Analysis of Hotel Booking Cancellations

Executive Summary

City Hotel and Resort Hotel have been facing high cancellation rates in recent years, negatively impacting their revenue and room utilization. This analysis aims to identify the factors contributing to these high cancellation rates and provide strategic recommendations to reduce cancellations and enhance hotel efficiency.

Business Problem

City Hotel and Resort Hotel are experiencing significant challenges due to high cancellation rates, which lead to reduced revenue and suboptimal room occupancy. Addressing this issue is crucial for improving revenue generation and operational efficiency.

Assumptions

1. No unusual events between 2015 and 2017 significantly impacted the data used in this analysis.
2. The information remains current and applicable for evaluating the hotels' plans.
3. Implementing the advised techniques will not result in unforeseen negative outcomes for the hotels.
4. The hotels have not previously implemented any suggested solutions.
5. Booking cancellations significantly affect revenue generation.
6. Other factors not related to cancellations are irrelevant to this analysis.
7. Clients make hotel reservations and cancellations within the same year.

Research Questions

1. What variables influence hotel reservation cancellations?
2. How can we reduce hotel reservation cancellations?
3. How can hotels use this information to make better pricing and promotional decisions?

Hypotheses

1. Higher cancellation rates occur when prices are higher.
2. Longer waiting lists lead to more frequent cancellations.
3. Most cancellations come from offline travel agent bookings.

Analysis and Findings

1. Data Loading and Initial Exploration

The dataset was loaded and initial exploratory steps were taken to understand its structure and content.

```
import pandas as pd
import numpy
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	st
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	

5 rows × 36 columns

```
df['reservation_status_date']=pd.to_datetime(df['reservation_status_date'])
```

```
df.describe(include='object')
```

```
for col in df.describe(include='object').columns:
    print(col)
    print(df[col].unique())
    print('-'*50)
```

```
df.isnull().sum()
```

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

```
df.isnull().sum()
```

```
df.shape
```

```
df.describe()
```

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

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

```
hotel
['Resort Hotel' 'City Hotel']
--------------------------------------------------
arrival_date_month
['July' 'August' 'September' 'October' 'November' 'December' 'January'
 'February' 'March' 'April' 'May' 'June']
--------------------------------------------------
meal
['BB' 'FB' 'HB' 'SC' 'Undefined']
--------------------------------------------------
country
['PRT' 'GBR' 'USA' 'ESP' 'IRL' 'FRA' nan 'ROU' 'NOR' 'OMN' 'ARG' 'POL'
 'DEU' 'BEL' 'CHE' 'CN' 'GRC' 'ITA' 'NLD' 'DNK' 'RUS' 'SWE' 'AUS' 'EST'
 'CZE' 'BRA' 'FIN' 'MOZ' 'BWA' 'LUX' 'SVN' 'ALB' 'IND' 'CHN' 'MEX' 'MAR'
 'UKR' 'SMR' 'LVA' 'PRI' 'SRB' 'CHL' 'AUT' 'BLR' 'LTU' 'TUR' 'ZAF' 'AGO'
 'ISR' 'CYM' 'ZMB' 'CPV' 'ZWE' 'DZA' 'KOR' 'CRI' 'HUN' 'ARE' 'TUN' 'JAM'
 'HRV' 'HKG' 'IRN' 'GEO' 'AND' 'GIB' 'URY' 'JEY' 'CAF' 'CYP' 'COL' 'GGY'
 'KWT' 'NGA' 'MDV' 'VEN' 'SVK' 'FJI' 'KAZ' 'PAK' 'IDN' 'LBN' 'PHL' 'SEN'
 'SYC' 'AZE' 'BHR' 'NZL' 'THA' 'DOM' 'MKD' 'MYS' 'ARM' 'JPN' 'LKA' 'CUB'
 'CMR' 'BIH' 'MUS' 'COM' 'SUR' 'UGA' 'BGR' 'CIV' 'JOR' 'SYR' 'SGP' 'BDI'
 'SAU' 'VNM' 'PLW' 'QAT' 'EGY' 'PER' 'MLT' 'MWI' 'ECU' 'MDG' 'ISL' 'UZB'
 'NPL' 'BHS' 'MAC' 'TGO' 'TWN' 'DJI' 'STP' 'KNA' 'ETH' 'IRQ' 'HND' 'RWA'
 'KHM' 'MCO' 'BGD' 'IMN' 'TJK' 'NIC' 'BEN' 'VGB' 'TZA' 'GAB' 'GHA' 'TMP'
 'GLP' 'KEN' 'LIE' 'GNB' 'MNE' 'UMI' 'MYT' 'FRO' 'MMR' 'PAN' 'BFA' 'LBY'
 'MLI' 'NAM' 'BOL' 'PRY' 'BRB' 'ABW' 'AIA' 'SLV' 'DMA' 'PYF' 'GUY' 'LCA'
 'ATA' 'GTM' 'ASM' 'MRT' 'NCL' 'KIR' 'SDN' 'ATF' 'SLE' 'LAO']
--------------------------------------------------
market_segment
['Direct' 'Corporate' 'Online TA' 'Offline TA/TO' 'Complementary' 'Groups'
 'Undefined' 'Aviation']
--------------------------------------------------
distribution_channel
['Direct' 'Corporate' 'TA/TO' 'Undefined' 'GDS']
--------------------------------------------------
reserved_room_type
['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'P' 'B']
```

2. Cancellation Rate

Calculated the cancellation rate to understand the proportion of bookings that were canceled.

```
cancelled_perc=df['is_canceled'].value_counts(normalize=True)
print(cancelled_perc)
```

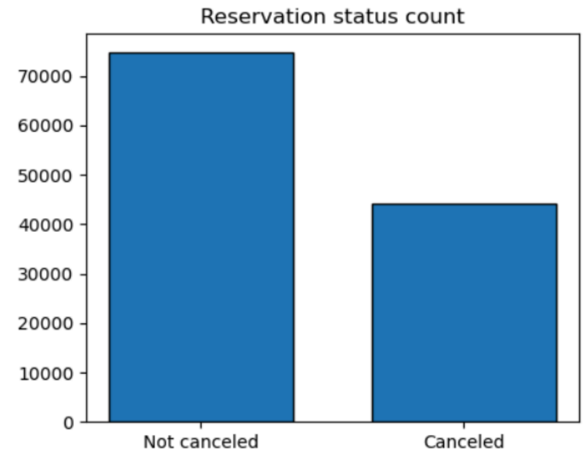
3. Reservation Status Count

A bar plot was created to visualize the count of canceled and non-canceled reservations.

```
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()
```

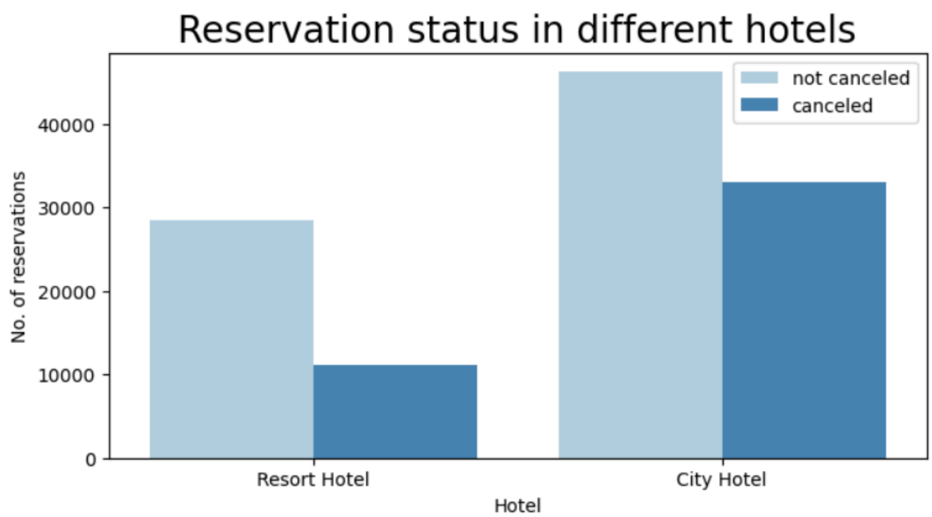
```
is_canceled
0    0.628648
1    0.371352
Name: proportion, dtype: float64
```



4. Reservation Status by Hotel Type

Explored the reservation status across different types of hotels (City Hotel and Resort Hotel).

```
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('No. of reservations')
plt.legend(['not canceled','canceled'])
plt.show()
```



5. Average Daily Rate (ADR) Analysis

Analyzed the average daily rate over time for both City Hotel and Resort Hotel.

```
resort_hotel=df[df['hotel']== 'Resort Hotel']
resort_hotel['is_canceled'].value_counts(normalize=True)
```

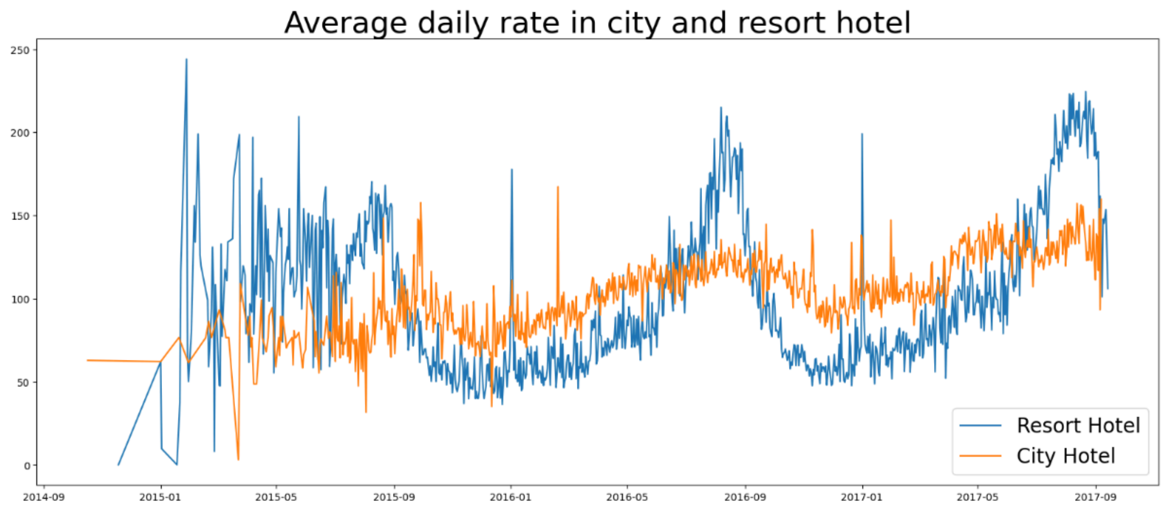
```
is_canceled
0    0.72025
1    0.27975
Name: proportion, dtype: float64
```

```
city_hotel=df[df['hotel']== 'City Hotel']
city_hotel['is_canceled'].value_counts(normalize=True)
```

```
is_canceled
0    0.582911
1    0.417089
Name: proportion, dtype: float64
```

```
resort_hotel=resort_hotel.groupby('reservation_status_date')[['adr']].mean()
city_hotel=city_hotel.groupby('reservation_status_date')[['adr']].mean()
```

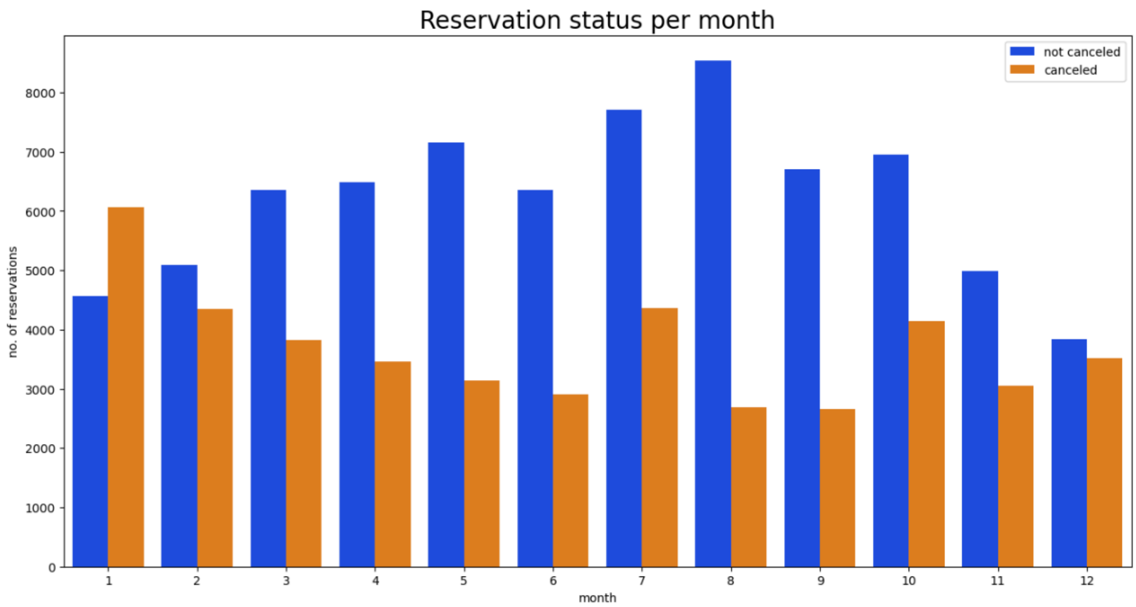
```
plt.figure(figsize=(20,8))
plt.title('Average daily rate in city and resort hotel',fontsize=30)
plt.plot(resort_hotel.index, resort_hotel['adr'], label='Resort Hotel')
plt.plot(city_hotel.index, city_hotel['adr'], label='City Hotel')
plt.legend(fontsize=20)
plt.show()
```



6. Monthly Reservation Status

Examined the monthly reservation status to identify seasonal patterns in cancellations.

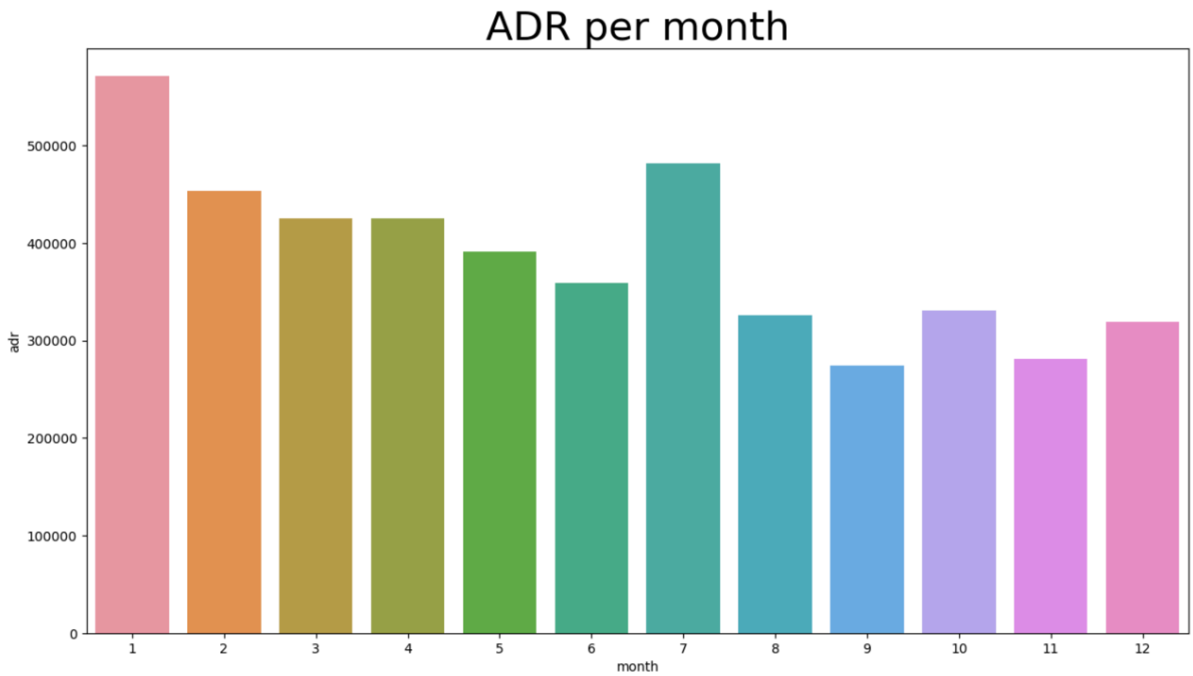
```
df['month']=df['reservation_status_date'].dt.month
plt.figure(figsize=(16,8))
ax1=sns.countplot(x='month',hue='is_canceled',data=df,palette='bright')
legend_labels, _ = ax1.get_legend_handles_labels()
ax1.legend(bbox_to_anchor=(1,1))
plt.title('Reservation status per month',size=20)
plt.xlabel('month')
plt.ylabel('no. of reservations')
plt.legend(['not canceled','canceled'])
plt.show()
```



7. ADR per Month for Canceled Reservations

Visualized the ADR per month for canceled reservations to identify any trends.

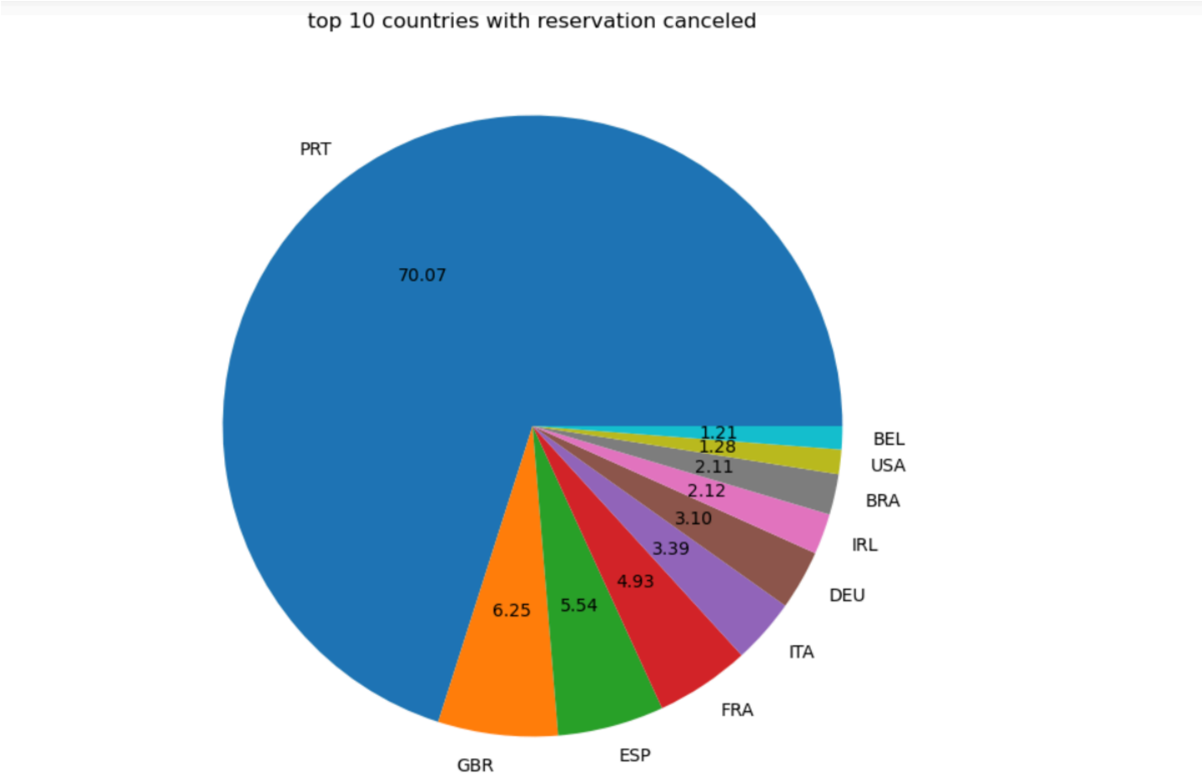
```
plt.figure(figsize=(15, 8))
plt.title('ADR per month', fontsize=30)
sns.barplot(x='month', y='adr', data=df[df['is_canceled'] == 1].groupby('month')[['adr']].sum().reset_index())
plt.show()
```



8. Top 10 Countries with Reservation Cancellations

Identified the top 10 countries with the highest number of reservation cancellations.

```
cancelled_data=df[df['is_canceled']== 1]
top_ten_country=cancelled_data['country'].value_counts()[:10]
plt.figure(figsize=(8,8))
plt.title('top 10 countries with reservation canceled')
plt.pie(top_ten_country,autopct='%.2f',labels=top_ten_country.index)
plt.show()
```



9. Market Segment Distribution

Explored the distribution of bookings across different market segments and analyzed cancellation rates by market segment.

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

```
market_segment
Online TA        56402
Offline TA/TO    24160
Groups           19806
Direct           12448
Corporate         5111
Complementary      734
Aviation           237
Name: count, dtype: int64
```

```
cancelled_data['market_segment'].value_counts(normalize=True)
```

```
market_segment
Online TA        0.469685
Groups           0.273979
Offline TA/TO    0.187484
Direct           0.043485
Corporate        0.022150
Complementary    0.002038
Aviation         0.001178
Name: proportion, dtype: float64
```

10. Average Daily Rate for Canceled and Non-Canceled Bookings

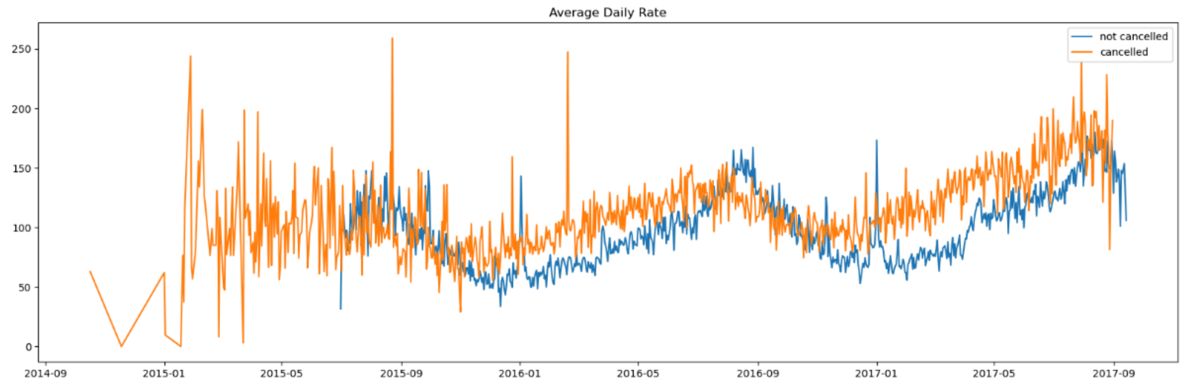
Plotted the ADR for canceled and non-canceled bookings over time to identify any significant differences.

```
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=(20,6))
plt.title('Average 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()
```

```
<matplotlib.legend.Legend at 0x1d6e06e62d0>
```



Recommendations

1. Adjust Pricing Strategies: Lowering room rates or offering special discounts based on location and demand can help reduce cancellations.

2. Targeted Promotions: Focus on marketing campaigns during high cancellation periods, such as January, to boost bookings and reduce cancellations.

3. Service Improvement: Enhance the quality and range of services offered, particularly in regions with higher cancellation rates, like Portugal, to encourage bookings and reduce cancellations.

4. Special Offers for Resort Hotels: Given the higher cancellation rates for resort hotels, especially on weekends or holidays, offering special discounts or packages during these times can help mitigate cancellations.

Conclusion

By understanding the factors influencing booking cancellations and implementing targeted strategies, City Hotel and Resort Hotel can significantly reduce their cancellation rates, thereby improving revenue and room occupancy. Continuous monitoring and adjustment of these strategies will be essential to maintain and improve the effectiveness of these measures.

Project Files

You can find the complete project files, including the dataset and analysis code, on my [GitHub repository].

Connect with Me

Feel free to connect with me on [LinkedIn] to discuss this project or explore collaboration opportunities.