Technical report

Airbnb in Amsterdam during COVID 19

Laura Trapero, Amsterdam, July 2021

Table of Contents

[Summary 2](#_Toc76544578)

[1. Introduction to the project 2](#_Toc76544579)

[1.1. Project description 2](#_Toc76544580)

[1.2. Objectives 2](#_Toc76544581)

[1.3. Questions & Hypotheses 3](#_Toc76544582)

[2. Data 3](#_Toc76544583)

[3. Analysis 3](#_Toc76544584)

[4. Conclusions 3](#_Toc76544585)

[5. Annex 4](#_Toc76544586)

# Summary

# 1. Introduction to the project

Since the start of the corona crisis in 2020, tourist movement have been worldwide changed.

Every country would announce restrictions, during this period, affecting the number of visitors arriving to the country. Obviously, that would be reflected in the touristic sector.

The hotels and accomodation industry are struggeling since then, to survive. The occupatation is directly proportional on severe the restriccion are in every moment.

Amsterdam, as other touristic destination cities, have been hardly hitted by the corona crisis.

Before corona, Amsterdam tourism used to be as follows[[1]](#footnote-1):

- the half of all foreign hotel guests in the Netherlands.

- visitors from abroad account for 84% of the total figure (nationally: 55%).

- the visitor mix is comprised as follows: 57% are European visitors, 27% are intercontinental visitors and 16% are Dutch.

- the seasonal peak for foreign visitors is traditionally in the periods April-May and July- August, with the peak period for domestic visitors being the months April, May and/or June.

- one-third of Amsterdam hotel overnight stays are for business purposes. The business market's seasonal peaks are the periods April-May and September-October. The recreational peak is July-August.

## 1.1. Project description

In this project we would like to explore how corona crisis is affecting the touristic sector in Amsterdam and for that, we will perform an exploratory analysis of the Airbnb bookings and prices in the city of Amsterdam. Amsterdam might be a good city to have some hints of the touristic corona pattern as it receives most of the non-national visitors in the Netherlands.

## 1.2. Objectives

Characterise the Airbnb locations spatial distribution by distrcit at this moment in the city, according to the number of bookings, locations, the safety, green spaces, prices and type of accomodation.

Explore how the number of locations, bookings, type of rental have changed during corona time.

Analysis of the evolution of prices before and during corona time to see if the industry is recovering.

## 1.3. Hypotheses

We expect to find:

- The most bookings in the disticts that are more centric and popular, with average prices and a good proportion between green areas and criminality.

* More Airbnb locations and bookings in the most popular and centric neighbourhoods.
* A relationship between bookings per district and green areas, price, safety and popular and centric neighbourhoods
* A decrease of locations due to the corona.
* An impact on the price between pandemic and pre pandemic with a decrease on rentals on single rooms vs entire apartment.
* A fluctuation on the prices during this period should follow a pattern according to the country restrictions of mobility and touristic activities.
* A fluctuation as well on the booking during this period following the same tendency.

# 2. Data

The data consists of several data sets.

The data sets from Aribnb haven been download from their official open data source. This data is just available for the las 365 days.

The older data was obtained from Kaggle rom people who perform studies on the previous years.

Those data sets were combined.

One has information about the different accommodations and its features as type of accommodation, total number of reviews, average reviews per month, location gps coordinates, neighbourhood location between other.

A second data set contains all the dates when a review was entered and the id of the accommodation from December 2018 till now.

The last one of the Airbnb data sets consists of a daily price, daily availability per accommodation. We have this information from December 2018 till December 2020 and from April 2021 till July 2021.

A second set of data sets are the ones obtain from de data open source from the city Hall of Amsterdam.

There is a data set with information about the crimes per district for the last 4 years and another that gives us information about the green areas as well per district.

A third data set was needed to define in which district was located the neighbourhoods of the listings data set from Airbnb.

# 3. Analysis

## 3.1. Characterisation of the Airbnb locations distribution by district

The majority of locations are found in the District West followed by Zuid, Centrum and Oost.

Those four districts are the most centric and well connected. Popular and trendy neighbourhood are contained on those districts. This confirm our assumptions.

We found less locations in Centrum perhaps because of the high cost of housing in this district.

We decided to use the variable average of reviews as an indicator of bookings, as if there is a review it has been a booking as well. The problem that is not the other way around and it can be delayed till 15 days after the stay.

We can observe more average reviews on Centrum. That can be due to the popularity of this area. It is followed by Zuidoost, Noord and Nieuw West. Zuidoost and Nieuw West are located further and are neighborhoods with more problematic areas. This can be the reason we find so many reviews, we should have to look if those are positive or negative to confirm those number. In this case, the reviews might not be a good indicator of number of bookings. Noord, as we said before is growing and popularity and cultural offer, being renovated and more diverse than before, offering a restructured area.

West, Oost and Zuid have similar number of reviews. Those are popular districts, and it would make sense they are demanded similarly.

When we look at the prices between the neighbourhoods, the numbers shows what we would expect, higher prices in the most popular, centric and with more accommodations in the district (Centrum, Zuid, West, Oost and Noord). Those neighbourhood have or a higher or similar price to the city median price.

We have checked as well how are the private, entire apartments distributed by district and all of them, except for Zuidoost, show much higher number of locations that are entire rentals.

For each district, data about green spaces and criminality was available. In terms on criminality (average number of total crimes in the las 4 years) the worst district is Centrum followed by Nieuw West and Zuidoost. That would make sense, as there is a high density of people in Centrum as well as tourist and criminals might act there more often. Surprisingly, Noord is the one with less. Probably for the reason we have already described before.

On the other hand, Nieuw West is the dstrict with more number and extension of green spaces, followed by Noord and Oost. After comes Zuid and West.

Zuidoost has a big extension but small number of them, meaning there are just a few but a couple of them are really big.

We wanted to know which features were explaining the number of bookings (in our case, reviews) per district. This could be used to know if they have preferences for certain areas depending on those features.

To explore this we have perform a correlation matrix. If we would have to do a multiple regression analysis, the average reviews per neighbourhood would have been the response variable and locations, green areas and criminality as independent variables. For now we will just explore, if it is intersting to continue this direction with the variable we have.

We expect to find correlation between bookings/reviews and the and prices (negative), positive with green areas, negative with criminality between districts and positive with number of locations.

The extension can influence in the space available for housing and the number of parks should be a positive factor to consider while choosing a location, as Amsterdam has beautiful parks in all the district, especially the ones more centric.

Noord, Zuid and Oost would be the districts with a better proportion between criminality and green areas.

We have hypothesis that the locations and bookings on the district would be related as well to the safety and green areas.

When we look at the correlation between locations and booking with our variables, we find strong correlation between:

Location- green area: negative strong correlation

The more space is taken by green areas, the less spaces there is for building and, therefore, a smaller number of locations in the district.

Location-price: positive correlation

The more locations in a district, the higher is the price. As we have seen also that those are the most centric, popular districts.

Reviews-num. of crimes: positive strong correlation. There could be an influence of bad reviews in the most problematic districts, making this variable not a good indicator for those.

Price-green area: negative strong correlation. The more surface of green areas, the further the districts are from the center where we find more locations and the price is higher.

We will not continue further with this analysis as our purpose is more exploratory and, to be able to perform this multiple regression analysis, we would have to choose between the intercorrelated independent analyisis and add, other intrinsic features from the appartment that we have not considered.

## 3.2. Explore how the locations, bookings, type of rental have changed during corona time

We observe a decrease an incremental decrease of locations from December 18-ril 2020- December 2020- April 2020. This might be showing the impact of the corona crisis. This is what we have expected knowing already how corona has affectes this type of industry.

We notice a clear decrease of both types of rooms. We just picked the ones that were reviewed. Both types of accommodations have been reduced during corona time, expect for the private rooms in 2020. This can be perhaps a bad indicator for booking as maybe with more affluence in that date and with the corona, users had the necessity to leave more often than before a review to explain their shared experience during corona times.

When we look at the number of accommodations of each type that has been reviewed, we also see that more or less follow the lockdown-release of measure pattern.

## 3.3. Analysis of the evolution of prices before and during corona time to see if the sector is recovering

A fluctuation on the prices during this period should follow a pattern according to the country restrictions of mobility and touristic activities.

Prices are fluctuating on our analysis following the lockdown-release of measure pattern. We see the big difference between the daily price median pre corona and during corona. It looks like it improves slowly after the lockdowns. It would be also interesting to visualize it against the corona patients per day.

It seems the prices are recovering during the last period.

A fluctuation as well on the booking during this period following the same tendency.

Our analysis shows a clear match on the changes of tendency for those two variables with the moments of the national lockdowns and the moments where the measures were softened. At this moment the number of reviews is really low, we don’t know if the booking would we as well. Peaks on special dates. Here we can see that people staying at private rooms they are more likely to leave reviews. Again, a sign to don´t use this variable as an indicator.

# 4. Conclusions

* If wel define as a good district the ones with good connexions to the center, with a good average price compare to the total average price in the city and not a lot of criminality and some green spaces, the best district in the city to locate a Airbnb are Zuid,West,Oost,Noord
  + There are more locations on centric (inside the ring), popular, trendy districts.
  + The prices are also on average higher on those.
  + Except for Centrum, the other districts have a good proportion between green areas and criminality.
* The number of reviews might not be a good indicator of the number of bookings or for preferences on bookings by type of accommodation.
* The descriptors chosen (locations, prices, green areas, criminality) might not be good to predict the number of bookings by district considering reviews as the response variable.
* We can see a clear influence of the corona crisis on the price and the number of reviews of the the moments where the national lockdowns were fixed, and measure were released. Prices seem to be slowly recovering but still under the pre corona one.
  + Decrease of the daily average price
  + More fluctuation on the average prices for the weekend (to compensate the decrease on the average prices after corona)

1. Informatic and statistic report. Tourism AMA 2019-2020. Government of the NL ([https://data.amsterdam.nl/publicaties/publicatie/tourism-ama-2019-2020/1cae9aa6-e53e-49fd-a4c9-3fc4ffb3653e/)](https://data.amsterdam.nl/publicaties/publicatie/tourism-ama-2019-2020/1cae9aa6-e53e-49fd-a4c9-3fc4ffb3653e/) [↑](#footnote-ref-1)