# Exploratory Data Analysis on Hotel Booking Data using Python

The goal of this data analysis was to find correlations between reservation cancellations and other variables and find ways that the hotels can improve their services or prices to counteract this issue.

I hypothesize that cancellations are correlated with prices which result in more cancellations during peak seasons.

Analysis
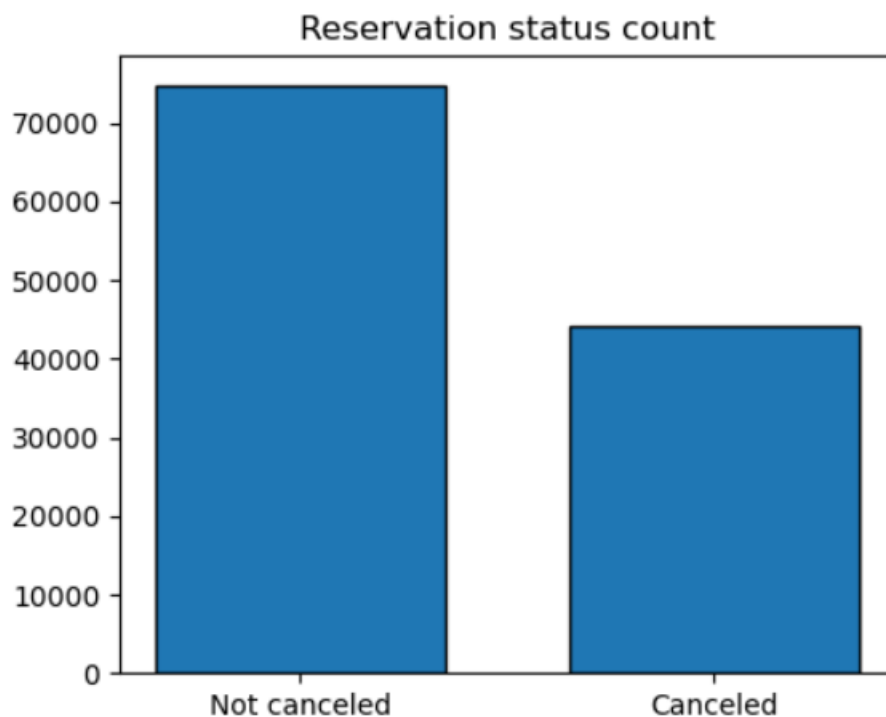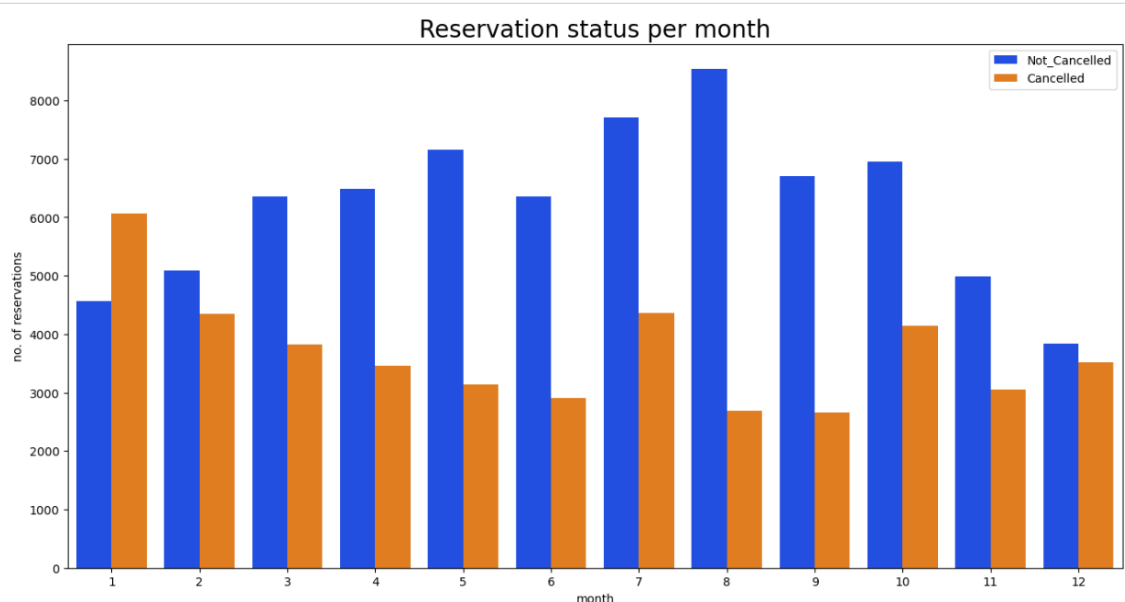

This bar graph displays the total count of canceled vs non-canceled reservations. A 37% rate of cancelation comes across as serious issue and it has a big impact on the total revenue.
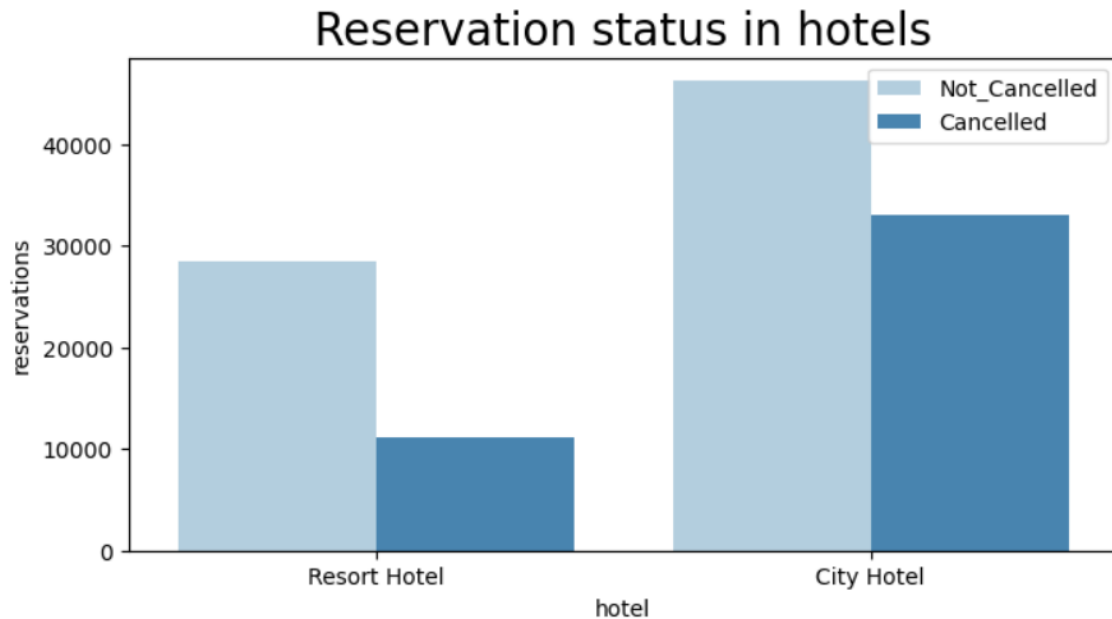
This clustered bar chart shows the total count of canceled vs non-canceled reservations between Resort and City hotels. This analysis displays a significantly larger rate of cancelation in city hotels (42%) when compared with resort hotels (28%).
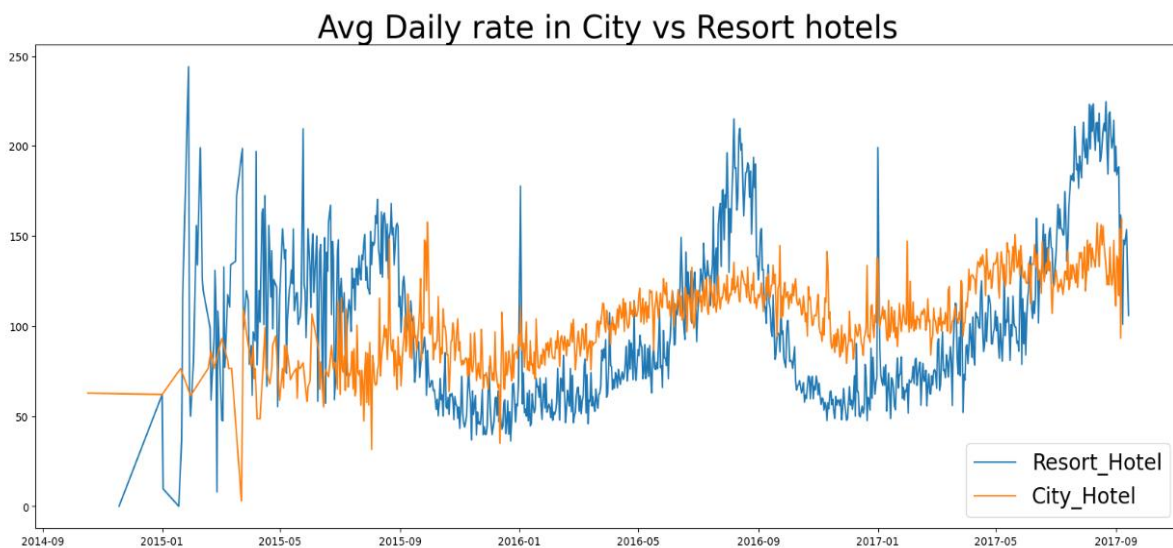
This line graph displays the Average Daily Rate (ADR) from 2015-2018 in Resort and City hotels. This information suggests that the city hotels usually have a higher ADR apart from some annual spikes in the Resort hotel ADR around the summer season.

Reservation status per month

This clustered bar graph displays the total count of canceled vs non canceled reservations across each month. This data suggests that the most cancelations take place during January despite it being one of the months with the lowest total reservations. Meanwhile, August seems to have the highest total number of reservations while maintaining comparatively lower cancelation rates.
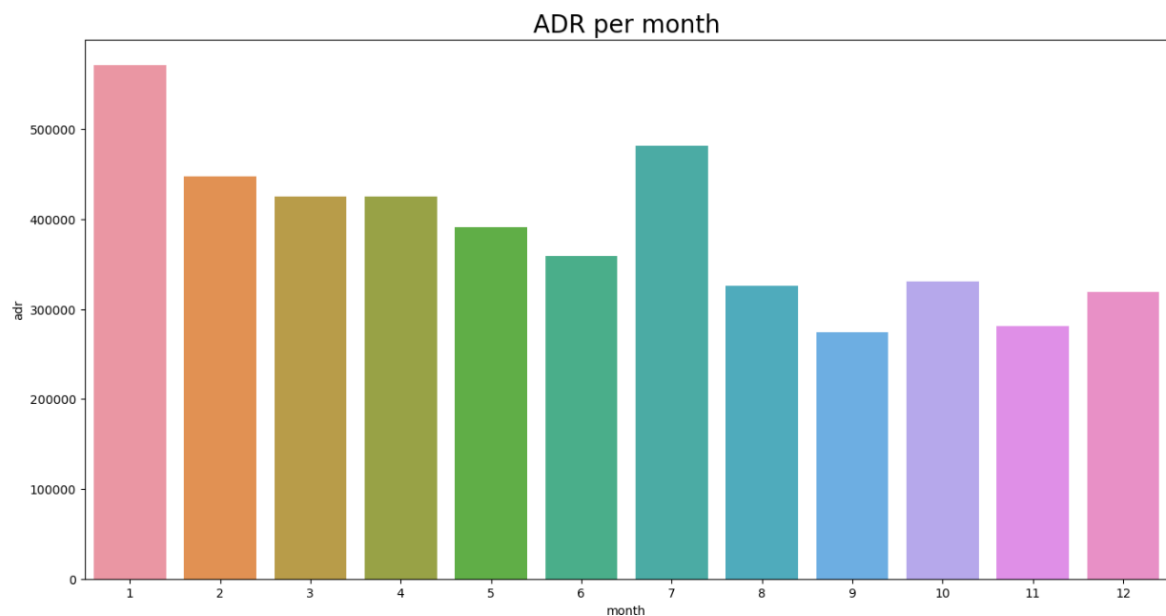
ADR per month

This bar charts shows the Average Daily Rate (ADR) per month. The insight from the data suggests that the highest rate of cancelations occurs when the prices are at their

highest and vice versa. Thus, the cost seems to be the variable with the largest effect on cancelations.

Top 10 countries with cancellations
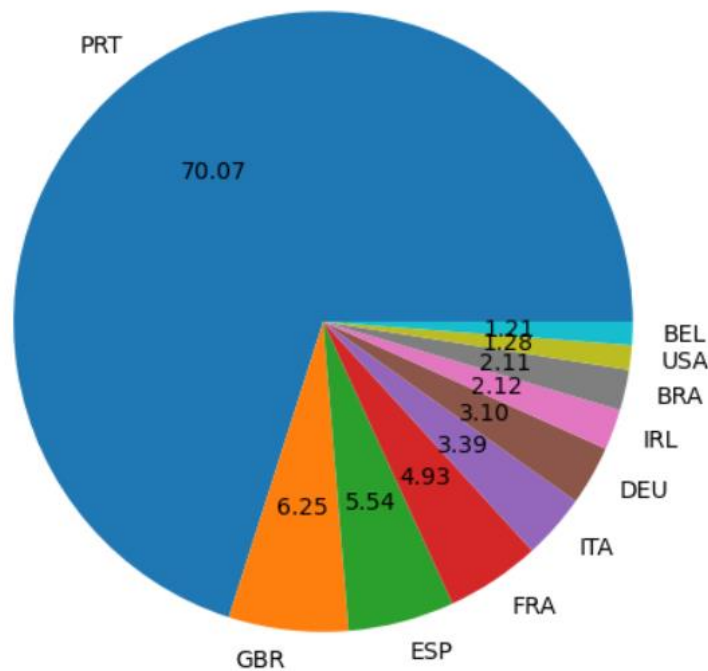
This pie chart displays the % of cancelations in the Top 10 countries with the highest rates of cancelations. Portugal seems to be the biggest offender in the rate of cancelations.

```
In [96]: df['market_segment'].value_counts(normalize=True)

Out[96]: Online TA        0.474377
         Offline TA/TO    0.203193
         Groups           0.166581
         Direct           0.104696
         Corporate        0.042987
         Complementary    0.006173
         Aviation         0.001993
         Name: market_segment, dtype: float64
```

```
In [97]: cancelled_data['market_segment'].value_counts(normalize=True)

Out[97]: Online TA        0.469696
         Groups           0.273985
         Offline TA/TO    0.187466
         Direct           0.043486
         Corporate        0.022151
         Complementary    0.002038
         Aviation         0.001178
         Name: market_segment, dtype: float64
```

The first data set shows the % of what source the guests used to book the hotels. Online booking and travel agents seem to be the biggest methods of booking.

Meanwhile, the second data set shows that guests using online bookings and travel groups have the highest rates of cancelations which makes sense since they make up a large percent of the total reservations.
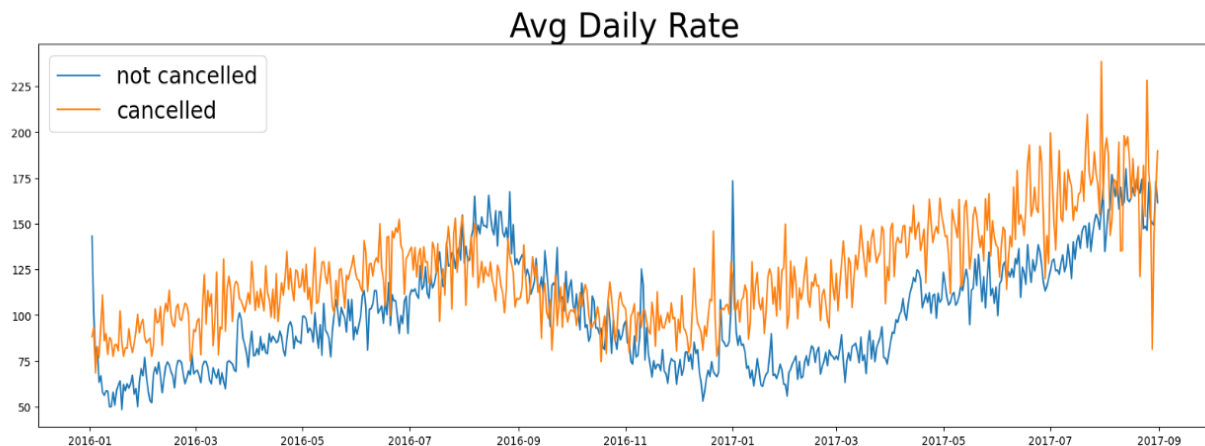
This line graph shows the ADR from 2016-2018 when looking at canceled vs non-canceled reservations. The trends show that the rate of cancelations increase with increasing ADR; thus, proving that price of rooms seems to be affecting the rate of cancelations.

Solutions/Insights

- To limit the increasing rate of cancelations, the hotels could work on pricing in locations most heavily impacted by cancelation or offer discounts and limited time offers to guests.
- Since resort hotels experience higher cancelation rates, they could implement changes in pricing during weekends and peak seasons.
- In months with higher cancelations like January, the hotels could offer discounted price to provide better incentive to guests to follow throw on their bookings.
- In places with high cancelation rates like Portugal, customer surveys would be beneficial in understanding the primary issue. Upgrading hotel facilities and offering better pricing could be a step in the right direction.