**City** **Hotel Data Analysis**

Unearthing useful insights from the vast amount of hotel data at your disposal is always going to be on-going and somewhat time-consuming – but it doesn’t have to be as complicated as it seems.

By focusing your efforts on specific data sets and getting creative with how you use them, data analysis will prove to be an actionable and transformative tool for your hotel.

The overall goal of data analysis is to track patterns at your property and set yourself up to make accurate predictions.

In turn you can plan and strategies towards your ultimate desire: increased revenue.

The more technology hotels use, the more data becomes available to them. For hotel property owners and sales managers, these metrics are so valuable. When you learn to organize, analyze, and apply the right hotel data, you can:

* Create a better experience for guests
* Book more rooms and group events
* Outsmart your competition
* Grow your bottom line

## 

## **Background**

Have you ever thought about when to book a hotel room in the year? Or is it the best time to stay for the best daily rate? What if you want to predict whether the hotel may receive too many special requests? The hotel reservation data set can help you explore these issues!

**Dataset:**

This data set contains booking information for a city hotel and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has from the data.

# We will try to answer the following Questions

1. How Many Bookings Were Cancelled?
2. Which is the busiest month for hotels?
3. From which country most guests come?
4. How Long People Stay in the hotel?
5. Which market segment makes the highest number of bookings?
6. What is the most preferred meal by customers?
7. Which type of customer usually come to the hotel?
8. Which room type is in most demand?

## **DATA CLEANING USING STRUCTURED QUERY LANAGUAGE (SQL)**

## **Data understanding:**

Graphical user interface, application, table, Excel

Description automatically generated

above picture are screenshot of fields and part of the data. The original data has 32 fields, each with 15,644 rows.

**Data Cleaning**:

1. Select a subset of the fields required for the user portrait.

Table

Description automatically generated

1. Clean data and check data integrity.

Graphical user interface, text, application

Description automatically generated

We can see that only a small part of the country field is missing, consider using the mode to fill.

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

It turns out that there are vacancies in the agent and company fields. As shown in the figure below, there are too many vacancies in the company field. Consider deleting the field.  
Missing agent means that there is no travel agency, and you can use sc to fill in (0 is not used here because it is not sure whether the id of the agent is 0, so English characters are used to fill in directly)

Text

Description automatically generated

Check whether the adr (average daily rate) has a negative value and delete the record if there is any.

Text

Description automatically generated

City table data cleaning is over.

**EXPLORATORY DATA ANALYSIS (EDA) AND VISUALIZATION WITH PYTHON**

We will perform exploratory data analysis with python to get insight from the data.

Let’s start:

**Import Packages**

First Import necessary packages and import the dataset

# Import packages  
  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# import and display dataset  
df=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel.csv")

# show first five rows  
print(df.head())

Display the dataset

A screenshot of a computer

Description automatically generated

Now let’s do the fun part, extract the information from our data and try to answer our questions.

1. **How Many Bookings Were Cancelled?**

Let’s write a python program to display the graph.

# import packages  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# display graph of hotel booking canceled or not  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel.csv")  
plt.figure(figsize=(12,4))  
sns.countplot(x='is\_canceled', data= df1)  
plt.legend(labels=["0.0= booking cancel" '\n' "1.0= booking not\_cancel"])  
plt.title('Booking canceled or not')  
plt.show()

Here the result

Chart

Description automatically generated

Here 0 indicates booking is canceled and 1 indicates booking is not canceled. Hence, we can see that most of the time booking are not canceled, that is good for hotel.

1. **Which is the busiest month for hotels?**

Now we want to know in which month the people book the hotel. To answer this question, we will select the arrival\_date\_month feature and get its value count.

Chart, bar chart

Description automatically generated

As we can see that most of the booking were made from month of April to June and peak in month of October. least booking was made at the start and end of the year.

City hotels should reasonably arrange the manpower and material resources of each quarter to ensure sufficient resource turnover during peak periods, reduce manpower and material resources during the off-season, and complete cost reduction.

1. **From which country most guests come?**

To see the country wise comparison, plot the **country**column.

Chart

Description automatically generated with medium confidence

Arrival of guests is high and above from the countries with code: DEU, ESP, FRA, GBR and PRT

Portugal, France, and Spain and Great Britain Germany are the top countries, most guests come from these 5 countries.

1. **How Long People Stay in the hotel?**

Let’s see the stay duration trend for hotel.

# How long most people stay in hotel.  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel.csv")  
df=df1['stays\_in\_week\_nights'].value\_counts().head(7)  
print(df)  
mylabels=["1=1st day","2=2nd day","3=3rd day","4=4th day","5=5th day","6=6st day","7=7th day"]  
plt.pie(df, labels=mylabels)  
plt.title("Number of Days People Stay in Hotel")  
plt.show()

Chart, pie chart

Description automatically generated

In City hotel, most popular stay duration is one, two, or three days.

Most people stay for one, two, or three. More than 60% of guests come under these three options.

1. **Which market segment makes the highest number of bookings?**

Most common booking by market segment.

# Import packages  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# total number of booking by market segment  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel Data.csv")  
plt.figure(figsize=(12,4))  
sns.countplot(y='market\_segment', hue = 'hotel', data= df1)  
plt.title('Total number of bookings by market segment')  
plt.show()

Chart, bar chart

Description automatically generated

As we can see that the greatest number of bookings done via Online through Travel Agent.

1. **What is the most preferred meal by customers?**

# most preferred meal by customer  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel Data.csv")  
plt.figure(figsize=(12,5))  
sns.countplot(x='meal', hue='meal', data= df1)  
plt.title('Most preferred meal by customer')  
plt.show()

Chart, waterfall chart

Description automatically generated

Most people were booked BB meal (Bed & Breakfast) and least people were booked HB meal (Half Board). Hotel can make a meal plan according to report.

1. **Which type of customer usually come to the hotel?**

This question will let us know which type of customer are usually come to the hotel. Is it Transient type or Transient party or Contract or Group?

# which type of customer usually come to the hotel  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel Data.csv")  
plt.figure(figsize=(12,5))  
sns.countplot(y='customer\_type', data= df1)  
plt.title('Type of customer')  
plt.show()

Chart

Description automatically generated

Most of the customers/guests were Transient type. More than 96% of customers were transient and least type of customers were in groups. Most of the guests are new guests, which is normal as a hotel.

1. **Which room type is in most demand?**

Let’s see most demanding room type.

# import packages  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
  
# which room type is in most demand  
df1=pd.read\_csv(r"C:\Users\manishgupta\Documents\City Hotel Data.csv")  
plt.figure(figsize=(12,5))  
sns.countplot(x='assigned\_room\_type', data= df1)  
plt.title('most demand room type')  
plt.show()

Chart, bar chart

Description automatically generated

First, let see A, B, D, E, F, G room type means.

A= One Room (one person or couple) B= Double Room

D= Triple Room E= Quad Room

F= Queen Size G= King size

As a following result most demand room type by customer is type A that is one room for one person or couple.

**Summary of conclusion**

* **City Hotel cancellation rate were low. Most of the time booking were not canceled and that is good for hotel.**
* **In City Hotel most booking was made from April to June and peak in month of October and least booking were made at start and end of the year.**
* **Portugal, France, Spain, and Germany were top countries from most guest come.**
* **In city hotel most popular stay duration is one, two and three days. Most people stay for one, two or three days. More than 60% of guest come under these three options.**
* **Guest uses different channel for making booking, out of which most preferred way were online through Travel Agent.**
* **Online TA market segment makes the highest number of bookings.**
* **Most people preferred BB meal (Bed & Breakfast).**
* **In city hotel most of the guest were Transient type. It can be seen that more than 96% of the guest are Transient.**
* **Room type A were most demand room.**
* **Couples are the most common guest for hotels; hence hotel can plan services according to couple needs to increase revenue.**