ETL REPORT:

**Extract**: *The sources of data that you will extract from.***Transform:** *The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc)***Load**: *The type of final production database to load the data into (relational or non-relational).***Analysis:** The final tables or collections that will be used in the production database.

**Background**:

We used Seattle Airbnb data we found through Kaggle and on the Airbnb website. The data was in csv form originally and we used all of the 2019 data that was available, which consisted of 8 CSVs (January – August). The csv's included data for each listing – neighborhood, price, review information, etc for each month. We also included a csv with availability data for 2019 to show dates that each listing was available or filled.

**ETL Process:**

1. This process included downloading each CSV from the Airbnb site and exporting them to Python using Jupyter Notebook.
2. Using pandas, we created a dataframe for each monthly CSV file.
3. Then we cleaned each dataframe to remove duplicate columns for aggregating
4. Re-named columns for merging all the CSVs together
5. Combined all dataframes into one main dataframe.
6. Once all of the 8 months of 2019 data were combined, we then exported into a CSV file.
7. We decided to store the data into a SQL database using pgAdmin 4 and created a table with the 2019 CSV file, setting the listing ID as the primary key.
8. We created another table with the 2019 availability listing CSV which enabled us to join with the main table we created and pull data based on availability for each listing.
9. By using a relational database, this allows us to more easily query on the data and extract data that we need.