**Problem Statement**

**Business Problem**

In recent years, City Hotel and Resort Hotel have seen high cancellation rates. Each

hotel is now dealing with several issues as a result, including fewer revenues and

less than ideal hotel room use. Consequently, lowering cancellation rates is both hotels’ primary goal to increase their efficiency in generating revenue, and for us to

offer thorough business advice to address this problem.

The analysis of hotel booking cancellations as well as other factors that have no bearing on their business and yearly revenue generation are the main topics of this report.

**Assumptions**

1. No unusual occurrences between 2015 and 2017 will have a substantial impact on the data used.

2. The information is still current and can be used to analyze a hotel's possible plans in

an efficient manner.

3. There are no unanticipated negatives to the hotel employing any advised technique.

4. The hotels are not currently using any of the suggested solutions.

**Research Questions**

1. What are the variables that affect hotel reservation cancellations?

2. How can we make hotel reservations cancellations better?

3. How will hotels be assisted in making pricing and promotional decisions?

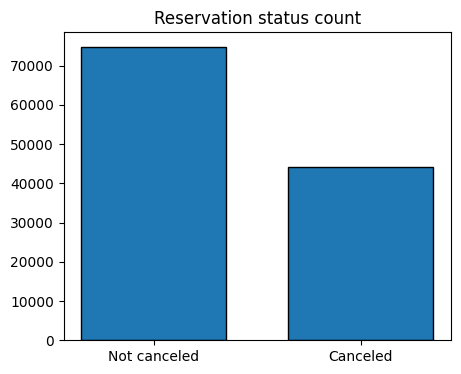
**Hypothesis**

1. More cancellations occur when prices are higher.

2. When there is a longer waiting list, customers tend to cancel more frequently.

3. Most clients are coming from offline travel agents to make their reservations

FINDINGS & Analysis



The accompanying bar graph shows the percentage of reservations that are cancelled

and those that are not. It is obvious that there are still a significant number of

reservations that have not been canceled. There are still 37% of clients who cancelled

their reservation, which has a significant impact on the hotels' earnings.

A picture containing text, screenshot, diagram, font

Description automatically generated

In comparison to resort hotels, city hotels have more bookings. It's possible that resort

hotels are more expensive than those in cities.

A picture containing screenshot, text, line, plot

Description automatically generated

The line graph above shows that, on certain days, the average daily rate for a city hotel

is less than that of a resort hotel, and on other days, it is even less. It goes without

saying that weekends and holidays may see a rise in resort hotel rates.

A graph of blue and orange bars

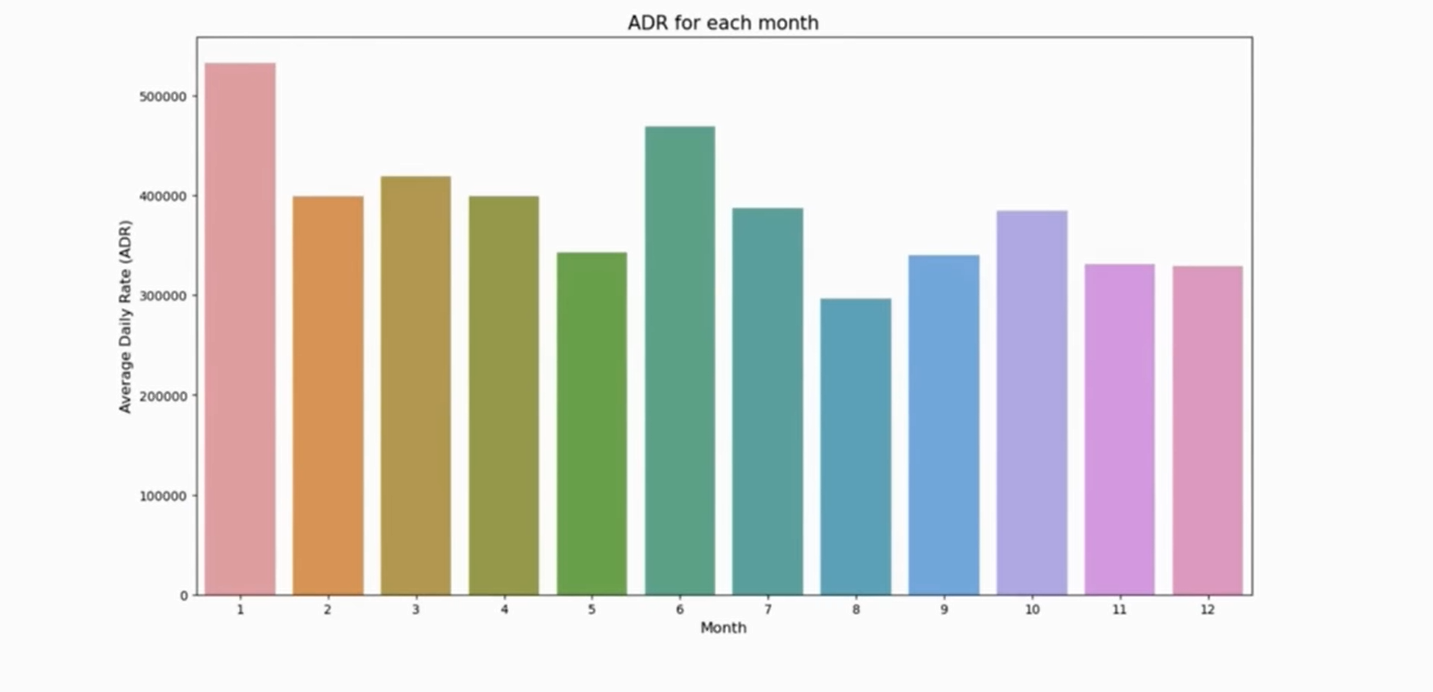
Description automatically generated with low confidence

I have developed the grouped bar graph to analyze the months with the highest and

lowest reservation levels according to reservation status. As can be seen, both the

number of confirmed reservations and the number of canceled reservations are largest

in the month of August. whereas January is the month with the most canceled reservations.



This bar graph demonstrates that cancellations are most common when prices are

greatest and are least common when they are lowest. Therefore, the cost of the

accommodation is solely responsible for the cancellation.

Now, let's see which country has the highest reservation canceled. The top country is

Portugal with the highest number of cancellations.

A picture containing text, screenshot, diagram, colorfulness

Description automatically generated

Let's check the area from where guests are visiting the hotels and making reservations.

Is it coming from Direct or Groups, Online or Offline Travel Agents? Around 46% of the

clients come from online travel agencies, whereas 27% come from groups. Only 4% of

clients book hotels directly by visiting them and making reservations.

A picture containing line, plot, font, screenshot

Description automatically generated

As seen in the graph, reservations are canceled when the average daily rate is higher

than when it is not canceled. It clearly proves all the above analysis, that the higher

price leads to higher cancellation.

Suggestions

1. Cancellation rates rise as the price does. In order to prevent cancellations of

reservations, hotels could work on their pricing strategies and try to lower the

rates for specific hotels based on locations. They can also provide some

discounts to the consumers.

2. As the ratio of the cancellation and not cancellation of the resort hotel is higher in

the resort hotel than the city hotels. So the hotels should provide a reasonable

discount on the room prices on weekends or on holidays.

3. In the month of January, hotels can start campaigns or marketing with a

reasonable amount to increase their revenue as the cancellation is the highest in

this month.

4. They can also increase the quality of their hotels and their services mainly in

Portugal to reduce the cancellation rate.

CODE:

#importing libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')

#loading dataset

df = pd.read\_csv('hotel\_booking.csv.zip')

#exploratory analysis

df.head()

df.tail()

df['reservation\_status\_date'] = pd.to\_datetime(df['reservation\_status\_date'])

df.info()

df.describe(include = 'object')

for col in df.describe(include = 'object').columns:

print(col)

print(df[col].unique())

df.isnull().sum()

df.drop(['company', 'agent'], axis = 1, inplace = True)

df.dropna(inplace = True)

df.isnull().sum()

df.describe()

df['adr'].plot(kind = 'box')

df = df[df['adr']<5000]

#data analysis visualization

cancelled\_perc = df['is\_canceled'].value\_counts(normalize = True)

print(cancelled\_perc)

plt.figure(figsize = (5,4))

plt.title('Reservation status count')

plt.bar(['Not canceled', 'Canceled'],df['is\_canceled'].value\_counts(), edgecolor = 'k', width = 0.7)

plt.show()

plt.figure(figsize = (8,4))

ax1 = sns.countplot(x = 'hotel', hue = 'is\_canceled', data = df, palette = 'Blues')

legend\_labels,\_ = ax1. get\_legend\_handles\_labels()

ax1.legend(bbox\_to\_anchor=(1,1))

plt.title('Reservation status in different hotels', size = 20)

plt.xlabel('hotel')

plt.ylabel('number of reservations')

plt.legend(['not canceled', 'canceled'])

plt.show()

resort\_hotel = df[df['hotel'] =='Resort Hotel']

resort\_hotel['is\_canceled'].value\_counts(normalize = True)

city\_hotel = df[df['hotel'] =='City Hotel']

city\_hotel['is\_canceled'].value\_counts(normalize = True)

resort\_hotel = resort\_hotel.groupby('reservation\_status\_date')[['adr']].mean()

city\_hotel = city\_hotel.groupby('reservation\_status\_date')[['adr']].mean()

plt.figure(figsize = (20,8))

plt.title('Average Daily Rate in City and Resort Hotel', fontsize = 30)

plt.plot(resort\_hotel.index, resort\_hotel['adr'], label = 'Resort Hotel')

plt.plot(city\_hotel.index, city\_hotel['adr'], label = 'City Hotel')

plt.figure(figsize = (20,8))

plt.title('Average Daily Rate in City and Resort Hotel', fontsize = 30)

plt.plot(resort\_hotel.index, resort\_hotel['adr'], label = 'Resort Hotel')

plt.plot(city\_hotel.index, city\_hotel['adr'], label = 'City Hotel')

df['month'] = df['reservation\_status\_date'].dt.month

plt.figure(figsize = (16,8))

ax1 = sns.countplot(x = 'month', hue = 'is\_canceled', data = df, palette = 'bright')

legend\_labels,\_ = ax1.get\_legend\_handles\_labels()

ax1.legend(bbox\_to\_anchor=(1,1))

plt.title('Reservation status per month', size = 20)

plt.xlabel('month')

plt.ylabel('number of reservations')

plt.legend(['not canceled', 'canceled'])

plt.show()

cancelled\_data = df[df['is\_canceled'] == 1]

top\_10\_countries = cancelled\_data['country'].value\_counts()[:10]

plt.figure(figsize = (12,10))

plt.title('Top 10 countries from where reservations are canceled')

plt.pie(top\_10\_countries, autopct = '%.2f', labels = top\_10\_countries.index)

df['market\_segment'].value\_counts()

df['market\_segment'].value\_counts(normalize = True)

cancelled\_df\_adr = cancelled\_data.groupby('reservation\_status\_date')[['adr']].mean()

cancelled\_df\_adr.reset\_index(inplace = True)

cancelled\_df\_adr.sort\_values('reservation\_status\_date', inplace = True)

not\_cancelled\_data = df[df['is\_canceled'] == 0]

not\_cancelled\_df\_adr = not\_cancelled\_data.groupby('reservation\_status\_date')[['adr']].mean()

not\_cancelled\_df\_adr.reset\_index(inplace = True)

not\_cancelled\_df\_adr.sort\_values('reservation\_status\_date', inplace = True)

plt.figure(figsize = (18,5))

plt.title('avg Daily Rate')

plt.plot(not\_cancelled\_df\_adr['reservation\_status\_date'],not\_cancelled\_df\_adr['adr'], label = 'not cancelled')

plt.plot(cancelled\_df\_adr['reservation\_status\_date'],cancelled\_df\_adr['adr'], label = 'cancelled')

plt.legend()