so anyone could look for the highest score to price/ category. It seems rather redundant to have done after the fact, however.

Once I was done cleaning the data I linked my jupyter notebook to PGSQL. I chose SQL because it's relational and I was looking for the relation between the stars and various columns. I integrated the df into tables in SQL and checked that they were there. That was about it.