**DATA MANAGEMENT PROJECT REPORT**

(Project Semester August-December 2020)

***ANALYZING HOTEL DATA FROM MAKEMYTRIP.COM***

Submitted by

Mansi Dattatray Kamthane

Registration No: 11802354

Section: KM077

Roll No: A16

Course Code : INT217

Under the Guidance of

**Miss. Komal Arora : 17783**

**Discipline of CSE/IT**

**Lovely School of computer science**

**Lovely Professional University, Phagwara**



**CERTIFICATE**

This is to certify that Mansi Dattatray Kamthane bearing Registration no. **11802354** has completed INT 217 project titled, “**Analyzing hotel data from makemytrip.com”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science**

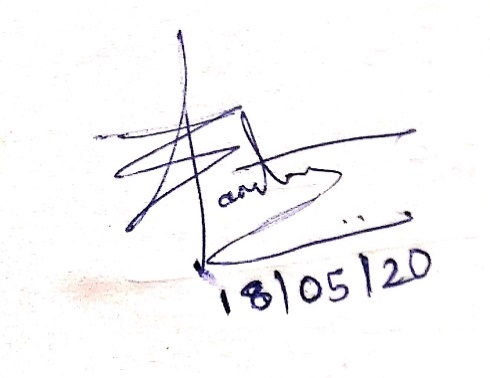
Lovely Professional University

Phagwara, Punjab.

Date: 18/11/2020

**DECLARATION**

I, Mansi Dattatray Kamthane, student of B.Tech under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 18/11/2020 

Registration No. 11802354 Mansi Kamthane

**ACKNOWLEDGEMENT**

I would like to express my special thanks and gratitude to the subject teacher **Miss. Komal Arora** who gave the opportunity and fruitful guidance to do this project on topic of “*Analysing hotel data from makemytrip.com*”. She guided me in preparation of the project at every interval of time and is project in charge. Finally, a word of gratitude to our family and friends who were always there with their support and encouragement.

By doing this project, I got enriched with many information which can help me in upcoming future projects. This Project shows up my Excel skills.

**Mansi Kamthane**

**TABLE OF CONTENT**

1. **Introduction**
2. **Objectives**
3. **Source of dataset**
4. **ETL Process**
5. **Analysis on dataset (for each)**
   * + **Introduction**
     + **General Description**
     + **Specific Requirements**
     + **Analysis Result**
     + **Visualization**
6. **List of Analysis with results**
7. **References**
8. **Bibliography**

**INTRODUCTION**

This project is based on analysis and visualization which is done with the powerful tool Microsoft Excel. In this project the dataset is analysed based on property type and its different attributes. Source of data is mentioned ahead and downloaded from Kaggle website. This analysis makes sure to bring different insights together and have respective objectives for growth of the business. In this project tableau is used for the concern of advanced visualization. The data is cleaned and filtered along with its manipulation for better results.

The use of Excel is widespread in the industry. It is very powerful data analysis tool and almost big and small business use excel in their daily functioning. Different types of formatting options are made available with easy calculations and graphics. Charts are been used for better visualizations.

Tableau drag and drop interface allows analysis quickly and interesting updates are made available along with functionalities of plotting unique graphs used in the concern of relevant information. It makes easy for preparation of dashboards and 3d graphs. Connecting to different data sources become easier and more effective.

Hence, along with the visualization this project makes sure to grasp the attention towards each objective holding pivot tables and pivot charts for user consideration. Screenshots of visualisations are inserted to make the report more effective.

**OBJECTIVES**

1. Finding available property type in respective cities along with customer reviews.
2. Summarizing the total number of property type, especially hotels in respective city.
3. Finding frequency of hotels according to stars and filtered by city as well as value given by people.
4. Analyzing the value of property type as review of people to look over the highlights.
5. Analyzing the outliers based on longitudes and latitudes along with layer of hotel ratings and summarizing the hotel summations in country.

**SOURCE OF DATASET**

The data is imported from [hotels-on-makemytrip](https://www.kaggle.com/PromptCloudHQ/hotels-on-makemytrip)

The above data is imported from Kaggle website. Downloading the data on local device and working on it is expected.

Link from local system for whole [**PROJECT**](https://d.docs.live.net/b4ae5011f6862f5e/Documents/project%20hotel.xlsx) in Excel

The data from the source is unfiltered and uncleaned with lost of glitches in it. It is necessary to clean the data and columns as per the convenience. Some of the cleaning is mentioned:

1. Nearly 10k duplicate rows have been deleted

2. co-ordinates which were wrongly placed are updated

3. count of reviews from different sources have been summarized

4. reviews of excellent, good and much more have been categorized from one column to other

5. room type are summarized in specific category

6. blank values are either filled with average or 0

7.invalid text data have been filtered and replaced wherever necessary.

8. unwanted data and columns are deleted like image URL column.

**ETL PROCESS**

ETL defines within itself- Extract – Transform – Load

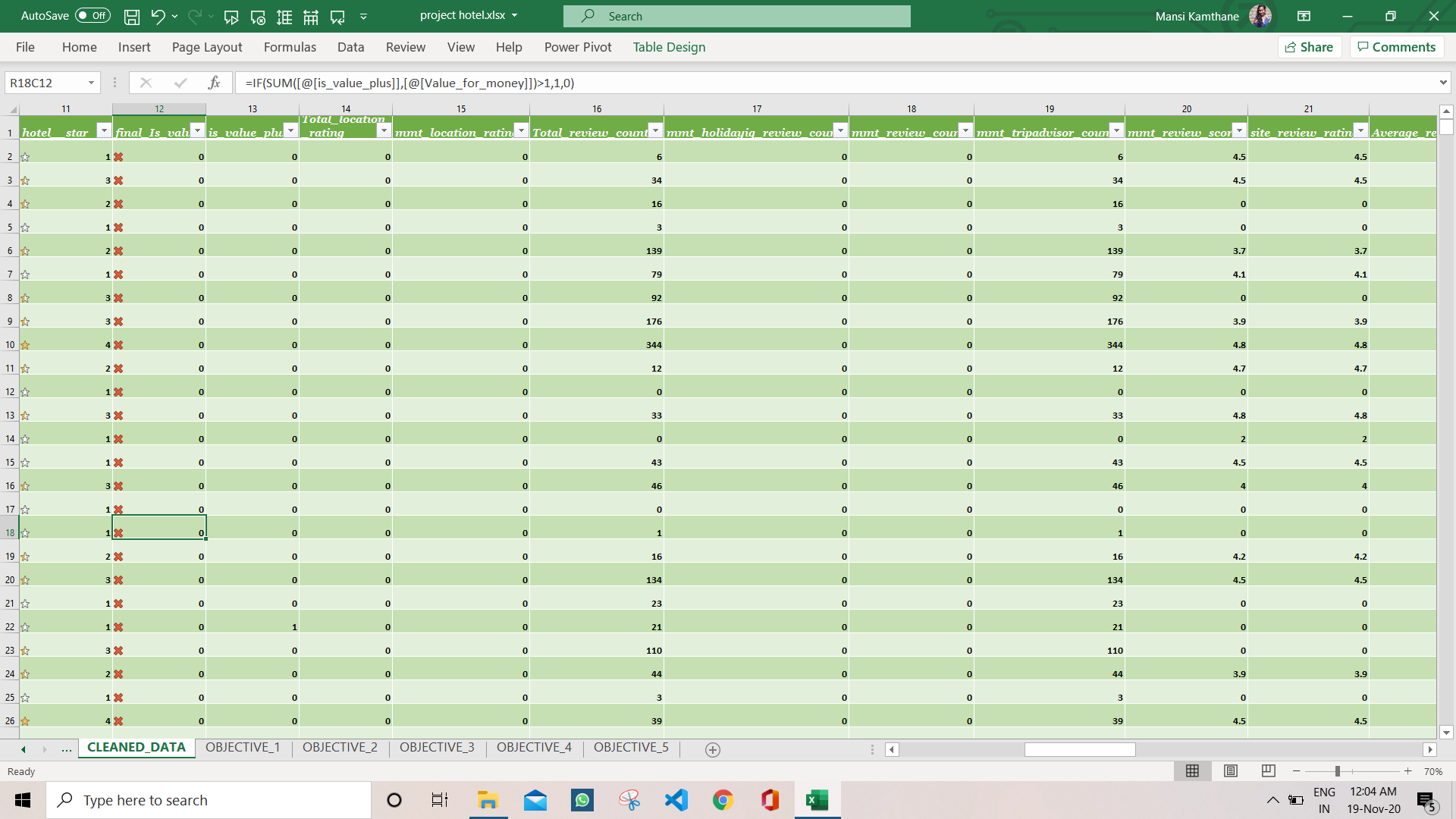
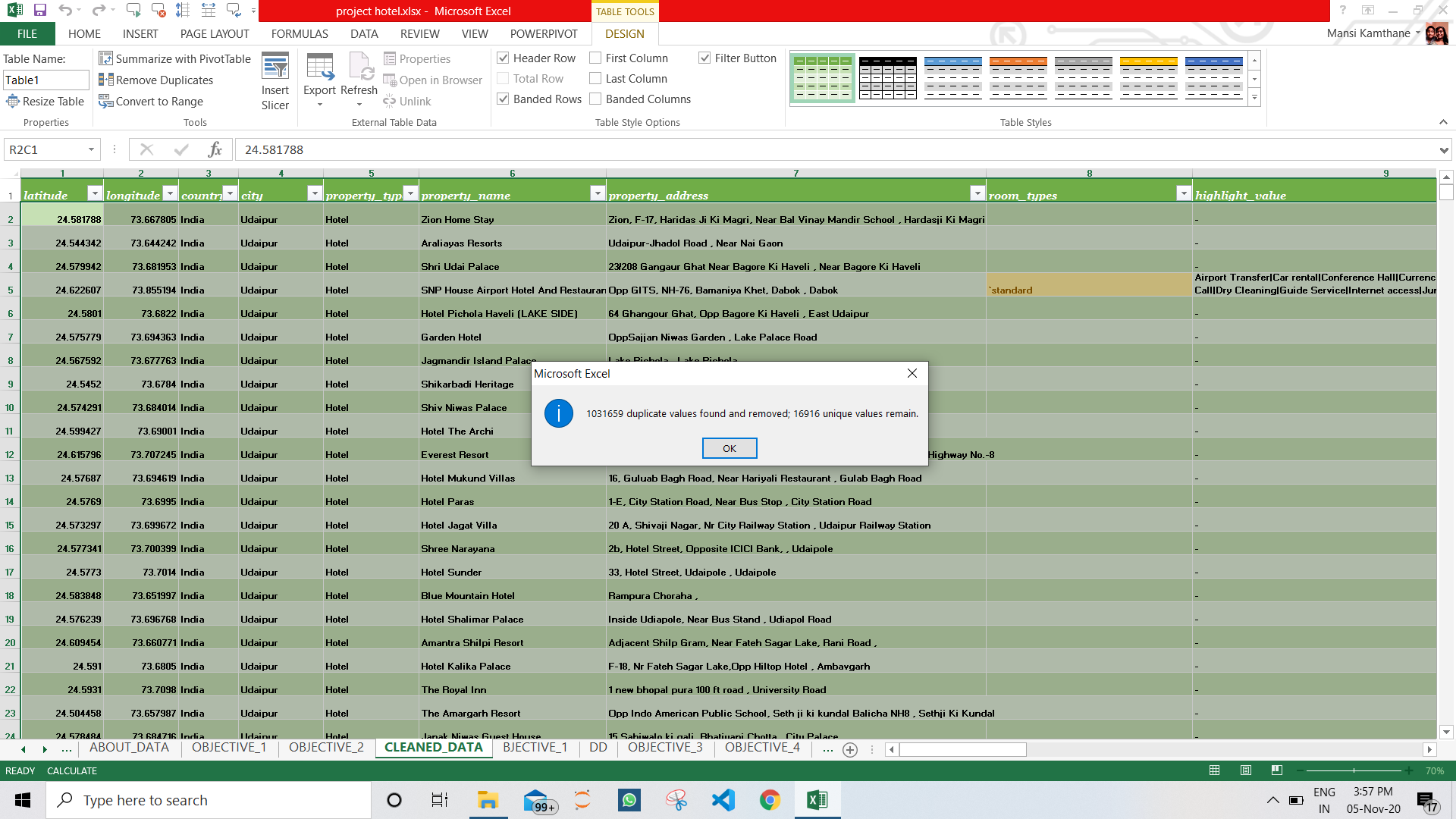
It is process that extracts the data from different source, then transformation of data which includes calculations, manipulation, and the filtering process. Finally, the data loading part which is ready for the working part and then the visualisations can be made from it.

In this project, data is extracted from the above source and some steps are to be performed as mentioned in above page for cleaning the data. Data is transformed into cleaned data by transforming some columns into text by defined separator by merging the columns using calculations such as average or summation. Many formulas such as VLOOKUP and Lookup are used along with pivot tables which completely transforms the data. Exporting of CSV data is done into Excel file and then later on as per the requirement the data is loaded into tableau for better visualisation.

The ETL Process finds the route to simplification from complicated and garbage data. The Main part is played on the transformation of data which leads to information from data and helps to finds the different insights in consideration with the growth of the business. Analysis should be done only after ETL process.

The Excel Project file consists the transformed data and the untransformed data or the data before ETL process can be found in CSV file.

**CLEANED DATA**

****

**ANALYSIS ON DATASET**

There are a variety of methods that may be utilized to analyze data. Many statistical packages are available, including Microsoft Excel, which is free and can often be used for simple, efficient analysis. Tableau prep and Tableau Desktop has additional functionalities for the analysis part.

With large data sets, manually counting or using a formula to count can be tedious and create opportunities for error. Pivot Tables will automatically sort data and list values, producing efficient and accurate information. To create a Pivot Table, select the data, click on the “Insert” tab, and then select “Pivot Table.”

Different analysis is done for different objective along with the charts and the visualizations. The description and other details will be given ahead. Every objective has specific insights for the data which results into beneficial analysis result and contributes towards the project aim.

For analyzing the data many functions as well as different types of charts are used. Data validation is used along with the filters. The sheets are protected, and user can make certain changes as per the providence of the ability to access the sheets as well as workbook. Sparklines are used to indicate the rising and lowering of data in certain columns. Conditional formatting as well as other fonts are used for better recognition. Tables are made wherever necessary as well as columns are sliced down as per requirement. Use of all tabs have been done in Excel such as Data, View, Review, Home, Insert, Power Pivot, Page Layout and many more.

**Finding available property type in respective cities along with customer reviews.**

1. **INTRODUCTION:**

With the help of columns property type, city and total reviews which are categorized into Excellent, Very Good, Average, Poor Terrible the visualization was made.

This is in need for the insight of the objective to make it clear and understand the criteria behind this. Growth can be considered if the other property type rather than hotel are taken into demand.

For the best results of the reviews and improve this, the steps are considered on the basis of the highlight values or the facilities they provide.

One can have hold as per the cities too where metropolitan cities shows great growth and better review too for the business of hotels. The highlighted values in these cities are satisfying and are well mannered which encourages customers for satisfaction.

The highlighted values and facilities can be found by lookup formulas and many other factors. The data is prior cleaned with required columns and then the task is to be taken ahead. Different colors are used for readers enthusiasm. Whereas the data is filtered by chart as well as pivot table. Slicers are added for user friendly data access and view the required data as per the user convenience. The 5 categories of reviews are initially modified and replaced by certain formulas and words which determine their existent in different directions.

1. **GENERAL DESCRIPTION:**

Initially the pivot chart is made by choosing the required columns. Here, there is need of selected columns that are city, property type and reviews by the people.

There is only one column of review where all type of reviews is mentioned together in single column. To simplify these complicate reviews, the data of that column is transformed into column to text option based on separator | which creates 5 columns with different reviews as Excellent, Very Good, Average, Poor, Terrible.

After getting all required data pivot chart is made along with the pivot chart. The sparklines are made based on pivot table summation calculation which shows rapid increase and decrease in the data of that review. Conditional formatting is applied where highest values are highlighted which shows the value heavily affecting the results.

Pivot chart is enabled with labels as well as specific bar graph is chosen to plot on plot area with filter of city. Design is changed with shadow, glow, and other effects. Objective is mentioned in the text box inserted which signifies the plotting and visualization.

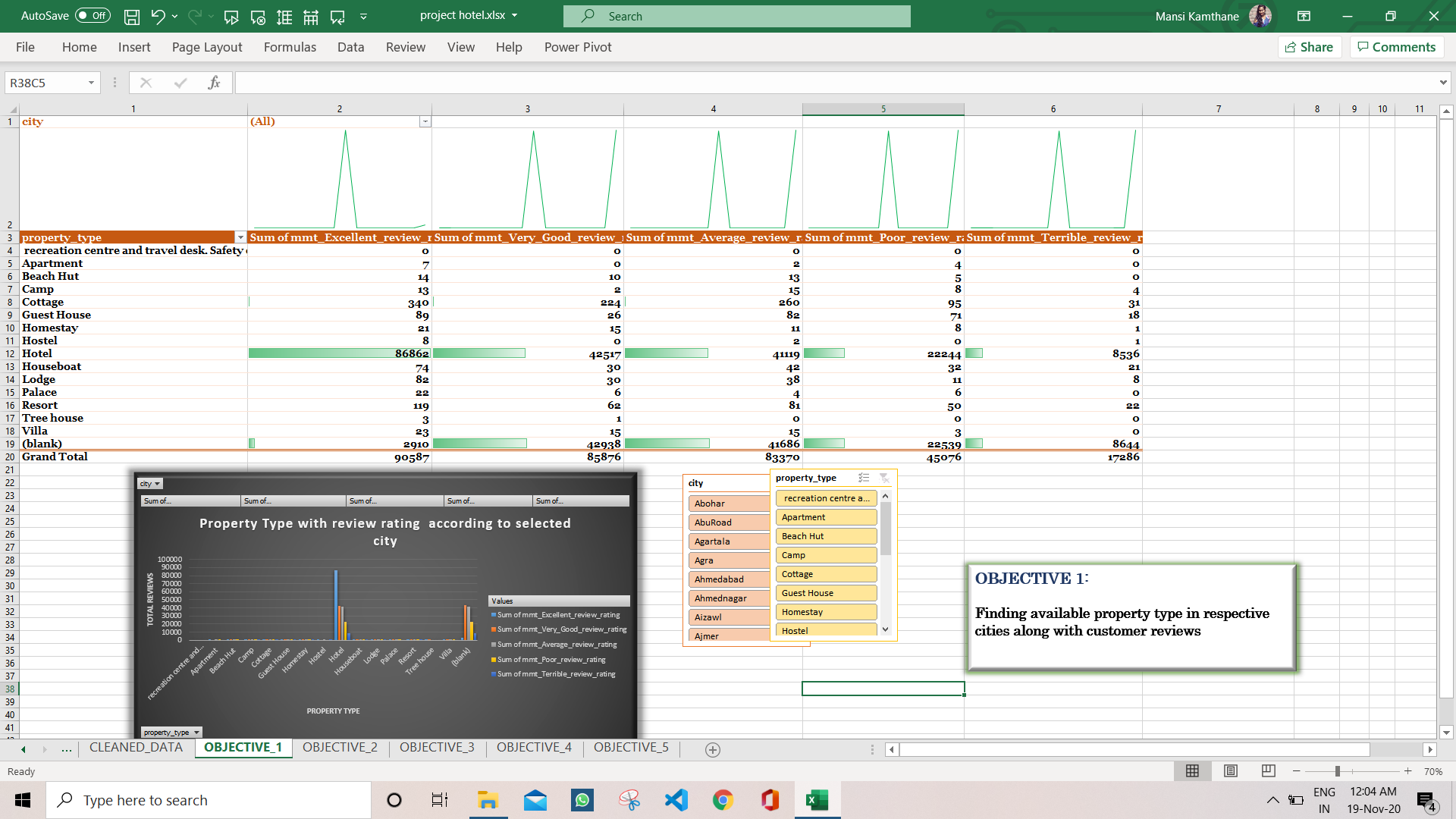
1. **SPECIFIC REQUIREMENTS:**

Knowledge of Pivot table and the sparklines, Slicer from Insert tab and the conditional formatting along with design of pivot chart.

1. **ANAYSIS RESULTS:**

It is observed that Excellent reviews were almost at the top whereas, Terrible were much less as compared. The highest reviews is shown at hotels which signifies the number of hotels to be more. The filter of city can be used to see the visualization according to the city.

There is huge demand at hotels but the terrible reviews here is more as compared to its average value. Hence, attention should be payed for the highlight values the hotel offers to improve the reviews of the customers.

****

1. **VISUALIZATION:**

Above Pivot chart shows the sum of ratings given by the customers from makemytrip.com which shows the hotels to be used most from other property type. Excellent rating is seen in hotels booked from the site. Blank section shows the rating which does not define for any specific property type.

According to city data changes and can be viewed as per the requirement. The reviews reach up to 90,000 counts and this shows the business growth. Hence, attention should be paid to other property type.

**Summarizing the total number of property type, especially hotels in respective city.**

1. **INTRODUCTION:**

In this objective, the insights are brought into light to know the use of hotels in respective area. It is assumed that the total number of hotels speaks about customer demand in respective cities. Count of hotels can be seen and visualized for user friendly. Similarly, the hotels as well as other type of property types can be considered.

Customers on site prefer hotels in locations which are tourist places. Lonavala, Pali and other such tourist cities have crowd of hotels which gradually increases the demand of hotels and increasing the growth of the owners. In such tourist places for more profit other property types can be constructed and hence, it will have better demand. The quality can be rise by the facilities they provide.

Therefore, the count of hotels shows the visits of customers and analyses the different aspects of tourism. In this analysis the data is cleaned and transformed as well as filtered for best possible results. Different sheet is used for pivot chart and tables. Further you may get familiar with its charts as well as the suspects its considering for. The insight is based on the counts of the hotels. It summarizes the count of property type in desired filters according to city and the layers of filtering process.

1. **GENERAL DESCRIPTION:**

As above scenario is described and hence the pivot chart and table are made available. The chart is in form of horizontal bar graphs. The graph plotting shows major city crowd such as Pali and Lonavala with high pitch of count of hotels which gives rise to growth in tourist area and ensures the satisfaction level of owner.

Two layer filtering is applied, one which is based on city and other based on property type which ensures the effectiveness and results in required objective form. The chart id based on the pivot table which is then designed in horizontal graphical way with the colour effect as well as the plotting designs which can enable labels too.

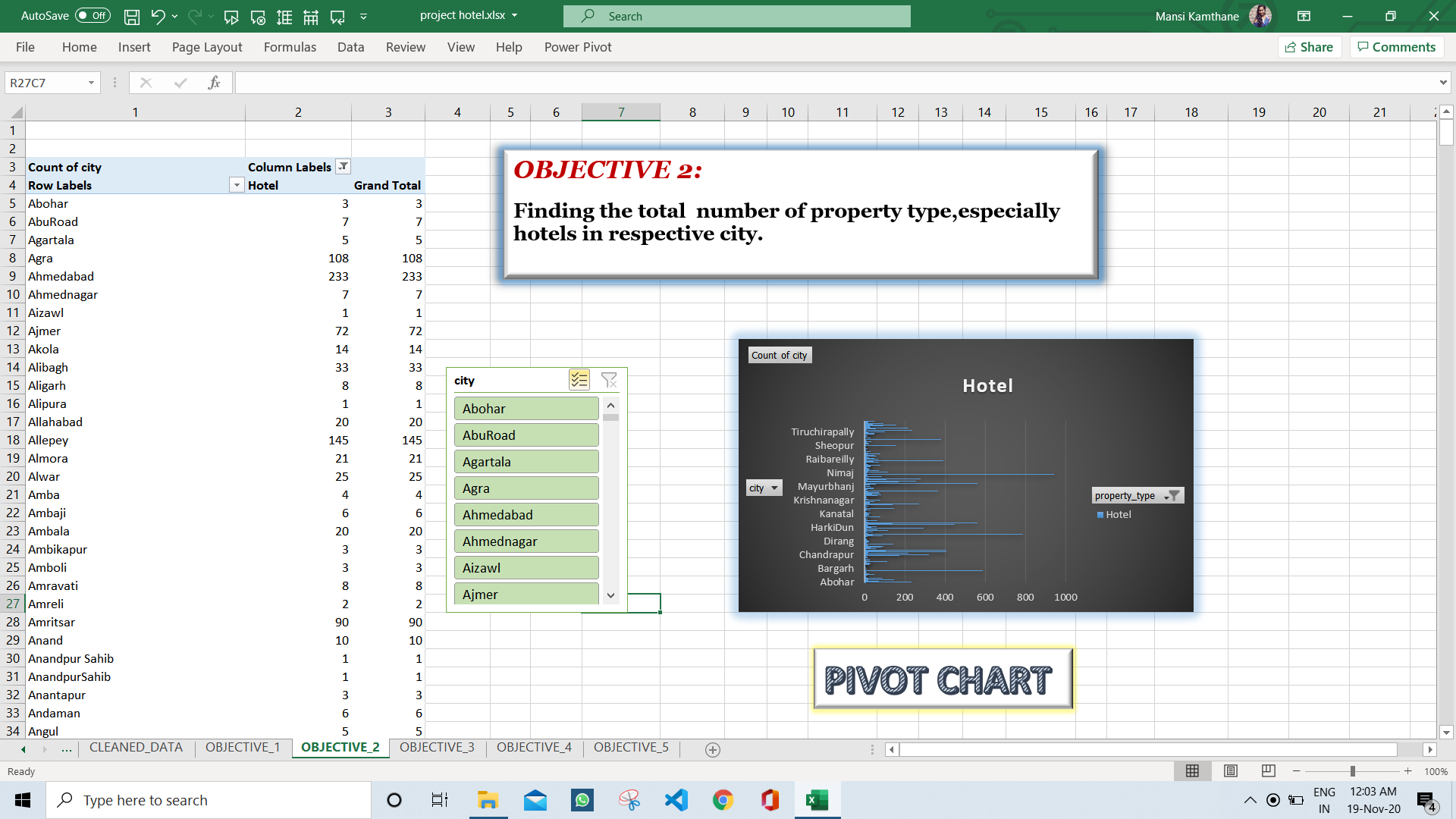
Slicers are used for user friendly filters and customized in such a way that it becomes easier to access required data. The changes in data can change the data in slicer too. One can break and connect the connection of data and the slicer. The slicing technique is enabled with special functionalities of excel.

1. **SPECEFIC REQUIREMENTS:**

Knowledge of Pivot table and the sparklines, Slicer from Insert tab and the conditional formatting along with design of pivot chart. Data tab should be known as well as the insert tab to enhance the use of slicers.

1. **ANAYSIS RESULTS:**

Pivot table is analysed where one can see the grand total which is count of the respective property type in respective cities. Goa, Cochin, Coorg, Mumbai, Bangalore such cities have count of nearly 600 hotels which attracts a great crowd. Hence, if any other hotel is to be constructed the same cities should be considered for better profit. As the count of hotel is directly proportional to the hotel demand by the customers.

****

1. **VISUALIZATION:**

All the cities are not visualized here as the filtering is selective. One can view according to the filters as explained in description part of the analysis. Glow, bevel and other border effects can be seen. Labels and caption is mentioned. The Blue colour horizontal bars show the counts as per the X-axis.

X-axis shows the count whereas the Y-axis shows the cities selected as per the filtered layers and the inner part shows the plotted area.

**Finding frequency of hotels according to stars and filtered by city as well as value given by people.**

1. **INTRODUCTION:**

The objective shows the frequency of hotels based on the star ratings of the hotels. The customers prefer 5 star hotels and average people wishes to go for 3 star hotels. Hence, it becomes easier for them to find the hotels as per their need by simply using slicer in user friendly way. Customers after choosing their own preferences they note their experience in the way by giving feedback of being valuable. They either say it was value plus or no to the feedback. The value plus also satisfies the customers by its worth in aspects of money and the living.

To know the satisfactory feedback strength as well as to know the 5 star hotels this insight Is brought into consideration. Plotting if this type of conclusions will help us to understand the categorization of hotel star rating and know the audience with their preferences. Value plus category helps to know the satisfaction level of the customers and also know the highlighted points to be changed as per the convenience.

Hence, categorization makes huge impact of the insight mentioned and will help owner to understand the aspects of the hotels while constructing the hotel. It also helps owner to understand the need of hotel rating as to be satisfied what should be the facilities provided.

1. **GENERAL DESCRIPTION:**

The column of value plus which holds the value as yes and no are manipulated into binary values for easy access. The columns are merged and cleaned for the duplicate values. Hotel star rating column is manipulated by changing the text into numeric values and applied with conditional formatting of star icon which signifies the column rating.

The pivot chart is made but for the count of hotels, property type is put in the values section of pivot table and hence the later on chart is prepared from pivot table. Specifically, slicers are added here for value plus column. Customer can go according to value Plus or the any way as per the convenience.

This shows the counts of hotels which is not value plus!

The above plot shows the hotels which are value plus. Hence, one can say that 3star hotels are value plus.

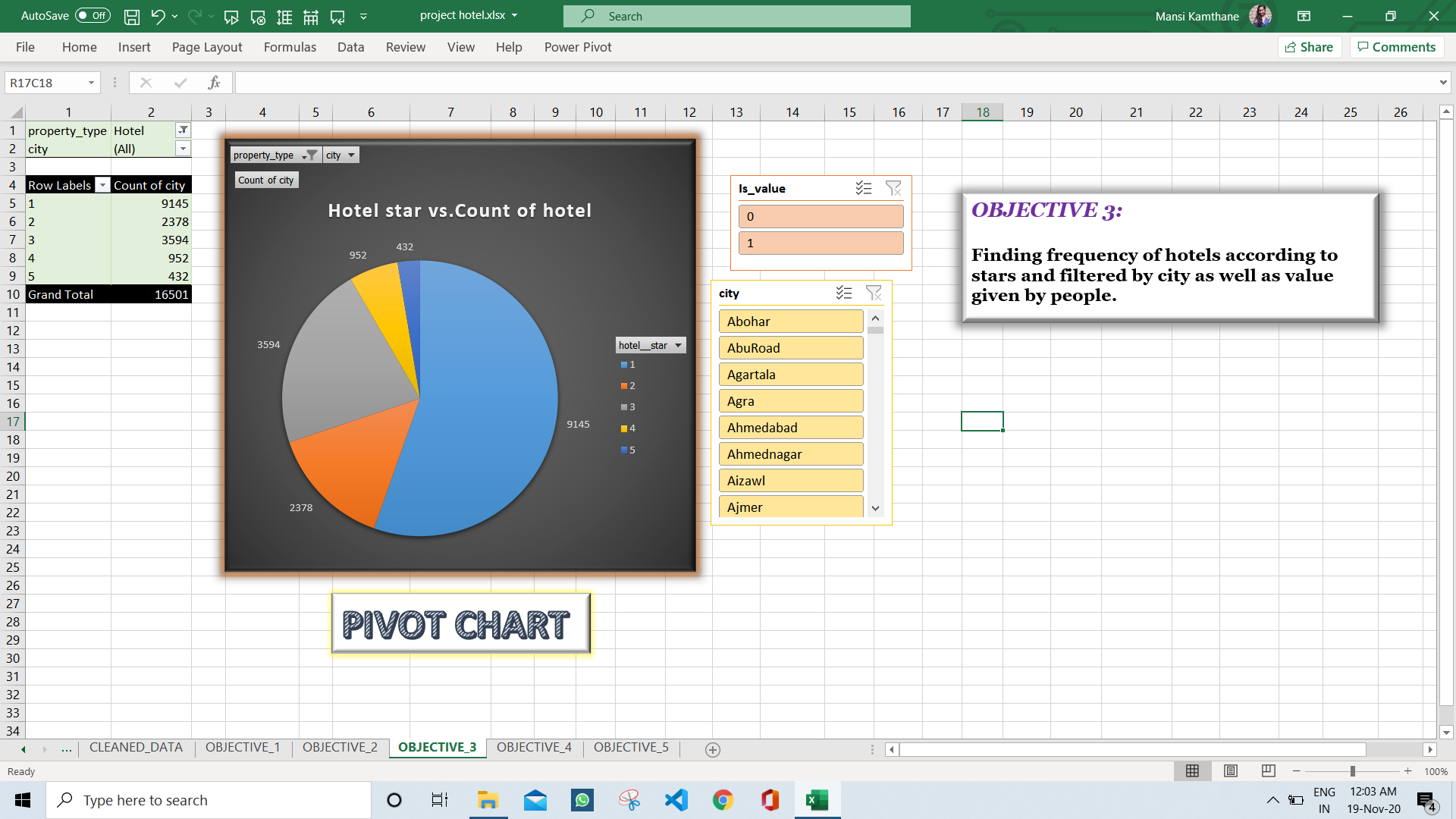
1. **SPECEFIC REQUIREMENTS:**

Knowledge of Pivot table and the sparklines, Slicer from Insert tab and the conditional formatting along with design of pivot chart. Familiar to data tab, view tab and design tab. Formulas and different types of charts specially pie chart.

1. **ANAYSIS RESULTS:**

There are two slicers where one denoted city and other value plus identification. The analysis results as there should be 5 star hotels to satisfy the need of customer satisfaction but the count is very less and hence should be increased if customer satisfaction is considered as priority. Average count is for 3 star hotels which runs smoothly and are satisfied too.

Whereas people in village areas or other areas prefer 1 star hotels where value is not considered but the rates are considered and hence are in huge crowd.

****

1. **VISUALIZATION:**

The colour label beside shows different colour which categorizes different star rating of the hotels, while the count on that category mentions the count of hotels. We can observe that the 1 star hotels are much in counts and hence explains the demand of the people. 5 star hotels are very less but when we see the satisfactory level by looking at slicer of value plus; it is observed that 5 star hotels are worthy as compared to 1 star hotels.

**Analyzing the value of property type as review of people to look over the highlights.**

1. **INTRODUCTION:**

There are many property type and hence to analyse the value of it by feedback of customer this insight has brought to book. In this, the total count of yes and no feedback are summarized using dashboard. The area coverage is considered for the data and further analysis is done.

To know the satisfactory feedback strength as well as to know the value of hotels this insight Is brought into consideration. Plotting if this type of conclusions will help us to understand the categorization of hotel value and know the audience with their preferences. Value plus category helps to know the satisfaction level of the customers and also know the highlighted points to be changed as per the convenience.

Hence, categorization makes huge impact of the insight mentioned and will help owner to understand the aspects of the hotels while constructing the hotel. It also helps owner to understand the need of hotel value as to be satisfied what should be the facilities provided.

People do prefer the value for money way!

1. **GENERAL DESCRIPTION:**

To visualize here the area plot is used and made understand the importance of facilities the customer provides. Hence, there is need oof customer feedback of yes in value plus category. The plot is designed and labelled well as per the caption and design. Binary values of value plus column is brought in use. Property type can be filtered accordingly. Column are merged and cleaned as per the requirement and filtered well.

Pivot table are made on basis of cleaned data and then the pivot chart Is made.

Slicers are used for user friendly filters and customized in such a way that it becomes easier to access required data. The changes in data can change the data in slicer too. One can break and connect the connection of data and the slicer. The slicing technique is enabled with special functionalities of excel.

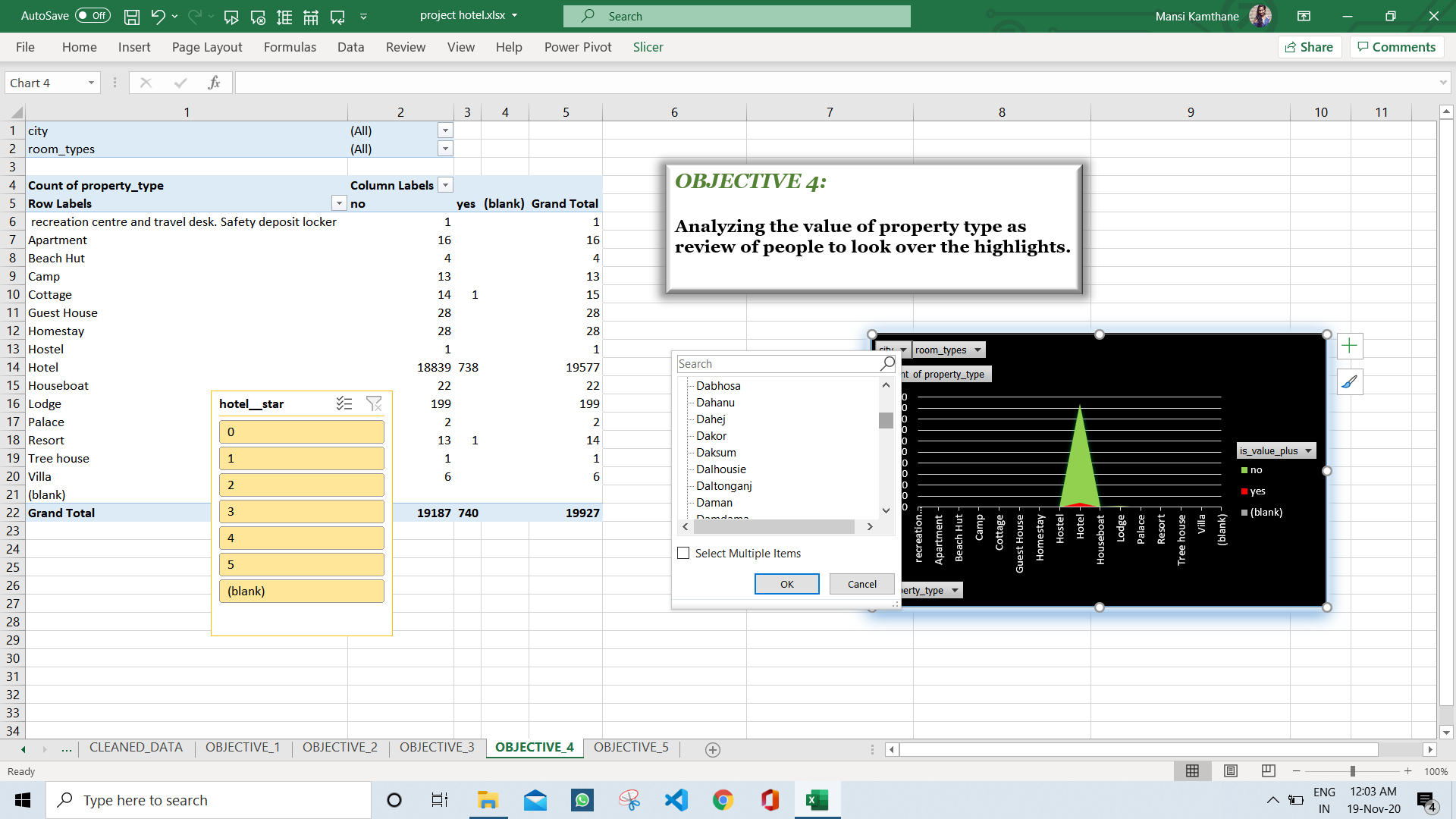
1. **SPECEFIC REQUIREMENTS:**

Knowledge of Pivot table and the sparklines, Slicer from Insert tab and the conditional formatting along with design of pivot chart. Different types of conditional formatting can be done. Data duplication can be filter out hence, knowledge to access is required.

1. **ANAYSIS RESULTS:**

City filter can be seen in the chart which filter accordingly. Count of hotel makes the huge impact of the summarizing of the categories. Hence by area graph this can be visualized. The graph results that the hotels are less which are value plus. Hence, there is urgent need to make the hotels worthy.

This visualization can also be done on room types maybe AC or Non-AC. The other attributes can create an another way to make it worthy.

****

1. **VISUALIZATION:**

Area covered by “NO” is large and huge as compared to the “YES” which says that the hotels are less worthy and have negative feedback by the customers.

The other property type nearly have any feedback and hence are nearly considered to be null values.

The visualization is made better by enabling design and making the theme visible. Gridlines are enabled to take the measure on Y- axis.

**Analyzing the outliers based on longitudes and latitudes along with layer of hotel ratings and summarizing the hotel summations in country.**

1. **INTRODUCTION:**

To identify outliers based on latitude and longitude, we plot the latitudes and longitudes to check whether the points are within the country or outside the country. These helps to find the outliers and know the wrong measures which can be rectified later on.

Data used should eb transformed earlier. Different types of aspects can be combined with these. All city at a glance can be viewed with combining different aspects. This helps in understanding the crowd over the city. Customer demand is always a priority. Hence, it becomes necessary to provide right information for positive feedback. Right measures of latitude and longitude can ensure about the right address through google maps.

Maps are types of charts. Heatmaps can also be brought in use to relate them with each other depending in any other category. Hence, it is necessary to assume the outliers to oppose the trend in any other type of category too. Lastly, in this objective the other factors are also highlighted and similarly displays the dashboard too. Here, slicers can also be added.

1. **GENERAL DESCRIPTION:**

Import the excel file of cleaned data into tableau to make the worksheets of visualizations. Green colour shows the numeric data and that of blue colour shows the labelled data. There is plot area, column and row; just by drag and drop operations can be done. Through show me panel different type of charts can be chosen. Size, colour, labels and much more can be customized.

3D-Maps can be plotted by using Excel Office 365 where range is to selected of country, region or cities to plot on a map. different colours, labels can be accessed as well as on can plot other factor on the same.

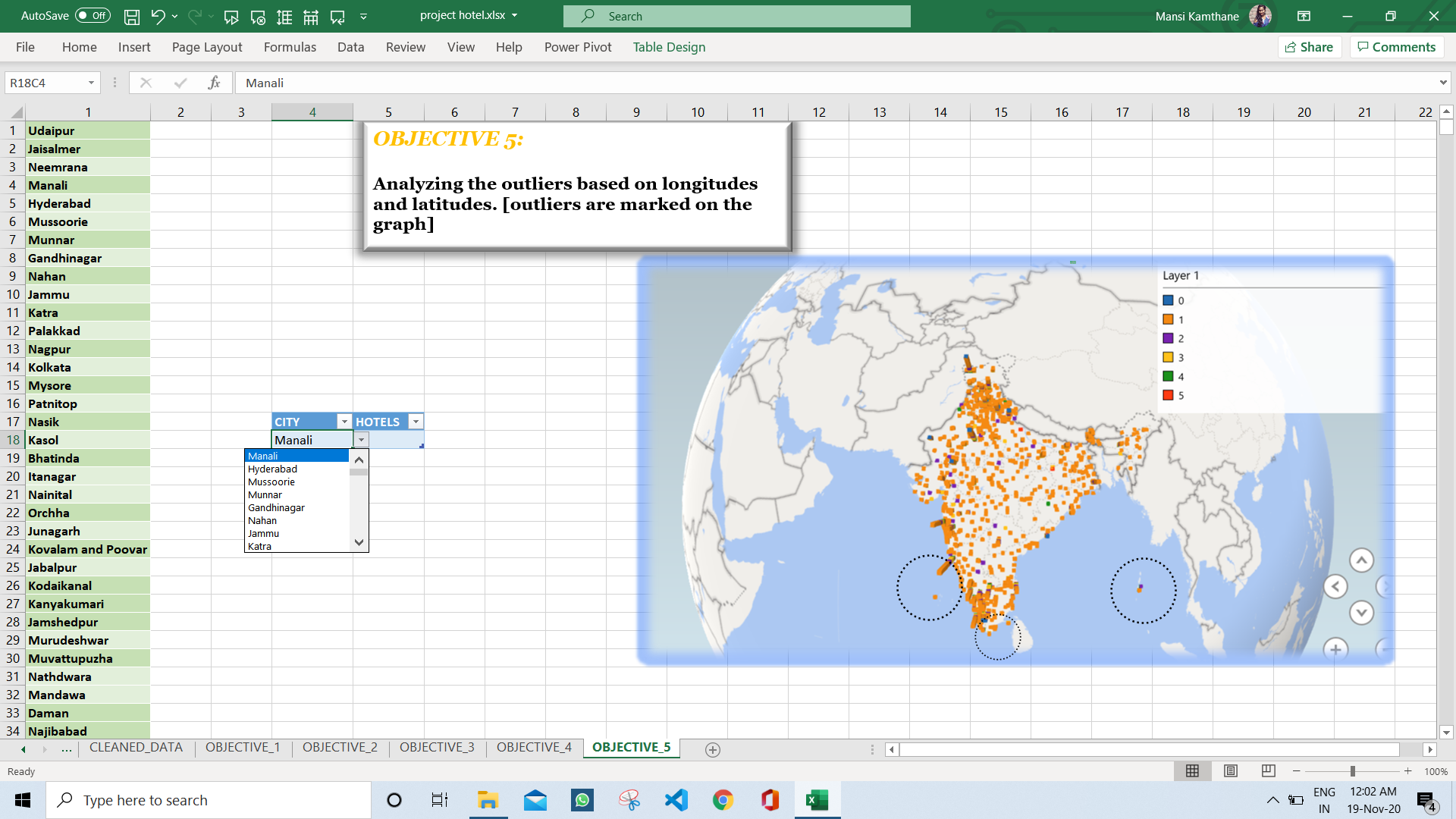
Count can be changed as sum, average, product and much more for customizing results. The type of data can also be chosen. Data validation can be applied. Colour platelet can be changed in design tab.

1. **SPECEFIC REQUIREMENTS:**

Knowledge of Pivot table and the sparklines, Slicer from Insert tab and the conditional formatting along with design of pivot chart. Tableau and its functionalities along with panel, filters, size show me and different aspects of the software. 3D- MAP in excel which can be studied with help of tutorials, plotting with help of latitudes and longitudes.

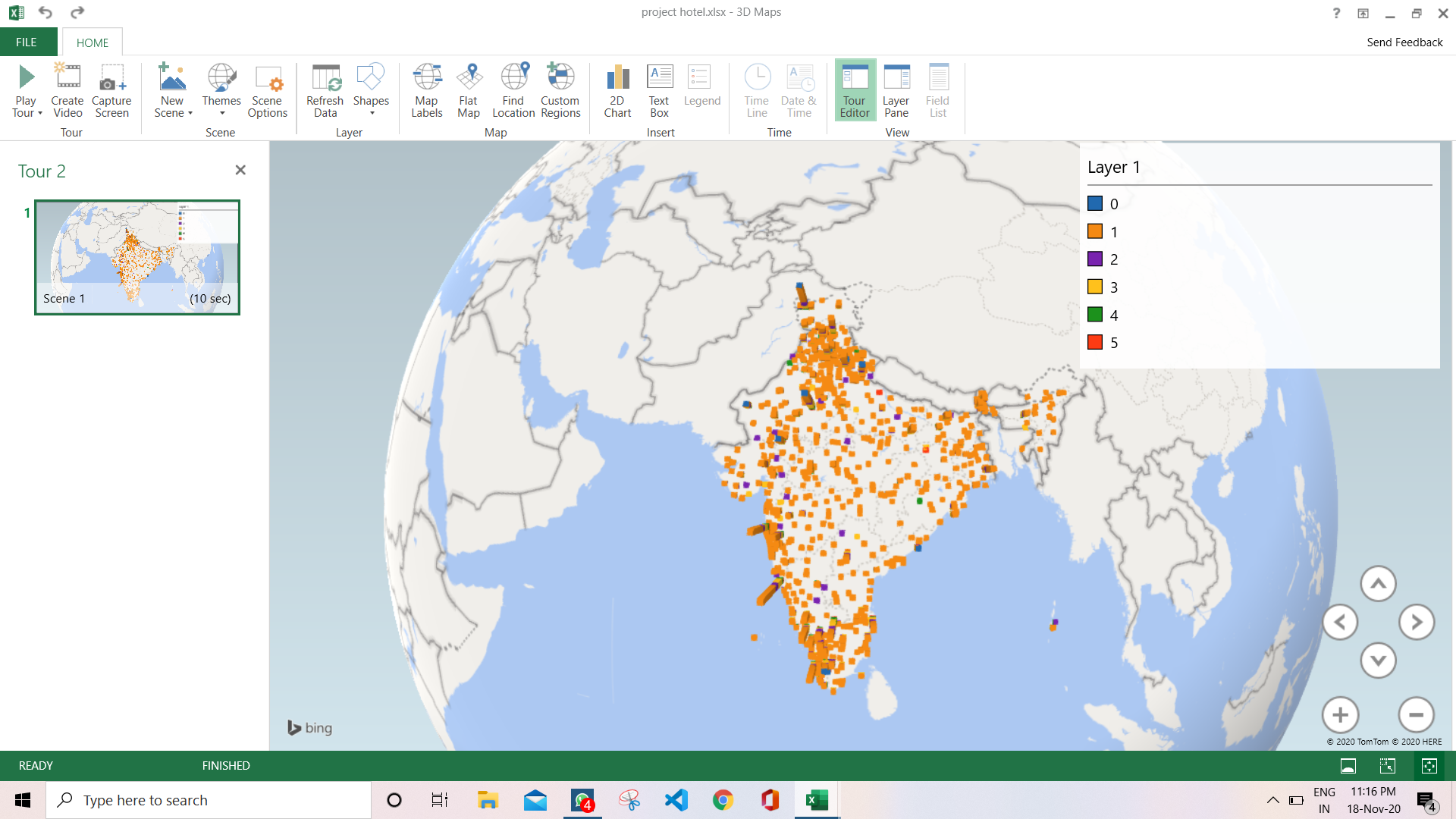
1. **ANAYSIS RESULTS:**

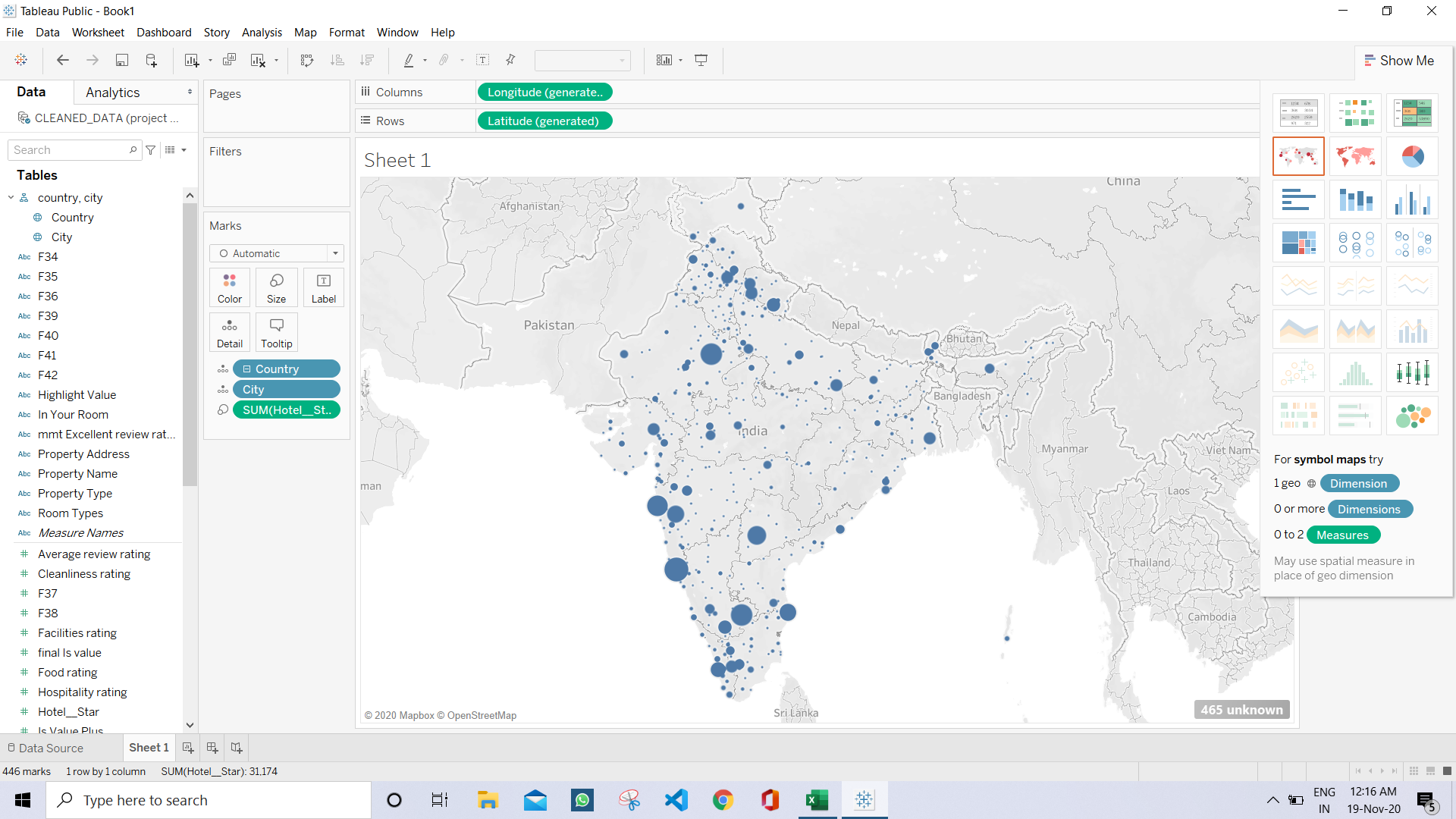
There were some outliers which should be noted down and the outliers shows the wrong measure of latitudes and longitudes which should be corrected. Data validation is added to see the hotel facilities along with the hotel name in respective city. The green table column shows unique city names nearly about 750.

****

1. **VISUALIZATION:**

The below visualization shows 3-D Map with different colours for different star ratings and plotted on map as bar graph. Respective colour shows the respective star rating hotels. Main concern here was to identify outliers.





The above visualization shows the sum of hotels which is done with help of tableau. This makes it easier with the size of bubble. Increase in size of bubble aims in increase of the count of hotels. Outliers can also be identified.

**REFRENCES**

<https://www.kaggle.com/PromptCloudHQ/hotels-on-makemytrip>

[https://www.excel-easy.com](https://www.excel-easy.com/)

**LIST OF ANALYSIS WITH RESULTS**

1. Finding available property type in respective cities along with customer reviews.

***RESULT:***

Only hotels have more excellent rating as compared to another property type on their average.

1. Summarizing the total number of property type, especially hotels in respective city.

***RESULT:***

Number of hotels are maximum in tourist area and the cities.

1. Finding frequency of hotels according to stars and filtered by city as well as value given by people.

***RESULT:***

Balancing the value, 3 star rated hotels are in profit and are valued.

1. Analyzing the value of property type as review of people to look over the highlights.

***RESULT:***

Hotels are not valued as per the worth which is less than the average votes.

1. Analyzing the outliers based on longitudes and latitudes along with layer of hotel ratings and summarizing the hotel summations in country.

***RESULT:***

There are some outliers who measures with wrong latitude and longitude.

**BIBILOGRAPHY**

I have used many resources as such as videos available on google and searched for many shortcuts used for excel which made by work easier. Once going through the tabs and functionalities will make it easier to access the MS. Excel. the version of Excel may vary which may or may not include some functionalities. I used the Office 365 which made it easier for navigation as well as more functionalities are made available.

Cited references are made available to know the browsing videos and tutorial for accessing Excel who is not aware about the software. Link for downloading Office 365 will be given at the end.

I had many doubts in enabling add-ins and choosing the right chart type for the desired analysis. On browsing it made easier to know which chart type should be used and where as well as when. Insights made it easier due to cleaning of data. Tutorials for tableau was browsed to understand the working of the software and to know the additional features.

Tableau tutorials are made available on the official site as well where necessary functionalities are explained in proper way with visualization.

Link for downloading tableau: <https://public.tableau.com/en-us/s/download>

Link for downloading Office 365: <https://www.microsoft.com/en-in/microsoft-365/get-office-and-microsoft-365-oem-download-page>