Topic 6. Airbnb & European cities. Berlin.
Group 12

### 1. Research Background & Motivation

With the increasing influx of tourism and rising visitor demands, European cities are experiencing high pressure. Airbnb positions itself as a P2P platform, aiming to enable residents to benefit from tourism. Our focus was to evaluate the actual impact of Airbnb on the city of Berlin answering the following research questions: What socio-economic impact do Airbnb apartments have on European cities? Can the position or the popularity of hotels/Airbnb apartments relate to Points of Interest, public transport, or other city features and can be predicted?

2. Data

Airbnb-related data was extracted from *insideairbnb.com*. Dataset contains all listings available in December 2023. Also, we extended our dataset by adding location information of points of interest (POI)[1], restaurants & hotels[2], and public transportation stops[3]. To measure the social impact we get social & health data[4] of all neighbourhoods of the city.

3. Exploratory & Spatial Analysis

Figure 1. Hotels vs Airbnbs locations heatmaps
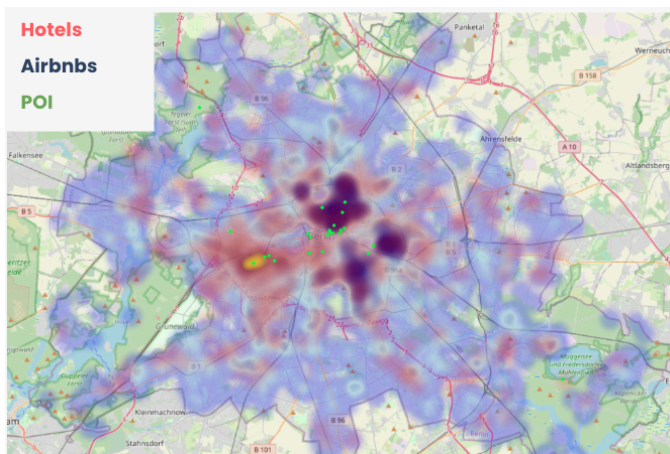
Initially, AirBnb dataset had shape (13.327; 75), after deleting listings that had no prices as well as some redundant features we ended up with the dataset of shape (9370; 26). Among the whole dataset, hotel-related properties correspond to 7.5% of listings; 25% of listings of "non-hotel" type are properties of shared type. Using heatmaps (figure 1) and spatial autocorrelation analysis we confirmed that hotels and Airbnbs are distributed differently around the city, with hotels mostly located mid-west, and Airbnbs - mid-east, adding to the overall tourism pressure in central districts. At the same moment, Airbnbs are penetrating more into residential areas compared to hotels, helping new districts to benefit from increasing tourism. Additionally, it was revealed that a high portion of Airbnb listings (50%) that were of "non-hotel" type are owned by hosts who have more than 1 posting and likely using platform for business purposes. This leads to a reduction in housing supply for locals and a decrease in taxes paid by these businesses.

4. Research Question 1: What socio-economic impact do hotels or Airbnb apartments have on cities in Europe? Berlin

As for the preprocessing, we standardized district names to align with all 138 Berlin district regions ("Bezirksregionen"), thus achieving excellent triangulation, and also imputed missing values using the median for complete data in regression analysis.

In an attempt to discover causality, we implemented Regression (OLS). We utilized a range of socio-economic indices as predictors, with Airbnb density as an additional key variable, in varied model configurations, with different socio-economic indices as response variables. We ensured model reliability by checking R^2, residuals' normality, and matrix conditioning. We conducted the Shapiro-Wilk test for residuals and checked for multicollinearity through correlation matrices. This allowed us to rely with considerable reliability on the significance levels of Airbnb density on socio-economic factors.

Conclusion: Airbnb presence does not significantly affect public health, mortality, healthcare needs, social environment, education, and social support. Airbnb shows a significant positive impact on the job market and employment conditions.

5. **Research Question 2: Is the position or the popularity of Airbnb apartments related to Points of Interest, public transport or other features of the city?**

For this question, additional data processing was needed. Data from AirBnB listings, TourPedia data and public transport stops was merged into one dataframe and aggregated by neighborhoods, extracted to GeoPandas dataframe. After these steps, statistics for each neighborhood were calculated, e.g. mean number of likes, cumulative quantity of POIs, mean number of public transport stops.

After, two models were built with target variables as month number or reviews: Random Forest Regressor & Linear Regression (with L2 regularization and without). Afterwards two feature importance metrics were applied: Recursive feature elimination for LinReg without and Feature Importance for Random Forest.

Conclusion: AirBnB popularity is closely linked to data from restaurants and points of interest, notably data provided by Foursquare. Restaurants have a more significant impact on popularity compared to points of interest, as indicated by Lasso Regression analysis. The presence of transport stops shows a weak correlation with the popularity of AirBnB apartments.

6. **Research Question 3: How well can good locations for a new Airbnb be predicted?**

The concept of a "good location" was elucidated by considering the product of the monthly review count and the price. To predict this variable we used the same dataset as for the previous question.

Conclusion: According to the results obtained, we can say that in general it is possible to predict location that will be successful on Airbnb however, there is still a room for improvement.

7. **Overall Conclusions**

- Airbnb apartments in Berlin are primarily concentrated in central districts, adding to the overall tourism pressure there.
- Airbnbs also exhibit a greater presence in peripheral districts compared to traditional hotels. This includes new districts, contributing to the tourism economy and positively influencing job market.
- High portion of Airbnb listings comprises commercial offers, leading to a reduction in housing supply for locals and a decrease in taxes paid by these businesses.
- Optimal locations for Airbnbs are associated with Points of Interest (POIs) and restaurants, while they do not demonstrate a strong connection to the public transportation network.
- According to the results of the modelling, we still see some room for improvement for existing models. For example, it will probably make sense to try data modelling not for city's neighbourhoods but for squares of 1 km x 1 km in width and length.