Fontys Hogescholen

Proposal
Price predictive model

AI4 – AI41

Ömer Faruk GÖKBAK

3782174

Table of Contents

1. Introduction and the Project goal

In this project, I will go through a genuine challenge for the rental market of Amsterdam in Airbnb, from the basics of gathering data, data cleaning, visualization, up to using machine-learning and hyperparameter tuning to develop valuation models for the city's houses. Airbnb is one of the most popular home-sharing platforms, in which landlords or renters can rent out a single spare room or their entire home to travelers for short-term accommodations. The price of properties is listed and expected to set price by the host. Throughout the project, I want to see if AI can help hosts by predicting as accurately as possible the price with the highest profit of a new listing given its attributes. This is the main purpose that I am interested in, but if I come up with more insight when exploring the data, I can focus on new areas as well.

Since Airbnb is an extended platform all around the world and has many city-based data, I need to start off researching which data is fit to work on the Amsterdam short-time rental market and find out the best and convenient results. This will be a part of the proposal in the domain understanding. I also need to think about which impact this technology can have on society and the way price gets distributed and lastly, I need to explore the data by performing EDA. After I have done the domain research, societal impact, and EDA I can come up with a project proposal about what I think is a viable version for an AI project to genuine challenge.

2. Domain understanding

Airbnb is an online marketplace that offers home rental services for travelers, which stands first for "Airbed & Breakfast". It is one of the world's largest alternative accommodation platforms, bringing together those who are looking for short-term accommodation and those who want to rent their house. Although the main idea is to create a home-sharing platform that offers travelers cheaper accommodation; Over time, Airbnb has become a popular platform used not only by travelers but by many people around the world.

Airbnb matches people in 2 different categories below which are also the stakeholders.

1. People looking for a short-term home.
2. Landlords who want to rent their homes for a short time to tenants

It brings individuals together on Airbnb through the postings they share. These advertisements are listed under 3 main titles according to different criteria.

- Entire House/Apt
- Private Room
- Shared Room

What makes it possible for Airbnb to be more budget-friendly than hotels is that the house is rent by private individuals, not hotels. Therefore, you do not charge extra fees (food, drinks). Even though prices are freely set by hosts, Airbnb recommends an average price ('base price') based on other housing prices in your area. In this case, the landlord, who sets a price above the average, will have to lower the price because he cannot receive demand. This is an indication of why Airbnb usage is more economical. Price is a vital topic in the hospitality industry [1], for a proper pricing strategy can lead a company or enterprise to the next level of financial success [2]. Price plays an important role in the sharing economy in the hospitality industries such as Airbnb because price impacts guests' lodging selection and also significantly impacts hosts' profits [3]. Therefore, being aware of the elements that affect price also extremely matters, since it can help landlords in setting a fair price such that both the households as well as the tenants benefit from the sharing economy. The main goal of the project is to predict the best base price depends on the amenities, facilities, or priorities of the house.

Data definition & Data understanding

The data is publicly available on the website http://insideairbnb.com/get-the-data.html. It contains data from listings in many cities across Europe and America. The data is sourced from publicly available information from the Airbnb site, so there are no privacy issues. The dataset was scraped on 8 February 2021 and contains information on all Amsterdam Airbnb listings that were live on the site on that date. The data is supplied in ASCII text files, with a CSV format. The files contain approximately 20,000 rows. Attribute specification:

- **Id**: Unique id for the listing
- **Host_id**: Unique id for the host
- **Neighbourhood**: Neighbourhood of the property
- **Accommodates**: The number of people the property accommodates
- **Bedrooms**: The number of bedrooms
- **Bathrooms**: The number of bathrooms
- **Beds**: The number of beds
- **Property_type**: Property types (e.g. apartment)
- **Room_type**: Room type (e.g. entire home, private room, shared room)
- **Latitude**: Location of the property the level of latitude
- **Longitude**: Location of the property the level of longitude
- **Amenities**: The presence or absence of a wide range of amenities (discussed in further depth in a previous post, but including items like TVs, coffee machines, balconies, internet, and parking, whether or not the property is child-friendly, allows self-check-in, or allows pets, and many others)
- **Maximum_nights**: Maximum number of nights a guest can stay for the rental
- **Minimum_nights**: Minimum number of nights a guest can stay for the rental
- **Price**: Nightly price for the rental
- **Security_deposit**: The amount required as a security deposit
- **Cleaning_fee**: The amount of the cleaning fee (a fixed amount paid per booking)
- **Guests_included**: The number of guests included in the booking fee
- **Extra_people**: The price per additional guest above the guests_included price
- **Availability_30**: How many nights are available to be booked in the next 30 days

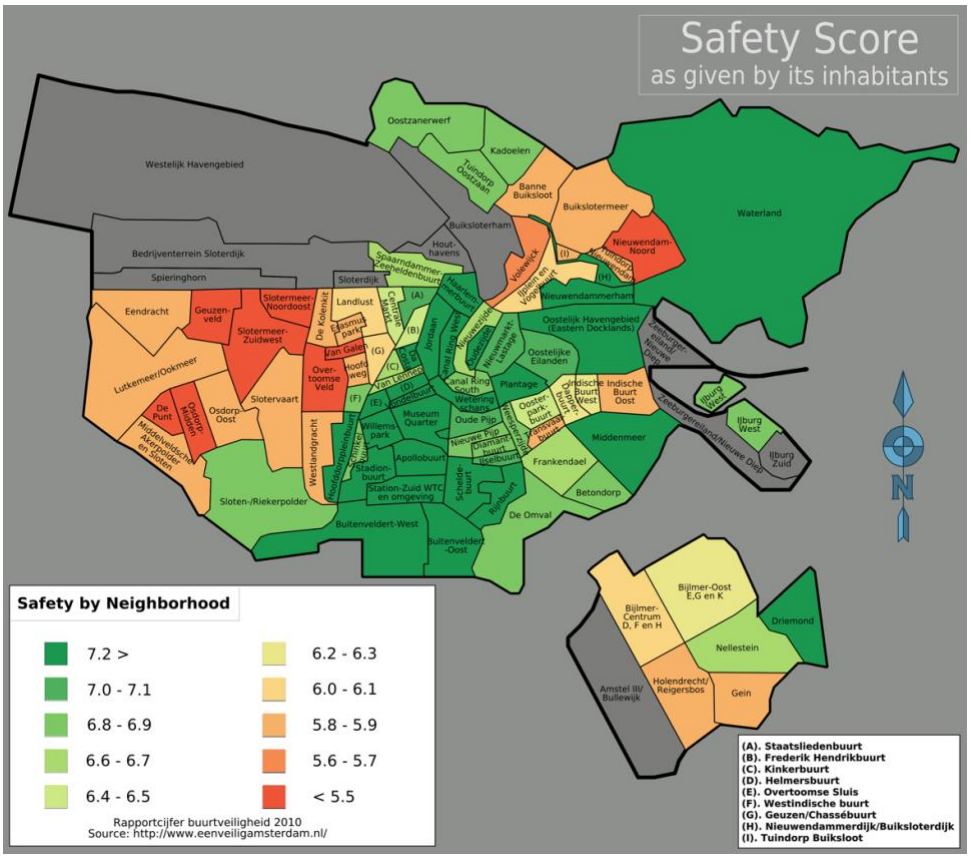
Neighborhoods in Amsterdam

Is Amsterdam safe? Areas to avoid in Amsterdam…

Amsterdam is ranked fourth in the world and third in Europe when it comes to personal security[5]. Check out the Safe Cities Index Report 2019.[6] That is not to say that no crime exists in Amsterdam – it does. Pickpockets and scammers roam around the city looking for easy targets (particularly in the Red-Light district) but, on the whole, the Dutch capital is a lot safer than most other European capitals.

Even previously 'rough' neighbourhoods such as Bos en Lommer, parts of Nieuw-West and Bijlmer, have improved significantly and unless you're roaming around late at night alone, you're highly unlikely to face any trouble or feel uncomfortable. Besides, these places have some great restaurants and cool cafes, clubs like De School and Radion, and independent boutiques popping up across the west of the city, the area is fast becoming the trendiest neighbourhood for young professionals and students. If you've booked an Airbnb in one of these neighbourhoods before reading some of the more negative press surrounding these parts of the city, don't panic! Just take the usual precautions that you would anywhere else.[7]

As there is no data on neighborhood safety in the Airbnb files, I searched for this online and came across the map below. As you can see, nowadays the western parts of the city are the most dangerous.



Amsterdam safety by neighborhood[8]

Accommodation tips

★ **First Time in Amsterdam?** Stay anywhere in the Old Centrum

★ **Layover in Amsterdam?** Anywhere near Zuid, Centraal, or Sloterdijk Stations, so you easily leave your bags.

★ **Best Neighbourhood in Amsterdam for Mature Travelers:** Oud-Zuid.

★ **Best Neighbourhood in Amsterdam for Families:** NOT near De Wallen! Oost, de Baarsjes, Plantage.

★ **Best Neighborhood in Amsterdam for Nightlife:** Near Leidseplein & Rembrandtplein

★ **Amsterdam on a Budget?** Consider staying in Haarlem.

★ **Most Romantic Neighborhood in Amsterdam:** Canal Belt & Oud-Zuid

★ **Local Neighborhood in Amsterdam:** De Pijp [9]

3. Societal impact

The societal impact assessment, which will be made to check whether all ethical areas of the technology have been covered, is made using the Technology Impact Cycle Tool which helps to determine potential social consequences of the technology. It makes this by divided ethics into different categories and asking related questions about each category. Based on the outcome of the assessment, it is decided whether or not to continue on the project.

QUICKSCAN - CANVAS

Airbnb price predictive model

NAME: Airbnb price predictive model
DATE: April 7, 2021 5:54 PM
DESCRIPTION OF TECHNOLOGY
A predictive model, that will help hosts by predicting as accurately as possible the price with the highest profit of a new listing in Airbnb given its attributes. Furthermore, it is also expected to increase the lease on Airbnb.

HUMAN VALUES

The new system does not affect anything about the way that the identity of the user is influenced, which is different from the current system.

TRANSPARENCY

Each step of the technology will be clarified transparently for each phase through the project by Jupyter notebooks.

IMPACT ON SOCIETY

A prediction tool, that will help the host by predicting as accurately as possible the price of a new listing in Airbnb given its attributes. As soon as attributes are filled, the model will compare attributes with similar data and predicts the best accurate price according to the features of the property. In the meantime, the model will count overpricing and underpricing issues to keep the highest profit for the host.

STAKEHOLDERS

- Host
- Tenant
- Airbnb

SUSTAINABILITY

This technology will use Google cloud services. According to Google, they invest in renewable energy in order to make their data centers more environmentally.

HATEFUL AND CRIMINAL ACTORS

The prediction model will not set the price by itself but will offer an advised price to the host, in order to find the best price for their properties. Therefore, I do not expect any breaking law issues from this model. However, it is counted that the model will affect free-market rates by predicting prices instead of human beings.

DATA

Missing values and outliers will be taken into account to avoid pitfalls in order to keep models performing as expected.

FUTURE

Utopia
The model can be extended for the other cities as well. Increasing on rented houses can be observed via the model.

Dystopia
Long-term leases local may be adversely affected by the increase in short-term rentals. Landlords may avoid long-term leases or increase their property prices further.

PRIVACY

The personal data in this model can be named as the name of the host and the coordinates of the house which are already public and accessible on the Airbnb official website to the visitors.

INCLUSIVITY

There is not any data in this technology that can be caused discrimination and bias.

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF A QUICKSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO

IMPROVEMENTSCAN - CANVAS

Airbnb price predictive model

NAME: Airbnb price predictive model
DATE: April 21, 2021 7:02 PM
DESCRIPTION OF TECHNOLOGY
A predictive model, that will help hosts by predicting as accurately as possible the price with the highest profit of a new listing in Airbnb given its attributes. Furthermore, it is also expected to increase the lease on Airbnb.

HUMAN VALUES

The main goal of the project is to keep safety and transparency to avoid harming human values. The process will be transparent about the kind of data used and any shortcomings of the software(e.g, data bias), and the scraped data is reliable and valid. Therefore, it is not being expected to change anything on human values.

TRANSPARENCY

I would like to clarify each step clearly what is done for what purpose.

IMPACT ON SOCIETY

I would like to inform the host that the price is provided by the system is the best accurate advised price and it is not mandatory to use the price in their listings. Thus, if the price doesn't satisfy the hosts, hosts could price listings whatever they want.

STAKEHOLDERS

I will create a model so that can offer the best price for both the tenant and the landlord, besides increases the lease on Airbnb.

SUSTAINABILITY

The main issue regarding sustainability in regard to this technology is the energy used for inputting and processing the data. By inputting and processing the data in fixed time intervals as opposed to constantly working with the data will help reduce the energy consumption, thus becoming more environmentally sustainable. Since this is technology is data-driven there is no specific lifespan, meaning that as long as there is relevant data the technology can continue working.

HATEFUL AND CRIMINAL ACTORS

After considering the various nefarious means that this data can be used for, the system will be using in the Airnbn while listing the property therefore, it would be safe with Airbnb since the security policies of Airbnb would protect the system and prevent accessing the bad actors.

DATA

The dataset is incomplete; this could be improved upon. I would advise Airbnb to have a better data collection system, to gain more benefits from AI systems. For now, this means that we have certain limitations to the system. Price changes on neither holidays nor festivals and sports competitions. So the best price prediction would be not as accurate as desired.

FUTURE

The ideal conditions, perfect predictions for all the properties. This would provide more trust in the predictive model. I would have a system where the technology would greatly aid the hosts and tenants but still be just a second measure to the real thing. In the worst conditions, a technology not very good at predicting the best price but hosts still trust it and can lose profit. They can be mislead to use the technology.

PRIVACY

The data is already public and accessible online on Airbnb. I am not in charge to keep data private. However, I am still responsible for the data that I use for the model, and the hostname which is not useful for predicting the price will be avoided using. Moreover, I will use the coordinates to reach the nearest accompanies to predict the best price for that neighborhood.

INCLUSIVITY

I would like to make this project applicable in other cities, not limiting to Amsterdam.

FIND US ON WWW.TICT.IO

THIS CANVAS IS PART OF THE TECHNOLOGY IMPACT CYCLE TOOL. THIS CANVAS IS THE RESULT OF AN IMPROVEMENTSCAN. YOU CAN FILL OUT THE FULL TICT ON WWW.TICT.IO

4. EDA

With the goal of figure out the effect of features on the price, the questions below are answered in the EDA.

1. *What is the overall distribution of prices?*
2. *How many listings do hosts have on average? How many multi-listing hosts are there?*
3. *What is the average number of people accommodated, bathrooms, bedrooms, and beds in Airbnb listings in Amsterdam, and how do prices differ?*
4. *What is the average price of extra people, security deposit, and cleaning fee in Airbnb listings in Amsterdam, and how do prices differ?*
5. *What are the most common property and room types, and how do prices differ?*
6. *What is the average number of amenities in Airbnb listings in Amsterdam, and how does the increase in numbers affect the price?*
7. Which areas have the most Airbnb properties, and which are the most expensive?

Neighborhood, room type, property type, number of the bedroom, bathroom, beds, and people accommodate, extra charges for extra services, and number of amenities. How can these features affect the price of listings? They are all explained in detail with the attached Jupyter notebook file.

i. Explored the dataset to better understand different fields.
ii. Presented different data visualization by using matplotlib, seaborn, and folium libraries.
iii. Identified the most relevant fields for my analysis.

5. Sources

1. Bull, A.O. Pricing a Motel's Location. Int. J. Contemp. Hosp. Manag. 1994, 6, 10–15. https://www.emerald.com/insight/content/doi/10.1108/09596119410070422/full/html
2. Hung, W.-T.; Shang, J.-K.; Wang, F.-C. Pricing determinants in the hotel industry: Quantile regression analysis. Int. J. Hosp. Manag. 2010, 29, 378–384. https://www.sciencedirect.com/science/article/abs/pii/S0278431909001029?via%3Dihub
3. Chiou, Y.-C.; Jou, R.-C.; Yang, C.-H. Factors affecting public transportation usage rate: Geographically weighted regression. Transport. Res. Pol. Pract. 2015, 78, 161–177.
4. *Amsterdam neighborhoods · Neighborhoods in Amsterdam*. Amsterdam Travel Guides. (2021). Retrieved 21 April 2021, from https://www.todaytourism.com/travel-guides/Neighborhoods-in-Amsterdam.html.
5. *Safety in the Amsterdam Area | Amsterdam*. Iamsterdam.com. (2021). Retrieved 21 April 2021, from https://www.iamsterdam.com/en/study/life-in-amsterdam/safety-in-amsterdam.
6. *Safe Cities Index 2019 Whitepaper*. Safe Cities. (2021). Retrieved 21 April 2021, from https://safecities.economist.com/safe-cities-index-2019/.
7. *Is Amsterdasafefe? Areas To Avoid In Amsterdam*. We Are Amsterdam. (2021). Retrieved 21 April 2021, from https://weareamsterdam.com/areas-to-avoid-in-amsterdam/.
8. *Amsterdam neighborhoods · Neighborhoods in Amsterdam*. Amsterdam Travel Guides. (2021). Retrieved 21 April 2021, from https://www.todaytourism.com/travel-guides/Neighborhoods-in-Amsterdam.html.
9. *Where to Stay in Amsterdam: Best Neighborhoods - By a Local*. Anna Everywhere. (2021). Retrieved 21 April 2021, from https://annaeverywhere.com/where-to-stay-in-amsterdam-neighborhoods/.