AMSTERDAM AIRBNB PRICE ANALYSIS

IBM CAPSTONE FINAL PROJECT

Xiaotong SHI

Contents

1

INTRODUCTION

1.1. PROBLEM DESCRIPTION

Amsterdam is the capital and most popular city of the Netherlands. It is colloquially referred to as the "Venice of the North", attributed by the large number of canals which form a UNESCO World Heritage Site [1]. Amsterdam is the heaven of art because of its high-density distribution of museums and art galleries. It has large amount of collections of Vincent Willem van Gogh and Rembrandt Harmenszoon van Rijn. Amsterdam is also famous of its open culture to sex and cannabis. According to [2], there were 19 million tourists visiting Amsterdam in 2019 which brings the high demands of hotels.

Airbnb is an online marketplace for arranging or offering lodging, primarily homestays, or tourism experiences [3]. More and more people choose to stay in a local house when they are travelling. The prices of the houses can vary a lot depending on the location, the size, the service or the surroundings of the houses.

In this project, I would like to use the data science techniques to solve the following question:
Which factors of the property could affect the rental price on Airbnb?

1.2. TARGET AUDIENCE

The answer of this question can be useful for:

- people from Amsterdam who wants to start renting their properties on Airbnb
- tourists who are going to visit Amsterdam and want to estimate the cost of accommodation

2
DATA

The data we will use are:

- The data of Airbnb properties in Amsterdam
- The venue information provided by Foursquare API
- The coordinate of Amsterdam center provided by Google Map

2.1. AMSTERDAM AIRBNB DATA

The Airbnb data of Amsterdam was retrieved from Kaggle [4]. The dataset includes the following files:

- calendar.csv: The calendar has 365 records for each listing. It specifies the whether the listing is available on a particular day (365 days ahead), and the price on that day.
- listings.csv: A listing is basically an advertisement. This file holds the most useful variables that can be used visualizations.
- **listings_details.csv**: This file holds the same variables as the listing file plus 80 additional variables.
- neighbourhood.csv: Simple file with the Dutch names of the neighbourhoods.
- reviews.csv: This is a simple file that can be used to count the number of reviews by listing (for a specific period).
- reviews_details.csv: This file holds the full details of all reviews, and can also be used for instance for text mining.
- **neighbourhoods.geojson**: This is the shape file that can be used in conjunction with interactive maps (such as Leaflet for R of the Python folium package).

In our case, we will use listings_details.csv and neighbourhoods.geojson for our analysis.

listings_details.csv describes properties on Airbnb using 96 features as shown in the Figure 2.1 including the information of host, the location and the description of the property, the reviews from the previous guests. Some of these features can directly affect the price property and some of them may not. By the feature engineering and regression modeling we are going to do later, we can figure out which features are more influential than the other by ranking the importance of these features.

```
ab_data.columns
```

```
Index(['id', 'listing_url', 'scrape_id', 'last_scraped', 'name', 'summary',
       'space', 'description', 'experiences_offered', 'neighborhood_overview',
       'notes', 'transit', 'access', 'interaction', 'house_rules',
       'thumbnail_url', 'medium_url', 'picture_url', 'xl_picture_url',
       'host_id', 'host_url', 'host_name', 'host_since', 'host_location',
       'host_about', 'host_response_time', 'host_response_rate',
       'host_acceptance_rate', 'host_is_superhost', 'host_thumbnail_url',
       'host_picture_url', 'host_neighbourhood', 'host_listings_count',
       'host_total_listings_count', 'host_verifications',
       'host_has_profile_pic', 'host_identity_verified', 'street',
       'neighbourhood', 'neighbourhood_cleansed',
       'neighbourhood_group_cleansed', 'city', 'state', 'zipcode', 'market',
       'smart_location', 'country_code', 'country', 'latitude', 'longitude',
       'is_location_exact', 'property_type', 'room_type', 'accommodates',
       'bathrooms', 'bedrooms', 'beds', 'bed_type', 'amenities', 'square_feet',
       'price', 'weekly_price', 'monthly_price', 'security_deposit',
       'cleaning_fee', 'guests_included', 'extra_people', 'minimum_nights',
       'maximum_nights', 'calendar_updated', 'has_availability',
       'availability_30', 'availability_60', 'availability_90',
       'availability_365', 'calendar_last_scraped', 'number_of_reviews',
       'first_review', 'last_review', 'review_scores_rating',
       'review_scores_accuracy', 'review_scores_cleanliness',
       'review_scores_checkin', 'review_scores_communication',
       'review_scores_location', 'review_scores_value', 'requires_license',
       'license', 'jurisdiction_names', 'instant_bookable',
       'is_business_travel_ready', 'cancellation_policy',
       'require_guest_profile_picture', 'require_guest_phone_verification',
       'calculated_host_listings_count', 'reviews_per_month'],
      dtype='object')
```

Figure 2.1: The features in Airbnb data

neighbourhoods.geojson contains the geometry information of neighborhood in Amsterdam. Using the folium package [5], we can visualize the price variations of neighborhoods on the map of Amsterdam. Map visualization can give us a direct feedback on how the location of property affect the price. On the other hand, this file also gives us the coordinates of neighborhoods centers which can be used to analyze the venues of neighborhoods by Foursquare[6] later.

2.2. FOURSQUARE API

We will use the FourSquare API [6] to explore the venues of each neighborhood in Amsterdam. Given a pair of coordinates, the Foursquare explore function can be used to retrieve the venues nearby. The categories and the number of venues can describe how convenient living in a neighborhood.

2.3. THE COORDINATE OF AMSTERDAM CENTER

The coordinate of Amsterdam center is (52.372952, 4.906080) provided by Google Map [7]. The coordinate will be used to initialize the center of the map when we visualize the neighborhood of Amsterdam using folium.

3

Methodology

In this section, we will explore the data we use. Data cleaning and feature engineering methods will be used to prepare the data that is ready to fit into the model.

3.1. Airbnb Data Cleaning

There are 96 columns in the Airbnb data. But there are columns not valuable for our analysis such as listing_url, scrape_id. So, we can firstly filter some columns that are not describing the property. After careful selection, 35 features are left as shown in the Figure 3.1. Then, we need to clean the data.

```
Index(['id', 'name', 'neighbourhood_cleansed', 'latitude', 'longitude',
       'property_type', 'room_type', 'accommodates', 'bathrooms', 'bedrooms',
       'beds', 'bed_type', 'square_feet', 'price', 'cleaning_fee',
       'guests_included', 'extra_people', 'minimum_nights', 'maximum_nights',
       'has_availability', 'availability_30', 'availability_60',
       'availability_90', 'availability_365', 'number_of_reviews',
       'review_scores_rating', 'review_scores_accuracy',
       'review_scores_cleanliness', 'review_scores_checkin',
       'review_scores_communication', 'review_scores_location',
       'review_scores_value', 'instant_bookable', 'is_business_travel_ready',
       'cancellation_policy'],
      dtype='object')
```

Figure 3.1: The features in Airbnb data

3.1.1. Drop the unavailable samples

We would like to focus on the properties that are available regularly. So, we can remove the properties that were not available in the past recent months.

3.1.2. Drop the Samples Missing Information

We would like to focus on the valid property only which means the property has a valid price and reviews from previous guests to verify the authenticity.

3.1.3. DEAL WITH MISSING VALUES

The real life data can be messy. There are always many missing data as shown in the Figure 3.2. We need to fill the missing data with meaningful values.

	Total	Percent
square_feet	7203	97.154033
cleaning_fee	965	13.015916
bathrooms	6	0.080928
beds	3	0.040464
bedrooms	3	0.040464
review_scores_value	2	0.026976
review_scores_location	2	0.026976
review_scores_communication	2	0.026976
review_scores_checkin	2	0.026976
room_type	0	0.000000
accommodates	0	0.000000
latitude	0	0.000000
property_type	0	0.000000
longitude	0	0.000000
neighbourhood_cleansed	0	0.000000
name	0	0.000000
bed_type	0	0.000000
cancellation_policy	0	0.000000
price	0	0.000000
is_business_travel_ready	0	0.000000

Figure 3.2: The missing values

- **Drop the feature having too many missing values**
  More than 97% of properties do not have the square feet recorded. So, we remove the square_feet from the data.
- **Fill the missing sub-reviews by corresponding review_scores_rating**
  The review_score_rating represents the average of reviews for all categories such as the location. If a review score from a kind of category is missing, we fill it by its corresponding review_score_rating.
- **Fill the missing bathrooms, bedrooms, beds by 0**
  If the values of bathrooms, bedrooms, beds are missing, we can understand there is no bathroom, bedroom, bed in this property. So, we fill them by 0.
- **Fill the missing cleaning fee by mode**
  We fill the missing cleaning fee by the mode which is 0$ in our case.

3.1.4. CORRECT HIGHLY SKEWED NUMERICAL FEATURES

The skewed distribution has the following disadvantages [12]:

- Highly skewed distributions are difficult to examine because most of the observations are confined to a small part of the range of the data.

- Outlying values in the direction of the skew are brought in toward the main body of the data when the distribution is made more symmetric.

For the features with high skewness, I use the box-cox transformation to correct the non-normal distribution while maintaining the information.

3.1.5. ENCODE CATEGORICAL FEATURES

We use one-hot encoder to encode the categorical features.

3.1.6. EXAMINE THE DISTRIBUTION OF PRICE

The price of the property is the value we are going to predict. However, the distribution of price does not follow the normal distribution as shown in the left two figures of Figure 3.3. Also, there are 5 outliers whose prices are over 2000. Considering the disadvantages of abnormality as described in Section 3.1.4, we remove the outliers and log-transform the price. The corrected distribution is shown in the right two figures of Figure 3.3.

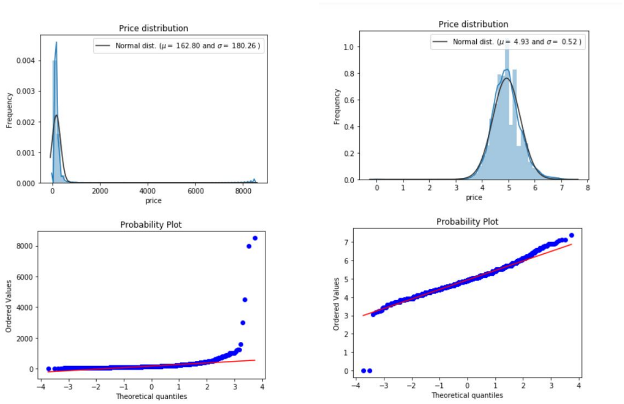

Figure 3.3: The distribution of the price

3.2. NEIGHBOURHOODS EXPLORATION

To have a initial sight how the neighbourhood affect the Airbnb properties. We use *neighbourhoods.geojson* and Python package folium [5] visualize the distribution of Airbnb properties in Amsterdam.

Figure 3.4 shows that the neighbourhood closer to the center, the more properties it has. The top 3 neighbourhoods are Oud-West, Centrum-West, De Pijp - Rivierenbuurt, Centrum-Oost, Zuid. And Figure 3.5 shows that the neighbourhood closer to the center, the higher median of the prices it has. Top 2 neighbourhoods are Centrum-Oost, Centrum-West. It not surprising that more people prefer to live close to city center and these neighbourhoods are also the places where many tourist attractions locate such as the Van Gogh Museum, Anne Frank Museum, Red Light District.

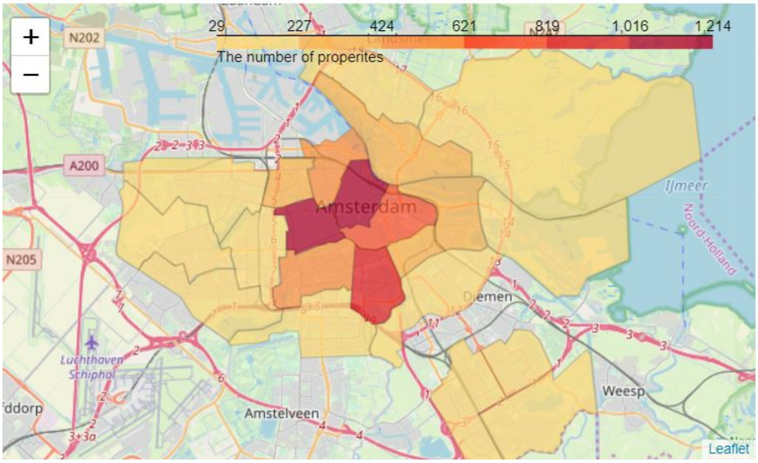

Figure 3.4: The count of properties in different neighbourhoods

3.3. VENUES EXPLORATION

Foursquare API can explore the venues near any location. Given a pair of coordinate, Foursquare explore function can be used to return the venues nearby with the restriction of radius 1km and limit 100. Each venue provided by Foursquare belongs to a category and a sub_category, and each category has an unique ID [8].

Venues in Foursquare belong to 10 main categories:

- 'Arts & Entertainment': '4d4b7104d754a06370d8125'
- 'College & University': '4d4b7105d754a06372d81259'
- 'Event': '4d4b7105d754a06373d81259'
- 'Food': '4d4b7105d754a06374d81259'
- 'Nightlife Spot': '4d4b7105d754a06376d81259'
- 'Outdoors & Recreation': '4d4b7105d754a06377d81259'
- 'Professional & Other Places': '4d4b7105d754a06375d81259'
- 'Residence': '4e67e38e036454776db1fb3a'
- 'Shop & Service': '4d4b7105d754a06378d81259'
- 'Travel & Transport': '4d4b7105d754a06379d81259'

The sub_category gives more detailed description. For example, the type of restaurant, such as Chinese restaurant.

In our case, the 10 main categories would be enough for us to understand the type of venues nearby. The more detailed categories are not very helpful for us to define a

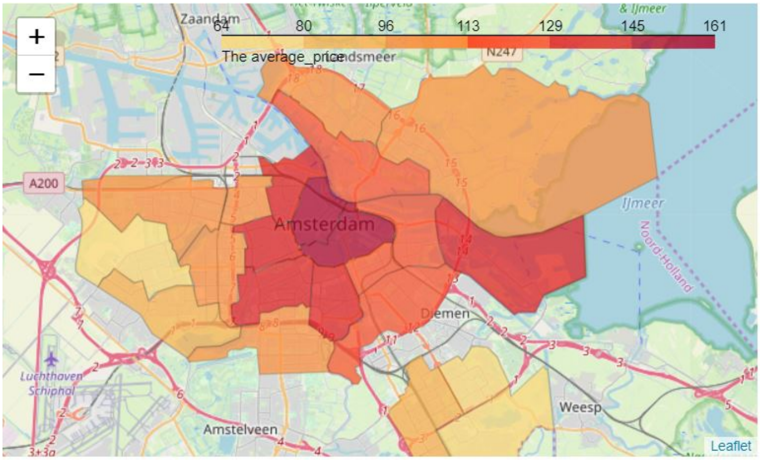

Figure 3.5: The median prices of different neighbourhoods

location. For example, we can say a place is suitable for people who has requirement for food if there are more than 20 restaurants nearby. It is not necessary to know whether there it's a Chinese restaurant or Italy restaurant.

There are three categories in the 10 main categories usually won't affect the choices of people when they are in travel: College & University, Professional & Other Places, Residence. So we won't consider the venues belonging to these there categories.

Unfortunately, because there are too many properties in the data and Foursquare has its limitation for the number of daily requests, it is impossible to explore the venues of every property. Alternatively, I assume the venue distribution of properties in the same neighbourhood can be similar, so I use the data of each neighbourhood center to represent the properties belonging to each neighbourhood.

Overall, for each neighbourhood, we will use Foursquare to explore the top 100 venues within 1km to the neighbourhood center. And the venues of each property is represented by the venues of its corresponding neighbourhood. The results is shown in the Figure 3.6. We divide each count by the total number of venues of each neighbourhood as shown in the Figure 3.7. This data can be used to describe the main type of venues of neighbourhoods.

	neighbourhood	Arts & Entertainment	Event	Food	Nightlife Spot	Outdoors & Recreation	Shop & Service	Travel & Transport
0	Bijlmer-Oost	9	0	9	5	8	16	7
1	Noord-Oost	0	0	1	2	2	1	0
2	Noord-West	8	1	4	1	7	14	11
3	Oud-Noord	3	1	38	6	10	27	12
4	IJburg - Zeeburgereiland	4	0	22	3	15	21	7

Figure 3.6: The number of venues in each neighbourhood

	neighbourhood	Arts & Entertainment	Event	Food	Nightlife Spot	Outdoors & Recreation	Shop & Service	Travel & Transport
0	Bijlmer-Oost	0.166667	0.000000	0.166667	0.092593	0.148148	0.296296	0.129630
1	Noord-Oost	0.000000	0.000000	0.166667	0.333333	0.333333	0.166667	0.000000
2	Noord-West	0.173913	0.021739	0.086957	0.021739	0.152174	0.304348	0.239130
3	Oud-Noord	0.030928	0.010309	0.391753	0.061856	0.103093	0.278351	0.123711
4	IJburg - Zeeburgereiland	0.055556	0.000000	0.305556	0.041667	0.208333	0.291667	0.097222

Figure 3.7: The percentages of different venues in each neighbourhood

3.4. MERGE FEATURES

We merge the venues features with the previous feature data we cleaned and remove the identification data such as id, name and coordinate.

Next, we would like to drop the features that have more than 95% same values for all the data. Meanwhile, we would also like to remove the feature value that rarely occur such as property_type_Boat, bed_type_Airbed.

Finally, we have our final data. It includes 41 numerical and categorical features as shown in Figure 3.8. It describes each property of Airbnb in Amsterdam from various aspects including the facilities, the location, the reviews, the extra fee and the service. In the next section, we are going train regression models to predict the prices using these features.

```
Index(['accommodates', 'bathrooms', 'bedrooms', 'beds', 'cleaning_fee',
       'guests_included', 'extra_people', 'minimum_nights', 'maximum_nights',
       'number_of_reviews', 'review_scores_rating', 'review_scores_accuracy',
       'review_scores_cleanliness', 'review_scores_checkin',
       'review_scores_communication', 'review_scores_location',
       'review_scores_value', 'Arts & Entertainment', 'Event', 'Food',
       'Nightlife Spot', 'Outdoors & Recreation', 'Shop & Service',
       'Travel & Transport', 'property_type_Apartment', 'property_type_House',
       'room_type_Entire home/apt', 'room_type_Private room',
       'instant_bookable_f', 'instant_bookable_t',
       'cancellation_policy_flexible', 'cancellation_policy_moderate',
       'cancellation_policy_strict_14_with_grace_period',
       'neighbourhood_Bos en Lommer', 'neighbourhood_Centrum-Oost',
       'neighbourhood_Centrum-West', 'neighbourhood_De Baarsjes - Oud-West',
       'neighbourhood_De Pijp - Rivierenbuurt', 'neighbourhood_Oud-Oost',
       'neighbourhood_Westerpark', 'neighbourhood_Zuid'],
      dtype='object')
```

Figure 3.8: Final Features

3.5. REGRESSION MODELLING

Given the information of the property, we would like to predict its price. In this section, we are going to fit our data into multiple common regression models and compare their performance on price prediction.

To evaluate the performance of models, we firstly separate the data into training set and test set. The training set will be used to select and train the model, and the test set will be used to test the performance of the model.

3.5.1. EVALUATION MATRIX

We use the root of mean-squared-error (rmse) [9] [10] to measure how close the prediction to the actual price. The lower rmse, the model is better.

$$RMSE = \sqrt{MSE(\hat{y})} = \sqrt{E((\hat{y}-y)^2)}$$

3.6. REGRESSION MODELS

We will fit out data into the following regression models:

- Lasso
- Ridge
- ElasticNet
- GradientBoostingRegressor
- XGBoosting Regressor
- LightGBM Regressor

We use the model from sklearn [11].

3.6.1. MODEL COMPARISON

We use the cross validation to compare the performance of models. The results are shown in Table 3.1. We can see boosting models (Gradient boosting, xgboosting, lightgm) show much better results than the other models. Ridge also works great and Ridge runs much fast than boosting models. So we will not consider lasso and Elastinet models anymore.

Model	Score Mean	Score Std
Ridge	0.3439	0.0309
Lasso	0.5075	0.0203
ElasticNet	0.4899	0.0200
Lightgbm	0.3279	0.0286
GradientBoosting	0.3474	0.0274
XGBoost	0.3298	0.0284

Table 3.1: The cross validation rmse scores

3.6.2. MODEL BLENDING

To get a better and robust model, we linearly blend the models. We give higher weights to the models having better performance.

4

Results

4.1. Regression Model Results

Let's fit the whole training data to our models. The rmse scores on the training set of these model are shown in the Table 4.1.

Model	Score
Ridge	0.3430
Lightgbm	0.2985
GradientBoosting	0.1969
XGBoost	0.2892

Table 4.1: The rmse scores on training set

We will give higher weights to boosting models due their better performance. The final regression model is blended in the following way:

$$FinalPrediction = 0.1 * Ridge + 0.3 * GradientBoosting + 0.3 * XGBoosting + 0.3 * LGBoosting$$

The rmse score of the blended model on training set and test set are 0.2603 and 0.3256. That shows our model can accurately predict the price given the information data of the property.

4.2. Feature Importance

Once we have the trained model, more importantly, we would like to know which factor effect the price most. This can be done by examining the coef_ attribute of linear regression model or the feature_importances_ attribute of boosting regression model which represent the importance scores of all features. We rank the features by their features' importance and keep the top 10 features only.

According to the above table, there is no surprise that the number of accommodates, the number of bedrooms and the room type are top 3 most important features to the

	1st Important Feature	2nd Important Feature	3rd Important Feature	4th Important Feature	5th Important Feature	6th Important Feature	7th Imp F
ridge	accommodates	room_type_Entire home/apt	review_scores_location	neighbourhood_Centrum-West	neighbourhood_Centrum-Oost	bathrooms	
GradientBoosting	accommodates	room_type_Entire home/apt	bedrooms	cleaning_fee	Nightlife Spot	extra_people	number_of_r
XGBoosting	room_type_Entire home/apt	accommodates	bedrooms	Nightlife Spot	review_scores_location	neighbourhood_Centrum-West	neighbourhoo
Lightgbm	number_of_reviews	extra_people	cleaning_fee	review_scores_rating	minimum_nights	accommodates	guests_ir

Figure 4.1: The top 10 importance features

price. Bigger house can accommodate more people therefore can be more expensive. Other related factors are the number of beds and the number of bathrooms.

On the other hand, reviews also affect the price. People tend to book the property that has more reviews and higher reviews. The number of reviews can describe the opening years and the popularity of the property and the review scores can reflect the condition and reliability of the property. The two most important reviews are for check-in and location.

Location is definitely another important factor to the price. The review scores for location can reflect the neighbourhood the property locates and the how convenience the public transportation nearby. Property locating in the centrum neighbourhood can be more expensive than the others. This can by verified by the previous visualization 3.4 of the medians of the prices of different neighbourhoods. We can see that it is true that the properties locating in Centrum-Oost and Centrum-West are more expensive than the others. It is worthy to mention these two neighbourhoods are the places where many famous places locate such as Anne Frank house, Red light district, Amsterdam museum. Also, it is interesting to see that IJburg - Zeeburgereiland also shows the high price even though it is not very close to the center. A reason could be the nice seaside scenery when living on the island.

On the other hand, model Lightgbm shows a kind of different result than the other three models. It says cleaning_fee and fee for extra people can also effect the price. If the host doesn't charge for the fee for cleaning, that might means the fee has already included in the price. On the other hand, if the extra guest is allowed, the average cost per person might won't be very high even the fee for extra people is charged.

Next, let's look at the venues features only. We can find something that is interesting

	1st Important Feature	2nd Important Feature	3rd Important Feature	4th Important Feature	5th Important Feature	6th Important Feature	7th Important Feature
ridge	Food	Travel & Transport	Shop & Service	Event	Nightlife Spot	Outdoors & Recreation	Arts & Entertainment
GradientBoosting	Nightlife Spot	Food	Shop & Service	Outdoors & Recreation	Arts & Entertainment	Event	Travel & Transport
XGBoosting	Nightlife Spot	Food	Shop & Service	Event	Outdoors & Recreation	Arts & Entertainment	Travel & Transport
Lightgbm	Nightlife Spot	Outdoors & Recreation	Arts & Entertainment	Shop & Service	Food	Travel & Transport	Event

Figure 4.2: The rank of venue categories

that whether there is a nightlife spot nearby can affect the price a most. A proper interpretation could be that the nightlife spots usually locate the centrum neighbourhoods so that the location is the actual influential factor.

It is not surprising that Food and Shop & Service are two important venues. Tourists might don't want to spend a lot of time on finding restaurants or shopping places. What really surprises me is that Travel & Transport is ranked as the least important feature by three models. If we compare the area of Amsterdam and other famous tourism cities in the world, for instance, new York, the area of New York is three times larger that Amsterdam. Amsterdam has intense public transportation network including bus, metro, tram, ship. And bicycling is a fashion way in the Netherlands to go out. So, people can always find a convenient way to go out no matter where they live.

5

Discussion

According to the our analysis, the price of the proper on Airbnb in Amsterdam mostly varies based on the number of accommodates/the number of bedrooms, the location and the extra fee (cleaning fee/fee for extra people). Bigger house and house close to city center or public transportation can be more expensive than the other. And the price can be higher if the extra fee is included in the price. Also, properties that locate near nightlife spots can be more expensive.

Here are my recommendations:

If you want to rent your house in Amsterdam on Airbnb:

- If you have a house in the popular neighbourhoods, then congratulations, you can rent your house in a good price.
- Ask for reviews from your guests. People trust on the previous review than your description and pictures. Showing on time when checking-in, being honest and providing good service (such as allowing extra guest) can really helpful on gaining higher review scores.

If you are going to visit Amsterdam and want to live in a local place:

- The more bedrooms and bathrooms in the house, the price can be more expensive. If you can find more travelling partners who don't mind sharing bedrooms or bathrooms with you, the average price per person can be very fair.
- It is always expensive to live in the center. If you really care about living close to center, I recommend to live in the neighbourhood of Oud-West which has the most number of houses but the median price ranks the 5th expensive neighbourhood. If you have an empty wallet, considering the mature public transportation network of Amsterdam, I recommend to live in the neighbourhood far away from the center. Even if you live the suburb, a less than 30-minutes train can take you to the center.

- Places having many restaurants and shopping places can be expensive. You can choose other places if you don't necessarily eat or shop nearby the place you live.
- If you are a night owl and a lover to night clubs of Amsterdam, unfortunately, you might need to prepare more budgets on accommodation.

There are might be other factors can effect the price such as the decoration of the property and the service the host can provide. Unfortunately, we can't go through the influence of all of possible factors due to the unavailability of the data. However, I believe my analysis is extremely valuable for the hosts from Amsterdam who wants to define the price of their proprieties or tourists who want to have a estimate of the accommodation cost based on their requirements and budgets. This analysis can also be applied to any other city in the world.

6

Conclusion

In this project, I analyzed the Airbnb houses in Amsterdam using the Airbnb data and the venue information of neighbourhoods provided by Foursquare API. I used several classic regression models to predict the price of the properties and analyzed the most important factors to the price of a property. Based on the above analysis, several recommendations were proposed to people who want to rent their house on Airbnb and people who to live in a local place in Amsterdam. This analysis can be adapted to any other city's Airbnb house price analysis.

This is the final project of my IBM professional certificate course on Coursera. I would like to thanks to myself for consisting motivating myself to explore better solution, and I also would like to thanks to the time of the person who is reviewing this work. Hope everyone can by happy and safe.

References

[1] https://en.wikipedia.org/wiki/Amsterdam.

[2] https://www.dw.com/en/how-amsterdam-is-fighting-mass-tourism/a-47806959.

[3] https://en.wikipedia.org/wiki/Airbnb.

[4] https://www.kaggle.com/erikbruin/airbnb-amsterdam.

[5] https://python-visualization.github.io/folium/.

[6] www.foursquare.com.

[7] https://www.google.com/maps.

[8] https://developer.foursquare.com/docs/build-with-foursquare/categories.

[9] https://en.wikipedia.org/wiki/Root-mean-square_deviationl.

[10] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html.

[11] https://scikit-learn.org/stable/supervised_learning.html#supervised-learning.

[12] John Fox. *Applied regression analysis and generalized linear models.* Sage Publications, 2015.