AMSTERDAM AIRBNB PRICE ANALYSIS

IBM CAPSTONE FINAL PROJECT

Xiaotong Shi

Problem

- There were 19 million tourists visiting Amsterdam in 2019 which brings the high demands of hotels.
- More are more people like to live in local house when they are travelling.
- The prices of houses on Airbnb varies from 50$ to 8000$

Which factors of the property could affect the rental price on Airbnb?

Target Audience

- 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

Data

- The data of Airbnb in Amsterdam provided by Kaggle
- The venue information provided by Foursquare API
- The coordinate of Amsterdam center provided by GoogleMap

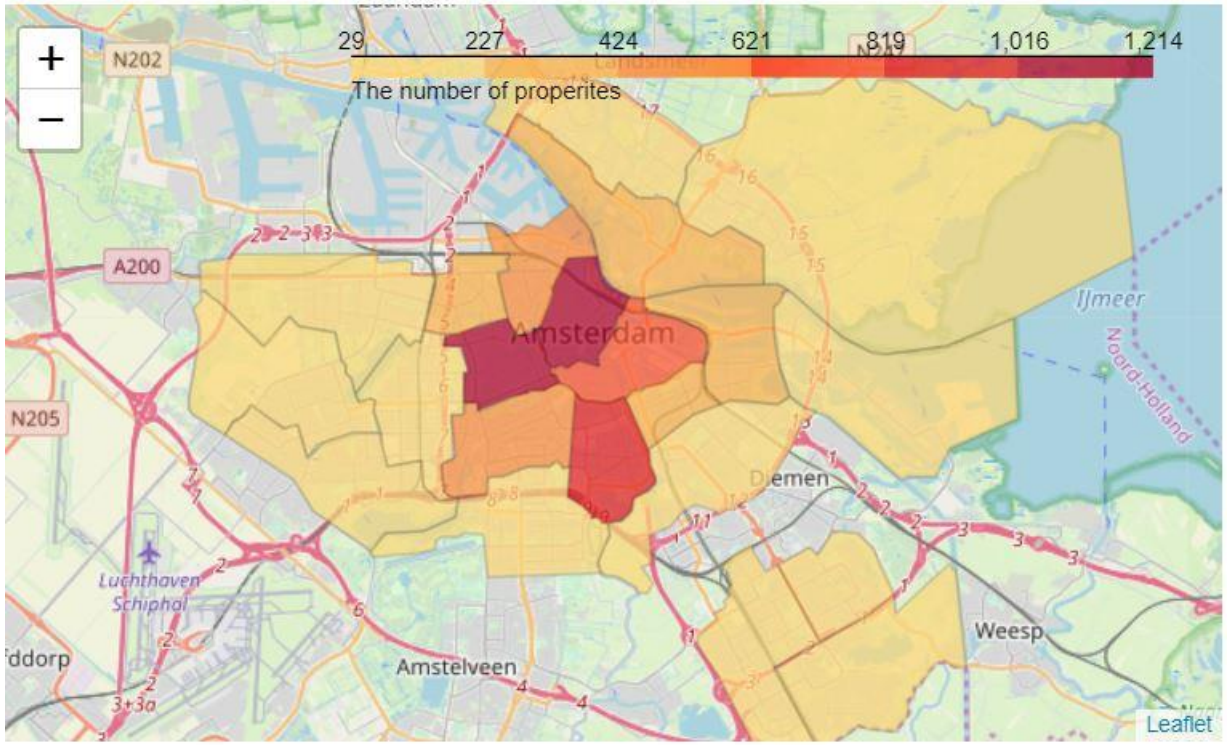
```
ab_data.columns
```

```
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')
```

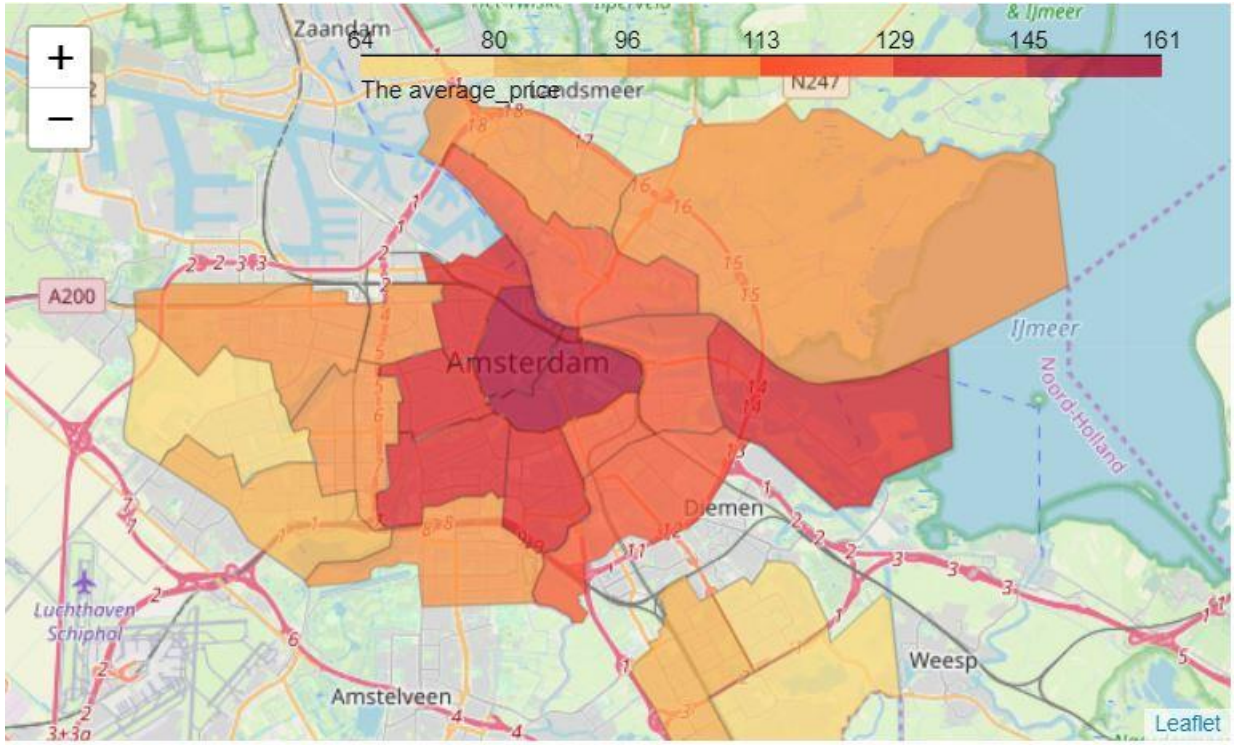
Data Cleaning

- Drop the unavailable samples
- Drop the samples missing important information
- Deal with missing values
- Correct Highly Skewed Numerical Features
- Encode Categorical Features
- Correct the abnormal distribution of the price

Neighborhood Exploration



The number of houses in each neighborhood



The median of the price of houses in each neighborhood

Venues Exploration

The main categories of venues provided by Foursquare:

- Arts & Entertainment
- Event
- Food'
- Nightlife Spot
- Outdoors & Recreation
- Shop & Service
- Travel & Transport

	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

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

The percentages of different venues in each neighborhood

Final Features

After the data cleaning and feature engineer, we merge all the features we want to keep.
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')
```

Regression Models

We fit our data into the following regression models:

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

We use the root of Mean Squared Error to evaluate these models:

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

Model blending

- We firstly exclude the models that didn't show good models: Lasso and ElasticNet.
- To get a better and robust model, we linearly blend the models. We give higher weights to the models having better performance.

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

The rmse cross-validation scores of models

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

Model Prediction

- 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

Feature Importance

	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

Important features:

- Accommodates / room_type / the number of bathrooms / the number of bedrooms
- Neighbourhood
- Number of reviews / reviews_scores_rating / reviews_scores_checkin / reviews_scores_location
- The fee for extra people / cleaning fee

Venue Importance

	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

Important venue categories:

- Nightlife spot
- Food / Shop & Service
- Outdoors & Recreation / Arts & Entertainment

Travel & Transport is surprisingly not very important.

Recommendations to the Host

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

- If you have a house in the popular neighborhoods, 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

Recommendations to the Tourists

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

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

Recommendations to the Tourists

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

- 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.

Conclusion

- Analyze the Amsterdam Airbnb data: data cleaning and feature engineering
- Analyze the venue data provided by Foursquare
- Fit data into multiple regression models and predict the price
- Rank the feature of the house by their importance
- Visualize the distribution of houses in different neighborhoods
- Give recommendations to the host and the tourist based on analysis.