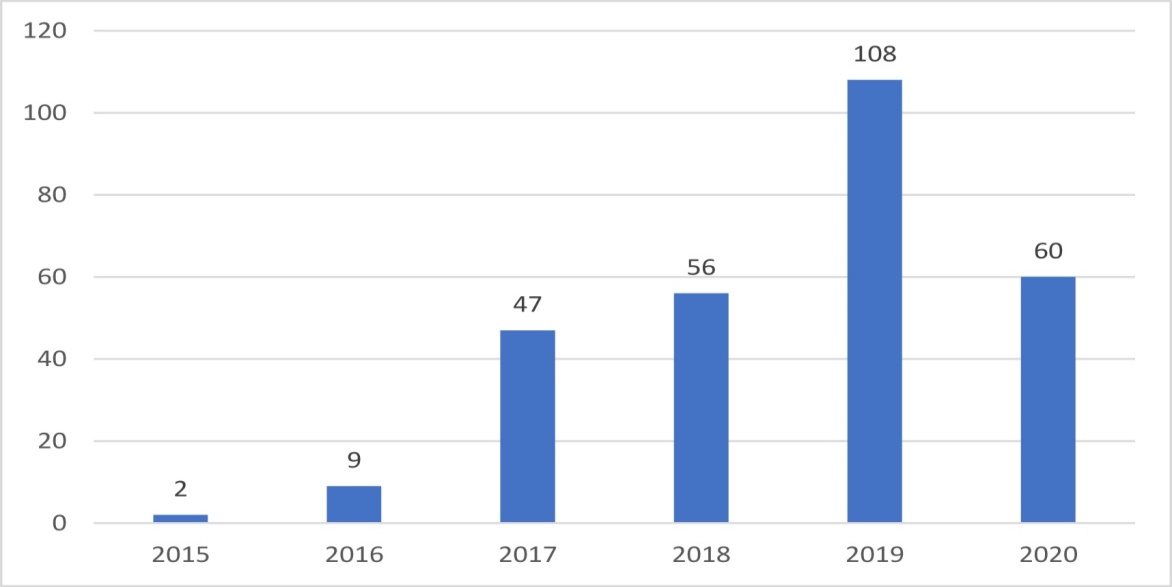
1. **Introduction**

In 2007, Airbnb - a company was established in 2007, whilst an important conference was held in San Francisco and then 3 roommates, saw an opportunity to generate a bit of income that could help them paying hire: they accommodated 3 convention attendees who couldn't discover any low-cost resort room in the city, offering them an airbed and breakfast. The name of company owners are B. Chesky, J. Gebbia and N. Blecharczyk. The idea of providing a list carrier for folks that had surplus area of their home, in order that they might hire out to travelers and tourists, changed into formally born. Airbnb is globally identified as the discoverers or inventors of the so-called “Sharing Economy” (additionally called “collaborative consumption”), constructed around the idea of peer-to-peer alternate of goods or services thru information technology, giving individuals the chance of renting access to their underused assets. Airbnb is essentially a two-sided Internet platform that works as a matchmaker among the hosts, the folks that provide a room or a whole house for short rental, and the visitors, travelers and tourists in search of a place to hire during their trip. The platform fees a carrier charge of 8% as much as 18% for every booking, to each the host and the guest, and it additionally has the role of making sure that transactions are secure and dependable for each events. Airbnb has reduced old-fashioned rational that barred property owners from sharing in the miniature leasing marketplace: first of all, it gives hosts the possibility of letting their hotels regarded to feasible visitors; secondly, it has resolved the friction associated to trust, performing as a assure among host and visitors. This characteristic is confident additionally thru the overview device at the platform: visitors can go away an overview inside 14 days after take a look at out and, to inspire unbiased and sincere comments, it's far published most effective after each event have finished their very own overview. This enterprise version has been identified as a disruptive force in the lodging enterprise (Guttentag, 2015), basically allowing small suppliers to compete with traditional providers such as hotels, most of all because of appealing prices (Lieber, 2011), however additionally due to numerous blessings that come from staying in a residential unit in preference to a resort room, including the chance for visitors to live an experience ‘like a local’, as the platform itself states on their website. However, many argue that Airbnb may now no longer be a direct competitor in the traditional lodging sector, as it has created a brand new institution of consumers, a “niche” of the traditional marketplace, encouraging tourists who may have now no longer been able to afford certain locations otherwise (Yglesias, 2012), and drawing them to neighborhoods and regions of the towns now no longer historically touristy.

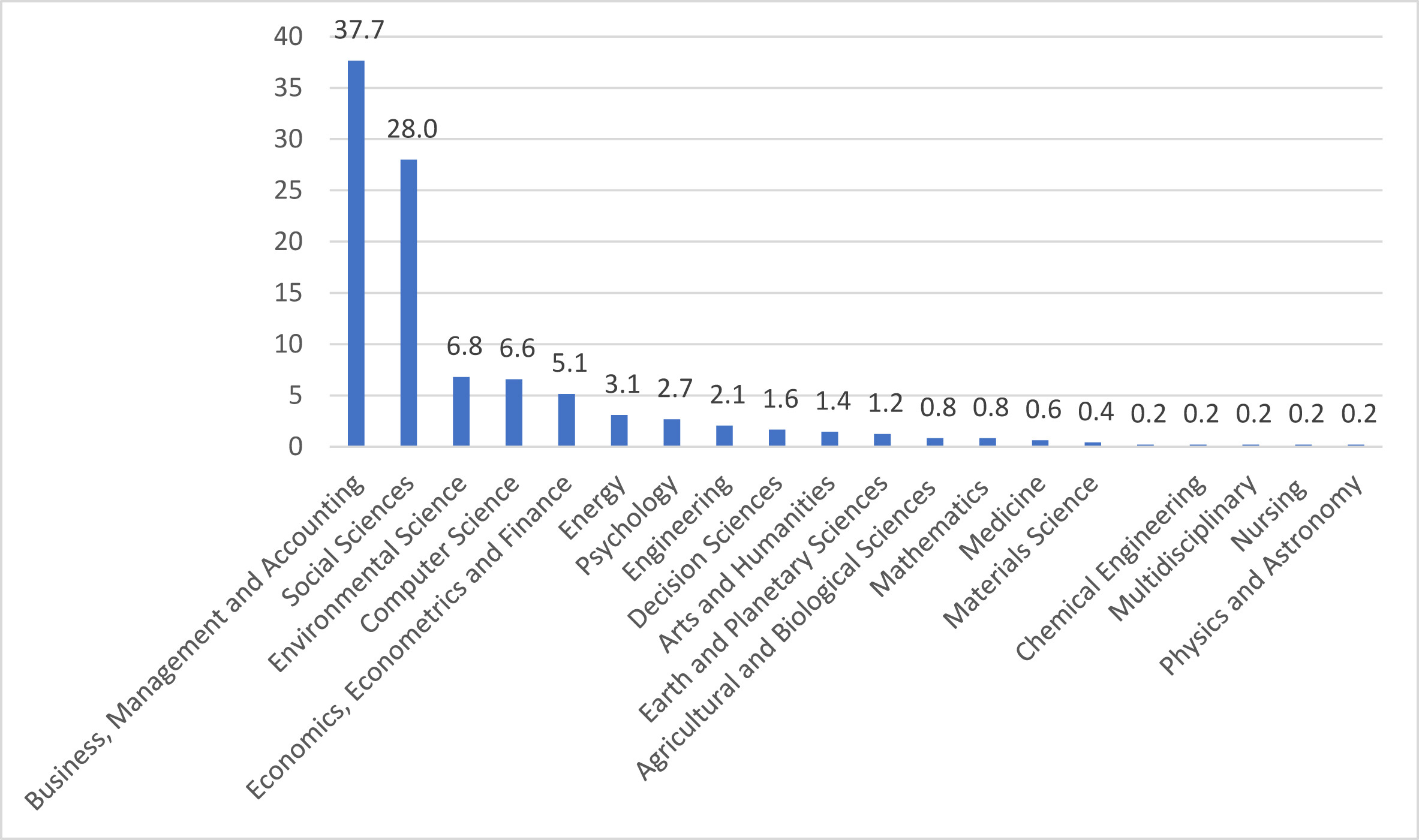
Since its growth in reputation over the past 10 years, Airbnb has been on the centre of many discussions of its possible bad externalities on towns regarding housing stress, especially focusing on its outcomes on gentrification and relocation of citizens. The main mechanism mentioned is the conversion of long-time period apartment gadgets into short-time period ones, because of the better profitability that domestic sharing platform can offer. This leads to a decrease in the variety of to be had gadgets within side the long-time period apartment market, which in the long run results in a general growth of rents. The battle for citizens to locate houses at sure places and the better variety of tourists crowding famous neighborhoods has been the focus of many papers and researches over the previous few years, with a number of those additionally introducing the hypothesis that those outcomes may also be meditated on residence charges with inside the actual property market (Gant, 2019). The Fig. 1 is shown below in which we can clearly conclude the numeral of research work published in a year.



**FIG-1: NO. OF RESEARCH PAPERS PUBLISHED BY EACH YEAR (HATI, 2021)**

* 1. **Home-Sharing Effects**

Analyzing the possible results of Home Sharing, the first one that comes to mind is the transformation of properties traditionally destined to long-time period condo, to tourist lodging on the quick-time condo marketplace: the extrade is simple for landlords, who don’t usually require huge investments on their properties, except in a few cases of renovation, even if they do require at least one person taking care of the place and of the guests’ 6 reception and eventual needs. If units are rented to the use that makes them extra profitable, the transfer must appear maximum while short-time period condo brings extra price to the proprietor as compared to the long-time period condo. If the transfer from a marketplace to the alternative happens for lots landlords at a sure location, this has a right away impact at the long-time period condo marketplace, wherein deliver lowers and the vacancy fees pass up, subsequently rising rents. However, there will also be a few externalities and indirect results to consider: domestic sharing draws a large range of vacationers to a neighborhood, growing the sales for neighborhood companies; this additionally will increase the range of available offerings in the area, making it extra attractive to stay in, consequently growing rents on each the quick and long-time period condo markets. However, if the range of vacationers is too elevated, this can lead to uncomfortable situations for citizens, making the neighborhood much less proper to stay in and consequently reducing rents. According to (Poterba, 1984), shopping for a residence may be regarded as shopping the existing price of future condo payments. In this sense, residence costs must be identical to the projected gift rate of rents for the same unit, accustomed for any tax insinuations, plagiarizing costs, preservation costs and physical decline. Kung and Barron, and (Proserpio, 2018) use this statements in their study to state that any impact domestic sharing has on long-time period condo fees, may be capitalized into residence costs. Moreover, domestic sharing additionally permits the proprietor to sell unused capacity on the short-time period marketplace, providing an additional potential income source: this may be regarded as a further growth of costs. Other results will be at the deliver of houses for sale, because the option to rent short-time period can also additionally affect owners’ propensity to determine to listing their houses for sales. The externalities can also be bad, as a few owners can be extra touchy to noisy neighbors than renters. The results of each rents and residence costs should then have an impact on the profitability fees, growing them as rents growth. Also, if a brand new profitability price is introduced, as the ratio of common annual short-time period rent (as opposed to long-time period) over common annual charge at a sure location, this additionally should have better values as compared to the identical index calculated with the long-time period rents, if short-time period rents absolutely are extra profitable. In Fig. 2, subject wise number of research papers published in AirBnB by analyzing different data as well as by applying different data mining techniques is shown.



**FIG-2: DISCIPLINE WISE NO. OF RESEARCH PUBLICATIONS ON AIRBNB (Hati, 2021)**

* 1. **Assignment Task**

As it is evident from Fig-1 and Fig-2, many authors have published different type of case studies using different data mining techniques like finding association between variables, anova, manova, clustering, decision trees as well as other different techniques and answered different type of questions depending on the domain of research.

* The question we need to find answer of in our assignment is that; whether user is permanently renting out the properties as hostels instead of occasionally?

1. **Data Analysis**

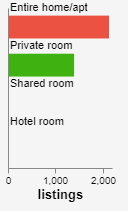
In order to answer the question of our assignment, first of all we are going to analyze the data of AirBnB which is provided to us. Data analysis is the process to find useful information out of data, by following the steps of cleaning, transforming and modeling data, on which business decisions can be made. May different data analysis methods are available like Regression Analysis, Factor Analysis, Neural Network and Data Mining. As we are more concerned with data mining in this assignment therefore I am going to elaborate it in detail.

Data mining is a way of statistical evaluation; this is the umbrella time span for designing measurements and experiences for added worth, course, and setting. By the use of exploratory factual assessment, an insight mining focuses to see conditions, relations, examples, and qualities to produce unrivaled information. While pondering how to investigate measurements, embracing an insights mining outlook is basic to progress - in that capacity, it's a spot this is all around definitely worth investigating in more detail. An exceptional use instance of measurements mining is datapine reasonable insights alarms. With the help of engineered insight and device learning, they offer electronic markers fundamentally founded absolutely on extraordinary guidelines or events inside a dataset. For instance, in the event that you're following convey chain KPIs, you might set a reasonable caution to cause while invalid or low-best measurements shows up. Thusly, you might be fit for drill down profound into the issue and interface it startlingly and actually. In the resulting picture, you could perceive how the reasonable alerts from datapine work. By placing in levels on step by step requests, meetings, and incomes, the cautions will tell you in the event that the aim turned out to be currently not generally gotten done or on the other hand assuming it passed the assumptions. Different available data mining techniques are clustering, data visualization, association, data warehousing, machine learning, data cleaning, classification, and prediction.

* 1. **Data Analysis of AirBnB**

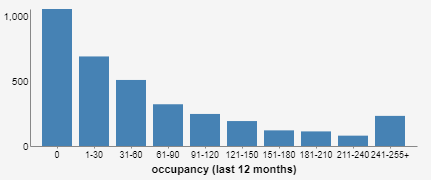
The dataset available in the file GM\_LISTING\_2021.csv contains the data about the listings, information of host as well features of property and the property reviews along with ratings. This dataset has 3,477 listings and contains 67 variables of each listing like listing\_id, host\_id, host\_name, bathrooms, bedrooms, beds, availability, min\_nights and max\_nights etc.

If we want to see the rental demand in Manchester, which is the value that how frequent, the rentals are booked through-out the year. The value of this variable is 59.8%. Moreover, the category of each rental is given in the form of chart-1 below. In the chart, it is clearly depicted that 2,054 are homes/apartments, 1,397 are private rooms and only 26 are shared rooms. By analyzing the following data we can generate opinion that most of the properties are permanently rented as hostels instead of rented occasionally but after analyzing more data, we will be able to generate opinion.



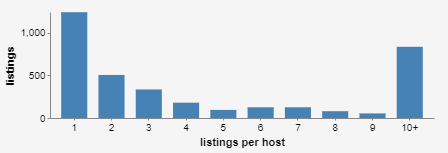
**CHART-1: NUMBER OF RENTALS PROPERTIES AS PER THE CATEGORY**

The analysis of minimum stay, price and review numbers are used for estimation of number of nights each room was booked and the generated income over the year. On average, each listing was booked for 63 nights, price per night was £143 and average income over the year is £6,071. In Chart-2, number of occupancy over the year is shown.



**Chart-2: OCCUPANCY OVER THE YEAR**

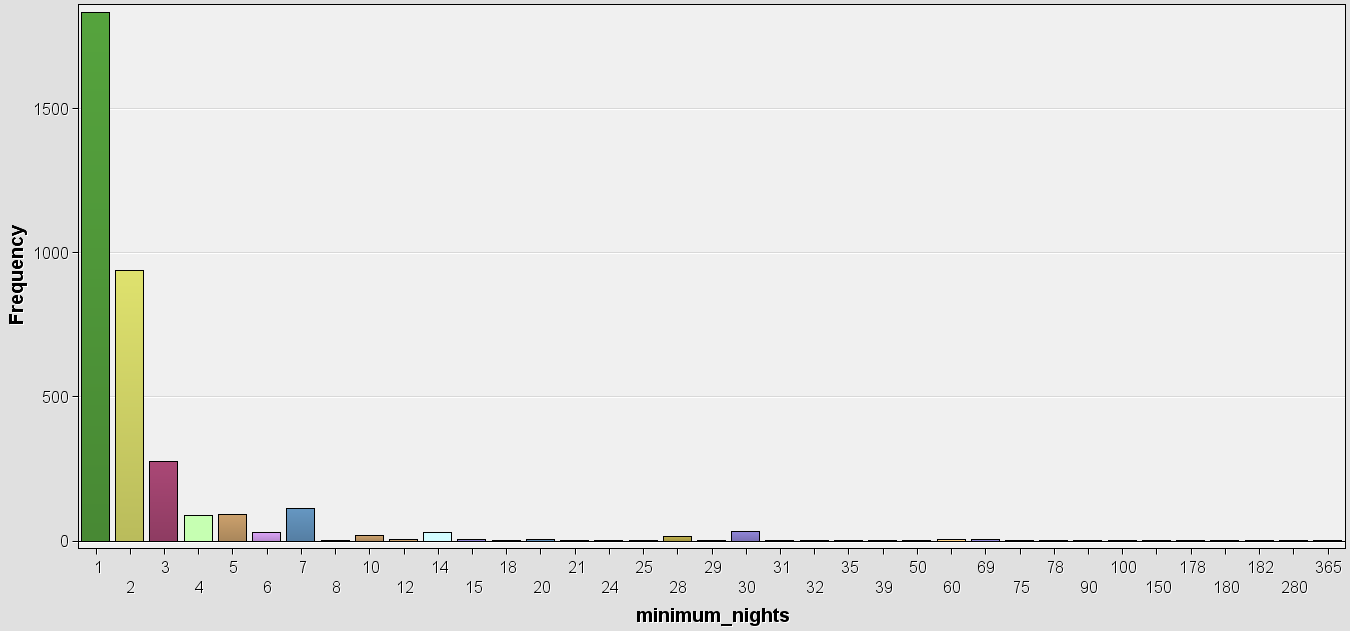
Some of the hosts of AirBnB advertise multiple listing like multiple available rooms in a sole apartment or multiple available apartments/houses. Hosts having multiple listing are most possibly running a business instead of renting out there residential property for short-term which is a violation of rule “short-term rental” which was designed to protect the residential housing in the cities or towns. After analyzing the data, it is evident that 65% hosts have multiple listings while only 35% people have advertised a single listing. Chart-3 analyzed number of listing per host.



**CHART-3: NUMBERS OF LISTING PER HOST IN SAS**

* 1. **Facts on Problem**

In order to protect the housing for residents, policy is to maintain short-term rentals as it is also the laws of cities and towns. It is also our top priority in this assignment too so here we are going to analyze the **minimum nights** variable. By looking at this variable, we can generate opinion about the said question that whether market is shifted towards long term stays. It is evident from the data that 96% listings, which are 3338, are on short-term rentals and remaining 4%, which are 139, are on long-term rentals. Chart-4 is showing minimum nights each listing is rented out and can be used for justification that 96% listings are on short-term rental and not violating any rules.



**CHART-4: MINIMUM NIGHTS EACH LISTING IS RENTED OUT IN SAS**

* 1. **Cluster Analysis**

Clustering or one can say cluster analysis in the task to group together the data which has similar properties or traits. It is statistical method and groups the data by finding the association between the data. Mainly a distance function is used to assess the similarity or differences and many other clustering algorithms which are based on different concept are available (Boccard, 2013). Similarity measures are first computed among observations and among clusters as soon as observations start to be assembled into groups or collections (Clusters). Numerous measures or metrics like Euclidean distance as well as correlation measure which can be used to compute the similarity for clusters, other measures can also be used for the same purpose like Manhattan distance or mutual information. Furthermore, numerous integration techniques which result in unique grouping styles are probable. Clustering outcomes are consequently extremely particular, as they suggestively be dependent upon the handlers’ adoptions. Old-fashioned clusters or groups’ evaluation is typically achieved to organization both interpretations and variables one by one however synchronized co-clustering (or bi-clustering) of the rows and the columns of the facts matrix establishes additionally an appropriate opportunity to look for biomarkers.

* + 1. **K- Means Clustering Algorithm**

K-Means Clustering is an unaided learning set of rules, which organizations the unlabeled dataset into stand-out groups. Here K characterizes the wide assortment of pre-portrayed bunches that need to be made withinside the cycle, like K=2, there can be groups, and for K=3, there can be 3 groups, etc. It licenses us to bunch the insights into exceptional organizations and a convenient way to figure out the kinds of offices withinside the unlabeled or unmarked dataset on its own special without the need for any preparation or training. It is a centroid-basically based completely set of rules, wherein each bunch is connected with a centroid. The preeminent objective of this arrangement of rules is to restrict the amount of distances among the insights factor and their relating bunches.

The k-way clustering set of rules specifically plays tasks:

1. Concludes the satisfactory fee for K middle factors or centroids with the aid of using an iterative process.
2. Assigns every statistics factor to its nearby k-middle. Those statistics factors that are close to the exclusive k-middle, create a group.

Hence every cluster has datum with a few cohesions, and it's miles farfar from different groups.

The running of the K-Means set of rules is characterized withinside the under advances:

Step-one: Select the amount K to decide the complete bunches.

Step-two: Select arbitrary K elements or centroids. (It could be not the same as the enter dataset).

Step-three: Allocate each record element to their nearest centroid, with an end goal to shape the predefined K groups.

Step-four: Compute the difference and area a spic and span centroid of each and every group.

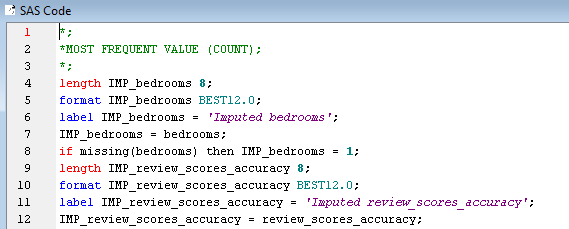
Step-five: Repeat the 1/3 stages, this implies that reassign each datum highlight the fresh out of the box new nearest centroid of each and every bunch.

Step-six: If any reassignment happens, then, at that point, visit step-four else visit step-seven.

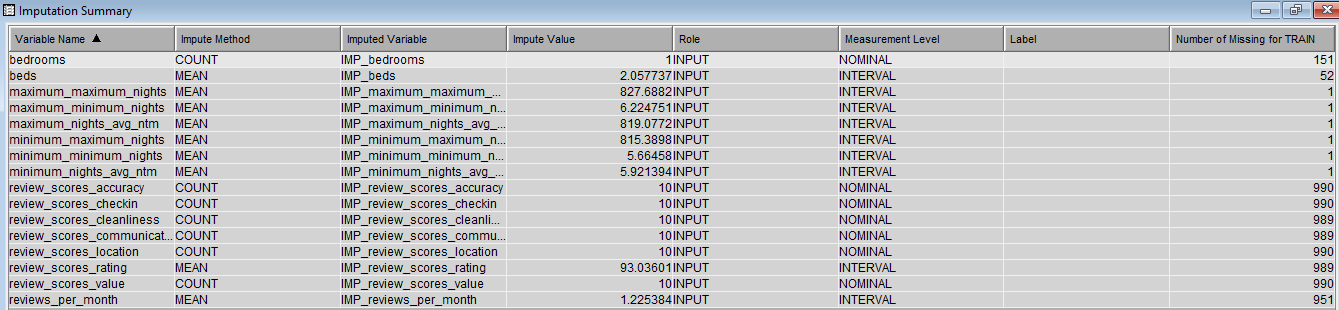
Step-seven: The form is prepared.

* + 1. **Data Cleaning**

Missing information is an appalling truth of data and statistics. The main problem in this data is also the missing values of columns in the data. Nonetheless, there are different ways of assessing and manage missing information is list-wise deletion (also known as complete case analysis), regression imputation, stochastic regression imputation, maximum likelihood, and multiple imputations. We have used regression imputation in SAS. Regression imputation (additionally referred to as provisional imply imputation) seals lacking values with anticipated values which are spawned after a regression equation. Varying values have a tendency to be associated so it marks spontaneous feel to apply this statistics to seal in lacking values. Nonetheless, this technique may be prejudice. It exaggerates the association among variables; also additionally take too lightly the variety of the data. In order to try and solve those issues, stochastic regression additionally makes use of regression equations to are expecting lacking data, however it additionally provides a commonly dispensed residual time period in an try and make amends for the herbal changeability in the information. In actual, this technique has the very best capability of being impartial out of the strategies mentioned so far. Nonetheless, this technique also can lessen the usual mistakes that may cause a better hazard of kind I error. In addition it does now no longer replicate the uncertainty of the lacking values. The code for imputation in the data is given in Fig-3 and summary of imputation is shown in Fig-4.



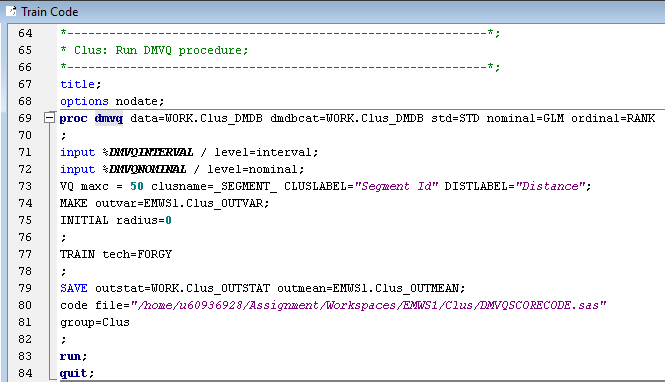
**FIG-3: CODE FOR IMPUTATION**



**FIG-4: SUMMARY OF REGRESSION IMPUTATION IN SAS**

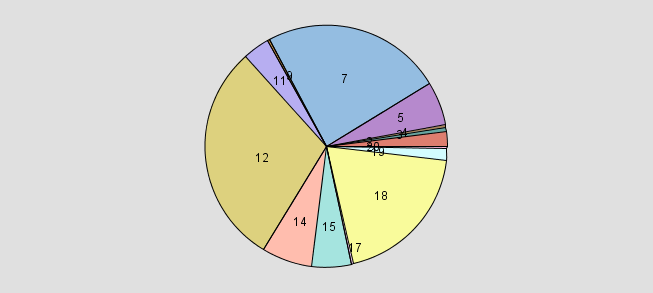
* + 1. **Clustering**

The code for applying clustering in SAS is shown in Fig-5.



**FIG-5: CODE FOR CLUSTERING**

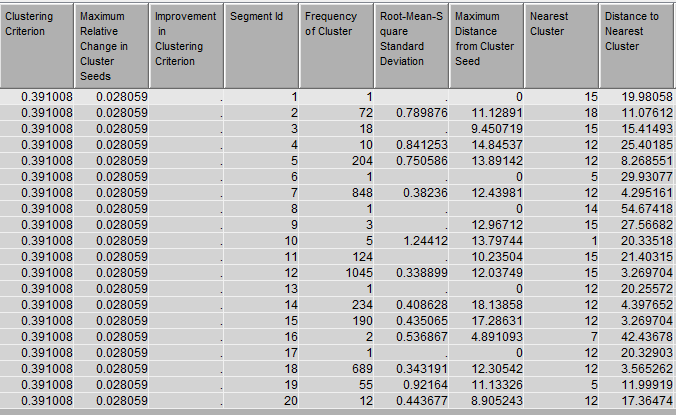
The Chart-5 below actually states the frequency of each cluster of the data. Total number of clusters which were developed are 48, and frequency of each cluster is actually showing in the form of color.



**CHART-5: FREQUENCY OF CLUSTERS IN SAS**

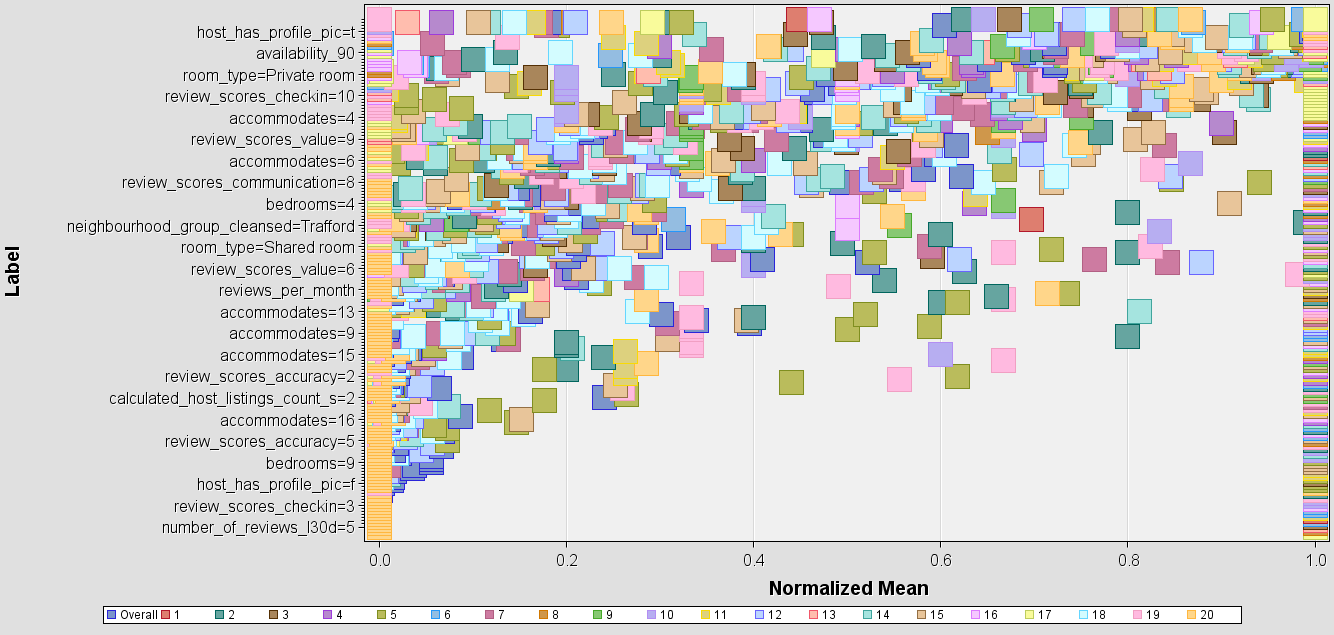
If we analyzed the cluster from Fig-6 which is showing mean statistics, it is shown that frequency of each clusters varies from 1 to 1045 means the minimum listing in each cluster is 1 and maximum listings in clusters are 1045. The standard deviation in each cluster and other variables like Maximum distance from cluster seed, Nearest cluster and distance to nearest cluster clearly depicting that clusters have less intra class variations and greater inter class variations which is the indication of great clusters.

However, the clusters which have only one listing can be treated as outliers and must not be considered for generation of model. That means, if we consider the clusters with only one listing as outliers, then we can say that we have five outliers listing in our data. In measurements, an exception is a records factor that contrasts significantly from various perceptions. An exception can be a result of changeability withinside the size or it can recommend exploratory mistake; the last option are once in a while rejected from the records set. An exception can reason outrageous issues in factual examinations.



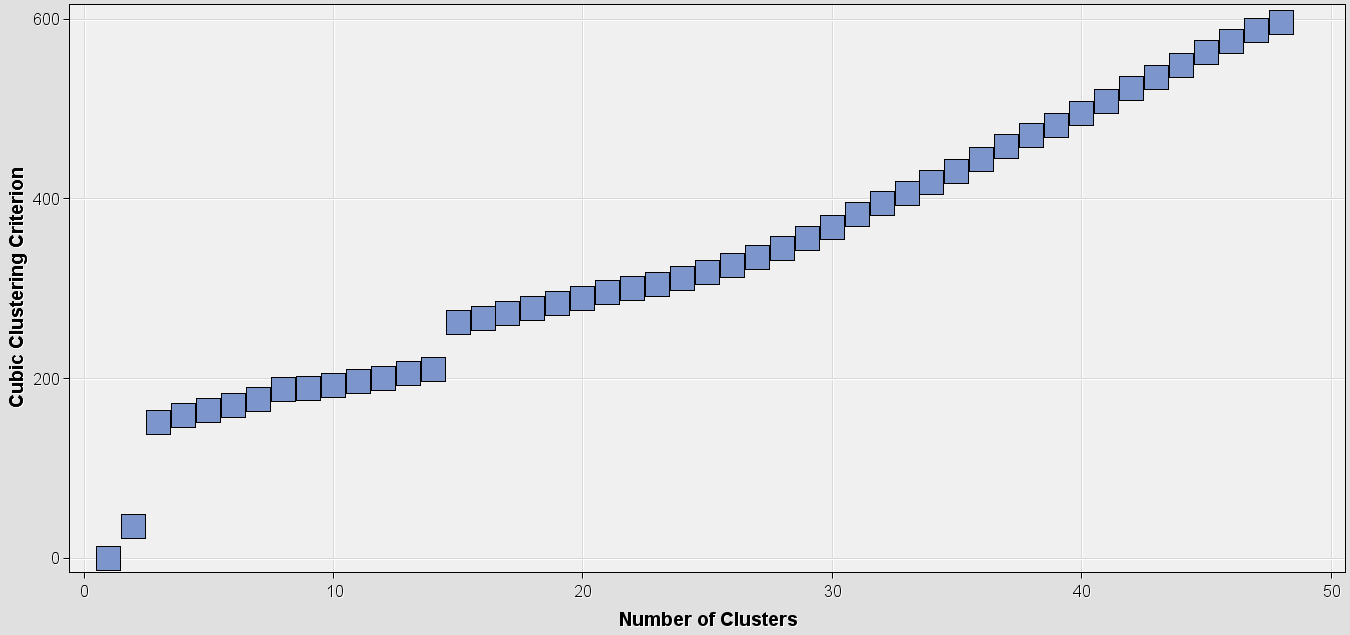
**FIG-6: MEAN STATISTICS SAS TABLE**

The input means’ plot on clusters is shown in Fig-7.



**CHART-6: INPUT MEANS PLOT IN SAS**

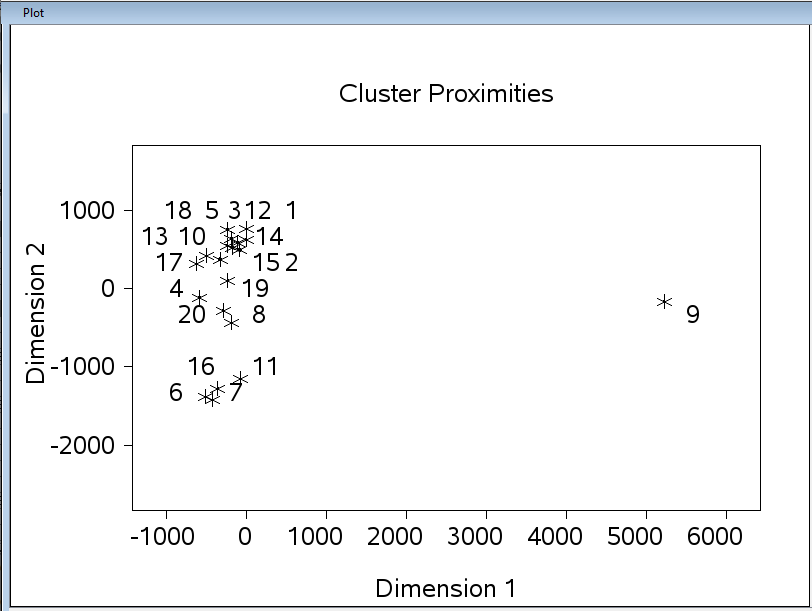
The (Cubic Clustering Criterion) CCC plot of each cluster is shown in Fig-5. On the horizontal side, number of cluster is shown however cubic clustering criterion is shown in vertical side. Cubic clustering criterion is may be used to estimate the wide variety of clusters the usage of Ward's minimal variance method, k -means, or different techniques primarily based totally on minimizing the within-cluster sum of squares.



**CHART-7: CCC PLOT OF CLUSTERS**

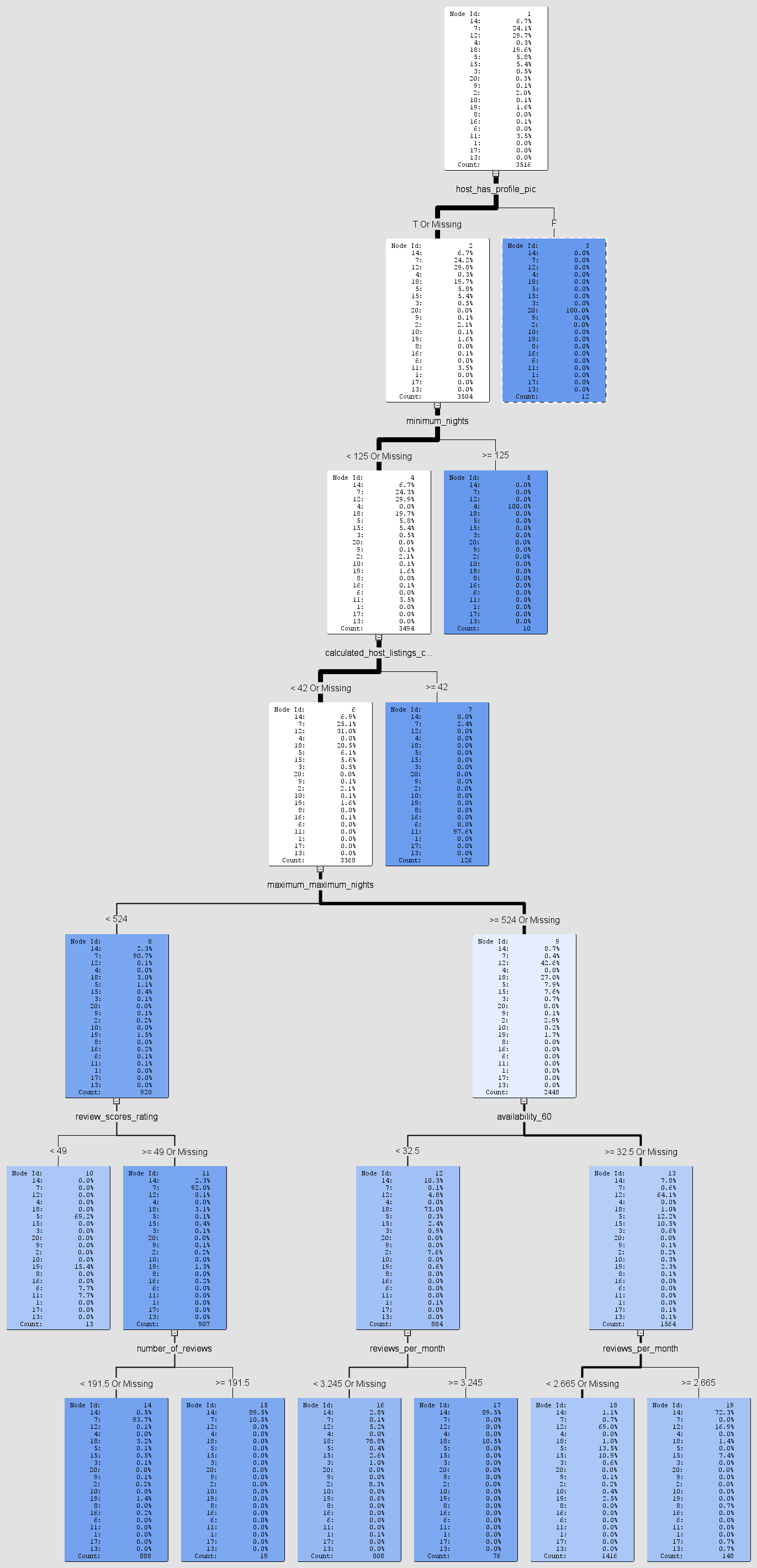
* + 1. **EVALUATION AND ANALYSIS**

The distance on which clusters were made is shown and plotted in Chart-8.

1. 

**CHART-8: CLUSTERS DISTANCE PLOT**

The tree diagram on the basis of which clusters were developed is shown in Fig-7. By looking at tree diagram, we can generate the rules on which the clustering was done on our data and very useful to analyze the complete picture of data.



**FIG-7: TREE DIAGRAM OF CLUSTERS**

The model tried to classify into 48 different groups and the rules on which the clustering was made is shown in Fig-7. On the basis of analysis it is observed that data is very useful and can be used to generate many different opinions like outliers available in the data, and business decision about hosts, there properties as well as the sentiment analysis of property listings. Short-term rentals are making more money for hosts but they have concerned about the vanishing of bookings or less bookings. Moreover, the decision which was the main objective of this assignment is also answered during the analysis phase that 96% listing are on short term rentals instead of long term. Moreover, 65% of the hosts have more than one property listing.

* 1. **Conclusion**

In this assignment, we have analyzed the data related to AirBnB for find the answers of different questions by using different data analysis and data mining methods. Initially we started with the introduction of AirBnB business and its evaluation and impact on the society. Then the analysis was conducted of different variables and there correlations were interpreted.

Cluster analysis can be used as a preprocessing step for labeling the data, features extractions or classification as well a stand-alone analysis for analyzing distribution of data, features’ examination as well as for analyzing the clusters individually. Better clusters are those which have less intra-class similarity and more inter-class similarity. We have used K-Means clustering concept for portioning of data with value of K set to 50 so that more clusters can be made and difference should be minimized.

The important question of assignment, whether the property is rental on short-term basis or as a permanent hostel, was also answered here. The main aim of this question was to protect the citizens and towns’ residence laws. The 96% property or housing listings on AirBnB are on short-term rental basis while only 4% listings are currently rented on a long-term residence or hostel basis which.

**References**

* Guttentag D., “Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector”, Routledge, Abingdon, 2015.
* [Matthew Yglesias](https://www.simonandschuster.com/authors/Matthew-Yglesias/85860222)., “The Rent Is Too Damn High”. ISBN13: 9781451663297
* Ron Lieber, “Airbnb’s Lodging Gets Tested, Yielding a Mixed Bag”. Nov1 11, 2011.
* Cocola-Gant A, Gago A. Airbnb, buy-to-let investment and tourism-driven displacement: A case study in Lisbon. *Environment and Planning A: Economy and Space*. 2021;53(7):1671-1688. doi:[10.1177/0308518X19869012](https://doi.org/10.1177/0308518X19869012)
* Poterba J., “Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach”, MIT Press, Cambridge, Massachusetts, 1984
* Barron K., Kung E., Proserpio D., “The Effects of Home-Sharing on House Prices and Rents: Evidence from Airbnb”, Social Science Research Network, 2018.
* Mead R., “The Airbnb Invasion of Barcelona”, The New Yorker, New York, 2019.
* Zukin S., “Landscapes of Power: From Detroit to Disney World”, University of California Press, Berkeley, 1993.
* Sorkin M., “Variations on a Theme Park: The New American City and the End of Public Space”, Macmillan Publishers, New York, 1992.
* Törnberg P., “Dark Disneyfication: Staging Authenticity on Airbnb”, Centre for Urban Studies, Amsterdam, 2019
* [Dann, D.](https://www.emerald.com/insight/search?q=David%20Dann), [Teubner, T.](https://www.emerald.com/insight/search?q=Timm%20Teubner) and [Weinhardt, C.](https://www.emerald.com/insight/search?q=Christof%20Weinhardt) (2019), "Poster child and guinea pig – insights from a structured literature review on Airbnb", [*International Journal of Contemporary Hospitality Management*](https://www.emerald.com/insight/publication/issn/0959-6119), Vol. 31 No. 1, pp. 427-473. <https://doi.org/10.1108/IJCHM-03-2018-0186>
* Hijrah Hati, Sri & Balqiah, Ezni & Hananto, Arga & Yuliati, Elevita. (2021). A decade of systematic literature review on Airbnb: the sharing economy from a multiple stakeholder perspective. Heliyon. 7. e08222. 10.1016/j.heliyon.2021.e08222.
* Julien Boccard, Serge Rudaz, in [Proteomic and Metabolomic Approaches to Biomarker Discovery](https://www.sciencedirect.com/book/9780123944467/proteomic-and-metabolomic-approaches-to-biomarker-discovery), 2013