

Predicting Coupon Redemption

Anushree Tomar

28 September 2019

XYZ Credit Card company regularly helps it's merchants understand their data better and take key business decisions accurately by providing machine learning and analytics consulting. ABC is an established Brick & Mortar retailer that frequently conducts marketing campaigns for its diverse product range. As a merchant of XYZ, they have sought XYZ to assist them in their discount marketing process using the power of machine learning.

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer's propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign.

ABC's promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer's marketing team to accurately design coupon construct, and develop more precise and targeted marketing strategies.

Datasets

The data available in this problem contains the following information, including the details of a sample of campaigns and coupons used in previous campaigns -

- User Demographic Details
- Campaign and coupon Details
- Product details
- Previous transactions

Problem Statement

Based on previous transaction & performance data from the last 18 campaigns, predict the probability for the next 10 campaigns in the test set for each coupon and customer combination, whether the customer will redeem the coupon or not?

Data Exploration

List of all given datasets-

```
## [1] "campaign_data.csv"          "coupon_item_mapping.csv"
## [3] "customer_demographics.csv"  "customer_transaction_data.csv"
## [5] "item_data.csv"             "test_QyYwdj.csv"
## [7] "train.csv"
```

Analyse all datasets-

##	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
## rows	28	92663	760	1324566	74066	50226	78369
## columns	4	2	7	7	4	4	5
## discrete_columns	3	0	4	1	2	0	0
## continuous_columns	1	2	3	6	2	4	5

```
## all_missing_columns 0 0 0 0 0 0 0
## total_missing_values 0 0 0 0 0 0 0
## complete_rows 28 92663 760 1324566 74066 50226 78369
## total_observations 112 185326 5320 9271962 296264 200904 391845
## memory_usage 5944 742664 36824 58318400 1780664 805368 1569408
```

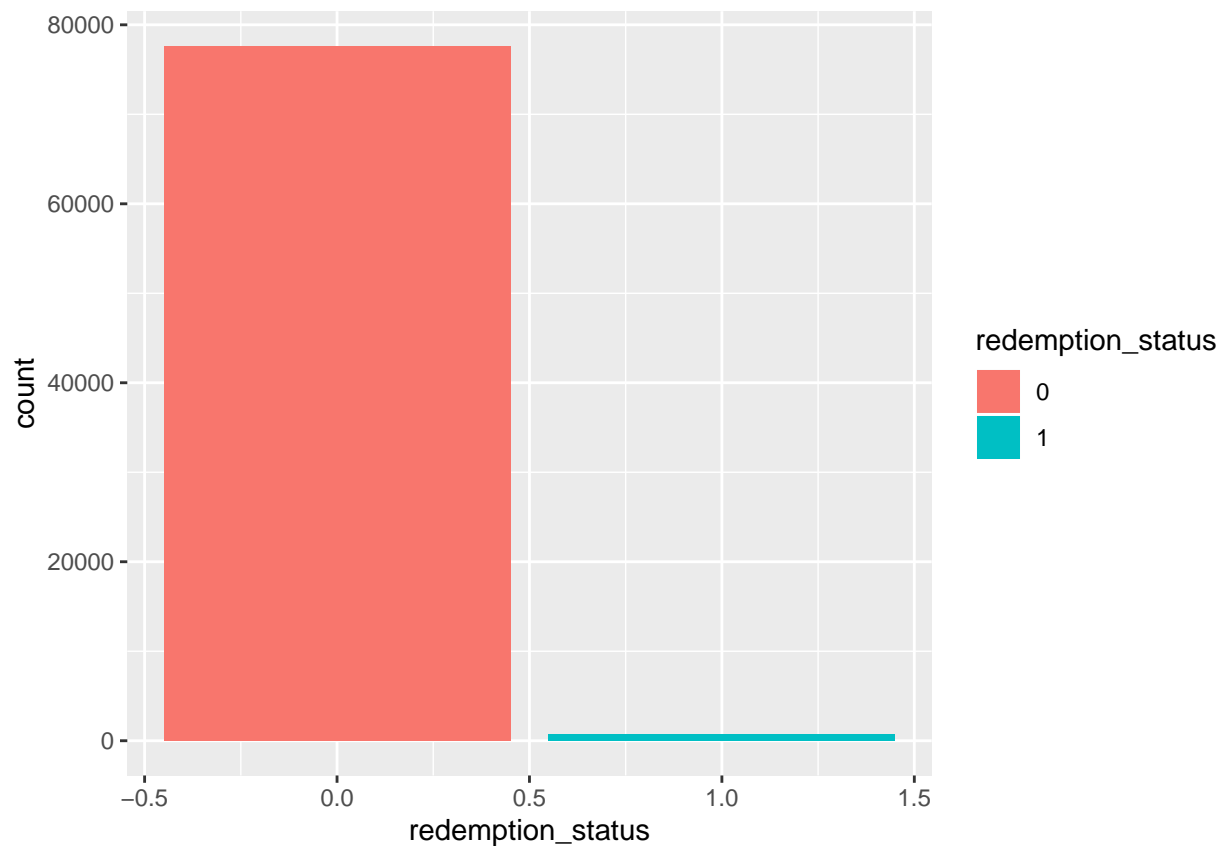
Observe all the data

Lets see the observations of the data.

```
## [1] "campaign_data.csv"          "coupon_item_mapping.csv"
## [3] "customer_demographics.csv"  "customer_transaction_data.csv"
## [5] "item_data.csv"             "test_QyjYwdj.csv"
## [7] "train.csv"

## [[1]]
##   campaign_id campaign_type start_date end_date
## 1:         24             Y  21/10/13  20/12/13
## 2:         25             Y  21/10/13  22/11/13
##
## [[2]]
##   coupon_id item_id
## 1:       105      37
## 2:       107      75
##
## [[3]]
##   customer_id age_range marital_status rented family_size no_of_children
## 1:          1      70+      Married      0          2
## 2:          6     46-55      Married      0          2
##   income_bracket
## 1:              4
## 2:              5
##
## [[4]]
##   date customer_id item_id quantity selling_price other_discount
## 1: 2012-01-02      1501   26830        1       35.26       -10.69
## 2: 2012-01-02      1501   54253        1       53.43       -13.89
##   coupon_discount
## 1:              0
## 2:              0
##
## [[5]]
##   item_id brand brand_type category
## 1:       1     1 Established   Grocery
## 2:       2     1 Established Miscellaneous
##
## [[6]]
##   id campaign_id coupon_id customer_id
## 1:  3          22      869          967
## 2:  4          20      389         1566
##
## [[7]]
##   id campaign_id coupon_id customer_id redemption_status
## 1:  1          13       27         1053              0
## 2:  2          13       116          48              0
```

Examine the Redemption Status



The Proportion of redemption status is unbalanced so we need to balance the status before creating any prediction model.

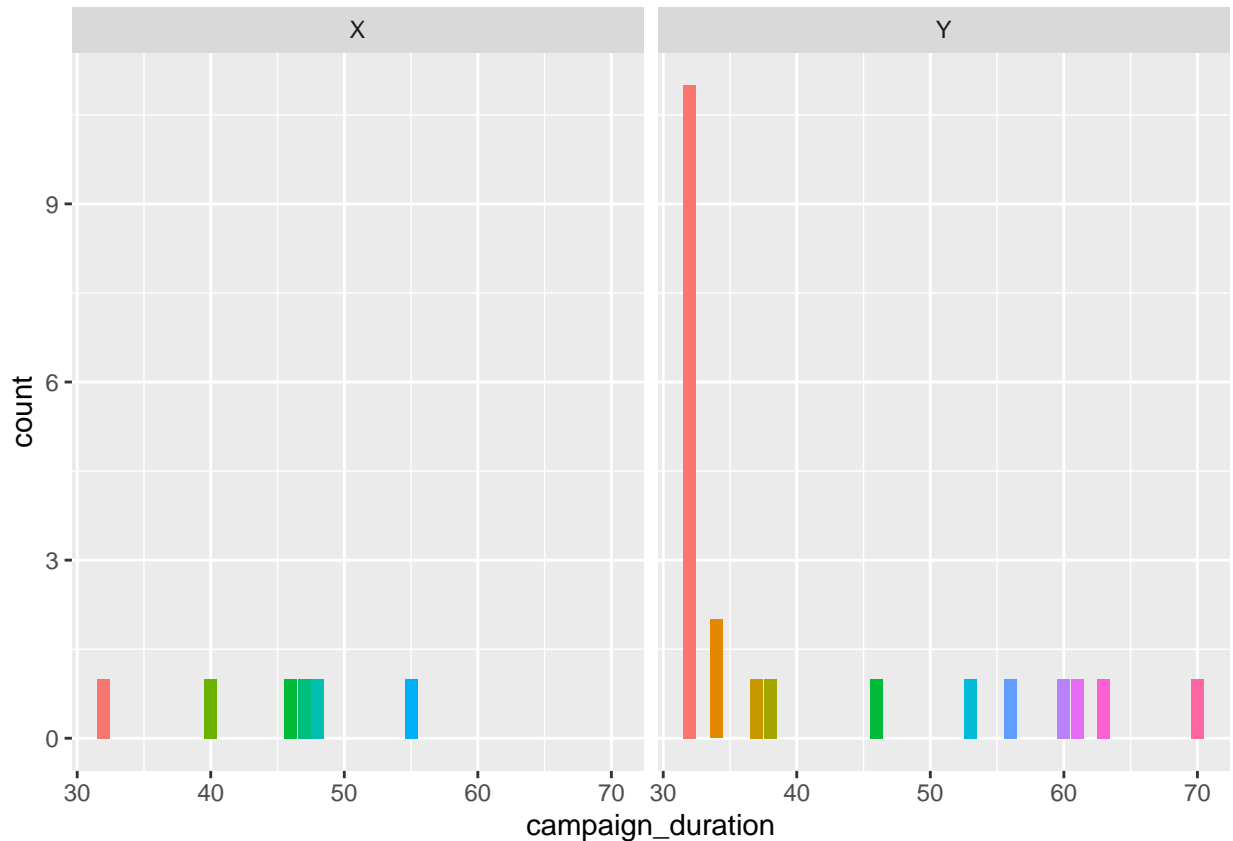
Campaign Data

Calculate the duration of campaign

```
##      campaign_id campaign_type campaign_duration
## 1:           24             Y           60 days
## 2:           25             Y           32 days
## 3:           20             Y           70 days
## 4:           23             Y           38 days
## 5:           21             Y           32 days
## 6:           22             X           32 days
```

Analyse campaign type

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



As from the above graph we can see that more duration spend on the campaign type Y

Combine train data and campaign data

In the train data we merged the campaign data by “campaign_id”

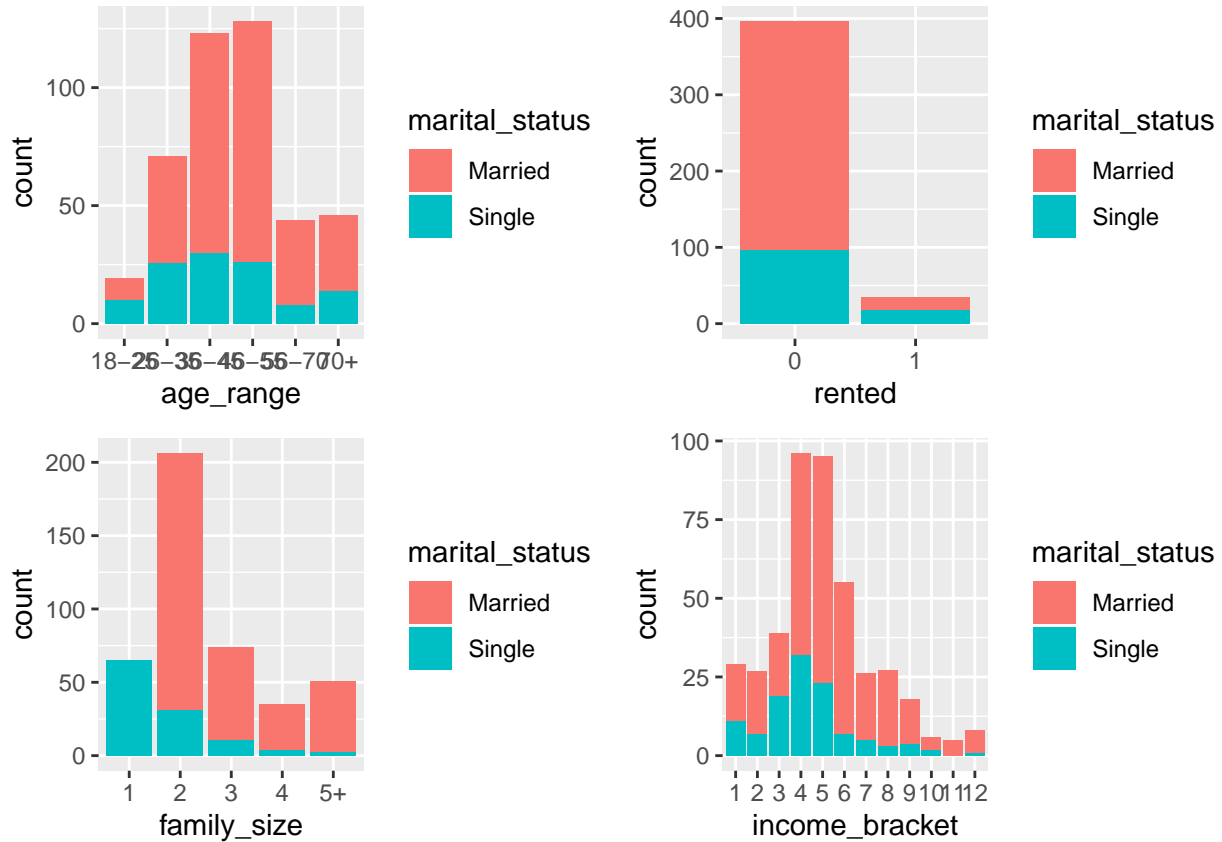
Customer demographics data

```
##   customer_id age_range marital_status rented family_size no_of_children
## 1           1      70+      Married      0         2          <NA>
## 2           6    46-55      Married      0         2          <NA>
## 3           7    26-35        <NA>      0         3           1
## 4           8    26-35        <NA>      0         4           2
## 5          10    46-55      Single      0         1          <NA>
## 6          11      70+      Single      0         2          <NA>
##   income_bracket
## 1              4
## 2              5
## 3              3
## 4              6
## 5              5
## 6              1

##   customer_id  age_range marital_status  rented  family_size
##           0         0         329         0         0
## no_of_children income_bracket
##           538         0
```

As we can see that some customer demographics data as *marital_status* and *no_of_children* is not available so we need to fill the missing values.

Visualise Customer demographics



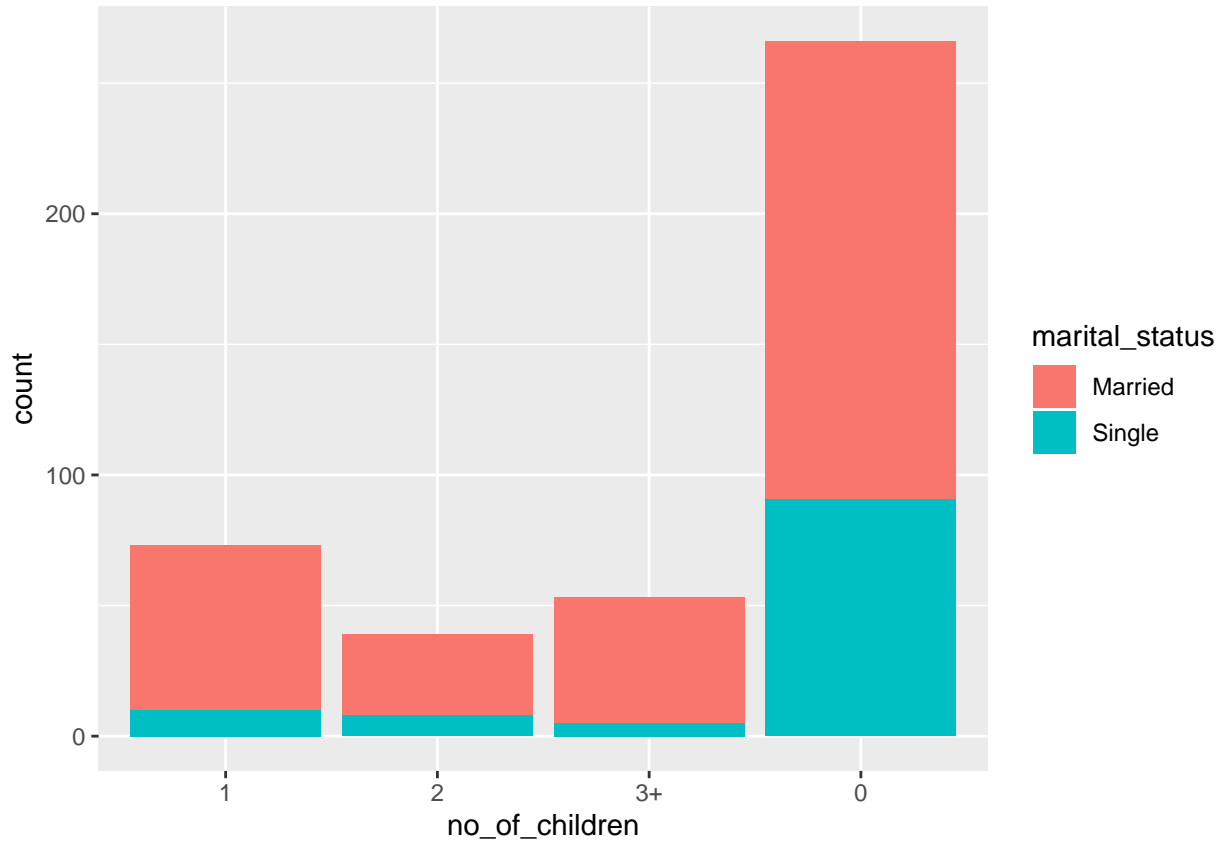
From the above graph we can see that most of the customers are married and having age range 46-55, 36-45, family size of 2 with income bracket of 4, 5 and not rented.

- Impute *no_of_children*

```
##
##  1  2  3+
## 107 55 60
```

There are 3 levels in the *no_of_children* 1, 2, and 3+ so we will fill the missing value with "0"

```
##
##  1  2  3+  0
## 107 55 60 538
```



From the above graph we can see that most of the married customer have children but most of single customer not having children.

- Impute *marital_status*

To impute *marital_status* we need to focus on *family_size* and *no_of_children*. most of the single customers have *family size*- 1 and 0 -*no_of_children*

```
##      customer_id age_range marital_status rented family_size
## [1,]          1         6              1      1           2
## [2,]          6         4              1      1           2
## [3,]          7         2              1      1           3
## [4,]          8         2              1      1           4
## [5,]         10         4              2      1           1
## [6,]         11         6              2      1           2
##      no_of_children income_bracket
## [1,]              4              4
## [2,]              4              5
## [3,]              1              3
## [4,]              2              6
## [5,]              4              5
## [6,]              4              1
```

Item data

```
##   item_id brand brand_type category
## 1      1     1 Established  Grocery
## 2      2     1 Established Miscellaneous
```

```
## 3      3      56      Local      Bakery
## 4      4      56      Local      Grocery
## 5      5      56      Local      Grocery
## 6      6      56      Local      Grocery
```

Coupon Item mapping data

```
##  coupon_id item_id
## 1      105      37
## 2      107      75
## 3      494      76
## 4      522      77
## 5      518      77
## 6      520      77

##  item_id coupon_id brand  brand_type      category
## 1      1      22      1 Established      Grocery
## 2      1      31      1 Established      Grocery
## 3      4      327     56      Local      Grocery
## 4      4      82      56      Local      Grocery
## 5      4      166     56      Local      Grocery
## 6      7      23      56      Local Pharmaceutical
```

In the coupon_item_mapping data we added the information of items by item_id

brand type

```
##
## Established      Local
##      78759      13904
```

Count of brand type by coupon id

```
##  coupon_id Established Local
## 1      1      39      0
## 2      2      2      0
## 3      3      17      0
## 4      4      24      0
## 5      5      7      0
## 6      6     3075     638

##  coupon_id Established      Local
##      623286      78759      13904
```

Count of Category by coupon id

```
##  coupon_id Bakery Dairy, Juices & Snacks Flowers & Plants Garden Grocery
## 1      1      0      0      0      0      9
## 2      2      0      0      0      0      2
## 3      3      0      0      0      0      17
## 4      4      0      0      0      0      24
## 5      5      0      0      0      0      0
## 6      6      0      259      0      0      3082
##  Meat Miscellaneous Natural Products Packaged Meat Pharmaceutical
## 1      0      0      30      0      0
## 2      0      0      0      0      0
```

```

## 3      0          0          0          0          0
## 4      0          0          0          0          0
## 5      0          0          0          0          7
## 6      2          0          0          242         0
##   Prepared Food Restauarant Salads Seafood Skin & Hair Care Travel
## 1          0          0          0          0          0
## 2          0          0          0          0          0
## 3          0          0          0          0          0
## 4          0          0          0          0          0
## 5          0          0          0          0          0
## 6          0          0          0         128          0
##   Vegetables (cut)
## 1          0
## 2          0
## 3          0
## 4          0
## 5          0
## 6          0

##           coupon_id           Bakery Dairy, Juices & Snacks
##           623286           100           1867
##   Flowers & Plants           Garden           Grocery
##           1963           286           36466
##           Meat           Miscellaneous           Natural Products
##           6218           184           6819
##   Packaged Meat           Pharmaceutical           Prepared Food
##           6144           25061           240
##           Restauarant           Salads           Seafood
##           1           100           2227
##   Skin & Hair Care           Travel           Vegetables (cut)
##           4924           44           19

```

Customer Transaction data

```

##           date customer_id item_id quantity selling_price other_discount
## 1 2012-01-02      1501    26830         1        35.26       -10.69
## 2 2012-01-02      1501    54253         1        53.43       -13.89
## 3 2012-01-02      1501    31962         1       106.50       -14.25
## 4 2012-01-02      1501    33647         1        67.32         0.00
## 5 2012-01-02      1501    48199         1        71.24       -28.14
## 6 2012-01-02      1501    57397         1        71.24       -28.14
##   coupon_discount
## 1          0
## 2          0
## 3          0
## 4          0
## 5          0
## 6          0

```

From the above data we will analyse customers purchasing pattern by RFM analysis then will calculate Final price after discount.

Summarize Final traindata

```

## redemption_status campaign_duration      X      Y

```



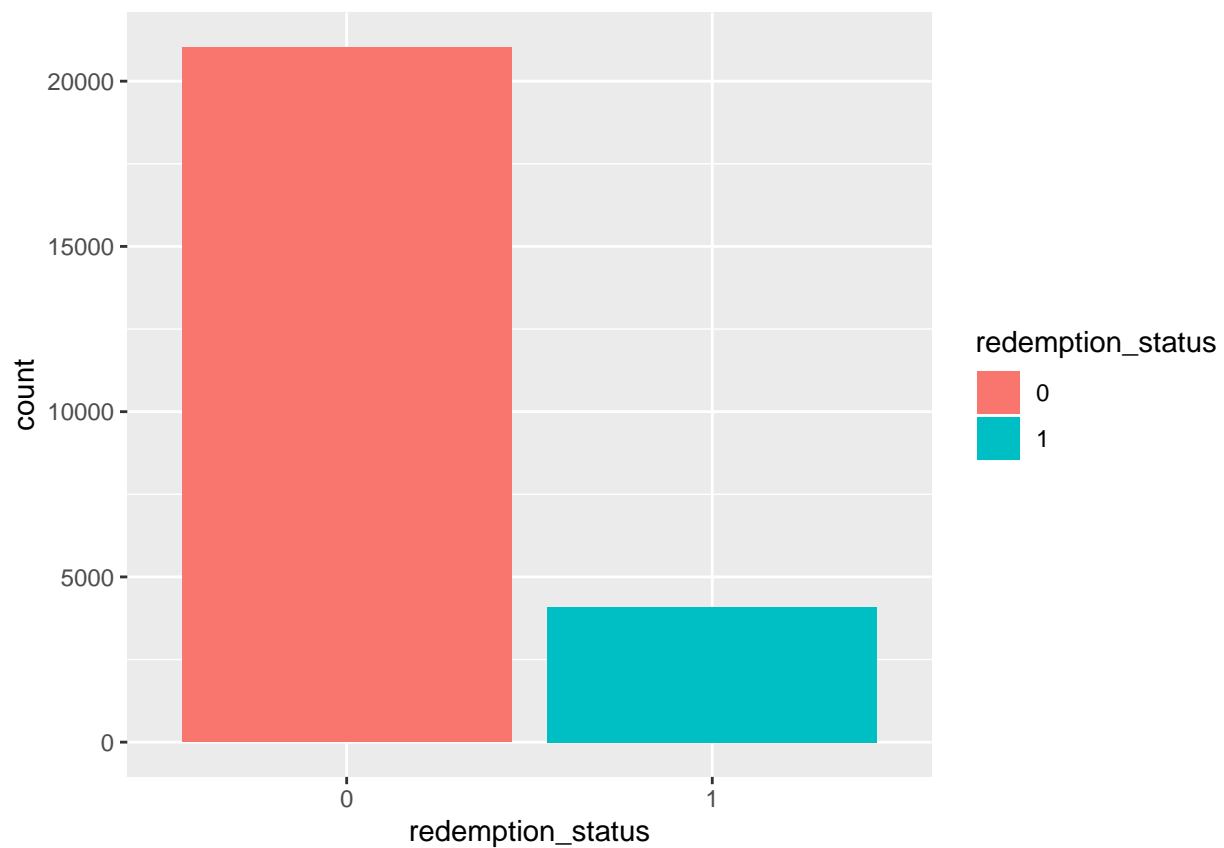
```

## 0:77640      Min.   :32.00      Min.   :0.0000      Min.   :0.0000
## 1: 729        1st Qu.:46.00      1st Qu.:0.0000      1st Qu.:0.0000
##              Median :47.00      Median :1.0000      Median :0.0000
##              Mean   :44.77      Mean   :0.7361      Mean   :0.2639
##              3rd Qu.:48.00      3rd Qu.:1.0000      3rd Qu.:1.0000
##              Max.   :63.00      Max.   :1.0000      Max.   :1.0000
##
## age_range    marital_status rented   family_size no_of_children
## 18-25: 2456   Married:61992   0:75864   1 :13612   1 : 6540
## 26-35: 8024   Single :16377   1: 2505   2 :51596   2 : 3547
## 36-45:11210                                     3 : 6267   3+: 3969
## 46-55:49862                                     4 : 3139   0 :64313
## 56-70: 3723                                     5+: 3755
## 70+ : 3094
##
## income_bracket Established      Local      Bakery
## 5 :45745   Min.   : 0   Min.   : 0.00   Min.   : 0.00000
## 4 : 8753   1st Qu.: 2   1st Qu.: 0.00   1st Qu.: 0.00000
## 6 : 4855   Median : 8   Median : 0.00   Median : 0.00000
## 2 : 3892   Mean   : 70   Mean   : 11.93   Mean   : 0.09081
## 3 : 3882   3rd Qu.: 27   3rd Qu.: 0.00   3rd Qu.: 0.00000
## 1 : 3456   Max.   :10744   Max.   :1171.00   Max.   :16.00000
## (Other): 7786
## Dairy, Juices & Snacks Flowers & Plants      Garden
## Min.   : 0.000      Min.   : 0.000      Min.   : 0.0000
## 1st Qu.: 0.000      1st Qu.: 0.000      1st Qu.: 0.0000
## Median : 0.000      Median : 0.000      Median : 0.0000
## Mean   : 1.868      Mean   : 1.708      Mean   : 0.2483
## 3rd Qu.: 0.000      3rd Qu.: 0.000      3rd Qu.: 0.0000
## Max.   :260.000      Max.   :652.000      Max.   :95.0000
##
## Grocery      Meat      Miscellaneous      Natural Products
## Min.   : 0.00   Min.   : 0.000      Min.   : 0.0000      Min.   : 0.000
## 1st Qu.: 0.00   1st Qu.: 0.000      1st Qu.: 0.0000      1st Qu.: 0.000
## Median : 6.00   Median : 0.000      Median : 0.0000      Median : 0.000
## Mean   : 34.38   Mean   : 5.869      Mean   : 0.1612      Mean   : 5.602
## 3rd Qu.: 18.00   3rd Qu.: 0.000      3rd Qu.: 0.0000      3rd Qu.: 0.000
## Max.   :4127.00   Max.   :2074.000      Max.   :37.0000      Max.   :2261.000
##
## Packaged Meat      Pharmaceutical      Prepared Food      Salads
## Min.   : 0.000      Min.   : 0.00   Min.   : 0.0000      Min.   : 0.00000
## 1st Qu.: 0.000      1st Qu.: 0.00   1st Qu.: 0.0000      1st Qu.: 0.00000
## Median : 0.000      Median : 0.00   Median : 0.0000      Median : 0.00000
## Mean   : 5.904      Mean   : 20.38   Mean   : 0.2327      Mean   : 0.07648
## 3rd Qu.: 0.000      3rd Qu.: 0.00   3rd Qu.: 0.0000      3rd Qu.: 0.00000
## Max.   :1676.000      Max.   :9616.00   Max.   :60.0000      Max.   :40.00000
##
## Seafood      Skin & Hair Care      Travel      Vegetables (cut)
## Min.   : 0.000      Min.   : 0.000      Min.   : 0.00000      Min.   :0.00000
## 1st Qu.: 0.000      1st Qu.: 0.000      1st Qu.: 0.00000      1st Qu.:0.00000
## Median : 0.000      Median : 0.000      Median : 0.00000      Median :0.00000
## Mean   : 1.924      Mean   : 3.431      Mean   : 0.03513      Mean   :0.01714
## 3rd Qu.: 0.000      3rd Qu.: 0.000      3rd Qu.: 0.00000      3rd Qu.:0.00000
## Max.   :590.000      Max.   :2170.000      Max.   :17.00000      Max.   :7.00000

```

```
##
##      totquant      totmonetary      Freqcust      recency
## Min.   :    88   Min.   : 3466   Min.   : 10.0   Min.   : 0.000
## 1st Qu.: 1593   1st Qu.: 48244   1st Qu.: 66.0   1st Qu.: 1.000
## Median : 31641   Median : 80766   Median : 91.0   Median : 3.000
## Mean   : 141279   Mean   :100069   Mean   :105.9   Mean   : 8.572
## 3rd Qu.: 184709   3rd Qu.:130575   3rd Qu.:135.0   3rd Qu.: 7.000
## Max.   :2617544   Max.   :529971   Max.   :386.0   Max.   :429.000
##
```

Balance the redemption status



Encode the categorical features of the data

```
##      redemption_status campaign_duration X Y age_range marital_status rented
## 1:                    0                48 1 0         4             0       0
## 2:                    0                32 0 1         3             0       0
## 3:                    0                48 1 0         3             0       0
## 4:                    0                32 0 1         1             0       0
## 5:                    0                48 1 0         2             1       1
## 6:                    0                61 0 1         3             0       0
##      family_size no_of_children income_bracket Established Local Bakery
## 1:              1              3              4           25      0      0
## 2:              1              3              4           18      0      0
## 3:              1              3              4           10      0      0
## 4:              2              0              2           12      0      0
```

## 5:	0	3	3	146	0	0
## 6:	1	3	4	9	0	0
##	Dairy, Juices & Snacks Flowers & Plants Garden Grocery Meat					
## 1:		0	0	0	25	0
## 2:		0	0	0	18	0
## 3:		0	0	0	0	0
## 4:		0	0	0	0	0
## 5:		0	0	0	144	0
## 6:		0	0	0	7	0
##	Miscellaneous Natural Products Packaged Meat Pharmaceutical					
## 1:	0		0			0
## 2:	0		0			0
## 3:	0		0			10
## 4:	0		0			12
## 5:	0		0			2
## 6:	0		0			2
##	Prepared Food Salads Seafood Skin & Hair Care Travel Vegetables (cut)					
## 1:	0	0	0		0	0
## 2:	0	0	0		0	0
## 3:	0	0	0		0	0
## 4:	0	0	0		0	0
## 5:	0	0	0		0	0
## 6:	0	0	0		0	0
##	totquant	totmonetary	Freqcust	recency		
## 1:	136261	109056.7	234	1		
## 2:	1732	111192.9	96	12		
## 3:	131204	119432.9	198	2		
## 4:	12346	210969.8	47	5		
## 5:	60494	164673.1	60	4		
## 6:	9720	140800.3	70	129		

Normalize the data

##	redemption_status	campaign_duration	X	Y
##	Min. :0.000000	Min. :0.0000	Min. :0.000	Min. :0.000
##	1st Qu.:0.000000	1st Qu.:0.4516	1st Qu.:0.000	1st Qu.:0.000
##	Median :0.000000	Median :0.4839	Median :1.000	Median :0.000
##	Mean :0.009252	Mean :0.4132	Mean :0.733	Mean :0.267
##	3rd Qu.:0.000000	3rd Qu.:0.5161	3rd Qu.:1.000	3rd Qu.:1.000
##	Max. :1.000000	Max. :1.0000	Max. :1.000	Max. :1.000
##	age_range	marital_status	rented	family_size
##	Min. :0.0000	Min. :0.0000	Min. :0.00000	Min. :0.0000
##	1st Qu.:0.4000	1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:0.2500
##	Median :0.6000	Median :0.0000	Median :0.00000	Median :0.2500
##	Mean :0.5349	Mean :0.2117	Mean :0.03248	Mean :0.2839
##	3rd Qu.:0.6000	3rd Qu.:0.0000	3rd Qu.:0.00000	3rd Qu.:0.2500
##	Max. :1.0000	Max. :1.0000	Max. :1.00000	Max. :1.0000
##	no_of_children	income_bracket	Established	Local
##	Min. :0.000	Min. :0.0000	Min. :0.0000000	Min. :0.000000
##	1st Qu.:1.000	1st Qu.:0.2727	1st Qu.:0.0001862	1st Qu.:0.000000
##	Median :1.000	Median :0.3636	Median :0.0007446	Median :0.000000
##	Mean :0.867	Mean :0.3536	Mean :0.0064617	Mean :0.009991
##	3rd Qu.:1.000	3rd Qu.:0.3636	3rd Qu.:0.0025130	3rd Qu.:0.000000
##	Max. :1.000	Max. :1.0000	Max. :1.0000000	Max. :1.000000

## Bakery	Dairy...Juices...Snacks	Flowers...Plants	
## Min. :0.000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.000000	Median :0.000000	Median :0.000000	
## Mean :0.005635	Mean :0.006822	Mean :0.002349	
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.000000	
## Max. :1.000000	Max. :1.000000	Max. :1.000000	
## Garden	Grocery	Meat	
## Min. :0.000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.000000	Median :0.001454	Median :0.000000	
## Mean :0.002347	Mean :0.007962	Mean :0.003551	
## 3rd Qu.:0.000000	3rd Qu.:0.004361	3rd Qu.:0.000000	
## Max. :1.000000	Max. :1.000000	Max. :1.000000	
## Miscellaneous	Natural.Products	Packaged.Meat	
## Min. :0.000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.000000	Median :0.000000	Median :0.000000	
## Mean :0.004384	Mean :0.002216	Mean :0.004229	
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.000000	
## Max. :1.000000	Max. :1.000000	Max. :1.000000	
## Pharmaceutical	Prepared.Food	Salads	
## Min. :0.000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.000000	Median :0.000000	Median :0.000000	
## Mean :0.002008	Mean :0.004214	Mean :0.00167	
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.000000	
## Max. :1.000000	Max. :1.000000	Max. :1.000000	
## Seafood	Skin...Hair.Care	Travel	
## Min. :0.000000	Min. :0.000000	Min. :0.000000	
## 1st Qu.:0.000000	1st Qu.:0.000000	1st Qu.:0.000000	
## Median :0.000000	Median :0.000000	Median :0.000000	
## Mean :0.003621	Mean :0.001479	Mean :0.001809	
## 3rd Qu.:0.000000	3rd Qu.:0.000000	3rd Qu.:0.000000	
## Max. :1.000000	Max. :1.000000	Max. :1.000000	
## Vegetables..cut.	totquant	totmonetary	Freqcust
## Min. :0.000000	Min. :0.0000000	Min. :0.00000	Min. :0.0000
## 1st Qu.:0.000000	1st Qu.:0.0005765	1st Qu.:0.08547	1st Qu.:0.1489
## Median :0.000000	Median :0.0123058	Median :0.14791	Median :0.2181
## Mean :0.002215	Mean :0.0530307	Mean :0.18389	Mean :0.2557
## 3rd Qu.:0.000000	3rd Qu.:0.0701849	3rd Qu.:0.24488	3rd Qu.:0.3351
## Max. :1.000000	Max. :1.0000000	Max. :1.00000	Max. :1.0000
## recency			
## Min. :0.000000			
## 1st Qu.:0.002331			
## Median :0.006993			
## Mean :0.020056			
## 3rd Qu.:0.016317			
## Max. :1.000000			

Predictive Model and Evaluation

XGBoost Model

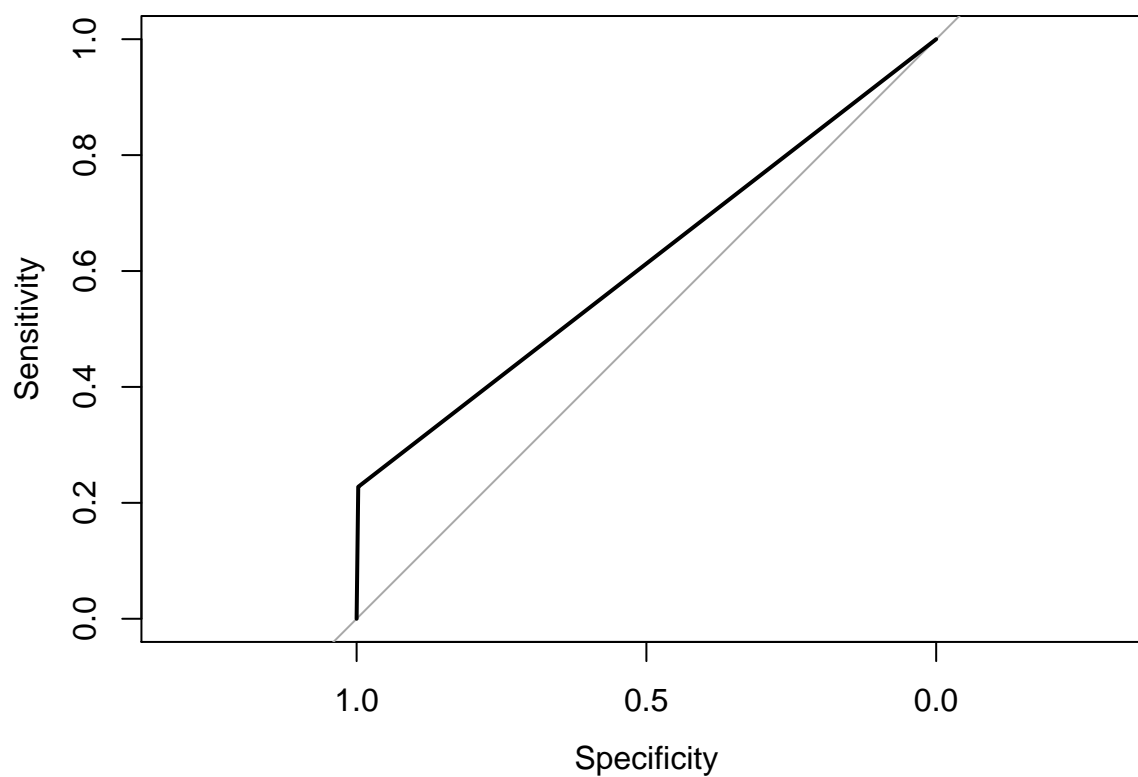
Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##           0 15485    43
##           1   112    33
##
##           Accuracy : 0.9901
##           95% CI : (0.9884, 0.9916)
##       No Information Rate : 0.9952
##       P-Value [Acc > NIR] : 1
##
##           Kappa : 0.2942
##
## Mcnemar's Test P-Value : 4.711e-08
##
##           Sensitivity : 0.434211
##           Specificity : 0.992819
##           Pos Pred Value : 0.227586
##           Neg Pred Value : 0.997231
##           Precision : 0.227586
##           Recall : 0.434211
##           F1 : 0.298643
##           Prevalence : 0.004849
##           Detection Rate : 0.002106
##       Detection Prevalence : 0.009252
##       Balanced Accuracy : 0.713515
##
##           'Positive' Class : 1
##
```

AUC

```
## Area under the curve: 0.6124
```

Plot ROC curve



Neural Network model

AUC

Area under the curve: 0.5785

Confusion Matrix

Confusion Matrix and Statistics

##

Reference

Prediction 0 1

0 15073 455

1 118 27

##

Accuracy : 0.9634

95% CI : (0.9604, 0.9663)

No Information Rate : 0.9692

P-Value [Acc > NIR] : 1

##

Kappa : 0.0729

##

McNemar's Test P-Value : <2e-16

##

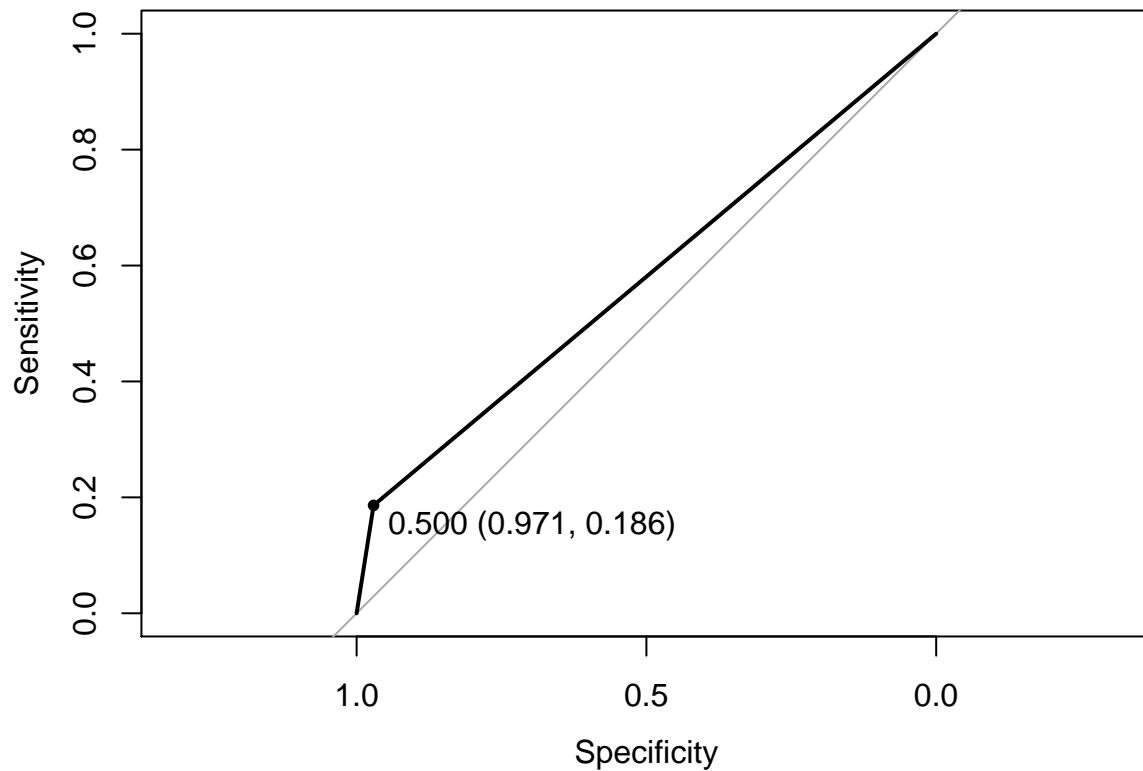
Sensitivity : 0.056017

```

##          Specificity : 0.992232
##      Pos Pred Value : 0.186207
##      Neg Pred Value : 0.970698
##          Precision : 0.186207
##          Recall    : 0.056017
##              F1    : 0.086124
##      Prevalence    : 0.030754
##      Detection Rate : 0.001723
##      Detection Prevalence : 0.009252
##      Balanced Accuracy : 0.524124
##
##      'Positive' Class : 1
##

```

ROC curve



Random Forest Model

AUC

```
## Area under the curve: 0.9304
```

Confusion Matrix

```

## Confusion Matrix and Statistics
##
##      Reference

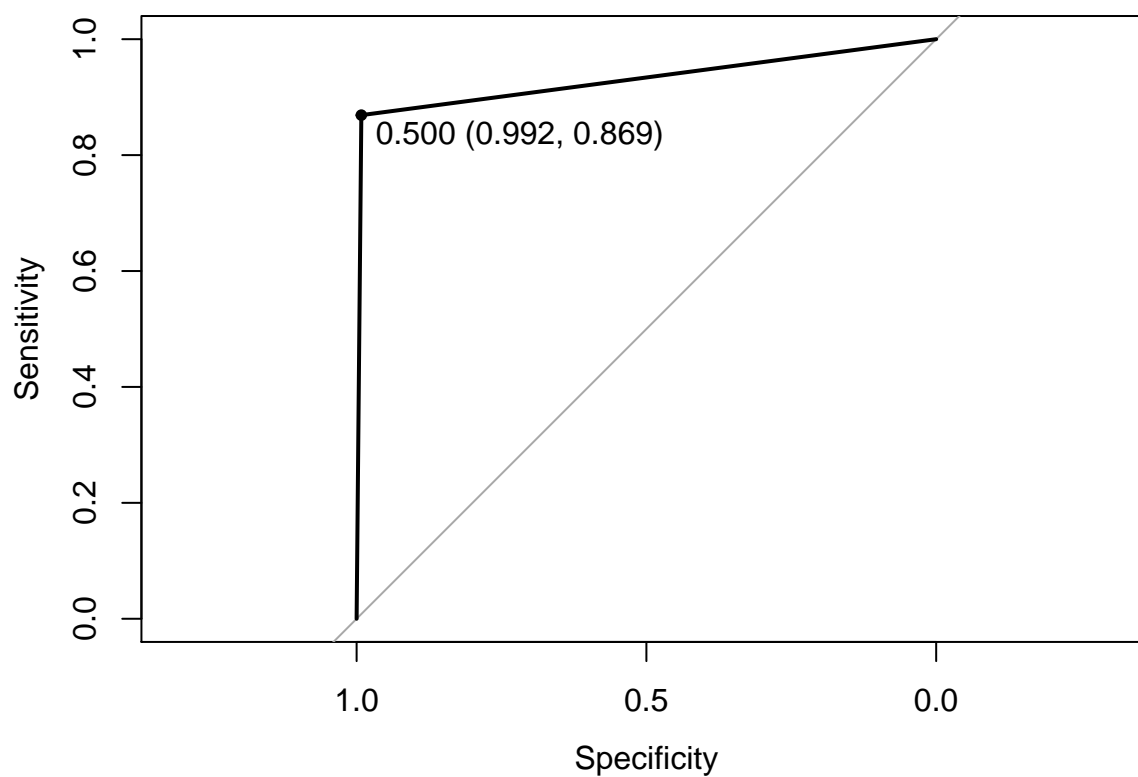
```

```

## Prediction      0      1
##              0 15402  126
##              1   19   126
##
##              Accuracy : 0.9907
##              95% CI : (0.9891, 0.9922)
##              No Information Rate : 0.9839
##              P-Value [Acc > NIR] : 1.146e-13
##
##              Kappa : 0.6304
##
## Mcnemar's Test P-Value : < 2.2e-16
##
##              Sensitivity : 0.500000
##              Specificity : 0.998768
##              Pos Pred Value : 0.868966
##              Neg Pred Value : 0.991886
##              Precision : 0.868966
##              Recall : 0.500000
##              F1 : 0.634761
##              Prevalence : 0.016079
##              Detection Rate : 0.008039
##              Detection Prevalence : 0.009252
##              Balanced Accuracy : 0.749384
##
##              'Positive' Class : 1
##

```


ROC curve



XGBoost Model

Train the Model without normalization

Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##           0 15481  47
##           1   118  27
##
##           Accuracy : 0.9895
##           95% CI : (0.9877, 0.991)
##           No Information Rate : 0.9953
##           P-Value [Acc > NIR] : 1
##
##           Kappa : 0.2418
##
##           McNemar's Test P-Value : 5.051e-08
##
##           Sensitivity : 0.364865
##           Specificity : 0.992435
##           Pos Pred Value : 0.186207
```

```

##          Neg Pred Value : 0.996973
##          Precision : 0.186207
##          Recall : 0.364865
##          F1 : 0.246575
##          Prevalence : 0.004721
##          Detection Rate : 0.001723
##          Detection Prevalence : 0.009252
##          Balanced Accuracy : 0.678650
##
##          'Positive' Class : 1
##

```

AUC

```
## Area under the curve: 0.9304
```

ROC curve

