DAO2702 Group Project

Tutorial Group 4 | Team name: printhack

Cui Tingwei	A0173218H
Leong Kai Yuan	A0171816A
Ngo Peh Choon, Nicholas	A0167378M
Sabrina-Jan Ong Wei Ting	A0172072M
Zhou Kai Jing	A0171366A

Table of Contents

Table of Figures

1.0 Case study description

In this section, we discuss our business case and the dataset we use for analysis and solving our case.

1.1 Our Business Case

Accommodation has always been one of the largest problems travellers face when visiting new places around the world (Lawton, 2017). Regardless of where you are, accommodation is one of the things we must have else we would risk spending homeless nights on our trip. One of the most prevalent problems travellers face with accommodation is the overcharging of their accommodation, which was made worse due to the language barrier in foreign countries.

However, this all changed in 2008, when Airbnb was established as an accommodation provider. Since its establishment in 2008, Airbnb has quickly gained popularity and has now grown to become one of the largest providers of accommodation for travellers (Medium, 2017). Today, Airbnb continues to provide large-scale accommodation services internationally and their business is valued at more than $25 billion (Momentum, 2016).

One of the reasons behind Airbnb's success is their ability to provide affordable accommodation across many different countries around the world. This highlights the importance of affordable accommodation and since accommodation accounts for approximately one-fifth of total travel costs (ValuePenguin, n.d.), price is definitely of big influence in travellers' decisions. Hence, we will be assuming the role of travel magazine DAO Magazine Co. with the purpose of coming up with an article on the factors affecting accommodation prices. To supplement us in the article writing, our aim in this project will be to provide insights on the different factors that affects accommodation prices to better understand what drives consumers' decisions when making travel plans, based on the dataset we have obtained.

For the purposes of this project, the country for analysis has been set at Amsterdam, which is one of the Top 30 countries in terms of international visitors (Wikipedia, 2018) and hence, is sufficiently significant for our analysis. Due to influences by the frugality of the Asian culture, we understand that many locals also tend to look at the lowest prices when planning their travel budget, including accommodation. Hence, we will also be extending our article to include an analysis and recommendation of the locations with best places to stay in Amsterdam.

1.2 Data Set

For our project, we will be making use of Airbnb's dataset obtained from the following source: www.insideairbnb.com/get-the-data.html for our analysis. Inside Airbnb is an alternative website to Airbnb which hosts publicly available data from Airbnb, and the data provided has been analyzed, cleaned and aggregated. Due to its ease of access and availability of the dataset, we have decided to make use of Inside Airbnb's data for Amsterdam in 2017 for our project rather than scrapping Airbnb's website for our required data.

1.3 Data Cleaning

We performed the following data cleaning techniques to make our data set more concise and comprehensive.

1.3.1 Extracting data

The datasets archived in Inside Airbnb were stored in monthly data files. Since we aim to analyze the full year data for Airbnb in Amsterdam, we have extracted the 12 data files from January to December 2017. By using the full year of data, it allows us to more effectively weed out any changes in price due to seasonality and hence, price seasonality is factored into our consideration when preparing our dataset for analysis.

1.3.2 Merging data

These data files were then merged using the Python function "pd.concat()" to obtain our main dataset for this project. This dataset will form our main source of data for further analysis into the trends and relationships of the identified variables that affects the prices of Airbnb listings in Amsterdam. After merging the dataset, we proceeded with the cleanup of data in preparation of our analysis.

1.3.3 Selecting Relevant Information

Amongst the initial 96 columns, we have identified 14 columns which we feel are the most crucial in affecting prices of Amsterdam's listings. The excess variables were removed for us to reach a simpler model with lower generalization error when the regression model is used to represent and predict prices for the different types of listings in Amsterdam.

The numerical variables "amenities", "bathrooms", "beds", "neighbourhood_cleansed" were chosen because we felt that any changes in these factors will directly affect the costs of providing the accommodation and hence, these factors will have a larger impact on minimum price in the listings.

On the other hand, variables such as “property_type”, “review_scores_communication”, “review_scores_location”, “review_scores_rating”, “review_scores_value” and “room_type” have an impact on the demand of Airbnb accommodation and hence, will also have an impact on how high the listings will be priced at.

Lastly, “calendar_last_scraped” and “id” are used as identifiers for each data point whereas “monthly_price” is retained for comparison against the daily prices listed. Together with “price”, these variables form our “cleaned” dataset.

1.3.4 Dropping NaN

As we understand and analyse the data, we realized that the dataset was “incomplete” as some cells in the dataset are empty and reflected as NaN values. In order to ensure a more complete analysis of the dataset, we decided to drop all rows that contain NaN values by filtering through each variable column and only retaining rows that contains data for all variables. This was done through the “data[data.notna()]” function. At the end of this step, another count of all NaN values that exists within each column was obtained using the “data.isna().sum()” function. Since all the rows with missing data have been dropped, the summary reflected that no columns contain missing data as we expected.

1.3.5 Data Transformation

To further simplify the dataset, data transformation is needed for a few of the existing variables to aid us in our analysis.

Review Scores

Firstly, there are 4 different review scores (communication, location, rating, and value) which are part of our dataset. To combine them into one metric for our analysis, we took the weighted average of all these review scores and created “review_score_average”, which will represent these 4 review factors for our analysis against price in the later part of our project. Subsequently, the initial 4 review scores are dropped to maintain a clean set of data.

Price

Secondly, the value of prices stored in our data contains the “$” sign, and some of the higher priced listings that cost more than $1,000 also contains commas. As such, we needed to clean the data columns containing price values. We begin by dropping the commas by applying the “.replace()” function to all the data values in our “price” and “monthly_price” columns. Next, since Python reads the price values from csv as strings and all price values do not contain the cents amount, we slice each data value to drop the dollar sign and cents amount, as well as convert them to integer values. The same was done for the "monthly_prices" column.

Amenities

Thirdly, the dataset also provides us with a list of all the amenities that each accommodation provides. For simplification, we decided to take the count of number of amenities provided under each listing. This was done by splitting different amenities by commas (since they are comma-separated) using the ".str.split(',')" function and then by applying the "len()" function to count the number of amenities for each listing. After these were done, we also dropped the "amenities" column to keep our dataset clean.

Date

Next, "calendar_last_scraped" is converted into the datetime format and the abbreviated month is extracted since that is the only relevant part of the calendar data that is required for filtering later on.

Data Duplication

Lastly, we have also identified that there is a possibility of data duplication since the same accommodation might be listed more than once throughout the year. This would lead to the over-representation of certain listings in our dataset since most of the variables of the listing such as "property_type" and "accommodates" would remain the same. In order to overcome this problem, we can firstly identify each unique accommodation by their "id". Following that, the average of all numerical variables for each listing will be taken after grouping them by their id. For categorical variables, since minimal changes, if not none, are expected throughout the year, the value from that accommodations latest listing will be taken. Finally, these newly transformed and cleaned values will form the final dataset for our analysis.

1.4 Assumptions made in Data Cleaning

In the process of data cleaning, there were a few assumptions made which are explained below.

1.4.1 Review Scores

We obtained the average of 4 types of review scores (communication, location, rating, and value) that were present in the data files. By using the average score for analysis, we are assuming that these 4 factors are equally important and hence, can influence list prices to the same extent. However, in reality, these might not be the case for all visitors who are looking for an accommodation and hence, they might not all influence prices to the same degree.

1.4.2 Amenities

The list of amenities provided for each listing was further simplified to the count of all amenities. This takes the assumption that all amenities provided costs the same. Hence, this would add the same amount to total costs and have the same effect of influencing price when scaled up or down linearly. However, it is obvious that this assumption is unlikely to hold true for the different kind of amenities that are provided. Nonetheless, this assumption was made in order to simplify our model and allow direct comparability against price.

1.4.3 Categorical Variables

Repeated categorical variable data values was taken to be the latest value for our model analysis. This carries the assumption that all the categorical variables under the same id listing did not change across the different months in which they were listed. However, this assumption might not hold true in reality. For example, the room type listed might change if the property were to undergo renovation.

2.0 Modelling Methodologies

In the next part of our project, we continue by running our cleaned dataset through three different regression models, best subset, forward stepwise, and backward stepwise, to obtain the best set of predictors for our final analysis of the price-influencing variables.

2.1 Backward Stepwise Model

First, we ran the cleaned dataset through the backward stepwise model for linear regression. This was done by taking 'price' as the dependent y-variable and running it against all other variables except for 'id'. For all our models, we will be using the Akaike Information Criterion (AIC) as our metric of finding the best set of predictors to fit the regression model for price. The backwards stepwise model starts with the full set of predictors and at each subsequent iteration with a set of k predictors (where k = p, p-1, p-2, ..., 1), it will iteratively drop each predictor and compare across all possible predictor sets within the same set of k variables to return the set of predictors with the lowest AIC since AIC measures the badness of fit of the set of predictors for the linear regression model. Subsequently, the AIC of the best set of predictors across predictor set of all sizes are compared to return a final best set of predictors for price under the backward stepwise model.

2.2 Forward Stepwise Model

Next, we also ran our dataset through the forward stepwise model for comparison purposes. The forward stepwise model begins with a null set of predictors. Subsequently, at each stage of k predictors (where k = 0, 1, ..., p-1), the model iteratively selects an additional predictor amongst all the remaining predictors to add to the current set of k-1 best predictors. All possible combinations of predictors with an additional predictor are then run through the regression model compared using the AIC. The set of predictors with the best (lowest) AIC value is then chosen for addition of subsequent predictors. At the end, the best set of predictors with the lowest AIC value amongst predictors sets of different sizes are selected to be the best predictor set for this model.

2.3 Best subset selection

The final model that we ran our dataset through was the best subset model. Under the best subset model, the same set of data and variables were also run through the model. The model uses the "itertools.combinations()" function to form all possible combinations of the predictors set that was passed through it. At each stage of k predictors (where k = 0, 1, ..., p-1), all possible combinations of predictor sets of size k are passed through the linear regression model and the best is obtained based on the model that returns the lowest AIC. Similar to the other two models, the lowest AIC is compared across models of different predictor set sizes and model with lowest AIC gives the best model from the linear regression.

2.4 Results from all Linear Regression Models

After running through all 3 models, all 3 outcomes were the same and we determined that the best set of predictors for predicting accommodation prices was the full set of predictors containing all the identified variables in our regression model. This set of predictors consists of "mean_reviews", "neighbourhood", "property_type", "mean_bathrooms", "mean_accomodates", "room_type", "mean_beds", "mean_amenities" and "mean_monthly_price".

2.5 Identification and Treatment of Multicollinearity

Aside from running our dataset through the linear regression model, there is also a need for us to verify any interactions between the different independent variables that might lead to multi-collinearity. To verify these interactions, we used the "corr()" function to run each variable against other variables. For the purposes of this project, we have decided on using 0.7 as the arbitrary cut-off for the acceptable correlation threshold. Following that rule, we identified that the variables "mean_beds" and "mean_accommodates" have potential interactions leading to multicollinearity since the correlation coefficient between these 2 variables is 0.839, which is much higher than the threshold value we have defined. Hence, there is a need for us to remove one of these variables to address the problem of multicollinearity.

In this case, we decided to remove "mean_beds" as we feel that "mean_accommodates" is a more overarching variable that would have accounted for "mean_beds" as it directly accounts for the number of people that can stay in the room. Also, planning for the number of people and accommodation can hold is usually one of the factors we consider first before thinking about the number of beds and any bed arrangements, thus signalling the importance of "mean_accommodates" over "mean_beds". Therefore, we will be removing "mean_beds" as a variable to address the issue of multicollinearity in our linear regression model.

2.6 Best subset selection re-run

After dropping the "mean_beds" variable, we then re-ran the further cleaned set of data through the best subset model once again. Similarly, the full set of predictors forms the best predictor set for our regression model. This set of predictors comprises: "mean_monthly_price", "mean_amenities", “mean_accomodates”, “mean_bathrooms”, “mean_reviews”, “neighbourhood”, “property_type” and “room_type”. By further analysis on the summary results of the best regression model obtained, we also found that at 5% significance level, “room_type”, “mean_monthly_price”, “mean_amenities”, “mean_accommodates”, “mean_bathrooms”, “mean_reviews”, “property_type” (more specifically for categories: T.Others, T,House, T.Boat), and “neighbourhood” (for categories: T.Centrum-West and T.Centrum-Oost), are the significant predictors of price.

3.0 Solution to our Business Case

In this section, we will be explaining how we derive the top 3 neighbourhoods for travellers to rent via Airbnb. Before we proceeded with providing insights and solutions for our business case, there is an additional step we implemented to touch up our data for analysis. Considering that the analysis in the subsequent sections will be on a price per unit variable basis, we will need to drop all zero values in the numerical variables to be analyzed as they will give an infinity value on the same basis. Since the only numerical variable with zero values in our current dataset is “mean_bathrooms”, we trimmed our dataset by dropping all rows of data containing values 0 for “mean_bathrooms” by replacing them with before proceeding with our analysis.

3.1 Methodology

We will be focusing on the categorical variable “neighbourhood” only and all numerical predictors except for “mean_monthly_price”. The reason why the other categorical variables are dropped is because our business problem is only concerned with the neighbourhood in which each Airbnb is in. As such, the other categorical variables can be seen as dummy variables. “Mean_monthly_price” is not used because it not only has a high correlation coefficient of 0.75777 with “price”, it does not help to answer our business question by comparing it against “price”.

Following the results of the regression model, we will group the data by neighbourhood and find the mean price per “mean_amenities”, “mean_accomodates” and “mean_bathrooms” within each neighbourhood. We will then rank the neighbourhoods in ascending order based on the price per numerical predictor.

As for “mean_reviews”, we only took into account listings with a mean review score of above 9.4 out of 10. Initially, we tried counting reviews with full marks but we realized that this was not a comprehensive number because there were listings in some neighbourhoods that did not receive a 10 out of 10 at all. Hence, through trial and error, we found that 9.4 was the highest number that allowed us to compare all 22 neighbourhoods. Hence, the neighbourhoods will be ranked by the number of listings with review scores above 9.4. The neighbourhood that is ranked first will have the greatest number of listings with a score above 9.4 while the neighbourhood that is ranked last will have the least count of review scores above 9.4.

Each rank will be assigned a value, with the first rank being assigned 22 points and the last rank being assigned 1 point. A total of 4 rankings will be done and ranking points will be assigned for each of the predictors which are numerical variables. The rank points are cumulative, and the neighbourhood which has the highest score will emerge as the overall best neighbourhood to stay.

3.2 Ranking by Price per Amenities

In this section, we will compare neighbourhoods according to the mean value of price per amenity offered.

3.2.1 Assumptions

We assumed that each amenity offered has the same cost, quality and quantity. This justifies our standardised basis of comparison.

3.2.2 Data Visualisation

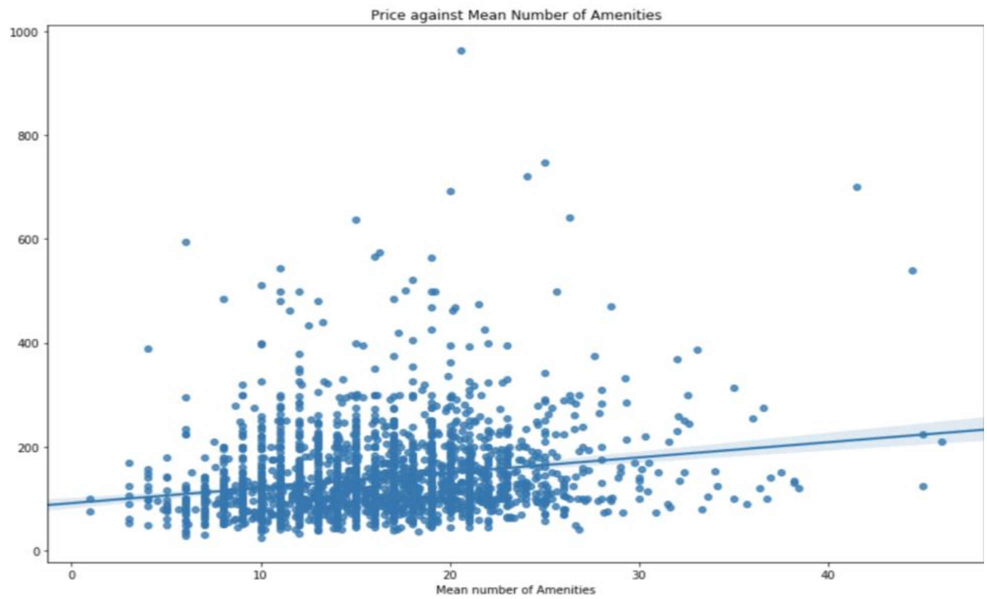

Figure 1: Regression plot of price against mean number of amenities

From the graph, the mean number of amenities for each property falls around the range of 10 to 20, regardless of the rental price of the property (refer to Figure 1). There are a few outliers that have more than 40 amenities, but this does not show a distinct relationship between price and amenities.

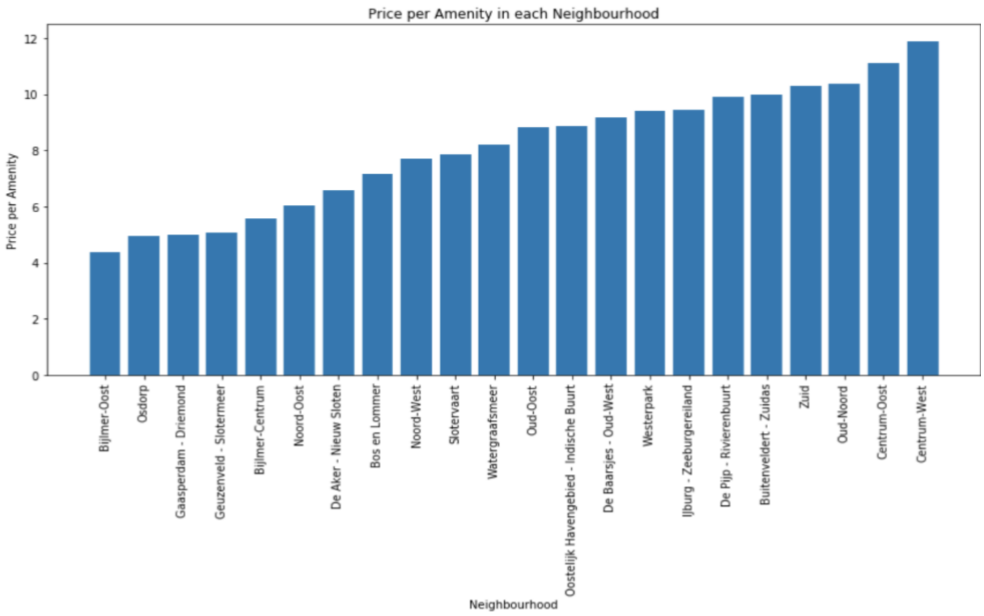

Figure 2: Bar graph of each neighbourhood and its price per amenity

From this bar graph, we will be able to see clearly which are the neighbourhoods with the lowest price per amenity (refer to Figure 2). In other words, neighbourhoods such as 'Bijlmer – Oost' and 'Osdorp' are most value for money when considering amenities only.

3.2.3 Results

Rank	Neighbourhood	Price per Amenity	Ranking Points	Accumulated Points
1	Bijlmer – Oost	$4.37	22	22
2	Osdorp	$4.95	21	21
3	Gaasperdam – Driemond	$4.98	20	20
4	Geuzenveld – Slotermeer	$5.07	19	19
5	Bijlmer – Centrum	$5.55	18	18

Figure 3: Table of top 5 nieghbourhoods ranked according to their corresponding price charged per amenity offered

The neighbourhoods have been ranked in ascending order of Price per Amenity and have been assigned ranking points based on their ranks. The top 5 neighbourhoods ranked by Price per Amenity are Bijlmer – Oost, Osdorp, Gaasperdam – Driemond, Geuzenveld – Slotermeer and Bijlmer – Centrum (refer to Figure 3). The table also summarises the respective ranking points assigned to each neighbourhood based on their rankings and their accumulated points.

3.3 Ranking by Price per Accommodates

In this section, we rank neighbourhoods according to the mean price charged per person the property is able to accommodate.

3.3.1 Assumptions

We assume the incremental space per person is the same. This means that no matter how big or small the property, we only take into account the number of people the property accommodates as presented in the listing.

3.2.2 Data Visualisation

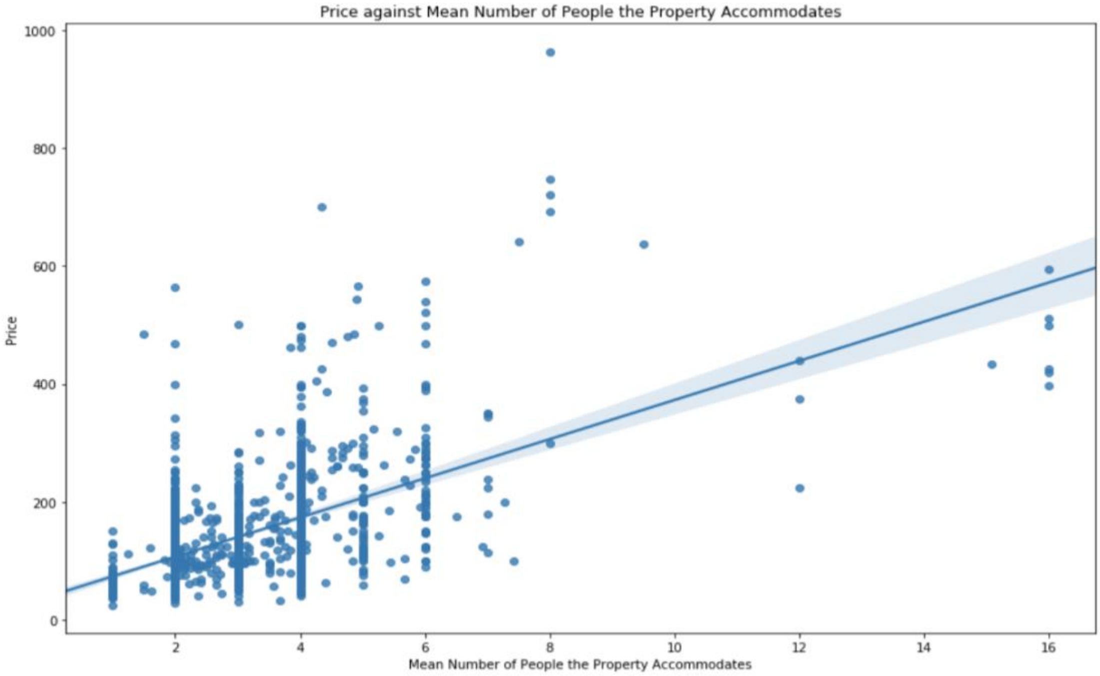

Figure 4: Regression plot of price against mean number of people the property accommodates

From the graph, we can understand how price varies with the mean number of people the property can accommodate (refer to Figure 4). We can also tell that the mean number of people most properties can accommodate hovers around 2 to 6 people, regardless of the rental price of the property. It might seem intuitive that the more people a property can accommodate, the more expensive the price of that listing would be. However, it is interesting to note that the highest rental price on the graph does not actually correspond with the highest number of people that the property can accommodate.

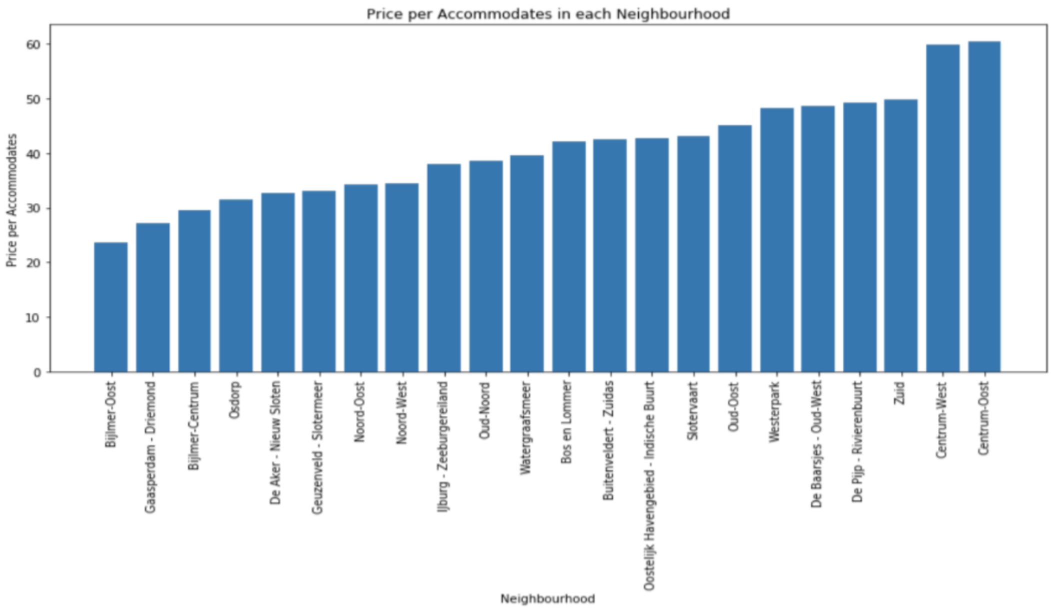

Figure 5: Bar graph of each neighbourhood and its price per person the property can accommodate

From this bar graph, we will be able to see clearly which are the neighbourhoods with the lowest price per the number of people it can accommodate (refer to Figure 5). In other words, neighbourhoods such as 'Bijlmer – Oost' and 'Gaasperdam – Driemond' are most value for money when we only look at the number of people the property can accommodate.

3.3.3 Results

Rank	Neighbourhood	Price per Accommodates	Ranking Points	Accumulated Points
1	Bijlmer – Oost	$23.48	22	44
2	Gaasperdam – Driemond	$27.12	21	41
3	Bijlmer – Centrum	$29.44	20	38
4	Osdorp	$31.51	19	40
5	De Aker– Nieuw Sloten	$32.67	18	34

Figure 6: Table of top 5 neighbourhoods ranked according to their corresponding price charged per person the property can accommodate

The neighbourhoods have been ranked in ascending order of Price per Accommodates and have been assigned ranking points based on their ranks. The top 5 neighbourhoods ranked by Price per Accommodates are Bijlmer – Oost, Gaasperdam – Driemond, Bijlmer – Centrum, Osdorp and De

Aker– Nieuw Sloten (refer to Figure 6). The table also summarises the respective ranking points assigned to each neighbourhood based on their rankings and their accumulated points.

3.4 Ranking by Price per Bathroom

In this section, we compare neighbourhoods according to the mean value of price per bathroom available.

3.4.1 Assumptions

We assumed that every bathroom is identical and have the same facilities such as a shower area, washbasin and toilet bowl.

3.4.2 Data Visualisation

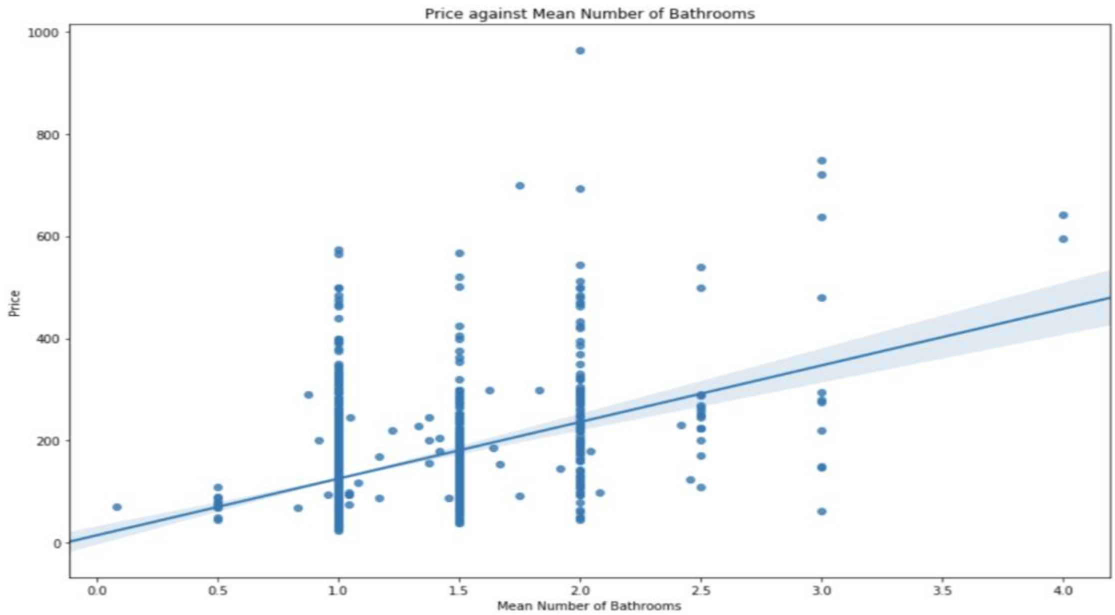

Figure 7: Regression plot of price against mean number of bathrooms

From the graph above, we can tell that the mean number of bathrooms clusters around 1 to 2 bathrooms, regardless of the rental price of the property (refer to Figure 7).

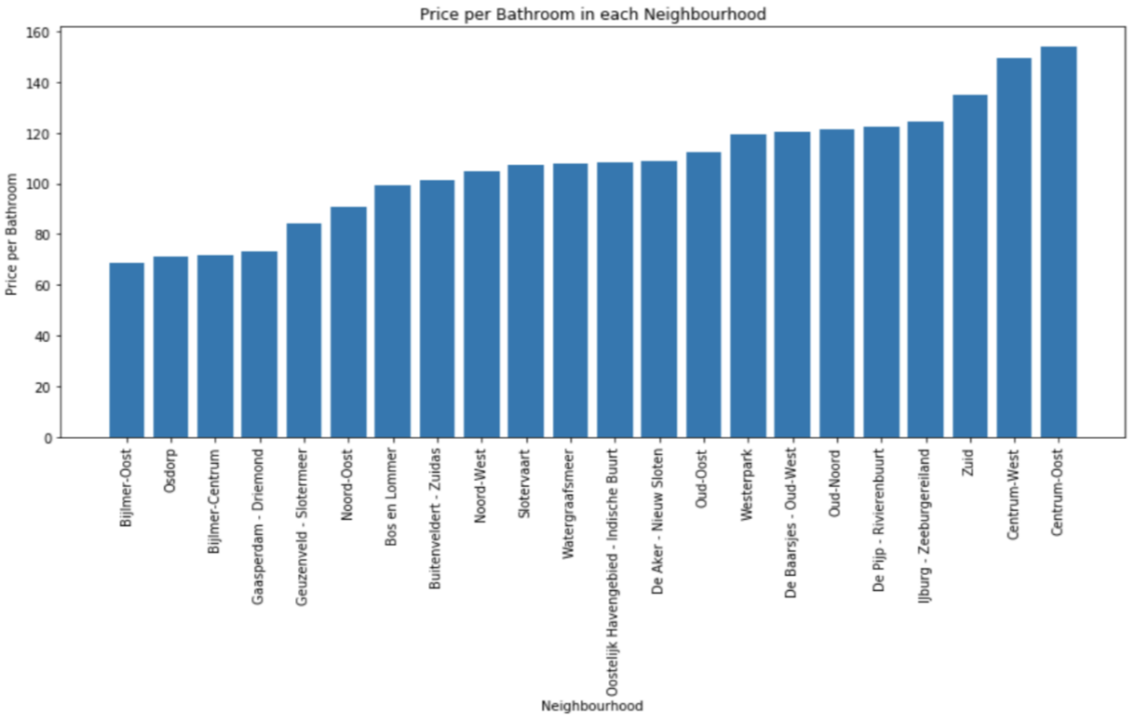

Figure 8: Bar graph of each neighbourhood and its price per bathroom

From this bar graph, we will be able to see clearly which are the neighbourhoods with the lowest price per bathroom (refer to Figure 8). In other words, neighbourhoods such as 'Bijlmer – Oost' and 'Osdorp' are most value for money when we only consider the number of bathrooms each accommodation provides.

3.4.3 Results

Rank	Neighbourhood	Price per Bathroom	Ranking Points	Accumulated Points
1	Bijlmer – Oost	$68.74	22	66
2	Osdorp	$71.26	21	61
3	Bijlmer – Centrum	$71.88	20	58
4	Gaasperdam – Driemond	$72.97	19	60
5	Geuzenveld – Slotermeer	$84.12	18	54

Figure 9: Table of top 5 neighbourhoods ranked according to their corresponding price charged per bathroom

The neighbourhoods have been ranked in ascending order of Price per Bathroom and have been assigned ranking points based on their ranks. The top 5 neighbourhoods ranked by Price per Bathroom are Bijlmer – Oost, Osdorp, Bijlmer – Centrum, Gaasperdam – Driemond and Geuzenveld

– Slotermeer (refer to Figure 9). The table also summarises the respective ranking points assigned to each neighbourhood based on their rankings and their accumulated points.

3.5 Ranking by Review Scores

In this section, we compare neighbourhoods according to their mean review scores.

3.5.1 Assumptions

We have aggregated the average of the 4 review scores: “review_scores_communication”, “review_scores_location”, “review_scores_rating”, “review_scores_value” into a single variable “mean_review_score”, which we will use to represent the overall review score. Here, we interpreted the overall review score as the level of satisfaction that the property brings when staying here, which is part of our assumption for the section.

3.5.2 Data Visualisation

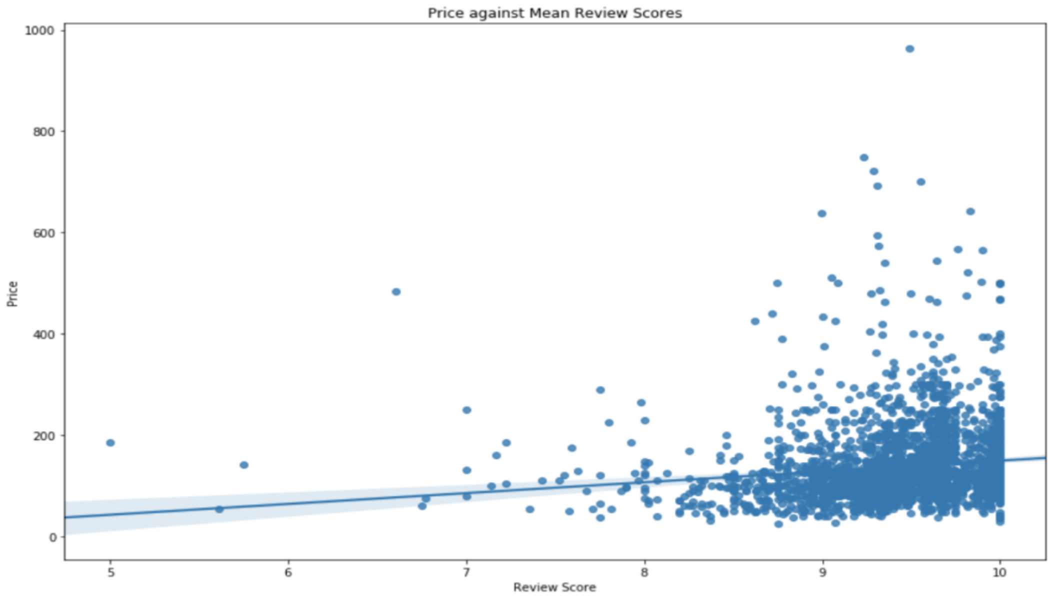

Figure 10: Regression plot of price against mean number of bathrooms

From the graph, we are able to understand that regardless of the price of the listing, the review clusters around the 9 - 10 range with a few outliers (refer to Figure 10).

3.5.3 Results

Rank	Neighbourhood	Review Count	Ranking Points	Accumulated Points
1	De Baarsjes – Oud West	290	22	9+5+7+22 = 43
2	Centrum – West	289	21	1+2+2+21 = 26
3	Centrum – Oost	219	20	2+1+1+20 = 24
4	De Pijp – Rivierenbuurt	216	19	6+4+5+19 = 34
5	Zuid	118	18	4+3+3+18 = 28

Figure 11: Table of top 5 neighbourhoods ranked according to their corresponding review score

Since a higher review score is generally more desirable, the neighbourhoods were ranked in descending order based on the number of listings in each neighbourhood which received a review score of 9.4 or more. Accordingly, each neighbourhood was assigned ranking points based on their ranks. The top 5 neighbourhoods ranked by review scores are De Baarsjes – Oud West, Centrum – West, Centrum – Oost, De Pijp – Rivierenbuurt and Zuid (refer to Figure 11). The table also summarises the respective ranking points assigned to each neighbourhood based on their rankings and their accumulated points.

From the table, we observe that neighbourhoods in the top 5 rankings for overall review score are totally different from the previous 3 rankings. One possible explanation behind this observation could be that expensive accommodation options tend to leave travellers with better experiences due to the convenience, amount of utilities and quality of service it provides. Hence, their overall quality of stay will also naturally be higher, resulting in a higher overall review score. This also indicates that there might be a negative correlation between the overall review score and price of accommodation listings.

3.6 Overall Rankings

Rank	Neighbourhood	Points for Amenity Ranking	Points for Accommodates Ranking	Points for Bathroom Ranking	Points for Review Ranking	Total Points
1	Bijlmer – Oost	22	22	22	4	70
2	Osdorp	19	21	21	5	66
3	Gaasperdam – Driemond	21	20	19	1	61

Figure 12: Table of top 3 neighbourhoods in Amsterdam to rent an Airbnb

By combining the 4 ranking results from above, we were able to aggregate the total number of ranking points for each neighbourhood in Amsterdam. This allows us to come up with an overall ranking across all the neighbourhoods (refer to Figure 12). From the table above, we can see that Bijlmer - Oost, Osdorp and Gaasperdam - Driemond are some of the best neighbourhoods to rent an Airbnb when travelling to Amsterdam.

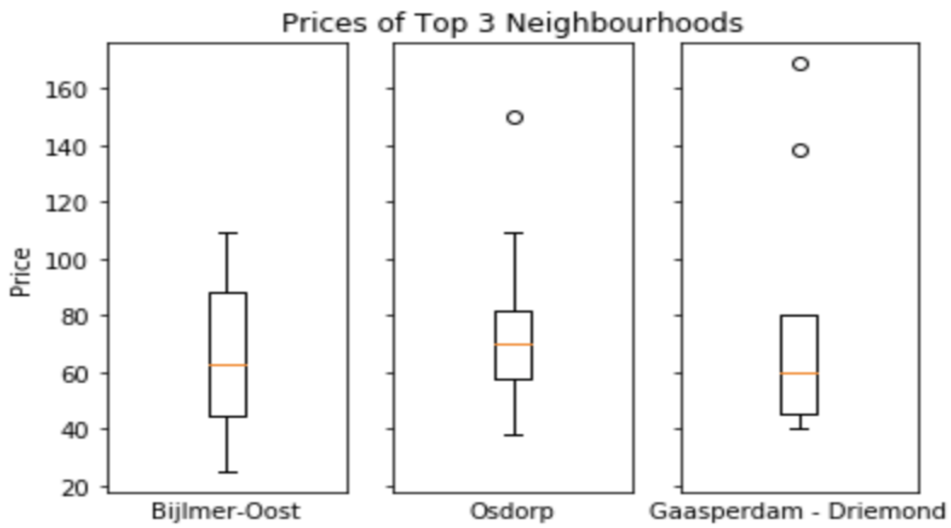

Figure 13: Box Plot of Prices of listings in Top 3 Neighbourhoods

Although the top 3 neighbourhoods provide the best value for money in terms of Airbnb accommodation, there are still variations in individual listing prices in each of these top neighbourhoods identified (refer to Figure 13). For instance, in Bijlmer - Oost, the median price is around $60 but it ranges from $25 to $110. Similarly, in Osdorp, the median price is close to $70, ranging from $40 to $110 with an outlier priced at $155. In Gaasperdam - Driemond, the median price is $60 and it has a lower limit of $40 and 2 outliers priced at $140 and $170. Hence, while these are the top neighbourhoods identified from our analysis, we must also take note within each neighbourhood, there are different accommodation options which are priced differently according to the difference in variables that are present.

4.0 Conclusion

In our DAO Magazine, we will recommend readers to rent listings from the following Amsterdam neighbourhoods: Bijlmer - Oost, Osdorp and Gaasperdam - Driemond. These locations give the best value for money as they holistically offer the lowest prices per amenity, bathroom, accommodation capacity. However, according to results of our analysis, there might also be a slightly higher risk that travellers visiting these areas might find accommodation that are less satisfactory than they desire or as according to what is advertised online due to the lack of sufficiently higher review ratings for listings in these areas.

As such, travellers who are looking at getting the best accommodation experiences and quality when they visit Amsterdam, and have less restrictions on their budget, should also consider other areas such as De Baarsjes – Oud West, Centrum - West, and Centrum - Oost, which are areas in Amsterdam with the greatest number of high average review scores as highlighted in Figure 11.

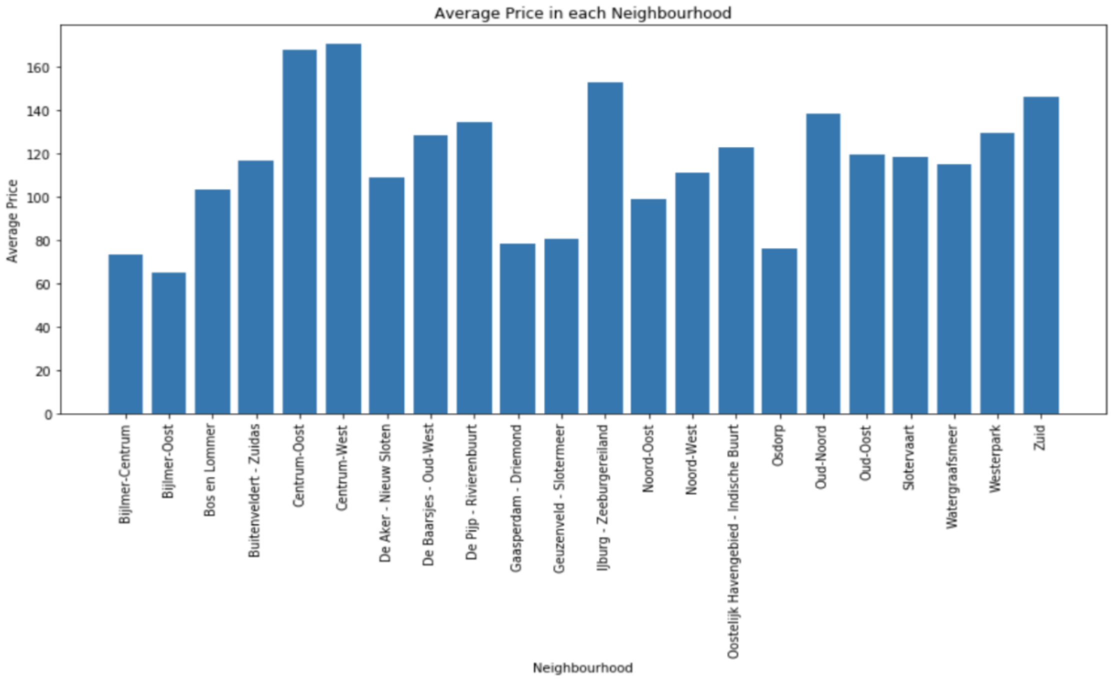

Figure 14: Bar chart of the average prices of listings in each neighbourhood

Furthermore, out of De Baarsjes – Oud West, Centrum - West, and Centrum - Oost (top 3 neighbourhoods in the ranking by review score), Centrum - West and Centrum - Oost have the highest average listing price (refer to Figure 14). On the other hand, De Baarsjes - Oud West has a high review score and a less expensive average listing price. Hence, the neighbourhood of De Baarsjes - Oud West will be one of the strongly recommended neighbourhoods in our article.

5.0 References

Lawton, Georgina (2017, March 06). 6 Common Travel Problems That Will Affect You At Some Point. Retrieved November 21, 2018, from https://www.liligo.com/travel-edition/6-common-travel-problems-that-will-affect-you-at-some-point-and-how-to-keep-going-12764.html

Medium (2017, January 17). Top 10 Airbnb Competitors and Alternatives 2018. Retrieved November 21, 2018, from https://medium.com/@airbnb_review/top-8-airbnb-competitors-not-who-your-might-think-417d6b3514e0

Momentum (2016, April). The surprising success of Airbnb. Retrieved November 21, 2018, from https://www.business.uq.edu.au/momentum/surprising-success-airbnb

ValuePenguin (n.d.). Average Cost of a Vacation. Retrieved November 21, 2018, from https://www.valuepenguin.com/average-cost-vacation

Wikipedia (2018). List of cities by international visitors. Retrieved November 21, 2018, from https://en.wikipedia.org/wiki/List_of_cities_by_international_visitors