Applying Data Mining Techniques to Direct Marketing

Spring 2019

ISM 6136 Data Mining

University of South Florida

By:

Priyanka Sariya

Hariharan Murali Krishnan

Anuradha Jain

Table of Contents

Introduction

Executive Summary

For any business to be successful it must find an appropriate way to approach their prospective customers and make them aware of the services/products. Marketing department tackles this problem for the business. In Direct marketing, the business makes use of various channels such as, SMS, email, websites, online advertisements, etc. The common constraints faced when dealing with this problem are resource, time and cost optimization. In this project we are trying to build a good model that would help us to predict if a customer will visit the store after a marketing campaign. This would help business to maximize the customer base through focused marketing campaign, targeting the right people and minimizing the cost of campaign. The analysis builds and compares the performance of boosted decision tree and neural networks model and suggests that boosted decision tree performs best on the dataset at hand.

The champion model suggests that the historical purchase pattern and the segment have significant effect on the success of the market campaign and should be considered while targeting customers in the campaign.

The model can be made even better by incorporating the average of cost of campaign person and profit margin per customer.

Business Objective

To better focus the marketing campaign funds, resources and time, data mining analysis is done on the historic direct marketing campaign data to uncover the hidden patterns in data and build a model that would better identify important attributes that contribute to the success of the campaign.

Data Sourcing and Tools Used

Data Source: The authenticity of the data is a very important aspect that has to be considered. The data used for this project was sourced from Azure ML data repository. Azure ML data platform is a collection of services which is used to store, prepare, ingest and analyze data.

Tools Used: R tool was used for data pre-processing and Azure ML was used for data modelling. Tableau was used for doing exploratory data analysis and visualization to get some insights on the data.

Data Mining Methodology

Before starting analysis on the data, it is important to have some idea about the data and derive insights from it. So, visualization was done in Tableau (graphs included in appendix)**.**

Pre-processing of Dataset

The dataset contains 64,000 records with *eight attributes and a response* (Visit) column. The response column is a binary variable that takes value 1 when a customer visited the store after the marketing campaign and 0 otherwise. (Graphs included in appendix)**.**

Feature Engineering

The dataset prepared for analysis has 9 columns, where 8 columns are predictors and 1 column *Visit* is response. The description of the variables is as follows:

Recency**:** The most recent shopping of the customer in the last 12 months. This is a categorical variable and will hold values between 1 and 12 only.

History**:** This contains the amount spent by the customer in their last visit

Marital Status: This column has 3 values - single male, single female, and married.

Zip_code**:** This is a categorical variable representing the customer's zip-code and can take values: Urban, Suburban and Rural

Newbie**:** This is a binary column which can take value 0 and 1

Channel**:** Categorical variable which can have 3 values – web, phone, or multichannel which represents the channel through which the customers were approached.

Segment**:** Categorical variable that will hold 3 values - Mens E-Mail, No E-Mail, Womens E-Mail

DM_category**:** This is a categorical variable which is the cluster of the customers in one of the 7 categories

Visit**:** This is our response variable where we have to predict if the customer will visit the store or not after direct marketing

The following columns were removed from the dataset

The *history_segmen*t column was removed because it was a simplification of *history* column.

The ***conversion*** and ***spend*** were eliminated because we were not using them in the analysis.

Mens and **Womens** columns are converted to a new column where 1 0 means Single Male, 0 1 is Single Female, and 1 1 is Married. The new column name is 'Marital Status'

The results of the exploratory data analysis suggests that the ratio of customers “visited” to “not visited” is very low and highly skewed. Using the dataset as it is to run the model would most likely have high accuracy but it will be highly biased to majority class in the target variable. This is an example of an unbalanced classification problem since the percentage of visits is only about 2% of the data. This is a classical trap in data mining which should be avoided.

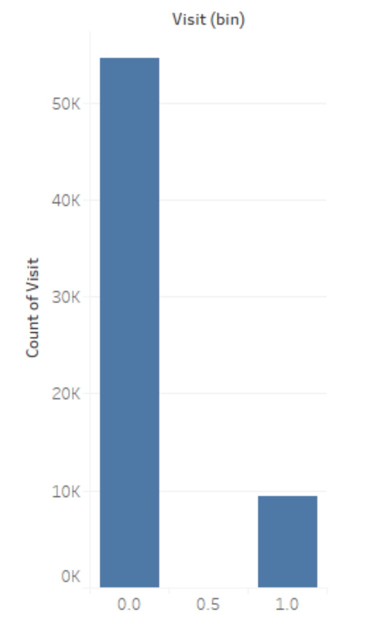


Figure 1: Histogram of Visit

 # visit == 0 : 54606

 # visit == 1 : 9394 this suggests the target variable is highly skewed.

Random under sampling of the majority class technique was employed for sampling. Sample was prepared by random sampling of 90% of the 1’s (which is 8454) and the same number (and not percent) of 0’s of the target variable. The training dataset has 16908 rows and 9 columns. The remaining data was used for test dataset, which had 47092 rows and 9 columns. This sampling strategy ensures equal distribution of classes in target variable and was used to perform data modelling in Azure ML.

The most worth or significant variables are as below in figure 2.

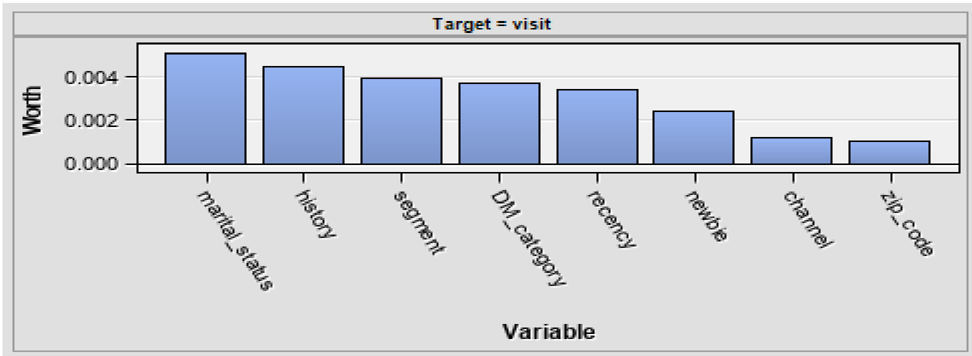


Figure 2: Variable Worth for the Direct Marketing Dataset

In the figure above, we see that *marital status, history, segment, DM category* and *recency* are more worthy than *newbie, channel* and *zip code*, with *zipcode* being the least significant variable.

Data Mining Models

A class of Supervised Classification techniques were used to classify the customers for marketing. The response variable was "visit" and all the other variables namely marital_status, history, segment, DM_Category, recency, newbie, channel, zip_code were used as predictors. We used three different modelling techniques for analysis.

1. Two-Class Boosted Decision Tree
2. Neural Network

The following graphic shows the overall project workflow:

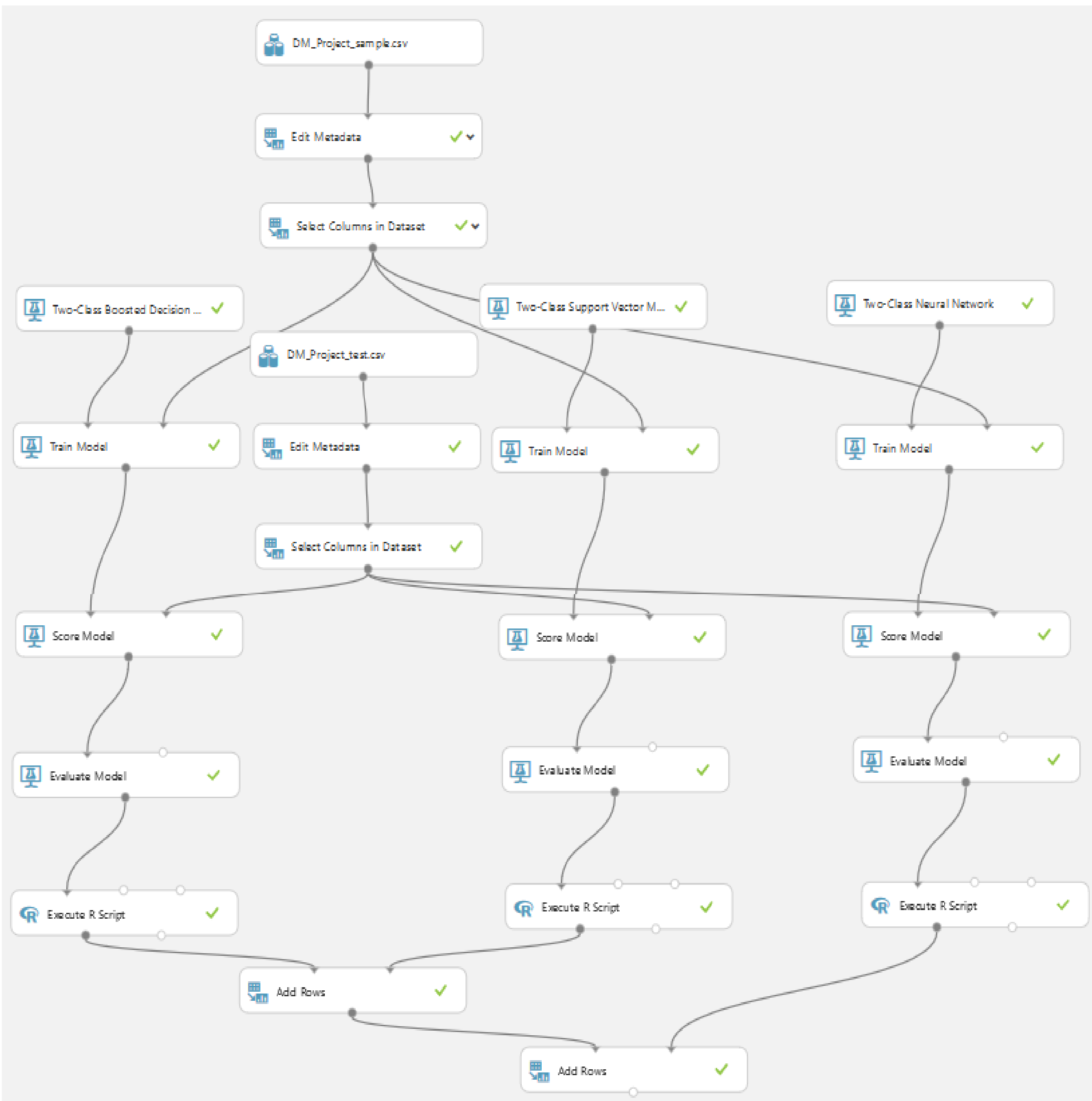


Figure 3: Decision Tree and Neural Network models in Azure ML

Model 1: Two-Class Boosted Decision Tree Classification

This classification creates a binary classifier using a boosted decision tree algorithm. A boosted decision tree is an ensemble learning method in which the second tree corrects for the errors of the first tree, the third tree corrects for the errors of the first and second trees, so on and so forth. Predictions are based on the entire ensemble of trees together that makes the prediction.

Boosted decision trees are the easiest methods with which we get high performance on a wide variety of machine learning tasks. However, they are also one of the more memory-intensive learners, and the current implementation holds everything in memory. So, a boosted decision tree model might not be able to process the very large datasets.

For the project, we used below hyper-parameters configurations:

Maximum number of leaves per tree: This indicate the maximum number of terminal nodes (leaves) that can be created in any tree. For our model, we set this to 10.

Minimum number of training instances required to form a leaf: This indicate the number of cases required to create any terminal node (leaf) in a tree. For our model, we kept this to 10.

Total number of trees constructed: This indicate the total number of decision trees to create in the ensemble. By creating more decision trees, you can potentially get better coverage, but training time will increase. For our model, we set this to 100.

Using the training dataset, our Two-Boosted Decision Tree Model is trained and then this resultant model is then tested on the Testing dataset. Below is the evaluation result:

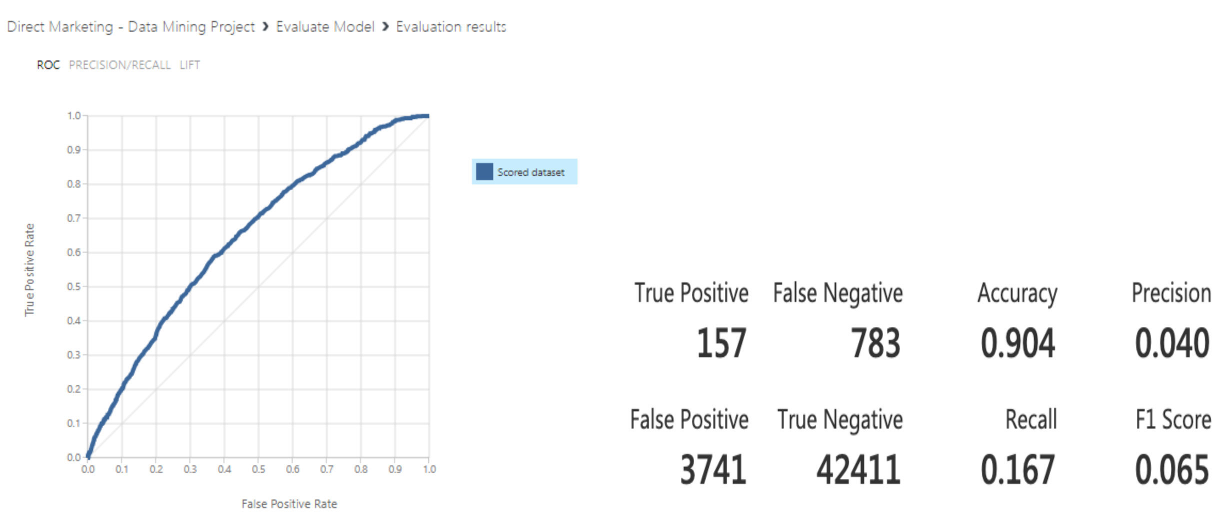


Figure 4: ROC and Confusion Matrix output of Decision Tree

Model 2: Two-Class Neural Network

This creates a binary classifier using a neural network algorithm that can be used to predict a target that has only two values. A neural network is a set of interconnected layers. The inputs are the first layer, and are connected to an output layer by an acyclic graph comprised of weighted edges and nodes. Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers. However, recent research has shown that deep neural networks (DNN) with many layers can be very effective in complex tasks such as image or speech recognition. The successive layers are used to model increasing levels of semantic depth.

The relationship between inputs and outputs is learned from training the neural network on the input data. The direction of the graph proceeds from the inputs through the hidden layer and to the output layer. All nodes in a layer are connected by the weighted edges to nodes in the next layer.

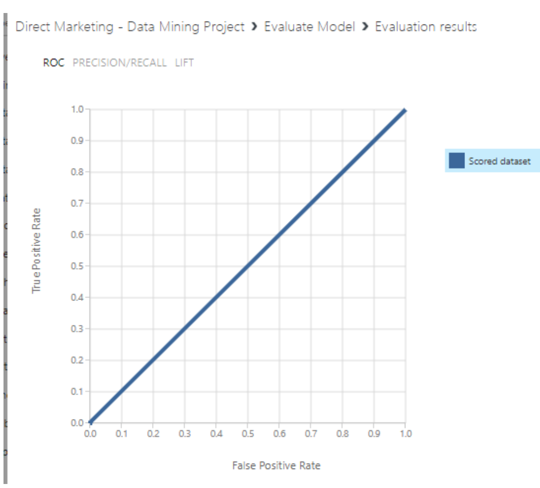
Below is the configuration of this project's Neural Network:

Number of hidden nodes: contains the number of hidden nodes. For our experiment, this value is set to 5.

Hidden layer specification: The output layer is fully connected to the hidden layer, and the hidden layer is fully connected to the input layer. For this experiment, this value is set to 'fully connected case'.

Number of learning iterations: specify the maximum number of times the algorithm should process the training cases. For this experiment, this value is 100.

Using the training dataset, our Two-Class Neural Network Model is trained and then this resultant model is then tested on the Testing dataset. Below is the evaluation result:



True Positive	False Negative	Accuracy
1	939	0.980
False Positive	**True Negative**	**Recall**
2	46150	0.001
Positive Label	**Negative Label**	
1	0	

Figure 5: ROC and Confusion Matrix output of Neural Network

Model Comparison:

S. No.	Model	Accuracy	Misclassification Rate	False Positive	False Negative
1	**Two-Class Boosted Decision Tree Classification**	**90.4%**	**9.6%**	**3741**	**783**
2	**Two-Class Neural Network**	**98%**	**2%**	**2**	**939**

Conclusion

In this project, two data mining techniques namely, boosted decision tree and neural networks model were built on direct marketing dataset in order to predict the success of the marketing campaign based on the marketing strategy and other attributes of the customers. Both the models are doing well.

Even though Neural Network model has a higher accuracy, it also has a higher false negative prediction when compared to the Decision Tree. The impact of falsely predicting a person as won't visit may result in loss of customer of the business which is very critical. Hence Boosted Decision tree is preferred.
The champion model suggests that the historical purchase pattern and the segment have significant effect on the success of the market campaign and should be considered while targeting customers in the campaign.

The model can be made better and more realistic by incorporating the cost of campaign per person and average profit gained per person into the model for maximum profit.

References

Y. Li, P. Murali, N. Shao and A. Sheopuri, "Applying Data Mining Techniques to Direct Marketing: Challenges and Solutions,

Suman, M & AnuRadha, T & Manasa Veena, K & , Greenfields. (2019). Direct Marketing with the Application of Data Mining.

https://gallery.azure.ai/Experiment/Binary-Classification-Direct-marketing-2

https://azuremlsampleexperiments.blob.core.windows.net/datasets/direct_marketing.csv

Appendix

The R code used to balance and pre-process the data is attached below.

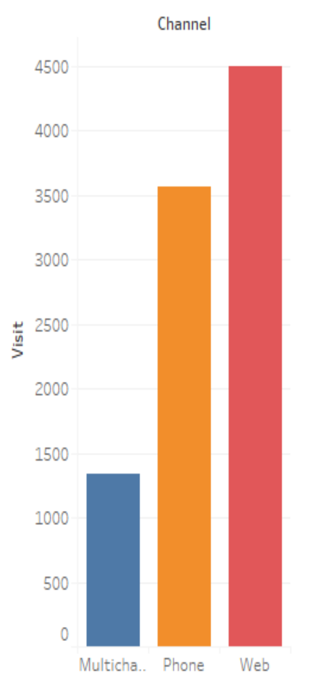
R Script

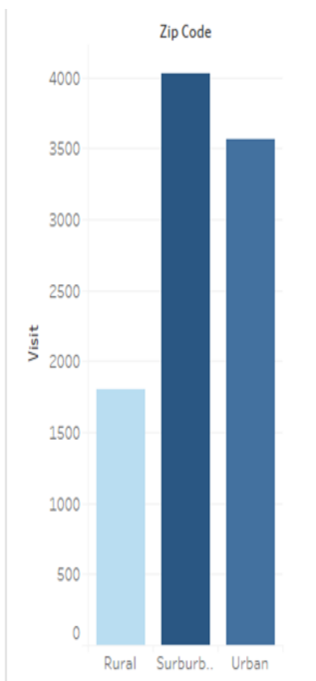
```
library(readr)

direct_marketing <- read_csv("direct_marketing.csv")

summary(direct_marketing)
length(direct_marketing$recency)

colnames(direct_marketing1)

# To remove history_segment,spend because they will not be used in the analysis.
direct_marketing1 <- direct_marketing[-2]
direct_marketing1 <- direct_marketing1[-11]

colnames(direct_marketing1)
# Doing feature engineering to handle mens and women column to make marital status
column
t <- direct_marketing1
t$marital_status <-
  ifelse(t$mens == 1 & t$womens ==1, "Married",
         ifelse(t$mens == 1 & t$womens ==0, "Single_Man",
                ifelse(t$mens == 0 & t$womens ==1, "Single_Women",0)))
summary(factor(t$marital_status))
direct_marketing1 <- t
colnames(direct_marketing1)

#To check the distribution of the target variable
nrow(direct_marketing1[ which(direct_marketing1$visit == 0 ),])
nrow(direct_marketing1[ which(direct_marketing1$visit == 1 ),])
# visit == 0 : 54606
# visit == 1 : 9394 this suggests the target variable is highly skewed.
hist(direct_marketing1$visit)

#Setting up the type of the variables
attach(direct_marketing1)
direct_marketing1$mens <- as.factor(direct_marketing1$mens)
direct_marketing1$zip_code <- as.factor(direct_marketing1$zip_code)
direct_marketing1$newbie <- as.factor(direct_marketing1$newbie)
direct_marketing1$channel <- as.factor(direct_marketing1$channel)
direct_marketing1$segment <- as.factor(direct_marketing1$segment)
direct_marketing1$visit <- as.factor(direct_marketing1$visit)
direct_marketing1$DM_category <- as.factor(direct_marketing1$DM_category)
direct_marketing1$marital_status <- as.factor(direct_marketing1$marital_status)
direct_marketing1$conversion <- as.factor(direct_marketing1$conversion)


#Handling unbalanced classification problem by Random Under-Sampling of the majority
class
visit1 = direct_marketing1[ which(direct_marketing1$visit == 1 ),]
visit0 = direct_marketing1[ which(direct_marketing1$visit == 0 ),]
set.seed(123)

sample_visit1 <- visit1[sample(1:nrow(visit1),floor(0.9*nrow(visit1)),replace=FALSE),]
```

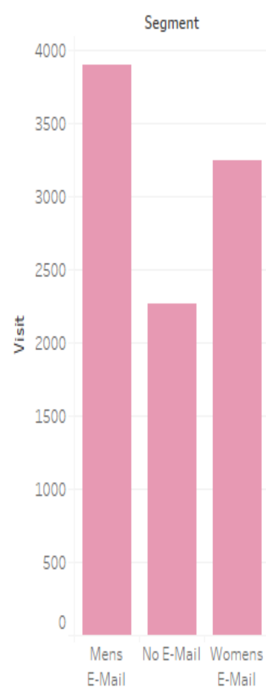
```
sample_visit0 <- visit0[sample(1:nrow(visit0),floor(0.9*nrow(visit1)),replace=FALSE),]
test_visit1 <- visit1[-
(sample(1:nrow(visit1),floor(0.9*nrow(visit1)),replace=FALSE)),]
test_visit0 <- visit0[-
(sample(1:nrow(visit0),floor(0.9*nrow(visit1)),replace=FALSE)),]
sample <- rbind(sample_visit1,sample_visit0)
test <- rbind(test_visit1,test_visit0)

write.csv(sample, file = "DM_Project_sample.csv", col.names = TRUE)
write.csv(test, file = "DM_Project_test.csv", col.names = TRUE)
```

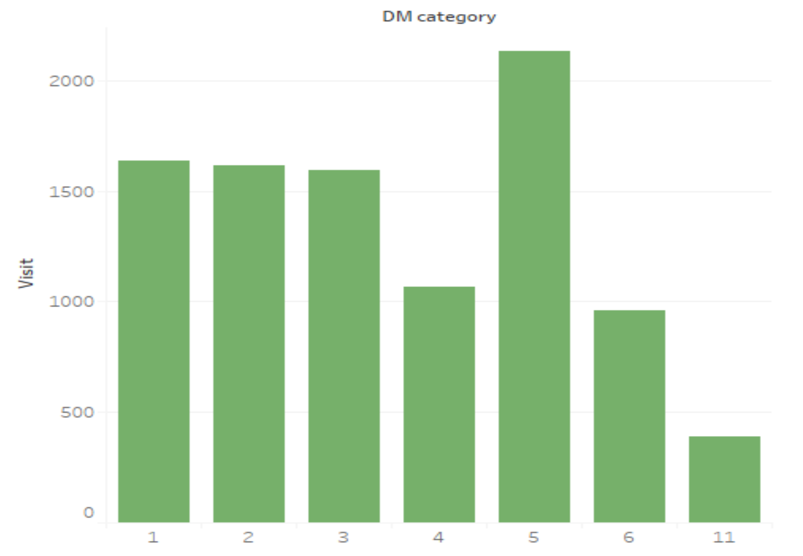
Data Visualization in Tableau

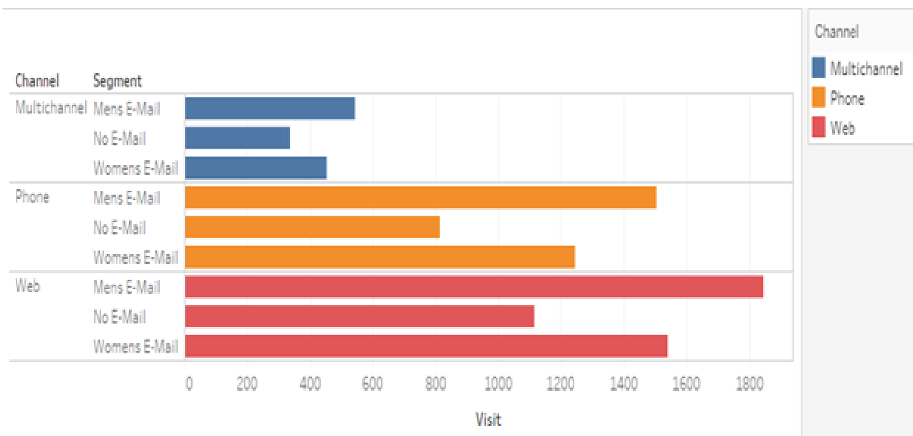
Below are the graphs for Channel, Zip Code and Segment Vs Visit respectively. Also, the graphs for DM category, Channel & segment together, and Recency Vs Visit are included.

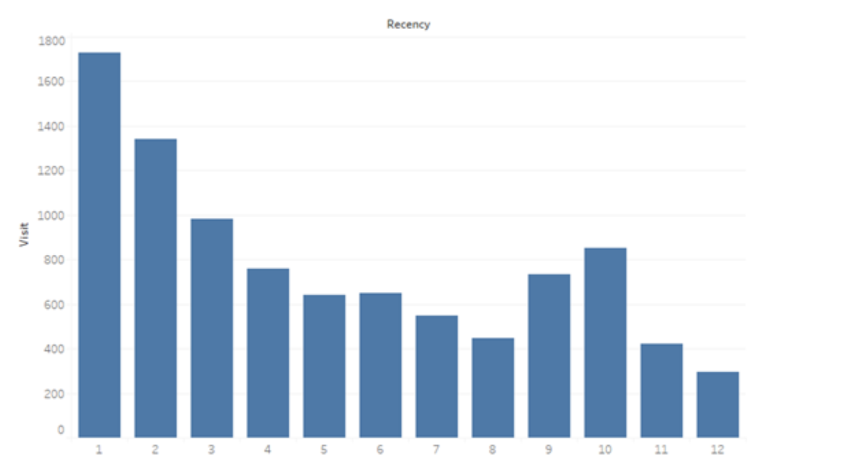







DM category
Visit
2000
1500
1000
500
0
1
2
3
4
5
6
11


Channel
Segment
Multichannel
Mens E-Mail
No E-Mail
Womens E-Mail
Phone
Mens E-Mail
No E-Mail
Womens E-Mail
Web
Mens E-Mail
No E-Mail
Womens E-Mail
0
200
400
600
800
1000
1200
1400
1600
1800
Visit
Channel
Multichannel
Phone
Web


Recency
Visit
1800
1600
1400
1200
1000
800
600
400
200
0
1
2
3
4
5
6
7
8
9
10
11
12