Amsterdam Airbnb Price Analysis

IBM Applied Data Science Capstone Final Project

Xiaotong Shi

Category

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. 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 vary a lot depending on the location, the size, the service or the surroundings of the houses.

So, 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 The data of Airbnb properties in Amsterdam

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 neighbouhoods
- 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** only for our analysis.

listings_details.csv describes properties on Airbnb using 95 features as shown in the **错误! 未找到引用源。** 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 feature engineering and regression modeling we are going to do later, we can which features are more influential than the other by ranking the importance of these features.

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.

```
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 1 The features in Airbnb Amsterdam data

2.2 The venue information provided by 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 provided by Google Map

The coordinate of Amsterdam center is (52.372952, 4.906080). The coordinate will be used to initialize the center of the map when we visualize the neighborhood of Amsterdam using folium.

3. Reference

[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