

User Guide – HotelViz Shiny App

This user guide is meant to guide the user on the effective way of navigating and using the app HotelViz - Visualization of the Hotel Dataset.

The app is divided into three sections, namely:

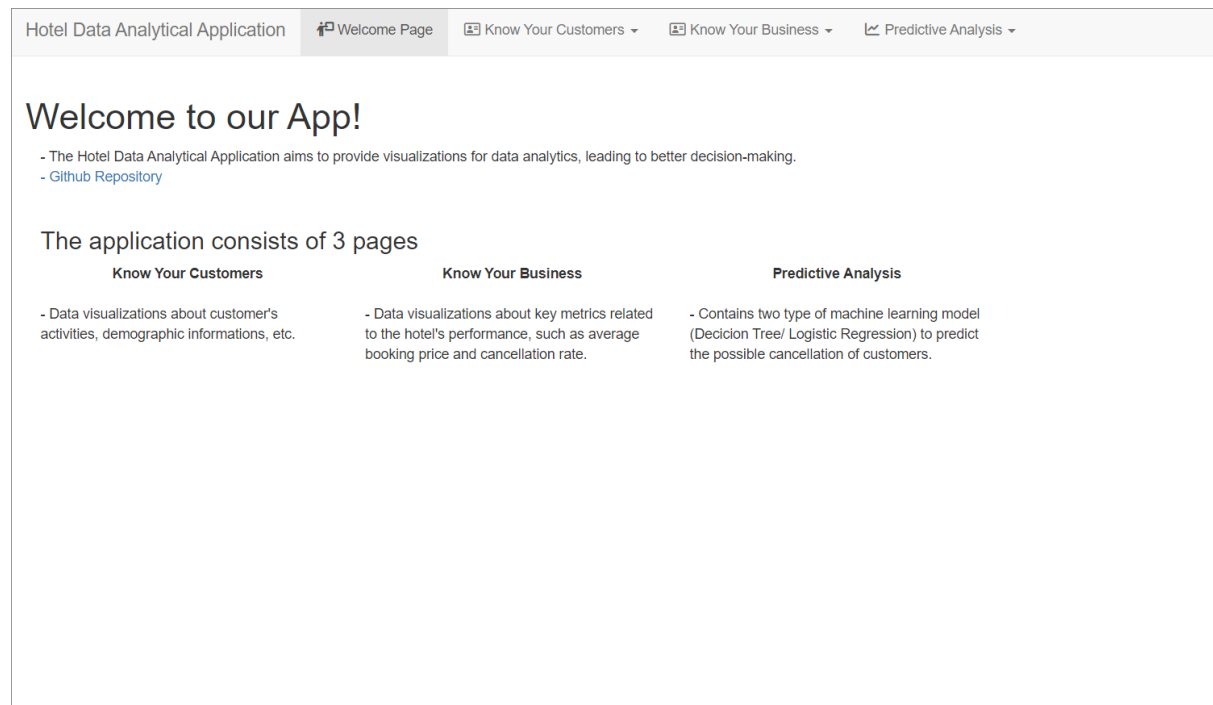
1. Know Your Customers
2. Know Your Business
3. Predictive Analysis

1. Introduction Page

1. Introduction Page

This page features a short description and introduction of the application.

<insert updated screenshot>



2. Know Your Customers

In this section of the app, visualizations meant to understand the distribution of Guests related parameters that help to define and understand hotel guests can be found. There are two tabs in this section, Exploratory Data Analysis and Inferential Analysis.

2.1 Exploratory Data Analysis

2.1.1 Summary Tab

In this summary table, an overall static summary table of the entire Hotel dataset is displayed.

1

2

3

4

Exploratory Data Analysis

Variable Selection

Select Hotel: All

Select Month: All

Select Continent: All

Select Booking Status: All

[Apply Changes](#)

Selection for Bubble Plot

Select Bubble Plot Color Variable: Continent

[Apply Changes](#)

What do we know about our Guests?

[Summary](#) [Understand](#) [Customer Type](#) [Read Me](#)

Overall Summary Table of Hotel Dataset

Characteristic	N	City, N = 79,330	Resort, N = 40,060
ADR, Mean	119,390	105	95
LeadTime, Mean	119,390	110	93
IsCanceled, n (%)	119,390	33,102 (42%)	11,122 (28%)

Reactive Summary Table

Show 10 entries Search:

	HotelType	IsCanceled	total_bookings	mean_adr	mean_lead_time
1	City	0	46228	105.745948342996	80.7027342736004
2	City	1	33102	104.687920367349	150.281221678448
3	Resort	0	28938	90.7889705577442	78.8376183564863
4	Resort	1	11122	105.787010429779	128.680543067794

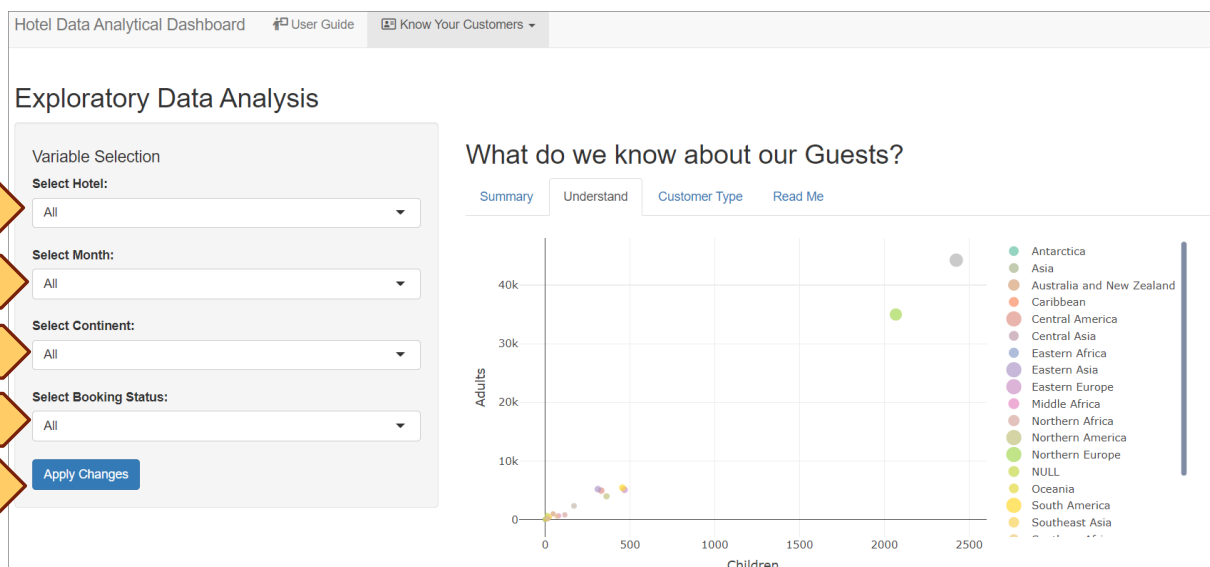
Showing 1 to 4 of 4 entries Previous 1 Next

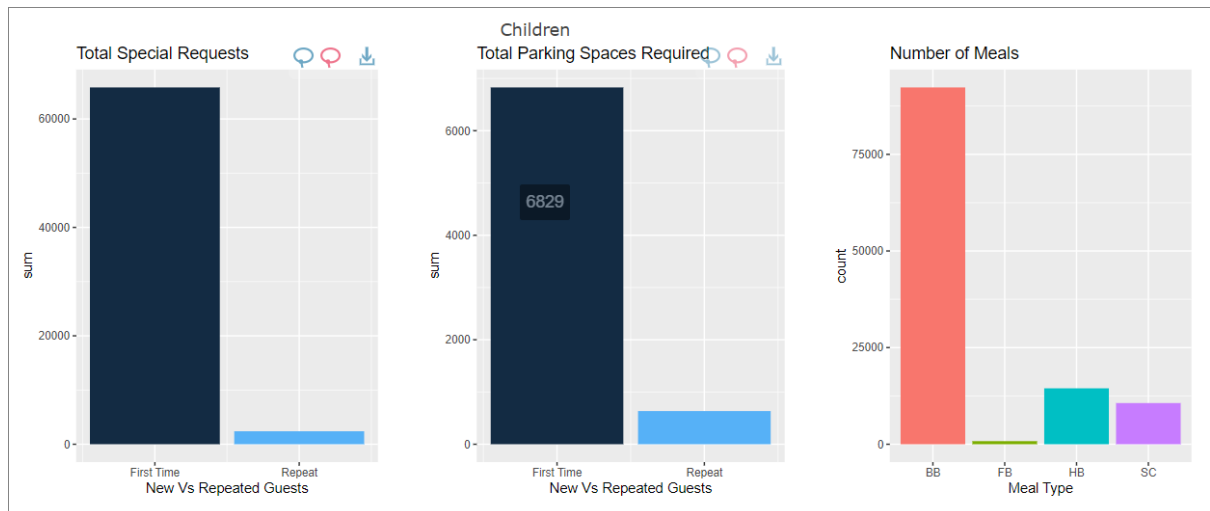
To use this tab effectively:

1. Select the desired Hotel Type (All, City, Resort)
2. Select Month (All or any month of the year)
3. Select Continent (All, or any continent)
4. Select Booking Status (All, Cancelled or Not Cancelled) in the side bar panel.
5. Click 'Apply Changes' to view filtered summary of data with calculated variables 'Total_bookings' to find the size of the data, 'mean_adr' to display the mean Average Daily Rate and 'Mean_lead_time' to display the average lead time that the filtered group has associated with it.

2.1.2 Understand Tab

In this tab, the app displays visualizations to attempt to understand the needs and wants of Hotel Guests.





With selection of the sidebar panel inputs, the various charts in the main panel are updated. The top chart is a bubble chart that shows the relationship between the number of Adults and Children + Babies for an idea of the profile of guest types according to the filters. The color of the bubbles reflects the continent profiles of the guests for further visualization. When hovered over the bubbles, the data points (x = No. of Children + Babies, y = No. of Adults) are displayed.

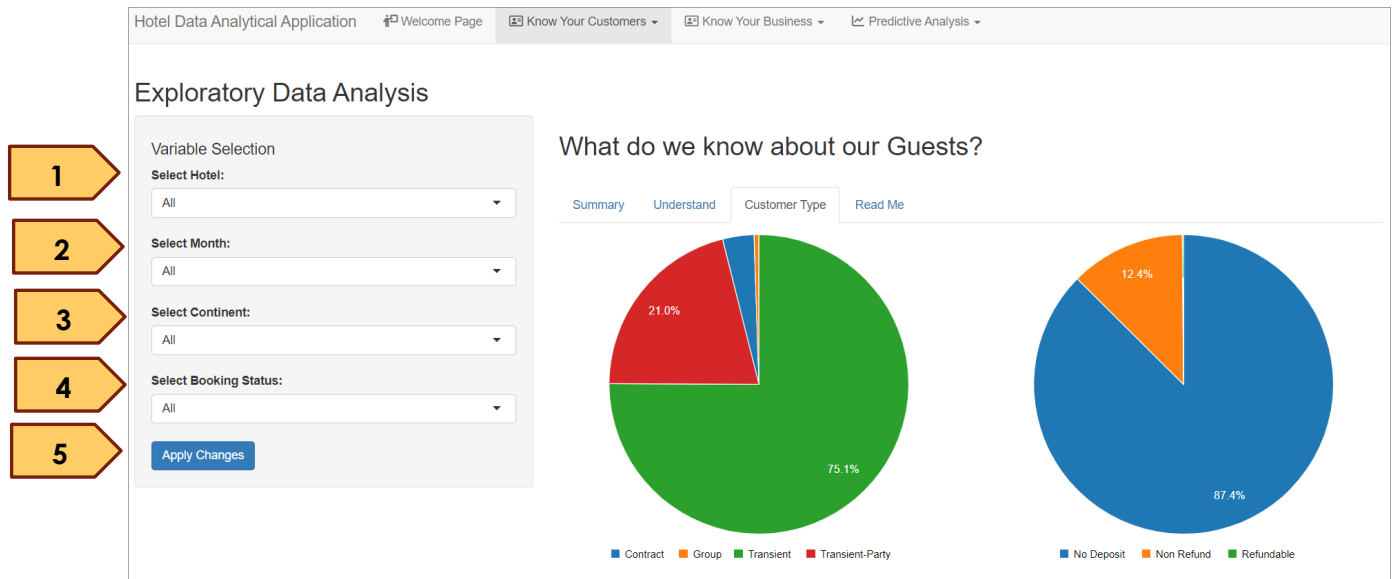
At the bottom, three bar plots are displayed. The first two display the variation in Total Special Requests and Total Parking Spaces Required respectively between first time and repeat guests. The third tab displays the variation in meal types specified by guests. By providing this additional layer of information, the Operations team can better utilize resources.

To use this tab effectively:

1. Select the desired Hotel Type (All, City, Resort)
2. Select Month (All or any month of the year)
3. Select Continent (All, or any continent)
4. Select Booking Status (All, Cancelled or Not Cancelled) in the side bar panel.
5. Click 'Apply Changes' to view the bubble plot and bar plot changes for operationally useful information.

2.1.2 Customer Type

In this tab, interactive pie charts are displayed about the customer type proportions.



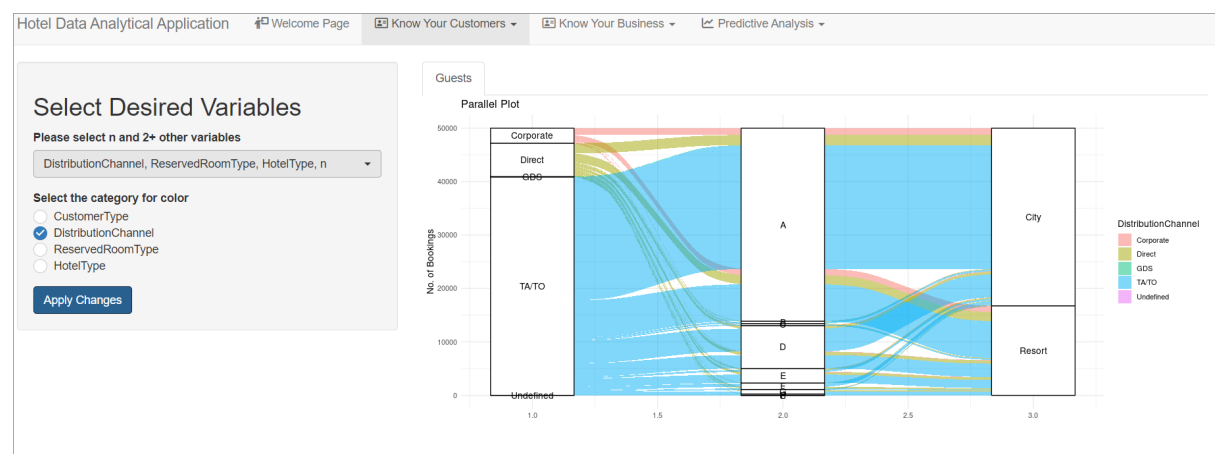
With selection of desired variables to check and apply changes, the proportion of guest types by Customer Type (Contract, Group, Transient, Transient Party) and Deposit Type (No Deposit, Non Refund, Refundable).

To use this tab effectively:

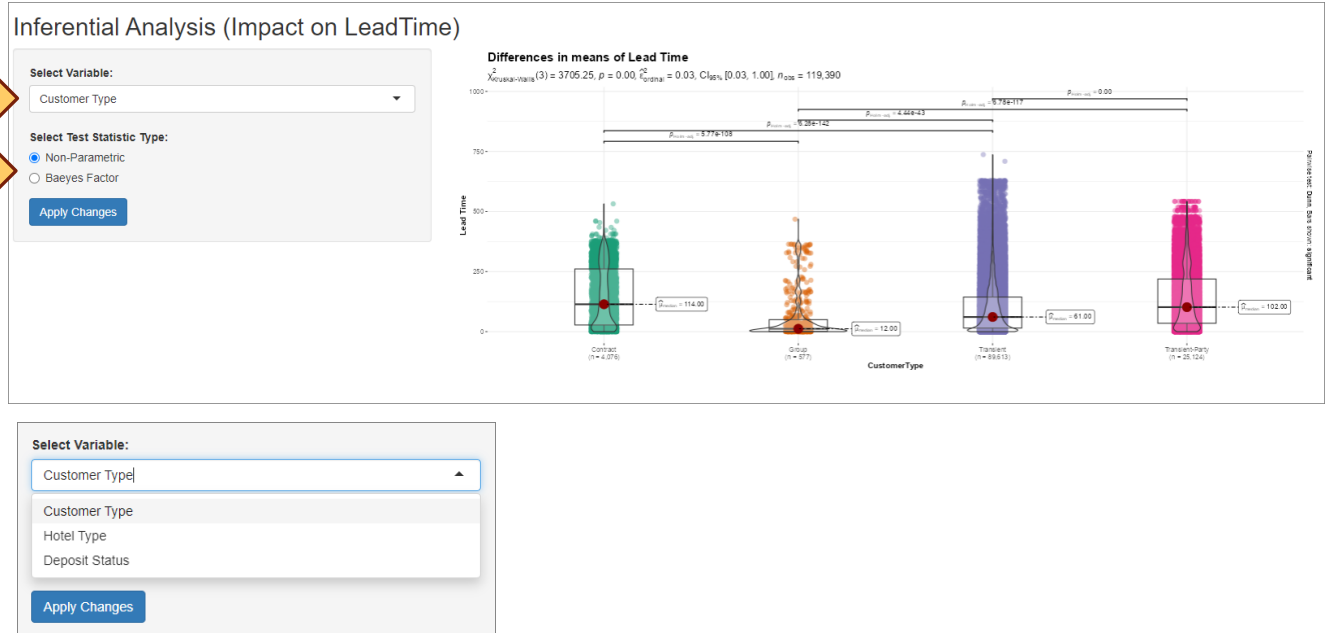
1. Select the desired Hotel Type (All, City, Resort)
2. Select Month (All or any month of the year)
3. Select Continent (All, or any continent)
4. Select Booking Status (All, Cancelled or Not Cancelled) in the side bar panel.
5. Click 'Apply Changes' to view the bubble plot and bar plot changes for operationally useful information.

2.2 Parallel Sets

In this tab,



2.1 Inferential Analysis

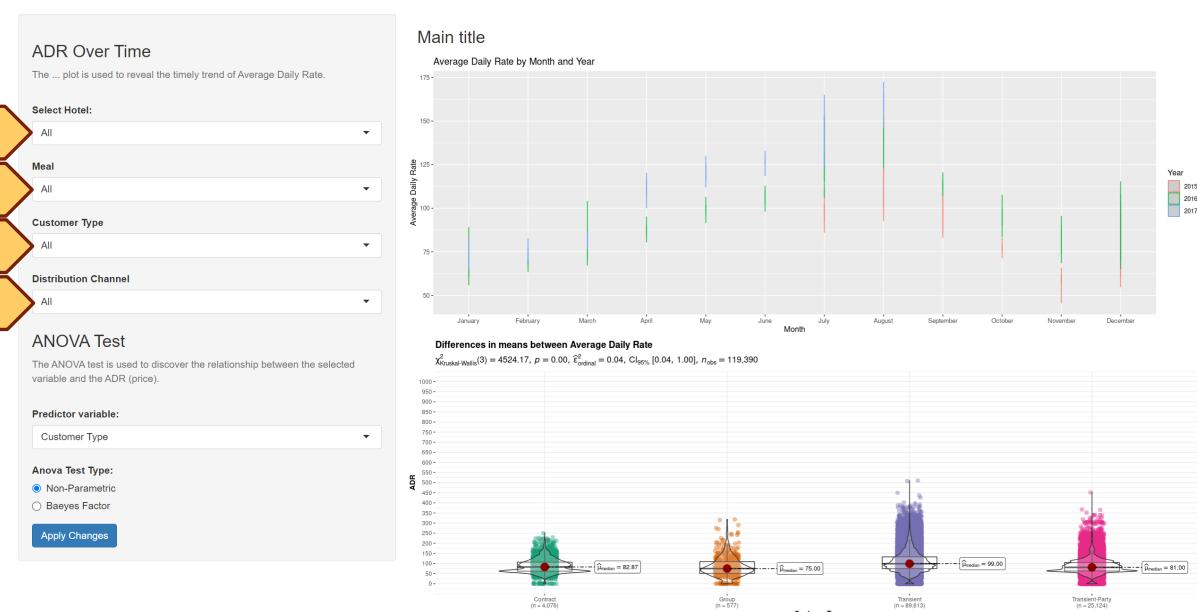


3. Know Your Business

There are two topics under “Know Your Business” drop down menu: Average booking price, and cancellation rate. In the first page, ADR (Average Daily Rate) Over Time will illustrate how the average daily rate changes by different months in a year. Users can apply different filters to see what the important factors are to consider when hotel business operators are making decisions.

3.1. Average booking price

In this page, an overall view of ADR over time will be plotted. There are two charts in a view for users to explore for their businesses.



To use this tab effectively:

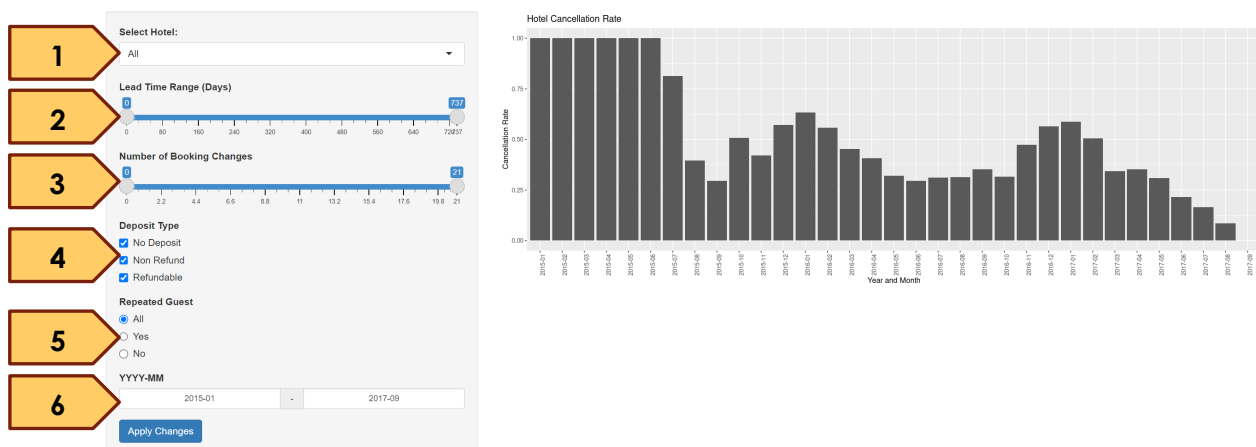
1. Select the desired Hotel Type (All, City, Resort)
2. Select Meal type (All HB, BB, SC, FB, undefined)
3. Select Customer type (All, transient, transient-party, contract, group)
4. ANOVA test: To discover the relationship between the selected variable and the ADR (price)
 - a. Predictor variable:
 - b. ANOVA test type: Non-parametric / Bayes factor
5. Click "Apply Changes"

In this page, there are two different charts in a view:

1. Average daily rate by month and year: This plots ADR by month in an x-axis. Users can check out different ADR by year from the color indicated in the legend.
2. Differences in means between ADR (Average Daily Rate): Users can find out the dispersion of means for ADR. Based on the predictor variable filter selection, the plot will show the spread of means. From this view, users can find out how spread means are, as well as significant outliers they have to take a look at.

3.2. Cancellation rate

In this page, an overall view of ADR over time will be plotted. There are two charts in a view for users to explore their businesses.



To use this tab effectively:

1. Select the desired Hotel Type (All, City, Resort)
2. Slide the slicer to adjust Lead time range
3. Slide the slicer to adjust the Number of booking changes
4. Select Deposit type (No deposit, no refund, refundable)
5. Select Repeated guest (All, yes, no)
6. Select the period range
7. Click "Apply Changes"

In this page, there is a bar plot to see the fluctuation trend of hotel cancellation rate. Based on the slicer and filters applied, users can choose the period they want to take a closer look at, or change input variables such as number of booking changes, deposit type and repeated guest.

4. Predictive Analysis

In this tab, predictive analysis to predict cancellations can be done by two methods: (1) Decision Tree and (2) Logistic Regression.

4.1 Decision Tree

To apply this method, the following steps need to be applied:

1. Select desired model features to be trained in the model. There are 18 variables available to select from. After which, hyperparameters can be selected.
2. Drag the slider input to determine the minimum size splits for the nodes.
3. Drag the slider input to determine the minimum bucket size allowed in the terminal node.
4. Drag the slider input to determine the maximum depth for the decision tree to reach.
5. Click 'Create Tree'
6. To test and validate the results from the trained model, click 'Test Tree' and the test results will be updated in the top right corner.

The Controls

5 & 6 Create Tree Test Tree

Model Features

1 Here are the choices of the predictors to choose:
No. of Non-Canceled Previous Bookings, No. of Previous Cancellations, ...

2 Here are some hyperparameters for tuning the DT:
Minimum Split
Control the minimum size allowed for the node to split further
Slider: 50 to 1,000 (set at 350)

3 **Minimum Bucket Size**
Control the minimum size allowed in the terminal node
Slider: 10 to 300 (set at 60)

4 **Maximum Tree Depth**
Control the maximum depth that the decision tree can reach
Slider: 2 to 15 (set at 8)

Training Results
training size is limited due to server memory capacity
Overall Accuracy: 75.2%
Precision: 67.1%
Recall: 65.9%

Outcomes	Actually Not Cancelled	Actually Cancelled	Total
Predicted Not Cancelled	5068	1269	6274
Predicted Cancelled	1206	2457	3663
Total	6274	3726	10000

Test Results
Overall Accuracy: 75.2%
Precision: 66.9%
Recall: 65.4%

Outcomes	Actually Not Cancelled	Actually Cancelled	Total
Predicted Not Cancelled	55810	14017	68892
Predicted Cancelled	13082	26481	39563
Total	68892	40498	109390

Decision Tree
Class 0: Predicted not Cancelled; Class 1: Predicted Cancelled

```
graph TD
    Root[DepositType = NDip,Rtn] -- yes --> Node1[LeadTime < 7.5]
    Root -- no --> Node2[PreviousCancellations < 0.5]
    Node1 -- 0 --> Leaf1[0]
    Node1 -- 1 --> Node3[CustomerType = Cnt,Grp,T,P]
    Node2 -- 0 --> Leaf2[0]
    Node2 -- 1 --> Node4[TotalOfSpecialRequests >= 0.5]
    Node3 -- 0 --> Leaf3[0]
    Node3 -- 1 --> Leaf4[1]
    Node4 -- 0 --> Leaf5[0]
    Node4 -- 1 --> Leaf6[1]
```

Classification

- The tree graph on the left depicts the suggested method by a trained decision tree model, which classifies whether a customer is likely to cancelled the order.
- The accuracy of the model can be referred by the training and testing accuracy on the top.
- Operation team can utilise this model as a guide and pay extra attentions to the customer who are likely to churn.

The decision tree will then be displayed and the Operations team can obtain an indicator of which feature would be more important to monitor for guests who may have higher propensity to cancel a booking.

4.2 Logistic Regression

To apply this method, the following steps need to be applied:

1. Select desired model features to be trained in the model. There are 18 variables available to select from.
2. Click 'Create Model'
3. To test and validate the results from the trained model, click 'Test Model' and the test results will be updated in the top right corner.

5 & 6

1

The Controls

Create Model Test Model

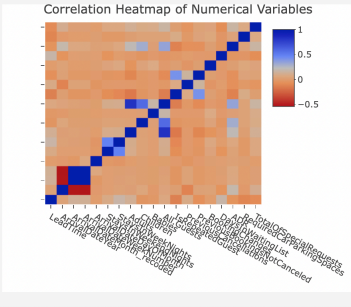
Model Features

Here are the choices of the predictors to choose:

No. of Non-Canceled Previous Bookings, No. of Previous Cancellations, ▾

Test For Multi-Collinearity

Below shows the correlations between each of the numerical variables. Highly correlated pairs need to be excluded from the logistic regression model.



Training Results

training size is limited due to server memory capacity

Overall Accuracy: 74.3%
Precision: 70%
Recall: 54.2%

Outcomes	Actually Not Cancelled	Actually Cancelled	Total
Predicted Not Cancelled	13526	4260	15692
Predicted Cancelled	2166	5048	7214
Total	15692	9308	25000

Test Results

Overall Accuracy: 73.9%
Precision: 68.9%
Recall: 53.9%

Outcomes	Actually Not Cancelled	Actually Cancelled	Total
Predicted Not Cancelled	13465	4272	15727
Predicted Cancelled	2262	5001	7263
Total	15727	9273	25000

Importance of Variables

