Rental Price Prediction

Aarthe Jayaprakash, Mohit Gupta, Satyen Amorkar, Srinath Ramachandran

Problem Statement:

There is no good property valuation tool to help a host come up with a justified price to charge for their property listing based on amenties and other services offered. This analysis aims at providing such a tool for hosts so they can find a suitable price to charge.

Data Description

The dataset was retrieved from https://www.kaggle.com/navaneesh/airbnb and contains information about property listings and services offered.

Dimensions: 74,111 rows * 29 columns

Exploratory Data Analysis

Distribution of Price

N = 74111   Price in US Dollars($)
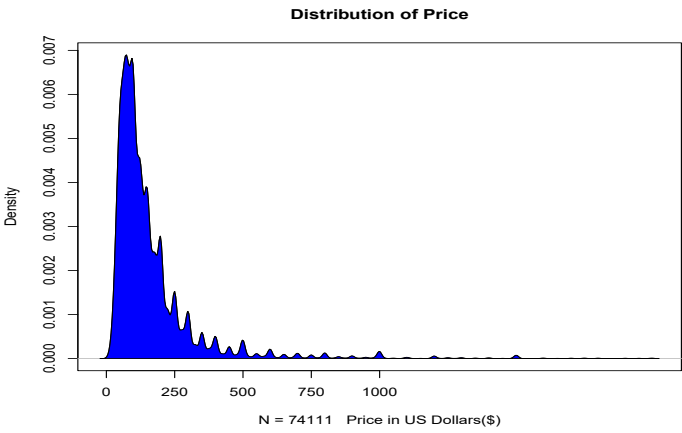
| summary	price
count	74111
mean	160.37
stddev	168.58
min	1
Q1	47.99
median	110.99
Q3	184.99
max	1998.99
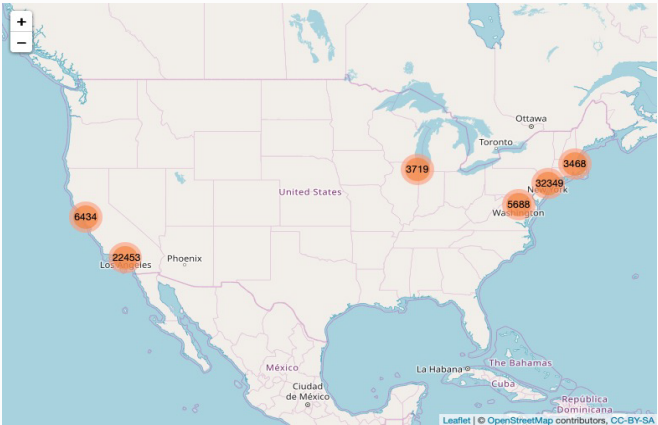
Data Preprocessing

We removed some redundant columns that did not provide useful information such as "thumbnail url", latitude, longitude, id's. Handled null values and converted all the columns to appropriate data types.

Feature Engineering

We created dummy columns from categorical variables such as amenties, property type, room type and cancellation policy so as to use them for regression models. We also converted dates to number of days to make it useful.

Pipeline

Vector Assembler → Standard Scaler → Models
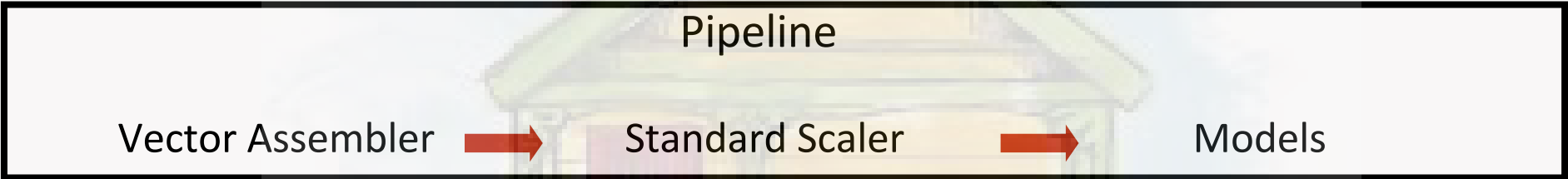
Linear Regression

A Linear Regression model describes a relationship between a dependent variable y and one or more independent variables x.

Evaluation Metric-RMSE
26.81

Random Forest Regression

A Random Forest is an ensemble technique capable of performing both regression and classification tasks. In this case, we have used Random Forest Regressor.
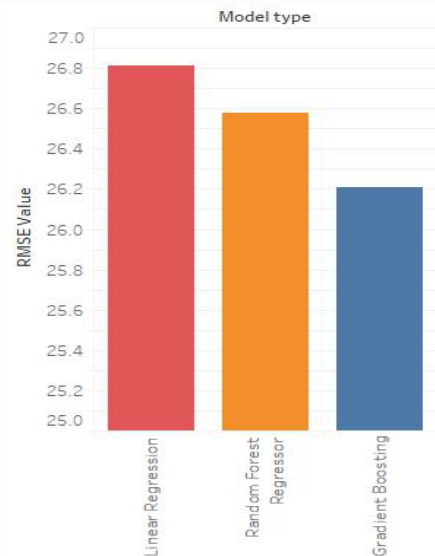
Evaluation Metric-RMSE
26.58

Gradient Boosting Regression

Gradient Boosting for regression builds an additive model in a forward stage-wise fashion; it allows for the optimization of loss functions.

Evaluation Metric-RMSE
26.21

Model Comparison

Inference