ETL Transform Process

Step 1: Import CSVs from Kaggle Website into Jupyter Lab Pandas.

Step 2: Read CSVs with “read\_csvs” into Pandas DataFrame

Step 3: Filter Homeless DataFrame set to only the measurement of “Total Homeless”

Step 4: Transform Homeless data: One major issue with the two different data sources in this project is that the home CSV lists data by timeframe into rows, while the Zillow Rent Price Data utilized columns to show progressive time. For example, the Homeless data set progressed through time as rows increased, while the Zillow Dataset time progressed through columns. To overcome this, we utilized the Pandas pivot table function and moved “States” to the first column set to index, then moved years into columns and populated values.

Step 5: After transforming the Homeless dataset we renamed columns to a format that we would continue with the Zillow data.

Step 6: Zillow Data Set: The Zillow dataset incorporated monthly rent prices by year for both cities and states. The Homeless Data set only had state information, because of this we used the GroupBy function in Pandas for State and set the values to Mean. This gave us consolidated average rent prices for every state. Next we filtered the Zillow dataset to only January of each year to match the Homeless dataset which was pulled annually in January.

Step 7: Zillow Data: Since the Zillow Data format is now matching to our homeless dataset we created a combination of a few datasets by performing DataFrame Joins using “State” as our key. The dataset listed on this website shows the Homeless count for 2011 & 2016 with the variance and % change. In addition, we also showed the States 2011 & 2016 rental price with the variance and % change.

In conclusion, the final data set compares the variance and % change for Homeless & Rental Prices by state. This allows users to quickly compare states changes in Homeless rate with a comparison to the states rent prices to see if there is any correlation. Finally, we also prepared a larger data frame that holds all Homeless Rates and Rental Prices from 2011 to 2016