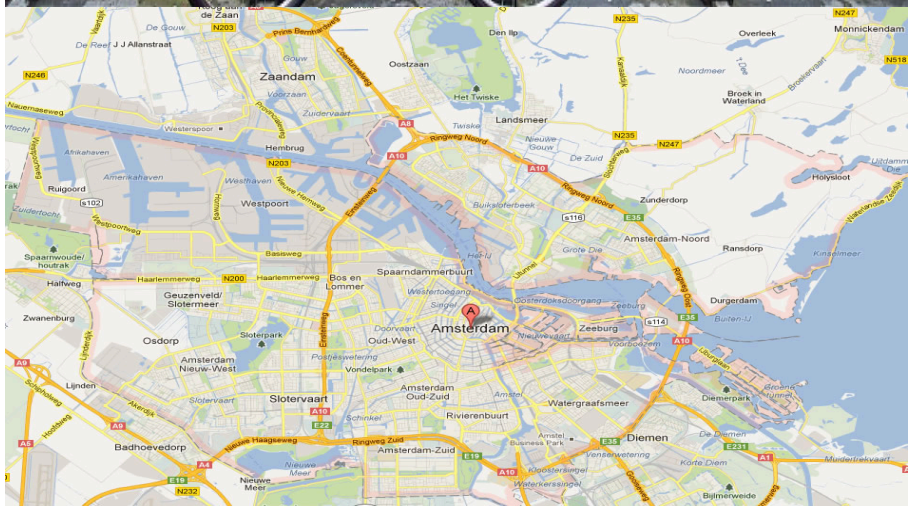

AirBnB Amsterdam

Market Performance Analysis

By Wendy Wong

1.The Business Problem

Identify the Problem

- What **key factors** will significantly affect performance and demand to host AirBnB in the local market in Amsterdam?

- The success of performance will depend on a service offering of a property type 'apartment', 'room type', 'neighbourhood' ,' host revenue', 'bed type' of real bed to compete with the hotel market. The listings with the highest number of positive reviews preferred private rooms and also entire houses or apartments that is centrally located to the city of Amsterdam.

Hypothesise goals

- **Hypothesis 1:** We propose that the following independent factors **'Property_type', 'review_scores_rating ','zipcode', ' city', 'bed_type', 'minimum_nights' and number of reviews'** will significantly affect **price** in Amsterdam for this specific market

Although multivariate statistical analysis (Principal component analysis or cluster analysis) was not conducted for this study, the insights from our data analysis concluded that the factors that affect price are dependent on the variables 'number of reviews' , 'property type', 'neighbourhood', 'bed type' and 'room type'.

- **Hypothesis 2**: To invest into AirBnB at a profit, we believe that the key factors that affect price are also correlated to profit (i.e. revenue) and provides an opportunity to invest

- **Hypothesis 3 :** We determine that the optimal price points to charge various **property types** and make a profit, are also related to the price points used to compete in the hotel market (i.e. house vs. apartment).

Further questions for identifying the correct data set

- What is the demand and **price point** to operate at profit for AirBnB in Amsterdam? The property types of house, apartment, boat or bed and breakfast and they all offer a room type of ' private room' or ' entire house/aprt'.
- What **property types** receive the most positive reviews? Apartments
- What **neighbourhoods** host the most listings? Explained in the data analysis
- How much **revenue** successful hosts generate? Explained in the data analysis
- Which **neighbourhood** should AirBnB invest in Amsterdam? The suburbs that highlighted the Top 10 hosts by revenue
- What **factors** are likely indicators of success for AirBnB to invest into Amsterdam at a profit? Invest into the Top 10 neighbourhood, a host revenue greater than $2000, has to offer a real bed (comparable to a hotel), The property type of apartment received the most positive reviews and number of reviews. Clients valued their privacy and room types such as 'private room' and 'entire house/apt' received the greatest number of reviews.

2. Identify the Dataset & Assumptions

Obtain the data

- Publicly available data was scraped from the website Inside AirBnB courtesy of Murray Cox a digital storyteller on the 3rd of September 2015.

Identify right data set

- Under the 'Get data' section of the *Inside AirBnB* website, the Amsterdam dataset was downloaded from a Microsoft Excel file (xlsx.) format
- **7834 observations ('n'), 33 variables (Qualitative and Numeric) and,**
- **Missing values (NA) - yes**

8547 for review scores (21%>15% tolerance)

review_scores_cleanliness_CLEAN	review_scores_checkin_CLEAN	review_scores_communication_CLEAN	review_scores_location_CLEAN	review_scores_value-CLEAN
Numeric	Numeric	Numeric	Numeric	Numeric
9	9	9	9	9
1708	1708	1711	1709	1711

Assumptions

- Availability of listings are based on a calendar year of 365 days
- The listing for 'Booked night' and 'Unavailable night' are counted once and are considered to be interchangeable terms that mean 'Unavailable'.
- Neighbourhood names for each listing are compiled by comparing the listings geographic coordinates
- Shared rooms and private rooms were excluded as a 'room_type' to compare the analysis of AirBnB for the hotel market
- AirBnB hosts occassionally live in the homes that they rent out
- Neighbourhood names for each listing are compiled by comparing the listings geographic coordinates (longitude and latitude)
- We used the **AirBnB New York Occupancy model** (i.e. "San Francisco Model") to estimate how often an AirBnB listing is being rented out and approximate the listings income.
- The Occupancy Model determines a 'Review rate of 50%' to convert reviews to estimated bookings
- An 'Average length of stay' is configured to each city and multiplied by the estimated bookings for each listing over a period
- Where there was no value provided for the 'Average length of stay' the missing value was substituted with 3 nights per booking
- Occupancy rate is capped at 70% for a highly occupied hotel
- Average daily rate or achieved rate is close to the requested rate
- When comparing AirBnB listings with traditional hotel rooms the following were excluded : tree houses, boats. Units that were included in the analysis for acceptable 'property_type' had to offer a **real bed** (not a couch or air mattress).

3. Methodologies for Cleaning Data

Cleaning methods used to remove erroneous data

- Retain the original copy of the raw data on a separate worksheet, create a second worksheet for cleaning data
- Freeze panes, examine column data with auto-filters (e.g. missing values), select all data and remove duplicates (unique values)
- Conditional formatting to highlight specific text or duplicate values
- Find and replace text – used the substitute function to alter text data
- UPPER and LOWER functions - Standardise and format text
- TRIM, LEN, TRUNC and CLEAN functions – Remove spaces, remove decimal places and removed non-print characters
- Delete irrelevant columns that are not required in the analysis
- ISBLANK - Used to check for any blank cell values in the data range.
- Absolute and relative values – Use these values to lock a particular cell reference for using formulas or locking a column within a selected data range.
- Inspect any unusual outliers and values that may be erroneous due to mistyping into the cell reference.

Missing Values

- Missing values – impute the missing values by taking an average of the known values
- **Treatment of Null values, outliers and sensitivities -** we did not remove the outliers or null values but retained all the data in the sample size to provide richness of insights. We measured the percentage of blanks if over 15% we imputed the values by taking an average of existing values.

Prepare data: Create necessary derived columns from the data (new data)

- Format, clean and combine data (text) – We joined text using CONCATENATE
- Create conditionals - IF, COUNTIFS, SUMIFS with AND and OR – to test a range of multiple criteria to return a value if true, otherwise returns a criteria that is FALSE
- LEFT, MID, RIGHT – save efficiency of typing values in columns by using these functions
- Auto-fill values in columns by double- clicking on the corner of a cell with formula (with + symbol) and auto fill column
- VLOOKUPS and HLOOIPS - to create new columns and rows and lookup existing values from the left most column or row
- INDEX, MATCH- to not be restricted by the slow processing time of VLOOKUPS and instead use INDEX and MATCH to return values within a data range.
- Advanced Filters – will save time instead of copying a pasting the columns and specifies a destination location to populate the new columns

5. Understanding Data and Data Limitations

Understanding of clean data

- To understand the data we read articles related to successful listings of AirBnB in related markets such as New York, the AirBnB website to obtain background information
- We performed **exploratory analysis** after data was cleansed
- Data verification –auto filters (check for blank values and unreadable text), data validation, formatting non-readable text (find and replace and text functions e.g. LOWER, CLEAN, SUSTITUTE)

Data limitations

- AirBnB only shows unit listings and does not make data available for actual rooms sold, occupancy or revenue per available room
- Data is available only for specific points in time because the owners can add or take down listings, hence the actual monthly supply of AirBnB units will fluctuate.
- Inventory is not a perfect representation of supply because some AirBnB inventory is offered by rental companies that offer multiple listings which are only offered online when they are available
- **VLOOKUPS** – inefficient looking at left most column and cannot process large datasets, used INDEX and MATCH
- **>15% tolerance level** of missing values : substitute with imputed average of existing values of review scores
- **Histogram**: requires that outliers be included in the dataset and different number of bins or intervals will give you a different visualisation and different story of the data (can be misleading)

5.Analysis of Data & Presenting the Results

Analyze the Data

- Identify trends and **outliers – Conditional Formatting (data bar and heat maps)** used to identify the Top 10 and best performing neighbouthoods, scatterplots will be used to identify outliers, line graphs will be used to identify the movement of trends
- Apply descriptive and inferential statistics
- Visualise and transform data :
- **Bar chart (view a few categories and identify specific target markets)**
- **Histogram –** used to present different stories and accepts only quantitative variables
- Pie charts

Presenting the Results

- Summarise findings (recommendations) with narrative, storytelling techniques
- **Metrics** and **Insights** to include in the **final deliverable** and **data handling summary includes:**
- 1.**Count, Min, Max, Average** from a possible dashboard that is refreshed overnight and automated
- 2.The demand and price point to achieve a profit
- 3.The neighbourhoods that host the most listings
- 4.Which neighbourhood that AirBnB should target and concentrate their integrated and digital marketing campaigns
- 5. Trend lines depicting the factors that are indicators of success to invest at a profit over time
- **Demand and price point** – Multiple Pivot tables will be toggled to select the information required to answer the business problem with different (filters or views). The price will be summarized by values using in built SUM and COUNT functions and shows values in %.

Next Steps

- Identify follow up problems and questions for future analysis – after particular neighbourhoods are identified and the market is tested, the evaluation of profits and success will be measured against other property types and against the hotel market overtime.