



# Price Analysis of Airbnb Listings - New York 2019

Bhuvan Malladihalli Shashidhara (msbhuvan@uw.edu)

Data 512: HCDS Final Project Presentation

# Data Description and Methodology



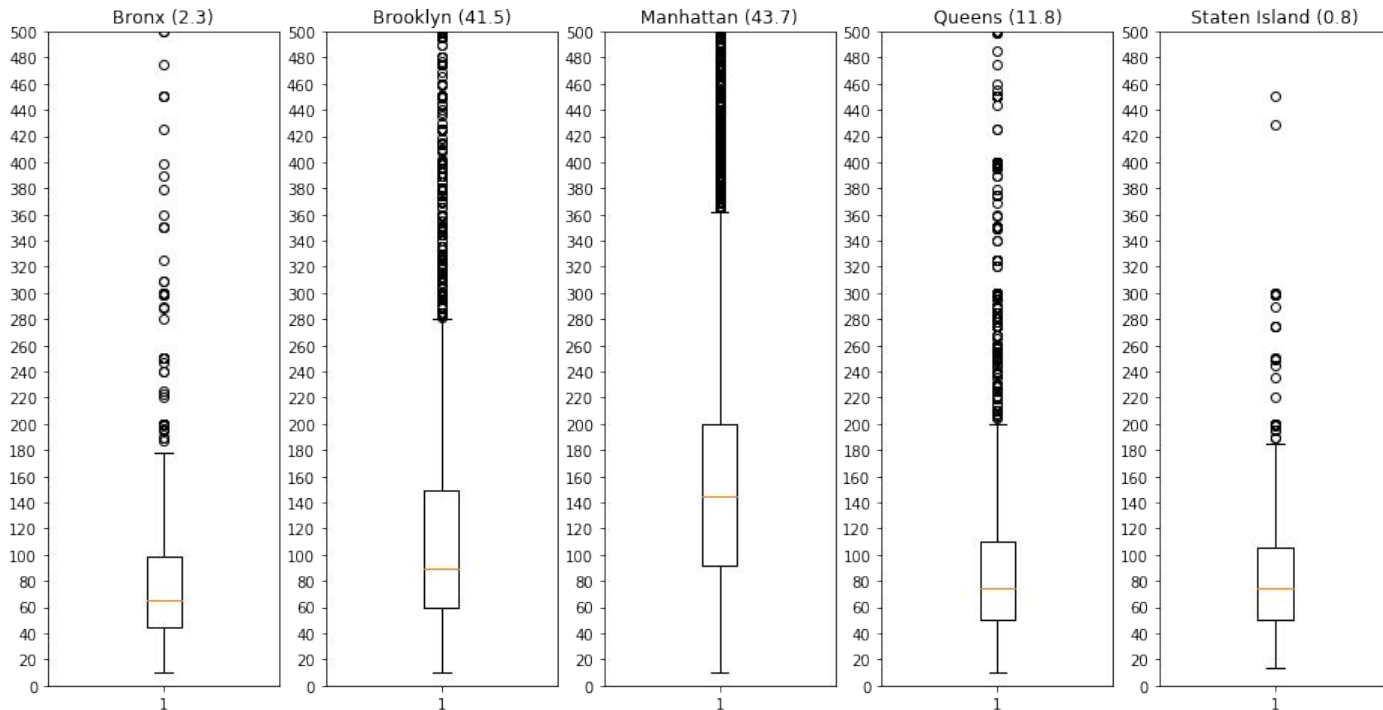
- [Kaggle dataset](#) (2019) released under [CC0 licence](#): Open data ([disclaimer](#)).
- Contains sensitive information - Personally identifiable:
  - Host name.
  - Listing name.
- **Anonymization:**
  - Removed the above two attributes.
  - Analysis and redistribution of anonymized version of the data.
- **Human-Centred Approach:** Types of stakeholders
  - Tenants or Tourists
  - Hosts
- **Scope:** Considering listings priced less than \$500 only.



# Tenant-Centered Research Questions

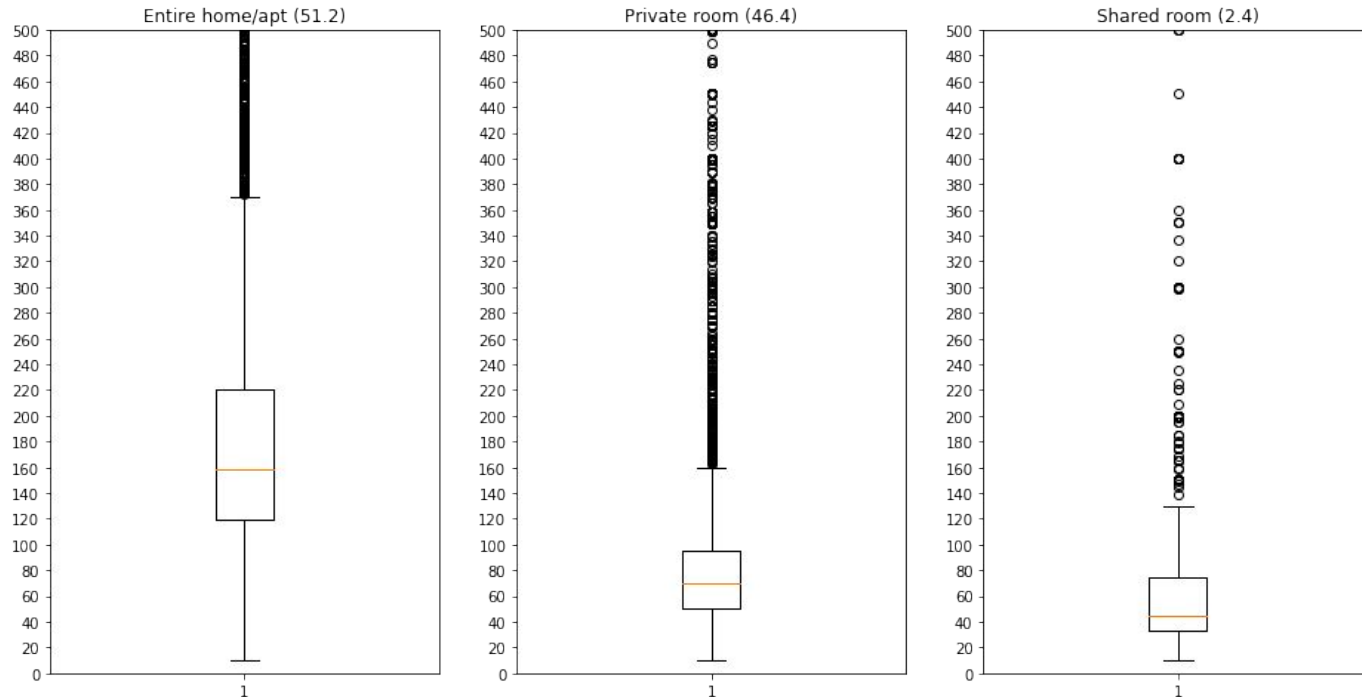
# Q1: How does the Airbnb prices distributed across the diverse set of neighborhoods in New York City?

Airbnb price distribution at different neighborhood groups (% prevalence) in New York - 2019



## Q2: How does the type of the property - shared room, private room or entire house - affect the prices of Airbnb listings in New York City?

Airbnb price distribution for different Room Types (% prevalence) in New York - 2019





# Host-Centered Research Questions

## Q3: Can we predict prices for Airbnb listings based on selected relevant attributes? What can we understand from the prediction model?



### Preprocessing:

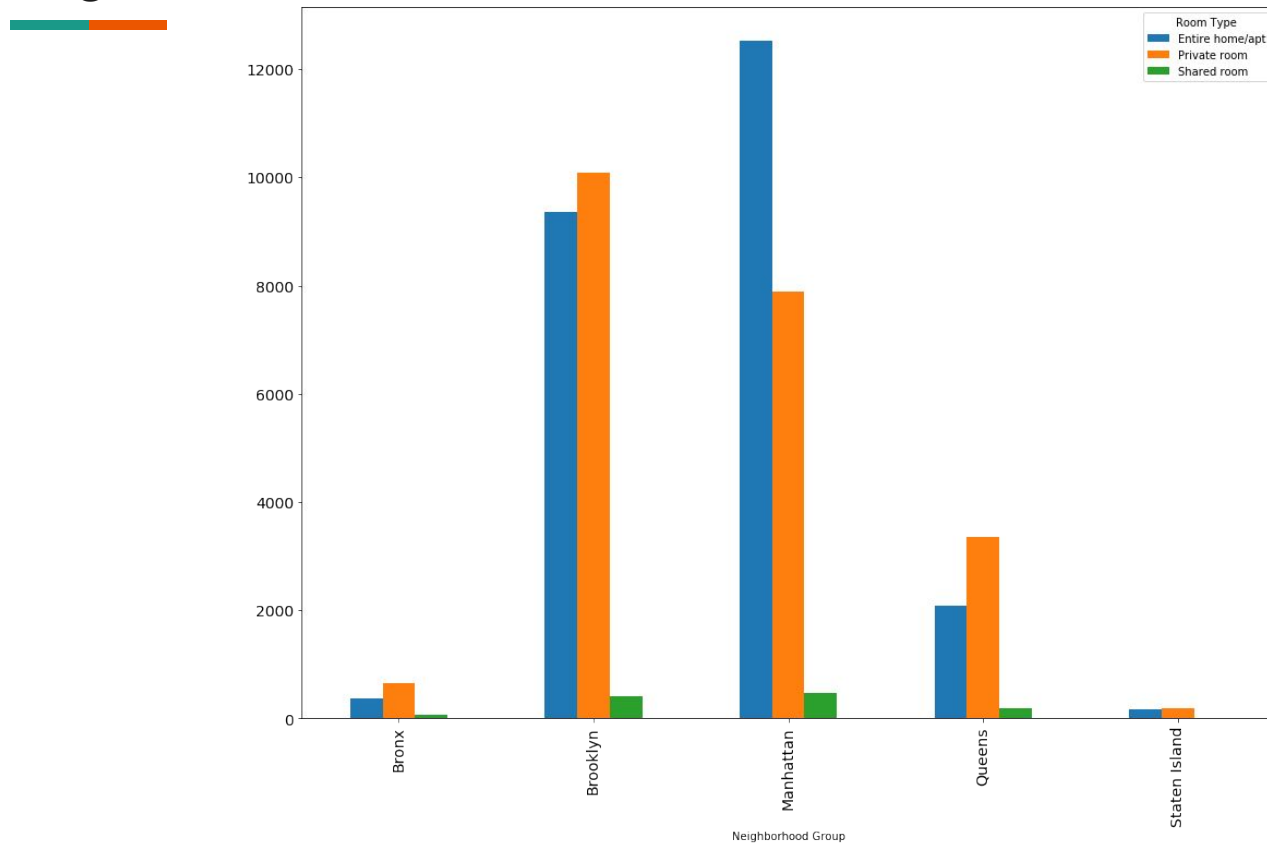
- Categorical Variables: One-hot encoding
  - Neighborhood group - Brooklyn, Bronx, Queens, Manhattan, Staten Island
  - Room type - Entire apt, Private room, Shared room
- Numeric: Z-score (mean center and scale by standard deviation)
  - Minimum nights
  - Number of reviews
  - Number of listings by same host

### Modeling: Linear Regression by Ordinary Least Squares (OLS)

### Results:

- R-squared: 0.38
- Lot of variance (outliers) in the price data, cannot be explained by linear models with these set of features.

## Q4: How are different types of property distributed in a given neighborhood?





# Conclusion and Future Work



- Provides an overview of the general trends and insights to tenants or tourists and hosts.
- Clear intent of the consumption of analysis due to stakeholder centric research questions.
- Limited scope, further analysis needed for drilled down insights!
- Alternate Research Potential: Various other attributes
  - Geo location: Latitude and Longitude
  - Minimum nights and Availability
  - Number of reviews
- Checkout the data and explore if interested!



# Thank you

Questions?