National College of Ireland

Data Warehousing and Business Intelligence Project

on

Hosting on Airbnb

Shareen Shahana Marottikal Khader
x18141404

https://youtu.be/yd9LiglDAY4

MSc/PGDip Data Analytics – 2019/20

Submitted to: Sean Heeney

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing

Student Name:	Shareen Shahana Marottikal Khader
Student ID:	x18141404
Programme:	MSc Data Analytics
Year:	2019/20
Module:	Data Warehousing and Business Intelligence
Lecturer:	Sean Heeney
Submission Due Date:	12/04/2019
Project Title:	Hosting on Airbnb

I hereby certify that the information contained in this (my submission) is information pertaining to my own individual work that I conducted for this project. All information other than my own contribution is fully and appropriately referenced and listed in the relevant bibliography section. I assert that I have not referred to any work(s) other than those listed. I also include my TurnItIn report with this submission.

ALL materials used must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is an act of plagiarism and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	April 12, 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table 1: Mark sheet – do not edit

Criteria	Mark Awarded	Comment(s)
Objectives	of 5	
Related Work	of 10	
Data	of 25	
ETL	of 20	
Application	of 30	
Video	of 10	
Presentation	of 10	
Total	of 100	

Project Check List

This section capture the core requirements that the project entails represented as a check list for convenience.

- ☒ Used LaTeX template
- ☒ Three Business Requirements listed in introduction
- ☒ At least one structured data source
- ☒ At least one unstructured data source
- ☒ At least three sources of data
- ☒ Described all sources of data
- ☒ All sources of data are less than one year old
- ☒ Inserted and discussed star schema
- ☒ Completed logical data map
- ☒ Discussed the high level ETL strategy
- ☒ Provided 3 BI queries
- ☒ Detailed the sources of data used in each query
- ☒ Discussed the implications of results in each query
- ☒ Reviewed at least 5-10 appropriate papers on topic of your DWBI project

Contents

Hosting on Airbnb

Shareen Shahana Marottikal Khader
x18141404

April 12, 2019

Abstract

Airbnb is one of the most popular online lodging reservation service Founded in 2008. Airbnb is running successfully in more than 75,000 cities worldwide, with upward of 5 million listings. This Report describes the research analysis carried out about hosting on Airbnb in 2018 across six cities applying advanced analysis tactics. While it is easy to become a host in Airbnb but being a super host is difficult. In this paper, leading Airbnb destinations such as London, Paris, Rome, Sydney, New York, and Rio de Janeiro are observed and concluded how a host can achieve success and return on investment through the online renting Industry. The Research observed that Airbnb overnight mean price is less compared to the Hotel and also how positive sentiments affect host registration trend Over the years. The Report also analyzed factors such as cancellation policy, host response rate, location, property type, and host identity verification were played a major role upgrading status of a host into a super host.

1 Introduction

Recent years we have seen a rapid increase in the number online platform that connects people in need of a product temporarily with people who have that product but are not using it. Some of the famous sharing economy platforms are uber, Airbnb, my taxi.

Airbnb, a peer-to-peer social website anyone who is ready to share a room or a property can register as a host. Similarly, Customers/Users can search for the listings online and book the desired room offered they like and in the process, both user and host are benefited. Airbnb has had enormous growth since its inception and has been impacting conventional hospitality businesses like Hotels. Just like when Ubers entrance made general taxi services market share decrease by about 25%. Similarly, the Hotel is the conventional business before the arrival of Airbnb. There were findings (Georgios Zervas 2015)(Ytreberg 2016) saying that a 10% increase in Airbnb supply decreases hotel revenue by 0.3%.

It is relatively very easy to become host in Airbnb while succeeding in business is very difficult. In this Report, we will be discussing how a host can be successful in their business. Following queries can serve the potential new Airbnb Hosts to make an informed decision.

(Req-1) Do people prefer Airbnb over Hotel?

(Req-2) Is it the best Time to register as host?

(Req-3) What makes a Host as a Super Host?

Source	Type	Brief Summary
insideairbnb.com	Structured	It provides Airbnb Listing Details
Statista	Un-Structured	It provides which cities are particularly popular on Airbnb
trivago.com	Un-Structured	It provides average overnight accommodation prices for a standard double room in popular cities
The Guardian	Structured	It provides sentiments of people over the years

Table 2: Summary of sources of data used in the project

2 Data Sources

Data Set used for the research is being collected from the following sources.

2.1 Source 1: insideairbnb.com

Inside Airbnb is a data portal for Airbnb, it allows you to explore how Airbnb is being used in cities around the world. Link of the data source is `http://insideairbnb.com/get-the-data.html`

Six major cities such as Paris, London, Rome, Sydney, Rio de Janeiro, and NewYork are chosen for this paper. Structured Data(CSV) is downloaded in the form of listings.csv.gz for all major cities listed above. Data Contains 95 columns however 15 columns that were cleaned are used in this paper.

Host Details, List Details, Review Details are major things collected for each listing.The total number of Airbnb listing for each city is given below.

Number of Listing	Cities
50221	NewYork
62685	Paris
33830	Sydney
75507	London
37246	Rio De Janeiro
29694	Rome

2.2 Source 2: Statista

Statista is another data source used in this paper. The Data source is in the form of image and it is an unstructured data. This Dataset contains twelve cities where Airbnb is most popular along with their active rentals, average daily rate. Link to Data Source is `https://www.statista.com/chart/14986/active-airbnb-listings-in-major-cities/`

2.3 Source 3: trivago.com

Trivago is a company specializing in an online tool to search for a hotel, lodging services. It has an average accommodation price for a standard double Hotel room overnight in pop-

ular cities around the globe. Link to Data Source is `https://businessblog.trivago.com/trivago-hotel-price-index/`. Data used from Trivago is an unstructured data downloaded in the form of pdf from the above link. Data set consist of average accommodation price over a night, city name. The Report considered Leading Airbnb destinations for comparing price.

2.4 Source 3: The Guardian

The Guardian is a famous newspaper in London, Sydney, and NewYork. Using rest API news articles, and blogs can be extracted in JSON format. This API can query the content based on keyword, start date, end date, relevance, etc. Sentiment analysis was done using guardian API for the keyword Airbnb from 2008 to 2018 for New York, Sydney, and London.Link to the Data source: `https://content.guardianapis.com/search?q=,Airbnb&api-key=2d203bae-d6cb-43bd-8897-3a9121759379`

3 Related Work

There have been numerous researches which have been conducted in analyzing how to host successfully on Airbnb

Gunter (2018) published a research paper on what makes an Airbnb host a super host. He says that San Francisco and Bay obtaining excellent ratings by good cancellation policy of the host, responsiveness of host, and good demand on Airbnb. The author says that the commercial host has more probability to obtain the super host status. The report analyzing attributes such as cancellation policies, host response rate, location, property type rented and identity verified by the host, etc.

Georgios Zervas (2015) published a research paper on The Impact of the Sharing Economy on the Hotel Industry by combining data from Airbnb and the Texas hotel industry. Authors estimated that each 10% increase in Airbnb supply results in a 0.37% decrease in monthly hotel room revenue. In Austin, where Airbnb supply is highest, the impact on hotel revenue exceeds 10%.

Ytreberg (2016) reported that a 10 % increase in Airbnb supply decreases hotel revenue by 0.3 %. This effect is around 20 times smaller than the effect on hotel revenue with an increase in hotel supply.

Ke (2017) says the majority of listing are entire homes not a single room, hence according to author Airbnb is more like a rental marketplace rather than a room sharing platform. The author also says that host who has multiple listing is mostly the early movers towards the starting of Airbnb and their listing is entire home/Apartment.

Heo (2019) published a research paper on What is happening in Paris? says that Hotel and Airbnb are not in direct competition and that their relationship might be more complex than previously thought.

According to Ert (2019) , in all cities Superhosts seem to receive a price premium of between 4% and 9% due to their status. However, there is no conclusive evidence that on Airbnb a Superhost badge is more important to hosts with less established reputation than to hosts with an established reputation.

Cheng (2018) reported that users tend to evaluate their experience based on past hotel experiences such as location, amenities and host. According to the author, the price is not a big consideration.

But Cheng (2018) research is opposed by Lin (2018) she says that Most family travelers need more than one room. A hotel may only allow one child to stay in an existing bed, which may not t a family with two kids, but paying for an additional room is costly. That price difference is the key indicator for their decision to choose Airbnb instead of a hotel. She also says that family doesnt want to spend more on hotel instead they want to experience themselves as a local so that they can even spend that money on other tourist activities. It also provides more space less cost, more privacy and more importantly a home experience.

4 Data Model

The following dimensions are used to create, analyze and visualize Airbnb fact table.

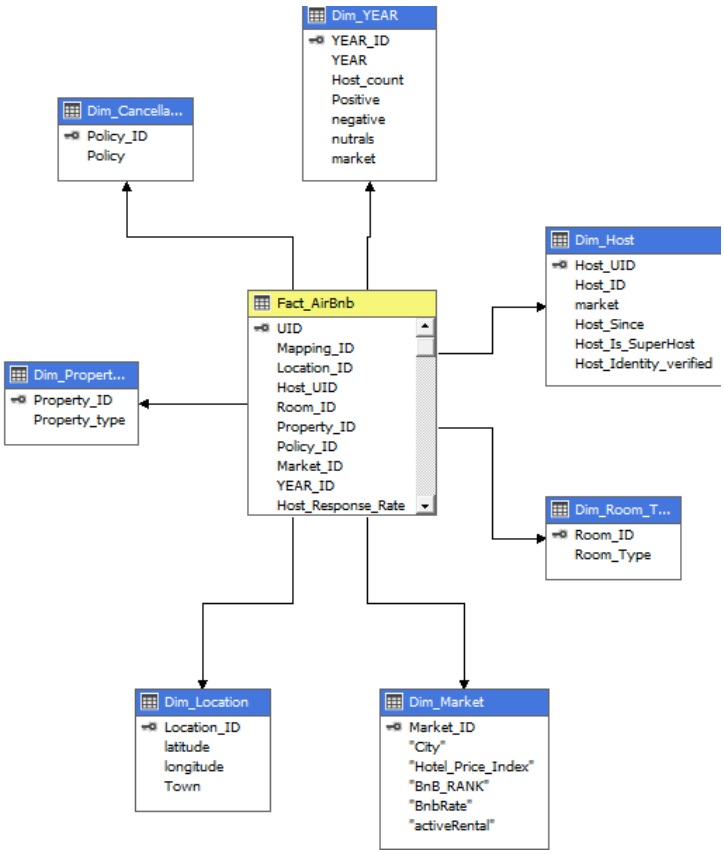


Figure 1: Star schema used Airbnb

4.1 Cancellation policies

This Dimension contains policies of the host while canceling the booking. It contains the policy name and its identifier. Analysis of this dimension gives detailed information about policies existing inside Airbnb and most commonly used policy by the host across cities. Data is collected from inside Airbnb for this dimension.

4.2 Year

This Dimension is used to analyze the sentiments as well as host registration trends yearly. It contains year,host Registration count and sentiments (positive ,negative ,neutral).Data for the dimension is collected from guardian API, Inside Airbnb, Statista.

4.3 Host

This Dimension contains Host Details such as Host since, Host Unique Id, If Host is a Super Host, Host identity verified or not, Market, etc. Host details of the Airbnb listing are found here. Source of the dimension is inside Airbnb, Statista.

4.4 Property Type

This Dimension contains a property identifier and Property Type. Source of the dimension is inside Airbnb. This dimension is used to analyze the property type listed inside Airbnb which can help to find the most preferred property type in Airbnb.

4.5 Room Type

Room Type Dimension contains an identifier and a Room Type. Source of the dimension is Inside Bnb. It is used to analyze the property type listed inside Bnb and can be used to find the most preferred room type in Airbnb

4.6 Location

This Dimension is the location of the listings in Airbnb. It has id, town, longitude, and latitude. The source of the dimension is Inside Airbnb.

4.7 Market

This Dimension contains city wise hotel price Index, as well as Bnb Rate. Fields are hotel price index, Bnb rate, Active listing, City, Identifier, etc. Source of the dimension is Inside Airbnb, Statista, and Trivago. It is used to analyze hotel price and Airbnb rate for city wise. These city details are linked with listing in Airbnb.

Airbnb Fact will be created based on the above dimensions in a star schema. Star schema is a commonly used schema to develop any DataWarehouse project. One or more fact table referring n number of dimension is called star schema. Measures are stored in Airbnb fact table. and I have seven dimensions connecting to Airbnb fact table.

5 Logical Data Map

Table 3: Logical Data Map describing all transformations, sources and destinations for all components of the data model illustrated in Figure 1

Source	Column	Destination	Column	Type	Transformation
1	room_type	Dim_Room_Type	Room_Type	Dimension	It transformed into alphanumeric format.
1	Latitude	Dim_Location	latitude	Dimension	It transformed into numeric format.
1	Longitude	Dim_Location	longitude	Dimension	It transformed into numeric format.
1	Neighborhood	Dim_Location	Town	Dimension	It transformed into alphanumeric format.
2 and 3	City	Dim_Market	City	Dimension	It transformed into alphanumeric format.
3	Hotel_Price_Index	Dim_Market	Hotel_Price_Index	Dimension	It transformed into numeric format.
2	BnB_RANK	Dim_Market	BnB_RANK	Dimension	It transformed into numeric format.
2	Bnbrate	Dim_Market	BnbRate	Dimension	It transformed into alphanumeric format.
2	activeRental	Dim_Market	activeRental	Dimension	It transformed into numeric format.
1 and 4	YEAR	Dim_YEAR	YEAR	Dimension	It transformed into alphanumeric format.
1	Host_Id	Dim_YEAR	Host_count	Dimension	It transformed into alphanumeric format.
4	Positive	Dim_YEAR	Positive	Dimension	It transformed into numeric format.
4	Negative	Dim_YEAR	negative	Dimension	It transformed into numeric format.
4	Neutrals	Dim_YEAR	Neutrals	Dimension	It transformed into numeric format.
1 and 4	Market	Dim_YEAR	Market	Dimension	It transformed into alphanumeric format.
1	Property_type	Dim_Property_Type	Property_type	Dimension	It transformed into alpha numeric format.
1	cancellation_policy	Dim_Cancellation_Type	Policy	Dimension	It transformed into alphanumeric format.
1	Host_ID	Dim_Host	Host_ID	Dimension	It transformed into alphanumeric format.

Continued on next page

Table 3 – *Continued from previous page*

Source	**Column**	**Destination**	**Column**	**Type**	**Transformation**
1	host_since	Dim_Host	host_since	Dimension	It transformed into alphanumeric format.
1	Market	Dim_Host	Market	Dimension	It transformed into alphanumeric format.
1	host_is_ superhost	Dim_Host	host_is_ superhost	Dimension	It transformed into bit format.
1	host_identity _verified	Dim_Host	host_identity_ verified	Dimension	It transformed into bit format.
1	Host_Response _Rate	Fact_AirBnb	Host_Response _Rate	Fact	It transformed into numeric format.
1	Price	Fact_AirBnb	Price	Fact	It transformed into numeric format.
1	Review_Score _Rate	Fact_AirBnb	Review_Score _Rate	Fact	It transformed into numeric format.
1	Review_Score _location	Fact_AirBnb	Review_Score _location	Fact	It transformed into numeric format.

6 ETL Process

The Word ETL is the acronym for Extraction Transformation and Loading. The Extraction, Transformation, and Loading are three functions of the ETL process. ETL is the process of collecting data from different sources in different forms (structured, unstructured and semi-structured), then transforming it into a suitable format by cleaning the data and loading it into the data warehouse. ETL process can be fully automated from R-integration to the cube deployment. The row data collected need to be modified and cleaned and transformed to achieve a quality data warehouse.

The SSMS (SQL Server Management Studio) is used to create a data warehouse and is used to automate the ETL process for the DW project. A Data Warehouse called Airbnb is created in SSMS integrated with SSIS through OLE DB connection. Prior to the execution of control flow task, all staging tables are truncated. The table is used for the data visualizations.

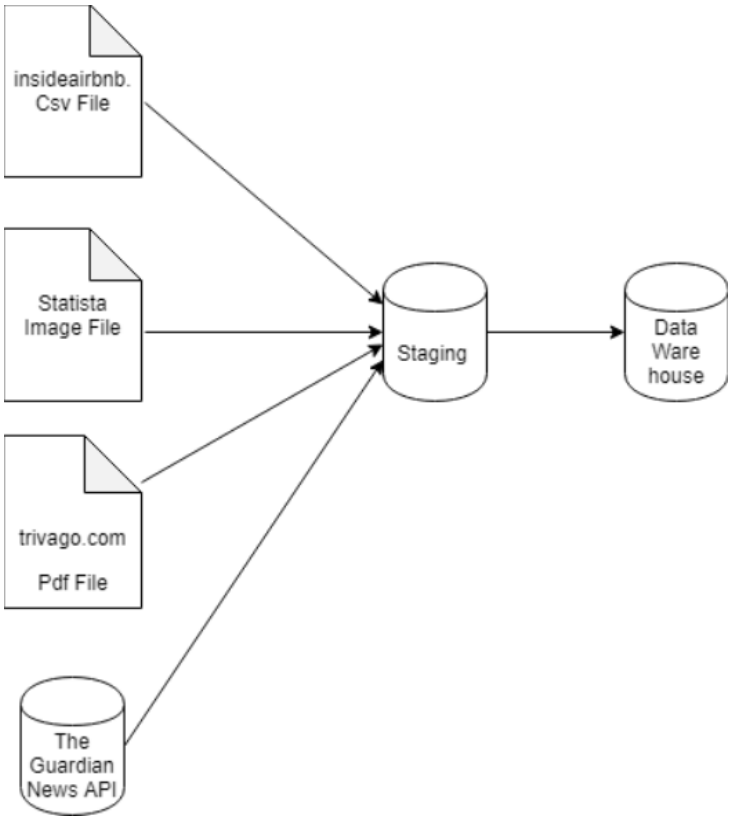


Figure 2: ETL process

This project is using four data sources. The collected data sources are structured CSV file from Inside Airbnb, unstructured image from Statista, unstructured pdf file from the site Trivago and structured data from the Guardian API. All four sources of data are cleaned and transformed into structured CSV files using R-Code.

The Execute process task in SSI will take care of the R code automation in order to avoid redundancy. The Data flow task in SSIS will take care of loading the flat file into the data warehouse staging tables. The Dimension table and the fact table creation of data warehouse are automated using the execute SQL task. The Cube deployment is automated by analysis services processing task in SSIS. The Manual intervention can be prevented as the process is automated from sever data collection to the cube deployment.

The Data warehouse will be more accurate if the process uses multiple data sources to build the data warehouse.

The projects control flow task using SSIS is given below. Sequence Container is used to group the task to execute sequentially. .

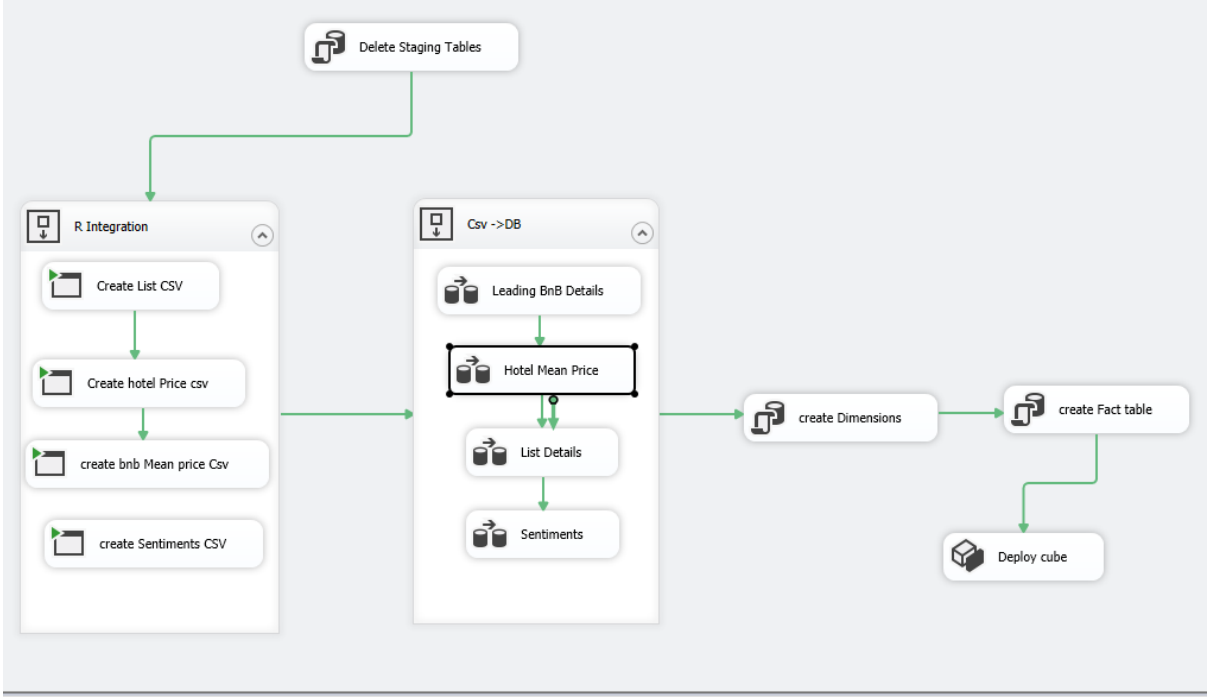


Figure 3: Control flow task of SSIS

The following paragraph explained how data was cleaned using R code and integrated with SSIS. Each time a new CSV file was created whenever the control flow executed.

The Data set from InsideAirbnb Contains 95 columns however 15 columns are used for the project. The remaining columns are removed and a new column called market is added for each city. Some of the columns are having a coma as part of their data hence column separator used for the CSV is a semicolon. Few columns like Host_Response_Rate, Review_Score_Rate, Review_Score_location are having NA values which are further cleaned by zero. The Price is having a $ character as the prefix which is cleaned by R code.The cleaned data frame is saved as an output CSV file.

The data set from Statista is unstructured data in the form of an image. The OCR technique is used to convert the image to text using R. collected text is further cleaned by removing a special character, useless texts. Then the pattern is identified and the required data is retrieved.also the dollar sign is removed from the price variable. The cleaned data frame is saved as an output CSV file using R.

The Trivago data set used was unstructured in the form of pdf. The pdf tools were used to convert pdf to text. Unwanted text and special characters removed from the retrieved text. also, the $ sign is removed from the price variable. And mean price calculated from the pdf for 2018. Finally, the data frame is written into a CSV file

The final data set used is The Guadian API. The Data is collected in the Rest API response.results.fields. The body of API is cleaned and used for sentimental analysis. Positive, negative and neutral values are collected as data frame along with the city and year. The data frame is saved into a CSV file. This is also done using R.

7 Application

The business requirements noted in Section 1 can be answered after the data warehouse processing is completed.

7.1 BI Query 1:Do people prefer Airbnb over the hotel?

The business requirement is to know whether people choose Airbnb Over hotel?.

There are several papers reported that a 10 % increase in Airbnb supply decreases hotel revenue by 0.3 %. The report also analyses that the family prefers to stay in Airbnb over Hotel due to the fact that they have to pay extra money for accommodating an extra kid in Hotel. Considering previous reports money seems to be a factor for the family. The contributing sources of data are www.trivago.com, www.insideairbnb.com, Statista

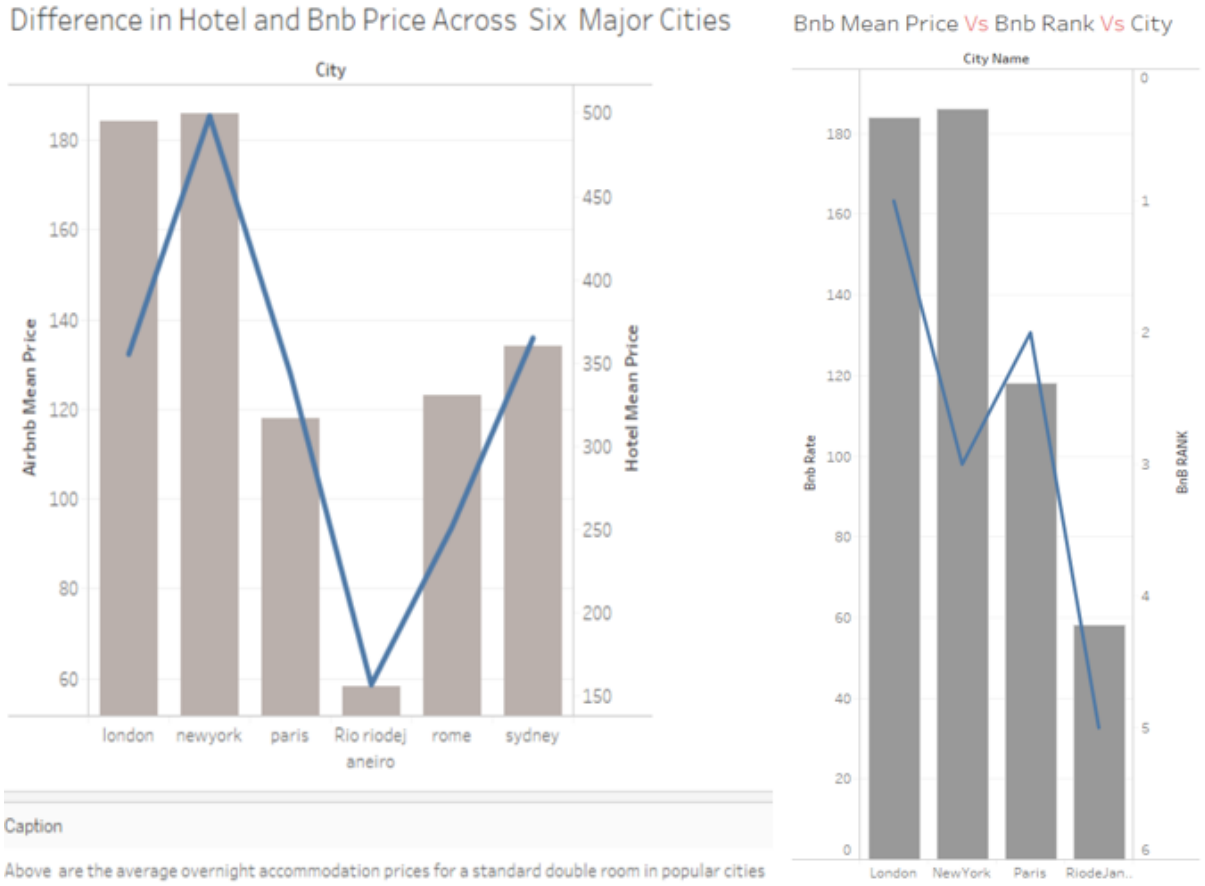


Figure 4: Results for Do people prefer Airbnb over the hotel.

Figure 4, clearly shows that Airbnb Average mean price/night is less than that of hotel mean price/ night across cities. It also analyzed that Airbnb price is directly proportional to hotel price. In 2018 Both Hotel and Airbnb price is maximum in NewYork and the minimum is in Rio de Janeiro. It is also observed that the popularity of Airbnb is also directly proportional to Airbnb Rate.

Hence as per reports price is major consideration it can be concluded that Airbnb definitely chooses over the hotel in major cities. It can also be concluded that anyone who is thinking to invest in Airbnb is a good option.

7.2 BI Query 2: Is it Right time to register as Host?

The next business requirement is to know is it the best time to register as a host?

the Report compared the sentiments along with the host registration across three cities. The three cities compared are London, NewYork, and Sydney. The contributing sources of data are The Guardian, www.insideairbnb.com

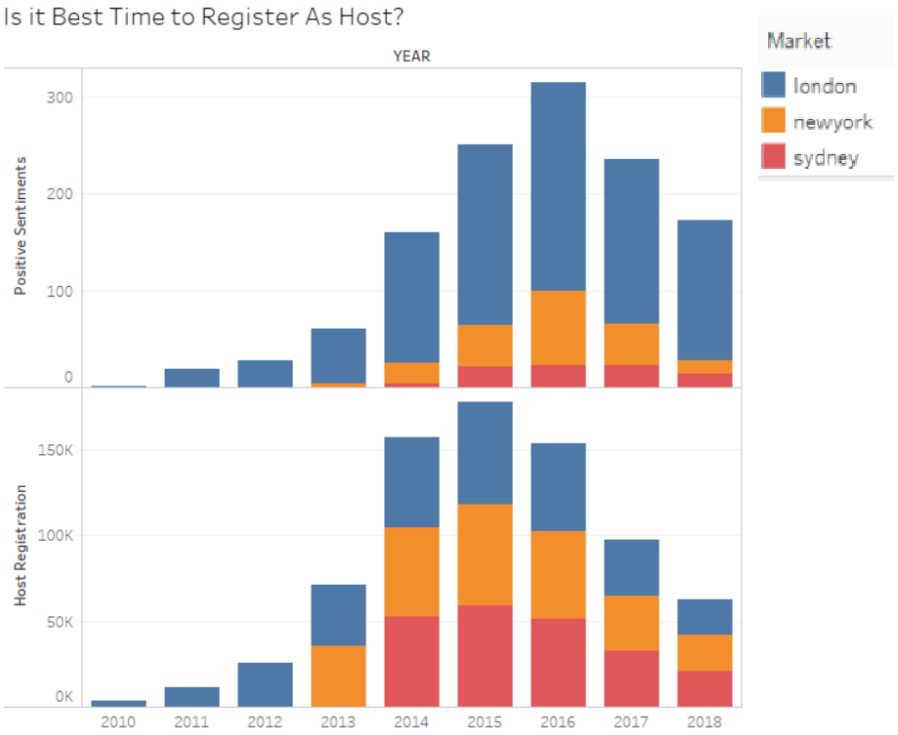


Figure 5: Is it Right time to register as Host?

Figure 5 shows that the Host registration is at its peak during 2015 and 2016 and gradually host registration trend started reducing. Similarly, it is noticed that positive sentiments towards Airbnb are also maximum in 2015 and 2016. hence it is concluded that the popularity of Airbnb is reducing over the years which resulted in to decrease in new registrations of the host.

It is also noted that Airbnb in London is more popular when compared to NewYork and Sydney. From the Figure, it is analyzed that host registration is reduced over the years. That doesnt mean that it could affect the revenue of Airbnb. The registered hosts since 2010 still continued hosting on Airbnb while the New Host registration is decreasing over time. The Reduced host registration trend indicates that it is not the right time to register as a host in Airbnb.

7.3 BI Query 3: What makes a Host as a Super Host?

The Third Business Requirement is to know what makes a host as a super host on Airbnb? the contributed data sources are Statista and www.insideairbnb.com

The Reports says that Host Listing price increased up to 10% if their status changed to Super host. Hence Report Analysed the factors common to super host such as host response rate, cancellation policy, identity verification, Property and Apartment Type, the common location most of the super host located, etc.

Figure 6 shows the factors that commonly haves for super hosts. From the below figure, it is observed that most the super host across the cities have 100% host response rate.

It also analyzed that the verified super host are more when compared to the super host with unverified identity. The commonly used policy among the super hosts is strict_14_with_grace_period. The most commonly used property type is Apartment and Apartment Type is Entire home/apt among super host.

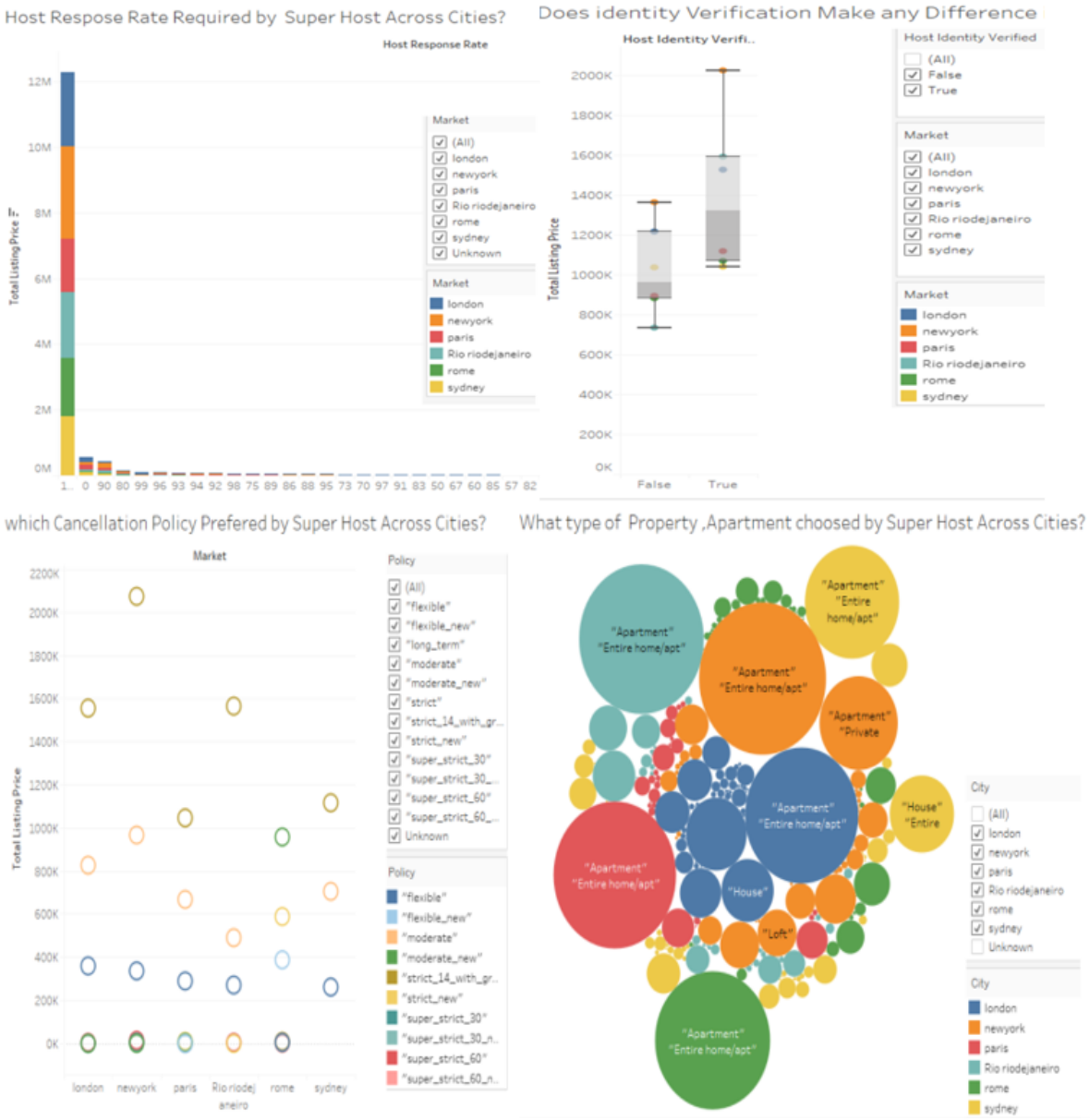


Figure 6: Result for What makes a Host as a Super Host?

Another factor is the location where most of the super host is located. From Figure 7, it can be concluded that the host most likely to become super host if chose a location with location review 10. Also, it is observed that Copacabana in Rio de Janeiro is the place where most of the super hosts are located. NewYork Williamsburg is the best place to start hosting. While Prati in Rome, Le Marais in Paris and Sydney in Sydney are the other best places in each metro city to start hosting.

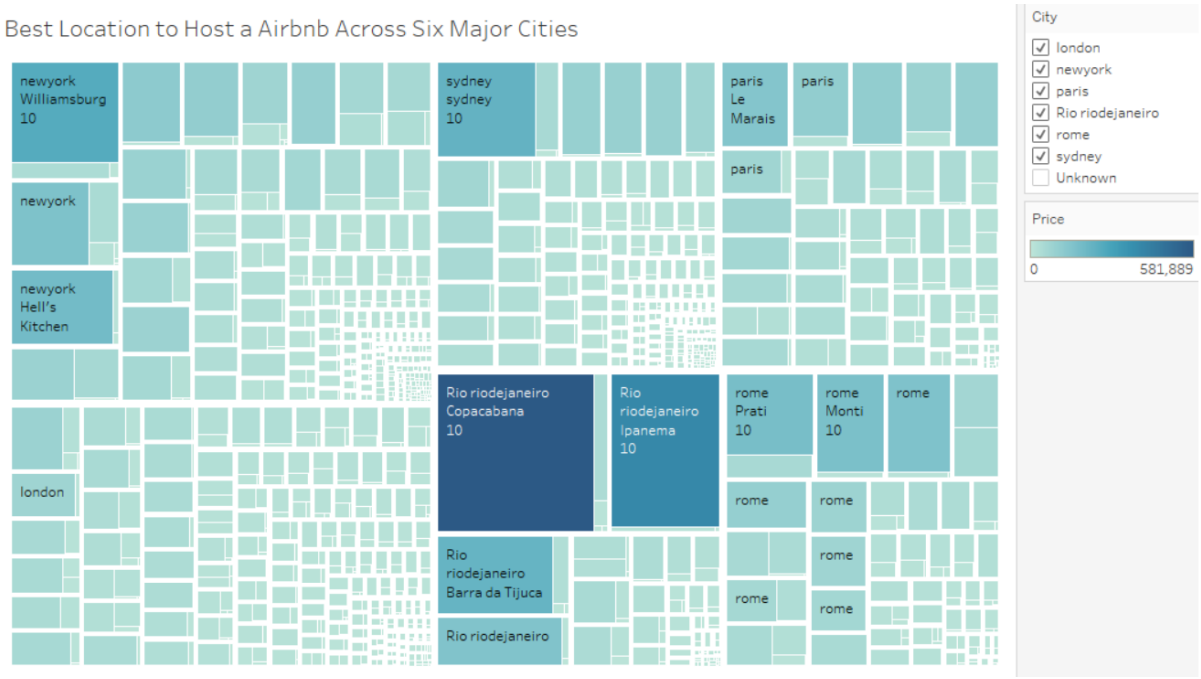


Figure 7: Location Review Vs City Vs Town

7.4 Discussion

This Report aims to find the factors favoring hosting on Airbnb. The project report observed before that hotel price is more when comparing to Airbnb price which is a key factor increasing Airbnb popularity among users. Research reported that price is the key indicator for the family's decision to choose Airbnb instead of a hotel as they may need to pay extra bed for accommodating their kids in the hotel (Lin 2018). The research is contradicted by another report saying users tend to evaluate Airbnb by location, host, and amenities but not by price(Cheng 2018). There are Other findings reporting that a 10% increase in Airbnb supply decreases hotel revenue by 0.3%(Georgios Zervas 2015)(Ytreberg 2016). From these, it can be concluded if people tend to choose Airbnb if hotel price increases. The familys on vacation choose to prefer Airbnb than Hotel But still, the hotel has it's on demand. It can seem that the above findings support (Heo 2019) report , Hotel and Airbnb are not in direct competition and that their relationship might be more complex

A Study found that obtaining excellent ratings by good cancellation policy of the host, responsiveness of host, and good demand on Airbnb. They also found that commercial host has more probability to obtain the super host status[1]. Another research reported super hosts to receive a premium price of up to 9%(Ert 2019). The report analyzing attributes such as cancellation policies, host response rate, location, property type rented and identity verified by the host, etc

Another report says that host who has multiple listing is mostly the early movers towards the starting of Airbnb and their listing is entire home/Apartment(Ke 2017). It explains the host registration trend is maximum at early stages of Airbnb gives a hint that multiple listing host is registered when positive sentiments are its peak.

8 Conclusion and Future Work

This Report aims to analyze the factors favoring hosting on Airbnb. It is observed that the Hotel price is more when compared to the Airbnb Overnight price across the cities. It is also observed that the Airbnb overnight price is directly proportional to the hotel overnight prices. as per previous researches, we concluded that price is a key factor in choosing Airbnb over the hotel.

The Report also analyzed that both positive sentiments and Host Registration is in its peak during 2015 and 2016. The popularity of Airbnb is reducing in 2018 when compared to previous years. It also concluded that it is not the right time if anyone to register as a host in Airbnb.

The Report also found some factors that help to achieve super host status to every host in Airbnb. It is observed that by increasing the host response rate, verifying the identification and choosing cancellation policy as strict_14_with_grace_period, renting an apartment as entire home and choosing a location with review score 10 for hosting can increase the probability to become a super host in Airbnb.

The future work can include the fact analyzing whether the price is a key indicator for choosing Airbnb or does any other factor included choosing an Airbnb instead of a hotel? It can also further research the user review and find out the reason for the downward trend of Airbnb among host over the years?

References

Cheng, M, J. X. (2018), 'What do airbnb users care about? an analysis of online review comments', *International Journal of Hospitality Management* **76**(A), 58–70.
URL: *https://www.sciencedirect.com/science/article/pii/S0278431917307491!*

Ert, E., F. A. (2019), 'The evolution of trust in airbnb: A case of home rental', *Annals of Tourism Research* **75**(1), 279–287.
URL: *https://www.sciencedirect.com/science/article/pii/S0160738319300040*

Georgios Zervas, Davide Proserpio, J. W. B. (2015), 'The impact of the sharing economy on the hotel industry: Evidence from airbnb's entry into the texas market', *Special Interest Group on Economics and Computation* pp. 637–637.
URL: *https://dl.acm.org/citation.cfm?doid=2764468.2764524*

Gunter, U. (2018), 'What makes an airbnb host a super host? empirical evidence from san francisco and the bay area', *Tourism Management* **66**(1), 26–37.
URL: *https://www.sciencedirect.com/science/article/pii/S026151771730242X*

Heo, C.Y., B. I. C. M. (2019), 'What is happening in paris? airbnb, hotels, and the parisian market: A case study', *Tourism Management* **70**(1), 78–88.
URL: *https://www.sciencedirect.com/science/article/pii/S0261517718300785*

Ke, Q. (2017), 'Sharing means renting?: An entire-marketplace analysis of airbnb', *'WebSci17, Troy, NY, USA* .
URL: *https://arxiv.org/pdf/1701.01645.pdf*

Lin, P. M. (2018), 'Is airbnb a good choice for family travel?', *Journal of China TourismResearch* .
URL: *https://www.researchgate.net/publication/329766979*$_{I}s_{A}irbnb_{a}G ood_{C}hoice_{f}or_{F}amily_{T}ravel$

Ytreberg, N. S. (2016), 'Competitive effects of airbnb on the norwegian hotel market', *Master in Economics University of Bergen* .
URL: *http://beccle.no/files/2016/12/Masteroppgave*$_{N}SY2.pdf$

Appendix

R code example

```
#####################################################################
# This code will download the data set from Inside airbnb website#
#it will further perform cleaning and Add/ modify the required #
#columns and all the data frames are merged into one file output #
#it into a CSV. #
#####################################################################

library(gtools)
library(Jmisc)
library(R.utils)
options(max.print = 999999999)

temp <- tempfile()
download.file("http://data.insideairbnb.com/united-states/ny/new-york-
    ↪ city/2018-09-08/data/listings.csv.gz", temp)
gzfile(temp, 'rt')
df1 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)

temp <- tempfile()
download.file("http://data.insideairbnb.com/united-kingdom/england/london
    ↪ /2018-09-10/data/listings.csv.gz", temp)
gzfile(temp, 'rt')
df2 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)

temp <- tempfile()
download.file("http://data.insideairbnb.com/france/ile-de-france/paris/20
    ↪ 18-09-10/data/listings.csv.gz", temp)
gzfile(temp, 'rt')
df3 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)

temp <- tempfile()
download.file("http://data.insideairbnb.com/brazil/rj/rio-de-janeiro/2018
    ↪ -09-14/data/listings.csv.gz", temp)
gzfile(temp, 'rt')
df4 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)

temp <- tempfile()
download.file("http://data.insideairbnb.com/italy/lazio/rome/2018-09-12/
    ↪ data/listings.csv.gz", temp)
gzfile(temp, 'rt')
```

```
df5 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)


temp <- tempfile()
download.file("http://data.insideairbnb.com/australia/nsw/sydney/2018-09-
    ↪ 09/data/listings.csv.gz", temp)
gzfile(temp, 'rt')
df6 <- read.csv(temp,stringsAsFactors = FALSE)
unlink(temp)

#Selecting Required columns from 95 columns

Data1 <- subset( df1, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96 ) )
Data2 <- subset( df2, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96 ) )
Data3 <- subset( df3, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96) )
Data4 <- subset( df4, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96 ) )
Data5 <- subset( df5, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96 ) )
Data6 <- subset( df6, select = -c(2:19,21:22,24:26,28,30:36,38,40:48,51,5
    ↪ 4:60,62:79,81:84,86:91,93:96) )

#adding market as last column
Data11<-addCol(Data1,value=c("market"="newyork"))
Data12<-addCol(Data2,value=c("market"="london"))
Data13<-addCol(Data3,value=c("market"="paris"))
Data14<-addCol(Data4,value=c("market"="Rio␣riodejaneiro"))
Data15<-addCol(Data5,value=c("market"="rome"))
Data16<-addCol(Data6,value=c("market"="sydney"))

#combaining all data frame
Data21<-smartbind(Data11, Data12)
Data22<-smartbind(Data13, Data14)
Data23<-smartbind(Data15, Data16)
Data24<-smartbind(Data21, Data22)
Data25<-smartbind(Data24, Data23)

#Writing into CSV
write.csv2(Data25,"C:/Users/MOLAP/Desktop/AirBnb/output/listings_1.csv",
    ↪ row.names = FALSE)
```

```
#####################################################################
#This Code will download image in the statista and using OCR it # #
   ↪ convert imge to text and then perform clean on those data output #
#into a csv file #
#####################################################################

library('tesseract')

image.Data<-ocr("https://infographic.statista.com/normal/chartoftheday_14
   ↪ 986_active_airbnb_listings_in_major_cities_n.jpg", engine =
   ↪ tesseract("eng"))
s1<-gsub("[[:punct:]]", "", image.Data)
image.Array<-strsplit(s1,'\n')
image.realData<-image.Array[[1]]
myData<-image.realData[4:15]
CityName<-NULL
activeRental<-NULL
BnbRate<-NULL
for (i in 1 :length(myData)){
  one.Data<-myData[i]
  words <- strsplit(one.Data, "␣")[[1]]
  ls<-unique(tolower(words))
  if(i==1)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[5]
  }
  else if(i==2)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[6]
  }else if(i==3)
  {name<-paste(ls[1],ls[2])
  activeRent<-ls[5]
  dailyRate<-ls[7]
  }
  else if(i==4){name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[5]
  }else if(i==5)
  {name<-paste(ls[1],ls[2])
  activeRent<-ls[5]
  dailyRate<-ls[7]
  }else if(i==6)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[5]
```

```
  }
  else if(i==7)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[6]
  }else if(i==8)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[5]
  }else if(i==9)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[6]
  }else if(i==10)
  {name<-paste(ls[1],ls[2])
  activeRent<-ls[5]
  dailyRate<-ls[7]
  }else if(i==11)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[6]
  }else if(i==12)
  {name<-ls[1]
  activeRent<-ls[4]
  dailyRate<-ls[6]
  }
  if(name=="new york")
  {
    name<-"newyork"
    print(name)
  }
  else if(name=="rio dejaneiro")
  {
    name<-"Rio riodejaneiro"
  }
  CityName<-c(CityName,name)
  activeRental<-c(activeRental,activeRent)
  BnbRate<-c(BnbRate,dailyRate)

}
BnB_ID<-c(1:length(activeRental))
BnB_RANK<-c(1:length(activeRental))
datafile<-data.frame(BnB_ID,BnB_RANK,CityName,activeRental,BnbRate)
write.csv(datafile,"C:/Users/MOLAP/Desktop/AirBnb/output/Leading_Bnb_
    ↪ Destinations_1.csv",row.names = FALSE)
```

```
#####################################################################
#Trivago.com pdf is downloaded and covert in to data frame with #
#the help of pdf tools,clean and then generate Output CSV #
#####################################################################

library(pdftools)
file<-pdf_text('C:/Users/MOLAP/Desktop/AirBnb/Table␣View.pdf')
data<-unlist(strsplit(file,"\n"))
data1<-data[2:8]
s1<-gsub("\t", "", data1)
s2<-gsub("\r", "", s1)
s3<-gsub("[[:punct:]]", "", s2)
s4<-gsub("Multiple","",s3)
s5<-gsub("EUR","",s4)
s6<-gsub("Euro","",s5)
s7<-gsub("values","",s6)
s8<-gsub("York","",s7)
s9<-gsub("de","",s8)
s10<-gsub("Janeiro","",s9)
mycol1<-NULL
mycol2<-NULL
mycol3<-NULL
mycol4<-NULL
mycol5<-NULL
mycol6<-NULL
mycol7<-NULL
City<-NULL
templist<-NULL
Hotel_Price_Index<-NULL
arraylen<-NULL

for(i in 1:length(s10))
{
  csvData<-unlist(strsplit(s10[i],"␣"))
  # print(csvData)
  for(j in 1:length(csvData))
  {
    if(""==csvData[j])
    {
      #print("Null found")
    }else
    {
      print(csvData[j])

      switch(i,mycol1<-c(mycol1,csvData[j])
             ,mycol2<-c(mycol2,csvData[j]),
             mycol3<-c(mycol3,csvData[j]),
```

```
            mycol4<-c(mycol4,csvData[j]),
            mycol5<-c(mycol5,csvData[j]),
            mycol6<-c(mycol6,csvData[j])
            ,mycol7<-c(mycol7,csvData[j]))


    }
  }}
templist[1] <- list(mycol1)
templist[2] <- list(mycol2)
templist[3] <- list(mycol3)
templist[4]<- list(mycol4)
templist[5] <- list(mycol6)
templist[6] <- list(mycol7)
for(k in 1:length(templist))
{
  hotelmeanval=0
  temparray<-templist[[k]]
  for(n in 1:length(temparray))
  {
    if(n==1)
    {
      print(temparray[n])
      if(!is.null(temparray) && temparray[n]=="New")
      {
        temparray[n]="NewYork"
      }
      else if(!is.null(temparray) && temparray[n]=="Rio")
      {
        temparray[n]="Rio␣riodejaneiro"
      }
      else if(!is.null(temparray) && temparray[n]=="London")
      {
        temparray[n]="london"
      }
      else if(!is.null(temparray) && temparray[n]=="Paris")
      {
        temparray[n]="paris"
      }
      else if(!is.null(temparray) && temparray[n]=="Rome")
      {
        temparray[n]="rome"
      }
      else if(!is.null(temparray) && temparray[n]=="Sydney")
      {
        temparray[n]="sydney"
      }
      City<-c(City,temparray[n])
```

```
    }
    else{
      intval<- as.integer(substr(temparray[n],2,4))
      hotelmeanval<-hotelmeanval+intval
      arraylen<- (length(temparray)-1)
    }

  }
  Hotel_Price_Index<-c(Hotel_Price_Index,hotelmeanval/arraylen)
}
Location_ID<-c(1:6)
outputData<-data.frame(Location_ID,City,Hotel_Price_Index)
write.csv(outputData,"C:/Users/MOLAP/Desktop/AirBnb/output/HOTEL_PRICE_
    ↪ INDEX.csv" ,row.names = FALSE)


#####################################################################
# Guardian API used to do sentimental analysis and the positive # #
    ↪ negative and nutral are saved into outputfilr #
#####################################################################

library(Jmisc)
library (jsonlite)
library (rJava)
library (rvest)
library(gtools)
library (syuzhet)

Reviews<-NULL
years<-NULL
totals<-NULL
positives<-NULL
negatives<-NULL
neutrals<-NULL
for(year in 2008:2018)
{
  s<-NULL
  for ( i in 1:50) {
    house<-"uk"
    url = paste("https://content.guardianapis.com/search?q=,Airbnb&api-
        ↪ key=2d203bae-d6cb-43bd-8897-3a9121759379&show-fields=body&
        ↪ production-office=",house,sep ="")
    url1 = paste(url,"&from-date=" , sep ="")
    FromDate<-paste(year,"-01-01",sep = "")
    url2 = paste(url1,FromDate , sep ="")
    url3 = paste(url2,"&to-date=" , sep ="")
    ToDate<-paste(year,"-12-31",sep = "")
```

```
    url4 = paste(url3,ToDate , sep ="")
    url5=paste(url4,"&page=",sep="")
    url6=paste(url5,i,sep="")
    tryCatch({
      content <- fromJSON (url6 , flatten = TRUE )
      tmp12 = as.data.frame(content)
      ifelse(i!=1,s <- smartbind (s,tmp12),s<-tmp12)

    }, error = function(err) {

      return()

    }, finally = {
    })
  }

  cleanBody<- function ( htmlString ) {
       return ( gsub("␣<.*?␣>", "", htmlString ) );
  }
  cleanone <- sapply(s$response.results.fields.body,function(row) iconv(
     ↪ row, "latin1", "ASCII", sub=""))


  s$cleanbody = trimws(cleanBody(cleanone))

  sent.value<- get_sentiment (s$cleanbody )

  max_positive <- s$cleanbody[sent.value == max(sent.value)]
  max_negative <- s$cleanbody[sent.value <= min(sent.value)]
  positive <- s$cleanbody[sent.value > 0]
  negative <- s$cleanbody[sent.value < 0]
  neutral<- s$cleanbody[sent.value == 0]

  total = length(positive) + length(negative) + length(neutral)
  positives<-c(positives,(length(positive)))
  negatives<-c(negatives,(length(negative)))
  neutrals<-c(neutrals,(length(neutral)))
  years<-c(years,year)
  Reviews<-c(Reviews,Review)
  s$bodyCleaned<-NULL
}
sentiments<-data.frame(years,positives,negatives,neutrals)
sentiments_final<-addCol(sentiments,value=c("market"="london"))
write.csv( sentiments_final,"C:/Users/MOLAP/Desktop/AirBnb/output/
   ↪ AirBnbSentimental.csv",row.names = FALSE )
```

SQL Code

```
#########################################################
# #
#create RawTables #
# #
#########################################################

CREATE TABLE [dbo].[Hotel_Mean_Price](
                              ["Location_ID"] [int] PRIMARY KEY,
                              ["City"] [varchar](50) NULL,
                              ["Hotel_Price_Index"] [varchar](50) NULL);

 CREATE TABLE [dbo].[Leading_Bnb_Dest](
                              ["BnB_ID"] [int] PRIMARY KEY,
                              ["BnB_RANK"] [int] NULL,
                              ["CityName"] [varchar](50) NULL,
                              ["activeRental"] [varchar](50) NULL,
                              ["BnbRate"] [varchar](50) NULL);

CREATE TABLE [dbo].[List_Details](
        ["id"] [int] PRIMARY KEY,
        ["host_id"] [int] NULL,
        ["host_since"] [varchar](50) NULL,
        ["host_response_rate"] [varchar](50) NULL,
        ["host_is_superhost"] [varchar](50) NULL,
        ["host_identity_verified"] [varchar](50) NULL,
        ["neighbourhood"] [varchar](50) NULL,
        ["latitude"] [varchar](50) NULL,
        ["longitude"] [varchar](50) NULL,
        ["property_type"] [varchar](50) NULL,
        ["room_type"] [varchar](50) NULL,
        ["price"] [varchar](50) NULL,
        ["review_scores_rating"] [varchar](50) NULL,
        ["review_scores_location"] [varchar](50) NULL,
        ["cancellation_policy"] [varchar](50) NULL,
        ["market"] [varchar](50) NULL);

CREATE TABLE [dbo].[Sentimental_Details](
        [Sent_ID] [int] PRIMARY KEY,
        [years] [varchar](50) NULL,
        [positives] [varchar](50) NULL,
        [negatives] [varchar](50) NULL,
        [neutrals] [varchar](50) NULL,
        [market] [varchar](50) NULL);
```

```
#########################################################
# #
#create Dimensions tables #
# #
#########################################################

IF OBJECT_ID ('dbo.Dim_Room_Type') IS NOT NULL
DROP TABLE Dim_Room_Type
CREATE TABLE [Dim_Room_Type](
       [Room_ID] [int] IDENTITY(1,1) NOT NULL,
       [Room_Type] [varchar](50) NULL
CONSTRAINT PK_Dim_Room_Type_Table PRIMARY KEY ([Room_ID])
)
INSERT INTO Dim_Room_Type([Room_Type])
SELECT distinct ["room_type"] from List_Details;


IF OBJECT_ID ('dbo.Dim_Location') IS NOT NULL
DROP TABLE Dim_Location
CREATE TABLE Dim_Location(
       [Location_ID] [int] IDENTITY(1,1) NOT NULL,
       [latitude] [varchar](50) NULL,
       [longitude] [varchar](50) NULL,
       [Town] [varchar](50) NULL,
CONSTRAINT PK_Dim_Location_Table PRIMARY KEY ([Location_ID]))
INSERT INTO Dim_Location([latitude],[longitude],[Town])
SELECT distinct ["latitude"],
              ["longitude"], REPLACE(REPLACE(["neighbourhood"],
              '"',["market"]),'"','')
from List_Details;


IF OBJECT_ID ('dbo.Dim_Market') IS NOT NULL
DROP TABLE Dim_Market
CREATE TABLE Dim_Market([Market_ID] [int] IDENTITY(1,1) NOT NULL,
       ["City"] [varchar](50) NULL,
       ["Hotel_Price_Index"] [float] NULL,
       ["BnB_RANK"] [int] NULL,
       ["BnbRate"] [int] NULL,
       ["activeRental"] [int] NULL,
CONSTRAINT PK_Dim_Market_Table PRIMARY KEY ([Market_ID]))
INSERT INTO Dim_Market(["City"],
                      ["Hotel_Price_Index"],
                      ["BnB_RANK"],
                      ["BnbRate"],
                      ["activeRental"])
 select replace(t.["CityName"],'"',''),
```

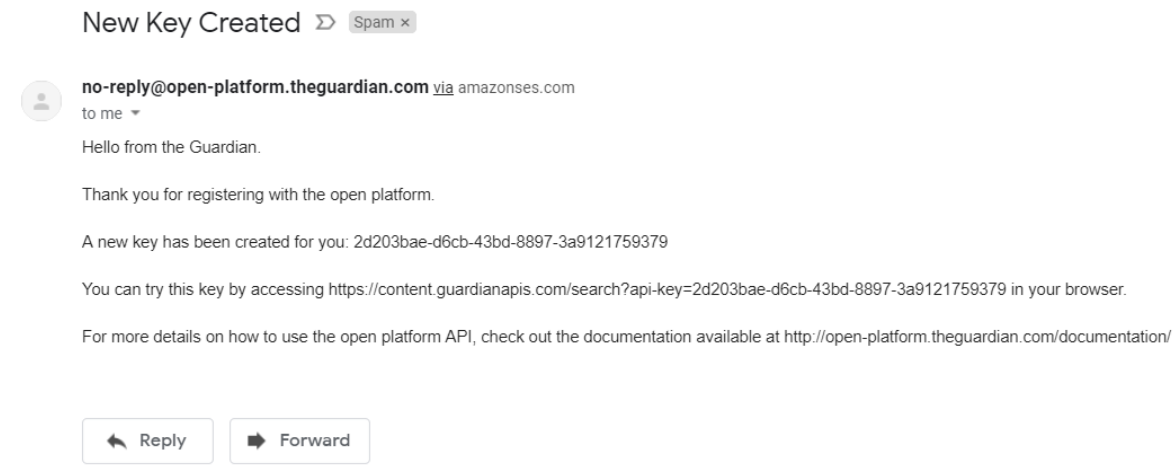
```
                Try_convert(float,s.["Hotel_Price_Index"]),
                t.["BnB_RANK"],
                replace(t.["BnbRate"],'"',''),
                replace(t.["activeRental"],'"','')
  from Leading_Bnb_Dest t,Hotel_Mean_Price s
  where s.["City"]=t.["CityName"];


IF OBJECT_ID ('dbo.Dim_YEAR') IS NOT NULL
DROP TABLE Dim_YEAR
CREATE TABLE Dim_YEAR(
       [YEAR_ID] int identity (1,1) ,
       [YEAR] varchar(10),
       [Host_count] int,
       [Positive] int,
       [negative] int,
       [nutrals] int,
       [market] varchar(50)
CONSTRAINT PK_Dim_YEAR PRIMARY KEY ([YEAR_ID]));
INSERT INTO Dim_YEAR([YEAR],
                  [Host_count] ,
                  [Positive] ,
                  [negative] ,
                  [nutrals] ,
                  [market])
select distinct p.years,
       sum(hostcount),
       p.positives,
       p.negatives,
       p.neutrals,
       p.market
from (select distinct year(cast(replace(p.["host_since"],'"','') as date)
   ↪ ) myyear,
        ["market"],
        COUNT(*) hostcount
  from List_Details p
 group by p.["host_id"],year(cast(replace(p.["host_since"],'"','') as
    ↪ date)),["market"]) n,Sentimental_Details p
 where p.market=replace(n.["market"],'"','') and p.years=n.myyear
  group by p.years,p.market,p.positives,p.negatives,p.neutrals;


IF OBJECT_ID ('dbo.Dim_Property_Type') IS NOT NULL
DROP TABLE Dim_Property_Type
CREATE TABLE Dim_Property_Type(
       [Property_ID] [int] IDENTITY(1,1) NOT NULL,
       [Property_type] varchar(50)
CONSTRAINT PK_Dim_Property_Type PRIMARY KEY ([Property_ID]))
```

```
INSERT INTO Dim_Property_Type( [property_type] )
select distinct ["property_type"] from List_Details;


IF OBJECT_ID ('dbo.Dim_Cancellation_Type') IS NOT NULL
DROP TABLE Dim_Cancellation_Type
CREATE TABLE Dim_Cancellation_Type(
       [Policy_ID] [int] IDENTITY(1,1) NOT NULL,
       [Policy] varchar(50)
CONSTRAINT PK_Dim_Cancellation_Type PRIMARY KEY ([Policy_ID])
)
INSERT INTO Dim_Cancellation_Type( [Policy] )
select distinct ["cancellation_policy"] from List_Details;


IF OBJECT_ID ('dbo.Dim_Host') IS NOT NULL
DROP TABLE Dim_Host
CREATE TABLE Dim_Host(
       [Host_UID] [int] IDENTITY(1,1) NOT NULL,
       [Host_ID] int,
       [market] varchar(50),
       [Host_Since] Date,
       [Host_Is_SuperHost] bit,
       [Host_Identity_verified] bit,
CONSTRAINT PK_Dim_Host PRIMARY KEY ([Host_UID]))
INSERT INTO Dim_Host([Host_ID],
                     [market],
                     [Host_Since],
                     [Host_Is_SuperHost],
                     [Host_Identity_verified] )
select distinct ["host_id"],
    replace(["market"],'"',''),
    CAST(REPLACE(["host_since"],'"','') As Date) as host_since,
    CAST(IIF ( REPLACE(["host_is_superhost"],'"','') ='t',1,0) AS BIT) as
       ↪  host_is_superhost,
     CAST(IIF ( REPLACE(["host_identity_verified"],'"','')='t',1,0) AS
       ↪ BIT) as host_identity_verified
 from List_Details;
```

```
#################################################################
# #
#create fact table #
# #
#################################################################

IF OBJECT_ID ('dbo.Fact_AirBnb') IS NOT NULL
DROP TABLE Fact_AirBnb
CREATE TABLE Fact_AirBnb(
    [UID] int identity(1,1),
        [Mapping_ID] [int],
        [Location_ID] int,
        [Host_UID] int,
        [Room_ID] int,
        [Property_ID] int,
        [Policy_ID] int,
        [Market_ID] int,
        [YEAR_ID] int,
        [Host_Response_Rate] varchar(4),
        [Price] float,
        [Review_Score_Rate] varchar(4),
        [Review_Score_location] varchar(4),
CONSTRAINT PK_Fact_AirBnb PRIMARY KEY ([UID])
)
INSERT INTO Fact_AirBnb( [Mapping_ID],
        [Location_ID] ,
        [Host_UID] ,
        [Room_ID] ,
        [Property_ID] ,
        [Policy_ID] ,
        [Market_ID] ,
        [YEAR_ID],
        [Host_Response_Rate],
        [Price],
        [Review_Score_Rate] ,
        [Review_Score_location] )
select a.["id"],
      b.Location_ID,
          c.Host_UID,
          d.Room_ID,
          e.Property_ID,
          f.Policy_ID,
          g.[Market_ID],
          h.[YEAR_ID],
          REPLACE(REPLACE(REPLACE(a.["host_response_rate"],'%',''),'"',''
              ↪ ),'N/A',0) as host_response_rate,
          Try_convert(float,substring(REPLACE(a.["price"],'"',''),2,len(
              ↪ REPLACE(a.["price"],'"',''))))) as [Varchar to float] ,
```

```
        replace(["review_scores_rating"],'NA',0) as review_scores_
            ↪ rating,
        replace(["review_scores_location"],'NA',0) as review_scores_
            ↪ location
from List_Details a
join Dim_Location b
on b.latitude=a.["latitude"] and b.longitude=a.["longitude"]
join [AirBnb].[dbo].[Dim_Host] c
 on c.Host_ID=a.["host_id"]
join Dim_Room_Type d
on a.["room_type"]=d.Room_Type
join Dim_Property_Type e
on a.["property_type"]=e.Property_type
join Dim_Cancellation_Type f
on a.["cancellation_policy"]=f.Policy
join Dim_Market g
on g.["City"]=REPLACE(a.["market"],'"','')
join [Dim_YEAR] h
on year(cast(replace(a.["host_since"],'"','') as date))= h.[YEAR]
```

Screenshots


New Key Created Spam ×

no-reply@open-platform.theguardian.com via amazonses.com
to me

Hello from the Guardian.

Thank you for registering with the open platform.

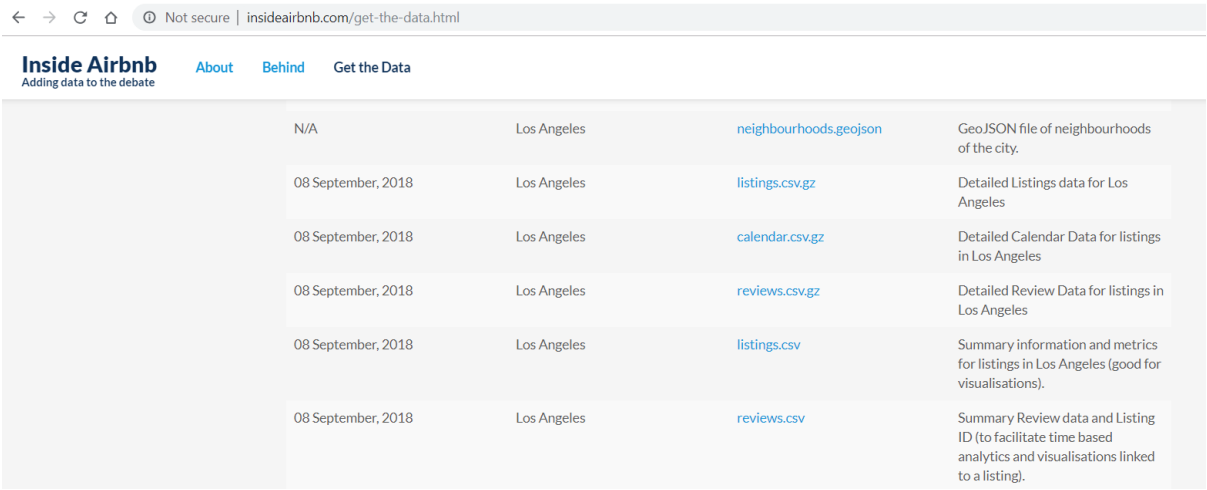
A new key has been created for you: 2d203bae-d6cb-43bd-8897-3a9121759379

You can try this key by accessing https://content.guardianapis.com/search?api-key=2d203bae-d6cb-43bd-8897-3a9121759379 in your browser.

For more details on how to use the open platform API, check out the documentation available at http://open-platform.theguardian.com/documentation/

Reply Forward

Figure 8: Approval Email from The Guardian

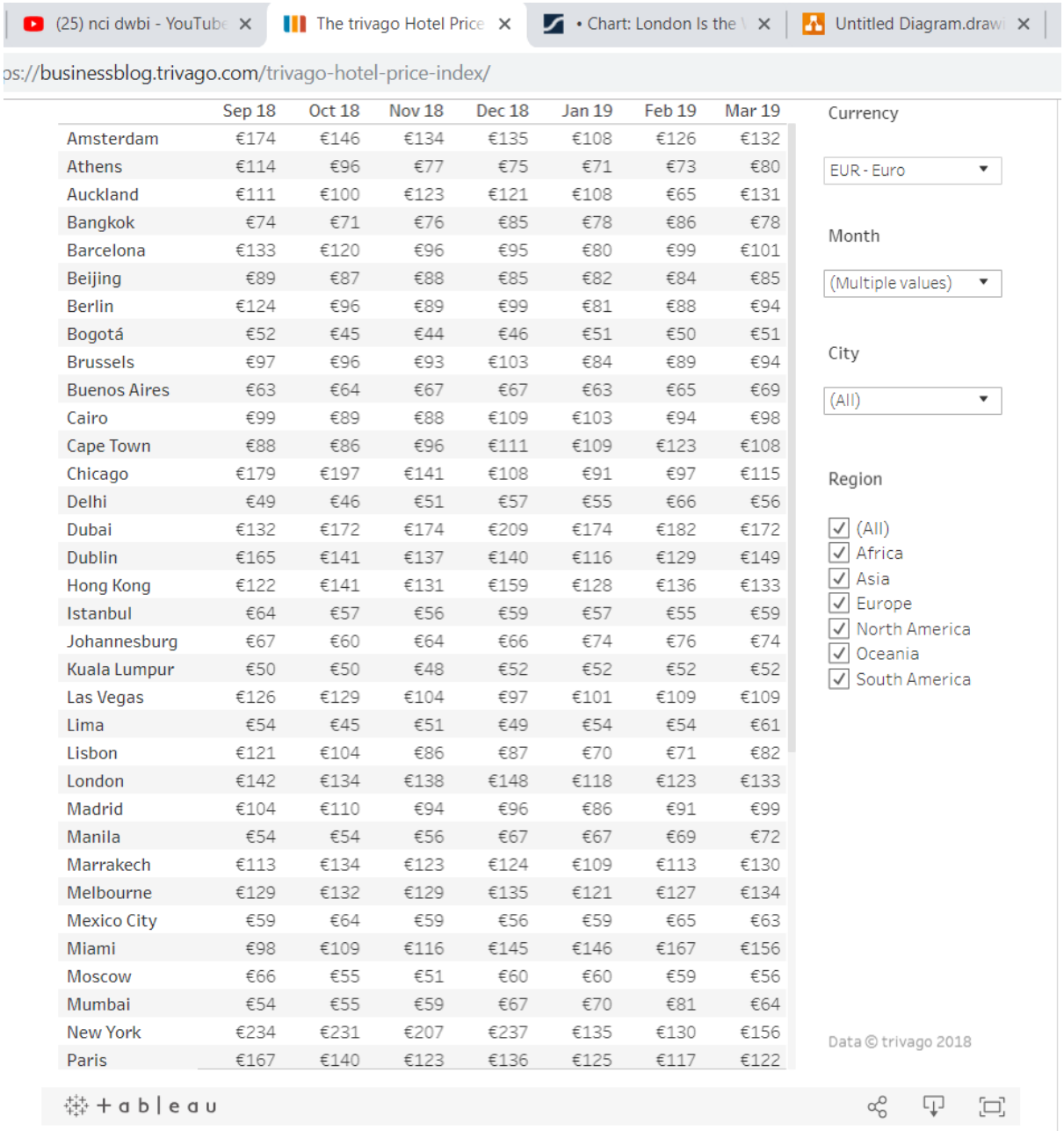

Not secure | insideairbnb.com/get-the-data.html

Inside Airbnb
Adding data to the debate

About Behind Get the Data

N/A	Los Angeles	neighbourhoods.geojson	GeoJSON file of neighbourhoods of the city.
08 September, 2018	Los Angeles	listings.csv.gz	Detailed Listings data for Los Angeles
08 September, 2018	Los Angeles	calendar.csv.gz	Detailed Calendar Data for listings in Los Angeles
08 September, 2018	Los Angeles	reviews.csv.gz	Detailed Review Data for listings in Los Angeles
08 September, 2018	Los Angeles	listings.csv	Summary information and metrics for listings in Los Angeles (good for visualisations).
08 September, 2018	Los Angeles	reviews.csv	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).

Figure 9: Listing details from Inside Airbnb



	Sep 18	Oct 18	Nov 18	Dec 18	Jan 19	Feb 19	Mar 19
Amsterdam	€174	€146	€134	€135	€108	€126	€132
Athens	€114	€96	€77	€75	€71	€73	€80
Auckland	€111	€100	€123	€121	€108	€65	€131
Bangkok	€74	€71	€76	€85	€78	€86	€78
Barcelona	€133	€120	€96	€95	€80	€99	€101
Beijing	€89	€87	€88	€85	€82	€84	€85
Berlin	€124	€96	€89	€99	€81	€88	€94
Bogotá	€52	€45	€44	€46	€51	€50	€51
Brussels	€97	€96	€93	€103	€84	€89	€94
Buenos Aires	€63	€64	€67	€67	€63	€65	€69
Cairo	€99	€89	€88	€109	€103	€94	€98
Cape Town	€88	€86	€96	€111	€109	€123	€108
Chicago	€179	€197	€141	€108	€91	€97	€115
Delhi	€49	€46	€51	€57	€55	€66	€56
Dubai	€132	€172	€174	€209	€174	€182	€172
Dublin	€165	€141	€137	€140	€116	€129	€149
Hong Kong	€122	€141	€131	€159	€128	€136	€133
Istanbul	€64	€57	€56	€59	€57	€55	€59
Johannesburg	€67	€60	€64	€66	€74	€76	€74
Kuala Lumpur	€50	€50	€48	€52	€52	€52	€52
Las Vegas	€126	€129	€104	€97	€101	€109	€109
Lima	€54	€45	€51	€49	€54	€54	€61
Lisbon	€121	€104	€86	€87	€70	€71	€82
London	€142	€134	€138	€148	€118	€123	€133
Madrid	€104	€110	€94	€96	€86	€91	€99
Manila	€54	€54	€56	€67	€67	€69	€72
Marrakech	€113	€134	€123	€124	€109	€113	€130
Melbourne	€129	€132	€129	€135	€121	€127	€134
Mexico City	€59	€64	€59	€56	€59	€65	€63
Miami	€98	€109	€116	€145	€146	€167	€156
Moscow	€66	€55	€51	€60	€60	€59	€56
Mumbai	€54	€55	€59	€67	€70	€81	€64
New York	€234	€231	€207	€237	€135	€130	€156
Paris	€167	€140	€123	€136	€125	€117	€122

Figure 10: Hotel price Details from trivago

Figure 11: Leading bnb details from statista