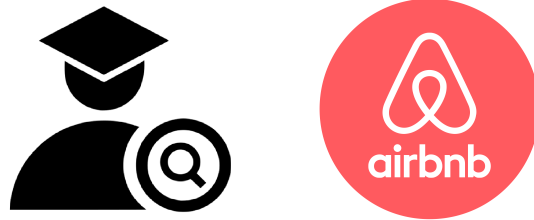


The relationship between the start of the academic year and the AirBnB market in student cities in Western Europe

2022-10-15

The relationship between the start of the academic year and the Airbnb market in student cities in Western Europe

This project uses archival data from the Airbnb website to investigate the relationship between the beginning of the academic year and the Airbnb market in five major Western-European cities.



Contents

1. Introduction
 - 1.1 *Problem background and research motivation*
 - 1.2 *Problem statement and research questions*
 - 1.3 *Conceptual model*
2. Methodology
 - 2.1 *Data*
 - 2.2 *Research method*
 - 2.3 *Analysis*
3. Results
4. Conclusion
5. Discussion
6. References

1. Introduction

1.1 Problem background and research motivation At the start of the academic year, many university students need temporary accommodation while they are arranging their documents and looking for permanent housing. Therefore, they often stay temporarily at Airbnb accommodations. In a recent study on the determinants of Airbnb listings' performance, Sainaghi et al. (2021) concludes that seasonality, and certain events can have a significant impact on the prices of listings. Hence, this project aims to investigate how

the Airbnb market in various Western-European cities is related to the period of the start of the academic year. The research will focus on analyzing the relationships between the start of the academic year and the Airbnb prices, as well as the maximum nights and accommodation can be booked for. In addition, the effect of the type of host (ordinary vs professional) and type of accommodation (room vs apartment) on the abovementioned relationships will be investigated.

The research will be executed for five major student cities in three Western-European countries: Belgium (Antwerp, Brussels), Germany (Berlin), The Netherlands (Amsterdam, Rotterdam).

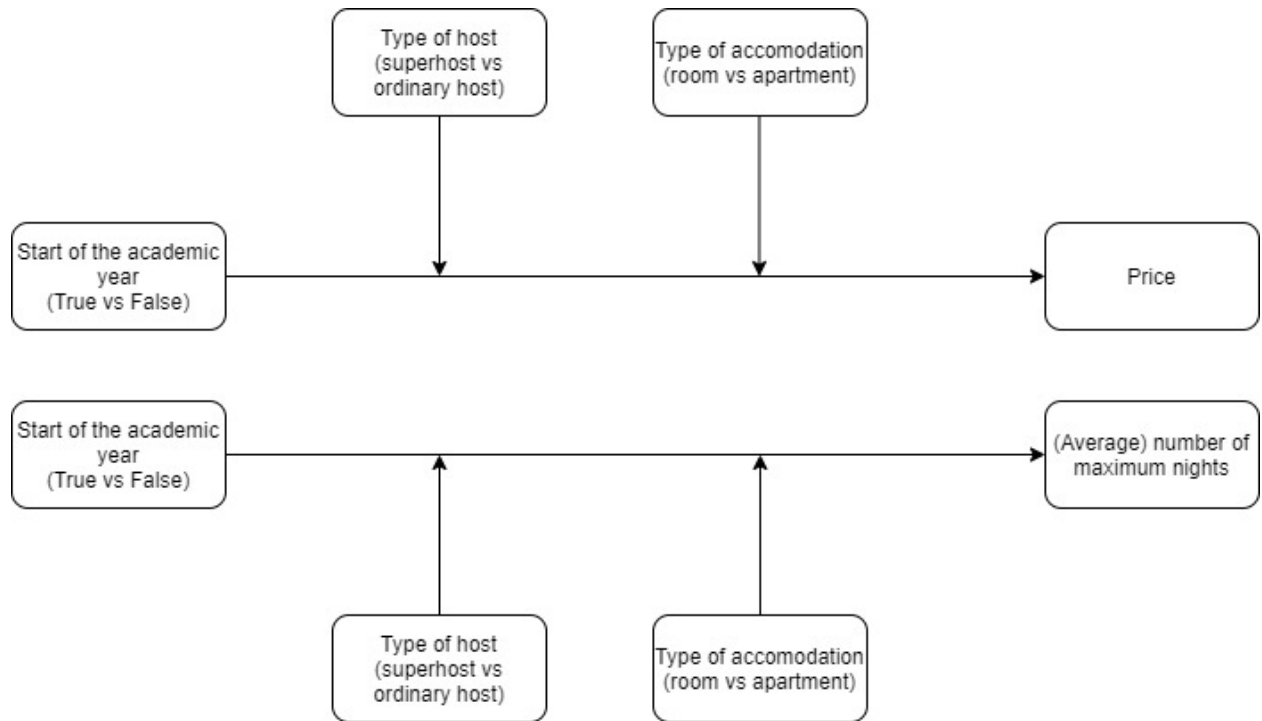
1.2 Problem statement and research questions Because of the arrival of many university students who have to arrange accommodation at the start of their university career, we expect that the period of the academic year affects the Airbnb market in Western-European cities. Therefore, we compose the following main research question:

To what extent does the start of the academic year affect the Airbnb market in popular student cities in Europe?

To investigate this problem statement, we look at the following sub-questions that aim to uncover the specific relationships between the start of the academic year and the two aspects of the Airbnb market of interest (price and average nights):

- How are the Airbnb prices related to the period of the academic year (beginning vs regular period)?
- How does the relationship between Airbnb prices and period of the academic year depend on the type of host (ordinary vs superhost)?
- How does the relationship between Airbnb prices and period of the academic year depend on the type of accommodation (room vs apartment)?
- How is the maximum number of nights per accommodation related to the period of the academic year (beginning vs regular period)?
- How does the relationship between the maximum number of nights per accommodation and period of the academic year depend on the type of host (ordinary vs superhost)?
- How does the relationship between the maximum number of nights per accommodation and period of the academic year depend on the type of accommodation (room vs apartment)?

1.3 Conceptual model To answer our research questions, we designed the following conceptual model. We hypothesize that there will be positive relationships between the start of the academic year and the price, and the start of the academic year and the average number of nights a listing is available for. In addition, we expect that these relationships will be moderated by the type of host (whether a host has the “superhost” badge or not) and the type of accommodation (an entire apartment as opposed to a private, shared or hotel room).



2. Methodology

2.1 Data For this project we used publicly available archival data from Airbnb. In total, ten datasets were retrieved from the website www.insideairbnb.com, two for each city of interest. For each of the five cities, two datasets were downloaded; namely, “listings.csv” and “calendar.csv”, both containing different variables. The listings dataset contained information per listing about the type of property and its features. The calendar dataset contained information on the price of the property on different dates and its availability. We combined these datasets together based on the listings id to create one dataset for each city. This data was then explored (see: “Data exploration raw data”). Later a selection of the variables we were interested in was made as well as selecting only data from the two time periods of interest (Aug/Sep 2022 and Mar/Apr 2023). This is described in more detail under “Data Transformation”.

Data exploration raw data First, the raw data was explored. We looked at all variables of interest to our research question. For the metric variables we made boxplots and took the summary statistics. For the categorical variables we took the count of each one to get an insight into their comparative numbers.

Maximum Nights

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	1	35	365	617	1125	1125
Amsterdam	1	22	180	493	1125	1825
Brussels	1	90	365	631.8	1125	1555
Berlin	1	29	365	518.2	1125	9999
Rotterdam	1	30	365	515.8	1125	1125

All dataset have the logical min. maximum number of nights at 1. The 3rd quartile number is the same for all cities. Brussels, Berlin and Rotterdam all have the same median. All the max. And 3rd quartile numbers don’t make much sense as they are so high that it would mean you could rent the property for

several years in a row. Likely this is some automatically filled in number or people don't think about it. Based on the differences between the median and mean for every city, the mean likely isn't representative and heavily influenced by the large outliers. But even the median is very high as apparently you can rent out the property for a full year.

Price

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	14	60	85	129	129	5800
Amsterdam	0	115	160	218.8	249	18400
Brussels	0	52	75	102.4	104	4112
Berlin	0	40	61	93.52	100	24999
Rotterdam	0	75	109	143.3	155	3261

For price there are very different outputs for all cities. Three cities have 0 as their minimum price, which doesn't make sense as that would mean the property was available for free. So likely this means that the property is rented out or not available. Looking at the median and mean Amsterdam is the most pricy city. Followed by Rotterdam, Antwerp, Brussels and then Berlin. The max. For many of the cities is also insanely high, so far off from the median and mean that once again this is likely some automatic fill out to indicate it's not available or a scrapping error. For all cities the price compared from the 1st quartile to the 3rd quartile is doubled, this can be because they are different types of rooms and have different price ranges. Or increase in price based on the tourist season or not.

Minumum Nights

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	1	1	2	5.625	3	500
Amsterdam	1	2	3	4.61	4	1001
Brussels	1	1	2	10.56	5	1125
Berlin	1	2	3	13.29	5	1124
Rotterdam	1	1	2	5.976	4	360

The minimum nights variable is similar for all cities. The mean once again differs greatly from the median and each city has a very large max. Values. So likely the outliers impact the mean strongly. The 2-3 minimum number of nights makes sense as cleaning and such gets more expensive and more to deal with the more often you have to do it. Demanding guest stay 2 nights is to lower the workload for the property manager.

Categorical variables

	Room_type					Host_is_superhost
Dataset:	Entire home/apt	Hotel room	Private room	Shared room	False	True
Antwerp	586,172	3650	142,715	1825	575,952	158,410
Amsterdam	1,541,395	27740	671,600	11680	1,858,215	394,200
Brussels	1,480,440	14600	550,055	8030	1,729,005	324,120
Berlin	3,535,025	55115	2,334,175	67525	5,231,180	757,010
Rotterdam	231,775	4745	111,690	5840	278,495	75555

Antwerp and Rotterdam are the smallest datasets. Berlin is the largest followed by Amsterdam and then Brussels. For all dataset the room_type show the same ratio for each type. Entire home has the largest

occurrence count, followed by Private room, then Hotel room and finally Shared room. So as the difference in counts are so large the decision was made to group Hotel room, private room and shared room together as one group.

Data Transformation Every two datasets per city were merged into one, based on listing ID. Afterwards, the data was explored and prepared.

A selection was made for all cities to take the dates from August 15 to September 15, 2022, as well as for March 15 to April 15, 2023. The first period corresponds to “the start of the academic year”, whereas the second is the alternative/ control period to compare the changes in the DVs. End of March/Beginning of April was chosen as a period with no important milestones for students, nor major national holidays or religious occasions to avoid biases that could impact the variation in the DVs.

We only selected the variables which were needed for the analysis when loading the data for each city into a dataframe. Therefore, we ended up with the variables ‘id’, ‘price’, ‘room_type’, and ‘maximum_nights’. We also created two new variables by taking the mean average of the price and the mean average of the maximum number of nights.

For the availability variable, we chose to use every data point regardless of the data point being listed as available or not. This was done because the ones listed as not available might have been rented out, and those listings were important for the results as we expect a higher demand for AirBnBs at the beginning of the academic year. Therefore, excluding those observations would bias the results.

After the selection of the data we also created a new dummy variable called `dummy_month_august`, which takes the value of 1 if the observation is for August/September (start of academic year) and 0 if it is in March/April.

Furthermore, we turned the variable `room_type` into a dummy variable called “room”. There are 4 different room types given; however, as three of them refer to single rooms, they were considered as one type. The dummy variable is coded as “Entire home/apartment” = 1, and 0 = “Private room” & “Shared room” & “Hotel room”. We also theorized grouping these together made sense, as these types would be the most popular among students, because these accommodations are usually meant for 1 or 2 people.

For the Amsterdam dataset, for example, these were the occurrences of each type:

Entire home / apartment	Hotel room	Private room	Shared room
8446	152	3680	64

After making boxplots and checking the summary output for our variable of interest, we removed the outliers for `average_price` and `average_nights` for all cities, as well as the outliers for the minimum number of nights for Berlin, as all these variables had extreme outliers which biased the mean and skewed the data.

We used a multiple methods to examine the datasets after transforming them.

The final variables in the datasets are as follows:

Variable	Description
<code>id</code>	Airbnb’s unique identifier for the listing
<code>host_is_superhost</code>	Boolean [t=true; f=false]
<code>room_type</code>	Categorical variable displaying the the type of room, 4 options are given; Entire home/apartment, Hotel room, Private room or Shared room.

Variable	Description
room	Binary variable. Transformed room_type, listing “Entire home/apartment” as 1 and “Private room”& “Shared room”& “Hotel room” combined as 0.
price	Daily price in local currency, numeric.
average_price	Mean price for each listing over 30 days, numeric.
maximum_nights	Maximum number of nights a listing can be rented out.
average_nights	Mean maximum_nights for each listing over 30 days, numeric.
minimum_nights	Minimum number of nights a listing can be rented out.
dummy_month_aug	Dummy variable listing August/September listings as 1 and March/April as 0.

Data exploration after transformation *Price*

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	14	65	80.88	90.95	110	213.28
Amsterdam	0	111.5	155	173.4	220.5	397
Brussels	0	51.58	70	74.83	93.42	158.12
Berlin	0	40	61	93.69	100	24999
Rotterdam	25	75	109.9	144.1	158.3	1312.5

For price there is a large range of values. Berlin, for instance, contains some rooms listed as 24999. Such extreme values might be due to the fact that the person renting out the property filled in the price incorrectly, or does not want to rent out the property, but have it listed. Some properties which have a minimum price of 0 might also be listed but not intended to be rented out. This is the case for Amsterdam, Brussels and Berlin. The prices also differ greatly per city, Amsterdam’s median price is more than twice the median price for Berlin. Amsterdam is the most expensive city of the five based on median and mean. As there are some quite extreme outliers, the median is more representative of the majority of the market.

Average nights

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	1	1	2	1.882	2	6
Amsterdam	1	2	2.3	2.364	3	4.5
Brussels	1	1	2	2.289	3	6
Berlin	1	2	3	13.17	5	1124
Rotterdam	1	1.25	2.0	6.087	4	360

The average nights variable shows great differences between the cities. In Amsterdam, Antwerp and Brussels, the mean average nights is around 2 nights, but in Berlin and Rotterdam this number is much higher. This might be because Berlin and Rotterdam both have high extremes, with 1124 days and 360 days as maximum respectively. On average, we can say guests prefer longer stays in Berlin and Rotterdam.

Minimum nights

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	1	1	2	5.519	3	500
Amsterdam	1	2	2	2.764	3	31
Brussels	1	1	2	2.844	3	999
Berlin	1	2	3	11.46	5	1124
Rotterdam	1	1	2	5.507	4	360

The mean minimum nights for all cities is 2 nights, except for Berlin where the median is 3 nights. The maximum minimum nights is much lower in Amsterdam compared to the other cities. In all other cities, there are some listings that can only be rented for a year or more per booking.

Maximum nights

Dataset:	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Antwerp	1	35	365	617	1125	1125
Amsterdam	1	21	180	496.3	1125	1825
Brussels	1	60	365	608.1	1125	1555
Berlin	1	29	365	581.2	1125	9999
Rotterdam	1	30	365	515.8	1125	1125

For the maximum nights we can see that on average median number of nights is a year for most cities. Only Amsterdam has fewer listed. This is likely because Amsterdam has a law about only being allowed to rent out your property for 30 days unless you get a permit. Rotterdam also has this rule but the allowed number of days are 60. So their median should be lower than the other cities as anyone listing their property without the permit can't rent it out for long. So it's quite odd then that Rotterdam still has this same high median as the other cities. This could mean that permits are perhaps easier to get there than in Amsterdam, or that this isn't checked as often and people are breaking the rules. All of the cities appear to have a somewhat insanely high number for their max. Number of days, as this would mean you could rent the property for several years at a time.

2.2 Research method This project will uses archival data which implies the research is correlational. Although the regression analysis could be used to derive causal relationships, further research would be needed to claim causal effects.

2.3 Analysis We have two models, each with three independent variables ("start of the academic year", "type of host", and "type of accomodation"), and a dependent variable ("price" in the first model and "number of maximum nights" in the second model). Since there are three categorical explanatory variables, and a continuous dependent variable in each model, we perform multiple linear regressions to analyze the relationships (Sekaran & Bogue, 2016). As we investigate the relationships for each of the two models with data for 5 different cities, there are ten linear regressions in total. Due to the fact that the focus of our research is to identify whether the start of the academic year, represented by "dummy_month_aug" is related to the DVs, and also to see whether this relationship depends on the moderators, we have introduced interaction effects between the IV of the study and the moderators ("room" and "host_is_superhostTRUE").

3. Results

Regression output explanation

When running the regression we have checked the effect of the independent variable namely 'dummy_month_aug' for each of the five cities on the two dependent variables; 'average_price' and 'average_nights'. As the focus

of our research is to identify whether the start of the academic year, represented by ‘dummy_month_aug’, has an effect on the DVs and also to see whether this effect depends on the moderators, we have introduced interaction effects between the IV of the study and the moderators, namely ‘room’ and ‘host_is_superhostTRUE’.

First model: DV = price

For Antwerp, there is no significant effect observed for the relationship between the start of the academic year and the price. Room type and superhost do appear to have a significant effect on the price. For entire apartments (value 1) the price is higher by €28.43 compared to the other room types. When the host is a superhost the price difference increases by €6.25 compared to prices of accommodations offered by ordinary hosts. The interactions between the independent variables are insignificant.

For Amsterdam, there is no significant effect observed for the relationship between the start of the academic year and the price. The room type has a highly significant effect, with a p-value below 0.001, and shows an increase of €54.22 when the accommodation is an entire apartment compared to a single room. The effect of host_is_superhost is also significant; however, the results indicate that superhosts’ listings are associated with lower prices (€8.11 difference). The interactions are not significant.

For Berlin, there is no significant effect observed for the relationship between the start of the academic year and the price. The moderator “room type” has a significant effect with a p-value below 0.01. The relationship is positive and indicates prices for entire apartments are €37.74 higher compared to the prices of rooms. In addition, listings by superhosts are pricier, with a €14.34 difference compared to listings by ordinary hosts.

For Brussels, there is no significant effect observed for the start of the academic year on price. The room type is highly significant with a p-value below 0.001 and shows a difference of €26.18 when the room type is entire apartment compared to single room. The effect of host_is_superhost is also significant, indicating that prices are higher by €3.74 for listings by superhosts compared to ordinary hosts. The interactions are not significant.

For Rotterdam, there is no significant effect for the start of the academic year on price. The effect of room type is significant and indicates prices of entire homes are higher by €39.48 compared to prices of single rooms. The effect of the moderator host_is_superhost is also significant and indicates that listings by superhosts are lower in price by €8.47 compared to listings by ordinary hosts.

Second model: DV = average (maximum) nights

For Antwerp, there is no significant effect observed for the start of the academic year on average nights. Only the effect of the moderator host_is_superhost is significant, meaning that superhosts rent out accommodations for 131.14 maximum nights more on average compared to ordinary hosts.

For Amsterdam, there is no significant effect observed for the start of the academic year on average nights. Only the effect of the moderator host_is_superhost is significant, with a 152.34 difference in favour of superhosts compared to ordinary hosts. In addition, entire apartments are rented out for 135.13 fewer maximum nights on average compared to single rooms.

For Berlin, no significant effect is observed for the relationship between the start of the academic year and the average (maximum) nights. Once again the moderators’ individual effects are significant, with 88.54 more average nights for superhosts compared to ordinary hosts, and 47.40 increase in number of average nights for entire homes compared to single rooms.

For Brussels, no significant effect is observed for the relationship between the start of the academic year and the average number of maximum nights. The moderators have significant individual effects, with 71.11 higher number of average nights for superhosts compared to ordinary hosts, and 80.45 more average nights for entire homes compared to single rooms.

For Rotterdam, there is no significant effect observed for the relationship between the start of the academic year and the average number of maximum nights. None of the effects of the moderators are significant either.

4. Conclusion

Based on the problem background, it was expected that there would be a positive relationship between the start of the academic year and the average price of Airbnbs, as students in need of temporary housing use Airbnb's services to find accommodation, resulting in higher demand and hosts increasing prices. In addition, we hypothesized that the average nights would also be higher at the start of the academic year, since students looking for permanent accommodation might need to rent and Airbnb for few weeks or even months. Finally, it was expected that the type of room and the type of host would influence the relationship between the start of the academic year and each of the two dependent variables. Nevertheless, no significant effects were found for the relationship between the start of the academic and the Airbnb prices, nor for the relationship between the start of the academic and the average maximum nights for which an accommodation is available. The interaction effects between the start of the academic year and the moderators were not significant for none of the models either. Although, initially not considered in the research questions, there were some significant effects of the moderators on price and on average maximum nights. The prices were higher for superhosts compared to ordinary hosts for Antwerp, Brussels and Rotterdam. Entire homes were associated with higher prices compared to rooms for all cities. The average number of maximum nights was higher for superhosts compared to ordinary hosts for all cities besides Rotterdam. The average number of maximum nights is also higher for entire homes compared to rooms for Amsterdam, Berlin and Brussels.

Based on the results, we conclude that:

- How are the Airbnb prices related to the period of the academic year (beginning vs regular period)?

We cannot derive meaningful conclusions regarding the relationship between the beginning of the academic year and the Airbnb prices due to the large p-values of the regression outputs.

- How does the relationship between Airbnb prices and period of the academic year depend on the type of host (ordinary vs superhost)?

This interaction did not have a significant effect for any of the cities. However, for most cities there was a positive effect caused by superhost. With superhost listings associated with a price higher by 3.74 euros to over 14 euros compared to listings by ordinary hosts, depending on the city. Nevertheless, the effect is opposite for Amsterdam and Rotterdam; therefore, further research needs to be done to draw conclusions.

- How does the relationship between Airbnb prices and period of the academic year depend on the type of accommodation (room vs apartment)?

This interaction effect is not significant for any of the cities. Nevertheless, the type of accommodation does have an individual positive effect on price for all of the cities.

- How is the average maximum number of nights per accommodation related to the period of the academic year (beginning vs regular period)?

The results of the analysis did not show significant individual effects of the relationship between the start of the academic year and the average maximum number of nights. Therefore, this research question also remains unanswered.

- How does the relationship between the maximum number of nights per accommodation and period of the academic year depend on the type of host (ordinary vs superhost)?

This interaction effect is not significant for any of the cities. However, for all of the cities besides Rotterdam there is a positive relationship between the superhost variable and average maximum number of nights.

- How does the relationship between the maximum number of nights per accommodation and period of the academic year depend on the type of accommodation (room vs apartment)? This interaction effect is not significant for any of the cities. Nevertheless, for Amsterdam, Berlin, and Brussels there is a positive relationship between roomtype and average maximum number of nights, meaning that entire homes are mostly rented out for a larger number of maximum nights (on average) compared to single rooms.

5. Discussion

Overall, this research did not find enough significant effects to draw meaningful generalizable conclusions. There are various reasons that could have caused this. First of all, this project used archival data that has not been collected for the purpose of this research. Therefore, we could not control for the quality of the data. Another possible explanation can be that not many students use Airbnb's services in these cities. We also have to keep in mind we are comparing the data from 2022 to 2023. So the prices listed for March/April might be changed closer to that time.

Some of the individual effects of the moderators were significant, but since they were not consistent for all cities, we cannot make conclusions without additional analysis.

In conclusion, further research is needed to confirm or reject the relationships between the variables in our models. It might be useful to repeat the research later in time, as well as for other student cities. In addition, two-way ANOVA could be performed for each model since the independent variables are categorical, whereas each DV is interval (Sekaran & Bogue, 2016), or the models could be tested together using a MANOVA analysis in order to take both dependent variables into consideration at once.

6. References

- Sainaghi, R., Abrate, G., & Mauri, A. (2021). Price and RevPAR determinants of Airbnb listings: Convergent and divergent evidence. *International Journal of Hospitality Management*, 92, 102709.
- Sekaran, U., & Bougie, R. (2016). *Research methods for business: a skill-building approach* (Seventh). John Wiley & Sons.