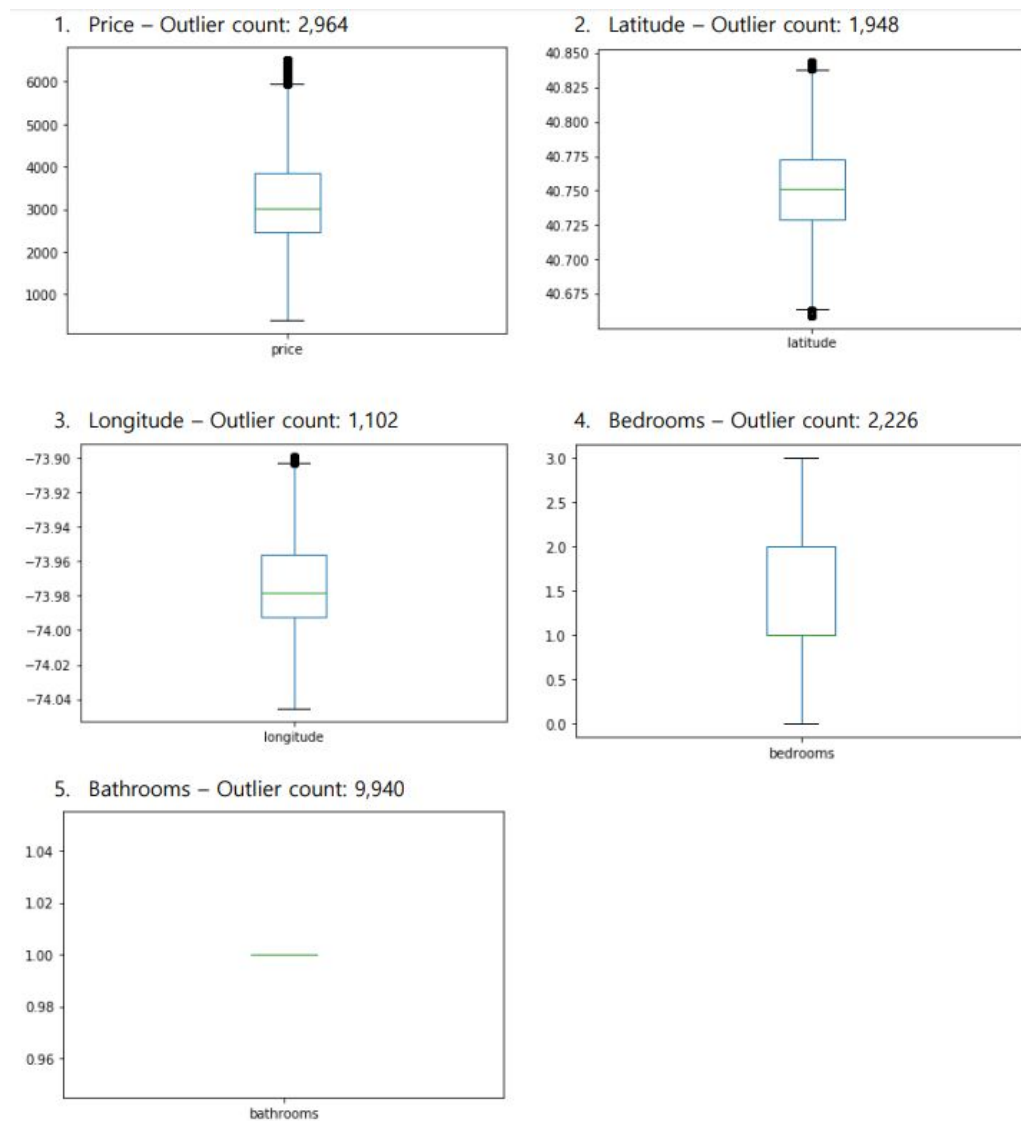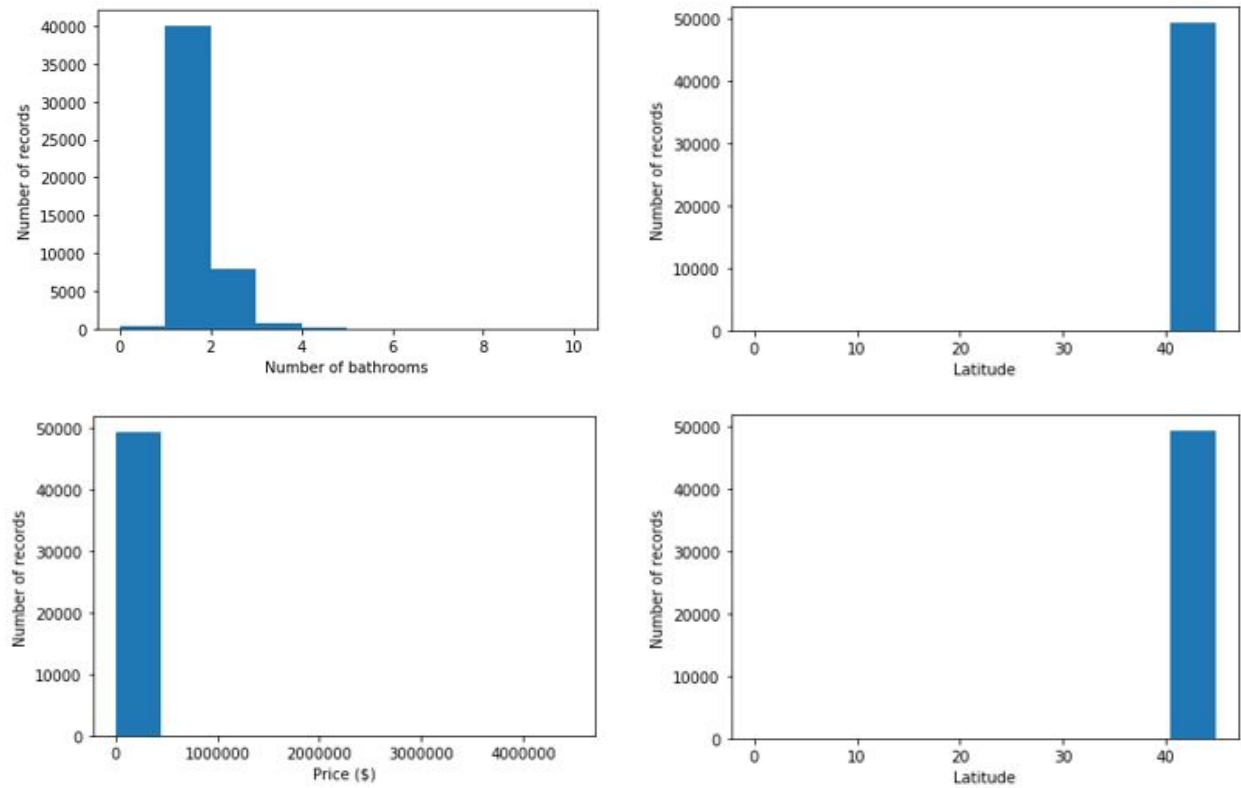# Appendix

A. Correlation rank of the numerical attributes in the initial dataset

```
interest_level_num    1.000000
bedrooms              0.074349
longitude             0.032485
bathrooms            -0.020075
latitude             -0.027469
price                -0.202009
```
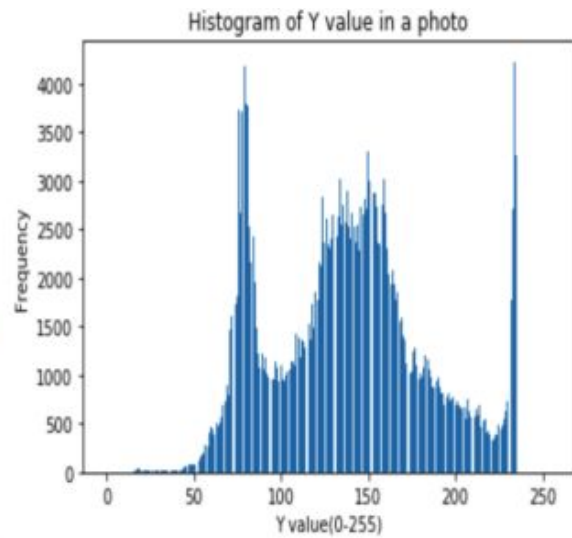
B. Boxplots of the numerical attribute in the initial dataset.

1. Price – Outlier count: 2,964

2. Latitude – Outlier count: 1,948

3. Longitude – Outlier count: 1,102

4. Bedrooms – Outlier count: 2,226

5. Bathrooms – Outlier count: 9,940

C. Histograms of the attributes in the initial dataset before outlier removal



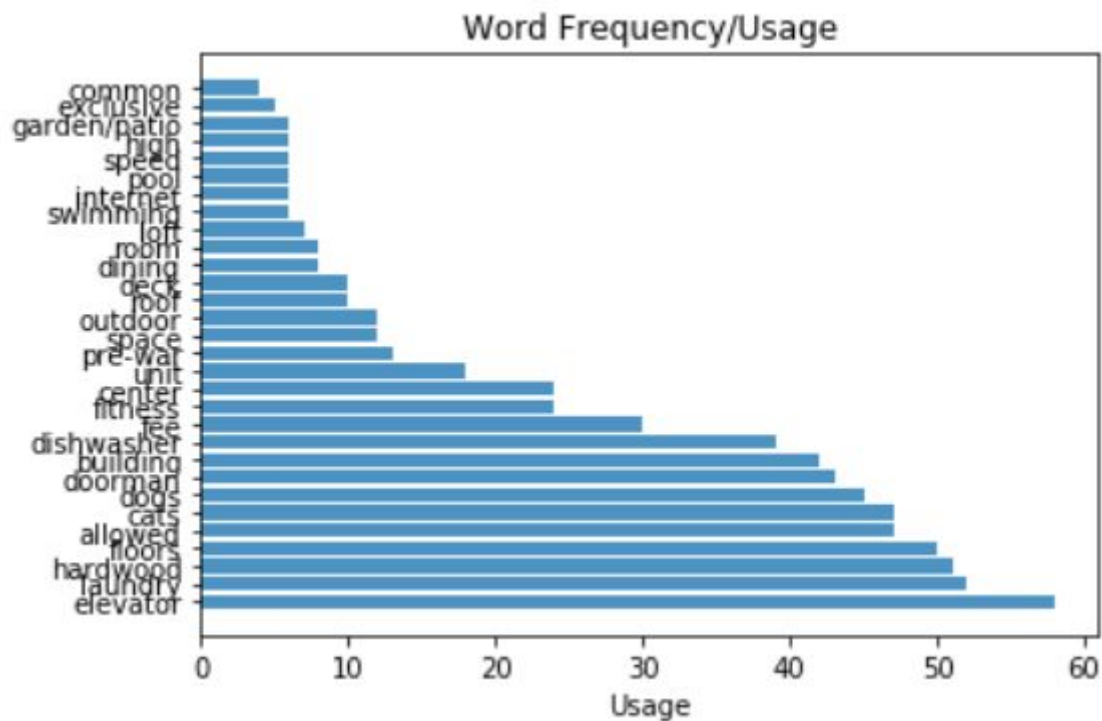D. Image extraction process
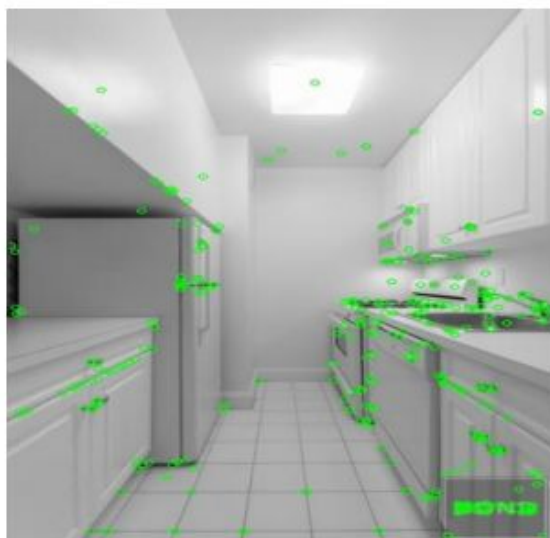


(E.g Features extracted using OpenCV)

E. Frequency of features from text extraction.

## Word Frequency/Usage



F. Manager skills and rent price by interest level

```
*Manager Skills by Interest Level Classes*
For high interest level:
The average manager skill is 0.0352
The median manager skill is 0.0000
The standard deviation is 0.4336
The sum of all the manager skill is 100.27
The dataset has 2849 number of records
-------------------------
For medium interest level:
The average manager skill is -0.1921
The median manager skill is -0.2237
The standard deviation is 0.3745
The sum of all the manager skill is -1651.61
The dataset has 8596 number of records
-------------------------
For low interest level:
The average manager skill is -0.6515
The median manager skill is -0.7188
The standard deviation is 0.3397
The sum of all the manager skill is -16262.66
The dataset has 24961 number of records
```

```
*Rent Price by Interest Level Classes*
For high interest level:
The average rent price is $ 2589
The median rent price is $ 2495
The standard deviation is $ 777
The dataset has 2849 number of records
-------------------------
For medium interest level:
The average rent price is $ 2869
The median rent price is $ 2759
The standard deviation is $ 772
The dataset has 8596 number of records
-------------------------
For low interest level:
The average rent price is $ 3184
The median rent price is $ 3090
The standard deviation is $ 877
The dataset has 24961 number of records
```

G.   Correlation heat map with additional attributes