Report on Data Analysis of Hotel Bookings

Overview

The data analysis focuses on a comprehensive dataset from a hotel booking system, examining several key metrics such as Average Daily Rate (ADR), cancellation rates, and market segmentation. The analysis was conducted using Python, with libraries including Pandas, Matplotlib, and Seaborn to manipulate the data and generate visual insights.

Data Preparation

The dataset was first loaded from a CSV file, `hotel_booking.csv`. Initial steps involved viewing the first and last few records, understanding the shape and structure of the data, and removing unnecessary columns such as `name`, `email`, `phone-number`, and `credit_card` to streamline the dataset.

Data types were adjusted with the `reservation_status_date` column converted to datetime format to facilitate time-series analysis. Missing values in the `company` and `agent` columns were identified and these columns were subsequently dropped from the dataset due to their high nullity.

Exploratory Data Analysis (EDA)

- The dataset was examined for any null values, which were addressed by removing rows where data was missing.
- Descriptive statistics were generated for both numerical and categorical data. This helped in understanding the distributions and unique values in columns such as `country` and `market_segment`.
- A box plot was created for the `adr` (Average Daily Rate) to identify outliers, and entries with an `adr` above 5000 were removed as they were deemed erroneous.

Visualization and Insights

- **Reservation Status**: A bar plot illustrated the count of canceled versus not canceled bookings. Cancellation rates were also visualized for different hotel types (City and Resort), showing distinct cancellation patterns.
- **Average Daily Rate (ADR) Trends**: Line plots were generated to show the ADR over time for both canceled and not canceled bookings, providing insights into pricing trends and their relationship with booking cancellations.
- **ADR by Month**: Seasonal variations in ADR were explored by aggregating data monthly. A colorful bar plot represented the monthly ADR, highlighting peak and low seasons.
- **Country Analysis**: A pie chart was used to depict the top 10 countries with the highest number of cancellations, offering a geographical perspective on where the majority of cancellations were originating.

- **Market Segmentation**: The distribution of different market segments was analyzed, with findings suggesting that certain segments had higher cancellation rates, potentially indicating segments that might require targeted marketing strategies.

Time-Specific Analysis

Data was filtered for the years 2016 to 2017 to focus on this specific period. Detailed time-series analysis was performed for both canceled and not canceled bookings, providing targeted insights for these years.

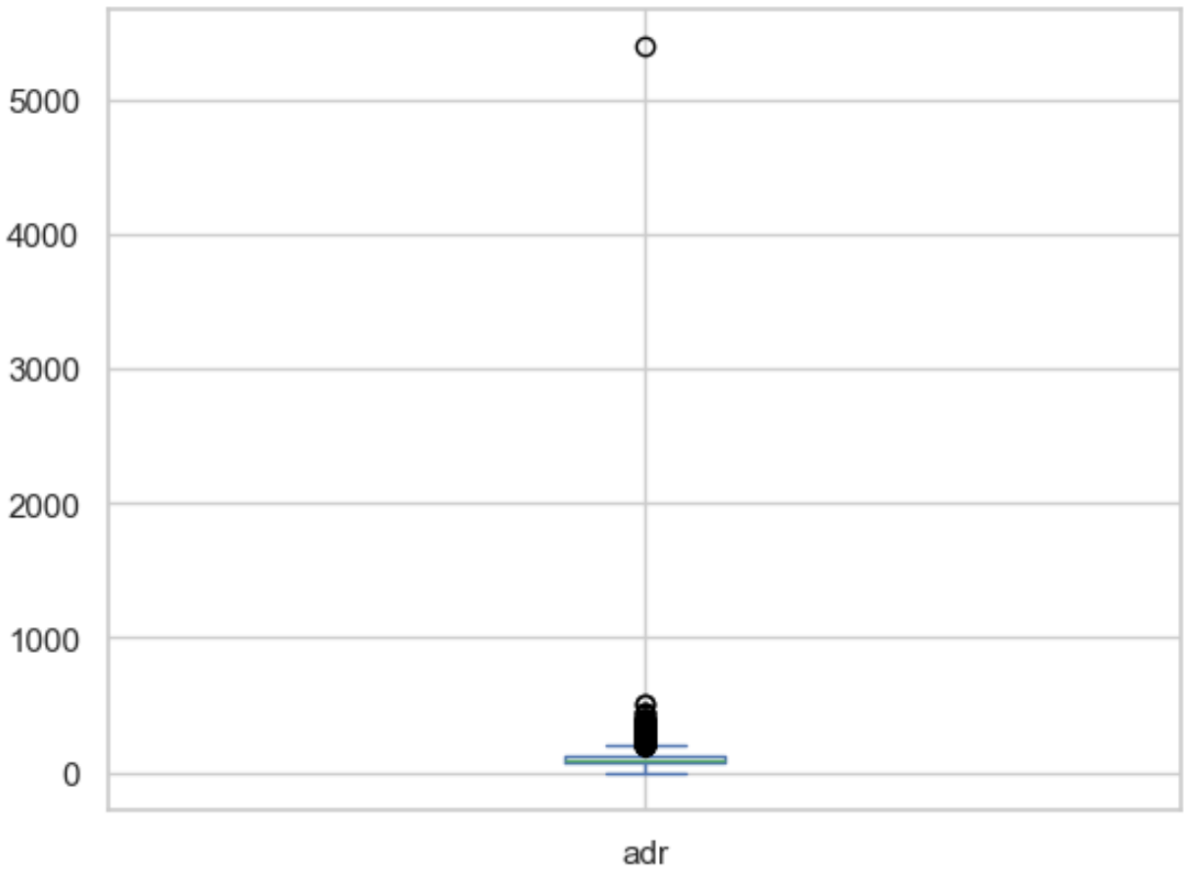
Conclusions

The analysis highlighted several key areas for potential improvement and further investigation:

- **Cancellation Patterns**: Understanding the reasons behind high cancellation rates in specific months and by market segments could help in devising better booking policies or promotional strategies.
- **Seasonal Pricing**: Insights into ADR fluctuations could aid in optimizing pricing strategies to increase revenue during low seasons and maximize profitability during peak periods.
- **Customer Geography**: Knowing which countries most cancellations come from might help tailor international marketing campaigns more effectively.

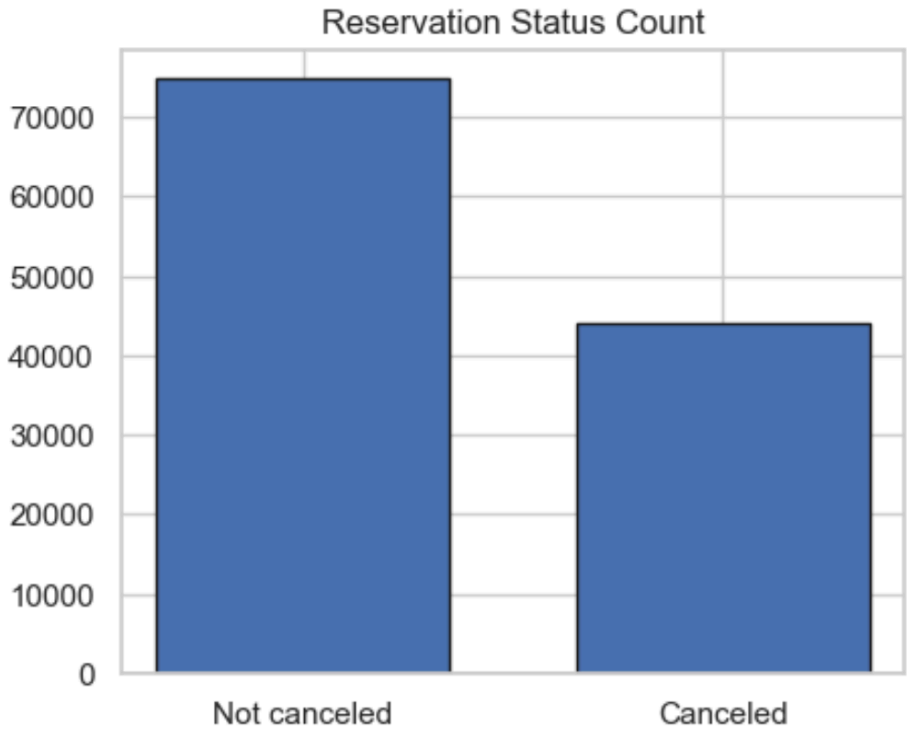
Recommendations

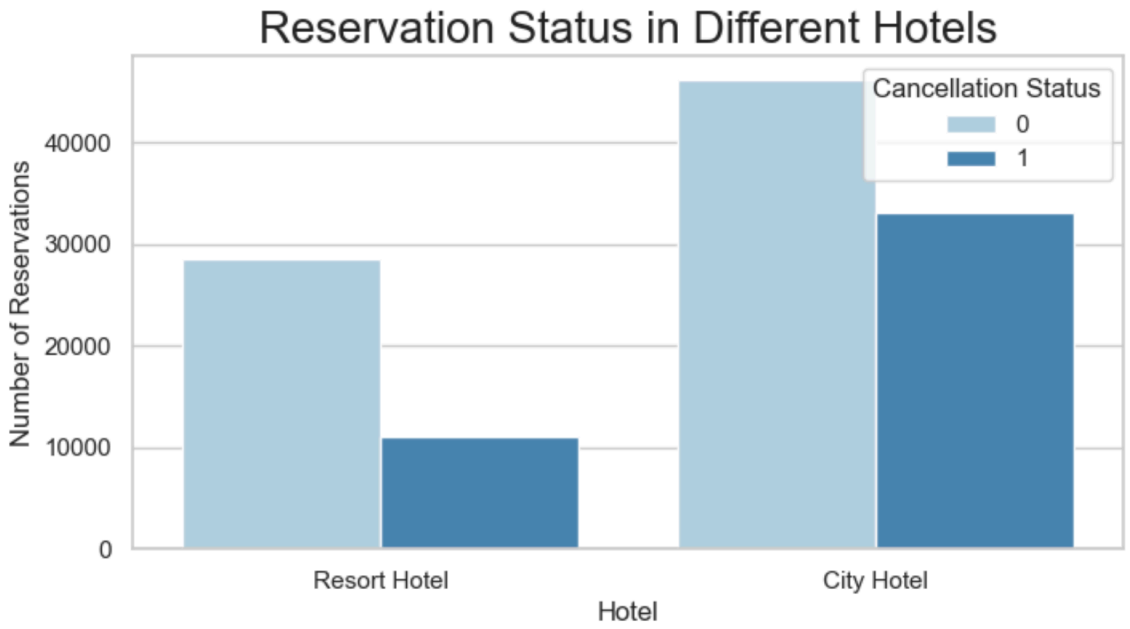
- **Dynamic Pricing Strategy**: Implement a dynamic pricing model that adjusts rates based on demand and cancellation trends observed in the data.
- **Focused Marketing**: Develop targeted marketing campaigns for segments with high cancellation rates and for international markets with high cancellation incidences.
- **Customer Retention Initiatives**: Introduce loyalty programs or incentives to reduce cancellation rates, particularly in segments or periods prone to higher cancellations.

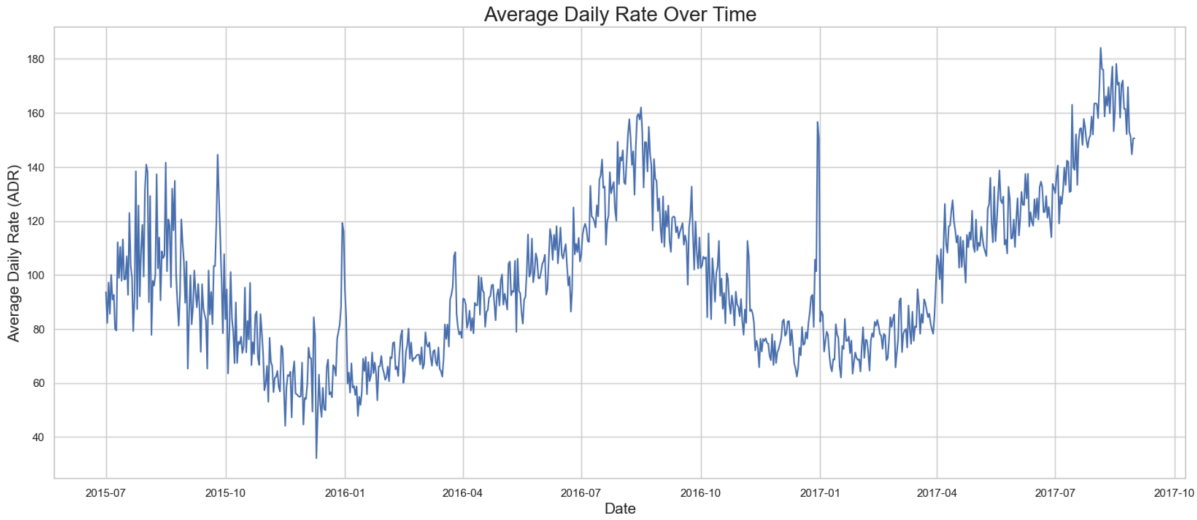

5000
4000
3000
2000
1000
0
adr

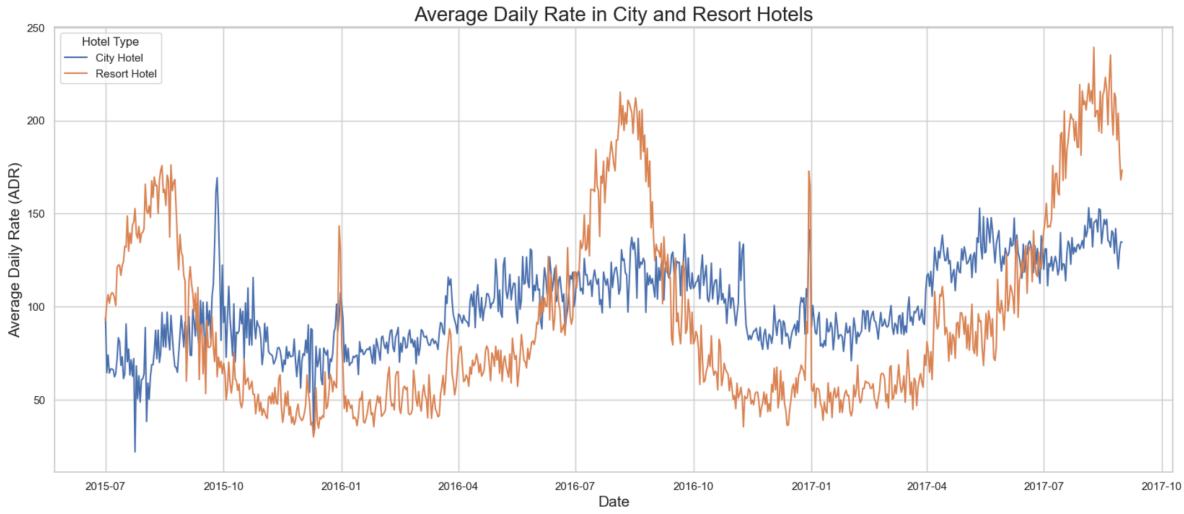
```
is_canceled
0    0.628653
1    0.371347
Name: proportion, dtype: float64
```

Out[128]: <BarContainer object of 2 artists>

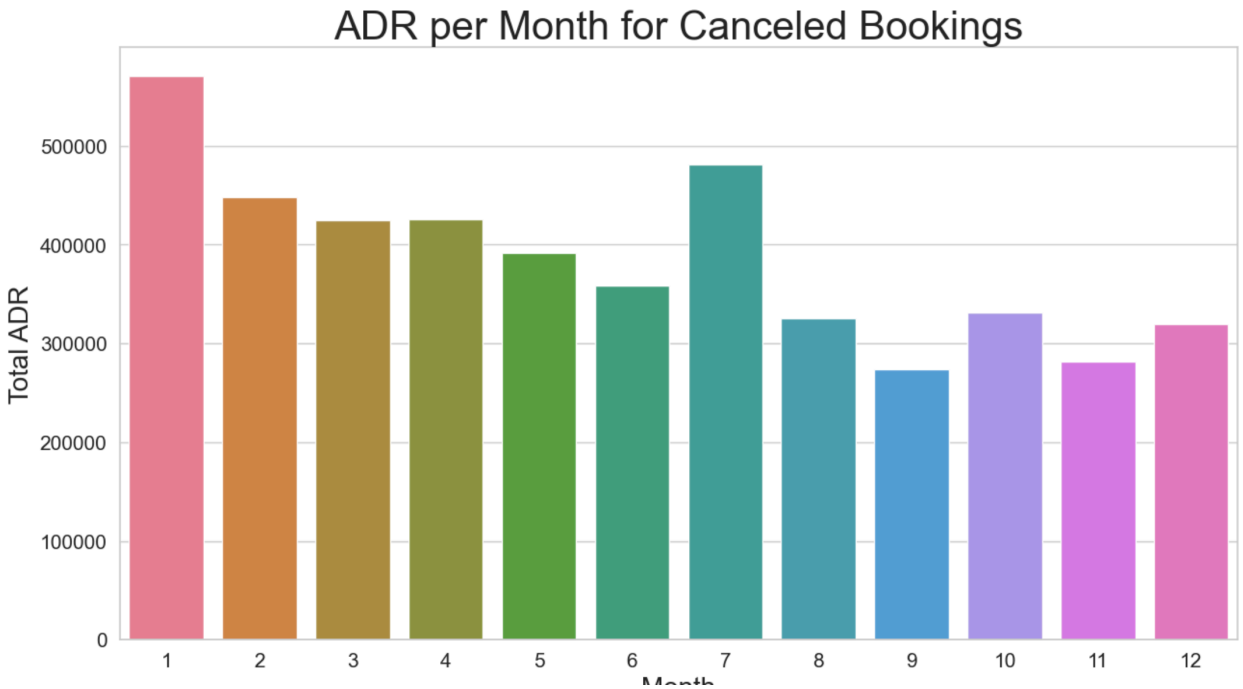


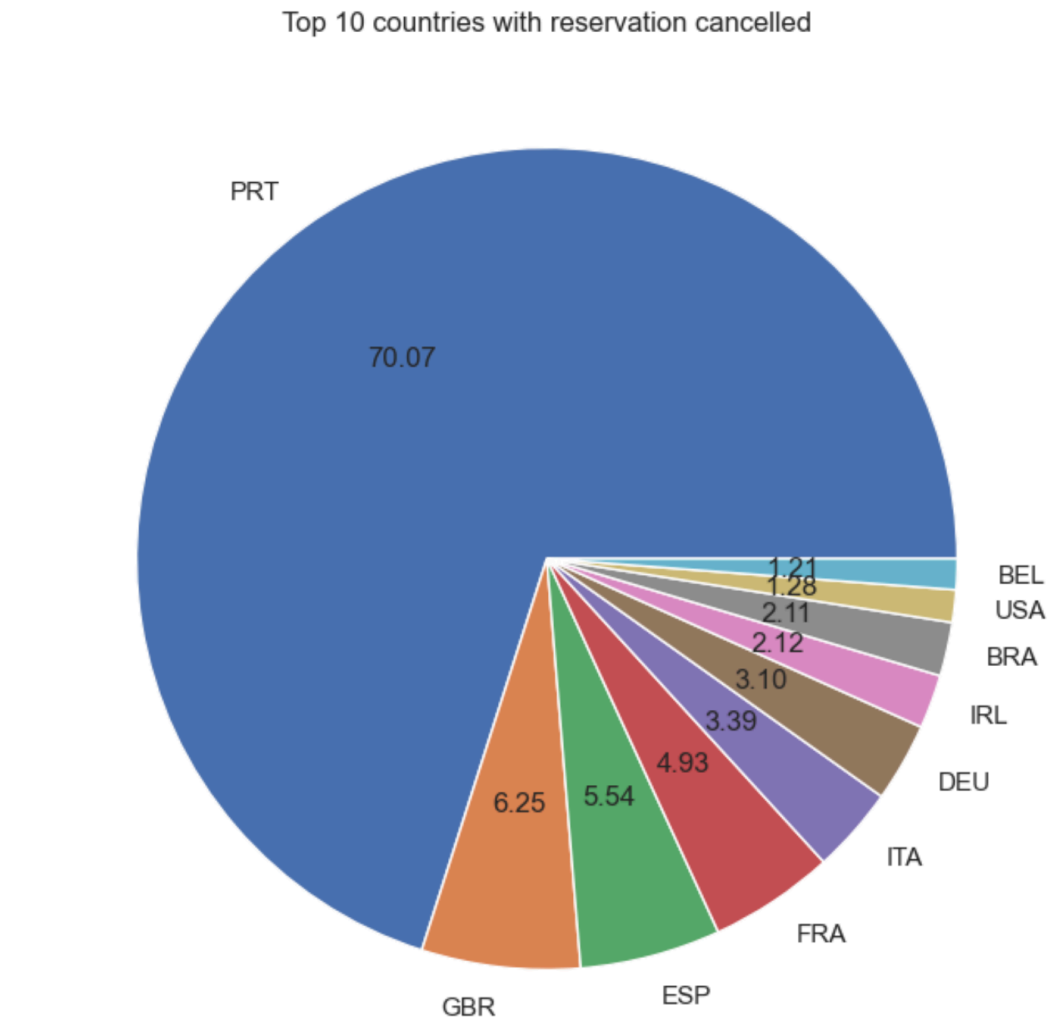


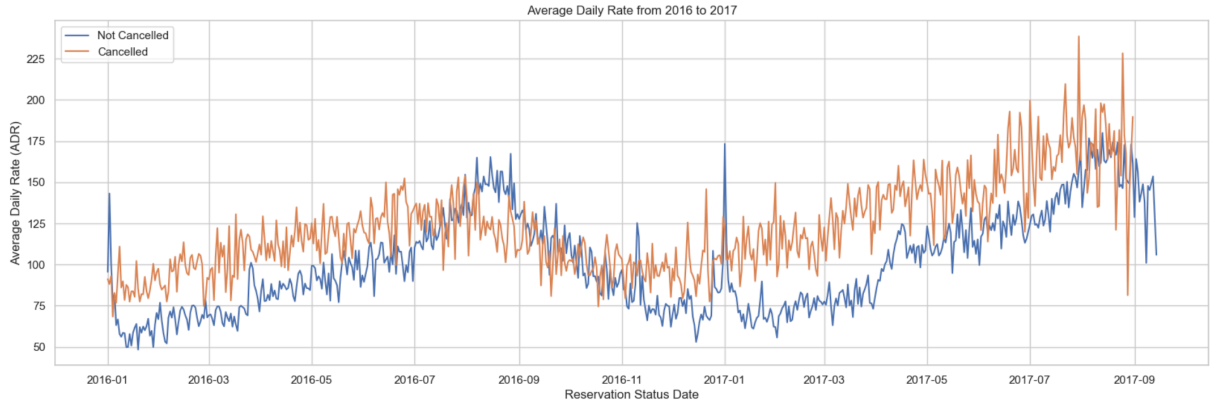

Average Daily Rate Over Time
Average Daily Rate (ADR)
180
160
140
120
100
80
60
40
2015-07
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01
2017-04
2017-07
2017-10
Date


Average Daily Rate in City and Resort Hotels
Hotel Type
City Hotel
Resort Hotel
Average Daily Rate (ADR)
250
200
150
100
50
2015-07
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01
2017-04
2017-07
2017-10
Date


Reservation Status Per Month
Reservation Status
Not Canceled
Canceled
Number of Reservations
8000
7000
6000
5000
4000
3000
2000
1000
0
1
2
3
4
5
6
7
8
9
10
11
12
Month


ADR per Month for Canceled Bookings
Total ADR
500000
400000
300000
200000
100000
0
1
2
3
4
5
6
7
8
9
10
11
12


Top 10 countries with reservation cancelled
PRT
70.07
GBR
6.25
ESP
5.54
FRA
4.93
ITA
3.39
DEU
3.10
IRL
2.12
BRA
2.11
USA
1.28
BEL
1.21


Average Daily Rate from 2016 to 2017
Not Cancelled
Cancelled
Average Daily Rate (ADR)
225
200
175
150
125
100
75
50
2016-01
2016-03
2016-05
2016-07
2016-09
2016-11
2017-01
2017-03
2017-05
2017-07
2017-09
Reservation Status Date