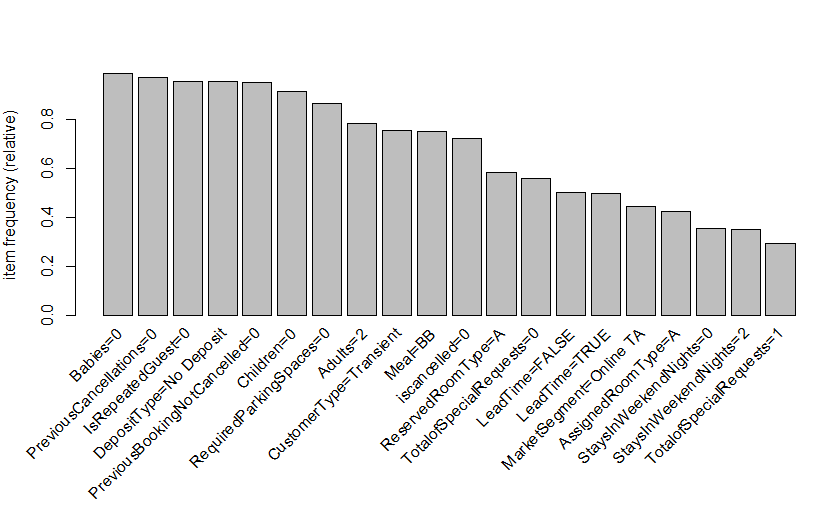
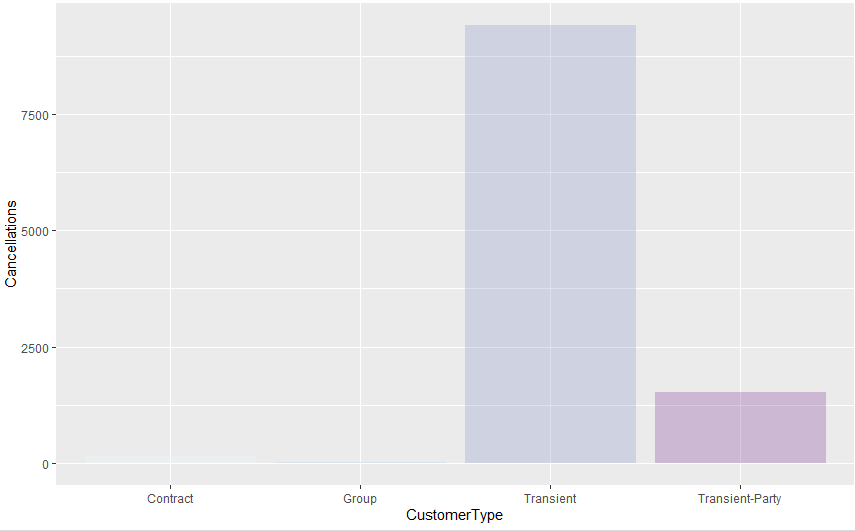
**Item Frequency plot**

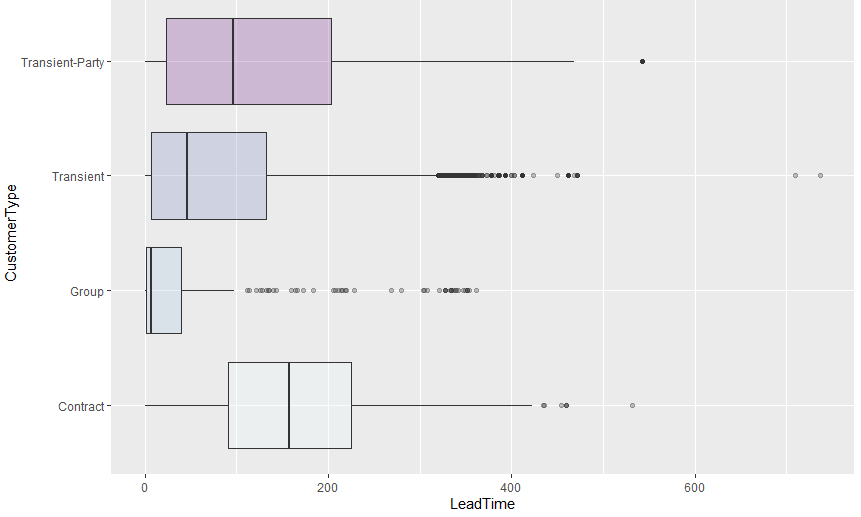


This bar plot displays the relative frequency of the different variables in the dataset. The higher the relative frequency the higher is the probability of the variable being of statistical significance.

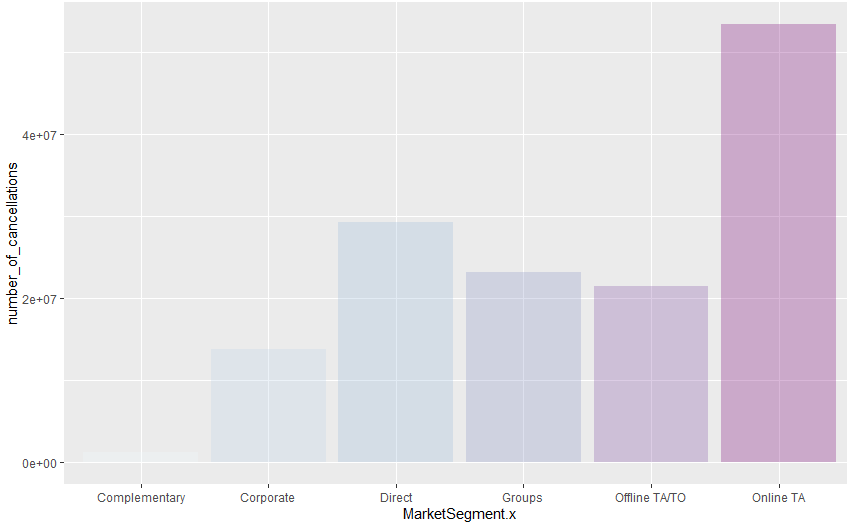
**Bar Plots and Box Plots**



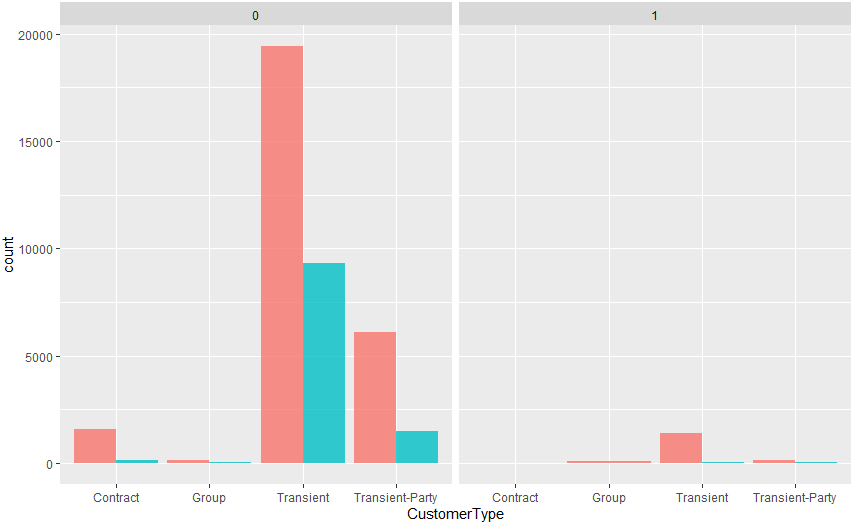
The customer type transient has the highest cancellations amongst all the customer types. This bar plot when looked at standalone can be misleading as there is no information on how many bookings are of customer type transient. The boxplot provides more insight into customer types and their respective lead time.



Lead time and customer type are the variables with the high statistical significance. Finding the relationship between the variables can be a helpful insight into predicting cancellations. The customer type ‘contract’ has a higher lead time. Our ML algorithm suggests that higher lead time leads to higher number of cancellations. This means that the probability of contract customer type being a cancellation is higher than any other customer type.



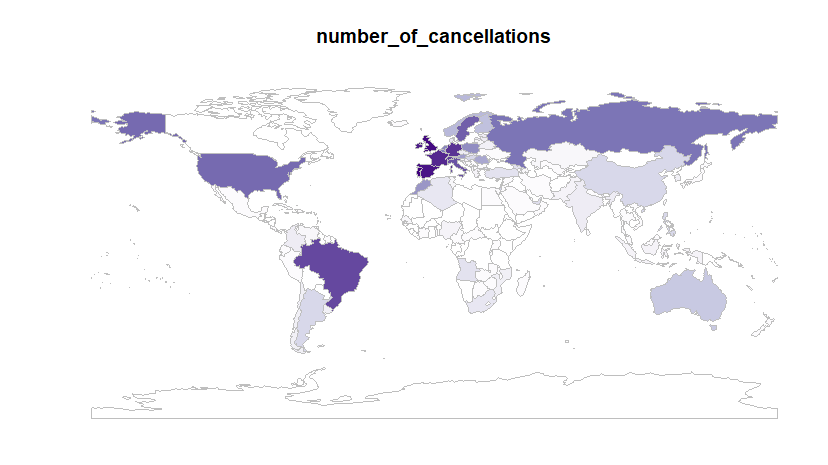
This bar plot shows the number of cancellations for different market segments. This bar plot shows that the market segment Online TA has the highest number of cancellations.



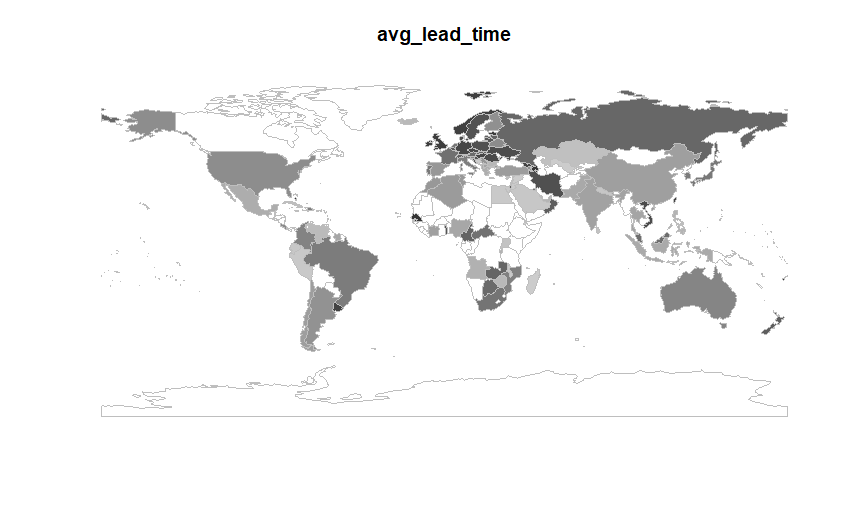
The number of cancellations and number of no cancellations for different customer types are displayed. The bifurcation for repeated customers is tracked. 0 represents new customers, 1 represents repeated customers.

This graph tells us that repeated customer cancellations are low through out the customer types. This graph tells us that 1/3rd of the transient customers that are new customers are likely to be a cancellation. This can be extrapolated to other customer types as well.

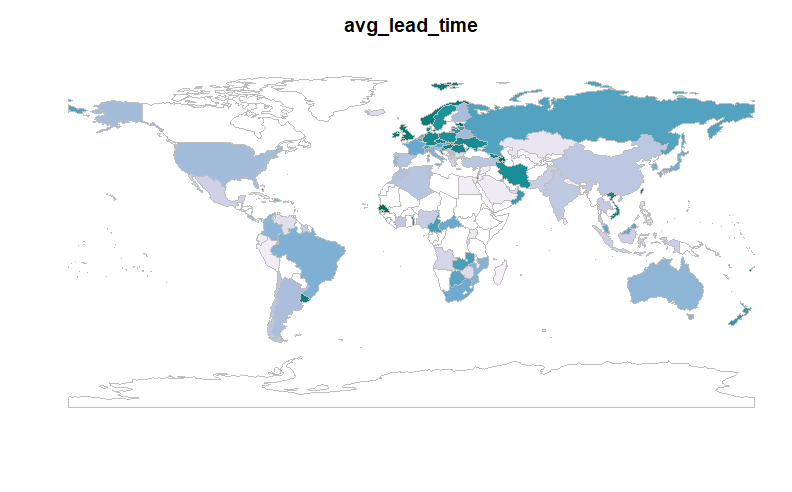
**Maps**



This map displays the mean lead time gradient. Since the lead time is one of the higher significance variables, knowing the mean lead time for different countries will be a helpful insight. This information can used to predict the probability of higher number of cancellations. Hotels in Europe tend to have longer lead times compared to other countries.

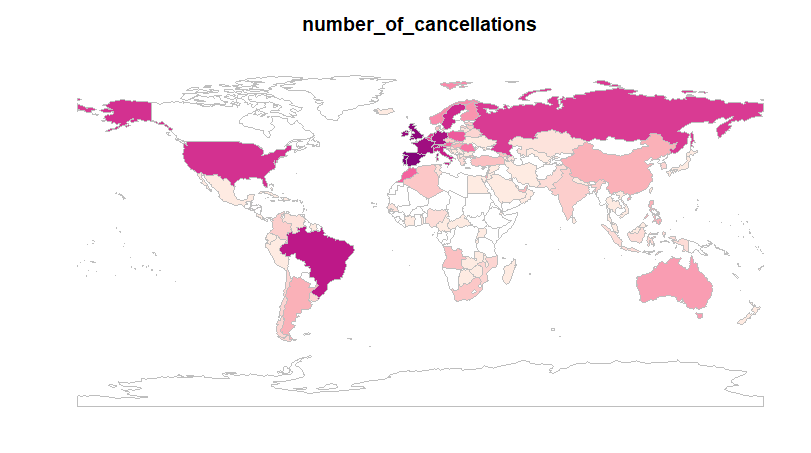


This plot displays the number of cancellations for different countries of the world. The number cancellations are higher in the European countries which is supported by our analysis of higher lead time in the European countries.

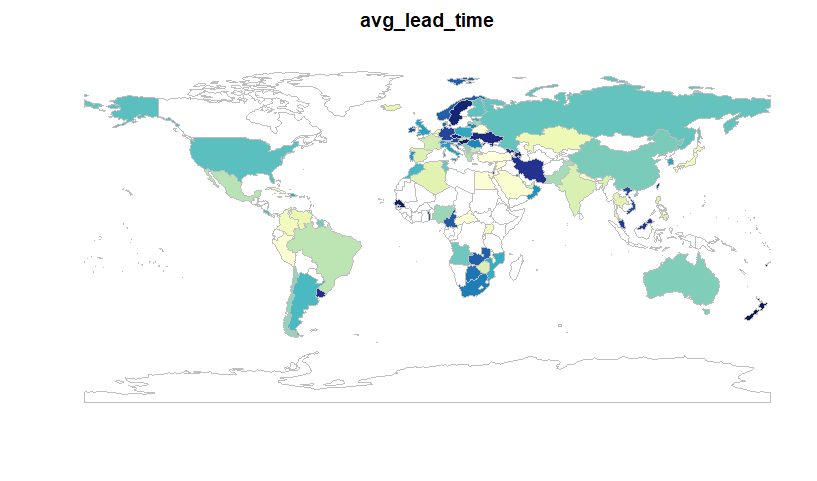


The No Deposit has a higher statistical significance and hence finding the mean lead time and number of cancellations when the deposit type is no deposit can be important insight into opening and expanding the business.

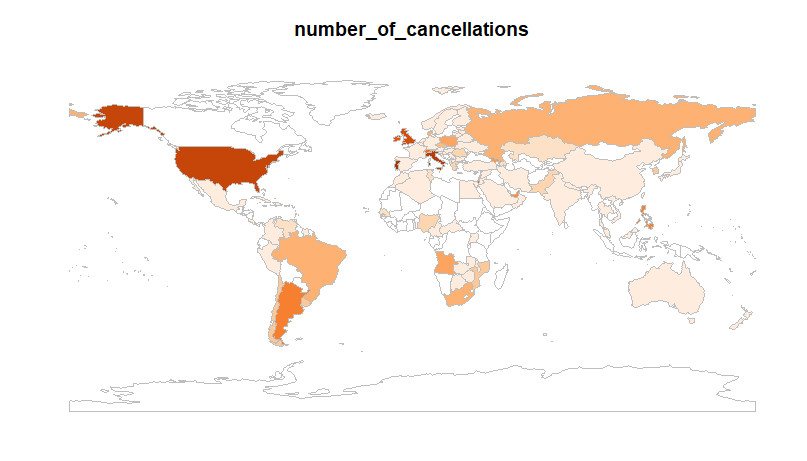
The lead time for the no deposit observations higher in the European countries and Russia.



The plot represents the number of cancellations in different countries when the deposit type is no deposit. The number of cancellations are higher in United States, European countries and Russia.

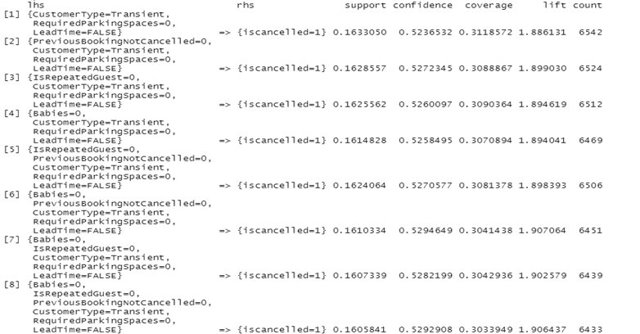


As seen in the barplots, the customer type transient is a high impact variable and its relation with lead time shows us that the probability of higher cancellations are In European Countries, USA, Russia etc

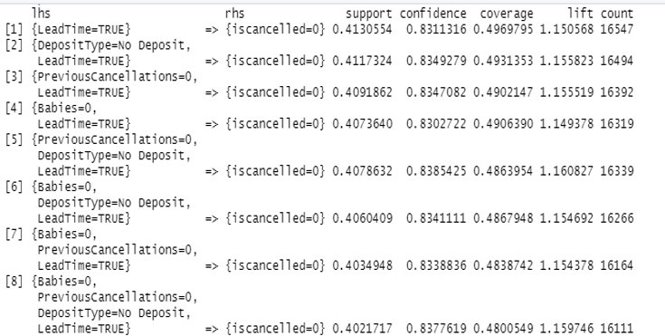


The number of cancellations for transient customer type are highest in USA and European countries, which confirms our assumption from the above plot.

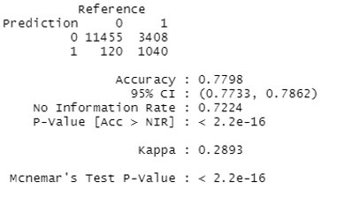
Our unsupervised ML model tells us the different rules for when the booking is canceled. As we can see, Lead time, Is repeated guest, customer type, previous cancellations are the high impact variables



 The rules for when the there is cancellation. As we can see, lead time and deposit type and previous cancellations are high significance variables



Supervised Machine Learning model



The model predicts if the provided observation is a canceled. 78 % accuracy is sufficient for the business purpose. The No information rate does not fall in the 95% CI rate indicating that the model is good. Statistical significance of the model is also high.