Predicting Manhattan Airbnb Rental Prices

A Guest and Host Solution

By: Michael Li

METIS

Manhattan Rent Prices are too ***** high!

Two Problems

- Finding cheap short term housing
- Renting out idle space efficiently
  - ...to even afford the rent

Airbnb Solution

- Guests: Find housing accommodations
- Hosts: Rent out their idle space for extra income

How to **save money** on rentals and how to **maximize return** on idle space on Airbnb?

Using **Linear Regression** to **Predict Price** and find **Key Features**

Methodology

Price Per Night vs. Features

Gathering Data/Web Scraping

Web-scraped Airbnb Listing Data

Beautiful Soup and Selenium

Gathering Data/Web Scraping

Web-scraped Airbnb Listing Data

Beautiful Soup and Selenium

Data Cleaning

Removed outliers such as a **$4000 / night penthouse suite**

Focusing on typical Airbnb Rentals

Removed listings with **0 Reviews**

Pandas

Gathering Data/Web Scraping

Web-scraped Airbnb Listing Data

Beautiful Soup and Selenium

Data Cleaning

Removed outliers such as a **$4000 / night penthouse suite**

Focusing on typical Airbnb Rentals

Removed listings with **0 Reviews**

Pandas

Exploratory Data Analysis

Created plots and visualizations

Looked for trends and relationships

Engineered **new features** such as: **“New Listing”**

Dropped features

Matplotlib and Seaborn

Gathering Data/Web Scraping

Web-scraped Airbnb Listing Data

Beautiful Soup and Selenium

Data Cleaning

Removed outliers such as a **$4000 / night penthouse suite**

Focusing on typical Airbnb Rentals

Removed listings with **0 Reviews**

Pandas

Exploratory Data Analysis

Created plots and visualizations

Looked for trends and relationships

Engineered **new features** such as: **“New Listing”**

Dropped features

Matplotlib and Seaborn

Model Training and Validation

Split Data off for Testing

Created a basic linear regression base model

Compared Base model With other more complex models

Cross-Validated most promising models

Gathering Data/Web Scraping

Web-scraped Airbnb Listing Data

Beautiful Soup and Selenium

Data Cleaning

Removed outliers such as a **$4000 / night penthouse suite**

Focusing on typical Airbnb Rentals

Removed listings with **0 Reviews**

Pandas

Exploratory Data Analysis

Created plots and visualizations

Looked for trends and relationships

Engineered **new features** such as: **“New Listing”**

Dropped features

Matplotlib and Seaborn

Model Training and Validation

Split Data off for Testing

Created a basic linear regression base model

Compared Base model With other more complex models

Cross-Validated most promising models

Testing

Tested best performing model:

Linear Regression Lasso Regularization

Results

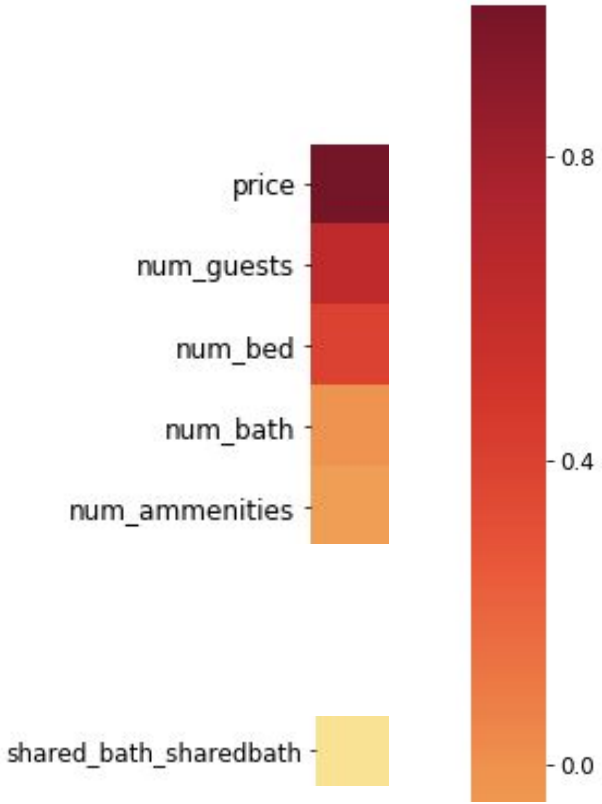
Price Per Night = Accomodations + Location + Listing Type + Reviews

Findings:

Price and Feature Correlations

- Positive Relationship
  - **Number of Guests**
  - **Number of Beds**
  - Downtown
  - **Review Ratings**

- Negative Relationship
  - **Shared Baths**
  - Uptown
  - Shared and Private Space

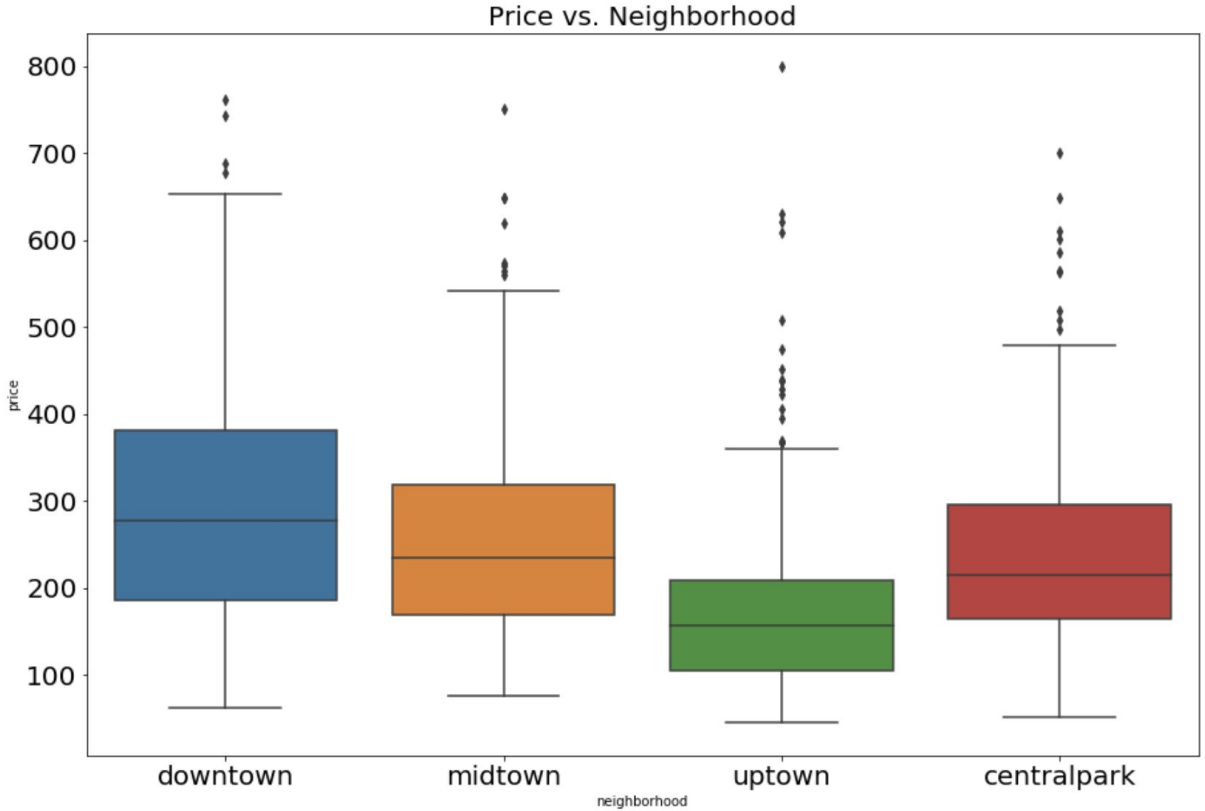


Findings:

Price and Feature Correlations

- Positive Relationship
  - Number of Guests
  - Number of Beds
  - **Downtown**
  - Review Ratings

- Negative Relationship
  - Shared Bath
  - **Uptown**
  - Shared and Private Space

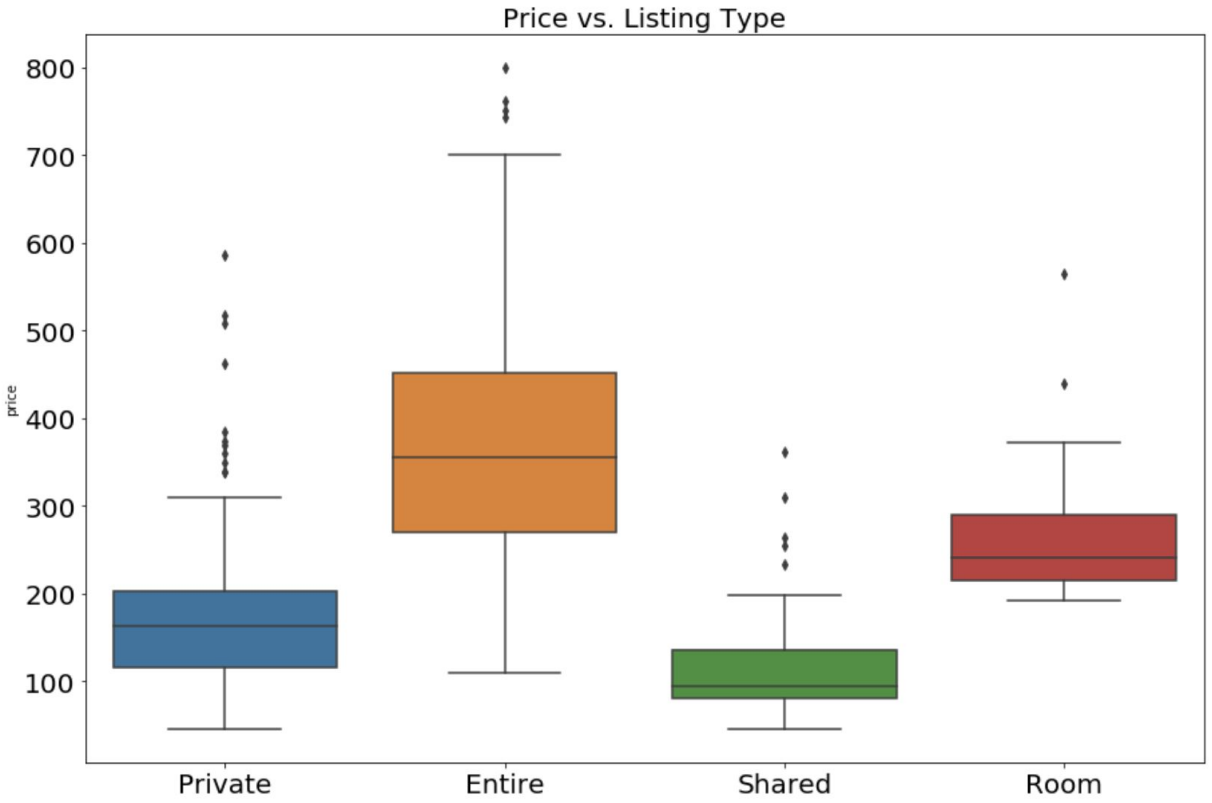


Findings:

Price and Feature Correlations

- Positive Relationship
  - Number of Guests
  - Number of Beds
  - Downtown
  - Review Ratings

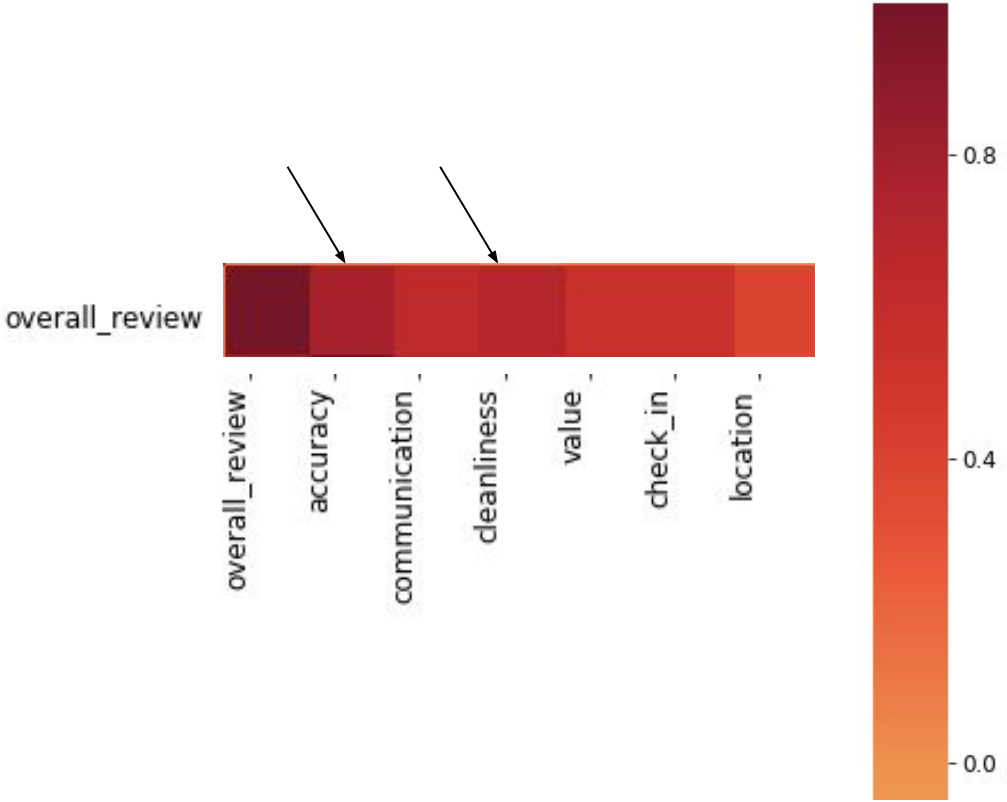
- Negative Relationship
  - Shared Bath
  - Uptown
  - **Shared and Private Space**



Reviews

Overall Reviews Very Important

- **Out of 5 Stars**
- **Important Components**
  - **Accuracy**
  - **Cleanliness**



Best Model

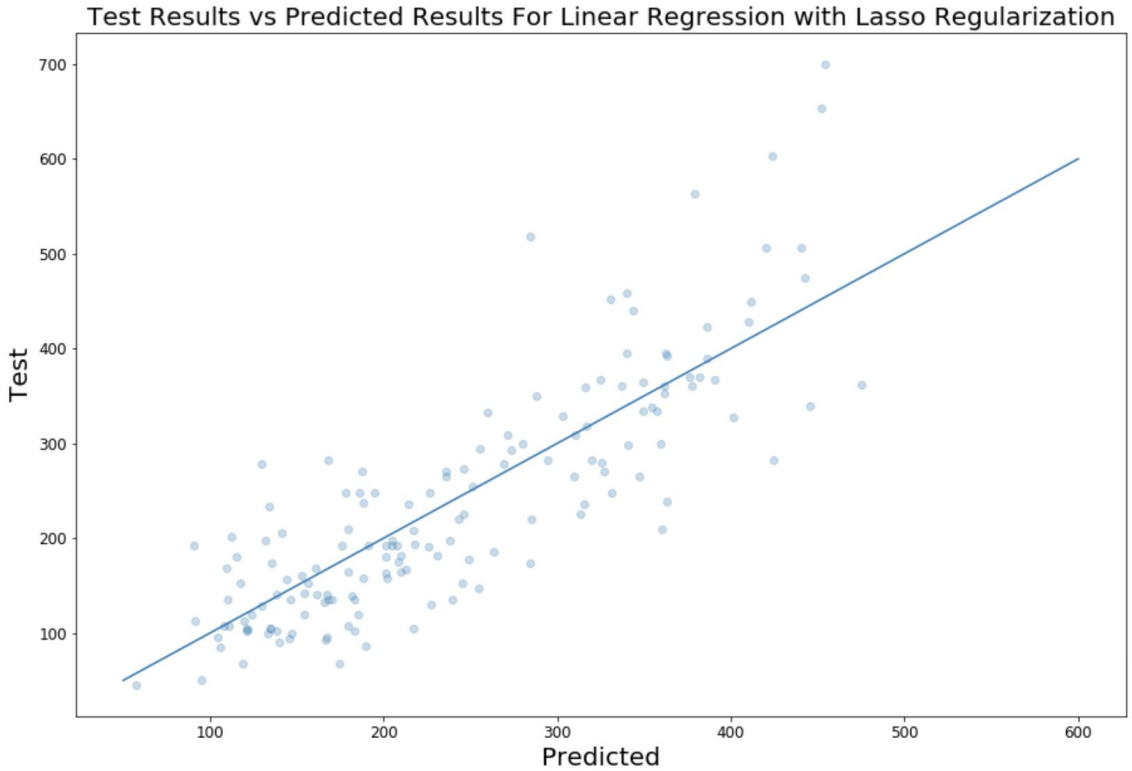
Linear Regression with Lasso Regularization

R^2 : 0.706

MAE: $50/Night
Mean: $250/ Night

Notable Coefficients:

- Guests: **$ 48**
- Beds: $ 8
- Overall Review: **$ 12**
- Cleanliness: $ 5
- Downtown: $ 9
- Uptown: **- $ 27**
- Private Room: **- $41**
- Shared Room -**$ 32**
- Shared Bath - **$ 20**
- **Intercept: $240**



Takeaways:

Hosts:

- Highest Guest Accommodation **+$48**
- Beds **+$8**
- Overall Review Important **+$12**
  - Cleanliness
  - Descriptions/Accuracy
  - Communication
- Private Bathroom **+$20**

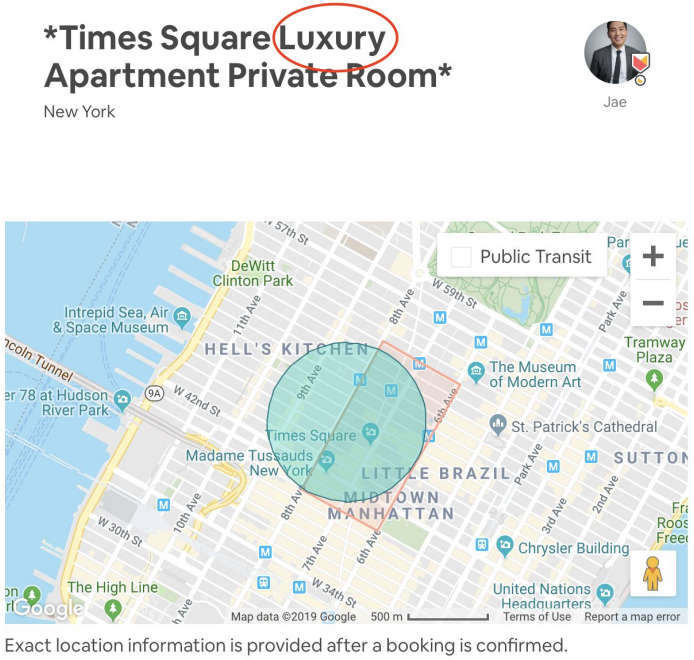
Guests:

- Private or Shared Listings **-$30/40**
- Uptown **-$27**
- Shared Bathroom **-$20**

Baseline price: **$240 /Night** in Manhattan
(Holiday Season)

Future Work

- GPS/Image Recognition
  - Photo of Google Maps Given
  - Exact Location Not Given
  - Proximity to Landmarks and Attractions
- Rental Price Comparison
  - Neighborhood Rental Rates
- NLP
  - Luxury Rentals
  - Special Amenities
- Host/Customer Breakdowns



Exact location information is provided after a booking is confirmed.

Questions?

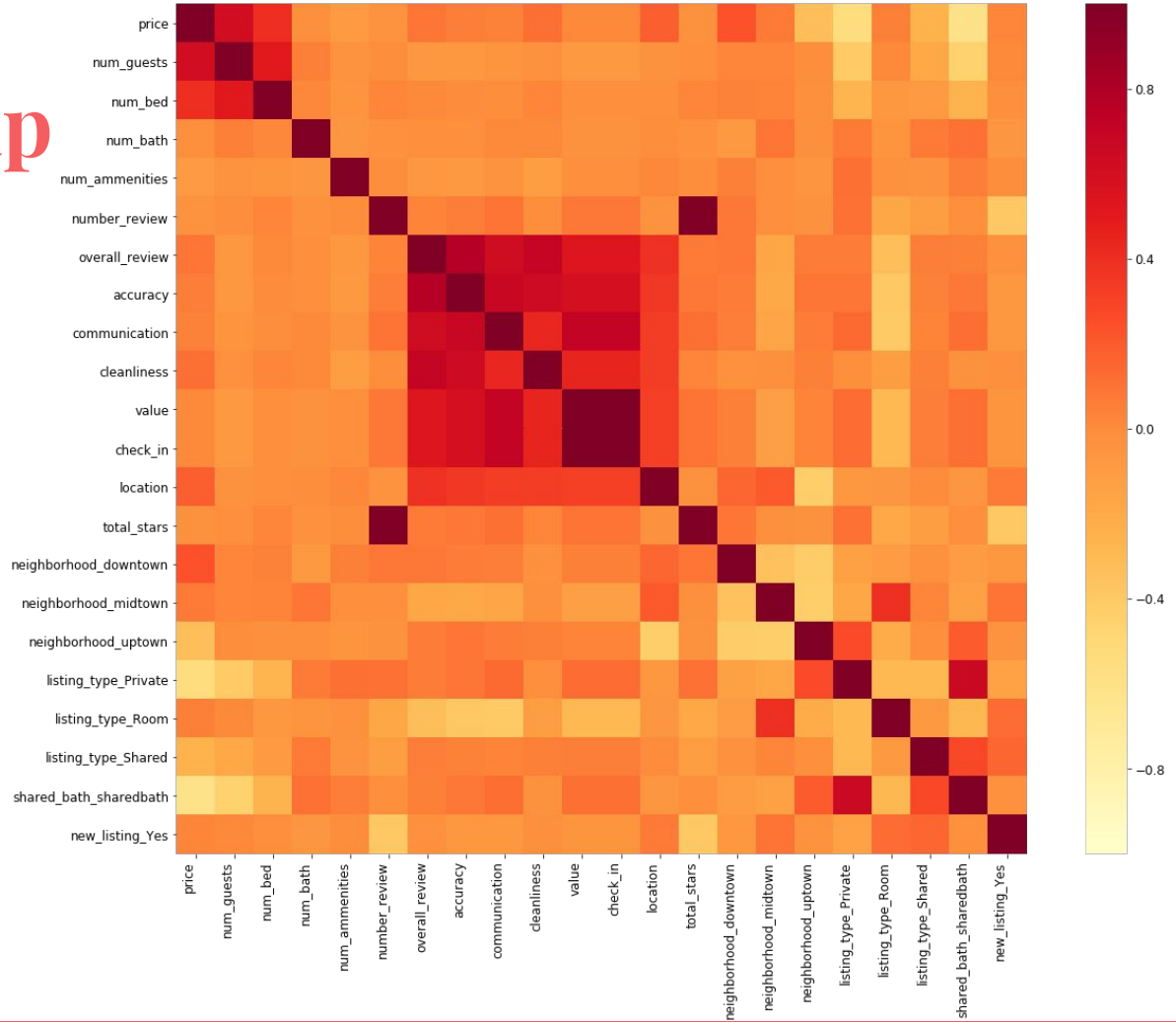
Appendix: Web Scraping

- Two Scrapers:
  - One for Search Result Links
  - Other for Data in listing links
- Dynamic Page
- Classes change
  - Resizing Window
  - Mouse Hover
  - Etc
- Bot Detection
  - RIP -- 3 IP addresses
  - Random Time Delays and Stops Necessary
- Selenium
  - Page needs time to load
    - Classes could be loaded, but data not populated

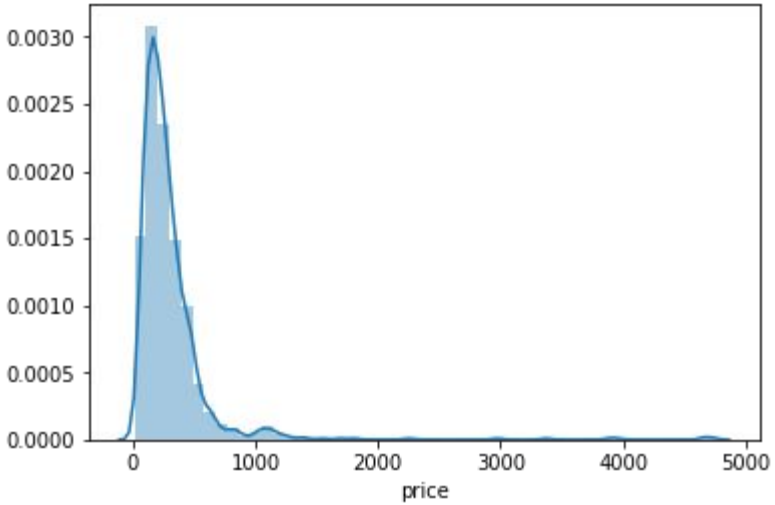
Appendix: Interesting Points

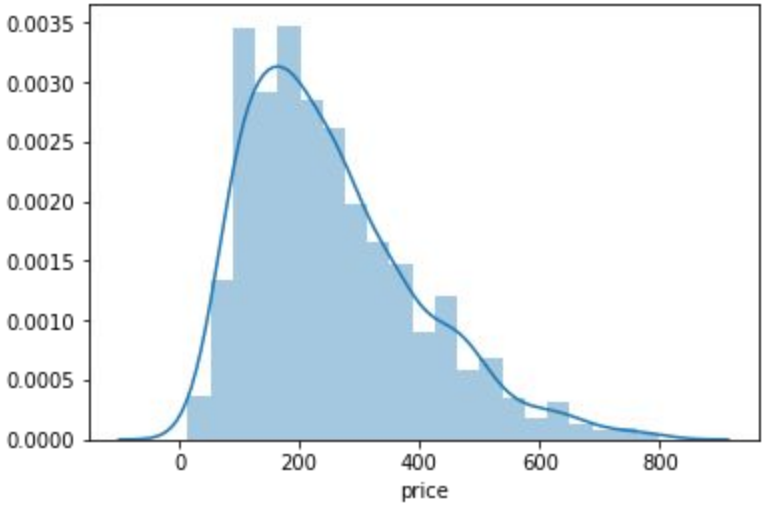
- Private Rooms are more negatively correlated with price than shared rooms
- Despite Being:
  - More Expensive on Average
  - Intuitively more valuable
- Private Rooms:
  - Many Private Rooms Have Shared Baths
  - Most Uptown Listings are Private Rooms
    - Uptown listings are much cheaper
  - Collinearity of these factors could contribute to the greater negative correlation

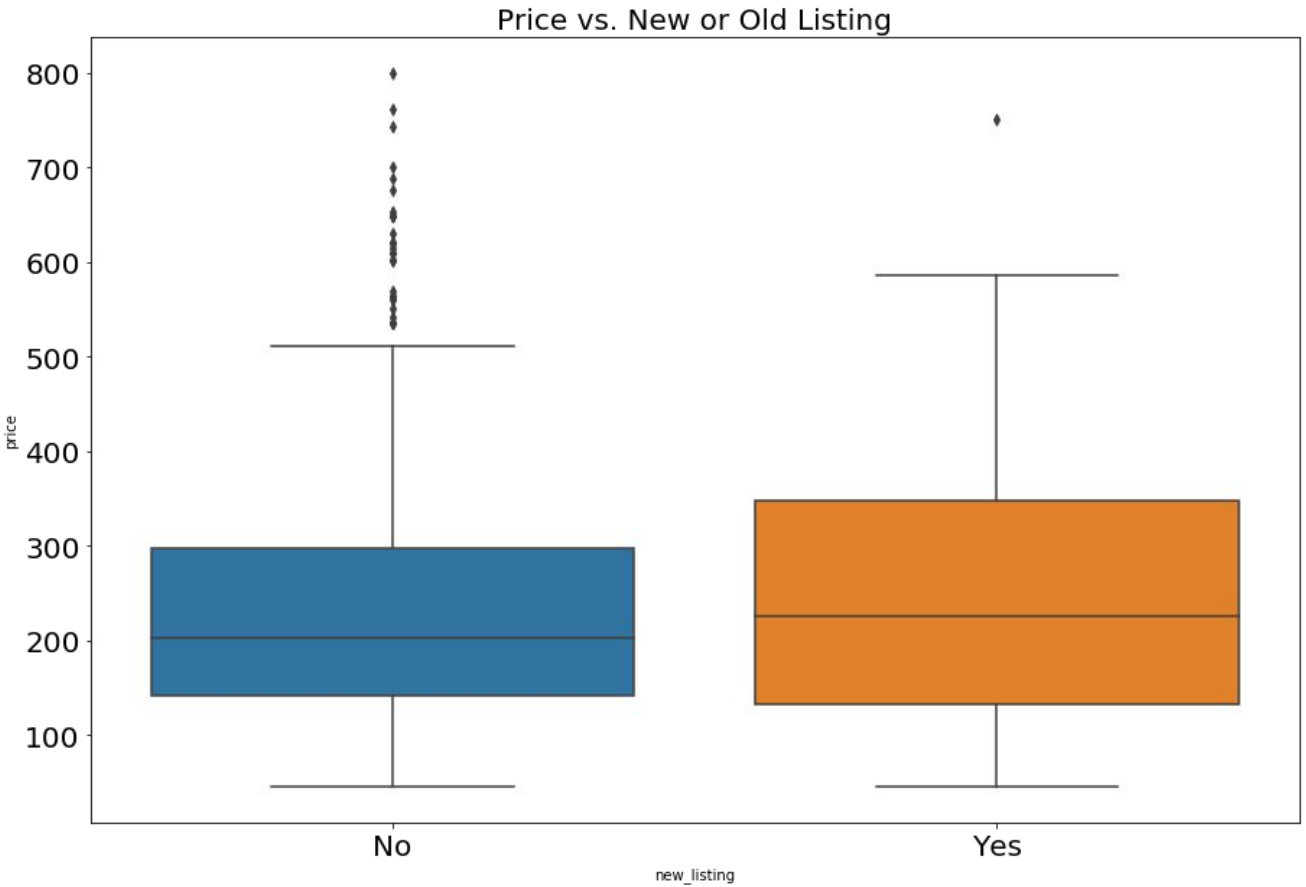
Full Heatmap



Appendix: Distribution of Price






Price vs. New or Old Listing
800
700
600
500
400
300
200
100
price
No
Yes
new_listing