



NEW YORK AIRBNB RENT PREDICTION

ZEIAD SABRA

BUSINESS PLAN

RightPrice wants us to make a model to advise tourists visiting New York City on the optimal price for a place to stay in New York city.

A tourist provides us with Information about the place and we provide him/her with the optimal price using our model.

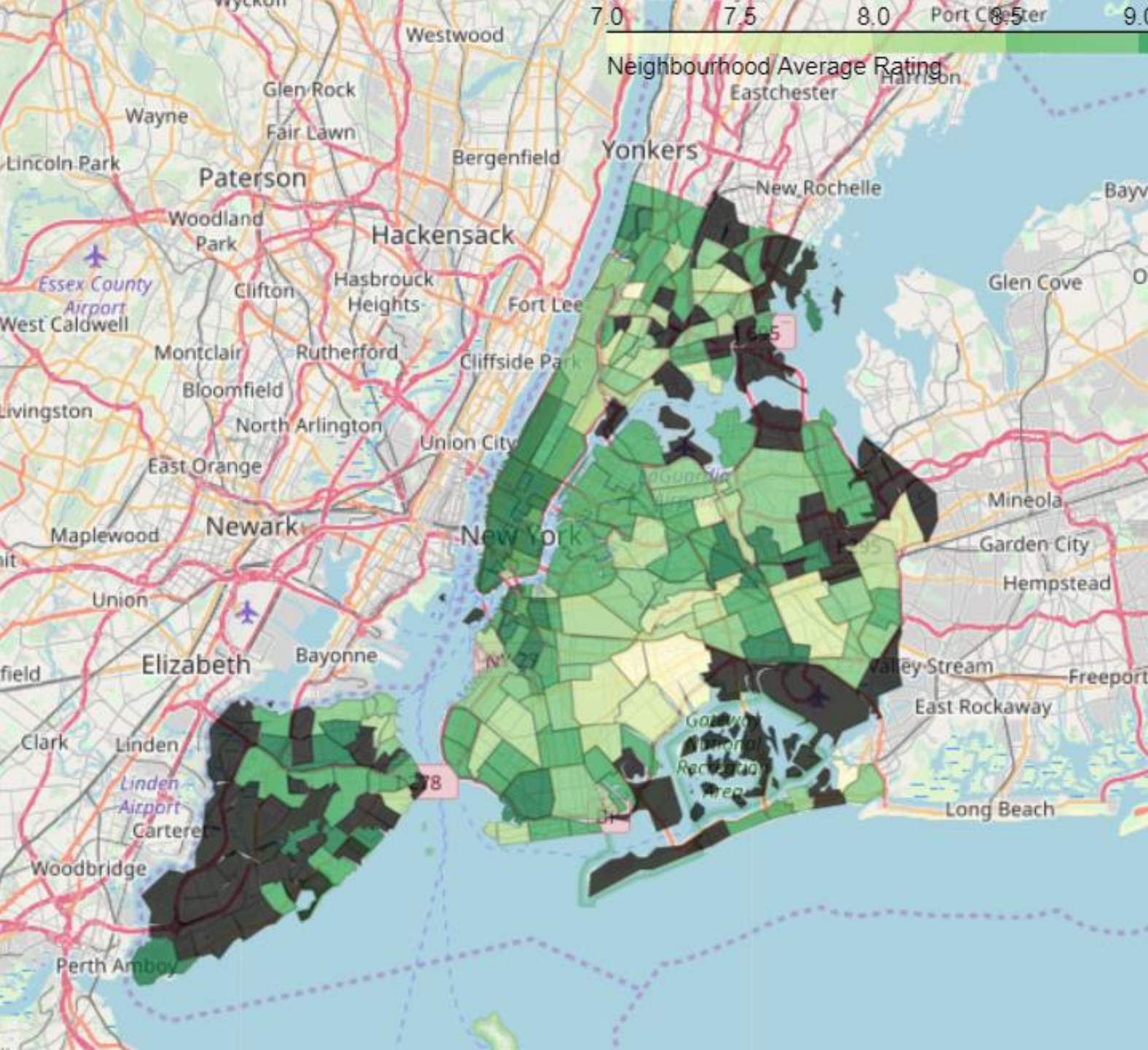
Our goal is to build a model that give an estimate of the rent of a place in New York City using available data.

- The Desired outcomes are:
- A model for calculating rental prices.
- A description of the most relevant features of the model.
- Cluster the Neighbourhoods based on the Rent, Venues, and location.

name	price	review_scores_location	latitude	longitude	bedrooms	room_type	bathrooms	property_type
Charming Studio - Central Park	150.0	10.0	40.781561	-73.971238	0.0	Entire home/apt	1.0	Apartme
Rockaway Bungalow by the Bay	60.0	9.0	40.591061	-73.814242	1.0	Private room	1.0	Apartme
Cozy Mexican Inspired Private Room	97.0	10.0	40.779410	-73.969830	1.0	Private room	1.0	Apartme
Modern 1BD with exposed brick	100.0	10.0	40.655026	-73.962212	1.0	Entire home/apt	1.0	Apartme
Manhattan Cozy BR Apartment \$60	60.0	9.0	40.873336	-73.911239	1.0	Entire home/apt	1.0	Apartme

SAMPLE DATA

We use this rental data of airbnb to build a model to predict the rent of Airbnb listings.



NEIGHBORHOOD RATING

- The rating of neighborhoods according to the Airbnb dataset. As you can see, Manhattan has the highest ratings


```
lm.coef_
```

```
array([-3.08488254e+02, -3.63734513e+13,  3.62237541e+13,  9.45296264e+12,  
       -6.95372282e+13,  7.39605121e+13, -2.58560734e+13,  5.47320789e+12,  
       -1.64310732e+13, -1.44266629e+13, -1.43607209e+13, -8.06380684e+12,  
       -2.82280340e+13, -1.34287739e+13, -2.18677749e+12, -1.19971614e+13,  
        4.59203255e+12, -7.84858086e+11, -6.68466228e+12,  4.76997154e+11,  
       -5.27380499e+11,  1.19534987e+12, -1.79806988e+13, -3.73228478e+12,  
        5.85455020e+12,  9.90299285e+12,  5.25122449e+12, -6.65150514e+12,  
       -9.17819305e+11, -1.29577667e+12,  1.09770582e+13, -1.84705534e+12,  
        2.82398857e+12,  1.20667669e+11,  3.22107762e+11,  4.78895382e+12,  
        2.25259985e+12,  3.00865116e+11,  1.17264463e+12, -3.06625814e+12,  
       -4.54004767e+12,  1.04183915e+12, -5.23047838e+12, -2.42187532e+12,  
        2.17752250e+12,  8.13457431e+11,  6.76133398e+12,  9.20961124e+11,  
       -3.79413280e+11, -4.04722027e+12,  4.47022404e+12,  7.34115146e+11,  
       -2.38197639e+11,  2.95118697e+12,  2.36512523e+12, -1.90717192e+12,  
        7.07166925e+10,  8.11688095e+11,  2.76766303e+12,  4.93846536e+12,  
        5.21320931e+10, -2.57588365e+12, -4.40594449e+12,  6.96624553e+11,  
       -1.62175532e+13, -3.17850182e+12,  2.75742649e+11,  1.51865003e+11,  
       -3.17128190e+11,  1.25013677e+12, -2.51075576e+12,  2.83893454e+12,  
        8.53521179e+12,  3.31745992e+12, -2.41441475e+12, -1.52028993e+13])
```

BUILDING A MODEL

- We used Linear Regression to train a model on the data, however, the model overfitted the data.

```
: lm = LinearRegression()  
lm.fit(X_train, y_train)
```

```
: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,  
                   normalize=False)
```

```
: lm_mse_train = mean_squared_error(lm.predict(X_train), y_train)  
print("Training Error: ", lm_mse_train)
```

Training Error: 28956.189363203102

```
: lm_mse_val = mean_squared_error(lm.predict(X_val), y_val)  
print("Validation Error: ", lm_mse_val)
```

Validation Error: 2.565207319572144e+20

```
coef = coef[abs(coef) != 0]
coef
```

```
: review_scores_location      16.663473
   bedrooms                   67.694707
   bathrooms                   100.748614
   guests_included             -5.210580
   cluster_0                   -34.035317
   cluster_2                   -22.423239
   cluster_4                   14.294990
   room_type_Entire home/apt    98.683553
   property_type_House          -11.154660
   property_type_Loft           4.339545
   dtype: float64
```

```
: lss.intercept_
-7009.935750371902675
```

```
Lasso()
fit(X_train, y_train)
```

```
alpha=1.0, copy_X=True, fit_intercept=True, max_iter=
maximize=False, positive=False, precompute=False, rand
action='cyclic', tol=0.0001, warm_start=False)
```

```
mse_train = mean_squared_error(lss.predict(X_train), y_
"Training Error: ", lss_mse_train)
```

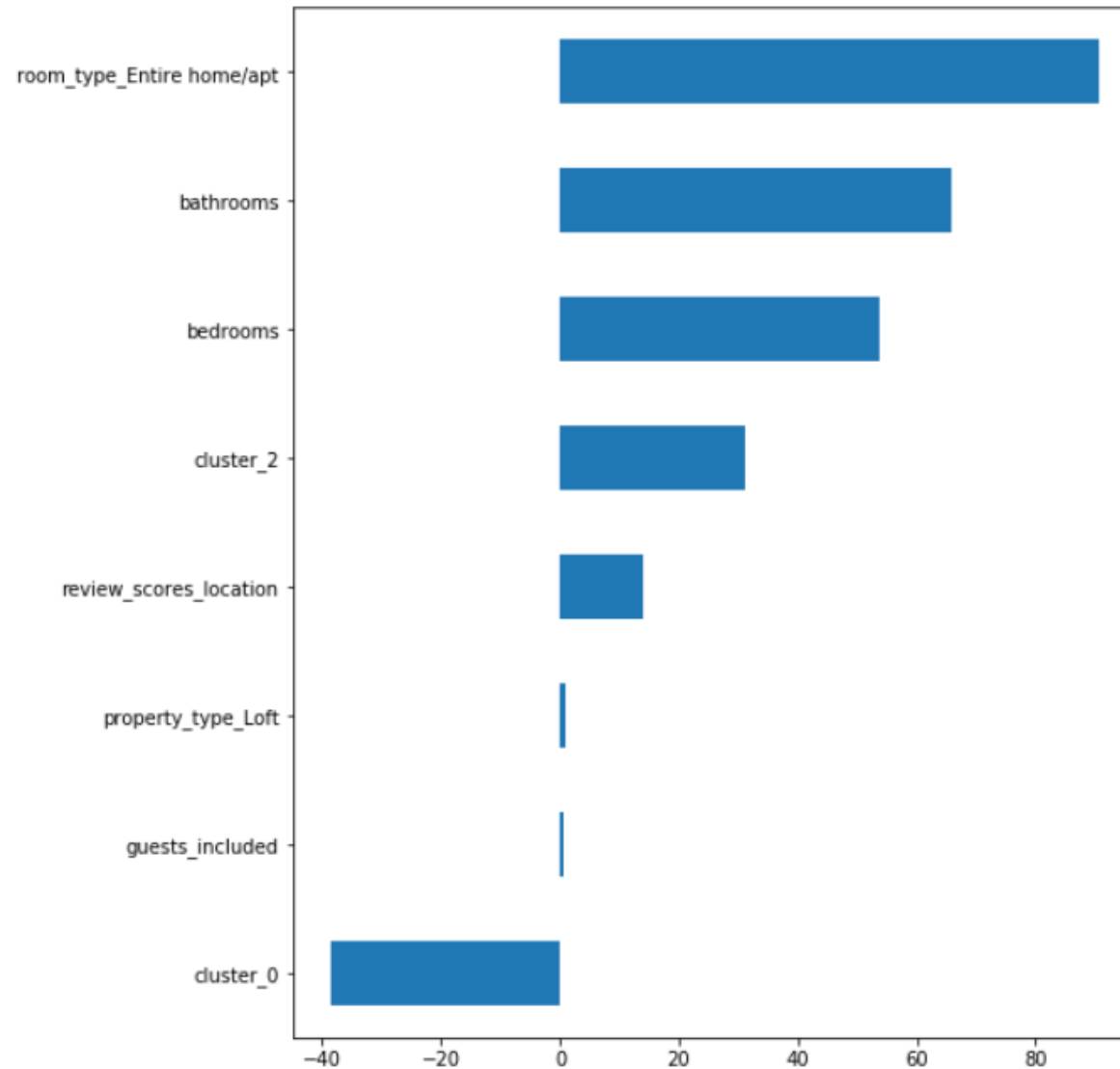
```
Training Error: 30757.109686446474
```

```
mse_val = mean_squared_error(lss.predict(X_val), y_val)
"Validation Error: ", lss_mse_val)
```

```
Validation Error: 37797.071490769726
```

BUILDING A MODEL

- Since the unregularized Linear Regression Overfitted the data, we use L1 Lasso Regression to avoid overfitting.



FEATURE SELECTION

- Our Lasso model is also used for Feature Selection.
- Here are the features it selected
 - Entire home or Apartment increase the rent.
 - Number of Bathrooms
 - Number of bedrooms
 - Type of Neighborhoods
 - Cluster 0 seem to be undesired/poor
 - While cluster 2 seem to be expensive, Includes Manhattan and the rich areas.

CONCLUSION

- While Our model was decent, it was not good enough to be deployed for a business.



That's all Folks!