Airbnb Price Prediction City of Boston

Group 6 - Final Group Project
August 2021

Team Members

- Syed Ahmed
- Pascal Duchesneau
- Raissa Fondjo
- Geetha Shanthibushan
- Lalchand Shivraj

Overview

- Customers select rental based on price, convenient location, reviews, and household amenities.
- Estimating a price for short time stay lodging is a difficult task.
- Therefore, our team decided to develop a price prediction model using machine learning.

Objective

- Build a model that predicts the ideal price of a property taking into account listing features.

Presentation

- Technologies
- Data Exploration & Analysis
- Database
- Machine Learning
- Dashboard

What we used

Dataset

kaggle

Database

PostgreSQL

mongoDB

Language

python

Tools

Visual Studio Code

jupyter

GitHub

Technologies

QUICK DBD

+ a b | e a u

HEROKU

GitHub Pages

Algorithms

Linear Regression

Decision Tree

Random Forest

Deep Learning

Preliminary Data Analysis

What are the key factors that affect the price of the rental property?

- Room Type
- Property Type
- Number of Reviews
- Amenities

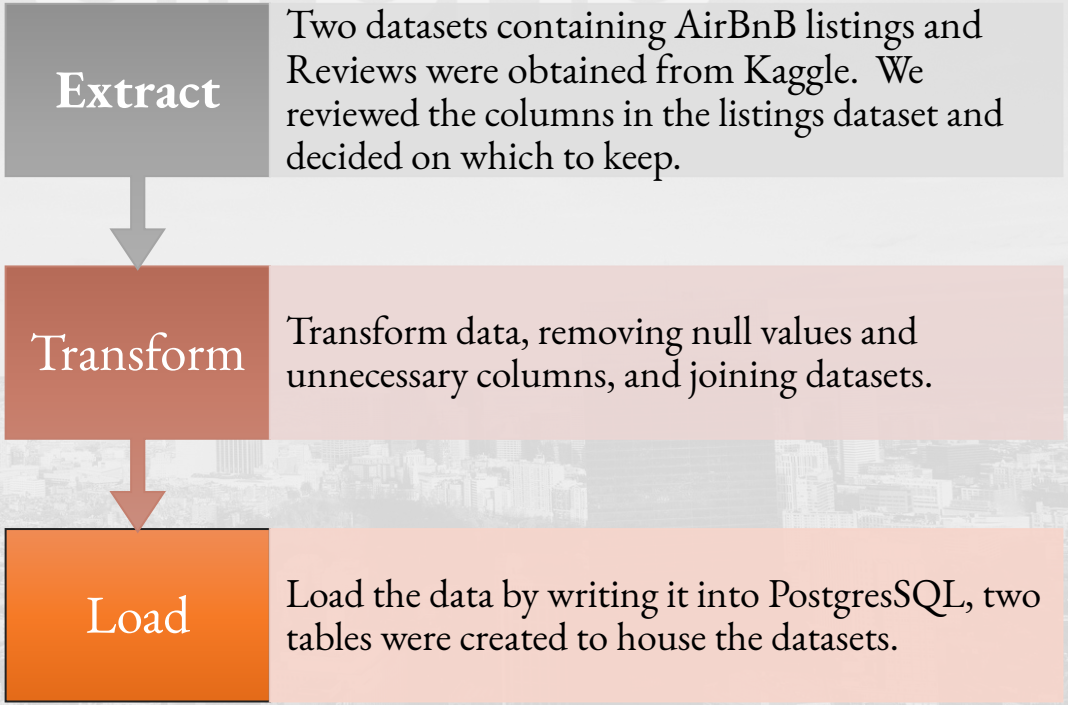
Which neighborhoods in Boston have the highest rental prices?

- Neighborhoods are separated into 25 districts.
- The price contributing factor differ significantly across the neighborhoods because of the variation in neighborhoods types.

Popular neighborhood comparison with Number of listings, Average Price, Review Scores

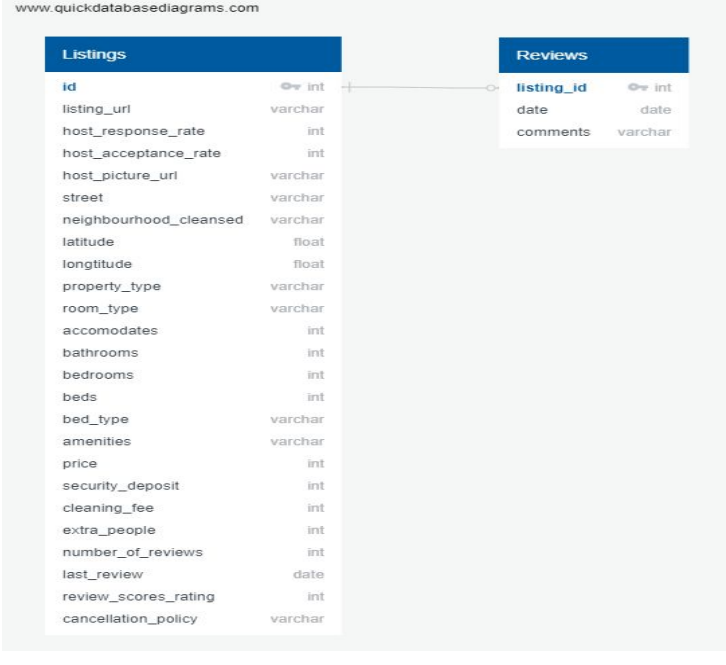
- Neighborhood with fewer listings have higher average listing price and review scored below median.
- Neighborhood with higher listings have lower average listing price and reviewed scored above median.

Data Exploration
The ETL Process



Database

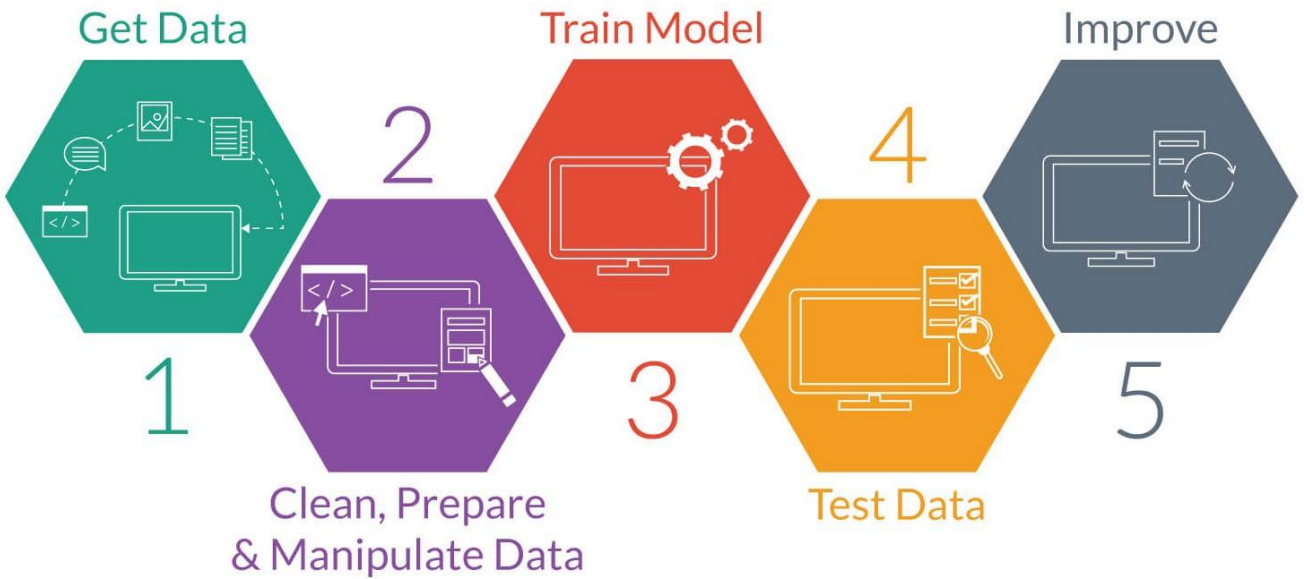
Used postgres by way of pgAdmin as our database to house the tables of data.



In MongoDB a cluster was created to access the data for multiple users.

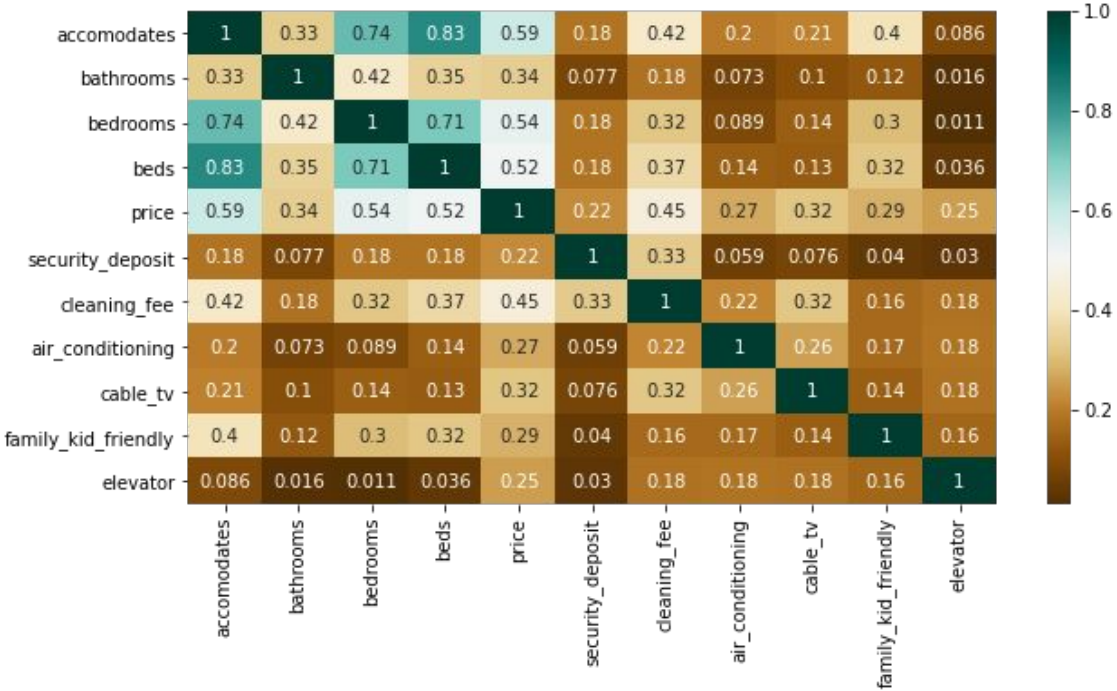
The dataset was then imported into a database created on MongoDB, using MongoDB Compass

Machine Learning



Exploratory Analysis:

Heatmap: Features Correlation

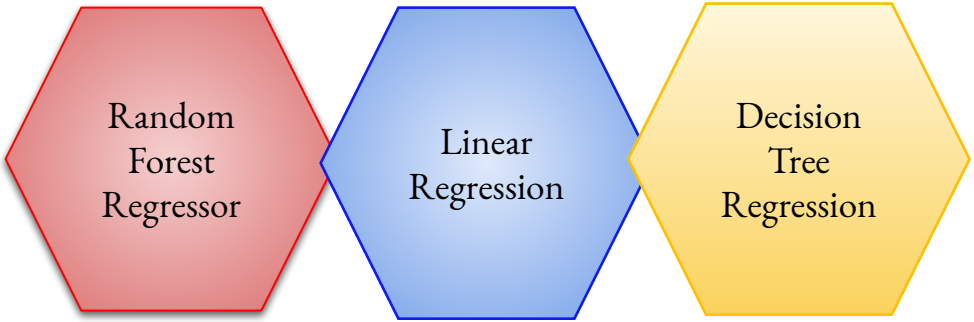


X variables included the features shown in the correlation matrix above, and our 'y' variable was price.

Split of Train & Test Data

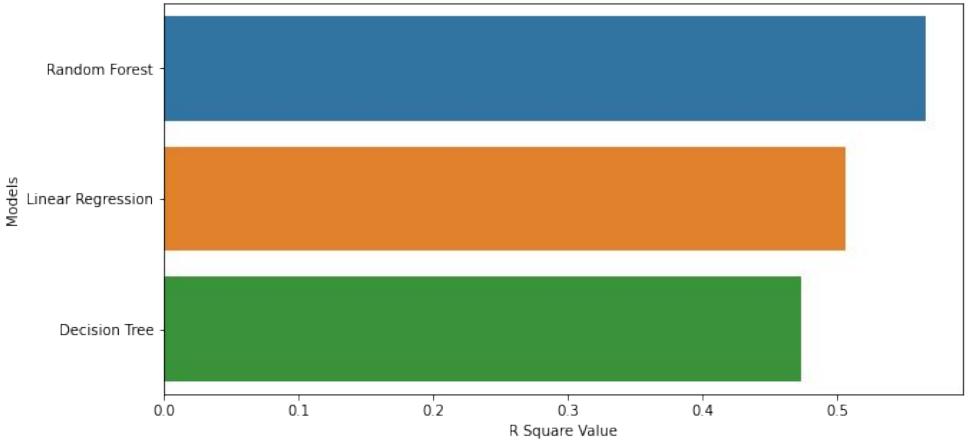
Statistical Modeling Methodology

Regression Models



Regression Models:

R Square Value



Model Optimization

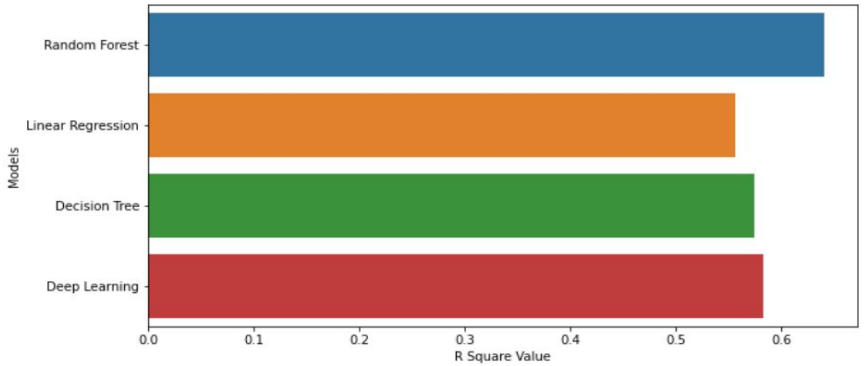
Statistical Modeling Methodology

Train Data 67%	Test Data 33%

Final Results

Final Results

Here is a chart comparing the R squared value of all the optimized models



- Random Forest R Square Value : 0.6401468962074119
- Deep Learning R Square Value : 0.5832696557887808
- Decision Tree R Square Value : 0.5747005190100378
- Linear Regression R Square Value : 0.5567145646136205

Recommendation

Recommendation for future analysis

- Use a logarithmic scale on the target variable
- Supplement data with more current data from Airbnb
- Add unsupervised learning to cluster patterns within the dataset, which could uncover relationships within variables
- Performing a qualitative assessment of how reviews affect Airbnb price

Dashboard

For the dashboard segment of this project, we will be using:

- The Visual Analysis: Tableau
- The Interactive Elements: Heroku
- The Website: Price Prediction

Improvements

What we would have done differently

- Explore more data sources to find more factors that could influence rental price.
- Data source from InsideAirbnb instead of Kaggle.

Airbnb Price Prediction City of Boston

Thank you!