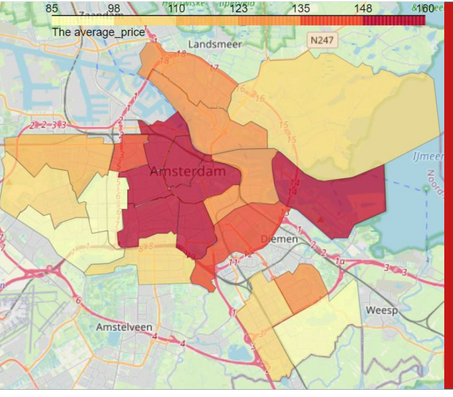


Knowledge Engineering - Group 12

Amsterdam Airbnb Price Prediction

TU/e

Client

Potential Airbnb owner looking for factors contributing to pricing

Goal

Analyze optimal factors that contribute towards pricing and build model for airbnb price prediction.

Data Sources

- Kaggle(https://www.kaggle.com/datasets/erikbruin/airbnb-amsterdam)
  Airbnb listings, reviews and neighbourhood data
- Inside Airbnb Portal (http://insideairbnb.com/get-the-data/)
- Exploded amenities from listings and create new dataset with (listing_id, amenities)
  Amenities

Data Integration & Cleaning

- Imputed missing values (numerical features → mean & categorical features → most_frequent strategy)
- Applied Log transformation for “price” column for normal distribution and for better statistical analysis
- Merged listings dataset with neighbourhood geojson to get details of airbnb prices w.r.t neighbourhood
- Created Choropleth map of amsterdam neighbourhood with avg prices by extracting lat/lon from neighbourhoods.geojson dataset
- Merged listings with reviews dataset to get listings reviews/comments
- Text preprocessing in reviews data (tokenization, stop words removal, excessive whitespace removal, lowercase)
- Calculated reviews/comments subjectivity and polarity to get better price analysis based on user’s reviews
- Dropped overfitting features (Drop the features that have more than 95% same values for all the data.)

Data Analysis & Knowledge Graphs

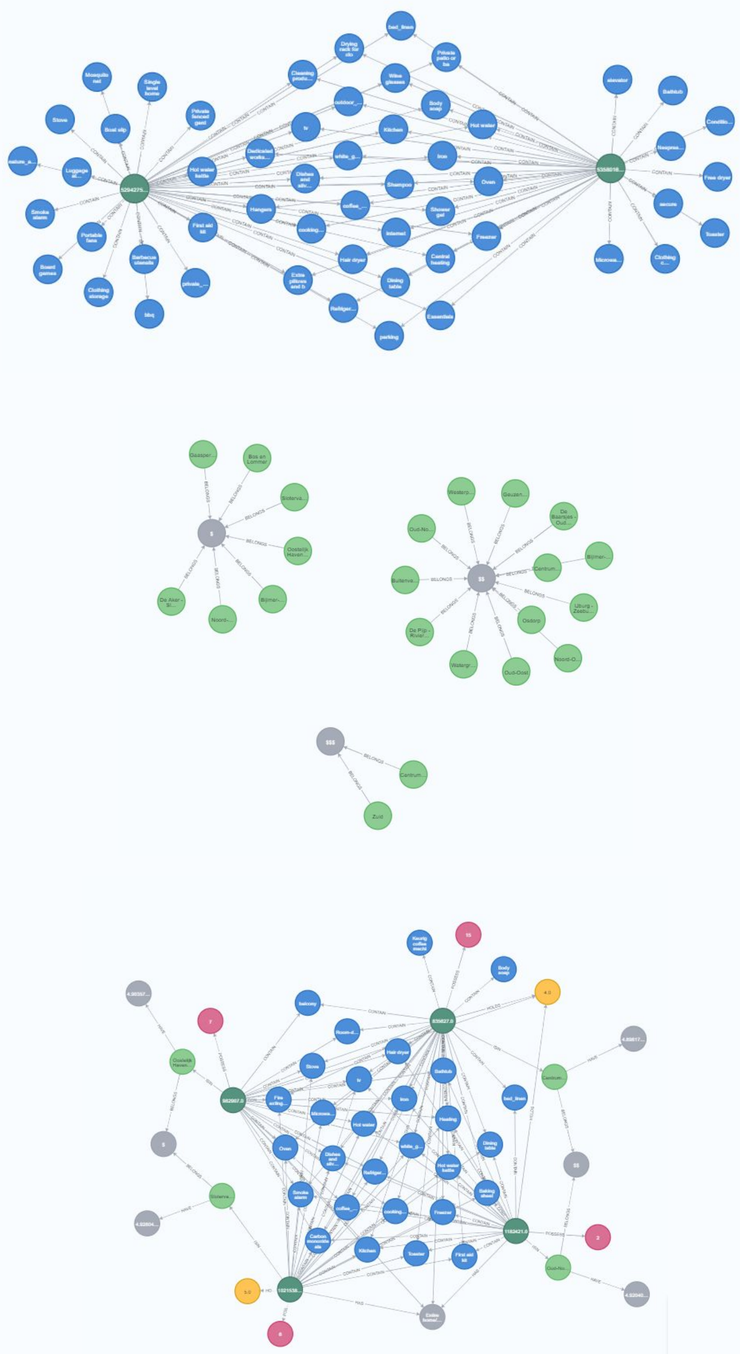
Question

➔ Find the common amenities offered by listings in Osdorp with the highest review score

➔ Find in which price category($, $$, $$$) each neighbourhood belongs

➔ Top listing per neighbourhood based on reviews, rating offered and optimal price/per neighbourhood and print 4 along with their features such as room type, amenities and count of reviews they offer for readability sake.

Knowledge Graph



Prediction analysis

A **regression based model (XGBoost)** was built to provide the client with **an appropriate price for their airbnb** based on the features they offer to match the current market in Amsterdam. An example is provided below:

Neighbourhood	Centrum-Oost
Room_type	Entire home
Accommodates	4
Bedrooms	2
Minimum_nights	2
Maximum_nights	180
Availability_365	329

Predicted Price: $ 281.78

An additional attribute, we thought the client might find convenient to know which features hold most importance in airbnb pricing.

Check our project on GitHub!