Prediction Challenge

By Romil Patel

Introduction

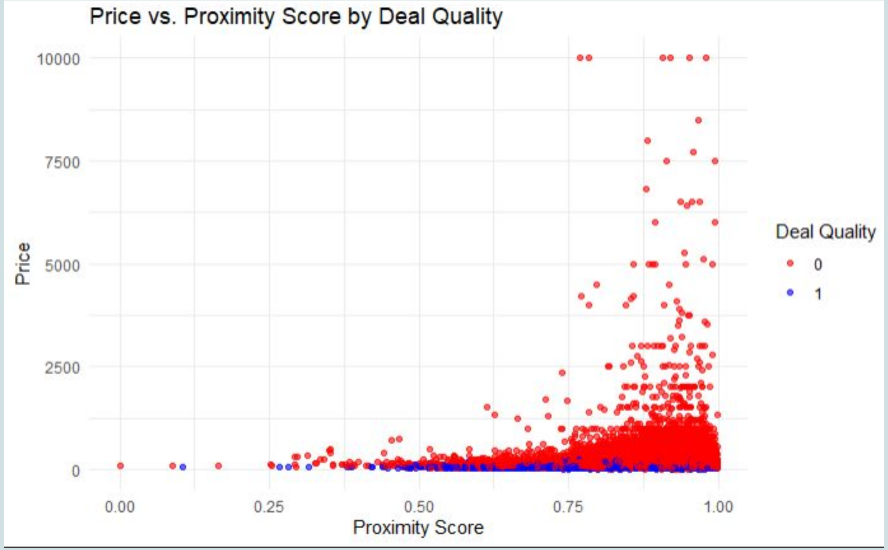
- Introduction to the project, focusing on the integration of Airbnb listing data with Points of Interest (POI) data.
- Overview of data processing steps: converting geographical data, calculating distances, and preparing data for analysis.
- Mention of the creation of new categorical decision columns based on proximity, attractiveness, demand-supply dynamics, and seasonal variations.
- Brief outline of the plan to develop a sophisticated model that encapsulates these diverse aspects to assess Airbnb listings.

The Four Pillars of Our Analysis

- Explanation of the four new columns we created: Proximity Score, Attractiveness Index, Demand-Supply Ratio, and Seasonal Adjustment.
- Proximity Score: Shows how close an Airbnb is to points of interest. Closer is usually better.
- Attractiveness Index: Combines how guests rate the Airbnb and how many reviews it has.
- Demand-Supply Ratio: A measure that reflects the balance between how many Airbnbs are available and how many people want to stay in them, specific to each neighborhood.
- Seasonal Adjustment: Accounts for changes in Airbnb's popularity across different times of the year.

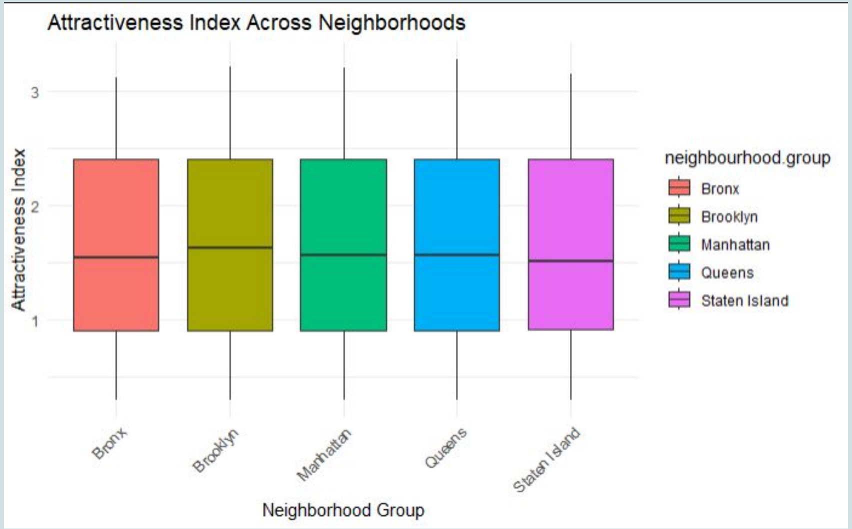
Price vs. Proximity Score by Deal Quality

- This plot shows the relationship between the price of Airbnb listings and their proximity scores, differentiated by deal quality. Red dots represent listings considered not a good deal (Deal Quality = 0), while blue dots are good deals (Deal Quality = 1). As the proximity score increases, indicating closer distance to points of interest, the price tends to increase, especially for those listings marked as good deals.



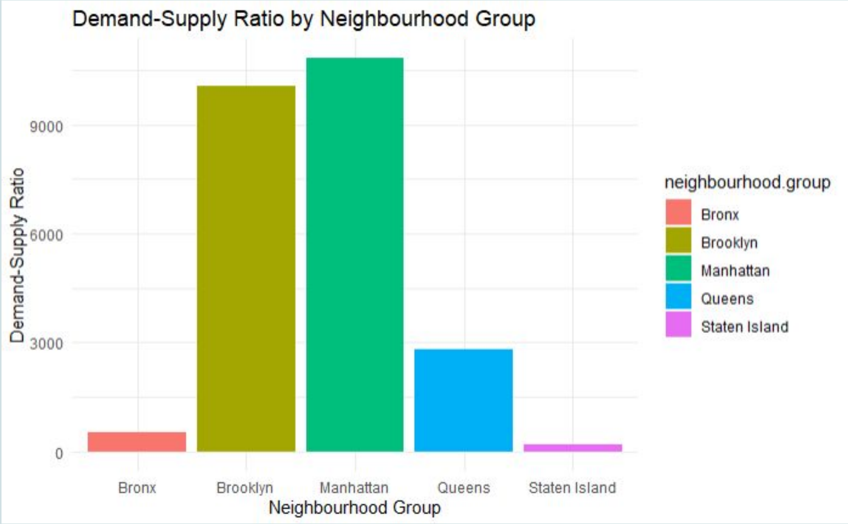
Attractiveness Index Across Neighborhoods

- This boxplot compares the attractiveness index of Airbnb listings across different New York neighborhoods: Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each box represents the range of attractiveness scores within that neighborhood, with the line in the middle showing the median score. This visual suggests that the attractiveness is relatively even across neighborhoods, with slight variations.



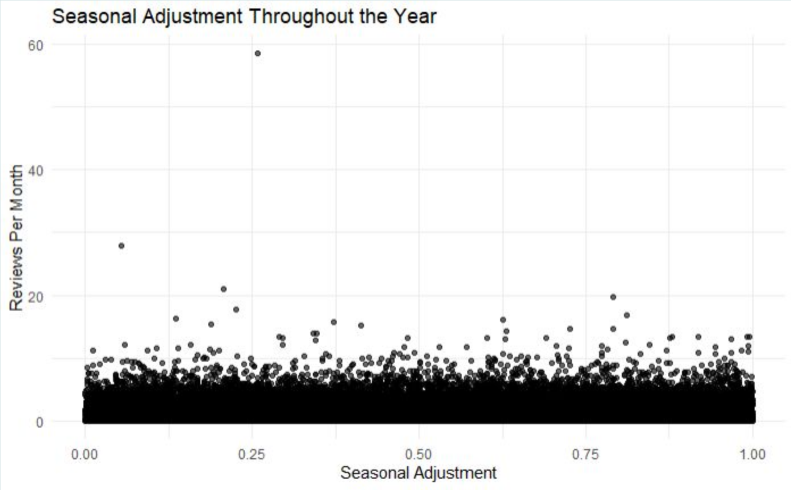
Demand-Supply Ratio by Neighborhood Group

- The bar chart illustrates the demand-supply ratio for Airbnb listings in various New York neighborhoods. The height of the bars indicates the ratio, with Manhattan and Brooklyn showing much higher demand relative to supply compared to the Bronx, Queens, and Staten Island. This could imply that Manhattan and Brooklyn are more popular among Airbnb users.



Seasonal Adjustment Throughout the Year

- Display of guest reviews per month against seasonal adjustment values.
- Dense clustering of data points towards the lower end of the seasonal adjustment scale.
- Some outliers indicating months with exceptionally high review counts.
- Insights from this visualization to inform more accurate seasonal trend modeling.



Logistic Regression Model Summary

````
```{r}
# Convert 'Deal.Quality' to a binary variable
Airbnb$Deal_Quality_Binary <- ifelse(Airbnb$Deal.Quality == "Good", 1, 0)

# Then, you can use logistic regression
binary_model <- glm(Deal_Quality_Binary ~ number_of_bedrooms * proximity_score +
                    attractiveness_index * demand_supply_ratio +
                    seasonal_adjustment + is_manhattan,
                    family = binomial(link = "logit"),
                    data = Updated_Airbnb)

summary(binary_model)
```
````

```
Call:
glm(formula = Deal_Quality_Binary ~ number_of_bedrooms * proximity_score +
    attractiveness_index * demand_supply_ratio + seasonal_adjustment +
    is_manhattan, family = binomial(link = "logit"), data = Updated_Airbnb)

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                               0.52854    0.38210   1.383    0.167
number_of_bedrooms                       -1.73182    0.26100  -6.635 3.24e-11 ***
proximity_score                          -0.31373    0.43491  -0.721    0.471
attractiveness_index                      0.80262    0.03146  25.515  < 2e-16 ***
demand_supply_ratio                      -0.05983    0.10522  -0.569    0.570
seasonal_adjustment                      -0.00564    0.04516  -0.125    0.901
is_manhattan                             -3.72925    0.05209 -71.592  < 2e-16 ***
number_of_bedrooms:proximity_score        0.40866    0.29996   1.362    0.173
attractiveness_index:demand_supply_ratio  0.03691    0.05386   0.685    0.493
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 54275  on 48894  degrees of freedom
Residual deviance: 35837  on 48886  degrees of freedom
AIC: 35855

Number of Fisher Scoring iterations: 6
```

- Our logistic regression analysis reveals the significant factors that contribute to an Airbnb listing being classified as a good deal. The number of bedrooms, the attractiveness index, and whether the listing is in Manhattan have emerged as particularly strong predictors. Although we've included interaction effects between various features, they appear to have a less substantial impact. The next steps involve validating the predictive power of our model and making refinements where necessary.

Conclusion

- Recap of the project's aim to integrate Airbnb listings with points of interest and various influencing factors.
- Confirmation of key predictors for a listing's deal quality, with emphasis on location, attractiveness, and room features.
- Acknowledgment of the insights gained from the visual analyses, emphasizing the importance of proximity and neighborhood attractiveness.
- Reflection on the successful application of logistic regression to predict deal quality.
- Emphasis on the practical implications for Airbnb hosts and the platform in optimizing listings and pricing strategies.


THANK YOU