Analysis of hotel cancelations.

By Daria Gavrilova

Introduction
Data
Analysis
Conclusion

Introduction
Cancellations are one of the biggest problems faced by a hotels. That is why it is so important to understand clients behaviour in order to implement strategies aimed at combating and preventing losses caused by unexpected cancelations.

It is common mind that online travel agencies such as Booking, Expedia, Agoda and etc. are one of the most important part of an a segment market for the hotels. They make process of reservations and cancelations easy and and accessible. So I will try to prove or reject a fact that Online TA are the reason of high cancelation tendency.

Data
Database consists of two datasets with hotel demand data. One of the hotels (H1) is a resort hotel and the other is a city hotel (H2). Both datasets share the same structure, with 31 variables describing the 40,060 observations of H1 and 79,330 observations of H2. Each observation represents a hotel booking. Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were canceled.
https://www.sciencedirect.com/science/article/pii/S2352340918315191

Analysis
In the graph below is shown reservation status of two types of hotels. We can see that for both hotels cancelations present considerable amount of a total reservations. However Resort Hotel receives less cancelations than a City Hotel.
City Hotel - 41%
Resort Hotel - 27%

Cancelation tendency in years.

Reservation status by hotel type — Reservation status by year
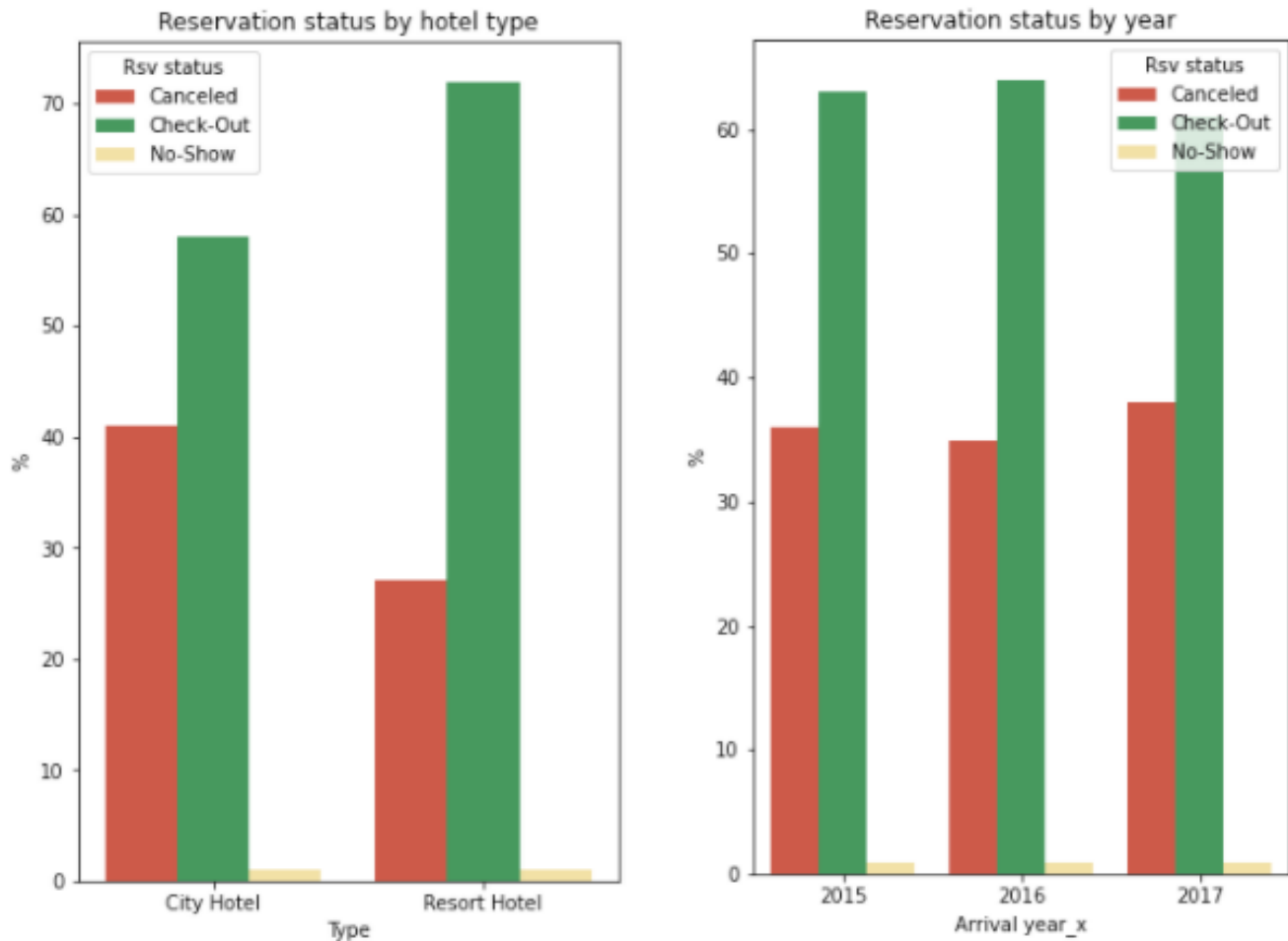
It is common mind that cancelation tendency increase in years. But as it is shown in a graph above cancelation tendency keeps its level during years and do not change statistically significant.
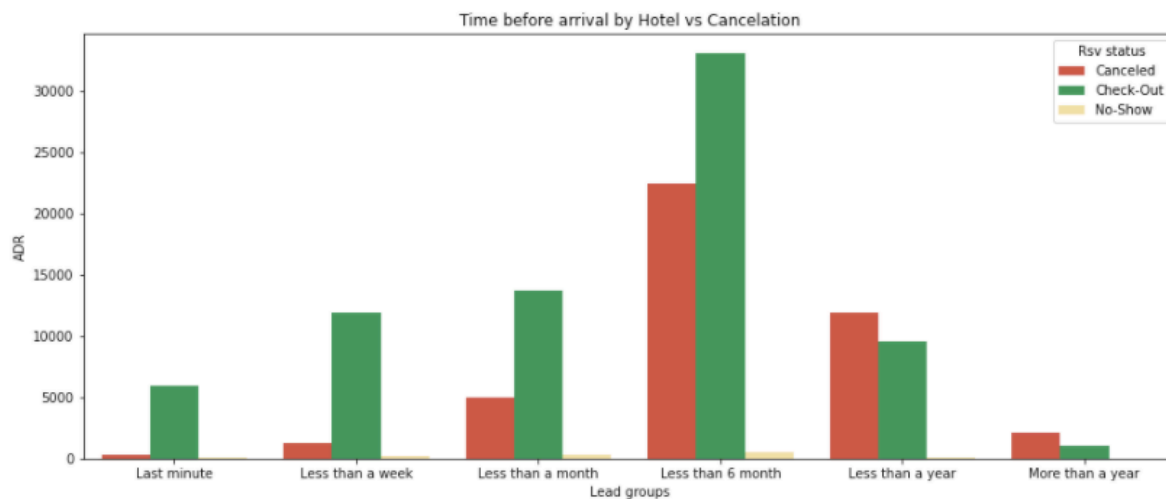
2015 - 36%
2016 - 35%
2917 - 38%

When most cancelations are made?

I grouped all the cancelations by 6 categories.
- Last minute (made on the arrival day)
- Less than a week
- Less than a month
- Less than six month
- Less than a year
- More than a year

Time before arrival by Hotel vs Cancelation

A graph above shows that the biggest amount of cancelations are done in a "Less than 6 month" time category. But the biggest percents of cancelation in compare with received amount is in "Less than a year" and "More than a year" categories. Last minutes reservations are rarely canceled.
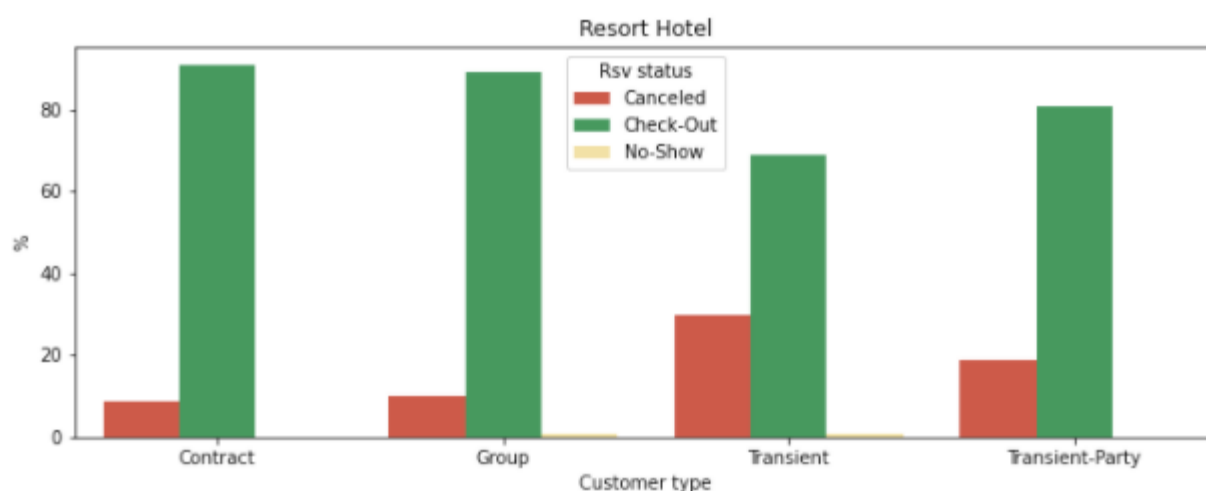
Customer type
For City Hotel most cancelation are done by Transient and Contact customer types.
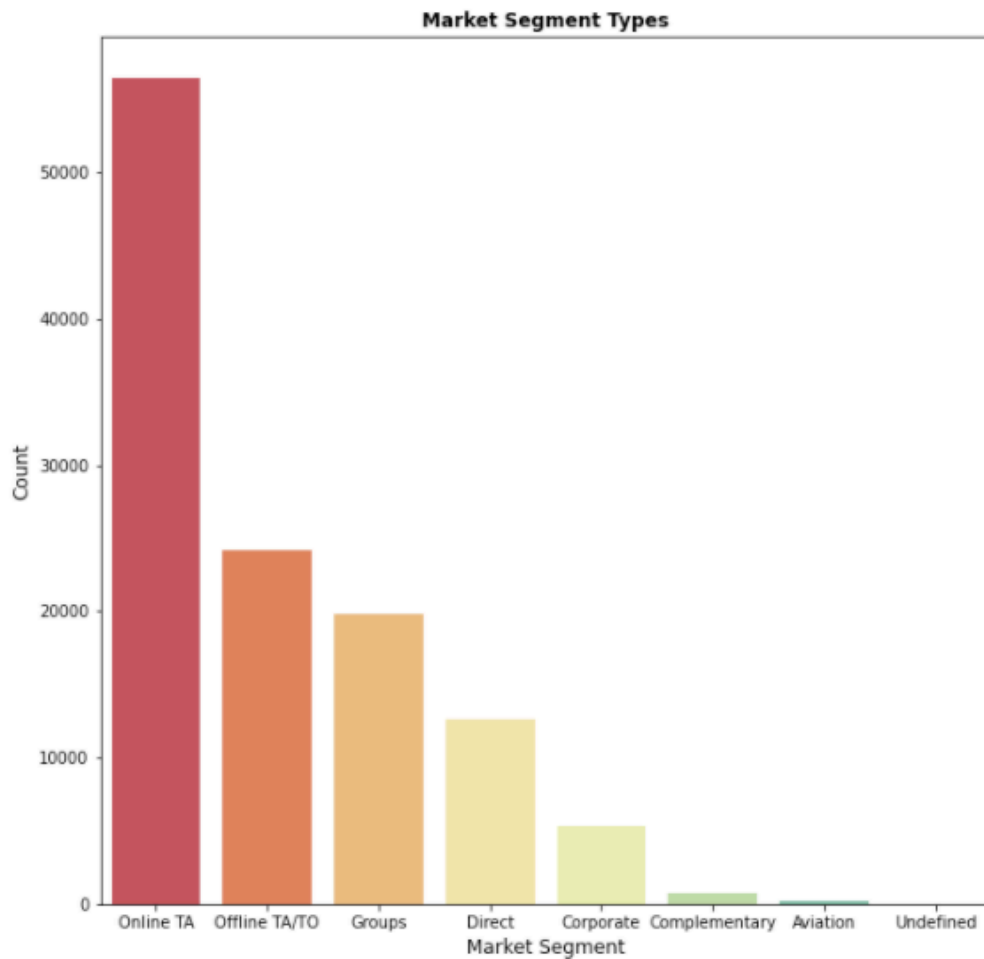Contract customer - 47%
Transient - 44%
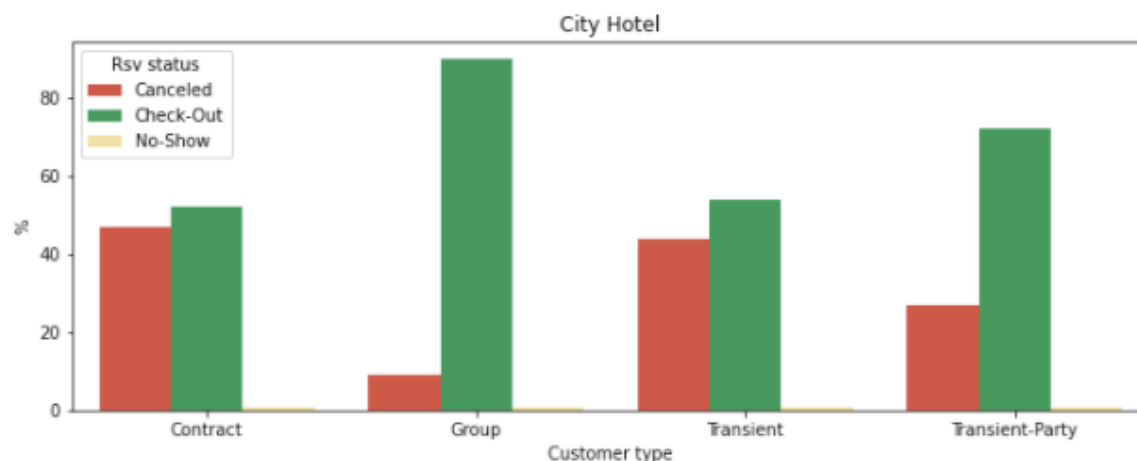
City Hotel

In Resort hotel: Transient - 30%

Resort Hotel

Market segment
Online TA is most important part of Market segment.

Market Segment Types

How it is shown in graphs bellow reservation of Online TA are not most cancellable. Talking about City Hotel.

There're only 2 "Undefined" reservations and all 2 are canceled. So we will not take them in account.

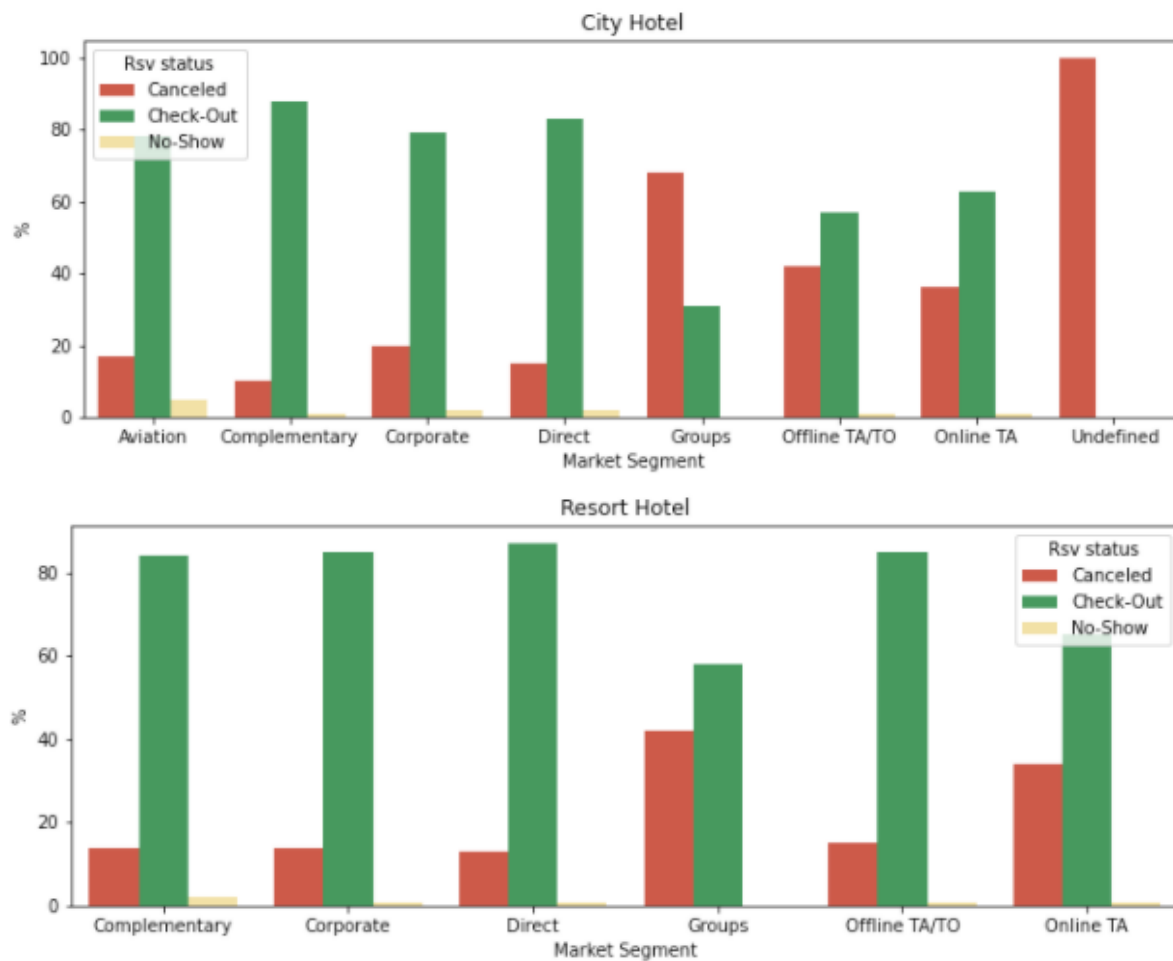Online - 36% of cancelations

Offline agencies - 42%

Groups - 68%

Resort Hotel receives less cancelations and it situation is:
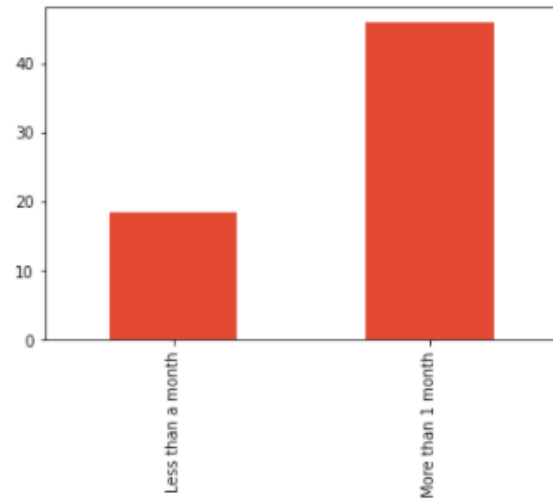
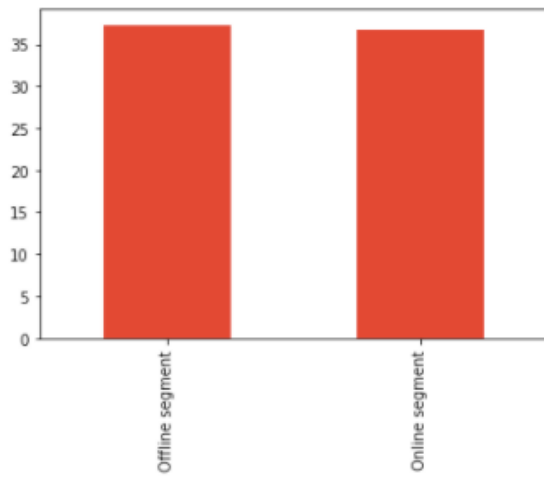Online - 34%

Offline agencies -15%

Groups - 42%

Statistical analysis

In order to understand if if Online TA brings more cancelations, I created to categories "Online" and "Offline". My "Null hypothesis" is -"Cancelations from both segments are equals". With a help of Proportions Z Test I received a p-value equal to 0.029 Difference is statistically significant, but real difference is very low.So "Null hypothesis" could be rejected. As it is shown in a graph below. "Offline" category brings more cancelation as an "Online".

My second "Null hypothesis" is "Percents of cancelations of a lead group "More than 1 month" is equal to cancelations of a lead group "Less than 1 month". P-value is equal to 0.0. Difference is statistically significant. "Null hypothesis" could be rejected.

Conclusion

Cancelations in hotels not depend that much on online segment. Reservation lead time, customer type and offline segment have more influence on cancelations.