GOVT. OF INDIA- RNI NO. UPBIL/2014/56766
UGC Approved Care Listed Journal
ISSN 2348-2397

SPECIAL ISSUE

शोध सरिता

An International Multidisciplinary Quarterly
Bilingual Peer Reviewed Refereed Research Journal

• Vol. 7 • Issue 25 • January to March 2020

National Conference On
"Current Trends in Management - Changes & Challenges"
27th & 28th February - 2020

Organized by
Yashaswi Education Society's International Institute of Management Science Chinchwad,
Pune In association with Associations of Indian Management School (AIMS)

Editor in Chief
Dr. Vinay Kumar Sharma
D. Litt. - Gold Medalist

SANCHAR
Educational & Research Foundation

CONTENTS

ISSN - 2348-2397
APPROVED UGC CARE
SHODH SARITA
SHODH SARITA
Vol. 7, Issue 25, January to March 2020
Page Nos. 116-119
AN INTERNATIONAL BILINGUAL PEER REVIEWED REFEREED RESEARCH JOURNAL

ANALYSIS OF CUSTOMER'S HOTEL BILLS USING SEABORN OF PYTHON

Prof. Prashant N. Wadkar*
Dr. Shivaji D. Mundhe**

ABSTRACT

The analysis of Hotel bills is the main research for this research paper. The Secondary data has taken online for this study. The main objective of this study is to analyze the data using payhon and its library seaborn. This Analysed the data and to predicts the prospective customers and make the facilities in the hotels accordingly. Here we have used the Python and Seaborn library for visualization. This study also reveals the trend of the customers their friends and family.

Keywords: *Python, Seaborn,*

Introduction :

Now a days any type of business needs to satisfy current customers and think to attract the new customers. Not only this they have to study the customer paying habits, capacity. Not only this it is also necessary to do gender wise analysis of customers. In this study the secondary data of Hotel is taken for on line analysis. The Objective of the study is stated as below.

1. Objectives :

1. To study the behavior of Hotel Customer.
2. To analyze Billing details.
3. Predicting business potential for hotel industry.

2. Significance of the study :

The significance of the study is mainly important because the ultimate aim of Hotel industry is to satisfy their customers. So to satisfy them it is necessary to analyze them and give the service accordingly and increase the profit in the business.

3. Research methodology adopted :

This is study on secondary data, so instead of primary data the main focus is on secondary data what we received and studying it by applying python library seaborn.

4. Hypothesis

H1 The no. of customer's are more whose family/friend size coming at hotel is 2

H2. No. of customers are more on Saturday and Sunday.

H3. The maximum billing amount of customer is in between 10 to 20 dollar

5. Data Analysis :

The Secondary Data which we received online, for this we proceed further for analyzing by different algorithms/coding in python by making use of seaborn library. Below are the outputs after applying algorithm/coding in python.

*Research Student, Asst. Prof. ASM's IIBR, Pune, India
**Research Guide Director IIMS, Pune, India

5.1 Importing seaborn library

```
import seaborn as sns
                    #Importing seaborn library
```

5.2 Loading tips database into dataframe df

```
df=s1.load_dataset('tips')
  # Loading tips database into dataframe df
```

5.3 Displaying dataframe

```
df
  # Displaying dataframe. (only limited rows
                              shown here)
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
5	25.29	4.71	Male	No	Sun	Dinner	4

244 rows × 7 columns

Table 1. Output showing Records in a Hotel billing Database.

The above output shows that there are 244 rows and 7 columns as shown above

5.4. Displaying structure information of the dataset

```
df.info()                    # Displaying
      structure information of the database
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
total_bill    244 non-null float64
tip           244 non-null float64
sex           244 non-null category
smoker        244 non-null category
day           244 non-null category
time          244 non-null category
size          244 non-null int64
dtypes: category(4), float64(2),
int64(1)
memory usage: 7.0 KB
```

Table 2. The Structure of the dataset

5.5 Getting head information of the dataset

```
df.head()                    # Getting
      head information of the dataset
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

Table 3. Output by head function

5.6 The number of males and females in the dataset

```
df['sex'].value_counts()
  Male     157
  Female    87
  Name: sex, dtype: int64
sns.countplot(x='sex', data =df)
```

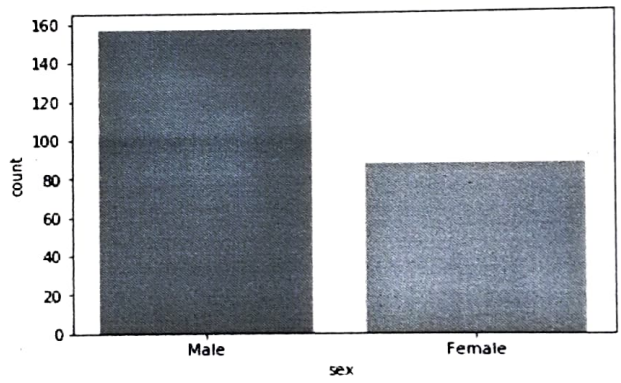

Chart 1. Number of males and Females

The above output depicts that there are 157 Male customers and 87 Female Customers who paid the bill.

5.7 Mean of the total bill

```
df['total_bill'].mean()
19.785942622950824
```

The 19.78 is the mean total bill paid by the customers.

5.8 Number of male & female non-smokers and smokers

```
df[(df['smoker']=='No') & (df['sex']=='Male')].shape[0]  #Male non-smokers
```

O/P : 97

```
df[(df['smoker']=='Yes') & (df['sex']=='Male')].shape[0]  #Male smokers
```

O/P : 60

```
df[(df['smoker']=='No') & (df['sex']=='Female')].shape[0]  #Female non-smokers
```

O/P : 54

```
df[(df['smoker']=='Yes') & (df['sex']=='Female')].shape[0]  #Female smokers
```

O/P : 33

```
sns.countplot('sex', data=df, hue='smoker')
```

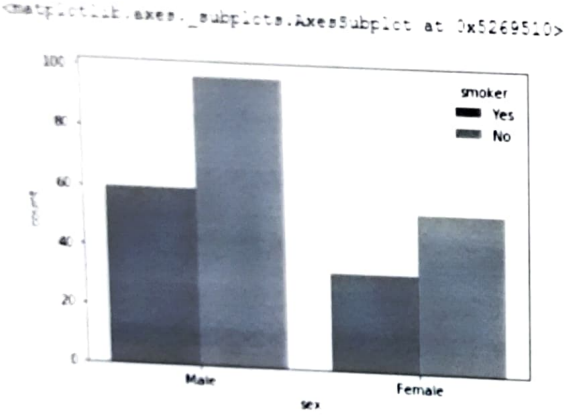

Chart 2. Number of male & female non-smokers and smokers

```
sns.countplot(x='smoker', data=df)
```

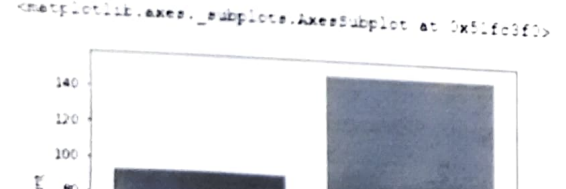

Chart 3. Total Number smokers and non-smokers

From above we observe that out of 157 males there are 97 non-smokers and 60 smokers. So the number of non-smokers are more as compared to smokers. Whereas out of total 87 Females there are 54 non-smokers and 33 are smokers.

1. Testing of Hypothesis:

H1: The no. of customers are more with family/friend size coming at hotel is 2

6.1 Family/friend size coming at hotel

```
df['size'].value_counts()
```

Family/friend size coming at hotel

O/p	Size	Count
	2	156
	3	38
	4	37
	5	5
	6	4
	1	4

Above output shows there are 156 billings where size of the family is 2, and for 38 billings for family size is 3, and 4 for 37, 5 for 5, 6 for 4, and 1 for 4. It depicts only 2 members billing is more than family size 3,4,5,6 and 1. From above it seems that the Hypothesis H1 is proved. So H1 is accepted.

H2: No. of customers are more on Saturday and Sunday.

6.2 No of Customers in a week

```
sns.countplot(x='day', data=df)
```

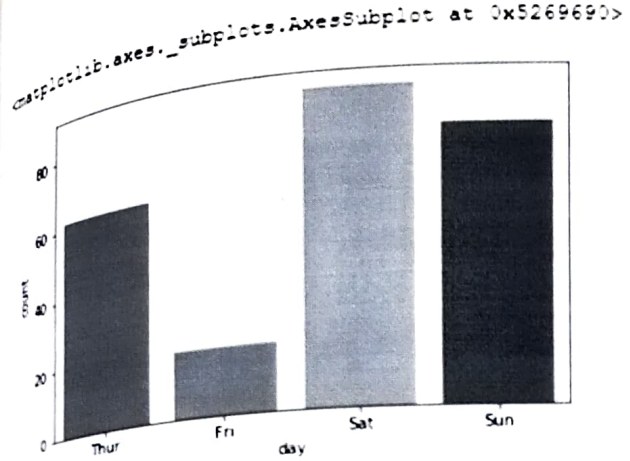

Chart 4. No of Customers in a week

Above chart shows there are more customers on Saturday and Sunday. From above chart it seems that the Hypothesis H2 proved. So H2 is accepted.

H3: The maximum billing amount of customer is in between 10 to 20 dollar

5.11 Chart for total bill by using distplot, kdeplot and rugplot simultaneously.

```
sns.distplot(df['total_bill'])
sns.kdeplot(df['total_bill'])
sns.rugplot(df['total_bill'])
```

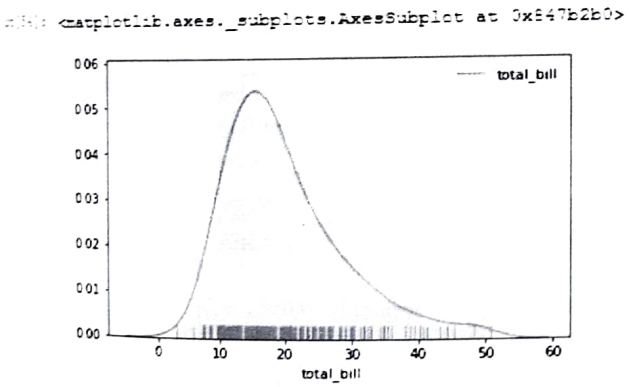

Chart 5. Chart shows there are more billings between 10 to 20 dollar

It has been observed from the above all three type of chart (3 in 1) that the maximum billing amount of customer is in between 10 to 20 dollar. Therefore the Hypothesis H3 is proved. So H3 is accepted.

7. Conclusion :

In above analysis of Hotel bills using the Python and seaborn we interpreted many of things as below. Firstly the Python and seaborn facilitated very well and done a lot of work. So it proved the best technology for analysis and visualizations. It has proved all hypothesis and they were like the H1 No. of customers are more whose family/friend size coming at hotel is 2 . Secondly The H2 No. of customers are more on Saturday and Sunday. And also H3. The maximum billing amount of customer is in between 10 to 20 dollar. So our research reveals that it is helped us to interpret the things and also prove the hypothesis. This will help to take further decisions for improving services and sustaining in hotel industry.

8. References :

1. Users' Sentiment Analysis in Social Media Context using Natural Language Processing by Nitish Bhardwaj1, Anupam Shukla2, Pradip Swarnakar3 1, 2, 3 (Department of Information Technology, ISBN:978-0-9891305-6-1 ©2014
2. Sentiment Analysis of Twitter Data using NLTK in Python by, Prateek Garg, Compuer Science and engineering department, Thapar university, June 2016, a thesis for M.Tech.
3. https://www.greycampus.com/blog/data-science/sentiment-analysis-on-twitter-tweets-using- Python
4. Study of Sentiment Analysis from Social Media using Python by Prof. Prashant N. Wadkar and Hidayat Peerjade ASM's International Conference on "Ongoing Research in Management & IT, Pune in the proceeding MARKETING MANAGEMENT & INFORMATION TECHNOLOGY ISBN : 978-93-89835-81-6, JANUARY, 2020
5. www.Kaggle.com