PREDICTING BROOKLYN APARTMENT RENTAL PRICES

By: Danielo Bennett

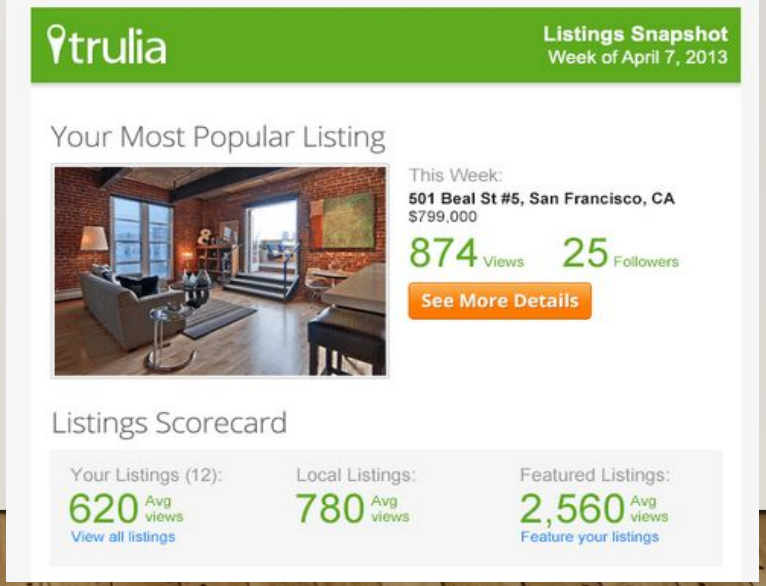
INTRODUCTION

- **Purpose:** Predicting rent can be useful for apartment building owners to know the **market value** of vacant units. This will better help them find the proper price of a listing in an inflated real estate market.

METHODOLOGY

BeautifulSoup + Se

trulia®



DATA SET

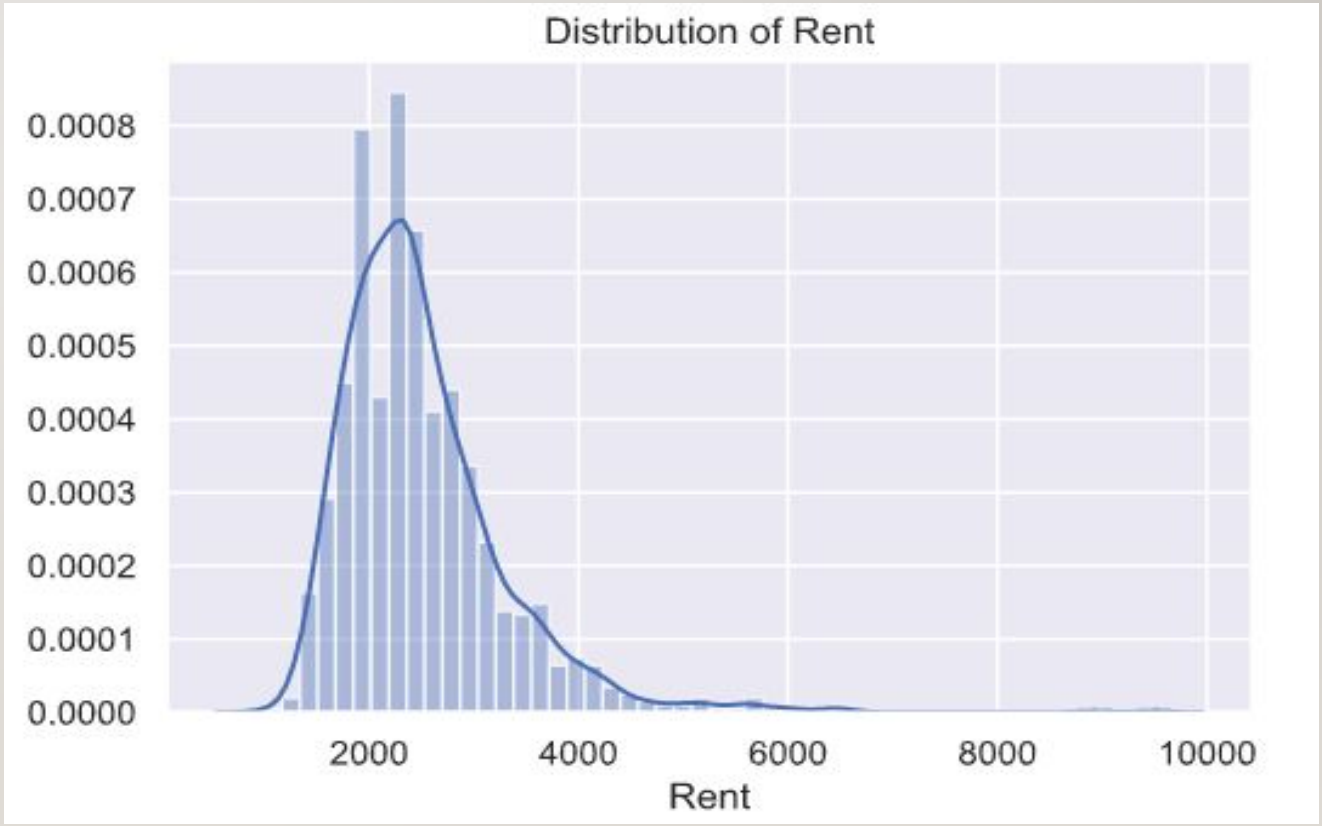
Data Set Parameters: ~**1200** Brooklyn apartment listings, **54** features

Some features such as:

- "Normal" features:
  - # of **bathrooms** & **bedrooms**
  - Number of **amenities**(doorman, fitness center, etc)
  - **Neighborhood(**categorical)
- "Interesting" Features:
  - Distance to **Starbucks**
  - **Elevation**
  - **Tree Coverage %**

DISTRIBUTION OF RENT VALUES

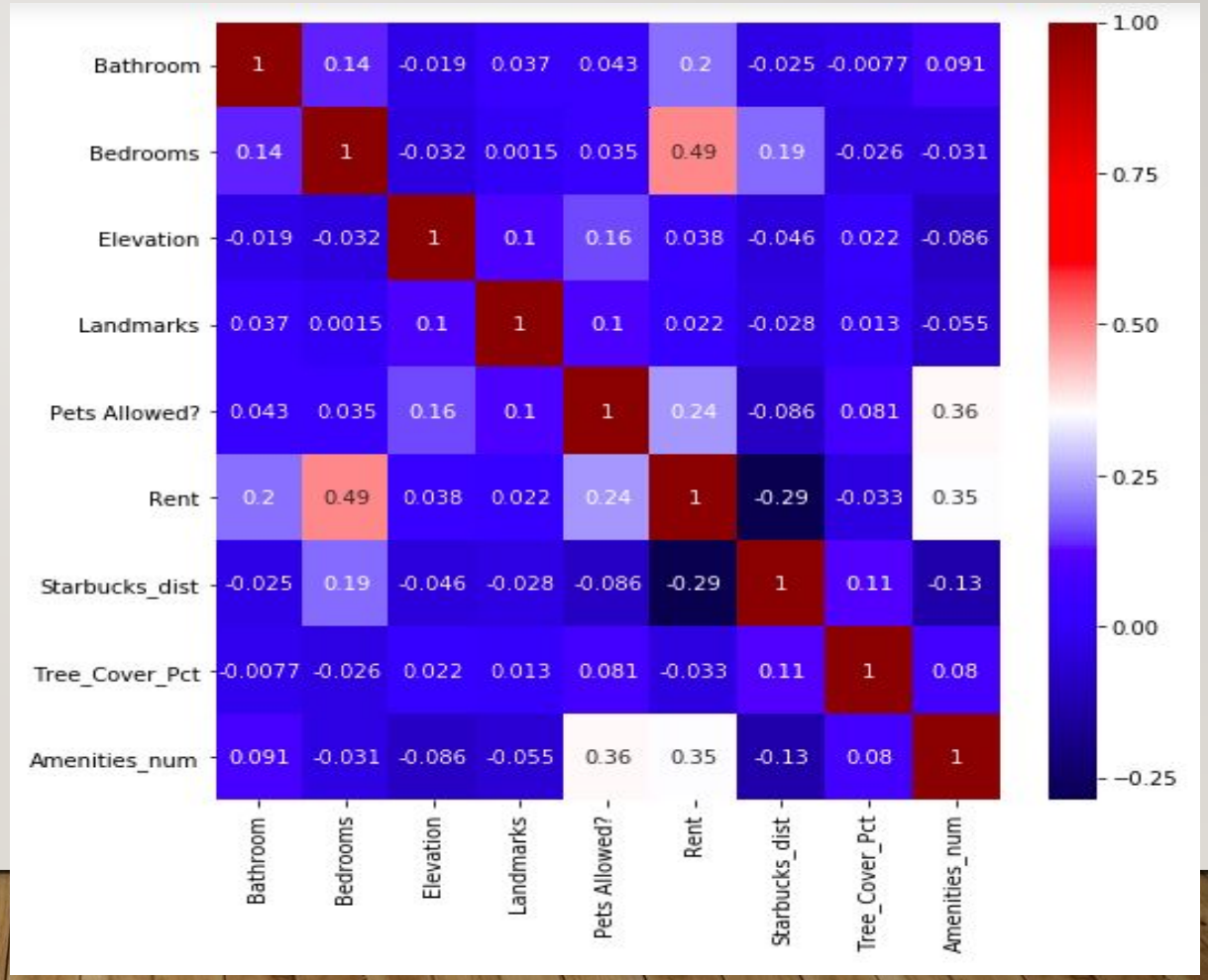
The majority of the rental prices range from $1800-$3200 a month.



CORRELATION MATRIX

Major Hypothesis:

- Bedrooms, Number of Amenities, & Starbucks distance seem to have the most pull
- Low overall multicollinearity in feature set



MODEL SELECTION PROCESS

Stage 1

Baseline Model:
Test R^2 = .62

Suspected overfitting

Stage 2

Cross Validation Model Comparison:
-Linear Regression
-Ridge Regression
-Lasso Regression

Optimized

Final Model

Final Model:
Lasso Regression
Test R^2 = .607

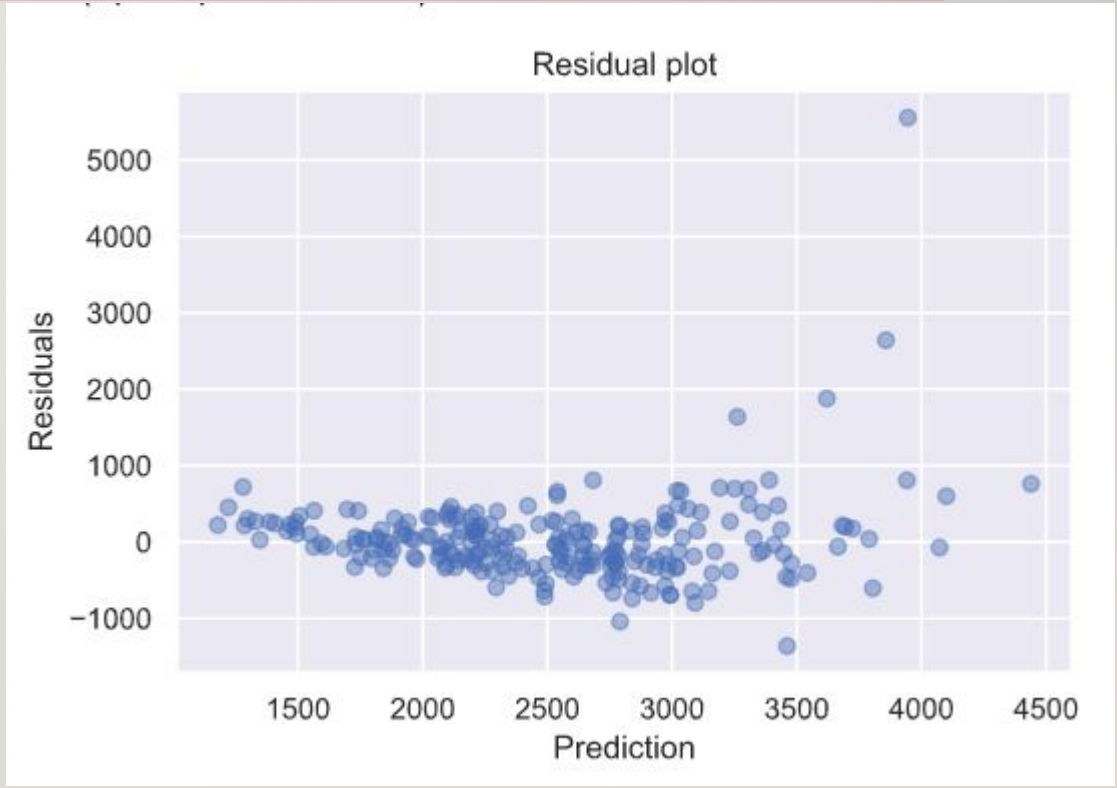
MAIN TAKEAWAYS

Top Features:

- **Number of Bedrooms** is the most influential feature
- **Starbucks Distance** has the most influence with a negative coefficient
- Certain **neighborhoods** have a premier due to popularity/closeness to Manhattan

Model is Prone to over predicting

- MAE: **$325.34**
- Mean of Rent: $**2508.89**



NEXT STEPS

-Add more **observations** to create a model that better generalize

-Extract Additional Features:
- -Distance to nearest train station
- -Include sqft of apartment
- -Replace number of amenities with individual amenities

- A more **evenly distribution** of listings across neighborhood

Q&A

THANK YOU! ANY QUESTIONS

APPENDIX A: TABLE OF MODEL SCORES

Model Name	Model Parameters	Additional notes	Training R^2	Test R^2	Valuation Method
Set 1					
Baseline v0 Incomplete	All columns excluding Neighborhood		0.5398976436	0.4635095642	80/20
Baseline v1	Added Neighborhood Dummies		0.772807155	0.6202703273	80/20
Set 2				CV mean score	
Linear Regression	unscaled			0.7304618515	Cross Val= 5 folds
Ridge Regression	scaled, alpha = 1			0.7324368651	Cross Val= 5 folds
Lasso Regression	scaled, alpha = 1			0.7347593317	Cross Val= 5 folds
Ridge Regression	scaled, alpha = (best alpha)	alpha = 7.84282206133768		0.7352731501	Cross Val= 5 folds
Lasso Regression	scaled, alpha = (best alpha)	alpha = 4.29700470432083		0.73529964	Cross Val= 5 folds
Set 3				CV mean score	
Ridge Regression w/ Polynomial Features (degree=2)	scaled, alpha=1, polynomial			0.7324368651	Cross Val= 5 folds
Lasso Regression w/ Polynomial Features (degree=2)	scaled, alpha=1, polynomial			0.7347593317	Cross Val= 5 folds
Ridge Regression	scaled, alpha = (best alpha), polynomial			0.7352731501	Cross Val= 5 folds
Lasso Regression	scaled, alpha = (best alpha), polynomial			0.73529964	Cross Val= 5 folds
Final			Training R^2	Test R^2	
Lasso Regression	scaled, alpha = (best alpha)		0.7660645509	0.6070599171	80/20

APPENDIX B: FEATURE COEFFICIENTS

```
Bathroom : 30.05
Bedrooms : 462.44
Elevation : 67.30
Landmarks : -0.00
Pets Allowed? : 2.89
Starbucks_dist : -149.58
Tree_Cover_Pct : 31.43
Amenities_num : 170.35
Bay Ridge : -40.61
Bedford Stuyvesant : 77.86
Bensonhurst : -54.64
Bergen Beach : -5.77
Boerum Hill : 185.91
Borough Park : -0.79
Brighton Beach : -12.57
Brooklyn Heights : 116.89
Brownsville : 18.47
Bushwick : 0.00
Canarsie : 0.00
```