



airbnb

# RECOMMENDATION ENGINE FOR NYC THROUGH SENTIMENT ANALYSIS

*Project developed by Nurgul Kurbanali kyzy*  
*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*

*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.*





# *Introduction*

*Airbnb, as in “Air Bed and Breakfast,” is an online platform service that lets property owners rent out their spaces to travelers looking for a place to stay.*



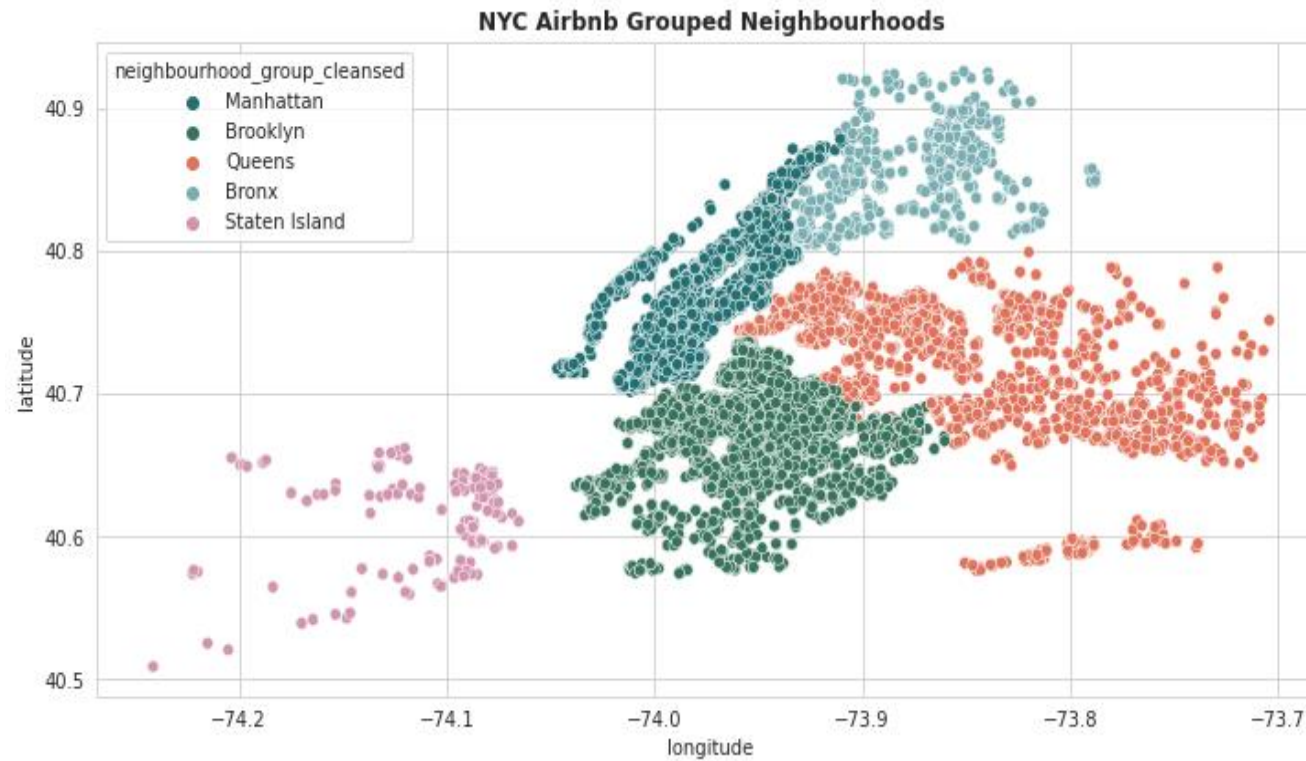
## *Business Statement*

*The main goal of this work is to analyze **sentiments** of users and **recommend** the most accurate listings for users based on their preferences in New York City.*





# Data Understanding



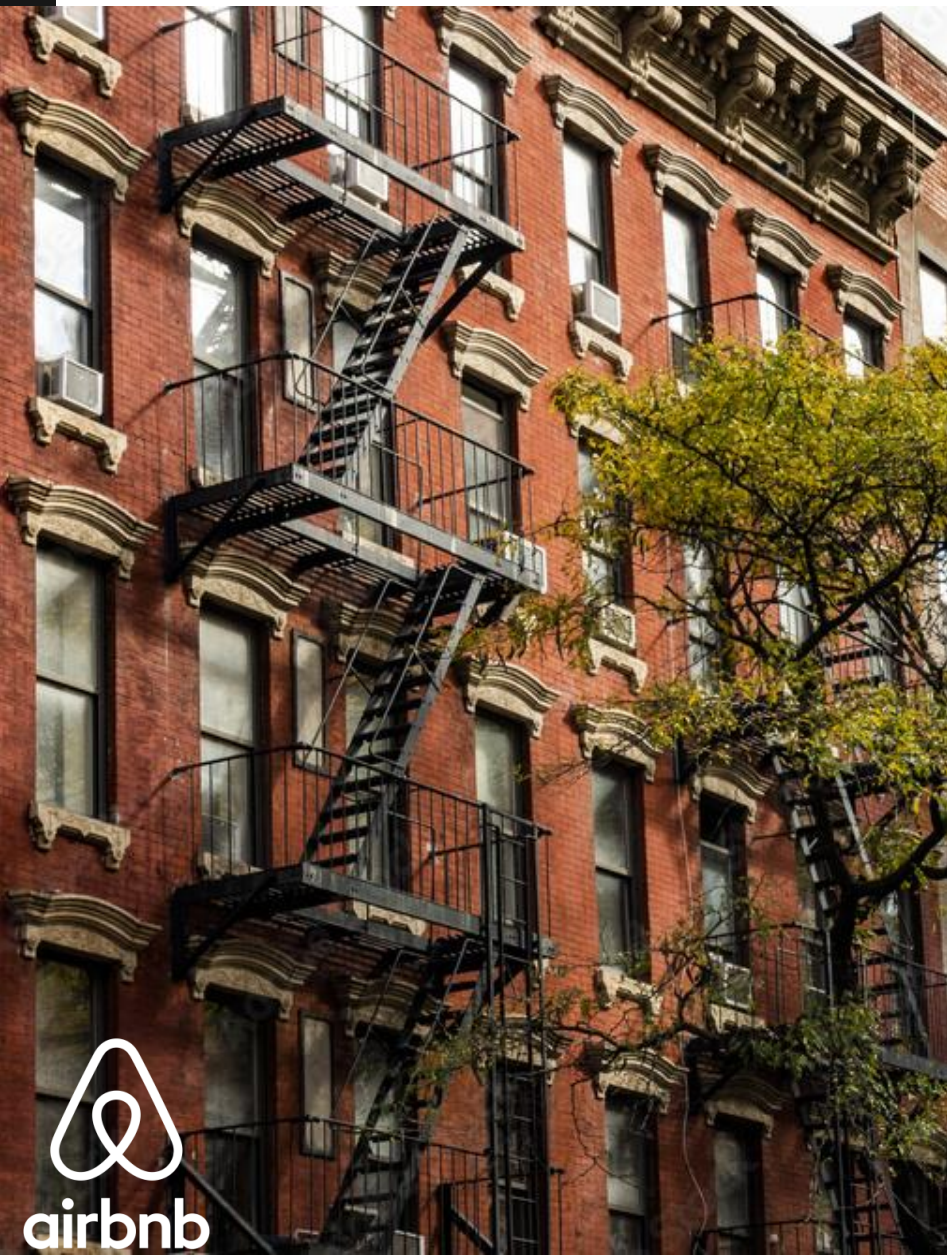
data source: *Inside Airbnb website*

used datasets: between Dec 2021 - Sept 2022

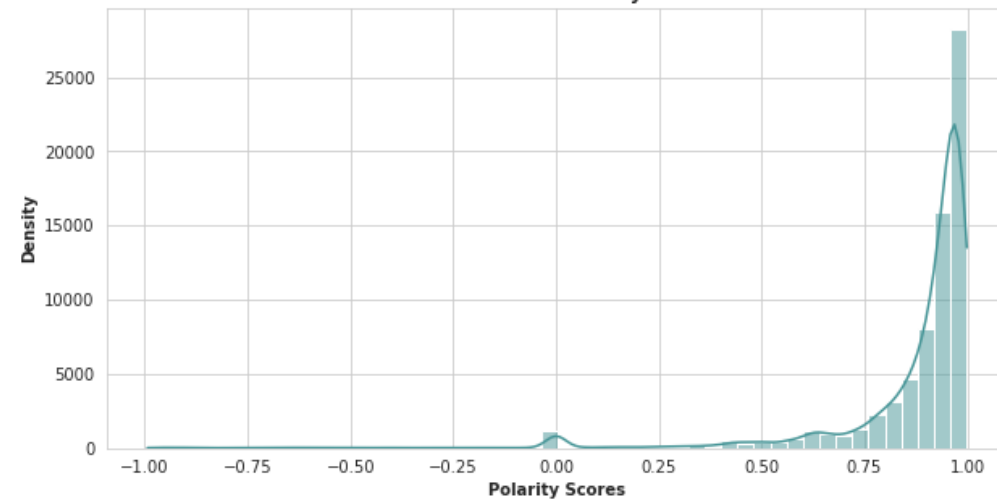
Used metrics from *listings and reviews* based on NYC



# Sentiment Analysis

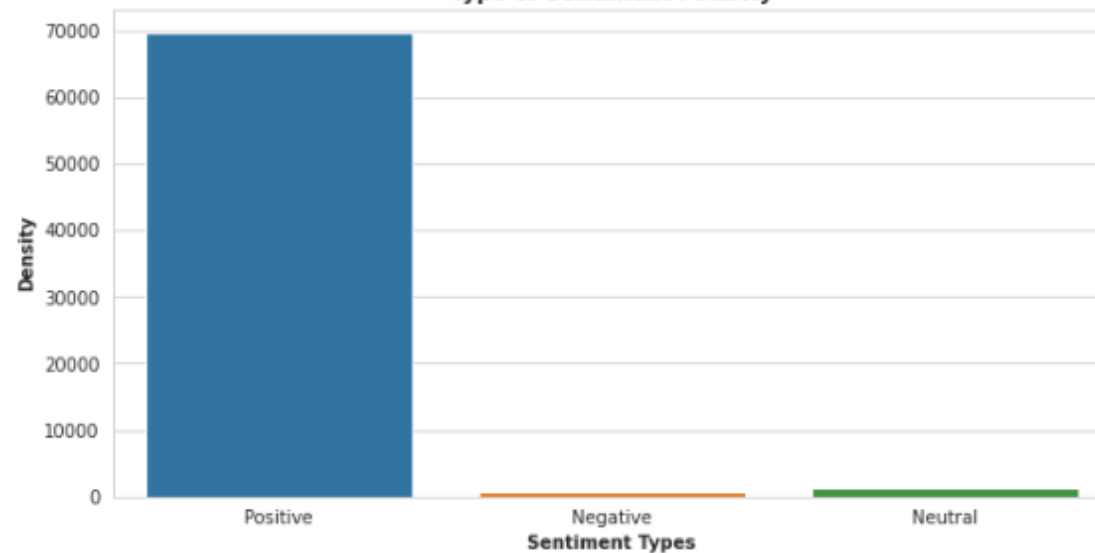


NYC Airbnb Review Polarity Distribution



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

Type of Sentiment Polarity



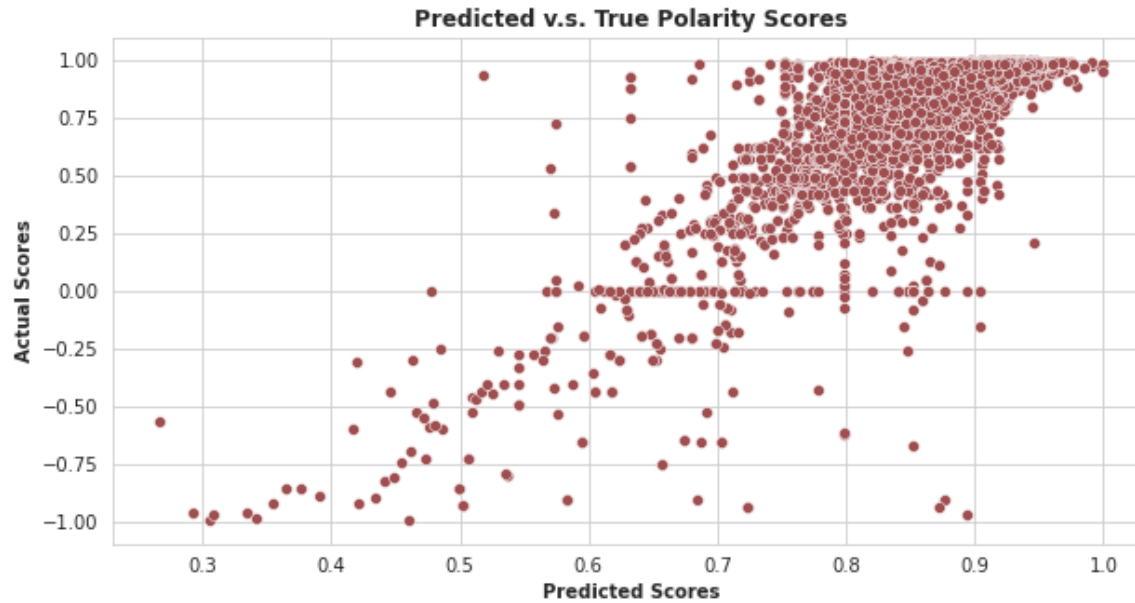






# *Recommendation Engine*

*How far predictions fall from measured true values (RMSE: 0.1629)*





# Recommendation Engine



*Sample Recommend listings for up to 10 Users along with their sentiment scores*

	user_id	recommended_listing - score
0	65425	[(140684, 0.9901070001691519)]
1	12192	[(504513, 0.8895133763758871)]
2	26785	[(5048661, 0.960613585255367)]
3	45936	[(6341938, 0.9078099828569064)]
4	16595	[(97130, 0.9265327691257963)]
5	38663	[(6364722, 0.9210822262482603)]
6	14314	[(63563, 0.9375356210693635)]
7	64365	[(43889727, 0.923928927426699)]
8	39572	[(950389, 0.9224122212787231)]
9	62427	[(9604553, 0.945914295478733)]



## *Limitations & Future Consideration*

*In order to produce better sentiment accuracy and increase models' performance*




- consider domain sensitivity information*
- get more data (+negative reviews)*
- balance dataset*
- try another common unsupervised learning algorithms*





# Thank You!

*Any Questions?*

-  [nurkamalova@gmail.com](mailto:nurkamalova@gmail.com)
-  [@kamalova](#)
-  [www.linkedin.com/in/nurgul-k-824876195](https://www.linkedin.com/in/nurgul-k-824876195)

*Travel in New York ☺*

