DATA ANALYTICS PROJECT ON AIRBNB CASE STUDY

A PROJECT REPORT

Submitted by

SURENDHAR R - 191401035

in the partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND BUSINESS SYSTEMS

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI-602105

ANNA UNIVERSITY: CHENNAI-600025

DECEMBER 2020

BONAFIDE CERTIFICATE

Certified that this project report **"DATA ANALYTICS PROJECT ON AIRBNB CASE STUDY"** is the bonafide work of **"SURENDHAR R (191401035)"** who carried out the project work under my supervision.

SIGNATURE   **SIGNATURE**

Dr. K.DEVAKI, M.E,PhD.,   **Mr.B.BHUVANESWARAN, M.E.,**

HEAD OF DEPARTMENT   SUPERVISOR

PROFESSOR   ASSISTANT PROFESSOR (SG)

Department of Computer Science and   Department of Computer Science and

Business Systems   Engineering

Rajalakshmi Engineering College   Rajalakshmi Engineering College

Chennai – 602 105   Chennai - 602 105

Submitted to project and viva-voce Examination held on _____

INTERNALEXAMINER   **EXTERNALEXAMINER**

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report.

Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E., F.I.E.,** and our respected Chairperson **Dr. (Mrs) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time.

We express our sincere thanks to **Dr. K.DEVAKI, M.E,PhD.,** Head of the Department of Computer Science and Business Systems for her guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Mr. BHUVANESWARAN, M.E.,** Assistant Professor (SG), Department of Computer Science and Engineering, Rajalakshmi Engineering College for her valuable guidance throughout the course of the project.

We are glad to thank our project coordinator, **Mrs.HELEN VIJITHA.P,** Associate Professor, Department of Humanities and sciences for her useful tips during our review to build our project.

Finally, we express our gratitude to our parents and classmates for their moral support and valuable suggestions during the course of the project.

ABSTRACT

In today's world, travelling is necessary for many reasons. It can be official Business meeting, attending important events, some people visit a place for inner peace and pleasure, etc. Generally, if he/she travel to any country, he/she will stay in a hotel. But majority of hotels cannot fulfil the customer needs. Most of the hotels are costly, sometimes customer don't like the food what they provide, the hotel rooms are fully booked, you cannot get a home experience.

Solution for this problem is provided by the Airbnb company. Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodation in that locale. Airbnb takes 3% commission of every booking from hosts, 6% and 12% from guests.

This analytics project aims at solving the Airbnb case study consists of more than 25 questions by analysing and finding insights from the Airbnb dataset. By using this project , Airbnb company can take further decisions for their Business improvement.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

EDA – Exploratory Data Analysis

CSV – Comma Separated Values

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Travelling is one of the daily parts of life. Most of the people travel from one country to another country or one city to another city for their own personal reasons. Airbnb is a company which gives better customer service than hotels.

ABOUT AIRBNB:

Airbnb is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in that locale. It currently covers more than 81,000 cities and 191 countries worldwide.

WIDE SELECTION:

Airbnb hosts list many different kinds of properties—single rooms, a suite of rooms, apartments, moored yachts, houseboats, entire houses, even a castle—on the Airbnb website.

FREE LISTINGS:

Hosts don't have to pay to list their properties. Listings can include written descriptions, photographs with captions, and a user profile where potential guests can get to know a bit about the hosts.

HOSTS CAN SET THEIR PRICE OWN:

It's up to each host to decide how much to charge per night, per week, or per month.

CUSTOMIZABLE SEARCHES:

Guests can search the Airbnb database—not only by date and location, but by price, type of property, amenities, and the language of the host. They can also add keywords (such as "close to the Louvre") to further narrow their search.

1.2 OBJECTIVES:

The main objectives of this project are:

➢ To explore more about the dataset.

➢ To provide visualization of data.

➢ To find and gain useful information from the dataset.

➢ To answer all the questions in the case study.

1.3 EXISTING SYSTEM:

The Airbnb dataset is a raw dataset which contains some unwanted and missing data in the form of table. The data has room_id, survey_id, host_id, room_type, country, city, borough, neighbourhood, reviews, overall satisfaction, accommodation, bedrooms, price, mainstay, name, last modified, latitude, longitude, location in the form of CSV file.

1.3.1 LIMITATIONS IN EXISTING SYSTEM:

- Raw data is difficult to interpret
- We cannot get insights from raw data

1.4 PROPOSED SOLUTION:

The main purpose of proposed solution is to process the raw data and use Python libraries to visualize it without much manual work. Also, with the help of Python libraries and with the help of visualizations, more insights and hidden information about the data can be derived. The data will be explored while looking for useful information.

1.4.1 ADVANTAGES OF PROPOSED SOLUTION:

- Better use of data
- EDA and Visualization
- Gaining Useful insights

CHAPTER 2
SOFTWARE SPECIFICATIONS

2.1 SOFTWARE SPECIFICATION:

The purpose of the Software Requirement Specification is to produce the specification of the analysis task and also to establish complete information about the requirement, behavior and also the other constraint like functional performance and so on. The main aim of the Software Requirement Specification is to completely specify the technical requirements for the software product in a concise and in unambiguous manner.

2.1.1 GOOGLE COLAB

Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs.

Colab notebooks are stored in Google Drive, or can be loaded from GitHub. Colab notebooks can be shared just as you would with Google Docs or Sheets. Simply click the Share button at the top right of any Colab notebook, or follow these Google Drive file sharing instructions. You can download any Colab notebook that you've created from Google Drive following these instructions, or from within Colab's File menu. All Colab notebooks are stored in the open source Jupyter notebook format (. ipynb).

2.1.2 REQUIREMENTS TO ACCESS GOOGLE COLAB

FEATURE	REQUIREMENTS
DISK SPACE	No need for disk space. It is a web-based IDE
SYSTEM ARCHITECTURE	Windows- 64-bit x86, 32-bit x86; MacOS- 64-bit x86; Linux- 64-bit x86, 64-bit Power8/Power9
OPERATING SYSTEM	Windows 8 or newer, 64-bit macOS 10.13+, or Linux, including Ubuntu, RedHat, CentOS 6+, and others

1) Start working with Colab you first need to log in to your google account, then go to this link https://colab.research.google.com.

2) **Opening Jupyter Notebook:**

On opening the website you will see a pop-up containing following tabs –
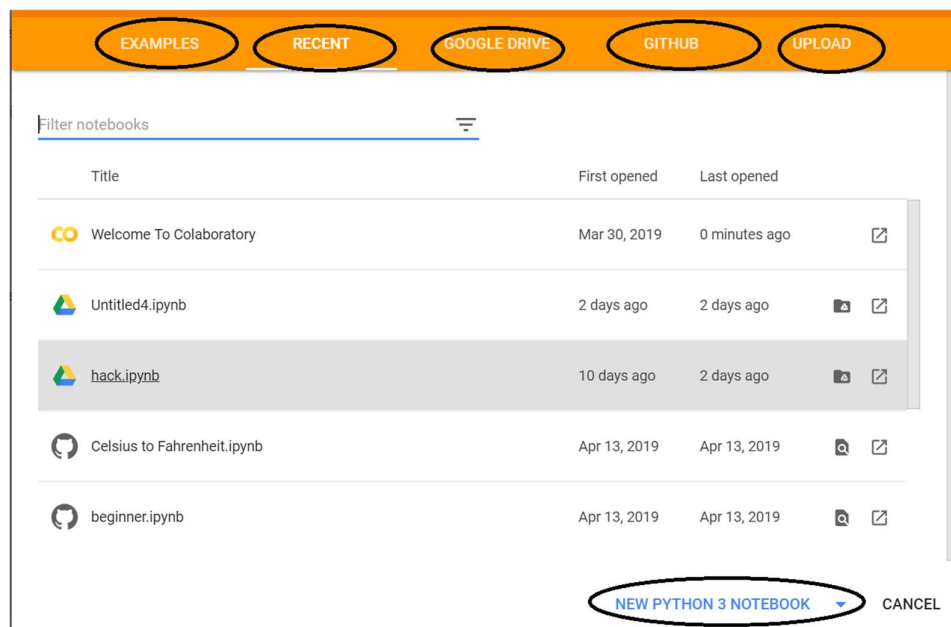
EXAMPLES: Contain a number of Jupyter notebooks of various examples.
RECENT: Jupyter notebook you have recently worked with.
**GOOGLE DRIVE:** Jupyter notebook in your google drive.
GITHUB: You can add Jupyter notebook from your GitHub but you first need to connect Colab with GitHub.
UPLOAD: Upload from your local directory.

Else you can create a new Jupyter notebook by clicking New Python3 Notebook or New Python2 Notebook at the bottom right corner.

Notebook's Description:

3) On creating a new notebook, it will create a Jupyter notebook with Untitled0.ipynb and save it to your google drive in a folder named **Colab Notebooks**. Now as it is essentially a Jupyter notebook, all commands of Jupyter notebooks will work here. Though, you can refer the details in Getting started with Jupyter Notebook.
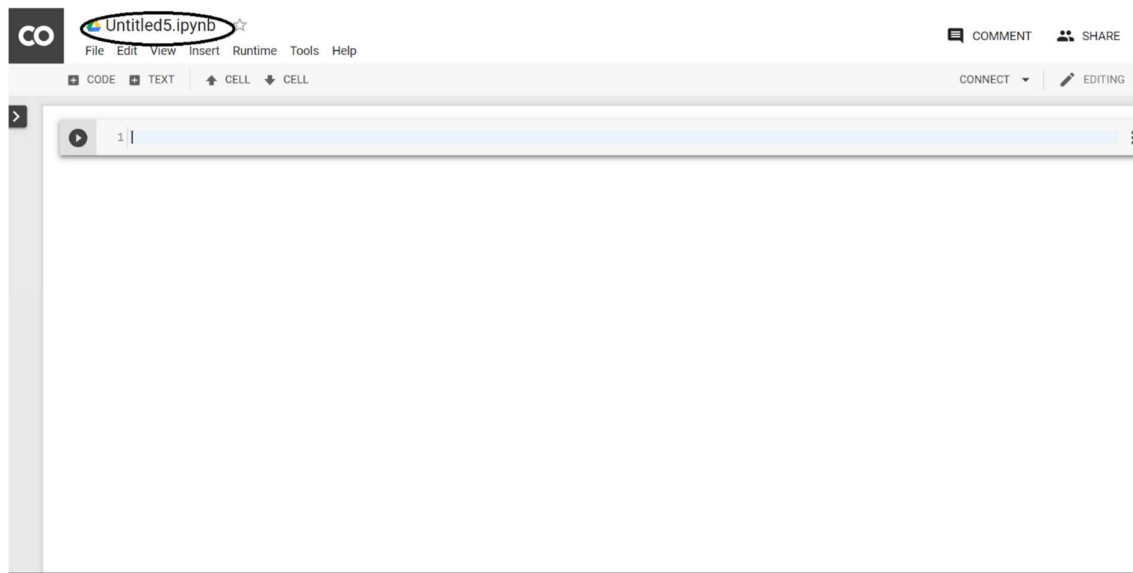
CHAPTER 3
PROJECT DESCRIPTION

3.1 GENERAL:

The goal of this analytics project is to solve the Airbnb case study consists of more than 25 questions by analysing and finding insights from the Airbnb dataset. Any data which are represented in Graphs and charts i.e. in visualized form will be easy to interpret. So appropriate data visualization is used in the required areas. By using this project, Airbnb company can take further decisions for their Business improvement, customer needs, what they prefer the most, how ratings, accommodation size, etc. affects the price rate.

3.2 MODULES:

1. Data Collection

2. Importing and Pre-processing Data

3. Data Analysis

3.2.1 MODULE DESCRIPTION:

1. Data Collection:

Data was collected from Kaggle, an online community of data scientists and machine learning practitioners.

2. Importing and Pre-processing data module:

In this module, we will import the data into the Jupyter Notebook. Then the pre-processing is done. In pre-processing, the shape of the dataset is analysed. This module is useful for gaining overall idea about the data which will help us in the upcoming modules. The data is checked for any NULL values in this step.

3. Data Analysis:

The Data Analysis part gives us insights that are hidden in the data. In this part, subject of courses, subscribers per courses, effect of price on number of subscribers, most profitable courses, most engaging courses, most popular courses, factors affecting price of course and many such factors are analysed and the possible insights that can derived from data are generated. The relationship between the variables are also discussed in this module. The visualization of these analysis is done and the appropriate graphs are generated.

CHAPTER 4
SYSTEM DESIGN

4.1 GENERAL:

The project Data Analytics on Airbnb case study dataset has various steps like collecting data, exploring data, pre-processing data, analysing data, drawing conclusions out of the analysis. This analysis will help the company to understand the users' needs and trend of the industry.

4.1.1 DEVELOPMENT ENVIRONMENT:

Hardware Environment

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the systems do and not how it should be implemented.

- Hard disk   : 1000GB
- Monitor   : LENOVO 1920*1080
- Ram   : 8GB
- Processor   : Intel(R) Core(TM) i5-8250U
- Processor speed   : 1.80GHz

Software Environment

The software requirements are the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team's and tracking the team's progress throughout the development activity.

- Operating system   : Windows 10 Home
- Languages   : Python
- IDE   : Google Colab

CHAPTER 5

IMPLEMENTATION AND RESULT DISCUSSION

5.1 Data Collection:

The Data was collected from Kaggle platform.

5.2 Importing and Pre-processing data:

The data is in a CSV file named 'airbnbdata.csv'. The dimension of data, basic details about data are checked and also the data is checked for the presence of NULL values.

5.3 Data Analysis:

Data is processed, analysed, visualized and answered all the questions in the case study.

5.4 Results:

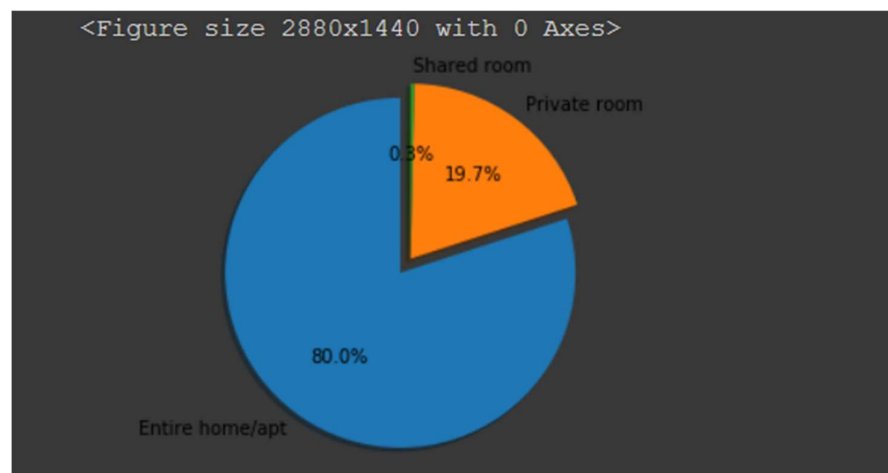Results obtained are

Question 1: What are the room type available?

Answer:

```
Entire home/apt      14937
Private room          3671
Shared room             63
Name: room_type, dtype: int64
[14937, 3671, 63]
```


Question 2: Drop the shared room type as its count is very low
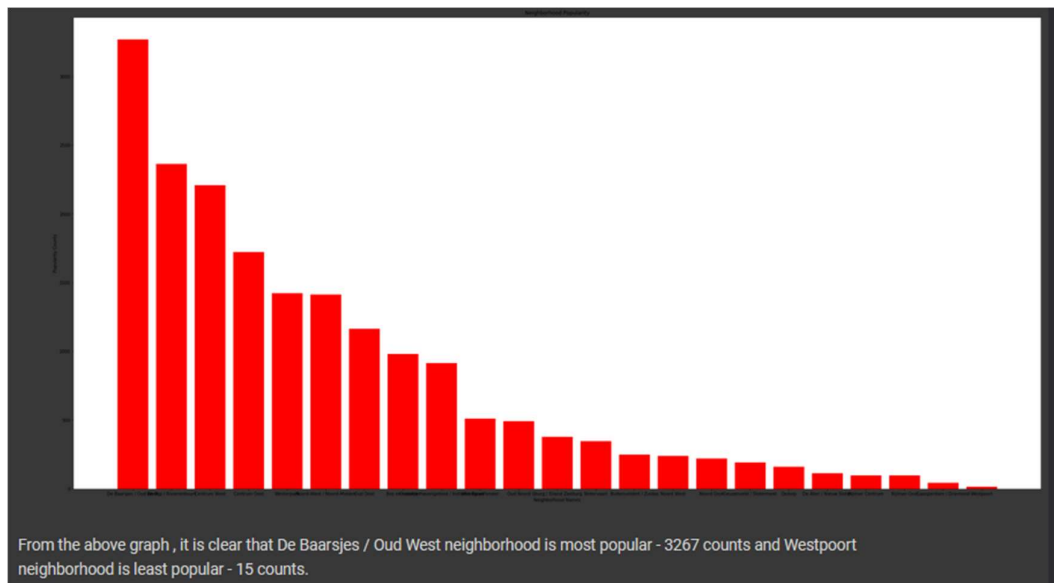
Answer:

```
Entire home/apt      14937
Private room          3671
Name: room_type, dtype: int64
```

Question 3: Take a count of each neighbourhood to show how popular they are.

Answer:

```
De Baarsjes / Oud West                         3267
De Pijp / Rivierenbuurt                        2364
Centrum West                                   2208
Centrum Oost                                   1721
Westerpark                                     1422
Noord-West / Noord-Midden                      1413
Oud Oost                                       1164
Bos en Lommer                                   980
Oostelijk Havengebied / Indische Buurt          915
Watergraafsmeer                                 512
Oud Noord                                       492
Ijburg / Eiland Zeeburg                         377
Slotervaart                                     345
Buitenveldert / Zuidas                          248
Noord West                                      240
Noord Oost                                      221
Geuzenveld / Slotermeer                         192
Osdorp                                          161
De Aker / Nieuw Sloten                          114
Bijlmer Centrum                                  98
Bijlmer Oost                                     97
Gaasperdam / Driemond                            42
Westpoort                                        15
Name: neighborhood, dtype: int64
```


From the above graph , it is clear that De Baarsjes / Oud West neighborhood is most popular - 3267 counts and Westpoort neighborhood is least popular - 15 counts.

Question 4: Number of reviews that is given for property

Answer:

```
0      2964
1      1494
2      1232
3      1096
4       923
5       872
6       732
7       679
8       587
9       525
11      494
10      473
12      395
14      362
13      340
15      308
16      299
17      261
18      235
20      226
```

<matplotlib.axes._subplots.AxesSubplot at 0x2cc7344a848>

From the above data and graph, histogram and counts of reviews 0,1,2,3,4,5 for property is comparatively high i.e

0 - 2964

1 - 1494

2 - 1232

3 - 1096

4 - 923

5 - 872
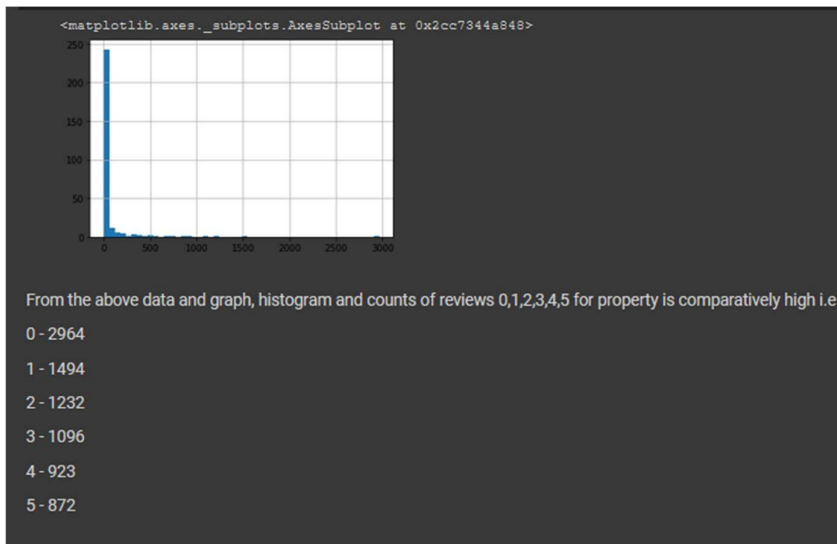
Question 5: What is the overall satisfaction count by each rating?

Answer:

```
5.0    7682
0.0    5698
4.5    4530
4.0     572
3.5     105
3.0      19
2.5       1
1.0       1
Name: overall_satisfaction, dtype: int64
```

Text(0.5, 1.0, 'Overall Satisfaction Ratings')
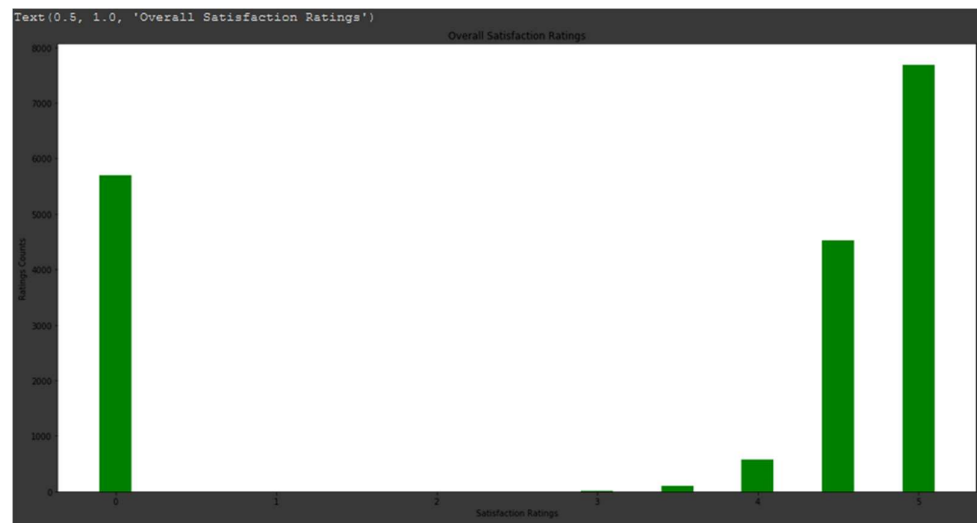
Question 6: Create a sub dataset where overall satisfaction is zero?
Answer:

	room_id	survey_id	host_id	room_type	city	neighborhood	reviews	overall_satisfaction	accommodates	bedrooms
34	3119298	1476	15847782	Entire home/apt	Amsterdam	Westerpark	1	0.0	8	3.0
36	10511291	1476	5421078	Entire home/apt	Amsterdam	Watergraafsmeer	1	0.0	8	3.0
37	18047871	1476	9302267	Entire home/apt	Amsterdam	Centrum West	0	0.0	4	4.0
38	7468232	1476	7215538	Entire home/apt	Amsterdam	Noord-West / Noord-Midden	2	0.0	4	2.0
40	5818567	1476	29561840	Entire home/apt	Amsterdam	De Aker / Nieuw Sloten	0	0.0	4	2.0
43	19980495	1476	53730026	Entire home/apt	Amsterdam	Noord West	0	0.0	5	2.0
44	20003728	1476	26861783	Entire home/apt	Amsterdam	Centrum Oost	0	0.0	4	2.0
47	14060573	1476	82283309	Entire home/apt	Amsterdam	IJburg / Eiland Zeeburg	2	0.0	4	4.0
48	19992596	1476	62831267	Entire home/apt	Amsterdam	Bijlmer Oost	0	0.0	4	3.0
49	19845595	1476	7601328	Entire home/apt	Amsterdam	Oud Oost	0	0.0	4	2.0
53	11998169	1476	1878355	Entire home/apt	Amsterdam	Centrum Oost	1	0.0	6	3.0
56	14102973	1476	20575748	Entire home/apt	Amsterdam	De Pijp / Rivierenbuurt	1	0.0	6	3.0
58	1650475	1476	6942600	Entire home/apt	Amsterdam	Watergraafsmeer	0	0.0	4	3.0
60	12031672	1476	53856038	Entire home/apt	Amsterdam	Buitenveldert / Zuidas	0	0.0	4	5.0

## Question 7: Describe this dataset using describe command
Answer:

	room_id	survey_id	host_id	reviews	overall_satisfaction	accommodates	bedrooms	price	latitude
count	1.860800e+04	18608.0	1.860800e+04	18608.000000	18608.000000	18608.000000	18608.000000	18608.000000	18608.000000
mean	1.120734e+07	1476.0	3.577515e+07	16.744895	3.305621	2.922399	1.432448	166.850924	52.365259
std	6.089422e+06	0.0	3.761321e+07	33.513610	2.212291	1.319493	0.880361	108.993577	0.015149
min	2.818000e+03	1476.0	2.234000e+03	0.000000	0.000000	1.000000	0.000000	12.000000	52.296200
25%	6.040021e+06	1476.0	7.110955e+06	2.000000	0.000000	2.000000	1.000000	108.000000	52.355247
50%	1.229705e+07	1476.0	1.986067e+07	6.000000	4.500000	2.000000	1.000000	144.000000	52.364615
75%	1.662402e+07	1476.0	5.202365e+07	17.000000	5.000000	4.000000	2.000000	192.000000	52.374790
max	2.000373e+07	1476.0	1.418319e+08	532.000000	5.000000	17.000000	10.000000	6000.000000	52.424980

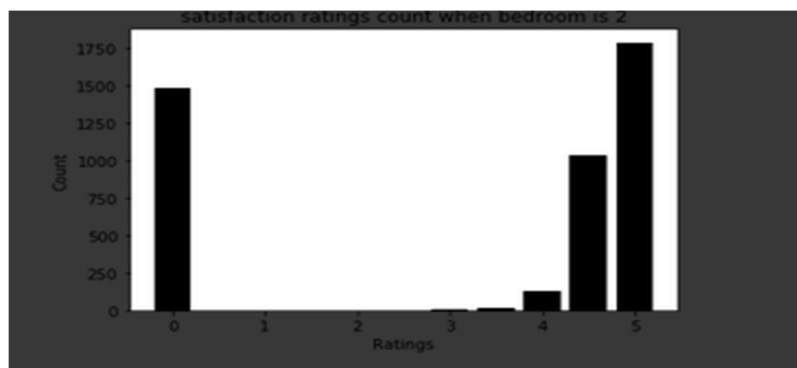## Question 8: What is the satisfaction ratings count when bedroom is 3?
Answer:

```
0.0    633
5.0    515
4.5    266
4.0     22
3.5      5
2.5      1
```


satisfaction ratings count when bedroom is 3
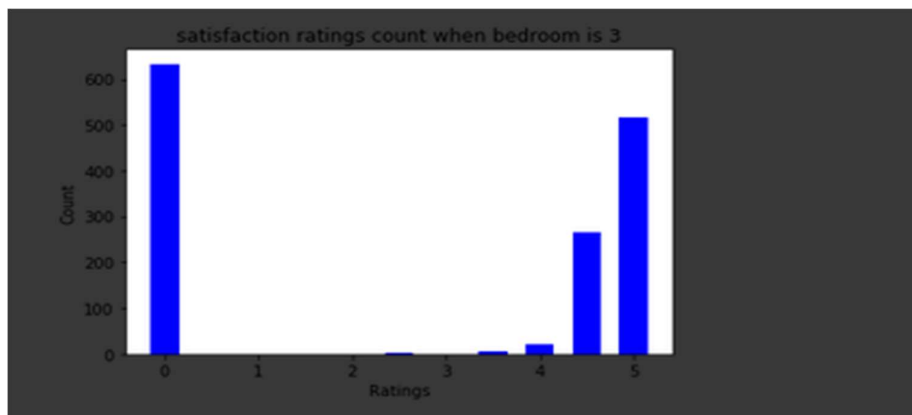
Question 9: What is the satisfaction ratings count when bedroom is 2?

Answer:

```
5.0     1785
0.0     1483
4.5     1030
4.0      126
3.5       16
3.0        6
Name: overall_satisfaction, dtype: int64
```


satisfaction ratings count when bedroom is 2

Question 10: What is the satisfaction ratings count where price is greater than equal to 150 but less than equal to 180?

Answer:

```
5.0     1632
0.0     1165
4.5      878
4.0      104
3.5       12
3.0        4
Name: overall_satisfaction, dtype: int64
```


Satisfaction ratings count where price is greater than equal to 150 but less than equal to 180
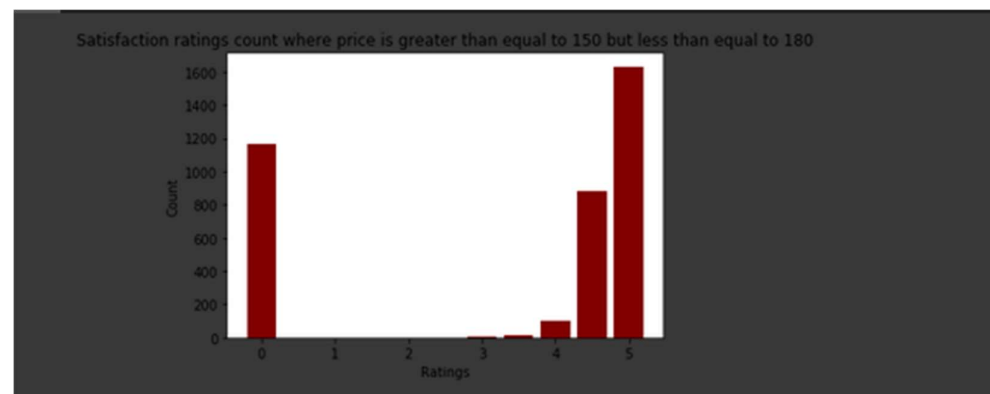
Question 11: Find out the count of each accomodation value to understand where is the most availability.

Answer:

```
2      9963
4      5555
3      1577
6       472
5       470
1       353
8       105
7        52
16       18
10       16
12        9
9         8
14        6
11        2
13        1
17        1
```


count of each accomodation value to understand where is the most availability

From the above graph it is clear that, the most available accomodation is 2 and 4
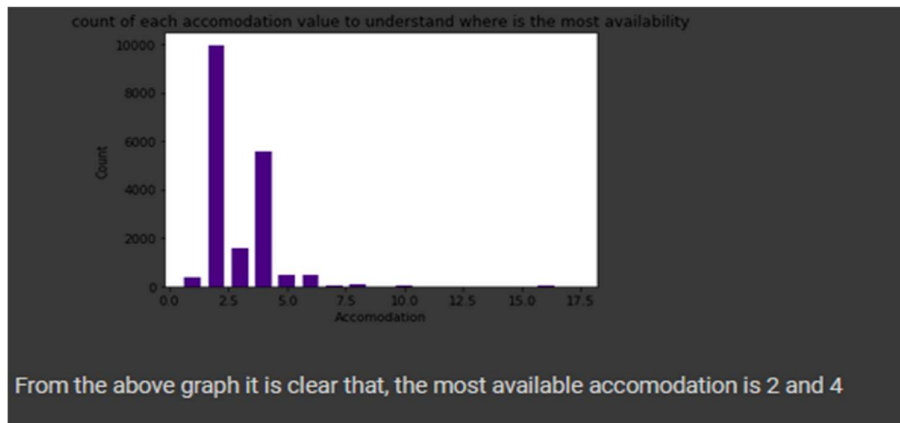
Question 12: On which price most of the properties available?

Answer:

	Price	Properties Available
0	119.0	1016
1	180.0	996
2	144.0	883
3	150.0	619
4	132.0	587
5	108.0	559
6	96.0	515
7	118.0	508
8	114.0	507
9	240.0	492
10	156.0	455

From the above dataframe it is clear that price 119.0 has more properties available
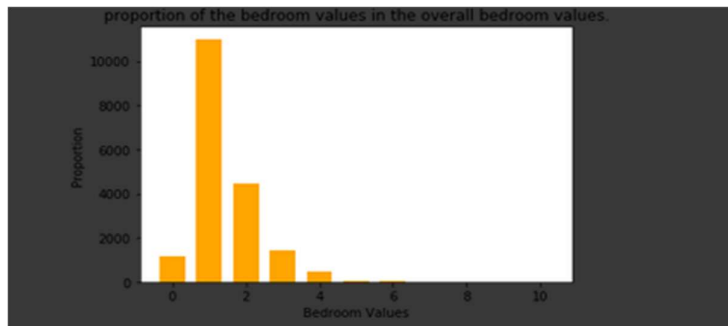
Question 13: What is the proportion of the bedroom values in the overall bedroom values?

Answer:

```
Total no. of Bedrooms available :  18608

Each bedroom proportion :
    Bedroom_values  Proportion
0              1.0       11005
1              2.0        4446
2              3.0        1442
3              0.0        1148
4              4.0         472
5              5.0          62
6              6.0          19
7             10.0           5
8              7.0           4
9              8.0           3
10             9.0           2
```


proportion of the bedroom values in the overall bedroom values.

Question 14: Drop the values when bedroom are equal to zero.

```
Dataframe after dropping 0 values in the bedrooms column :
    Bedroom_values  Proportion
0              1.0       11005
1              2.0        4446
2              3.0        1442
3              4.0         472
4              5.0          62
5              6.0          19
6             10.0           5
7              7.0           4
8              8.0           3
9              9.0           2
```

Question 15: Now check the info of entire dataset again.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 17460 entries, 33 to 18722
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   room_id               17460 non-null  int64
 1   survey_id             17460 non-null  int64
 2   host_id               17460 non-null  int64
 3   room_type             17460 non-null  object
 4   city                  17460 non-null  object
 5   neighborhood          17460 non-null  object
 6   reviews               17460 non-null  int64
 7   overall_satisfaction  17460 non-null  float64
 8   accommodates          17460 non-null  int64
 9   bedrooms              17460 non-null  float64
 10  price                 17460 non-null  float64
 11  name                  17460 non-null  object
 12  last_modified         17460 non-null  object
 13  latitude              17460 non-null  float64
 14  longitude             17460 non-null  float64
 15  location              17460 non-null  object
dtypes: float64(5), int64(5), object(6)
memory usage: 2.3+ MB
```
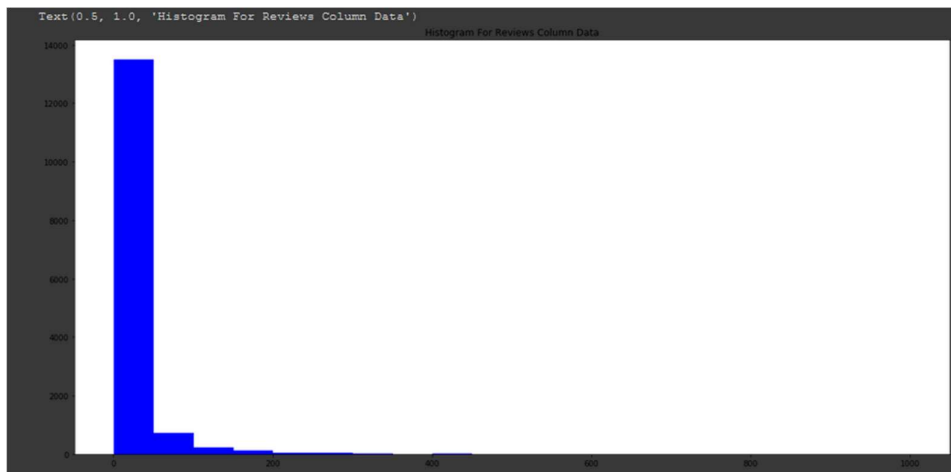
Question 16: Plot a histogram of reviews where bins are 20 and range is between 0 to 100.

Answer:
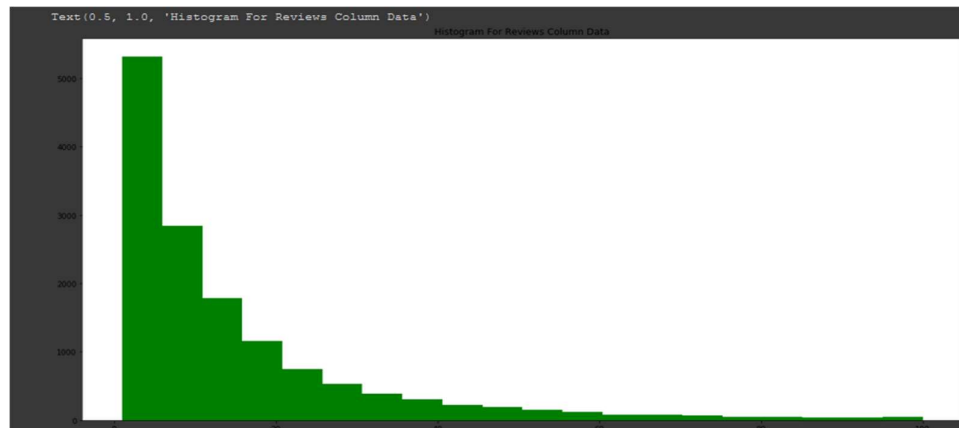
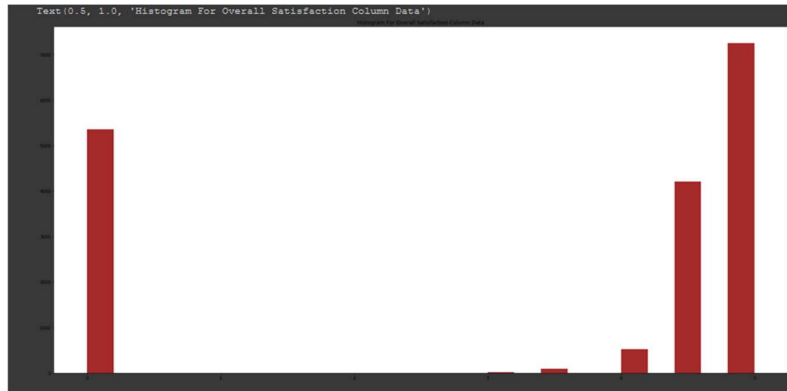**Question 17: Plot a histogram of reviews where bins are 20 and range is between 0 to 1000.**

Answer:

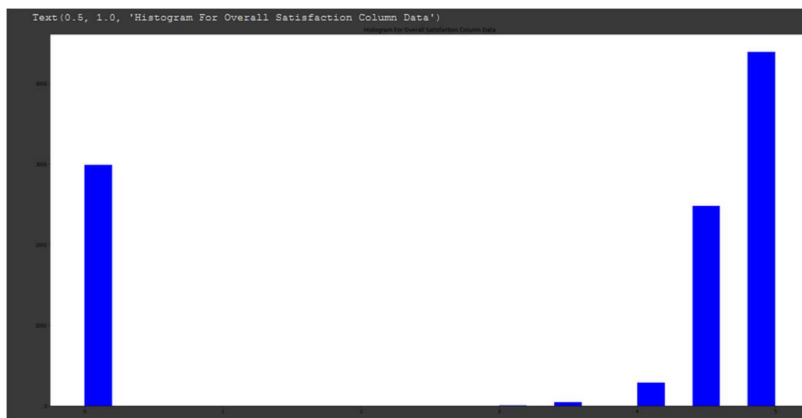**Question 18: Plot a histogram of accomodates where bins are 16.**

Answer:

Question 19: Plot a histogram of overall_satisfaction.

Answer:

Question 20: Plot a histogram of overall satisfaction where price is greater than equal to 100 and less than equal to 200.

Answer:

Question 21: Plot a histogram of overall satisfaction where price is less than 50.
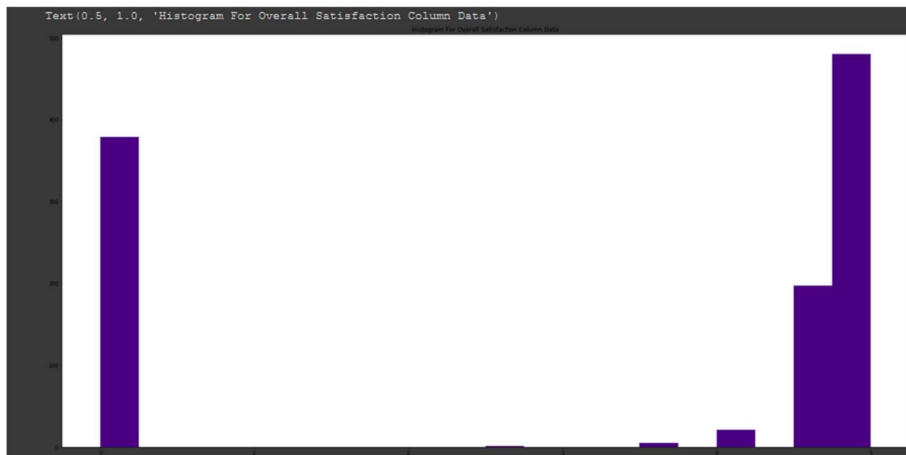
Answer:

Question 22: Plot a histogram of overall satisfaction where price is greater than 200.
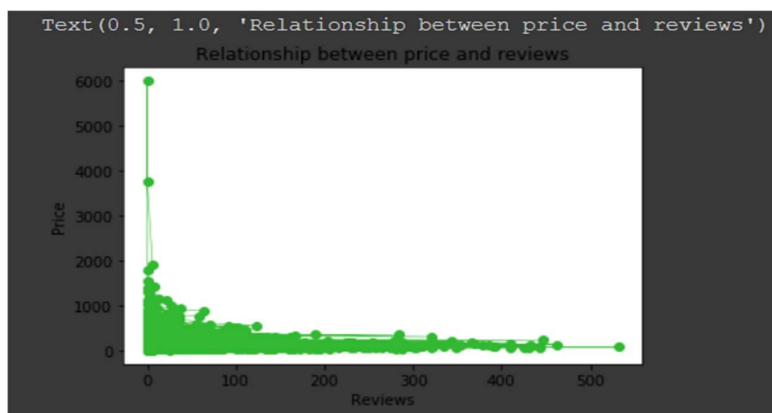
Answer:

Question 23: Plot a histogram of overall satisfaction where price is greater than 200.

Answer:

Question 24: Is there a relationship between price and reviews?
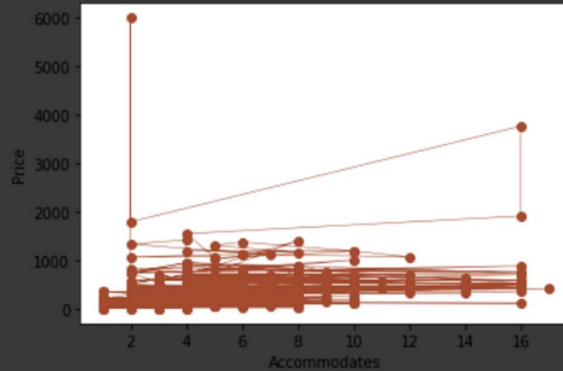
Answer:

- Inference:

    1.There is a negative trend relationship between price and reviews.

    2. If the no. of reviews increses, price will decreses.

Question 25: Is there a relationship between price and accommodates?

Answer:

Text(0.5, 1.0, 'Relationship between price and accommodates ')

Relationship between price and accommodates
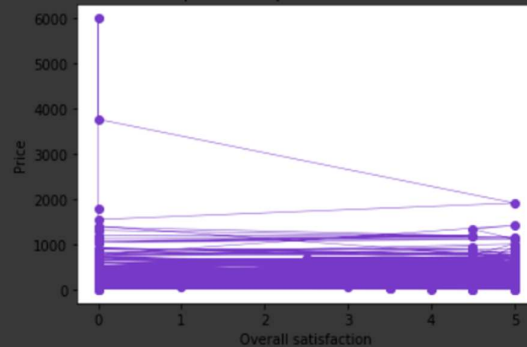

Pearsons correlation: 0.502

- Inference

    1. From the above graph and correlation value , it is clear that there is a slightly positive linear trend relationship between accommodates and price.

    2. If no. of accommodates increases, price will also increases

Question 26: Is there a relationship between price and Overall Satisfaction?

Answer:

Text(0.5, 1.0, 'Relationship between price and overall satisfaction ')

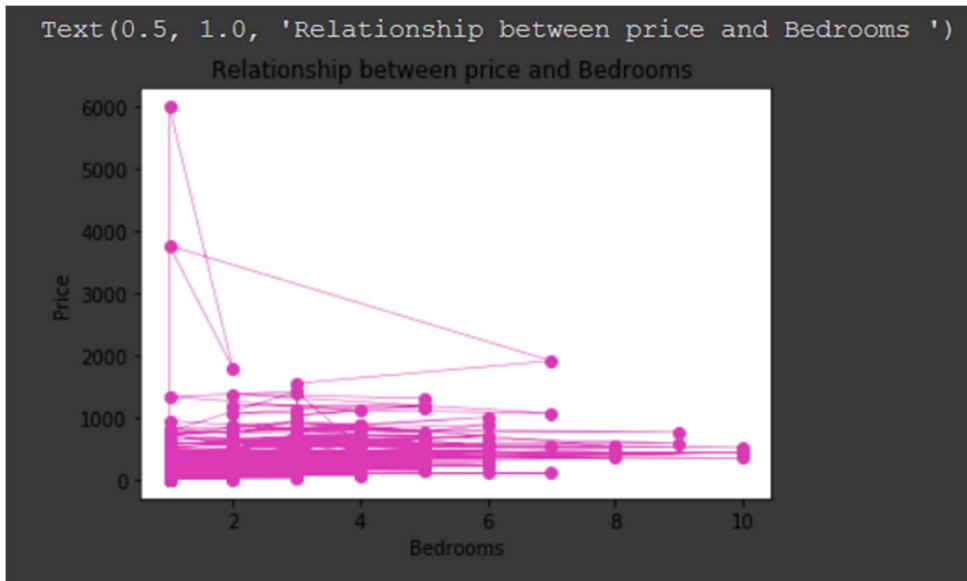Relationship between price and overall satisfaction

Pearsons correlation: -0.047

- Inference

    1. From the above graph and correlation value , it is clear that there is a negative linear trend relationship between overall_satisfaction and price.

    2. If overall satisfaction increases, price will decreases.

Question 27: Is there a relationship between price and Bedrooms?
Answer:


```
    Pearsons correlation: 0.459
```

Inference

1. From the above graph and correlation value , it is clear that there is no linear trend relationship between bedrooms and price.

2. The graph is uncorrelated

5.5 Conclusion:

Based on the data analysis and visualizations, several insights are derived and all the questions in the case study are answered.

Source Code:

```python
# 1. Import important numpy,pandas,matplotlib and seaborn libraries.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline

import warnings

warnings.filterwarnings('ignore') #to ignore warnings

# 2. Read the csv file and create pandas dataframe object airbnb

airbnb=pd.read_csv("airbnb data.csv")

# 3. Review first 5 rows

airbnb.head()

## Handling missing values

airbnb.isnull().sum()

# printing the 52 records containing nan values in name column in airbnb
dataframe.

airbnb[airbnb['name'].isnull()]

# using drop() attribute we can drop the unwanted rows and columns

airbnb.drop(columns=['country','borough','bathrooms','minstay'],inplace=True)

airbnb.isnull().sum()

# drop the nan values in the name column

airbnb.dropna(inplace=True)

airbnb.isnull().sum()

# 4. What are the room type available?

room_types=airbnb['room_type'].value_counts()

print(room_types)

room_counts=airbnb['room_type'].value_counts().to_list()

print(room_counts)
```

```python
#visualize using pie chart

labels = ['Entire home/apt','Private room','Shared room']

explode = (0.1, 0, 0)  # only "explode" the 1st slice

fig1, ax1 = plt.subplots()
ax1.pie(room_counts, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=90)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.

plt.show()
# 5.Drop the shared room type as its count is very low
airbnb.drop(airbnb.index[airbnb['room_type'] == 'Shared room'], inplace =
True)
airbnb['room_type'].value_counts()
# 6. Take a count of each neighborhood to show how popular they are.
airbnb['neighborhood'].value_counts()
# visualizing using bar graph
neigh_counts=airbnb['neighborhood'].value_counts().to_list()
figure=plt.figure(figsize=[40,20])
plt.bar(airbnb['neighborhood'].value_counts().index,neigh_counts,color='red',)
plt.xlabel("Neighborhood Names")
plt.ylabel("Popularity Counts")
plt.title("Neighborhood Popularity")
# 7. Number of reviews that is given for property
reviews_count=pd.Series(airbnb['reviews'].value_counts())
pd.set_option("display.max_rows", None, "display.max_columns", None)
```

```python
print(reviews_count)

reviews_count.hist(bins=50)

# 8. What is the overall satisfaction count by each rating

airbnb['overall_satisfaction'].value_counts()

#Visualize using bar chart


fig=plt.figure(figsize=[20,10])

plt.bar(airbnb['overall_satisfaction'].value_counts().index,airbnb['overall_satisfa
ction'].value_counts().to_list(),color="green",width=0.2)

plt.xlabel("Satisfaction Ratings")

plt.ylabel("Ratings Counts")

plt.title("Overall Satisfaction Ratings")

# 9. Create a sub dataset where overall satisfaction is zero

zero_sat_data=airbnb[airbnb['overall_satisfaction']==0.0]

# 10. Describe this dataset using describe command.

airbnb.describe()

# 11. What is the satisfaction ratings count when bedroom is 3

bedroom_3_data=airbnb[airbnb['bedrooms']==3.0]

bedroom_3_ratings=bedroom_3_data['overall_satisfaction'].value_counts()

bedroom_3_ratings

# Visualize using bar chart

sizes=bedroom_3_ratings.to_list()

labels=bedroom_3_ratings.index


plt.bar(labels,sizes,width=0.3,color='blue')

plt.xlabel("Ratings")

plt.ylabel("Count")

plt.title("satisfaction ratings count when bedroom is 3")
```

```python
plt.show()

# 12. What is the satisfaction ratings count when bedroom is 3

bedroom_2_data=airbnb[airbnb['bedrooms']==2.0]

bedroom_2_ratings=bedroom_2_data['overall_satisfaction'].value_counts()

bedroom_2_ratings

# Visualize using bar chart

sizes=bedroom_2_ratings.to_list()

labels=bedroom_2_ratings.index


plt.bar(labels,sizes,width=0.4
    ,color='black')

plt.xlabel("Ratings")

plt.ylabel("Count")

plt.title("satisfaction ratings count when bedroom is 2")

plt.show()


# 13. What is the satisfaction ratings count where price is greater than equal to
150 but less than equal to 180?

price_filter_data=airbnb[(airbnb.price>=150) & (airbnb.price<=180)]

price_filter_data_ratings=price_filter_data['overall_satisfaction'].value_counts()

print(price_filter_data_ratings)

# Visualize using bar chart

sizes=price_filter_data_ratings.to_list()

labels=price_filter_data_ratings.index


plt.bar(labels,sizes,width=0.4
    ,color='maroon')

plt.xlabel("Ratings")
```

plt.ylabel("Count")

plt.title("Satisfaction ratings count where price is greater than equal to 150 but less than equal to 180")

plt.show()

14. Find out the count of each accomodation value to understand where is the most availability

accomodation_counts=airbnb['accommodates'].value_counts()

accomodation_counts

#visualize this using bar chart

sizes=accomodation_counts.to_list()

labels=accomodation_counts.index

plt.bar(labels,sizes,width=0.7

    ,color='indigo')

plt.xlabel("Accomodation")

plt.ylabel("Count")

plt.title("count of each accomodation value to understand where is the most availability")

plt.show()

15. On which price most of the properties available

most_prop_available=airbnb['price'].value_counts()

print(most_prop_available)

data={'Price':most_prop_available.index.to_list(),'Properties Available':most_prop_available.to_list()}

most_prop_available_df=pd.DataFrame(data)

most_prop_available_df

16. What is the proportion of the bedroom values in the overall bedroom values.

```
bedroom_props=airbnb['bedrooms'].value_counts()

Totalbedrooms=bedroom_props.sum()

data={'Bedroom_values':bedroom_props.index.to_list(),'Proportion':bedroom_props.to_list()}

bedroom_props_df=pd.DataFrame(data)

print("Total no. of Bedrooms available : ",Totalbedrooms)

print('\n')

print("Each bedroom proportion : ")

bedroom_props_df

#visualize using bar chart

sizes=bedroom_props.to_list()

labels=bedroom_props.index.to_list()

plt.bar(labels,sizes,width=0.7
    ,color='orange')

plt.xlabel("Bedroom Values")

plt.ylabel("Proportion")

plt.title("proportion of the bedroom values in the overall bedroom values.")

plt.show()

#17. Drop the values when bedroom are equal to zero

airbnb.drop(airbnb.index[airbnb['bedrooms']==0.0],inplace=True)

#18. Now take the count of bedroom again which is without the zero value.

bedroom_props=airbnb['bedrooms'].value_counts()
```

```python
Totalbedrooms=bedroom_props.sum()


data={'Bedroom_values':bedroom_props.index.to_list(),'Proportion':bedroom_props.to_list()}


bedroom_props_df=pd.DataFrame(data)

print("Dataframe after dropping 0 values in the bedrooms column : ")

bedroom_props_df

# 19.Now check the info of entire dataset again

airbnb.info()

# 20.Plot a histogram of reviews where bins are 20 and range is between 0 to 100

fig=plt.figure(figsize=[20,10])

plt.hist('reviews',bins=20,range=(1,100),data=airbnb,color="green")

plt.title("Histogram For Reviews Column Data")

# 21.Plot a histogram of reviews where bins are 20 and range is between 0 to 1000

fig=plt.figure(figsize=[20,10])

plt.hist('reviews',bins=20,range=(1,1000),data=airbnb,color="blue")

plt.title("Histogram For Reviews Column Data")

# 22.Plot a histogram of accomodates where bins are 16

fig=plt.figure(figsize=[30,15])

plt.hist('accommodates',bins=16,data=airbnb,color="violet")

plt.title("Histogram For Accomodates Column Data")

# 22.Plot a histogram of overall_satisfaction

fig=plt.figure(figsize=[30,15])

plt.hist('overall_satisfaction',bins=25,data=airbnb,color="brown")

plt.title("Histogram For Overall Satisfaction Column Data")
```

23. Plot a histogram of overall satisfaction where price is greater than equal to 100 and less than equal to 200

filter_price=airbnb[(airbnb['price']>=100) & (airbnb['price']<=200)]

filter_price.head()

#plotting a histogram

fig=plt.figure(figsize=[30,15])

plt.hist('overall_satisfaction',bins=25,data=filter_price,color="blue")

plt.title("Histogram For Overall Satisfaction Column Data")

24. Plot a histogram of overall satisfaction where price is less than 50

filter_price=airbnb[(airbnb['price']<50)]

filter_price.head()

#plotting a histogram

fig=plt.figure(figsize=[30,15])

plt.hist('overall_satisfaction',bins=20,data=filter_price,color="red")

plt.title("Histogram For Overall Satisfaction Column Data")

25. Plot a histogram of overall satisfaction where price is greater than 200

filter_price=airbnb[(airbnb['price']>200)]

filter_price.head()

#plotting a histogram

fig=plt.figure(figsize=[30,15])

plt.hist('overall_satisfaction',bins=20,data=filter_price,color="black")

plt.title("Histogram For Overall Satisfaction Column Data")

26. Plot a histogram of overall satisfaction where price is greater than 200

filter_price=airbnb[(airbnb['price']>300)]

filter_price.head()

#plotting a histogram

fig=plt.figure(figsize=[30,15])

plt.hist('overall_satisfaction',bins=20,data=filter_price,color="indigo")

plt.title("Histogram For Overall Satisfaction Column Data")

#27. Is there a relationship between price and reviews

plt.plot(airbnb['reviews'],airbnb['price'],marker='o',c='#32b832',linewidth=0.5)

plt.xlabel("Reviews")

plt.ylabel("Price")

plt.title("Relationship between price and reviews")

Finding the relationship using Pearson Correlation method

from scipy.stats import pearsonr

correlation, _ = pearsonr(airbnb['reviews'],airbnb['price'])

print('Pearsons correlation: %.3f' % correlation)

#28. Is there a relationship between price and accommodates

plt.plot(airbnb['accommodates'],airbnb['price'],marker='o',c='#a34a2f',linewidth
=0.5)

plt.xlabel("Accommodates")

plt.ylabel("Price")

plt.title("Relationship between price and accommodates ")

#29. Is there a relationship between price and Overall Satisfaction

plt.plot(airbnb['overall_satisfaction'],airbnb['price'],marker='o',c='#783ac9',line
width=0.5)

plt.xlabel("Overall satisfaction")

plt.ylabel("Price")

plt.title("Relationship between price and overall satisfaction ")

#30. Is there a relationship between price and Bedrooms

plt.plot(airbnb['bedrooms'],airbnb['price'],marker='o',c='#db39b5',linewidth=0.5
)

plt.xlabel("Bedrooms")

plt.ylabel("Price")

plt.title("Relationship between price and Bedrooms ")

```
# Finding the relationship using Pearson Correlation method
correlation, _ = pearsonr(airbnb['bedrooms'],airbnb['price'])
print('Pearsons correlation: %.3f' % correlation)
```

CHAPTER 6
CONCLUSION AND FUTURE ENHANCEMENT

6.1 FUTURE ENHANCEMENT:

We can create a Machine Learning model that will predict the price of the Homes/Apartments which are rent out by the hosts by using Regression techniques and other predictive modelling techniques.

6.2 CONCLUSION:

Data can add a great value to companies and businesses when used in right way. In this modern world, we are generating a lot of data every day. If we use it to gain useful insights, it can prove to be of great worth. In short, Data is valuable to any company in any industry. In this project we have used data to understand more about the Airbnb prices, ratings and more. Thus, in this project data has helped us to understand more about Airbnb homes and prices,user needs and what they prefer.

CHAPTER 7

REFERENCES

1.   https://www.kaggle.com/

2.   https://pandas.pydata.org/docs/

3.   https://matplotlib.org/