Curtin College

Bentley Campus

**Documentation and Report**

**Assignment 1: Airbnb Data Analysis**

Mamert Vonn G. Santelices

Student ID: 90026174

# User Documentation

## Reading CSV Files

At the start of the program, it reads the Airbnb data from a “Dataset.csv” found within the directory of the program. It follows the following format, respectively:

**Host\_Response\_Rate**: host response to booking requests; **Host\_Identity\_Verified,Host**: identity verified before booking/listing property; **Host\_Total\_Listings\_Count,Total**: number of properties listed by the host; **Unique\_City \_ID**; **Location\_Verified**: the location of property was exact as per the listing; **Property\_Type**: property type code; **Room\_Type**: room type code; **Max\_Accomodation**: maximum number of people allowed to stay in; **Bathrooms\_Count**: number of bathrooms on the property; **Bedrooms\_Count**: number of bedrooms on the property; **Beds\_Count**: number of beds in the bedroom; **Bed\_Type**: type of bed i.e. single, double, etc; **Amenities\_Price**; **Price**: total price of booking; **Minimum\_Nights,Minimum**: nights you can book the property; **Number\_of\_Reviews**: number of reviews on the property; **Review\_Scores\_Rating**: review score for the property; **Instant\_Bookable**: is property instant book able; **Cancellation\_Rate**: rate of cancellation for the bookings; **Reviews\_per\_Month**: number of reviews on the property per month; and **Fraud\_Count**: number of fake/fraud bookings on the property.

As per the use of CSV, the data is separated through a comma, and the first line is the header which the program swiftly skips through. If the file is not found within the directory, then the program stops any subsequent operations and exits the program. Once the etl.py module extracts the data, it is stored in an In-memory Database (in the RAM) through an array of AirbnbStore instances. The array handily stores the data in one heap of memory. Wherein, the program uses indexes to access the data, since the program keeps the data as immutable. Hence, when filtering data the program return an array of indexes pointed to the array of AirbnbStore instances, like pointers used in a lower-level language. Moreover, the index and the value are stored in the program’s indexed queries. Wherein, the query is sorted using numpy based on the QueryElement’s value. Any ‘filter with’ and ‘order by’ functionality then refers to this sorted query to compare and transform data in the form of array of indexes pointed to the array of AirbnbStore instances. Also, the query is immutable, and no changes are ever applied to the indexed queries or the AirbnbStore array.

## Displaying Menu

The menu.py is the programs Menu API. After extract is successfully called by the etl.py in main, the menu.py starts the program by entering a while-loop with a try-and-catch when the Menu prompts for the following:

## Adding Filters

The filters make use of the Query classes of the program. The following filters are available for use: (1) Bedroom Counts; (2) Price; (3) Minimum Nights; (4) Review Score Rating; and (5) Cancellation Rate. You can exit these options by inputting 6 (Go Back) option.

## Printing Names

This option prints the data or the filtered data on a table outputted within the console. The data is usually ordered based on the latest filter applied on the data.

## Graphing the Results

This option graphs the data or the filtered data as the x-axis using the prompted attributes from the following: (1) Price; (2) Bedroom Counts; (3) Property Type; and (4) Review Scores Rating. Then, the program displays two to three graphs, of which the following y-axis attributes are price, amenities price, and cancellation rate.

## Reset Filters

This option resets all the filters applied on the data.

The program is exited when 5 (exit) is inputted in the main-loop.

For the program to start, the user is required to be using a Linux OS, and Python3 with the numpy package and matplotlib package installed. Within the src directory of the program, enter this command: ./src$ python3 airbnb.py

Program Dependency Graph

console.py

Calls Menu.start() on main()

menu.py

airbnb.py

Uses console.log; console.log\_loading; console.log\_loading\_size\_of

Uses AirbnbETL.transform() and AirbnbETL.load() functions

The program’s algorithm takes advantage of using a bitset data type. Through a bitset, the program can iterate through an array of Query, list, or ndarray and mark the indexes it found as True. It then compares that bitset with another Query, list, or ndarray of indexes by using the index to access the bitset’s element and check for True, if so, then add that index to the output array of filtered data. Conveniently, since the queries are already sorted by value, the programs generate a bitset from the filtered indexes and uses that query to output an ordered data based on the query’s values. The algorithm has the performance efficiency of O(m+n), wherein m is the size of first Query, list, or ndarray, and n is the size of the other Query, list, or ndarray. The disadvantage is that the program uses a lot of memory, since each query and bitset is the size of the array of AirbnbStore.

Use of numpy also greatly optimizes the program, since the sort function uses mergesort with O(nlog(n)) performance. And, accessing an array is O(1), while accessing the element of a list is O(n), wherein n is the number of elements visited.

get\_data\_size of the array of AirbnbStore

Stores an array of AirbnbStore. Uses \_\_str\_\_() and print functions of class

Uses query.fiter() and query.order\_data() functions

Calls AirbnbETL.extract() on main()

store.py

etl.py

query.py

# Abstract

The Boston, Massachusetts has plentiful of thriving Airbnb businesses. This report aims to provide insight on establishing an Airbnb business with the Boston, Massachusetts area. Hence, observations are made on identifying the attributes that help to successful Airbnb rental. The analysed data uses the dataset found within Kaggle, a machine learning dataset website, for the Airbnb listings at Boston. Additionally, an in-house develop data analysing software is used on the dataset for visualising and modifying the data. As a result, it has been found that the at most an Airbnb rental shouldn’t have more than or equal to five bedrooms, and the highest price that can be charged is USD ~15000. This is the model showed to be followed by the Airbnb listings with a review score rating of more than 80 and a cancellation rate below an average of one. Also, it’s been observed the property type of code two, seven, and ten are showed to perform less on the dictated metrics.

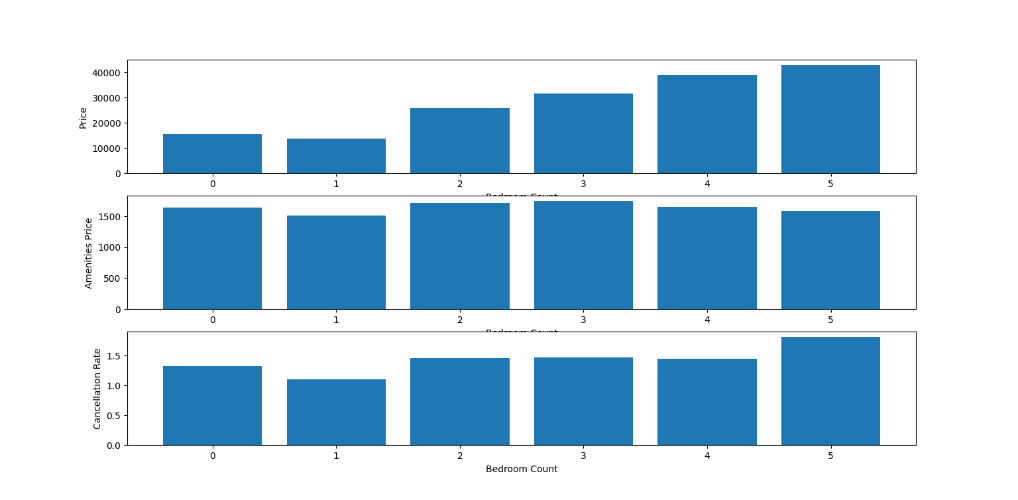
# Background

The Airbnb is $31 billion company, wherein its main source of profit is on allowing homeowners to earn a bit of extra money by renting (Aydin, 2019). On average, one entire home can earn $14,000 of annual earning (Aydin, 2019). The Boston Airbnb listings are analysed to identify the metrics that lead to a successful Airbnb listing. The observations made tries to create a correlation between several metric of the gathered data. Ideally, the observations can also identify factors that lead to a failing Airbnb listing.

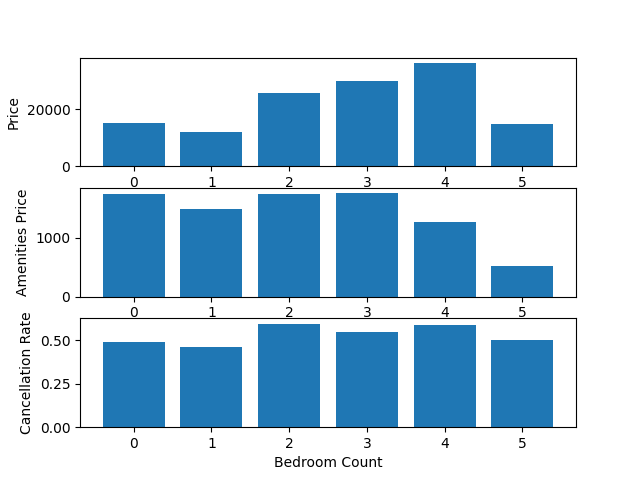
# Methodology

The Data Analysis uses the in-house developed Airbnb Data Analysis Software. The data used within the analysis was gathered by Kaggle from the Boston Airbnb listings. The software is a simple program that takes the average of price, amnesties price, and cancellations in order of a set x-axis. Thereafter, the data can be printed within the console or visualize by either a line graph or bar graph. The data analysis applied within the analysis is simple and doesn’t make use of variance or standard deviations. The program is limited without features, such as scatterplots and linear regression. All observations made are limited to be applicable only to the area of Boston, Massachusetts.

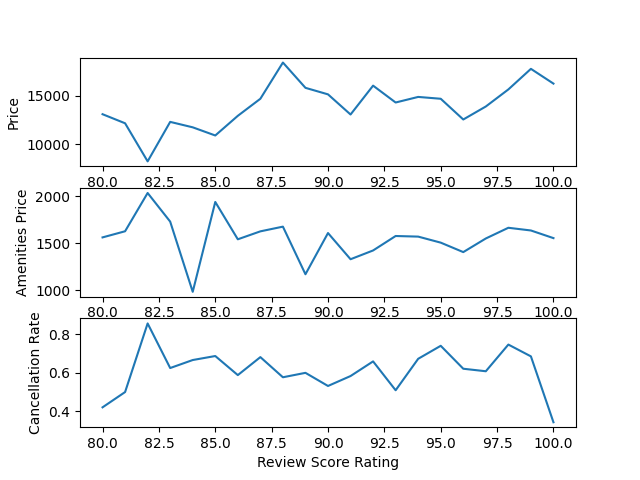
# Results

The table above shows that property type of code two and code seven are charging above average prices are also experiencing higher than average cancellation rates than the rest of the other properties. At least severe effect can also be observed from the property type of code ten. It is recommended that when listing a property for Airbnb to avoid properties of type code two, seven, and ten from the Boston Area.

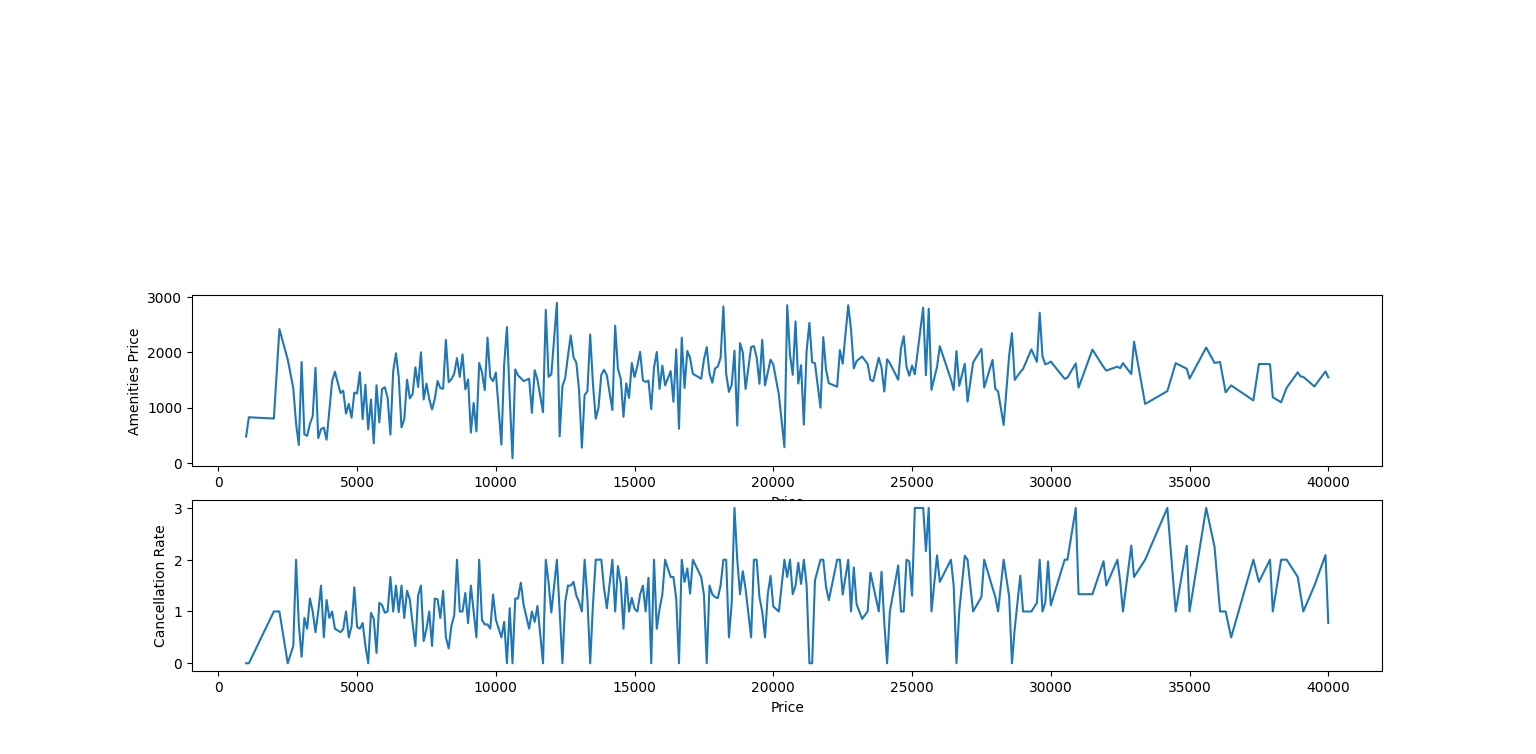
The table at the left shows that higher number of bedrooms allow higher rental prices.



On the other, filtering the data for only listings with above 80 score rating and an average of less than one cancellation rate show the table on the right. Based on the figures, we can infer that despite the higher charges that a five-bedroom count Airbnb listing are very likely to drop the score rating and with an observed higher cancellation rate.



Most successful Airbnb listing with a review score rating of higher than 80 and an average cancellation rate of less than one can charge on an average of USD ~15000 for rental and USD ~1500 for amenities prices.



The table shows that a higher prices front atleast USD ~15000 increases the average of cancellation rates and lowers the expected amenities price charges to customers

# Conclusions

When setting up a property to list to Airbnb within the Boston, Massachusetts is, the ideal bedroom count should be between 2 to 4 bedrooms. On the other hand, prices above USD ~15000 offers no benefits to the business. Since, five bedrooms and prices higher than USD ~15000 increases the likelihood of cancellations and decrease the expected amenities price that can be charged to customers. In addition, it had also been observed to reduce the review score rating of the establishment. Lastly, it has been observed to avoid property type of code two, seven, and ten because of the high average of cancellation rates and high prices.

# References

Kaggle. 2019. *Boston Airbnb Open Data*. Dataset. Publisher: Airbnb. https://www.kaggle.com/datasets/airbnb/boston.

Aydin, Rebecca. 2019. *How 3 Guys Turned Renting Air Mattresses in Their Apartment into a $31 Billion Company, Airbnb.* https://www.businessinsider.com/how-airbnb-was-founded-a-visual-history-2016-2.