**Project Report**

​

​

**1 Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).**

In this project we are using 2 CSV datasets:

Housing List: it includes more than 500 rental houses and apartments available in the market for each major city.

Data Analyst Job: it list out the current hiring positions related to data analytic and data science in the job market

**2 Transform: what data cleaning or transformation was required.**

We selected 5 major cities from both datasets: Chicago, San Francisco, San Diego, New York and Seattle

We first loaded both Datasets into 2 separate Data Frames, and then clean up the housing list data to calculate the Avg. monthly rental price and Avg. rent /square feet in 5 cities.

We found cities’ name are written slightly different in both two datasets, since the “city” name will be our primary key, we updated the “city” name in housing list dataset to align with the “city” name in Data Analyst.

​

**3 Load: the final database, tables/collections, and why this was chosen.**

​

After Data Cleaning we connected to local database to query the required the results. For querying, we merged the 2 tables on "City" and found Top Paid Jobs.