

# Finding Potential Customers to advertise Carribean Holiday Tour

---

Ravi Dahiya  
Finn Tan  
Khalid Gharib

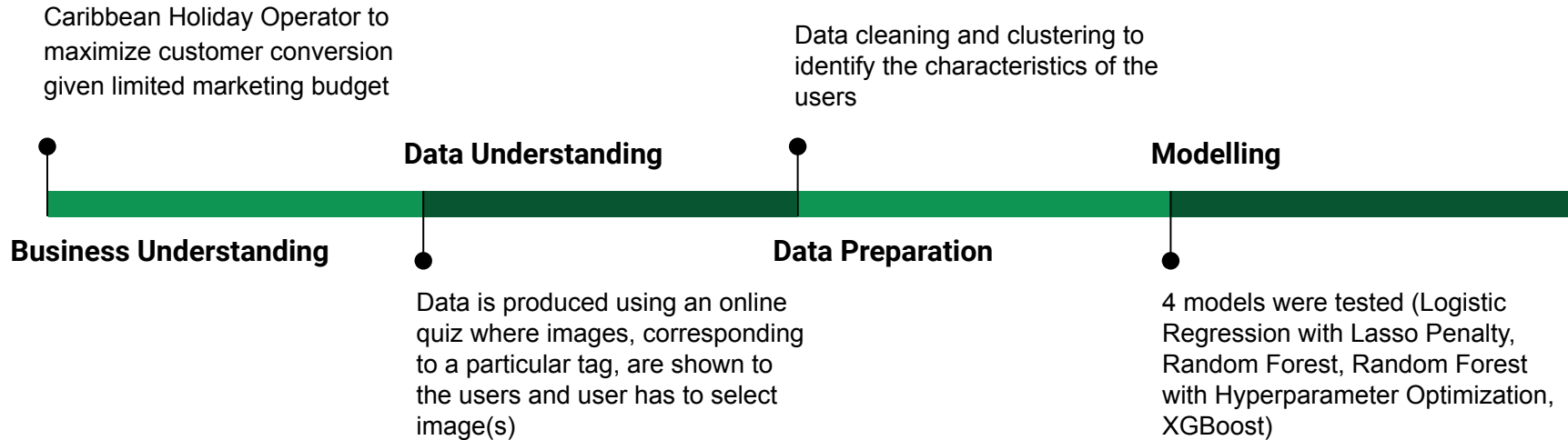
# The challenge

We are consultants hired by a Caribbean Holiday Operator to advertise their trips to potential customers online.

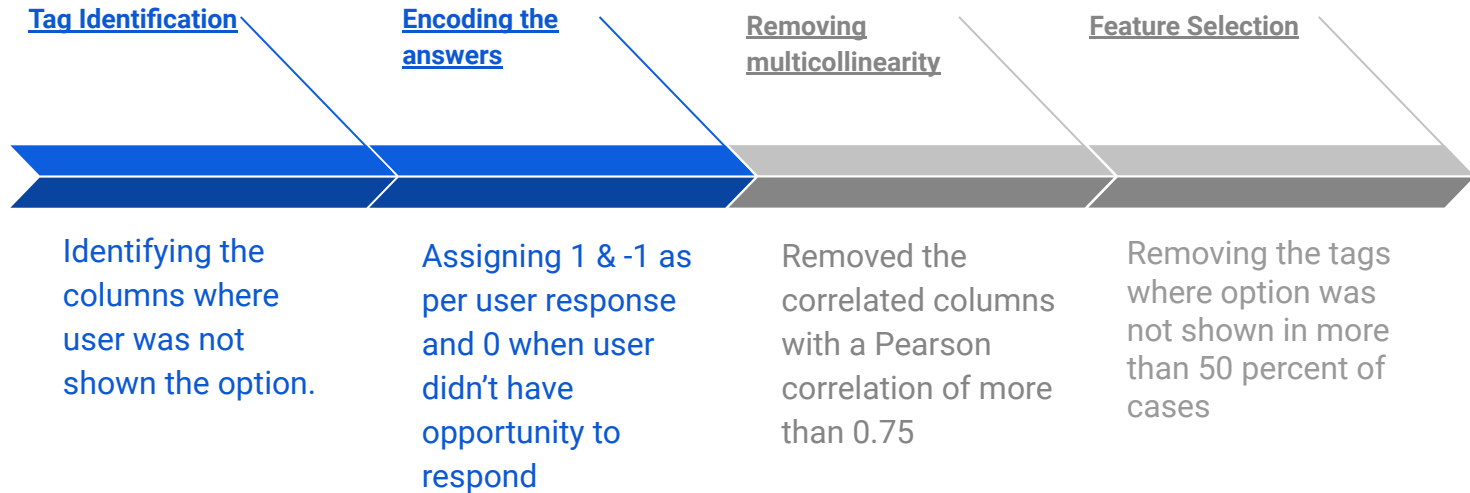
We have list of customers who have bought the tour from them in past. Now we have to predict which of the other users in the dataset are likely to be interested in their offers as well.

---

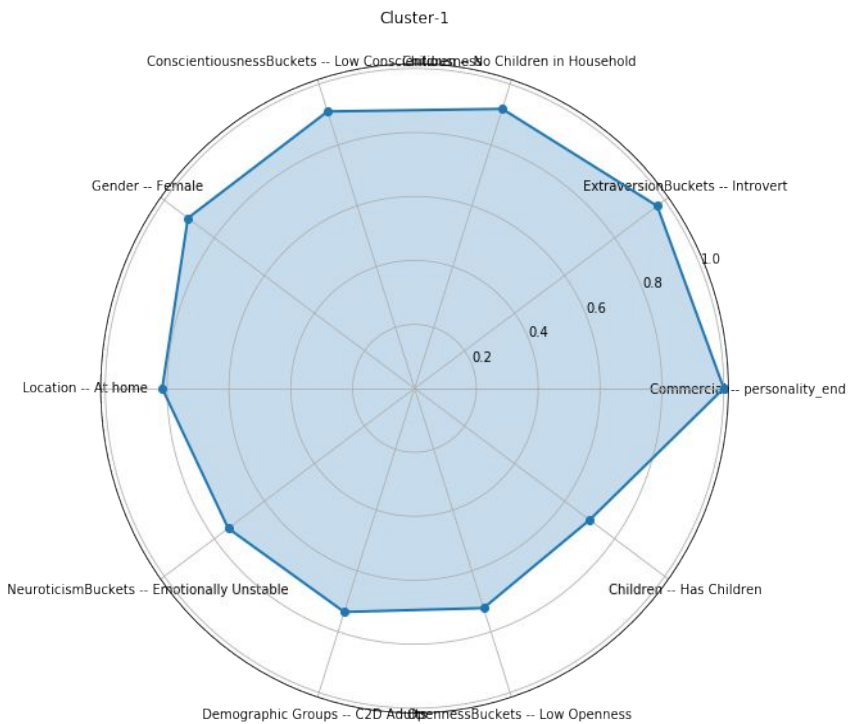
# Methodology



# Data Preparation

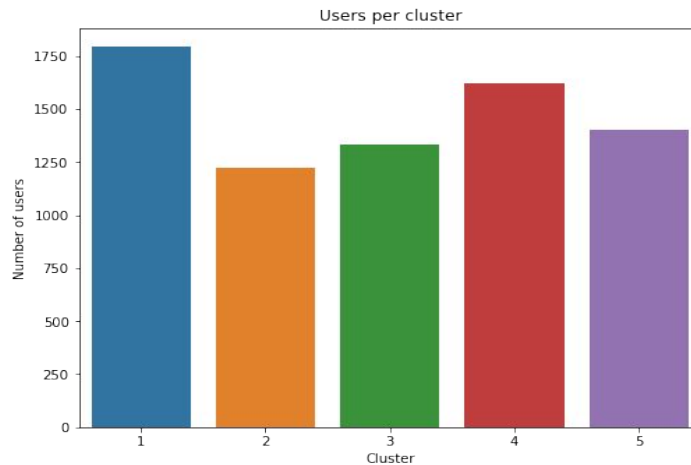


# Clustering

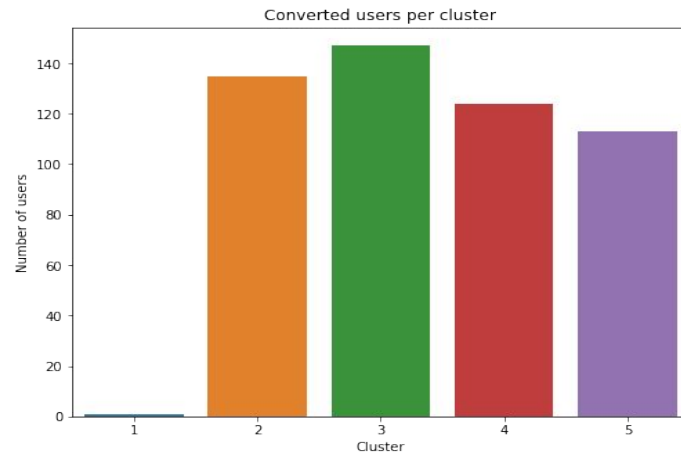


**Top-10 user Characteristics in Cluster-1**

## Clustering for the entire user database



## Clustering for the converted users



# Modelling

	Accuracy	Precision	Recall	F-1 Score
Logistic Regression with GridsearchCV (Lasso Penalty)	0.98	0.92	0.98	0.95
Ensemble - Random Forest	0.94	0.92	0.79	0.85
Ensemble - Random Forest with Hyperparamter Optimization	0.98	0.92	0.98	0.95
Ensemble - XGBoost	0.98	0.92	0.97	0.94

1. Classification scoring metric including Accuracy, Recall, Precision and F1 scores were compared.
2. 'Recall' was prioritized over other metrics because logically speaking, the cost of False Negative outweighs False Positive, representing the loss of potential sale vs loss of marketing cost.
3. All the models (except Random Forest) have similar performance.
4. Logistic regression(with Lasso Regularization) is computationally cheap and would be using only 9 columns instead of 226. So, Logistic regression with Lasso Regularization is used to find the final probabilities for the users.

# Results

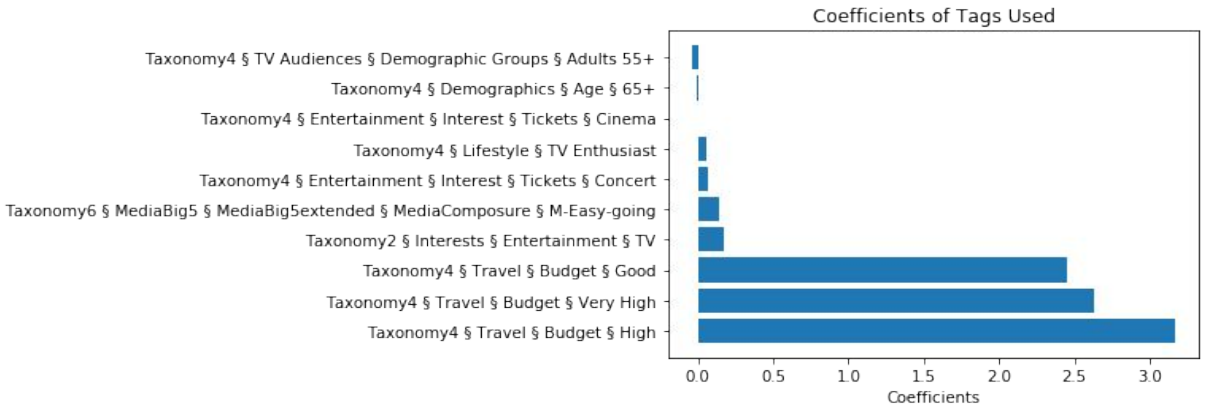
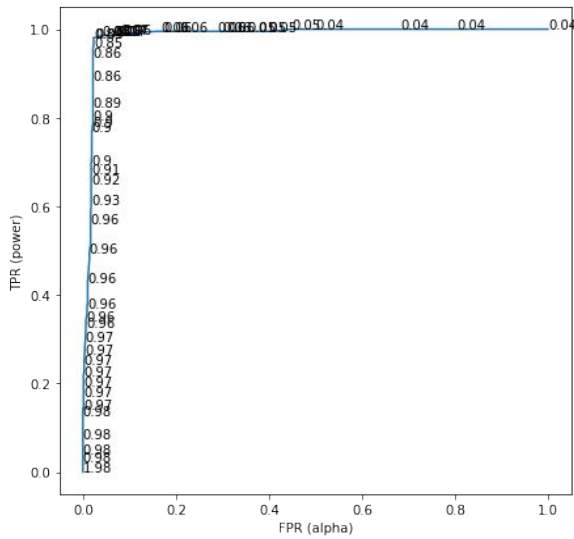
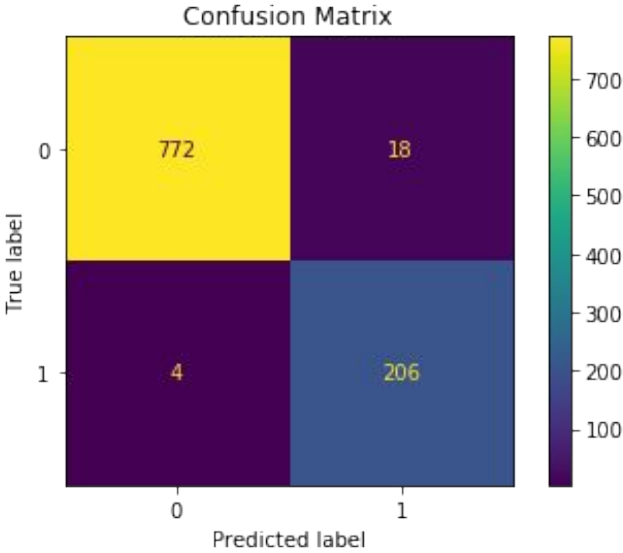
Best Penalty: {'C': 0.05}  
\*\*\*RESULTS SUMMARY\*\*\*

Model: Logistic Regression with Lasso Penalty  
Dataset: Validation Set  
Target distribution:

Class: 0 / Count: 790 / Pct: 79.0  
Class: 1 / Count: 210 / Pct: 21.0

Metric Scores:

Accuracy score: 0.98  
Recall score: 0.98  
Precision score: 0.92  
F1 score: 0.95



This is the most  
important takeaway  
that everyone has to  
remember.



# Conclusion

What is the conclusion of your experiment? Did the results support your hypothesis or predicted outcome? How will your findings help the area of science you've researched?

# What will I do next?

What will you do with your findings next? How will you further your research/findings?

