Around the Corner

Comparison of Airbnb listings and hotels in Berlin

Outline

Goal of the project

- Comparison of Airbnb and booking.com/Expedia data for Berlin (room for two persons for one night)
- Search and use of different data sources (API, web scraping, dataset)
- Get insights per district for Airbnbs vs. hotels regarding
    - Density of accommodations
    - Price per night
    - Price level

Data sources, problems and limitations

Sources:

- **API for Airbnb data**
- **Dataset for booking.com data** - using Octoparse
- **Web Scraping for Expedia data**

Problems/limitations:

➔ **10k results** although >40k available, data from 2017, lot of columns with similar but slightly different content

➔ **360 results**, but not all available hotels included, depends highly on the search filter, not all hotel prices were in the final output, depending on info on website assignment to area not always

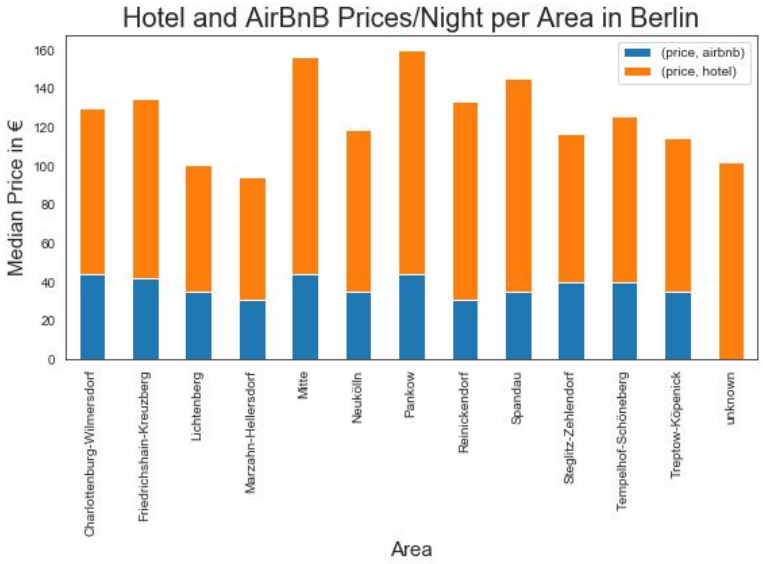
➔ **100 results**, very time consuming to build a working web scraper + result only around 100 hotels due to limitations from Expedia, around 40% had street name only instead of area

Database - data wrangling/cleaning and database structure

- All gained datasets required intensive cleaning
- Data missing in all datasets - decision which rows to drop (esp. missing prices, areas)
- Same format of datasets required before merging
- Final structure as follows
    - Three notebooks for getting + cleaning data, output per notebook is list with “source, name, price, area”
    - One final notebook for merging datasets, final cleaning and running analysis + plotting

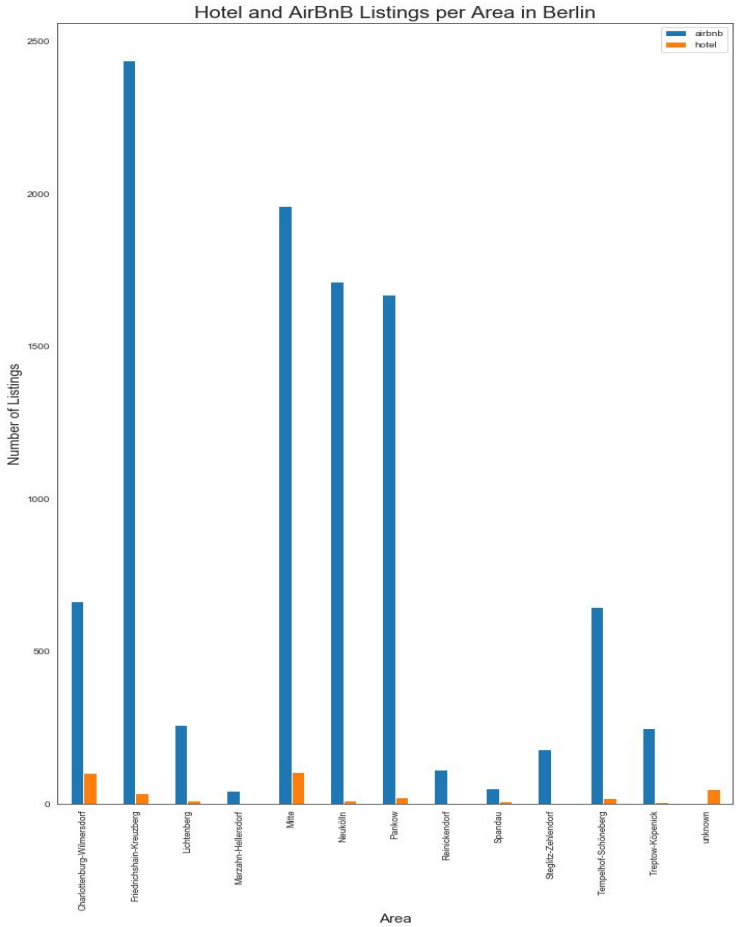
Main insights - price/night + price level

- Median price per night in Airbnb more than 50% lower than hotel price, true over all areas in Berlin
- Price level for Airbnb on average on similar level over all areas, hotel prices have higher variance
- Areas with highest price per night for a hotel room (Mitte + Pankow) have also the highest price per night in an Airbnb apartment



Main insights - density

- Fhain/X-Berg, Mitte, Neukölln, Pankow have by far the highest density of Airbnb listings
- Central areas of Berlin have most hotels (Mitte+Charlottenburg-Wilm.)
- Airbnbs are wider spread over the whole city than hotels



Open questions

- How do prices/# of available rooms (both for Airbnb and hotels) change throughout the year?
- Prices for hotel rooms depend on when the booking is made. How much does that influence the average price per night?
- How much did Airbnb prices/ listings go up from 2017 to today? If we'd have data from today, would the results differ?
- Where do you get more value for money? Hotels or Airbnbs?
    - Does the higher service level in hotels justify the higher prices?
- How can you compare the quality of an Airbnb to hotel ratings? Hotels have a standardized rating (1 - 5 stars) , the reviews on Airbnb are based on subjective opinions

Learnings

- Web scraping might be very time consuming but not effective
- API delivers data but still requires (intensive) data cleaning
- One time analysis of data only gives you a snapshot, to get deeper into a topic a constant or repeated analysis has to be run
- The complexity increases significantly if you use more data sources that connect/merge with each other

Questions?