

Does the infrastructure of Amsterdam make a difference?

Francisco Riaño Martinez, Jan van der Doe, Kamila Majdlenová, Pomme Verhagen, Yuetong Bi

27.03.2022

Introduction

Do factors like distance to/from tram stops, metro stops, central station, swimming pools, religious sites, parks, Schiphol airport, sport facilities influence the prices of AirBnB listings in Amsterdam? And does the effect of the several infrastructure components on AirBnB pricing differ depending on whether there is a holiday/weekend or not?

Motivation

Infrastructure is the key to developing a successful tourism destination. Tourism industry stimulates investments in new infrastructure, most of which improves the living conditions of local residents as well as tourists. The reason for diving into this research is that the distances to transportation hubs and busy areas are important considerations for tourists when choosing an accommodation. Also other infrastructure components like swimming pools, sport facilities, religious sites or parks can be factors which improve the living conditions of local residents as well as tourists. These components are for some people of vital importance in choosing a place to stay. While AirBnB plays an important role in the tourism industry of today, infrastructure components might influence prices of AirBnB listings accordingly. Moreover, the holiday season and weekends can be contributing factors to prices of AirBnB listings because the demand for accommodations increases. Therefore questions like: 'Does the distance to/from Amsterdam Central Station or the distance to/from Schiphol Airport in relation to the location of the AirBnB listing influence the prices of that AirBnB listing?' and 'does the effect of the several infrastructure components on AirBnB pricing differ depending on whether there is a holiday/weekend or not?' could be asked. This code is written in order to find answers to these questions which focuses on AirBnB listings in Amsterdam.

Background

There is a large amount of research done exploring AirBnB pricing and its relationship to location (Deboosere, Kerrigan, Wachsmuth & El-Geneidy, 2019; Gibbs, Guttentag, Gretzel, Yao & Morton, 2018; Gunter & Önder, 2017; Gutiérrez, García-Palomares, Romanillos & Salas-Olmedo, 2017). Study of Gunter & Önder (2017) show that the further the AirBnB listing is from the city center the less willing are the consumers to make the booking. If the demand decreases the more remote listings might have to lower their prices to attract consumers. This seems to be true in Barcelona where the price of the accommodation goes down the further it is from tourist attractions (Gutiérrez et al., 2017). Moreover, Deboosere et al. (2019) found similar results where neighborhood accessibility was affecting the price of accommodation in New York. However, the opposite response was observed for pricing of AirBnB listings during weekend (Gibbs et al., 2018). Compared to hotels where the price went down AirBnB prices increased during weekend (Gibbs et al., 2018). Therefore, it remains unclear how the infrastructure in Amsterdam influences the price and whether this relationship differs if there is a weekend/holiday or not.

Data

The data used in this project comes from the website insideAirBnB.com which is an open source activist project that provides publicly available data from AirBnB. Moreover, infrastructure information was collected from the city of Amsterdam website at maps.amsterdam.nl. Lastly, a new dataset was created with Dutch holidays.

Data Collection

This research focuses on price analysis of AirBnB in Amsterdam, the Netherlands between years 2012 and 2019. Therefore, only data from Amsterdam in 2012-2019 is downloaded from the website, namely, listings and calendar information. The downloaded files were stored in Google Drive to ensure the same dataset can also be accessed in the future. To download the data see data/download_data.R.

Data Processing

Several steps were taken in order to create a final dataset that will be used throughout the analysis. Firstly, in the holidays dataset a new variable “weekend” was created. Next, the datasets listings, distance, calendar and holidays were joined together creating a new merged dataset. Furthermore, two new constructs per each infrastructure point of interest were created. The first variable calculates the distance from each accommodation to metro, tram, swimming pool, Schiphol airport, Central station, place of religion, park, sport facilities. The second variable states how many of the eight above mentioned infrastructure measures can be found within 500 meters. The last step was to clean the data, during which missing values were reviewed and price variables were converted from character to numeric type. The final cleaned dataset that will be used for the analysis was saved to folder gen/output.

Methodology

The research method that is used in this project is a regression analysis. This method is used to predict the influence of the independent variables ‘Distance to/from tram stops’, ‘Distance to/from metro stops’, ‘Distance to/from Amsterdam Central Station’, ‘Distance to/from swimming pools’, ‘Distance to/from religious sites’, ‘Distance to/from parks’, ‘Distance to/from Schiphol Airport’ and ‘Distance to/from sport facilities’ on the dependent variable ‘Prices of AirBnB listings’. Furthermore the moderators ‘Holidays in NL’ and ‘Weekend’ will be included and analyzed. According to this model, the research questions can be answered.

Results

This paper studies several infrastructure factors affecting the price of AirBnB listings and calculate the correlation between the four variables of price, adjusted price, cs_dist (distance to/from AMS central station), and holiday by rcorr(type=“spearman”) function to calculate the Spearman correlation coefficients between the response variables.

By plotting the graphs between the variables schiphol_dist (distance to/from Schiphol airport), tram_dist (distance to tram stops), metro_dist (distance to/from metro stops), cs_dist (distance to/from AMS central station), swim_dist (distance to/from swimming pool), religion_dist (distance to/from the places of religion), park_dist (distance to/from park), and sport_dist (distance to/from sport facilities) and price to visualize a general overview of their correlations.

According to the plots it can be found that schiphol_dist is positively correlated with price, while the rest of the independent variables are negatively correlated with price. Since the value of dist is particularly large, the log() function is used to process the value for visualization again in order to prevent the influence of the huge value on the model.

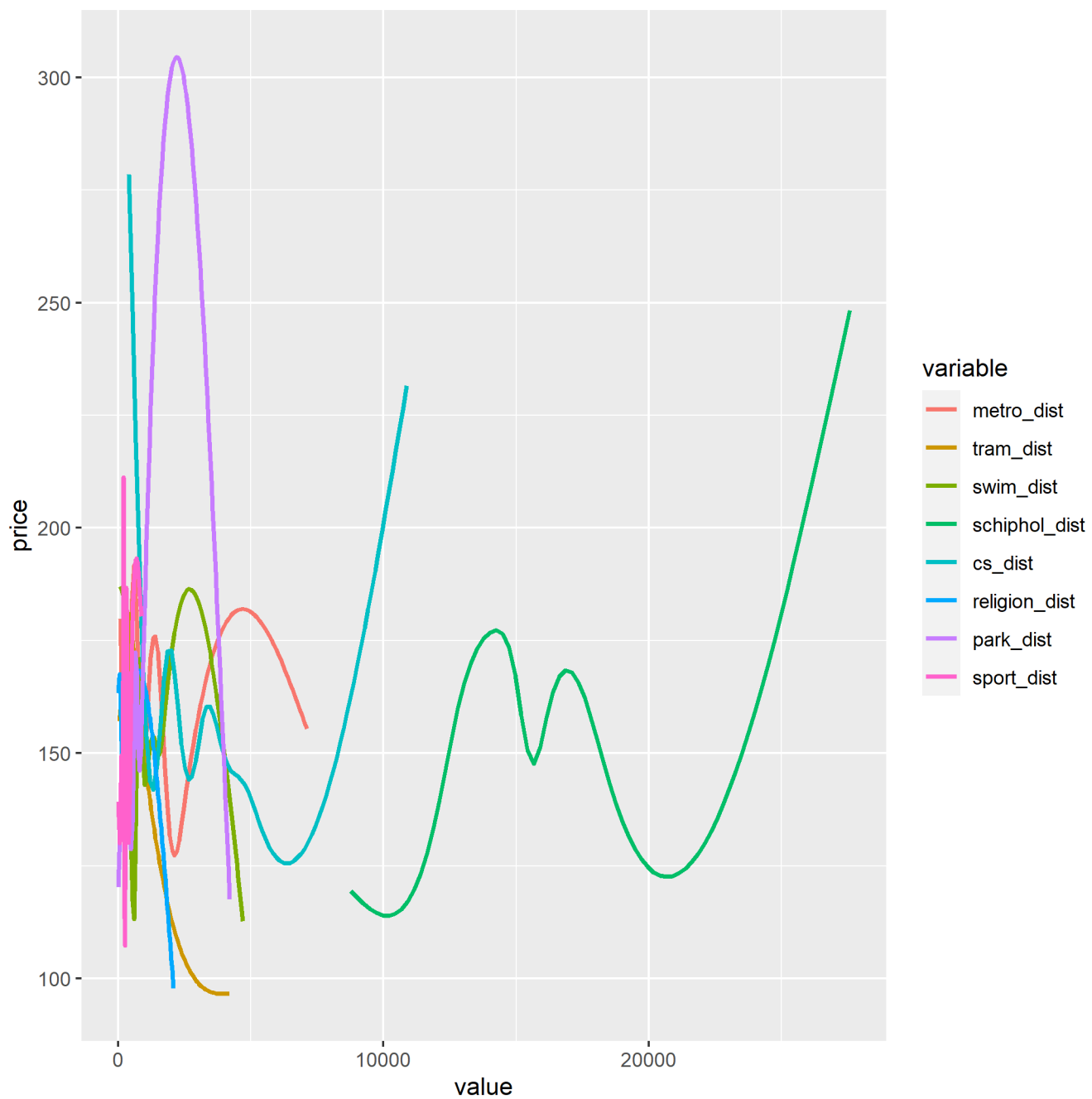


Figure 1: Relationship between price and distance

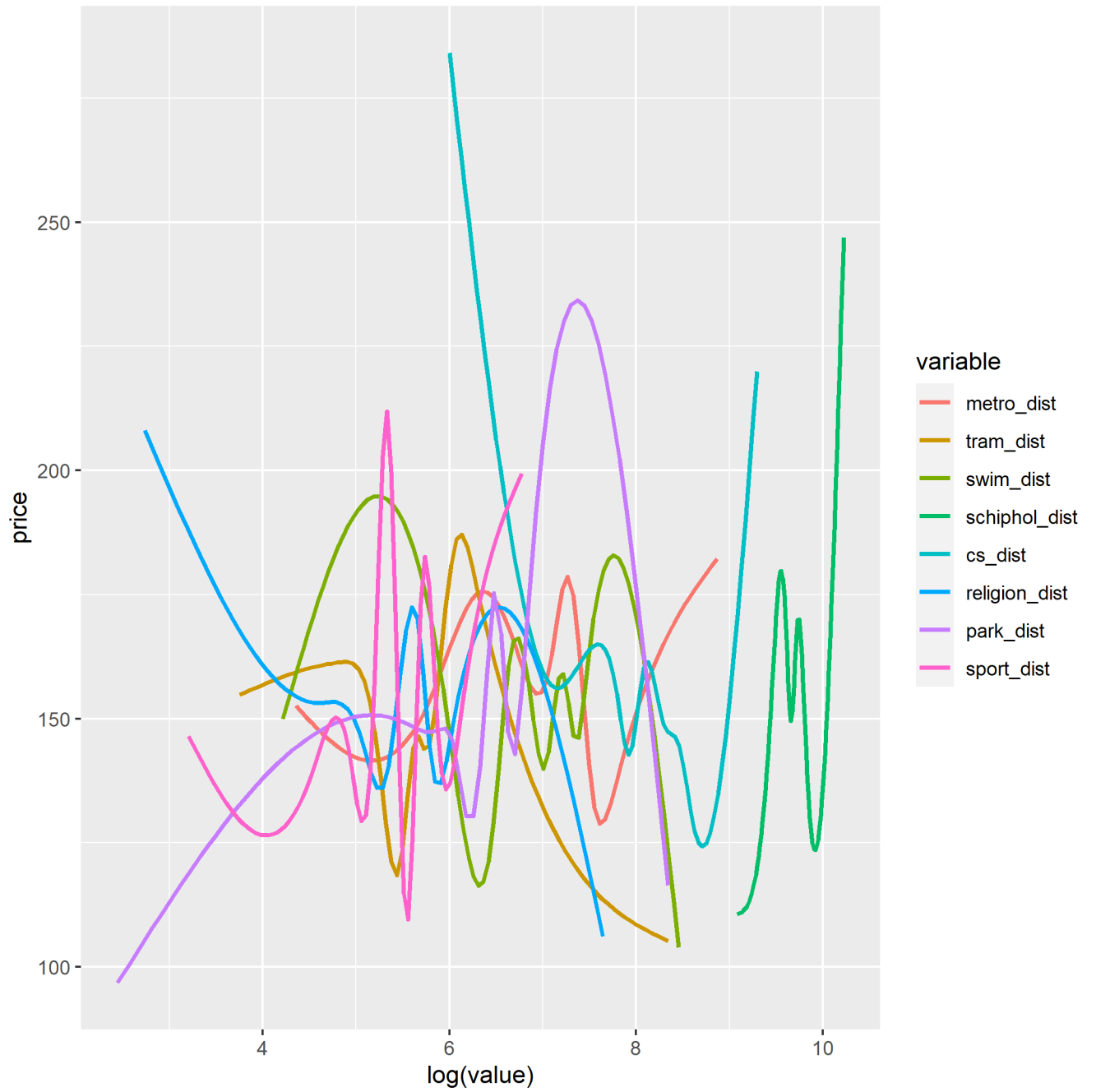


Figure 2: Relationship between price and log distance

Based on further visualization, it can be found that price around 175 is a significant inflection point that changes the direction. To be more specific, the listings with prices below 175 showing opposite results to those with prices above 175. For listings priced less than 175, there is a positive correlation between price and each variable. For the listings with price greater than 175, the relationship between price and the variables shows a negative relationship.

It can also be concluded that the infrastructure is often related to the distance. For further research, this paper analyzes the number of metros, airports, AMS central station, etc. within a 500-meter radius, and uses `slice_sample(n=100000)` to extract data from a sample of 100,000, uses the `total ASE` function of `ggplot` to map the x-axis and y-axis, and uses the `geom_col()` function to draw a histogram. According to the histogram of the number of each infrastructure and the price of the neighborhood within 500 meters, it can be found that the number conforms to the right-hand side normal distribution with a mean value of 0, and most of the numbers are distributed in the range of 0 to 3.

The study also considers the correlation between independent variables, the possible interaction factors would affect the result. Weekend and holidays (dummy variables) are set as two moderators and establish a `model__poly__log__Int` multiple regression model.

Based on the results of summary of multiple regression model (`model__poly__log__Int`), it can be shown that most of the variables are set within the 99.9% confidence interval, indicating that these variables have a significant correlation with price. The adjusted R square shows that how much variation of price of AirBnB is explained by the independent variables of infrastructure factors in the regression model. The adjusted R is 0.05734, which means multiple regression model is a good fit, that proves the effectiveness of the model. Through this model the price of AirBnB according to the different value of independent variables can be predicted.

Conclusion

To conclude, it can be stated that every distance-variable (Schiphol Airport, Central Station, tram stops, metro stops, swimming pools, religious sites, parks, sport facilities) has a significant effect on the prices of AirBnB listings in Amsterdam, whereas the interaction effect of holiday and weekend show no significant effect on the model. This explains that distances from and to important infrastructure sites play an important role in the price setting of AirBnB listings. The positive inflection point with regards to the distance to/from Schiphol airport could be due to the fact of noise pollution of the airport. It can be assumed that residents and tourists of AirBnB listings do not want to be as close to the infrastructure (in particular for Schiphol) while also staying not too far away. Hence, the positive inflection point at a price of 175. In the distance plots of cs, tram, metro and religion sites, a small increase at first is followed by a negative relationship between the distances and the prices of AirBnB. This can also be related to the fact of noises nearby the AirBnB listings while at the same time being mobile and staying close to the infrastructure sites. Parks, swimming pools and sport facilities show an immediate negative relationship between the distances of these infrastructure sites and the prices of AirBnB listings. This indicates the importance of these infrastructures sites nearby the AirBnB listings. All in all, the closer the infrastructure sites, the higher the prices of AirBnB listings (except for Schiphol airport). This shows that tourists and residents value the infrastructure sites to be nearby AirBnB listings which in turn drives prices up.

Managerial Relevance

Prices of AirBnB accommodation vary across cities and countries. There are various factors that can play a role in the price setting but this study focuses on the infrastructure as it is a crucial aspect when deciding where to travel. If the analysis shows that living close to certain kind of infrastructure leads to higher price, these insights can be further used by the management of AirBnB as well as the host for more effective price setting. Public policy makers can also benefit from this research. If there is a significant effect of infrastructure on pricing, they should invest in developing and improving it.

Regression model	
	Model 1
(Intercept)	-93850.911*** (2177.082)
logtram_dist	-149.872*** (7.228)
tram_number_in_500	-0.692+ (0.358)
I(logtram_dist^2)	14.708*** (0.683)
log(metro_dist)	199.405*** (8.059)
metro_number_in_500	-22.279*** (1.145)
I(log(metro_dist)^2)	-15.751*** (0.625)
log(cs_dist)	-1247.366*** (26.508)
cs_number_in_500	-161.886*** (4.981)
I(log(cs_dist)^2)	84.163*** (1.819)
logswim_dist	-124.179*** (8.798)
swim_number_in_500	6.603*** (1.227)
I(logswim_dist^2)	9.020*** (0.710)
logreligion_dist	-67.470*** (4.534)
religion_number_in_500	4.902*** (0.236)
I(logreligion_dist^2)	6.280*** (0.445)
logpark_dist	-49.917*** (4.886)
park_number_in_500	5.657*** (0.829)
I(logpark_dist^2)	6.584*** (0.449)
logschiphol_dist	20391.161*** (464.794)
I(logschiphol_dist^2)	-1049.862*** (24.135)
logsport_dist	10.597+ (5.414)
sport_number_in_500	-2.228*** (0.146)
I(logsport_dist^2)	-0.619 (0.535)
WeekendTRUE	28.290 (45.605)
Holiday	27.945 (49.034)
neighbourhood_cleansedBuitenveldert - Zuidas	-69.814*** (3.624)
neighbourhood_cleansedCentrum-Oost	-4.345 (2.913)
neighbourhood_cleansedCentrum-West	-35.533*** (2.891)
neighbourhood_cleansedDe Baarsjes - Oud-West	-23.195*** (2.208)
neighbourhood_cleansedDe Pijp - Rivierenbuurt	-36.442*** (2.541)
neighbourhood_cleansedGeuzenveld - Slotermeer	-59.624*** (5.442)
neighbourhood_cleansedJburg - Zeeburgereiland	167.798*** (6.153)
neighbourhood_cleansedNoord-Oost	-267.697*** (11.702)
neighbourhood_cleansedNoord-West	-300.223*** (8.288)
neighbourhood_cleansedOostelijk Havengebied - Indische Buurt	-47.105*** (4.345)
neighbourhood_cleansedOud-Noord	-107.430*** (4.462)
neighbourhood_cleansedOud-Oost	-46.656*** (3.263)
neighbourhood_cleansedSlotervaart	89.096*** (4.245)
neighbourhood_cleansedWatergraafsmeer	-106.485*** (4.022)
neighbourhood_cleansedWesterpark	-1.033 (2.658)
neighbourhood_cleansedZuid	6.048** (2.229)
log(metro_dist) * WeekendTRUE	0.286 (0.939)
logtram_dist * WeekendTRUE	-0.233 (1.025)
log(cs_dist) * WeekendTRUE	-1.286 (1.417)
logschiphol_dist * WeekendTRUE	-1.836 (4.308)
log(metro_dist) * Holiday	0.376 (1.010)
logtram_dist * Holiday	0.013 (1.102)
log(cs_dist) * Holiday	-2.322 (1.524)
logschiphol_dist * Holiday	-1.152 (4.632)
Num.Obs.	81940
R2	0.174
R2 Adj.	0.173
RMSE	85.62

Figure 3: Regression model

Limitations and Future Directions of Research

This study also has some limitations. The research was geographically limited to Amsterdam so future studies could focus on different countries or cities to test the generalizability of the results. Moreover, only data collected from AirBnB were analyzed in this study. Thus, focusing on other hospitality providers or another industry might bring interesting results. Furthermore, other factors that likely have effect on price could be included in the future studies. For example, reviews and ratings as well as review sentiment might be driving or influencing the relationship to some extent. Hence, building upon the current study by adding impact of reviews could provide new insights.

References

- Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). Location, location and professionalization: a multilevel hedonic analysis of AirBnB listing prices and revenue. *Regional Studies, Regional Science*, 6(1), 143–156.
- Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by AirBnB hosts. *International Journal of Contemporary Hospitality Management*, 30(1), 2–20.
- Gunter, U., & Önder, I. (2017). Determinants of AirBnB demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics*, 24(3), 270–293.
- Gutiérrez, J., García-Palomares, J. C., Romanillos, G., & Salas-Olmedo, M. H. (2017). The eruption of AirBnB in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*, 62, 278–291.