Hotel Booking Cancellation Prediction

Predict Hotel Cancellations

Prepared By:

Bharat

Objective

- To predict whether a booking will be cancelled or not using machine learning based on historical data.

- The goals of this project are:

✓ To find out the characteristic of customers who cancelled and finding a pattern in cancelled booking by doing an exploratory data analysis.

✓ Building classification machine learning model to predict cancellation, that has accuracy score around 0.8 - 0.9

Background Knowledge and Problem Statement

Background Knowledge

- Hotel industry is one of the faster growing businesses of tourism sector, especially with the rise of giant OTA that make booking a hotel as easy as is ever been.

Problem Statement

- The increased trend of cancellation from year to year has affected hotel not being able to accurately forecast occupancy within their revenue management, and the trend of cancellation also have causes hotel loss in opportunity cost (unsold room due to cancellation).
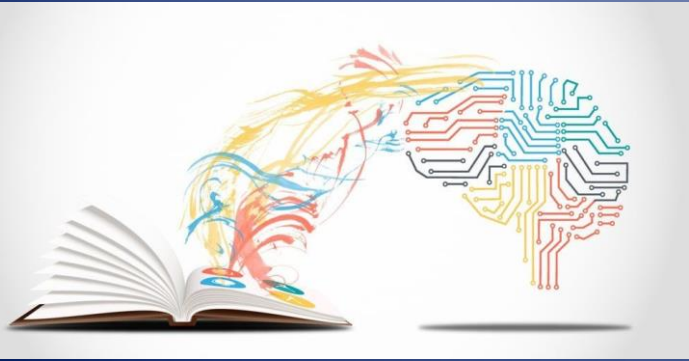
How Cancellation Affecting Hotels

➢ Loss of income in shape of unsold room (due to cancellation)
➢ Lower RevPAR (Revenue Per Available Room) when selling cheaper at the last minute

Questions to be answered

Limitations of Project

List of Question to help achieving the goal:

➢ How Market Segment Of Booking Affecting Cancellation

➢ How's a lead time of a booking affecting cancellation

➢ How's different deposit type affecting cancellation of a booking

➢ How does cancellation rate of booking from portugal and booking that's made outside portugal

➢ What Are The Other Factors that affecting cancellation of booking

➢ What machine learning algorithm that has the highest accuracy when it comes predicting hotel booking cancellations

➢ This hotel booking cancellation project only applied for hotel bookings in Lisbon Region and Algarve Region both location are located in portugal.

➢ Predicting cancellation with this web application outside both region might have not so accurate result due to the location constraint, different pattern of cancellation
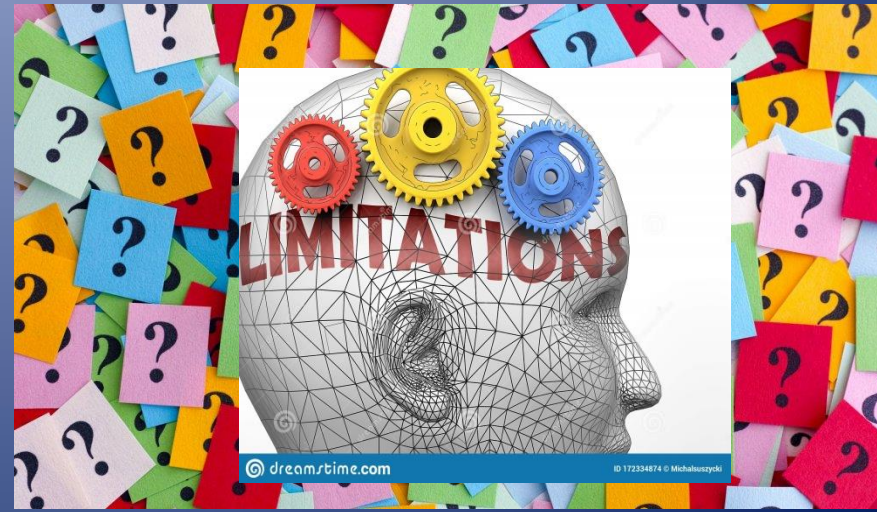
Data Analysis

Contents

- ➢ Cancellation Insights and Analysis:
  - ➢ Analysis on characteristic of cancelled booking & non cancelled booking
- ➢ EDA Recommendation:
  - ➢ Recommendation /Input on how to reduce cancellation based on the exploratory data analysis insight
- ➢ Machine Learning:
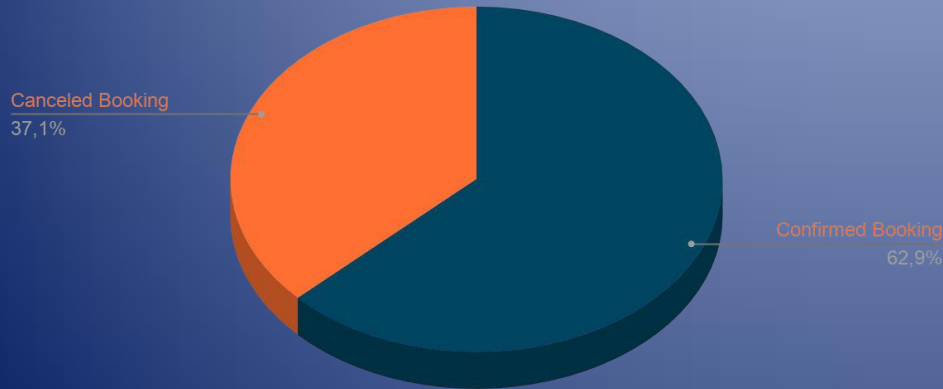 - ➢ Machine Learning process on predicting hotel cancellation from model building

Cancellation Insights and Analysis

General Cancellation Report

Booking Cancellation Report

Canceled Booking
37,1%

Confirmed Booking
62,9%

Based on the dataset we see that percentage of cancellation in portugal from **July 2015 - August 2017 is 37.1%,** which is slightly below industry standard around that year at **41.3 %**

Cancellation Rate Per Market Segment

Percentage of Booking Per Market Segment
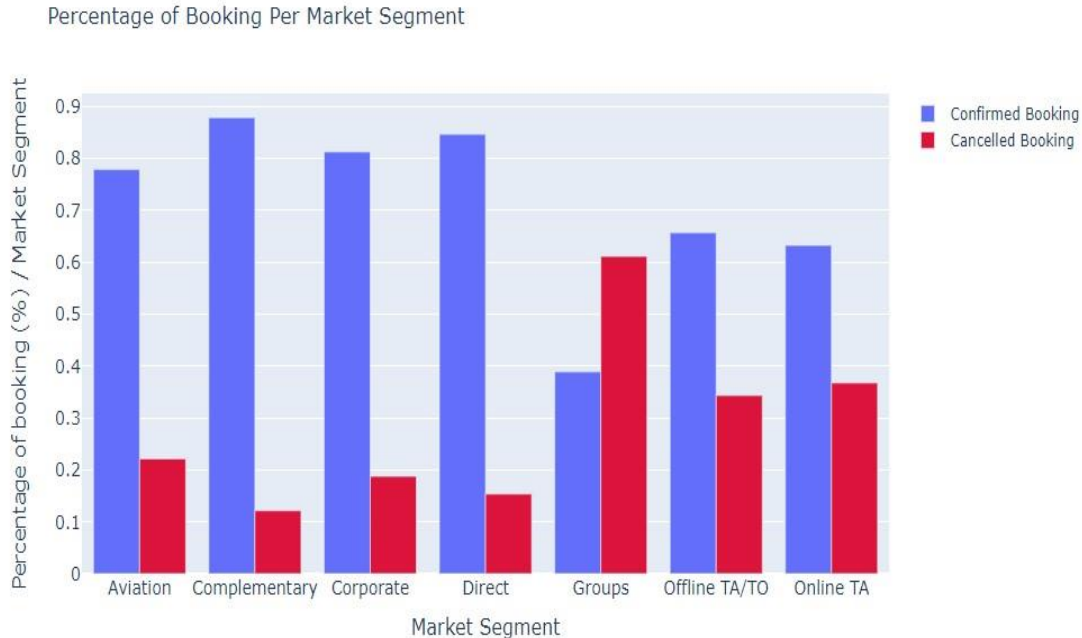
1. From our analysis we see that **Corporate , Direct , and Aviation** has a cancellation rate around **18 - 22 %** of their booking

2. **Travel Agent (Online /Offline)** has a cancellation rate around **34 - 36 %**

3. Lastly **Group** has the highest cancellation rate around **61 %**

Lead Time Month & Cancellation Rate

Percentage of Confirmed Booking and Cancelled Booking per Lead Time Month
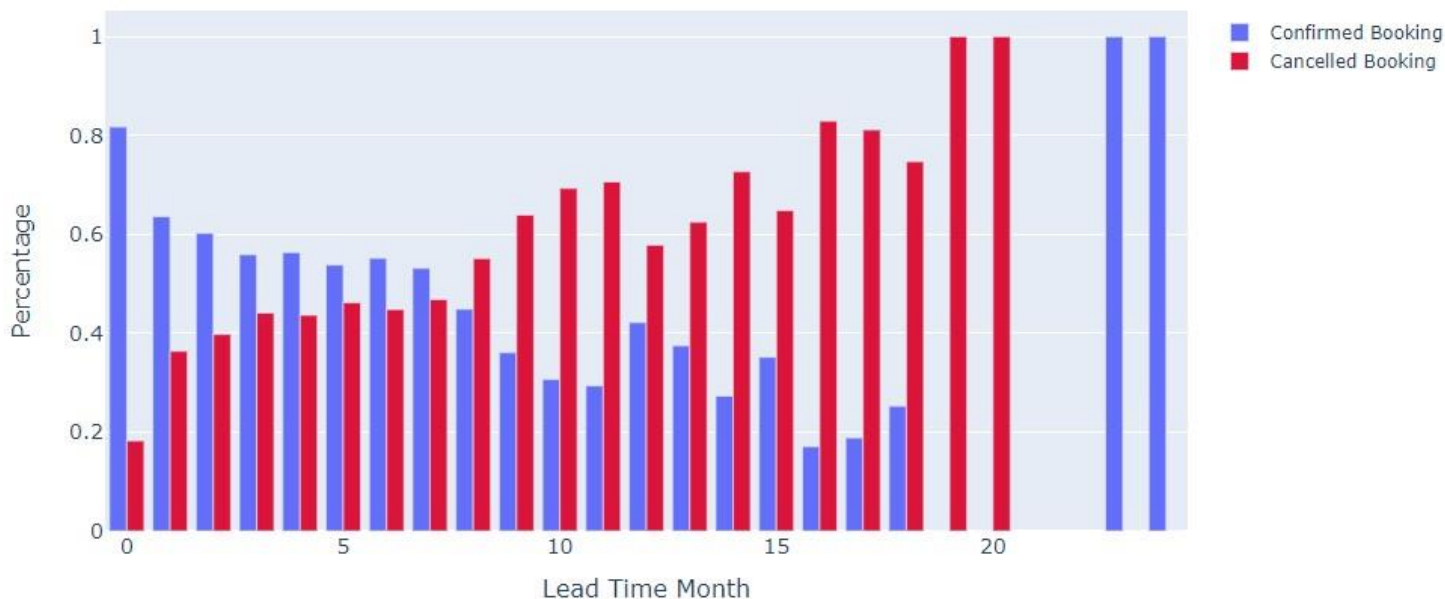
- Booking that has less than or equal **to 7 months** lead time have a higher confirmed booking rate **(>50%)** to it canceled rate

- Booking that has more than **7 months** have a higher cancellation rate compared to it's booking rate

- Cancellation is positively correlated with lead time

Deposit Type & Cancellation Rate

Percentage of Cancellation for each Deposit Type

- **Non Refund Deposit type** has the highest cancellation rate among all deposit type **(99.4 %)**

- **No Deposit** has the second highest cancellation rate at **(28.3%)**

- **Refundable Deposit** has the lowest cancellation rate among all of the deposit, it cancellation rate is at **(22%)**
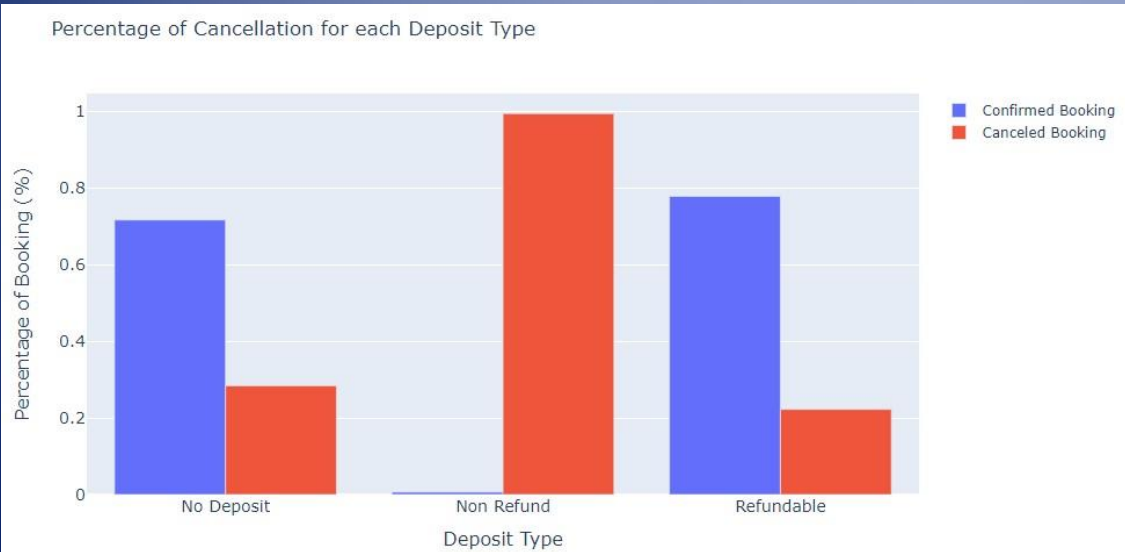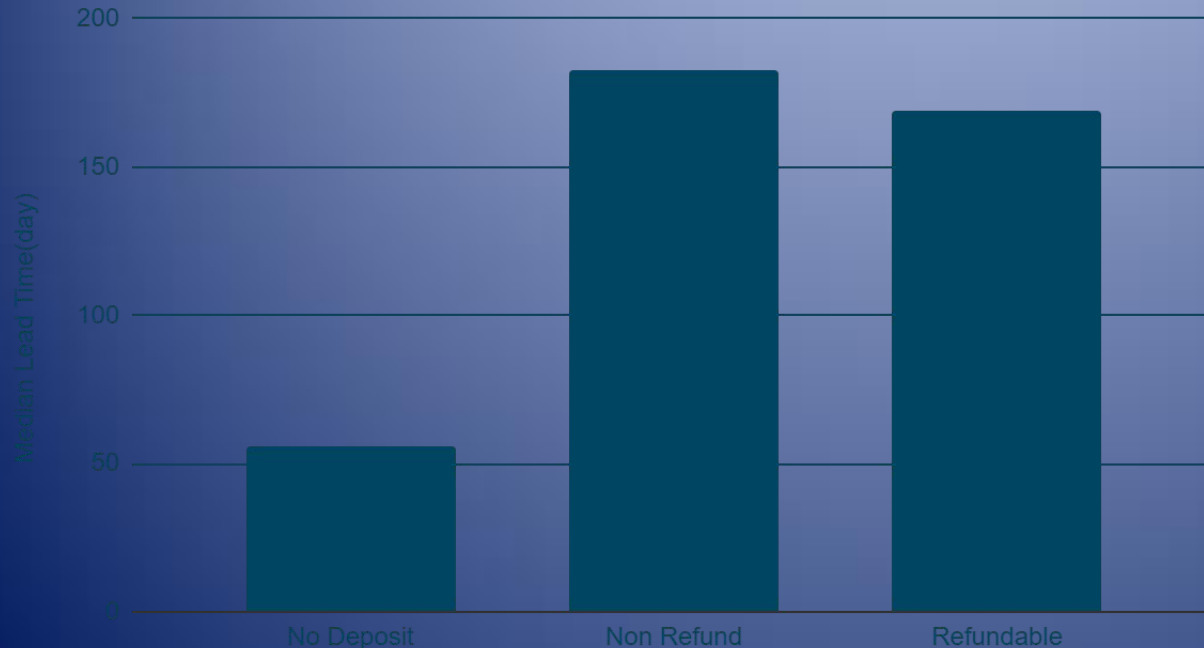
For the hotels this is nothing alarming since they don't lose revenue when no refund booking is canceled.
But it's always a good practice to question something that's extraordinary happening.
In this case, we take a closer look at the median lead time for each deposit type.

Median Lead Time for Each Deposit Type

Median Lead Time(day) for Each Deposit Type

➢ Based on the previous analysis we saw that non refund booking has the highest cancellation rate compared to other booking type.

➢ Based on lead time and cancellation analysis we saw that the longer lead time has higher cancellation rate compared to the shorter one

- Median Lead Time Non Refund 183 Days

- Median Lead Time Refundable 169 Days

- Median Lead Time for No deposit 56 Days

Previously Cancelled Booking & Cancellation Rate

Percentage of Previously Cancelled Booking

- Booking That's previously Canceled has 92% cancellation rate

- Booking that's never been canceled before has cancellation rate of 34%

This shows that booking that has been canceled before will more likely to be cancelled again

Booking Location and Cancellation rate

Percentage of Confirmed Booking and Cancelled Booking for each booking location

- International Booking has 24% Cancellation Rate

- Local Booking has 56% Cancellation Rate

This arise a question.

Why does local booking are more likely to be canceled compared to international booking ?

Booking Location & Previously Cancellation

Percentage of Previously Cancellation

Based on the previous analysis we saw that booking that has been canceled before has a cancellation rate of 92 %

- For international booking **99.5%** of international booking never been cancelled before.

- While for local booking **13%** of bookings has been canceled before.

This higher number of previous cancellation definitely is one of the factors why local booking has a higher cancellation rate compared to international booking

Repeated Booking & Cancellation Rate

Percentage of Confirmed Booking and Cancelled Booking for Repeat & Non Repeat

- Repeated Booking cancellation rate is around **14 %**

- Non Repeated Booking are more than 2 x more likely to be cancelled their booking compared to repeated booking **(37%)**

Conclusion :
Repeated guests are more likely to confirmed the booking compared to non repeated guests

Parking Space & Cancellation

Parking Space And Cancelled Booking

From July 2015 to August 2017 There are **7407 (6.2 %)** out of total booking that required a parking space .

- Out of **7407 bookings** that required a parking space. **There's not a single booking that's cancelled (0 Cancellation for booking that required a parking space)**

This conclude booking that required a parking space will most likely be confirmed

EDA Recommendation

EDA Recommendation

Only Non Refundable Deposit For Group Booking

- From the analysis we see that group booking has the highest cancellation rate.

- With this policy hotel wouldn't suffer any loss of revenue or loss of potential revenue due to cancelled group booking

Maximum Lead Time for Booking

- We see a pattern of booking that has more than **210 days** lead time are more likely to be canceled

- setting up a maximum lead time means users wouldn't be able to make booking that's too far in advance

Combination of Restriction

- Setting up maximum lead time for booking might have resulted in hotel visibility to the potential guest search.

- Combining deposit type lead time might help to get the hotel more exposure without higher risk of losing revenue. **(ex. Non Refundable Deposit only for booking >210 days in advance)**

EDA Recommendation

Increase Direct Booking Market Segment

- From this dataset we see that direct booking has the least cancellation rate at **15%** (*ouīside complimenīary)*

- with only being **10 %of the total booking**, the increase of direct booking might lead to less cancellation

Stricter Cancellation Policy For Previously Canceled Booking

- Booking that's previously canceled has cancellation rate of **92%**.

- looking at this pattern we see that booking that's previously canceled are most likely to be canceled again.

- to protect hotel from losing revenue due to this kind of cancellation, **hotels need to set booking payment in advance /non refundable deposit,** for booking that has been canceled before

Attracting Customers Who Drives

- **6%** of total booking required a parking space **(7407 bookings),** and out of **7407** bookings none of them were canceled**.**

- That's around **10%** of total confirmed booking. Hotel could promote to attract customer that drives
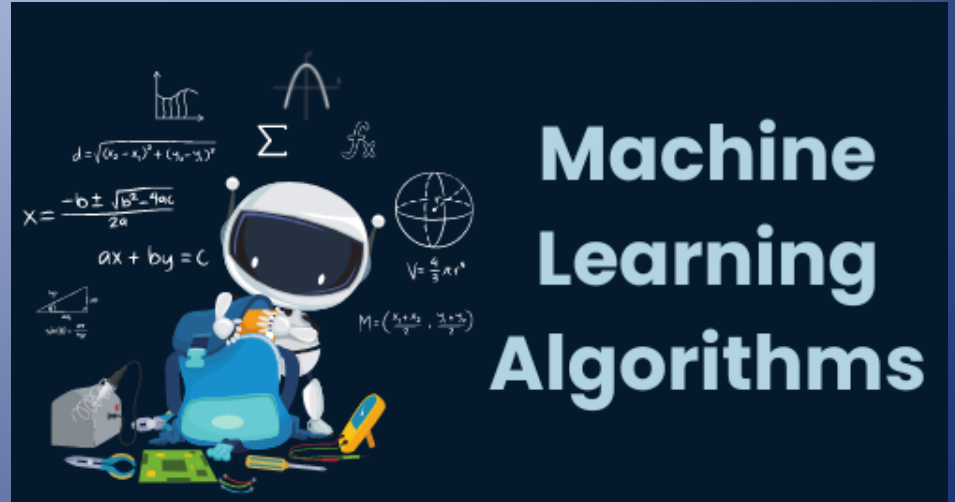
Machine Learning
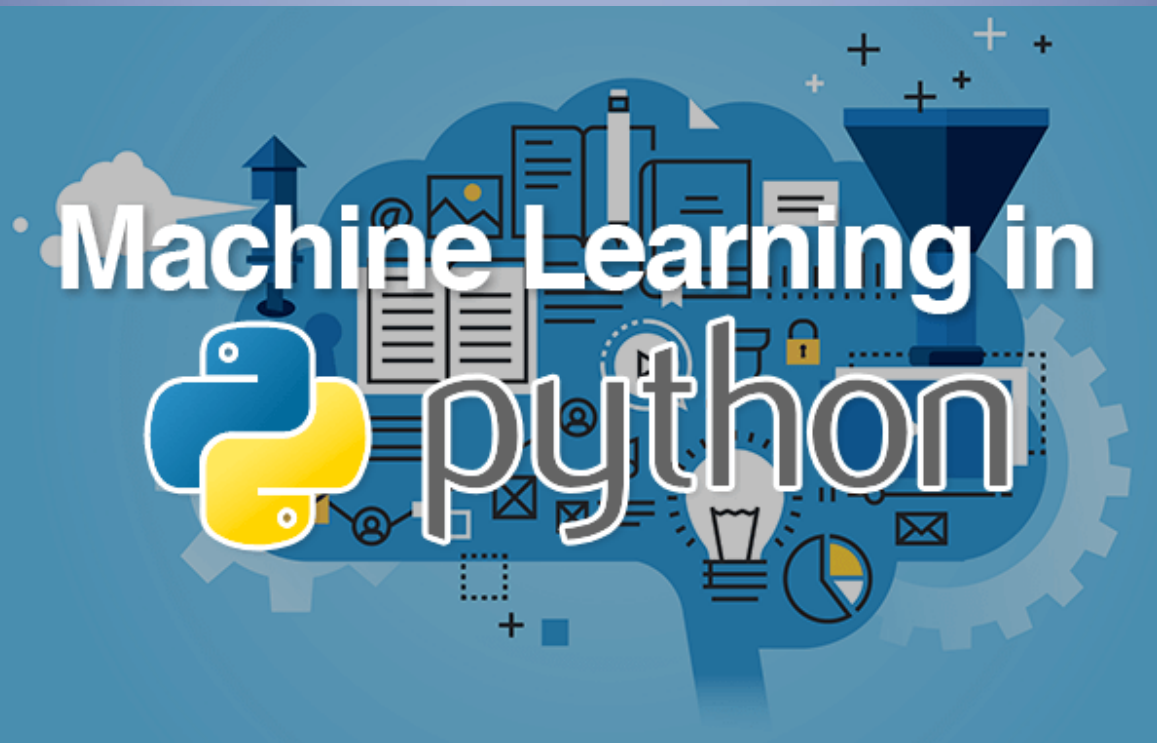
Machine Learning Algorithm

Algorithm Used:

1. Logistic Regression

2. KNeighbors Classifier

3. Decision Tree Classifier

4. Random Forest Classifier

Model Evaluation Matrix

	Logistic Regression	KNN	Decision Tree	Random Forest
Accucary	0.804135	0.835441	0.831828	0.840358
Recall	0.615210	0.740904	0.746345	0.762666
Precision	0.810875	0.800416	0.788717	0.797842
F1 Score	0.699620	0.769511	0.766946	0.779857

	training	testing
Logreg	0.802412	0.804135
KNN	0.879586	0.835441
Decision Tree	0.941978	0.831828
Random Forest	0.941957	0.840358

Accuracy as The Primary Evaluation Matrix

- Because the data is somehow balance between the target class **(63%) Confirmed, 37% Canceled, Every Class is equally important**
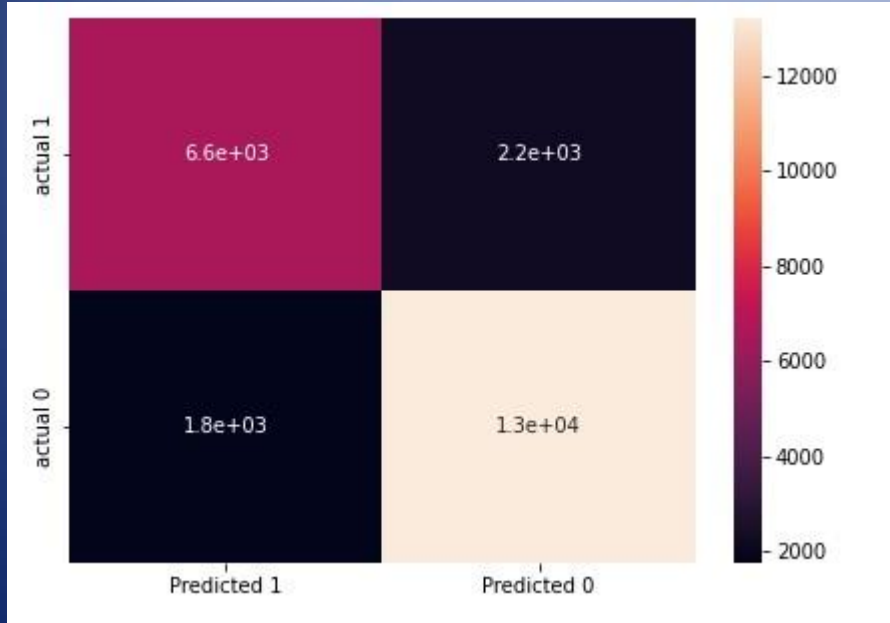
Confusion Matrix

Logistic Regression

KNearest Classifier

Confusion Matrix

Decision Tree Classifier

Random Forest Classifier

How this model will help hotels?

- This model will allow hotel managers / revenue manager to take actions on bookings that's identified as "potentially going to be canceled"

- These prediction models enable hotel managers to mitigate revenue loss derived from booking cancellations and to mitigate the risks associated with overbookings (reallocation costs, cash or service compensations, and particularly important today, social reputation costs).

- Booking cancellations model also allows hotel managers to **implement less rigid cancellation policies, without increasing uncertainty**. This has the potential to translate into more sales, since less rigid cancellation policies generate more bookings.