Different Factors that affect Hotel Reservation

Members -
Mekhal Raj
Rohan Gupta
Shi Wang
Zichen Wang
Yilun Wang

About this project

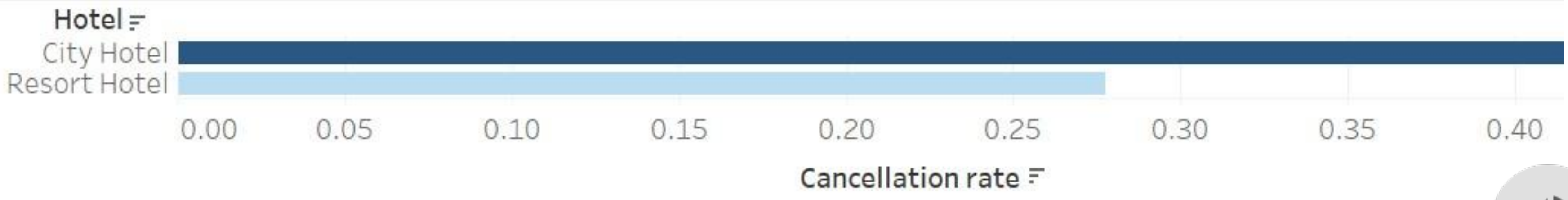
	Total_Sample_Number
0	119390

This data set contains booking information for city hotels and resort hotels, and includes information such as when the booking was made, length of stay, the number guests, the number of bookings, and whether the booking was cancelled or not among other things.

The cancellation rate of the two types of hotels

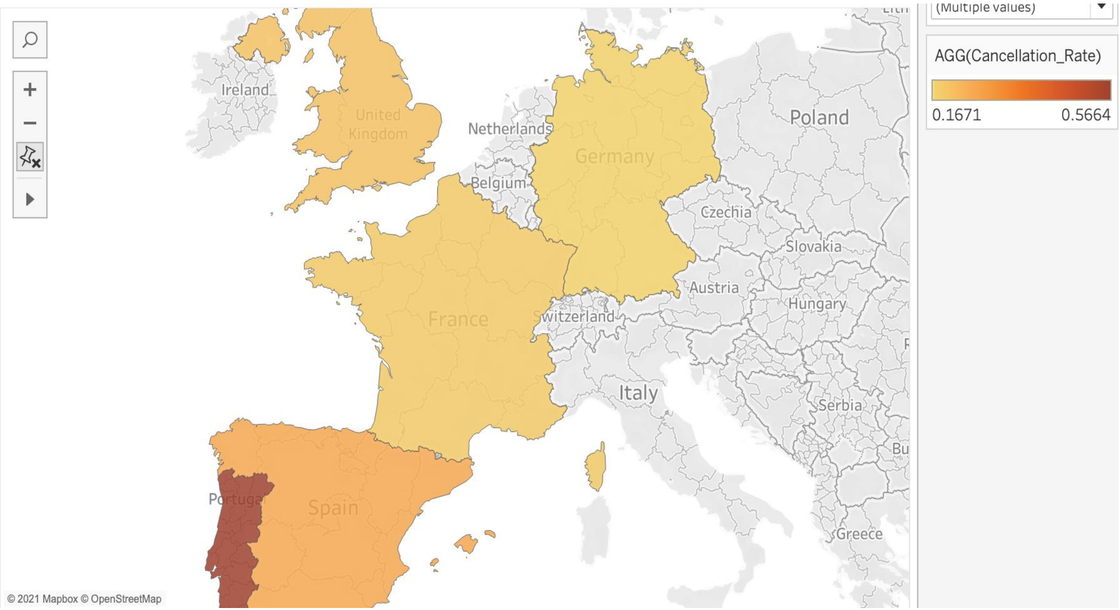
	Hotel_Type	Total_Number_of_Booking	Total_Number_of_Cancellation	Cancellation_Rate
0	Resort Hotel	40060	11122	0.277634
1	City Hotel	79330	33102	0.417270

Cancellation Rate Graph



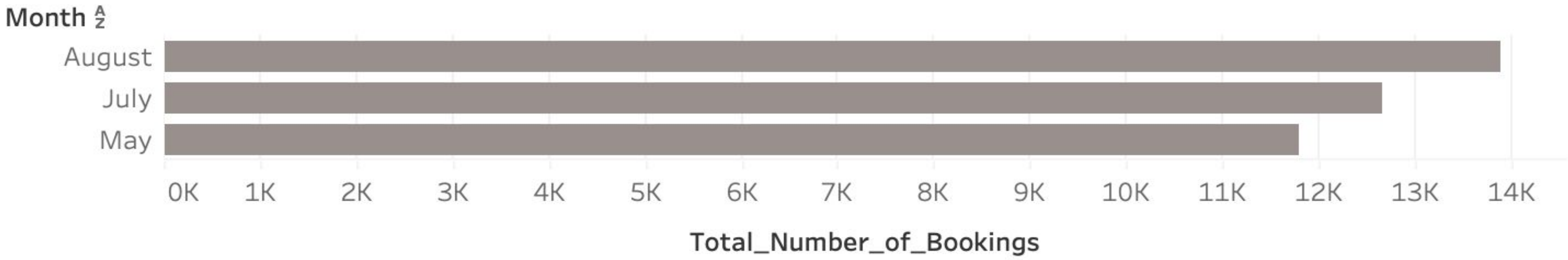
Countries with highest cancellation rates

	country	cancellation_rate
0	PRT	0.566351
1	GBR	0.202243
2	FRA	0.185694
3	ESP	0.254085
4	DEU	0.167147



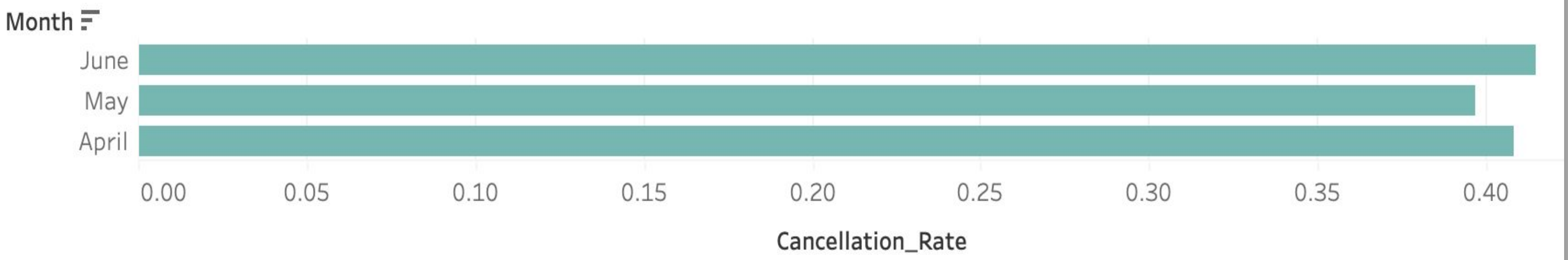
Months with the highest booking

	Month	Number_of_Booking	Proportion_of_Booking	Number_of_Cancellation	Cancellation_Rate
0	August	13877	0.116	5239	0.378
1	July	12661	0.106	4742	0.375
2	May	11791	0.099	4677	0.397



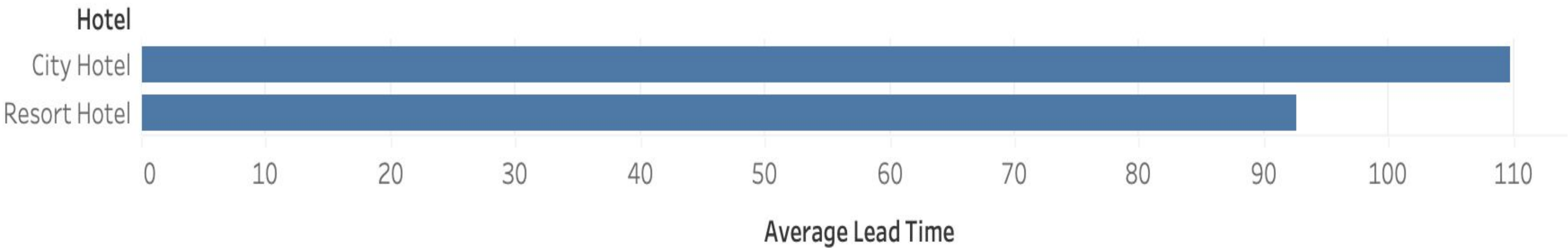
Months with the highest cancellation rate

	Month	Number_of_Booking	Proportion_of_Booking	Number_of_Cancellation	Cancellation_Rate
0	June	10939	0.092	4535	0.415
1	April	11089	0.093	4524	0.408
2	May	11791	0.099	4677	0.397

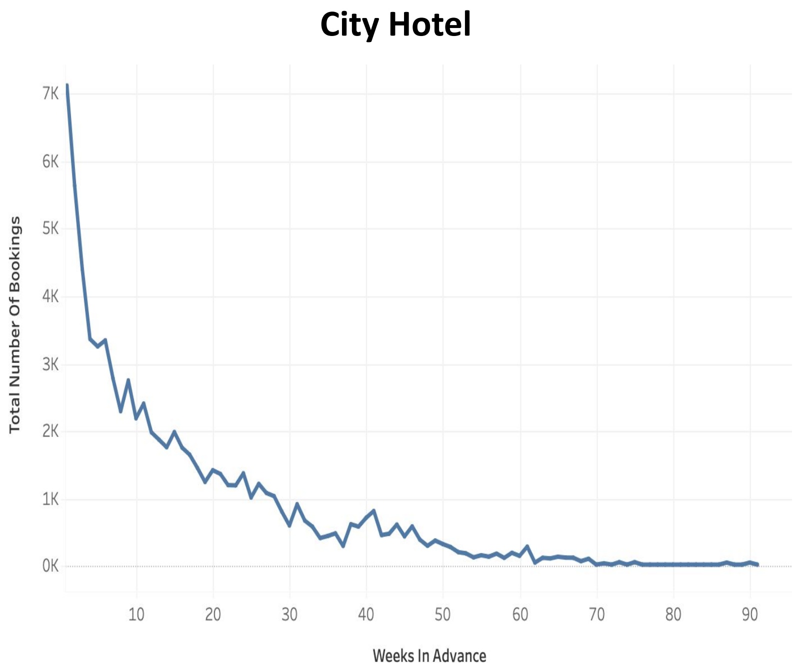


Average lead time for different types of hotel

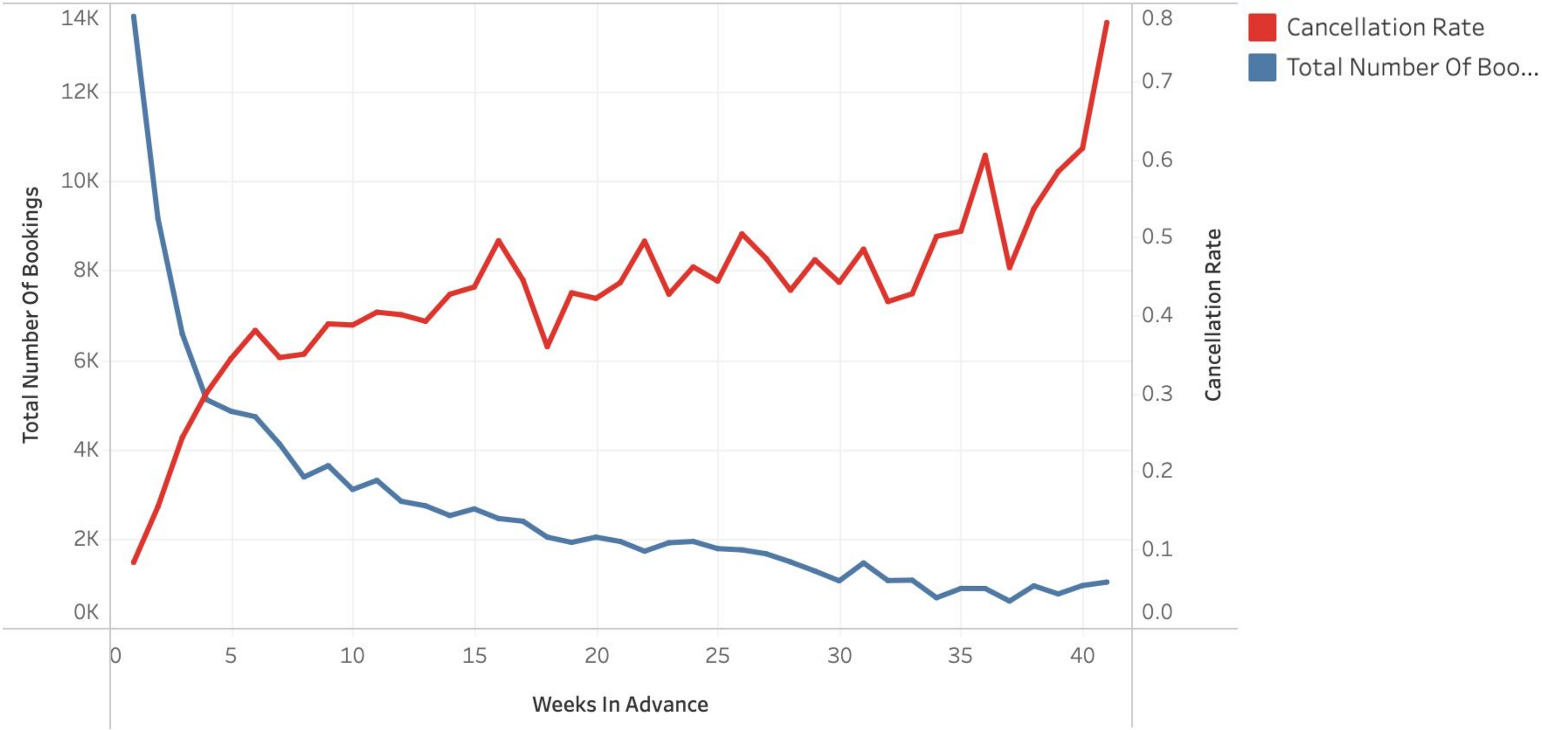
	hotel	avg_lead_time
0	Resort Hotel	92.675686
1	City Hotel	109.735724



Possible misunderstanding of average lead time



Cancellation Rate related to lead time

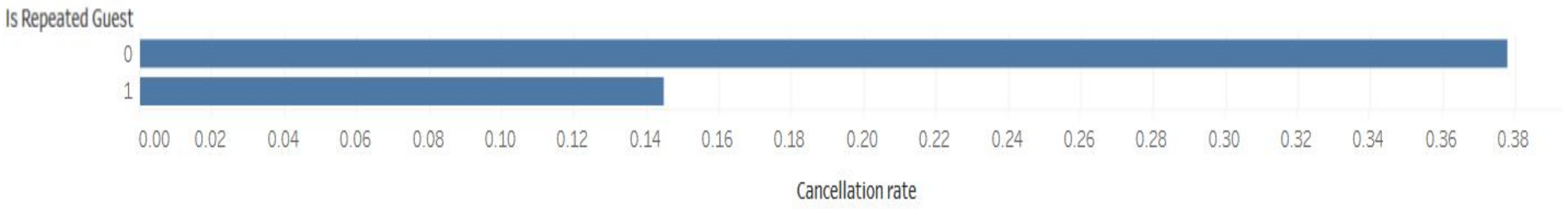


Cancellation Rate Related to unwanted change in Room Type

	Change_of_room_type	Cancellation_Rate
0	0	0.415629
1	1	0.053764

Cancellation rate related to whether a guest is a new customer or not

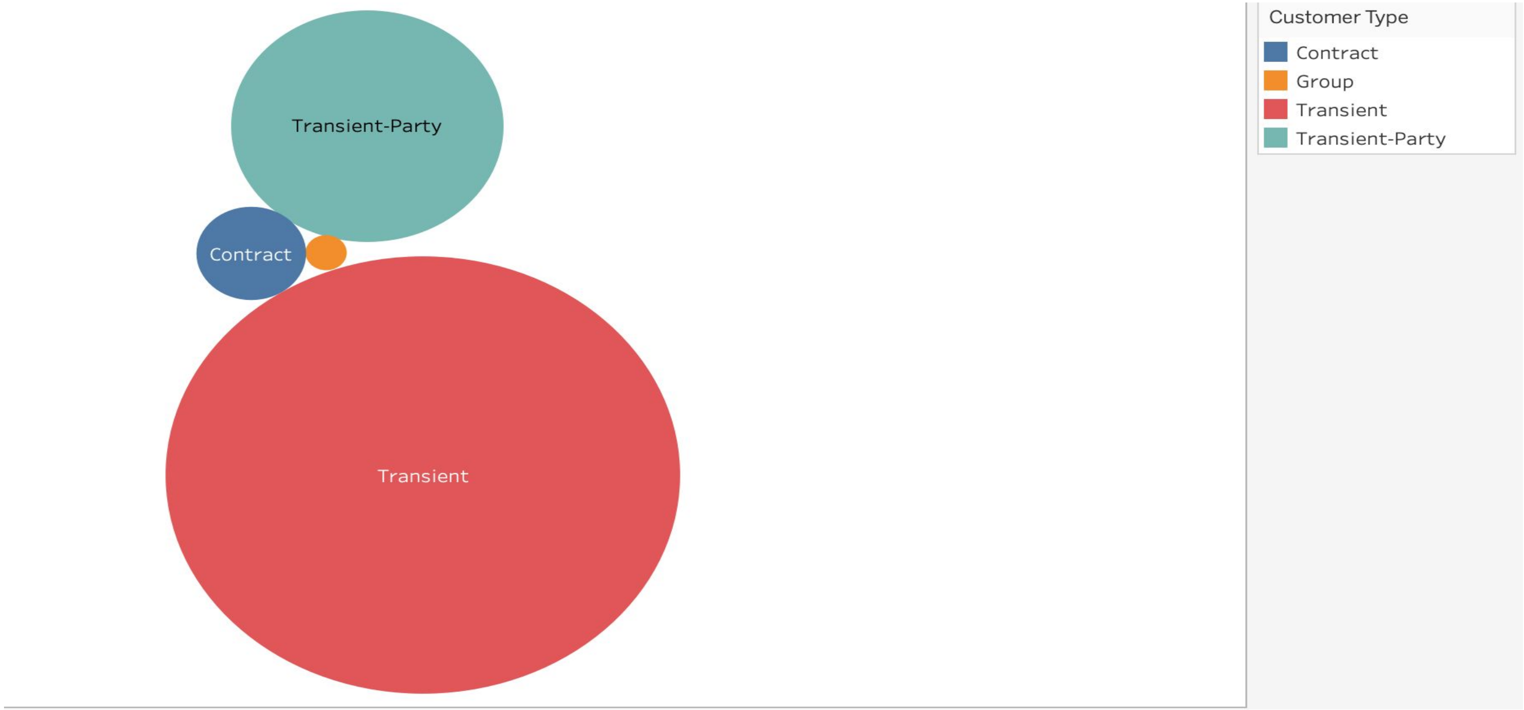
	is_repeated_guest	Cancellation_Rate
0	0	0.377851
1	1	0.144882



Customer composition: What are the different types of customers who make reservations

	customer_type	Description	Number_of_Booking
0	Contract	The booking has a contract associated to it	4076
1	Group	The booking is associated to a group	577
2	Transient	The booking is not part of a group	89613
3	Transient-Party	The booking is associated to other booking	25124

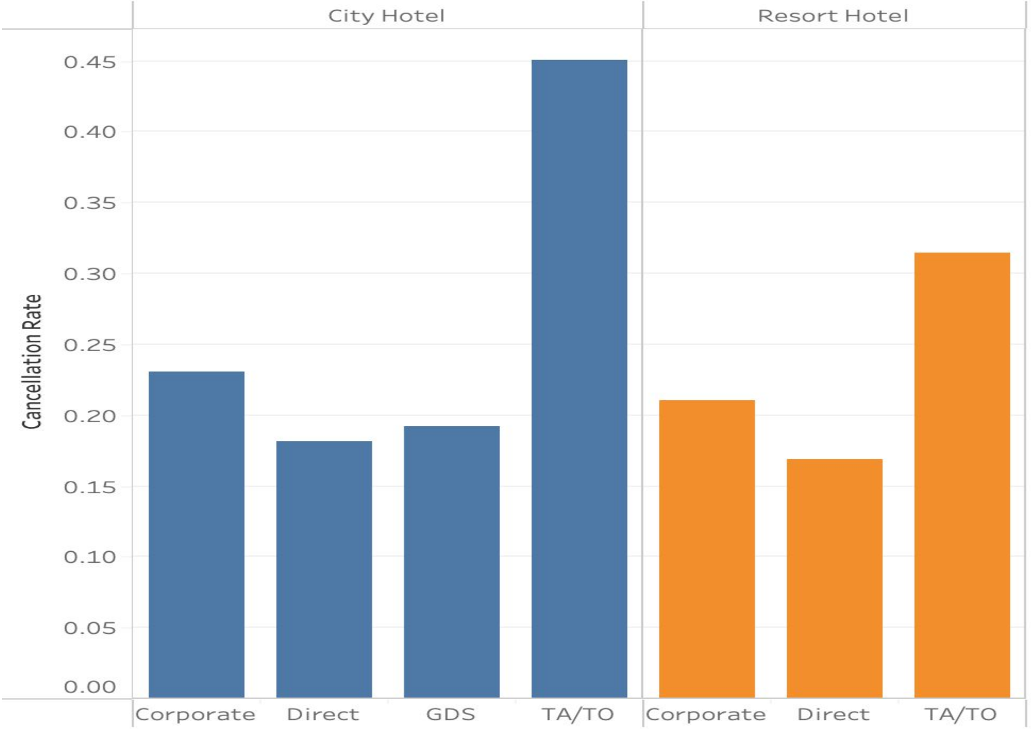
Different Customer types and their share in total number of bookings



Different distribution channels for different hotel types

Hotel	Distribution Channel	Cancellation rate	Is Canceled
City Hotel	Corporate	0.2306	786
	Direct	0.1817	1,232
	GDS	0.1917	37
	TA/TO	0.4503	31,043
	Undefined	1.0000	4
Resort Hotel	Corporate	0.2105	688
	Direct	0.1685	1,325
	TA/TO	0.3149	9,109
	Undefined	0.0000	0

Correlation Between Distribution channels and cancellation rates for different types of hotels



CONCLUSION

This dataset consists of hotel booking cases that are collected from all over the world, and divided into two hotel types: resort hotel and city hotel. We analyzed this dataset to determine the different variables that affected the booking and cancellation rate of the hotels.The variables that had the most impact on the hotel booking rate are **hotel type, peak season, lead time, repeated customers.**

During the peak season from June to August, people tend to book more hotels. The intended stay period is in August, which was indicated by the huge volume of bookings for August. People tend to book for the August season couple of months in advance and are more likely to change their plans as observed from the higher cancellation rate in June.

The next variable that impacts the cancellation rate of hotels is the lead time of booking. We observed from the dataset that a higher lead time leads to a tremendous increase in cancellation rate. For bookings with more than three weeks of lead time, the cancellation rates are almost 5 times than that of bookings with a lead time of 1 week.

The next variable that we investigated was the impact of a room change on the cancellation rate. We expected to see an increase in the cancellation rate when the hotel changed the room type of the customers. But the result was contradictory to our expectations. The cancellation rate of the people who didn't receive a room change was almost 8 times more than that of people who received a room change. This implies that the change in room type didn't compromise on the quality of the rooms and it still met the expectations of the majority of the customers. Additionally, this room change might have been an upgrade for some and hence the very low cancellation rate. We then investigated the booking behaviour of repeated customers. We found that repeated customers tend to be more satisfied with the hotel and have a lower cancellation rate.

Linear Regression

```
. regress is_canceled lead_time is_repeated_guest change_room_type previous_cancellations prev
> ious_bookings_not_canceled

      Source |       SS           df       MS      Number of obs   =   119,390
-------------+----------------------------------   F(5, 119384)    =   3838.21
       Model |  3855.89419         5  771.178838   Prob > F        =    0.0000
    Residual |  23986.8161   119,384  .200921532   R-squared       =    0.1385
-------------+----------------------------------   Adj R-squared   =    0.1385
       Total |  27842.7103   119,389  .233210014   Root MSE        =    .44824

------------------------------------------------------------------------------------------
             is_canceled |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
               lead_time |   .0011301   .0000124    91.09   0.000     .0011058    .0011544
       is_repeated_guest |  -.0849665    .008187   -10.38   0.000     -.101013   -.0689201
        change_room_type |   -.300206   .0039741   -75.54   0.000    -.3079952   -.2924167
  previous_cancellations |   .0512786   .0015636    32.80   0.000      .048214    .0543432
previous_bookings_not_can~d |  -.0099017   .0009626   -10.29   0.000    -.0117885    -.008015
                   _cons |    .289986   .0019648   147.59   0.000      .286135     .293837
------------------------------------------------------------------------------------------
```

Bigquery Machine Learning

We divide the data according to the year, use data in 2015 & 2016 as training set and use 2017 as test set.

Aggregate Metrics

Log loss	0.3841
ROC AUC	0.8976

Score threshold

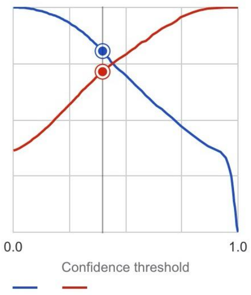
Positive class threshold	0.3978
Positive class	1
Negative class	0
Precision	0.7143
Recall	0.8059
Accuracy	0.8122
F1 score	0.7573

Confusion matrix

This table shows how often the model classified each label correctly (ir

True label	Predicted label 1	Predicted label 0
1	81%	19%
0	18%	82%

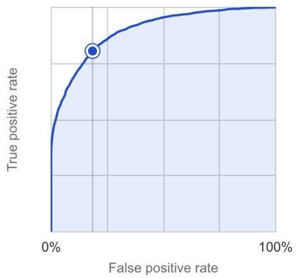
Precision-recall by threshold



Precision-recall curve

Area under curve: 0.864

ROC curve



Area under curve: 0.898

Bigquery Machine Learning

	predicted_is_canceled	predicted_is_canceled_probs	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	stays_in_weekend_nig
0	0	[{'label': 1, 'prob': 0.3086404582589439}, {'l...	City Hotel	1	56	2017	March	
1	0	[{'label': 1, 'prob': 0.4382903883846842}, {'l...	City Hotel	1	95	2017	April	
2	0	[{'label': 1, 'prob': 0.1951686134395547}, {'l...	City Hotel	0	1	2017	January	

3 rows × 25 columns

	Predicted	
Actual	Cancel	Not Cancel
Cancel	66.79%	33.21%
Not Cancel	15.79%	84.21%

References & Materials

- **Dataset Source:** https://www.kaggle.com/jessemostipak/hotel-booking-demand

- **Original Source:** Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019. https://www.sciencedirect.com/science/article/pii/S2352340918315191

- **Tableau Dashboard:** https://www.sciencedirect.com/science/article/pii/S2352340918315191