



RECOMMENDATION ENGINE FOR NYC THROUGH SENTIMENT ANALYSIS

Project developed by Nurgul Kurbanali kyz
5 December, 2022

Table of Contents:

- 1. Introduction*
- 2. Business Statement*
- 3. Data Understanding*
- 4. Sentiment Analysis*
- 5. Recommendation Engine*
- 6. Limitations & Future Consideration*



airbnb

Introduction

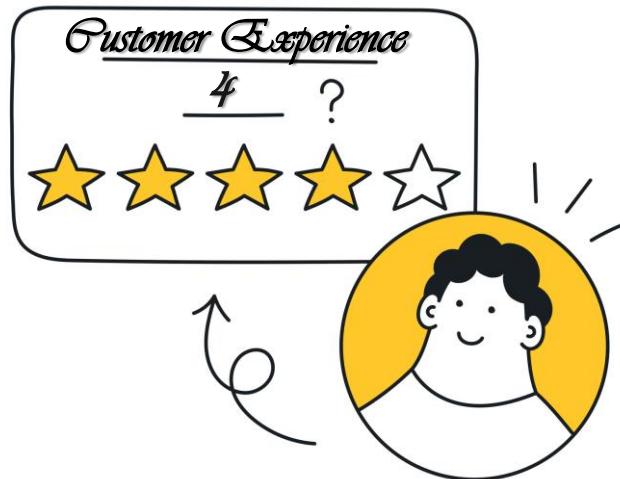
Airbnb, as in “Air Bed and Breakfast,” is an online platform that supports adding apartment/room to the platform as a host and booking an apartment/room as a user.

Given tens of thousands of available listings, competing for user’s attention, one of the most useful features that help attract users and ensure a constant repeat business flow is appropriate recommendation engine.



Business Statement

*The goal of this project is to analyze **sentiments** of users (based on their reviews) and recommend the most accurate **listings** (based on their preferences) for **Airbnb** users in NYC which stimulates demand and actively engage users. Recommendation Engine will be used by stakeholders(Airbnb) to increase their user experience by providing personalized user experience .*



Introduction

Customer Reviews play an important role in the customer's decision, and they are affected by other customers' reviews online, on blogs or over social networking platforms.

Sentiment analysis allows to understand the nuances in customer reviews and pinpoint where an issue (or positive experience) stems from. It turns customer reviews into insights.

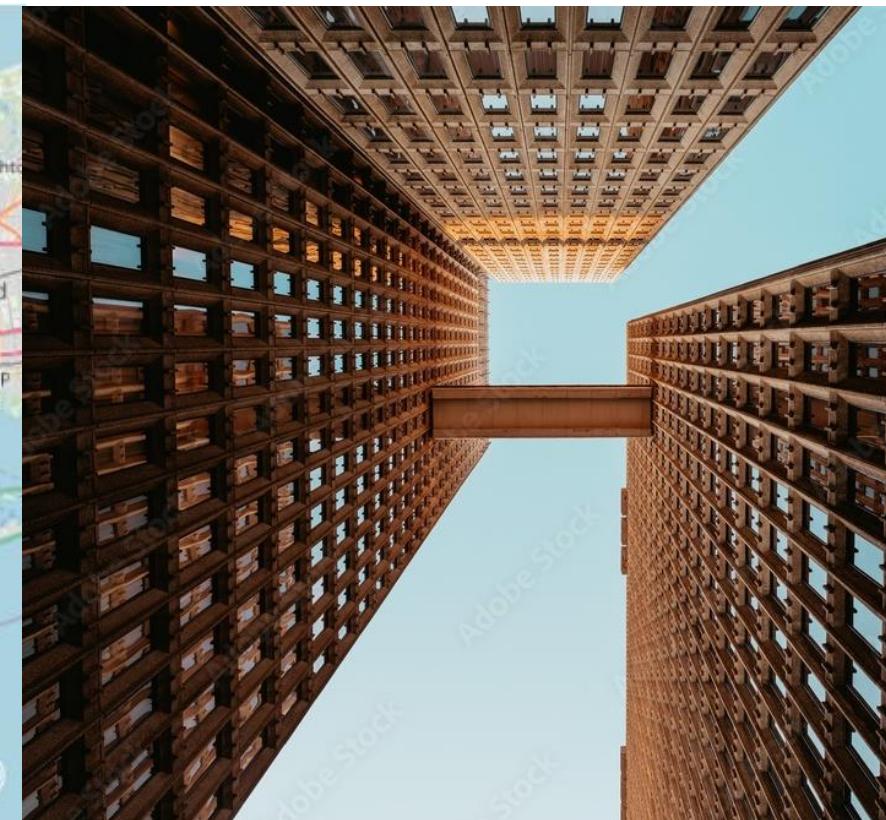
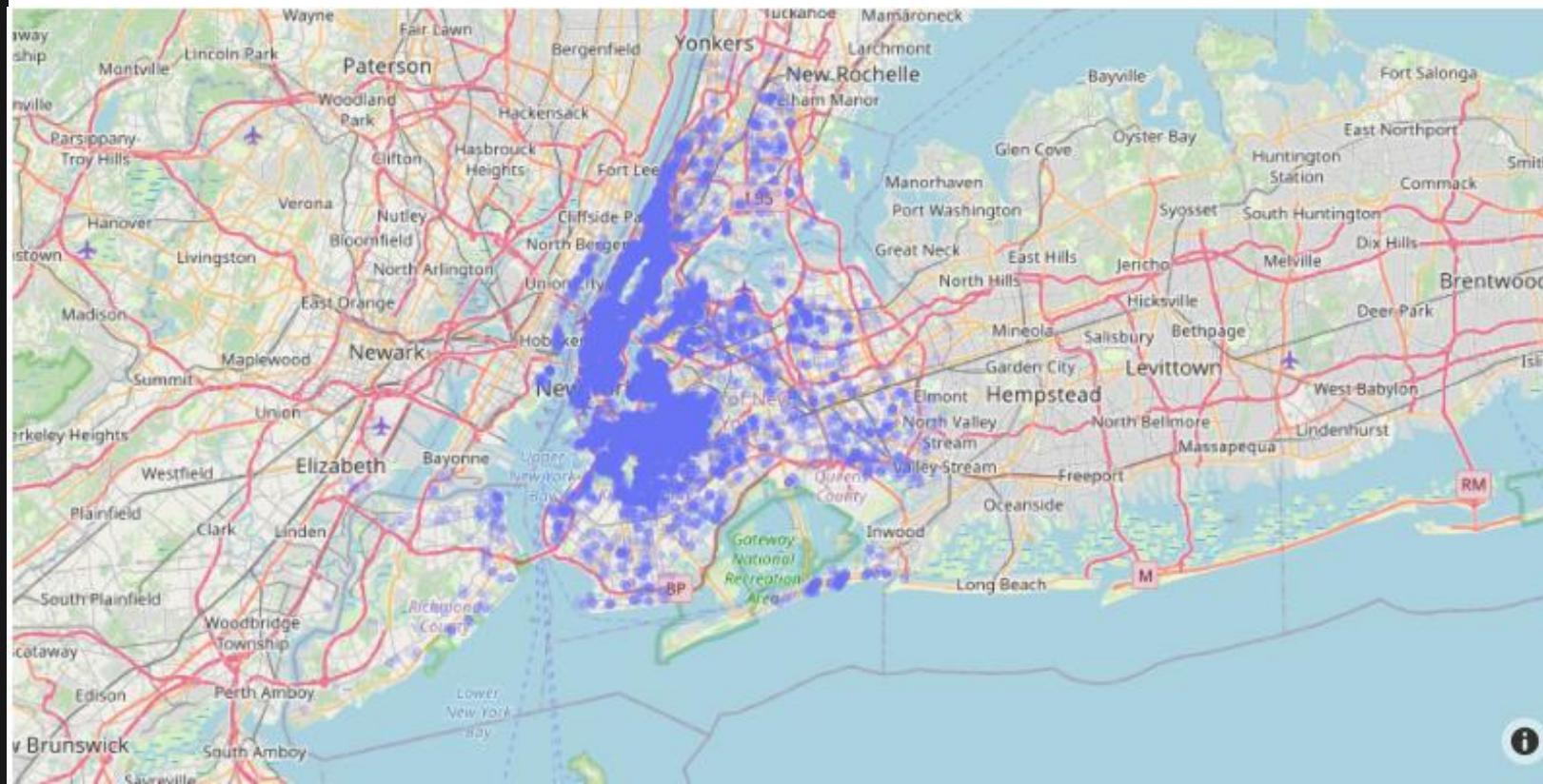


Data Understanding

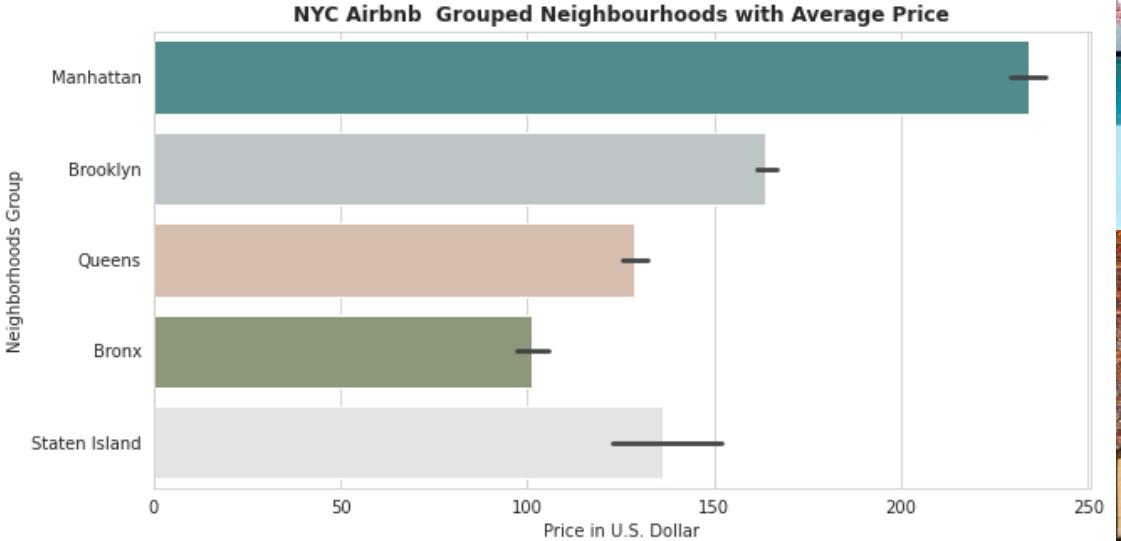
Data Source: [Inside Airbnb website](#)

Used Datasets: between Dec 2021 - Sept 2022

Used metrics from *listings and reviews* based on NYC



Data Understanding



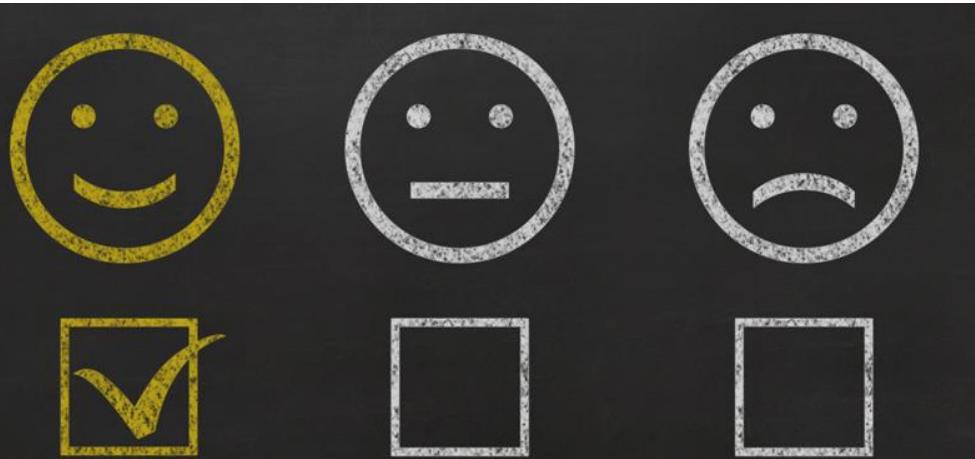
Top 30 Frequent Words About Amenities

Detailed description: This is a word cloud visualization titled 'Top 30 Frequent Words About Amenities'. The words are arranged in a grid-like pattern, with the size of each word indicating its frequency. The colors of the words vary, creating a visual hierarchy. The most prominent words include 'Air conditioning', 'Hot water', 'term stays', 'Wifi', 'Hangers', 'Carbon monoxide', 'Smoke alarm', 'Kitchen sound system', 'Bedroom door', 'Baking sheet', 'Clothing storage', 'Conditioner', 'TV Paid parking', 'Dishes', 'Dishwasher', 'Bathtub', 'Backyard HDTV', 'Lock', 'Stove', 'Microwave', 'Property', 'Parking garage', 'Darkening shades', 'Room darkening', 'Cable TV', 'Paid parking', 'TV', 'Carbon monoxide', 'Smoke alarm', 'Kitchen sound system', 'Bedroom door', 'Baking sheet', 'Clothing storage', 'Conditioner', 'TV Paid parking', 'Dishes', 'Dishwasher', 'Bathtub', 'Backyard HDTV', 'Lock', 'Stove', 'Microwave', 'Property', 'Parking garage', 'Darkening shades', 'Room darkening', 'Cable TV', 'Paid parking', 'TV', 'Carbon monoxide', 'Smoke alarm', 'Kitchen sound system', 'Bedroom door', 'Baking sheet', 'Clothing storage', 'Conditioner'.

Sentiment Analysis

Sentiment analysis (opinion mining), is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text.

Opinion mining can extract the **polarity** (amount of positivity and negativity)



Estimate Sentiment Polarity



VADER (Valence Aware Dictionary and Sentiment Reasoner)

Lexical database sentiment analysis tool that is optimized for social media sentiments.

It demonstrates the positivity and negativity scores, but also the degree (polarity) to which a sentiment is positive or negative

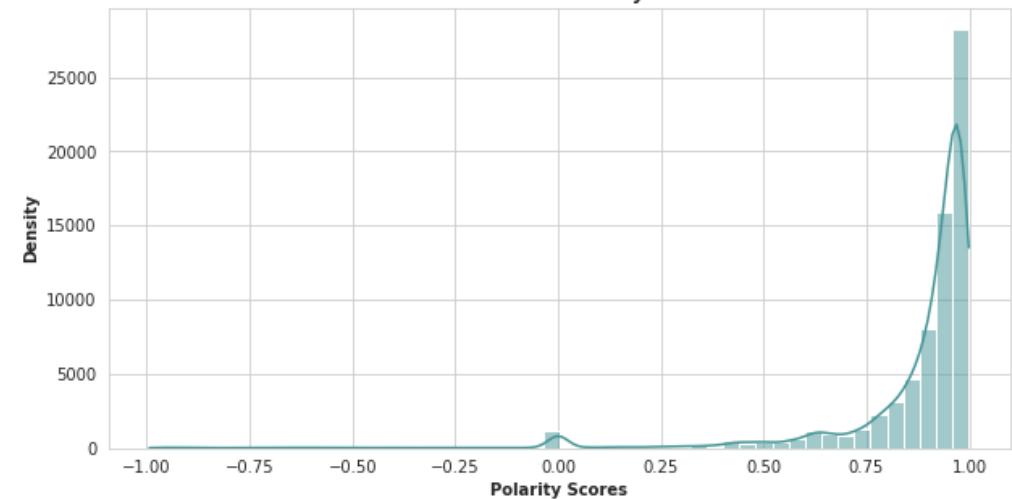
The compound score is the sum of positive, negative & neutral scores which is then normalized between -1 (most extreme negative) and +1 (most extreme positive).

Sentiment Analysis

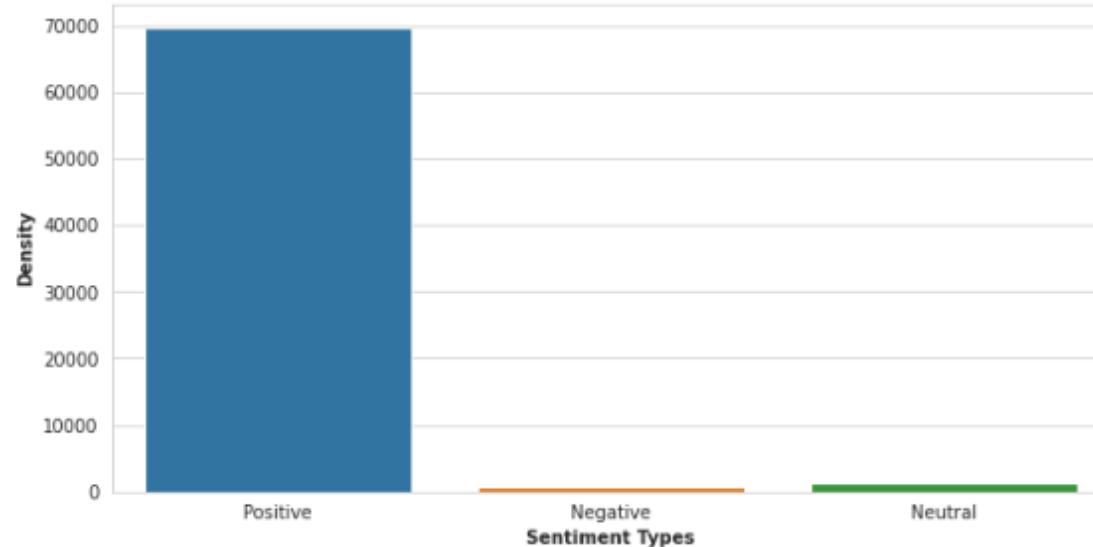


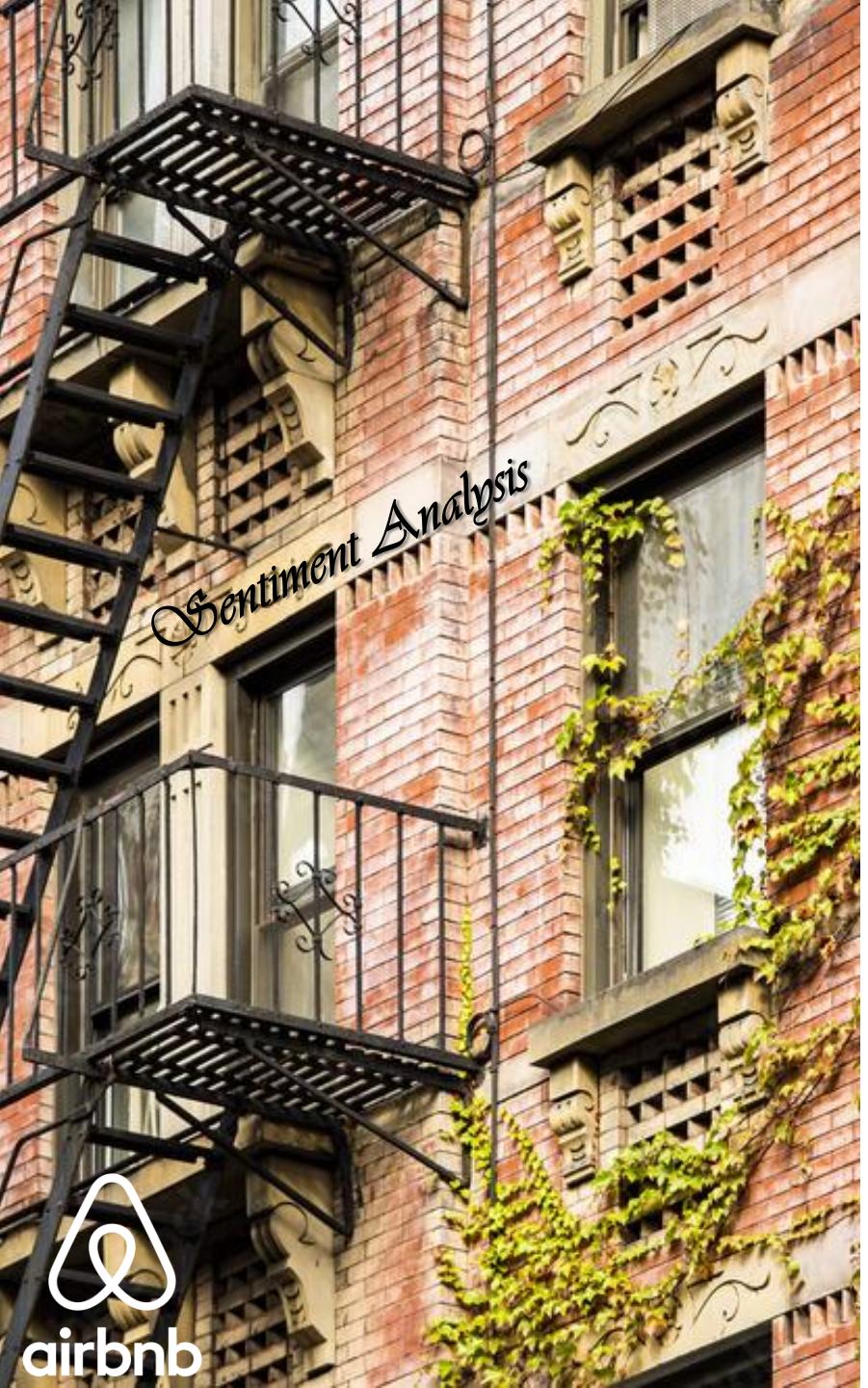
```
Positive      69674  
Neutral       1254  
Negative      708  
Name: sentiment_type, dtype: int64
```

NYC Airbnb Review Polarity Distribution



Type of Sentiment Polarity

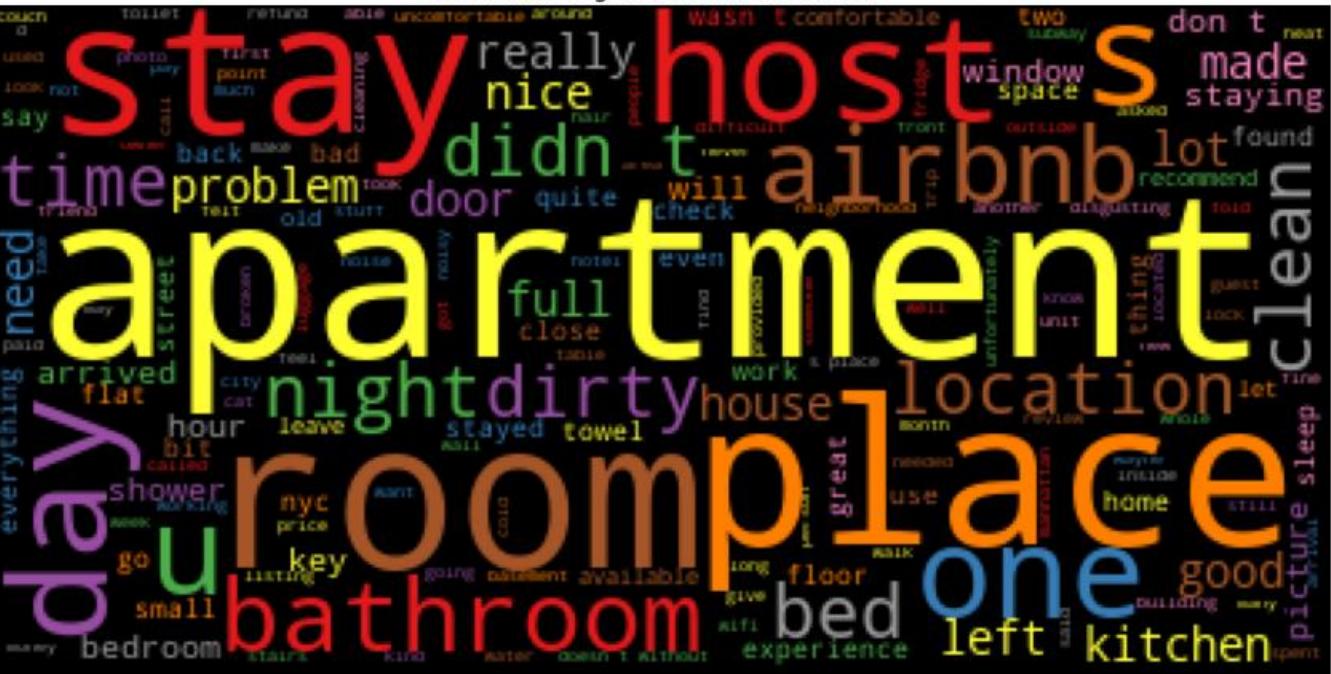




NYC Airbnb Positive Reviews Word Cloud



NYC Airbnb Negative Reviews Word Cloud



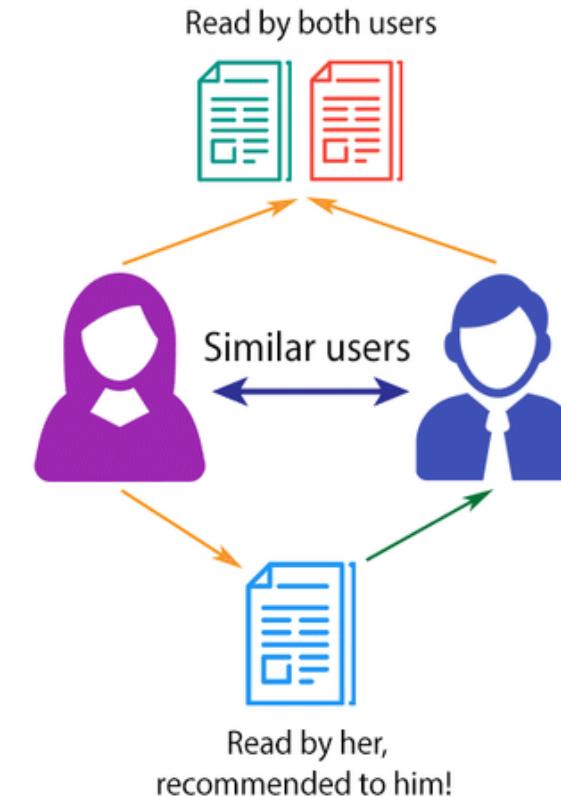
Recommendation Engine

The **Collaborative Filtering** method focuses on collecting and analyzing data on user behavior, to predict what a person will like, based on their similarity to other users.

The main idea for user-based CF is if two users have similar behavior in the past, e.g. have rated some items similarly, they will likely prefer similar items in the future.



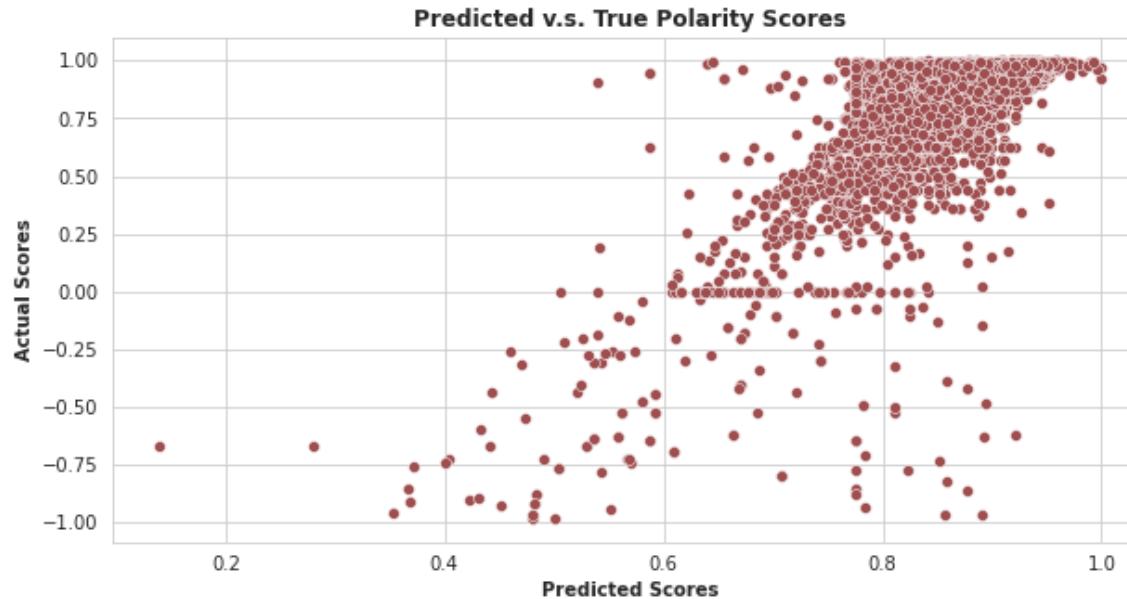
COLLABORATIVE FILTERING



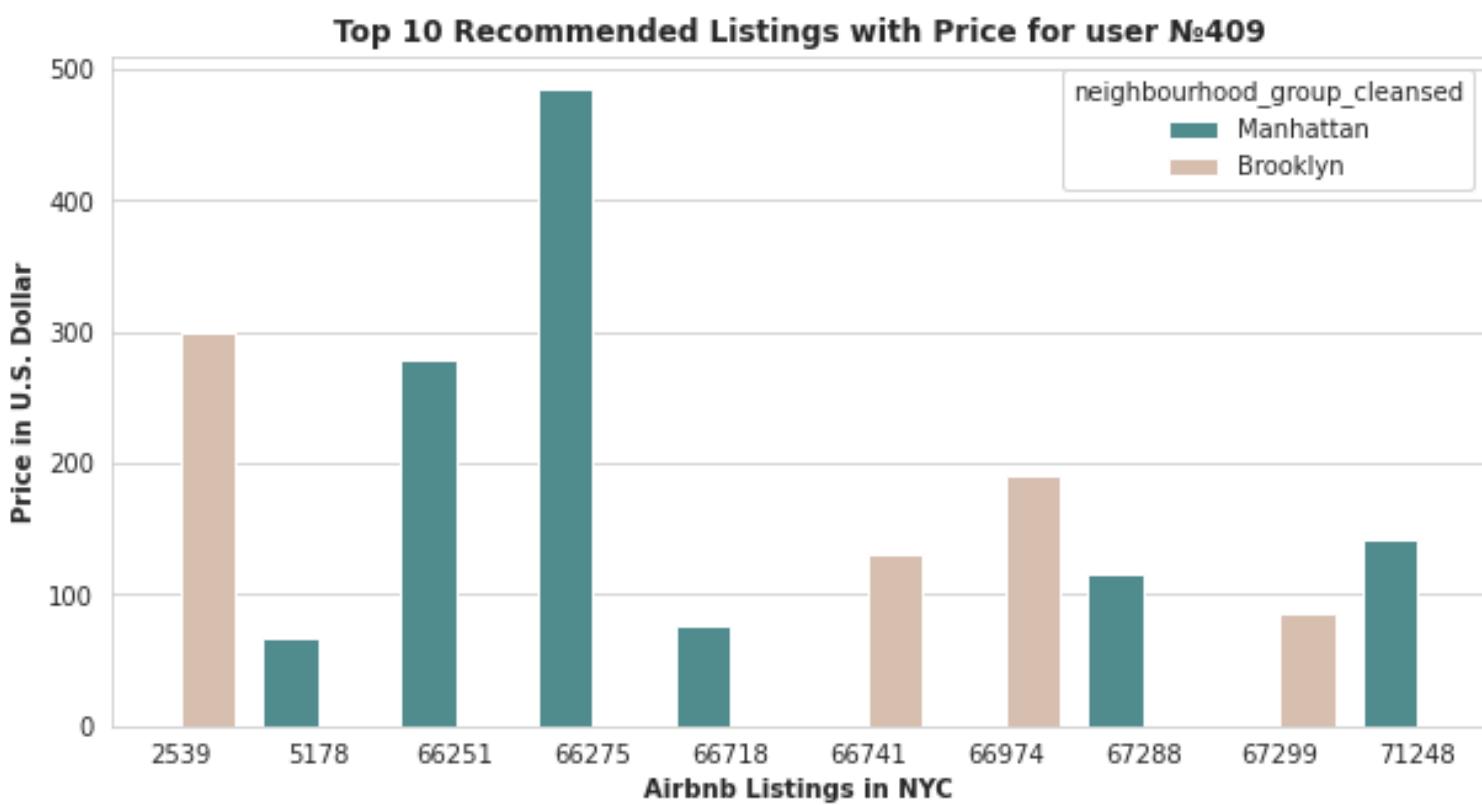
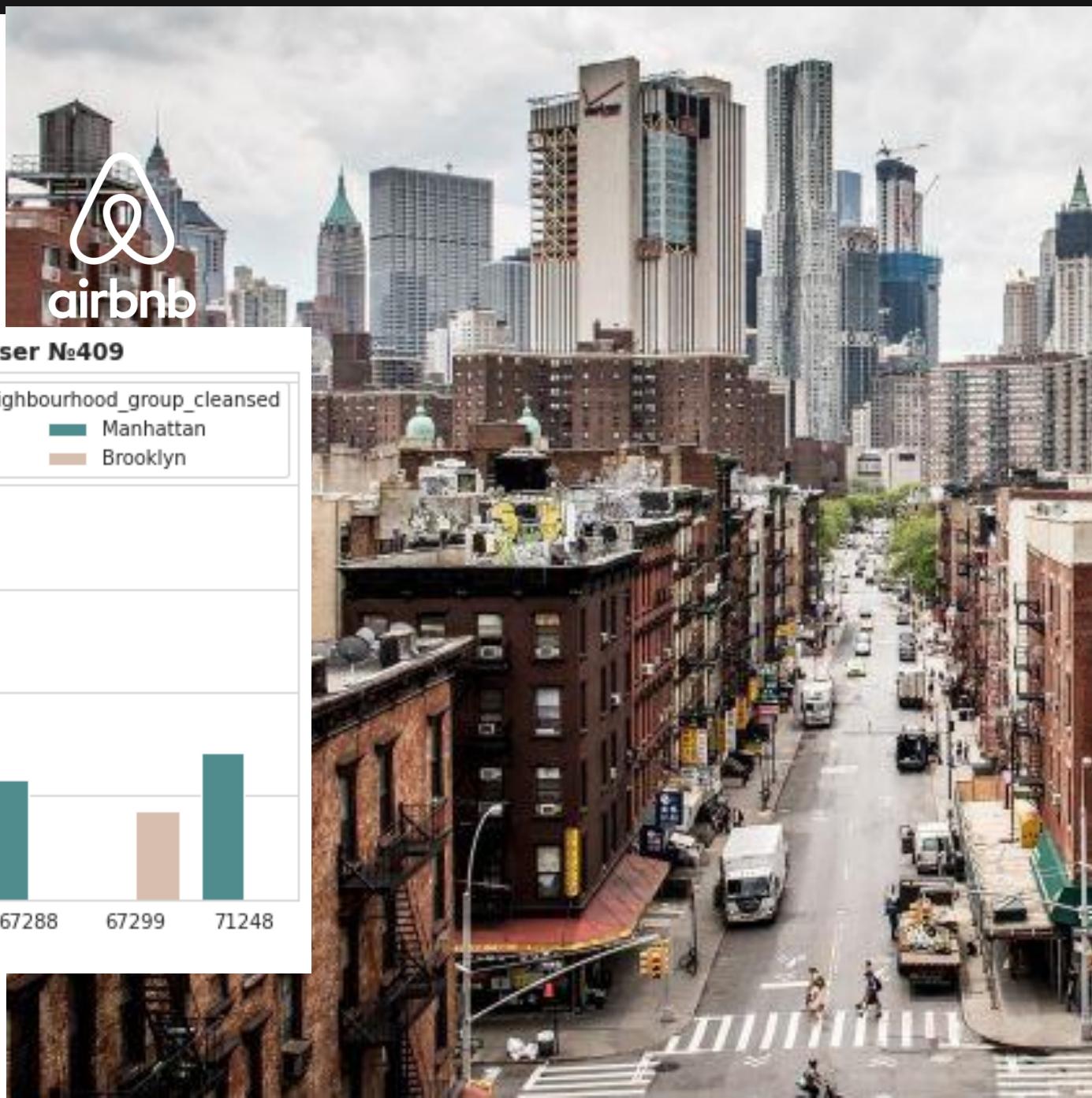
Recommendation Engine

The results show that our model outperforms in terms of the root mean squared errors of the generated predictions.

How far predictions fall from measured true values (RMSE : 0.1652)



Recommendation Engine



Use models which involve regularization like :

- 1. Matrix factorization using Alternating (optimization algorithms) least squares.*
 - 2. Matrix factorization using Bayesian Personalized ranking*
- Balance dataset with state-of-the-art oversampling or under sampling techniques such as: SMOTE during pre-processing*
 - Using Clustering Techniques- K-means and Self-Organizing Map (SOM)*
 - The time factor can play an important role in providing effective personalized recommendations and consequently improving the accuracy of predictions.*

Limitations & future Consideration



Thank You!

Any Questions?



nurkamalova@gmail.com



@kamalova



www.linkedin.com/in/nurgul-k-824876195

Travel in New York ☺

