Machine Learning Project

```
# Suppress dplyr summarise grouping warning messages
options(dplyr.summarise.inform = FALSE)

library(tidyverse)
library(tidymodels)
library(discrim)
library(vip)
library(rpart.plot)
library(recipes)
library(parsnip)
credit_card_df <- readRDS(url('https://gmubusinessanalytics.netlify.app/data/credit_card
_df.rds'))
```

Data Analysis

In this section, you must think of at least 5 relevant questions that explore the relationship between `customer_status` and the other variables in the `credit_card_df` data set. The goal of your analysis should be discovering which variables drive the differences between customers who do and do not close their account.
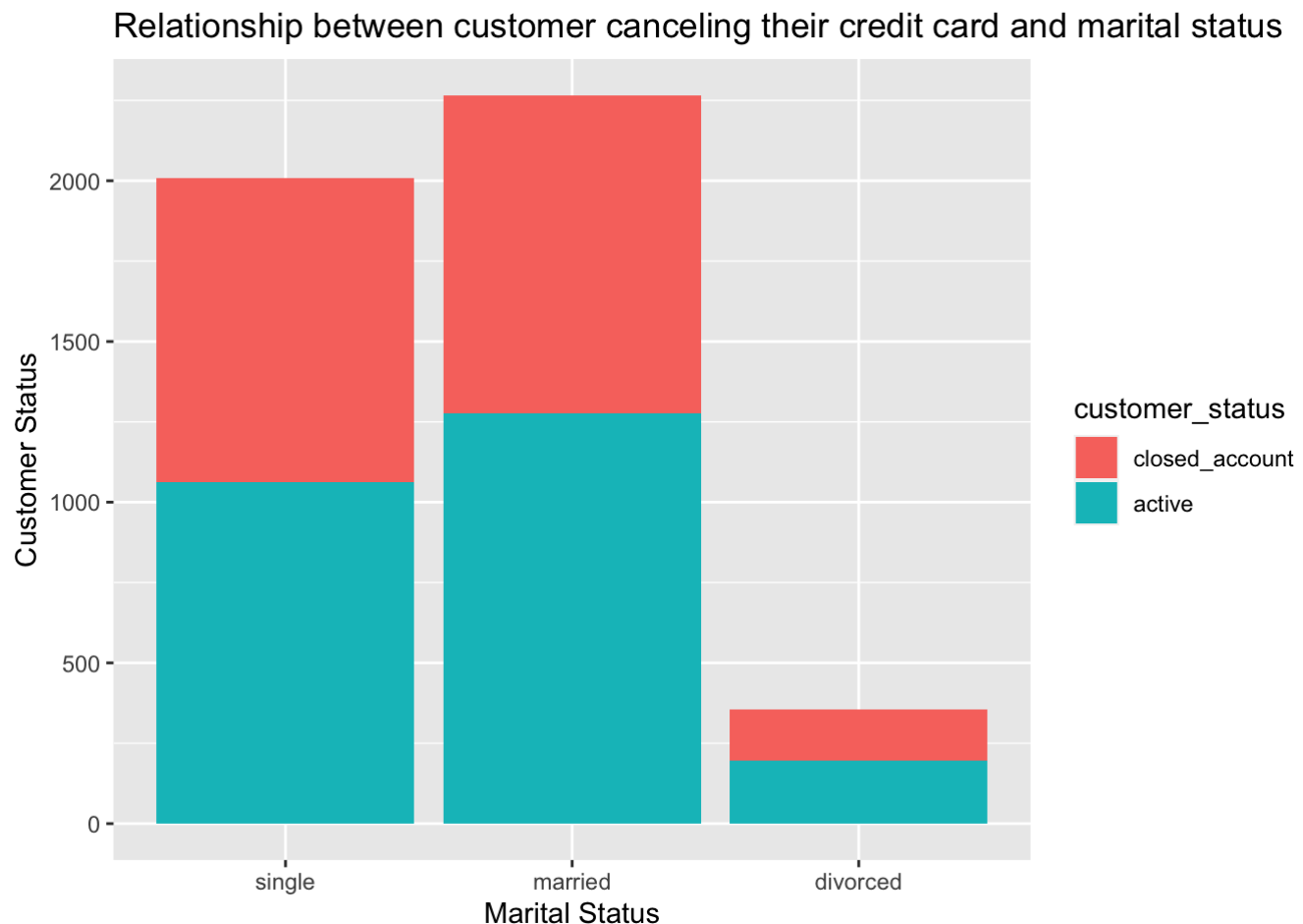
Question 1

Question: Is there a relationship between customer canceling their credit card and marital status?

Answer: Yes there is a relationship between the credit card canceling and marital status. At most half of the single people closed there account. 47% of the single closed there account, where as 43.6% of the married people closed there account. 44.9% of the divorced people closed account.

```
credit_card_df %>%
  group_by(marital_status,customer_status) %>%
  summarise(count = n()) %>%
  mutate(perc = 100*count/sum(count))
```

marital_status <fct>	customer_status <fct>	count <int>	perc <dbl>
single	closed_account	944	47.03538
single	active	1063	52.96462
married	closed_account	989	43.64519
married	active	1277	56.35481
divorced	closed_account	159	44.91525
divorced	active	195	55.08475

6 rows

```
ggplot(data = credit_card_df, aes(x = marital_status, fill = customer_status)) + geom_ba
r() +
   labs(title = "Relationship between customer canceling their credit card and marital s
tatus",
            x = "Marital Status", y = "Customer Status")
```


Relationship between customer canceling their credit card and marital status

Question 2

Question:Cancellation of credit card vs their annual income?

Answer: All most 2092 closed there credit account who's average income of closed account people is 61601.51. Count of people who have there credit account active is 2535 and there average income is 62842.83.

```
credit_card_df %>% group_by(customer_status) %>%
              summarise(n_customers = n(),
                        min_income = min(income),
                        first_quartile = quantile(income, 0.25),
                        avg_income = mean(income),
                        third_quartile = quantile(income, 0.75),
                        max_income = max(income),
                        sd_income = sd(income))
```

customer_status <fct>	n_customers <int>	min_income <dbl>	first_quartile <dbl>	avg_income <dbl>	third_quartile <dbl>
closed_account	2092	30198	35311	61601.51	79796.0
active	2535	30094	35496	62842.83	79858.5

2 rows | 1-6 of 8 columns

```
ggplot(credit_card_df, aes(x=income,y = customer_status
,fill=customer_status)) + geom_boxplot() + coord_flip()+
  labs(title = "Credit card cancellation depending on annual income ",
          x = "Annual Income",
          y = "Customer Status")
```


Credit card cancellation depending on annual income
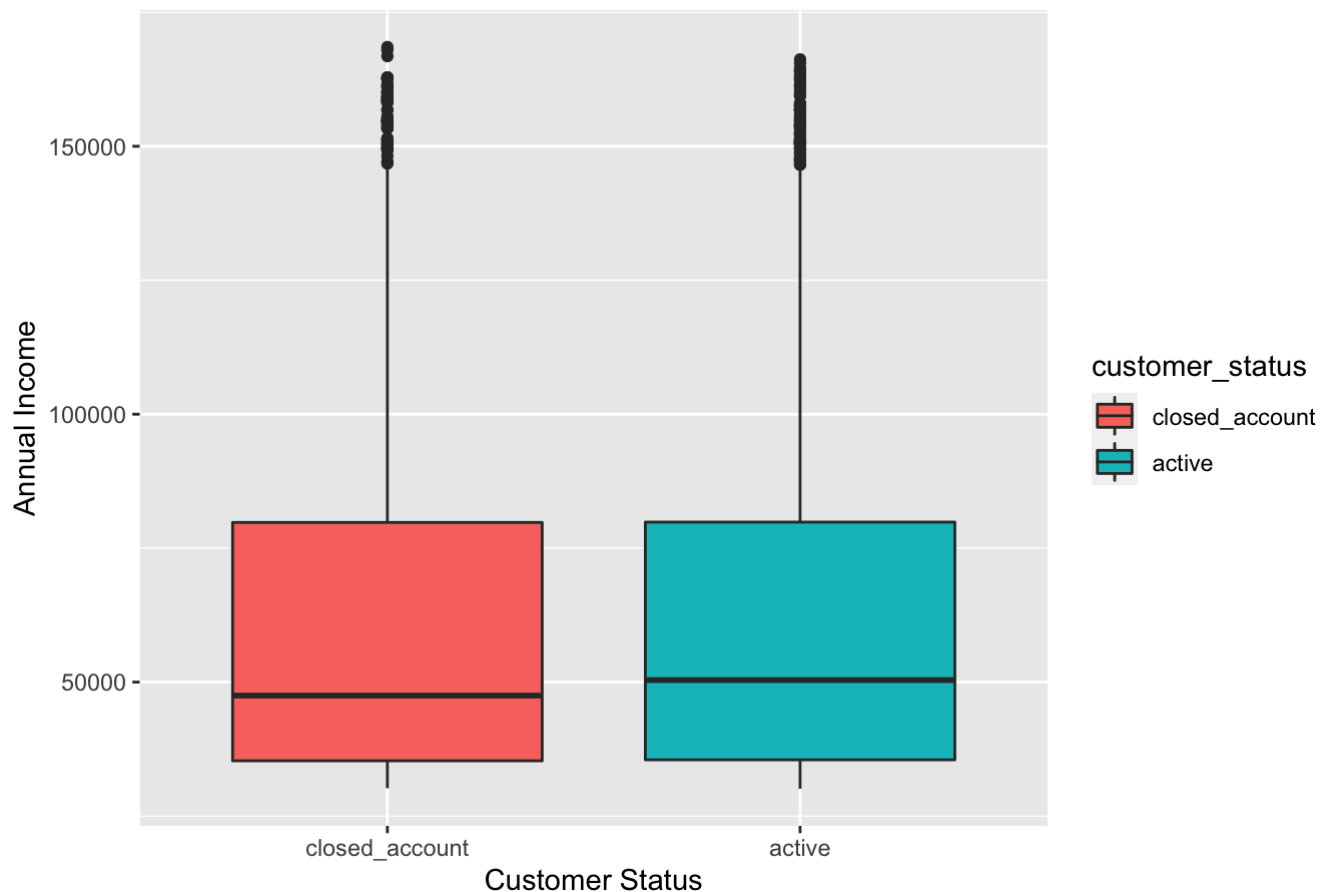
Question 3

Question:Does education level has an influence on cancellation of credit card?

Answer: When compared the relationship between the education level and cancellation of credit card we found nearly equal percent of people closed there account and active account. We also found that percentage of people who closed there account whose education level is doctorate is high of 7% which is 53.4% compared to active accounts of percentage 46.6%. In the remaining education levels the percentage of closed account is less than active accounts percentage.

```
credit_card_df %>%
  group_by(education,customer_status) %>%
  summarise(count = n()) %>%
  mutate(perc = 100*count/sum(count))
```

education	customer_status	count	perc
<fct>	<fct>	<int>	<dbl>
associates	closed_account	874	45.35547
associates	active	1053	54.64453
bachelors	closed_account	201	42.85714
bachelors	active	268	57.14286
masters	closed_account	891	44.66165
masters	active	1104	55.33835
doctorate	closed_account	126	53.38983
doctorate	active	110	46.61017

8 rows

Question 4

Question: Transactions_last_year vs total_spend_last_year with customer_status. Is there a relationship?

Answer: From the analysis we can see that there is effect between these two. 85% of peoples have transactions last year less than 60 who have canceled the credit card service. And those who have active credit card service 32% of them whose transactions last year is greater than sixty.

```
credit_card_df %>% group_by(customer_status) %>%
                summarise(n_customers = n(),
                        min_charges = min(transactions_last_year),
                        first_quartile = quantile(transactions_last_year, 0.25),
                        avg_charges = mean(transactions_last_year),
                        third_quartile = quantile(transactions_last_year, 0.75),
                        max_charges = max(transactions_last_year),
                        sd_charges = sd(transactions_last_year),
                        pct_less_60 = mean(transactions_last_year <= 60))
```

customer_status	n_customers	min_charges	first_quartile	avg_charges	third_quartile
<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
closed_account	2092	10	37	45.03776	51
active	2535	11	54	68.45602	82

2 rows | 1-6 of 9 columns

```
credit_card_df %>% group_by(customer_status) %>%
                summarise(n_customers = n(),
                          min_charges = min(total_spend_last_year),
                          avg_charges = mean(total_spend_last_year),
                          max_charges = max(total_spend_last_year),
                          sd_charges = sd(total_spend_last_year))
```

customer_status	n_customers	min_charges	avg_charges	max_charges	sd_charges
<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
closed_account	2092	510	3120.827	10583	2344.730
active	2535	893	4596.663	17498	3464.805
2 rows					

```
ggplot(credit_card_df, aes(x= transactions_last_year , y= total_spend_last_year)) +geom_
point(aes(colour= customer_status))+ facet_wrap(~customer_status,nrow=2)+geom_abline()+g
gtitle("Transactions_last_year vs total_spend_last_year with customer_status")
```


Transactions_last_year vs total_spend_last_year with customer_status
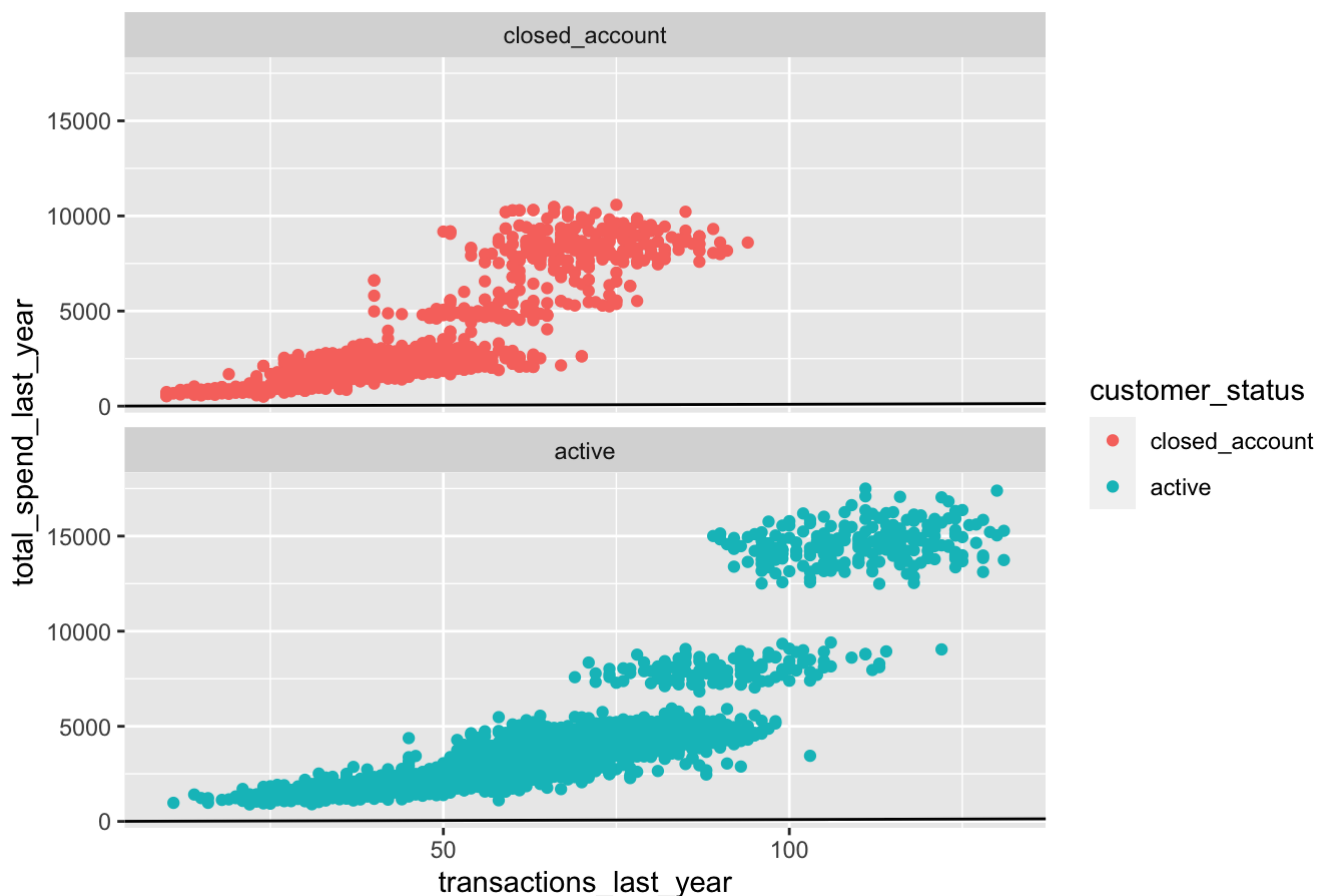
Question 5

Question:Most credit card cancellations are from which credit card type?

Answer: There is link between the card type and the cancellation of credit card service. Because we can see that 58.6% of blue card holder closed there account. Compared to the other card types silver and gold which has less percentage of credit card cancellation is less than active credit card service.

```
c_type <- credit_card_df %>%
  group_by(card_type,customer_status) %>%
  summarise(count = n()) %>%
  mutate(perc = 100*count/sum(count))

c_type
```

card_type <fct>	customer_status <fct>	count <int>	perc <dbl>
blue	closed_account	1497	58.68287
blue	active	1054	41.31713
silver	closed_account	296	25.34247
silver	active	872	74.65753
gold	closed_account	299	32.92952
gold	active	609	67.07048
6 rows			

```
#plot
ggplot(c_type, aes(x = factor(card_type), y = perc*100, fill = factor(customer_status)))
+
  geom_bar(stat="identity", position = "dodge", width = 0.7) +
  labs(title = "Card type vs Custmer status",
       x = "Card Type", y = "Customer Status", fill = "Customer Status")
```

Card type vs Custmer status
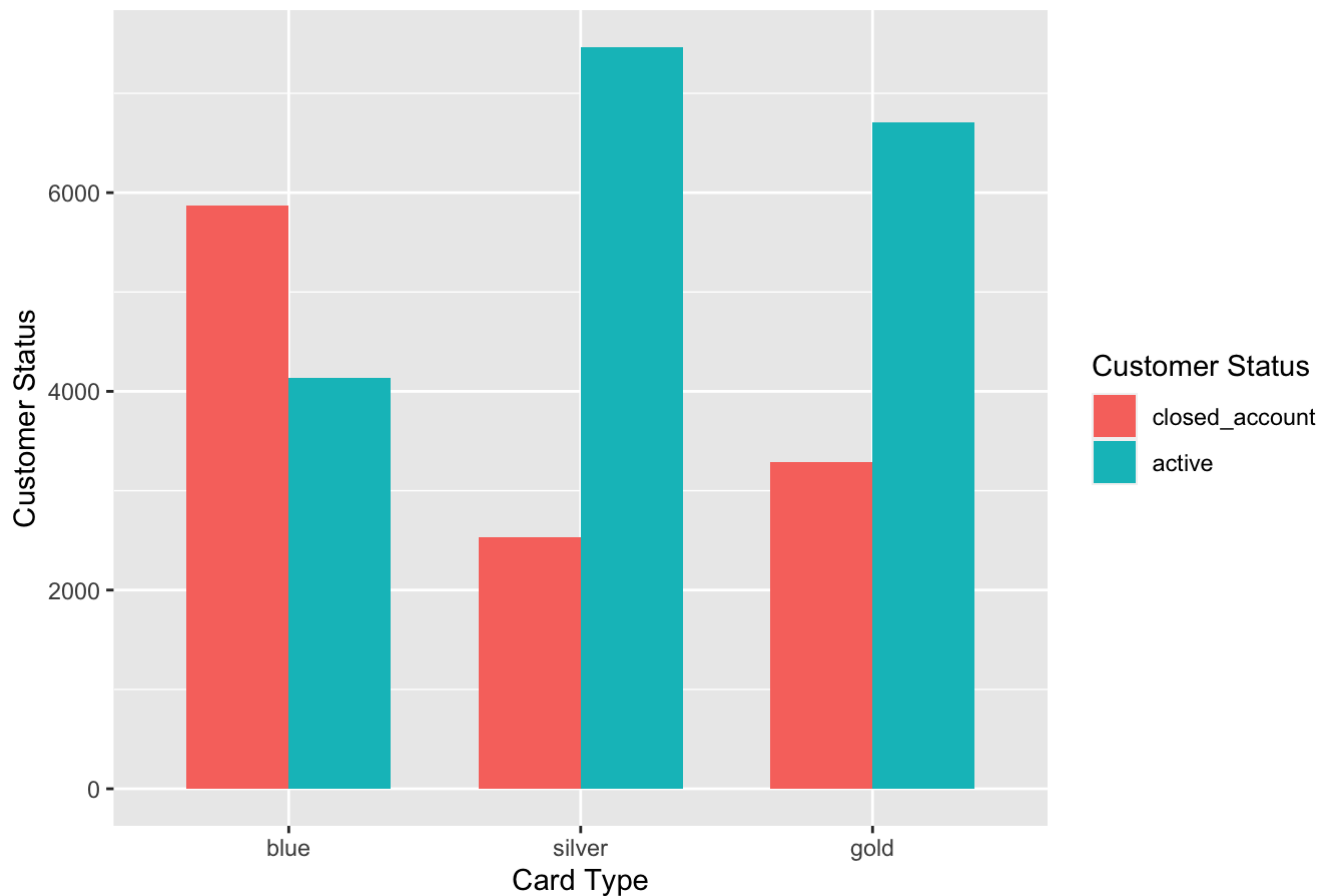
#Question 6:

Question: Are there any findings when compared with customer status, total accounts and months_inactive_last_year

Answer: From the graph we can say that people with 2 and 3 accounts has closed there credit card account. Also people who did not use there card for 3 months mostly closed there credit card account.

```
credit_card_df %>% group_by(customer_status) %>%
            summarise(n_customers = n(),
                      min_charges = min(total_accounts),
                      avg_charges = mean(total_accounts),
                      max_charges = max(total_accounts),
                      sd_charges = sd(total_accounts))
```

customer_status	n_customers	min_charges	avg_charges	max_charges	sd_charges
<fct>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
closed_account	2092	1	3.291587	6	1.583954
active	2535	1	3.936095	6	1.517544
2 rows					

```
credit_card_df %>% group_by(customer_status) %>%
                summarise(n_customers = n(),
                        min_charges = min(months_inactive_last_year),
                        avg_charges = mean(months_inactive_last_year),
                        max_charges = max(months_inactive_last_year),
                        sd_charges = sd(months_inactive_last_year))
```

| customer_status <fct>	n_customers <int>	min_charges <dbl>	avg_charges <dbl>	max_charges <dbl>	sd_charges <dbl>
closed_account	2092	0	2.691205	6	0.9044547
active	2535	0	2.261538	6	1.0089575

2 rows

```
ggplot(data = credit_card_df, aes(x = months_inactive_last_year, fill = customer_statu
s)) +
    geom_histogram(aes(y = ..density..), color = "white", bins = 20) +
    facet_wrap(~ customer_status, nrow = 2) +
    labs(title = "Credit Card cancellation depending on the months inactive ",
        x = "months_inactive_last_year", y = "customer_status")
```


Credit Card cancellation depending on the months inactive
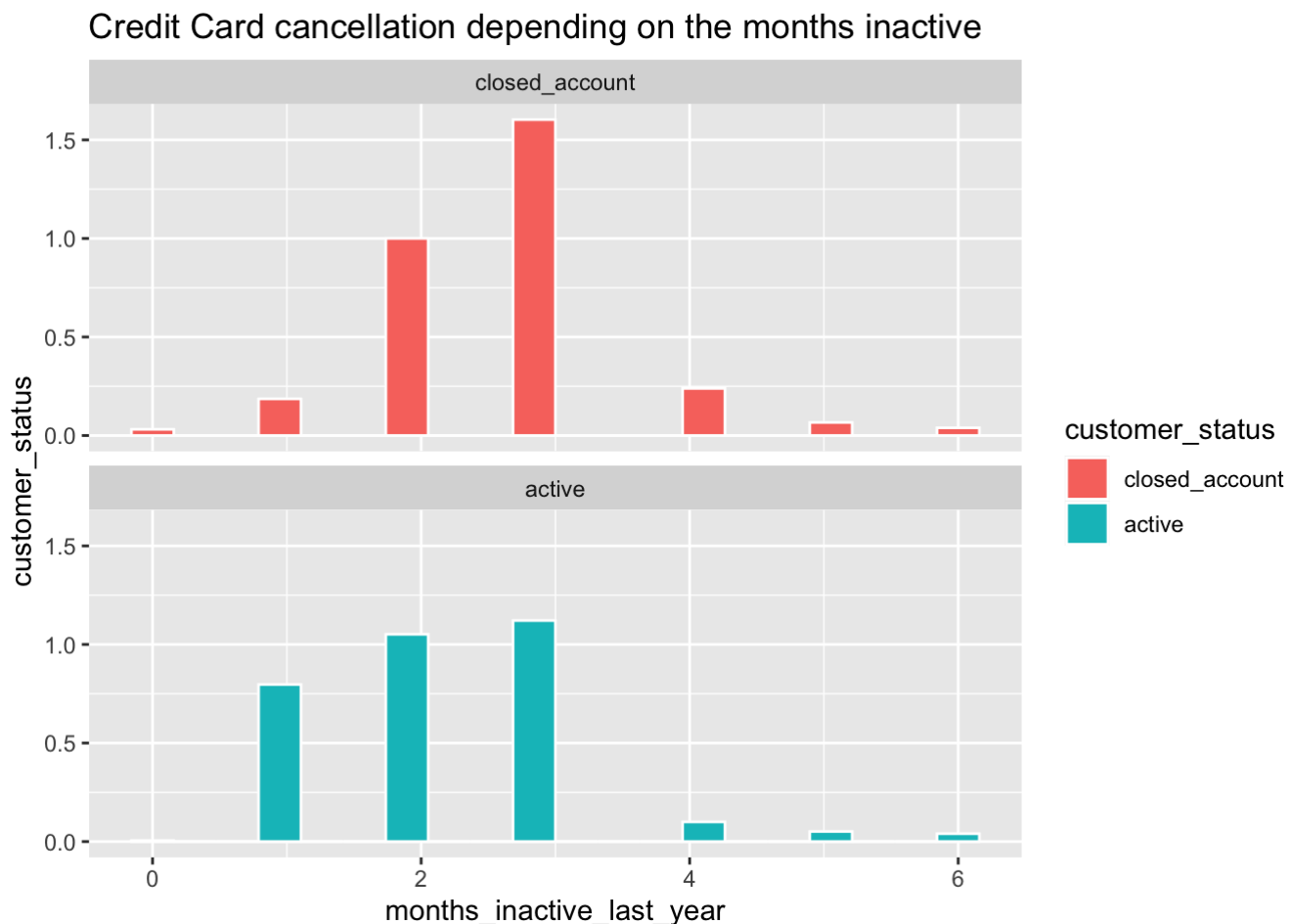
```
ggplot(data = credit_card_df, aes(x = total_accounts, fill = customer_status)) +
    geom_histogram(aes(y = ..density..), color = "white", bins = 20) +
    facet_wrap(~ customer_status, nrow = 2) +
    labs(title = "Credit Card cancellation depending on the total accounts ",
         x = "total_accounts", y = "customer_status")
```


Credit Card cancellation depending on the total accounts
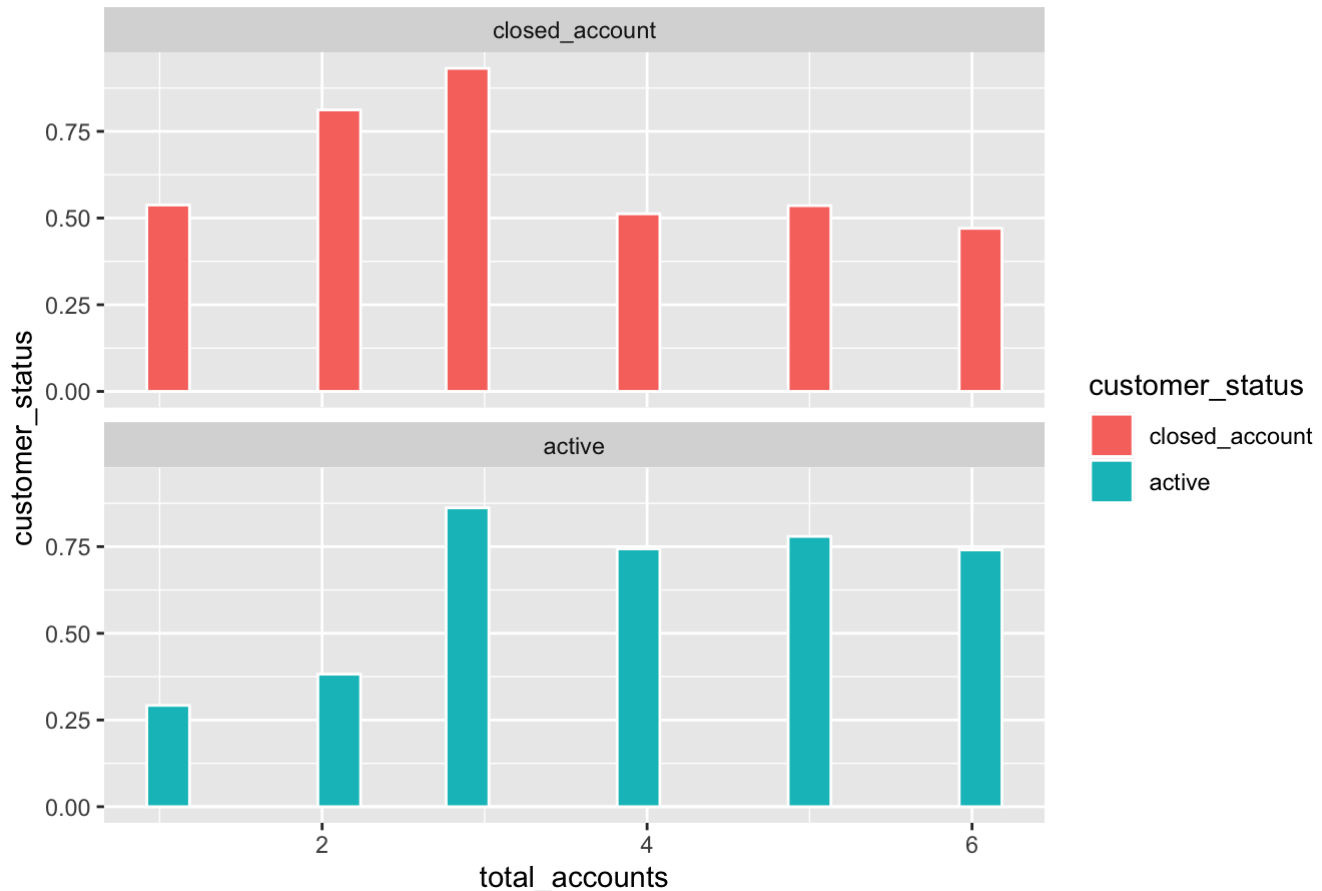
Machine Learning

In this section of the project, you will fit **three classification algorithms** to predict the outcome
variable, `customer_status` .

You must follow the machine learning steps below.

The data splitting and feature engineering steps should only be done once so that your models are using the
same data and feature engineering steps for training.

- Split the `credit_card_df` data into a training and test set (remember to set your seed)
- Specify a feature engineering pipeline with the `recipes` package
  - You can include steps such as skewness transformation, correlation filters, dummy variable encoding
    or any other steps you find appropriate
- Specify a `parsnip` model object
  - You may choose from the following classification algorithms:
    - Logistic Regression
    - LDA

- QDA
- KNN
- Decision Tree
- Random Forest
- Package your recipe and model into a workflow
- Fit your workflow to the training data
  - If your model has hyperparameters:
    - Split the training data into 5 folds for 5-fold cross validation using `vfold_cv` (remember to set your seed)
    - Perform hyperparamter tuning with a random grid search using the `grid_random()` function
    - Refer to the following tutorial for an example - Random Grid Search (https://gmubusinessanalytics.netlify.app/lesson-08-r-tutorial.html#Hyperparameter_Tuning14)
    - Hyperparameter tuning can take a significant amount of computing time. Be careful not to set the `size` argument of `grid_random()` too large. I recommend `size` = 10 or smaller.
    - Select the best model with `select_best()` and finalize your workflow
- Evaluate model performance on the test set by plotting an ROC curve using `autoplot()` and calculating the area under the ROC curve on your test data

```
set.seed(56)

cc_split <- initial_split(credit_card_df , prop = 0.75,
                            strata = customer_status)

cc_training <- cc_split %>% training()

cc_test<- cc_split  %>% testing()

# Create cross validation folds for hyperparameter tuning
set.seed(56)

cc_folds<- vfold_cv(cc_training, v = 5)
```

Feature Engineering

```
cc_recipe<- recipe(customer_status ~., data = credit_card_df) %>%
            step_YeoJohnson(all_numeric(), -all_outcomes()) %>%
            step_normalize(all_numeric(), -all_outcomes()) %>%
            step_dummy(all_nominal(), -all_outcomes())
```

#Model 1

Logistic Regression

```
logistic_model<- logistic_reg() %>%
            set_engine('glm') %>%
            set_mode('classification')
```

Creating a Workflow

```
logistic_wf<- workflow() %>%
              add_model(logistic_model) %>%
              add_recipe(cc_recipe)
```

Fitting Model

```
logistic_fit<-  logistic_wf%>%
                last_fit(split = cc_split)
```

Collecting Predictions

```
logistic_results <-  logistic_fit%>%
                collect_predictions()
logistic_results
```

id	.pred_closed_account	.pred_active	.r…	.pred_class	customer_statu
<chr>	<dbl>	<dbl>	<int>	<fct>	<fct>
train/test split	3.450369e-01	6.549631e-01	3	active	closed_account
train/test split	2.350118e-01	7.649882e-01	20	active	active
train/test split	6.982217e-01	3.017783e-01	21	closed_account	closed_account
train/test split	6.556262e-02	9.344374e-01	23	active	active
train/test split	5.564445e-03	9.944356e-01	36	active	active
train/test split	6.310508e-01	3.689492e-01	38	closed_account	closed_account
train/test split	8.600884e-01	1.399116e-01	39	closed_account	closed_account
train/test split	3.898251e-01	6.101749e-01	41	active	active
train/test split	7.945724e-01	2.054276e-01	44	closed_account	closed_account
train/test split	9.746486e-01	2.535144e-02	48	closed_account	closed_account

1-10 of 1,157 rows | 1-6 of 7 columns   Previous **1** 2   3   4   5   6 … 116 Next

ROC Curve

```
## ROC Curve
roc_curve(logistic_results , truth = customer_status , estimate = .pred_closed_account )
%>%
  autoplot()
```
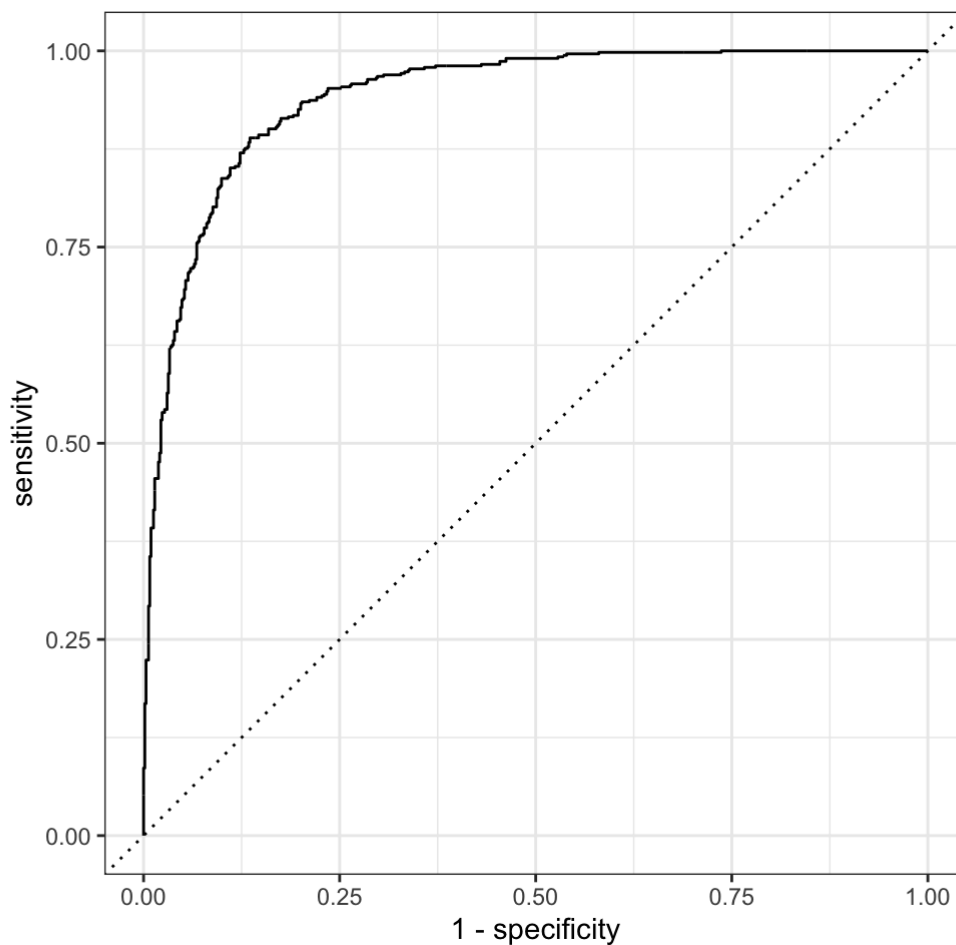
```
# ROC AUC
roc_auc(logistic_results, truth = customer_status, .pred_closed_account)
```

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
roc_auc	binary	0.9418032

1 row

```
# Confusion Matrix
conf_mat(logistic_results, truth = customer_status, estimate = .pred_class)
```

```
##                    Truth
## Prediction      closed_account active
##    closed_account           455     78
##    active                    68    556
```

Model 2

#KNN

```
library(kknn)
set.seed(310)
knn_model <- nearest_neighbor(neighbors =tune()) %>%
             set_engine('kknn') %>%
             set_mode('classification')
```

##Creating Workflow

```
knn_wf <- workflow() %>%
          add_model(knn_model) %>%
          add_recipe(cc_recipe)


k_grid <- tibble(neighbors = c(10, 15, 25, 45, 60, 80, 100, 120, 140, 180))


set.seed(310)

knn_tuning <- knn_wf %>%
              tune_grid(resamples = cc_folds, grid = k_grid)

best_k <- knn_tuning %>%
          select_best(metric = 'roc_auc')

final_knn_wf <- knn_wf %>%
                finalize_workflow(best_k)
```
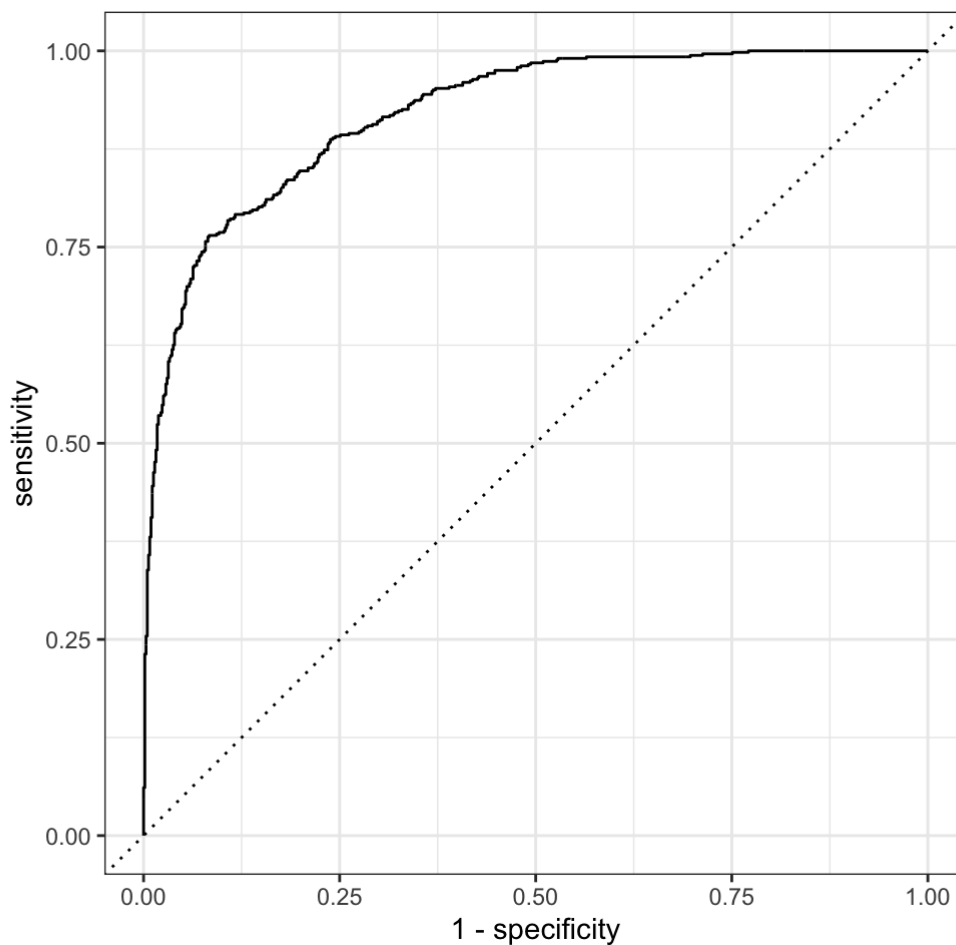
Fitting Model

```
knn_fit <- final_knn_wf %>%
           last_fit(split = cc_split)
```

Collecting Predictions

```
knn_results <-   knn_fit %>%
                 collect_predictions()
```

ROC Curve

```
roc_curve( knn_results , truth = customer_status , estimate = .pred_closed_account ) %>%
  autoplot()
```

```
roc_auc(knn_results, truth = customer_status, .pred_closed_account )
```

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
roc_auc	binary	0.9202068

1 row

```
conf_mat(knn_results, truth = customer_status, estimate = .pred_class)
```

```
##                   Truth
## Prediction     closed_account active
##    closed_account        419     95
##    active                104    539
```

```
my_metrics<- metric_set(accuracy,f_meas)
my_metrics(knn_results,truth=customer_status,estimate=.pred_class)
```

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
accuracy	binary	0.8280035

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
f_meas	binary	0.8081003

2 rows

Model 3

#Decision Trees

#Model Specification

```
tree_model <- decision_tree(cost_complexity = tune(),
                            tree_depth = tune(),
                            min_n = tune()) %>%
              set_engine('rpart') %>%
              set_mode('classification')
```

Creating the Workflow

```
tree_flow <- workflow() %>%
                add_model(tree_model) %>%
                add_recipe(cc_recipe)
```

#Hyperparamter tuning

```
## Create a grid of hyper parameter values to test
t_grid <- grid_regular(cost_complexity(),
                       tree_depth(),
                       min_n(),
                       levels = 2)

t_grid
```

cost_complexity	tree_depth	min_n
<dbl>	<int>	<int>
1e-10	1	2
1e-01	1	2
1e-10	15	2
1e-01	15	2
1e-10	1	40
1e-01	1	40
1e-10	15	40

cost_complexity	tree_depth	min_n
<dbl>	<int>	<int>
1e-01	15	40

8 rows

#Tuning hyperparameters

```
## Tune decision tree work flow
set.seed(65)

t_tune <- tree_flow %>%
            tune_grid(resamples = cc_folds,
                      grid = t_grid)
```

Show the top 5 best models based on roc_auc metric

```
t_tune %>% show_best('roc_auc')
```

cost_complexity	tree_depth	mi...	.metric	.estimator	mean	n	std_err
<dbl>	<int>	<int>	<chr>	<chr>	<dbl>	<int>	<dbl>
1e-10	15	40	roc_auc	binary	0.9502096	5	0.005600466
1e-10	15	2	roc_auc	binary	0.9109364	5	0.004790632
1e-10	1	2	roc_auc	binary	0.7636351	5	0.007256002
1e-01	1	2	roc_auc	binary	0.7636351	5	0.007256002
1e-01	15	2	roc_auc	binary	0.7636351	5	0.007256002

5 rows | 1-8 of 9 columns

```
## Select best model based on roc_auc
best_tree <- t_tune %>%
            select_best(metric = 'roc_auc')

# View the best tree parameters
best_tree
```

cost_complexity	tree_depth	min_n	.config
<dbl>	<int>	<int>	<chr>
1e-10	15	40	Preprocessor1_Model7

1 row

```
final_tree_flow <- tree_flow %>%
                          finalize_workflow(best_tree)
```
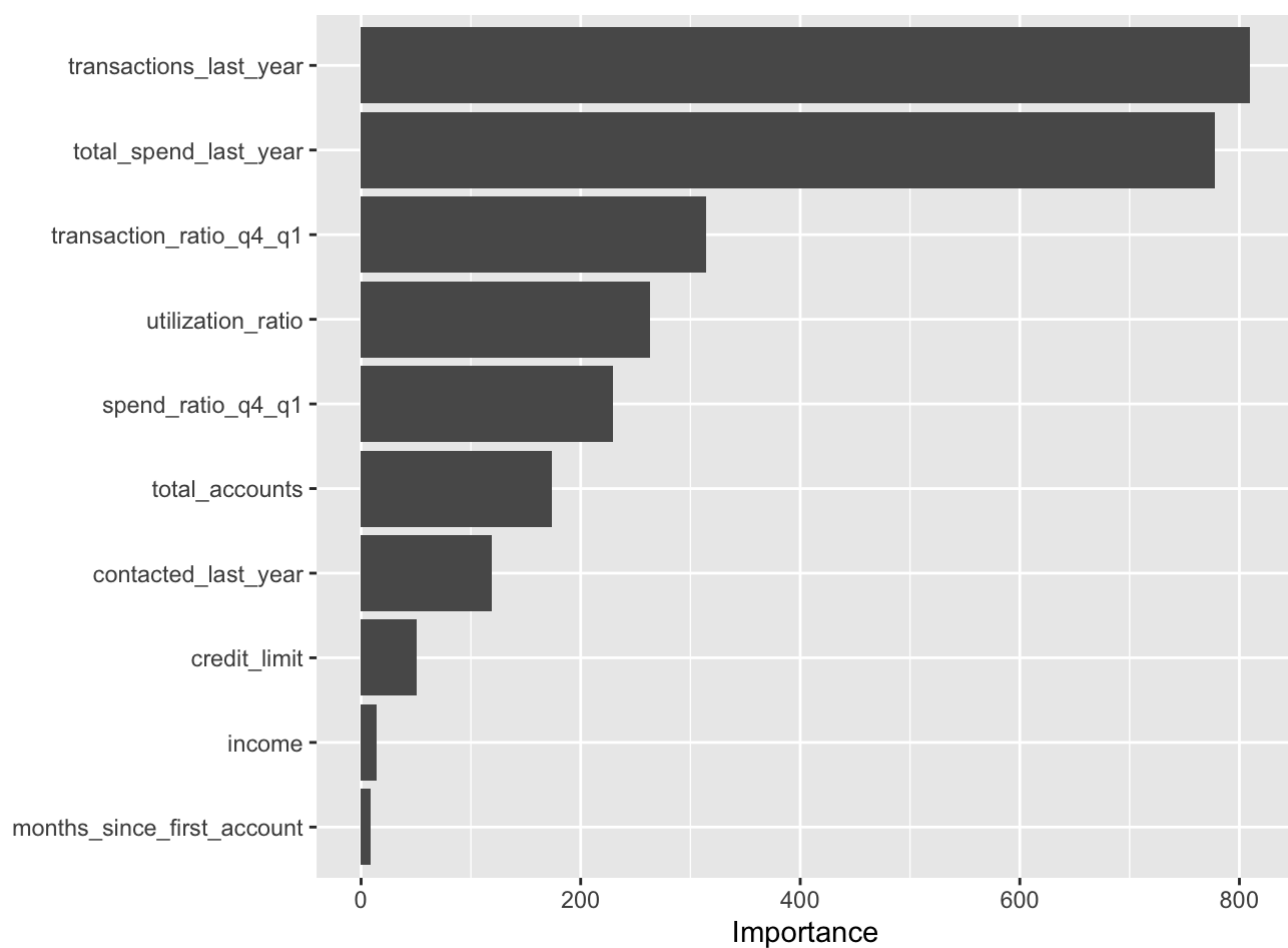
#Fittiing the Model
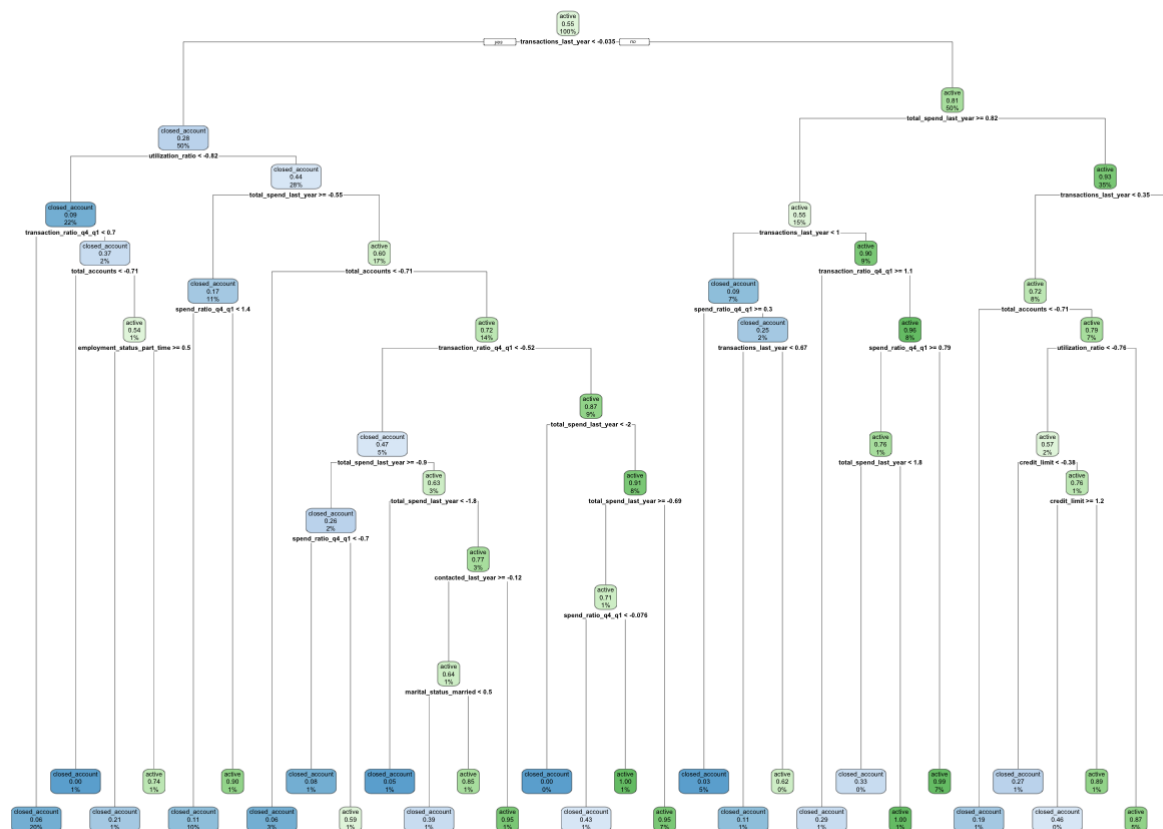
```
tree_wf_fit <- final_tree_flow %>%
                 fit(data = cc_training)
```

```
tree_fit <- tree_wf_fit %>%
              extract_fit_parsnip()
```

```
vip(tree_fit)
```



```
rpart.plot(tree_fit$fit, roundint = FALSE)
```

#Train and Evaluate With last_fit()

```
tree_last_fit <- final_tree_flow %>%
                last_fit(cc_split)
```

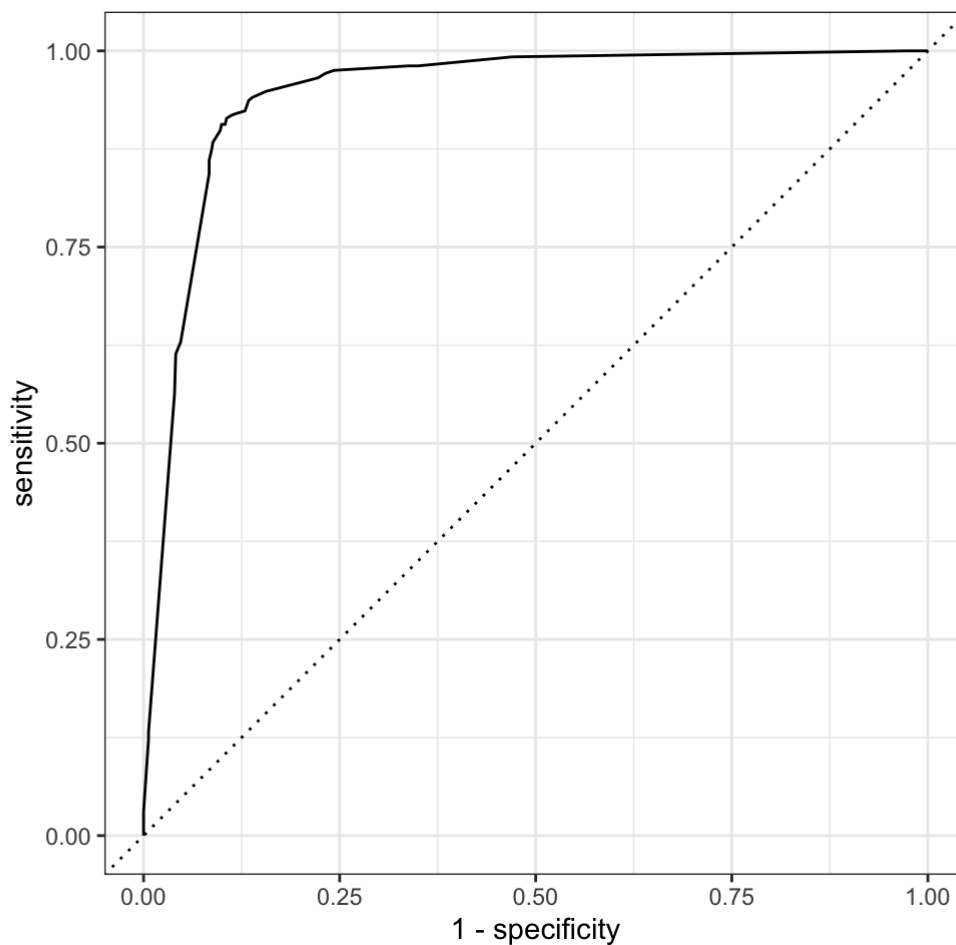```
tree_last_fit %>% collect_metrics()
```

| .metric | .estimator | .estimate | .config |
|---------|------------|-----------|---------|
| <chr> | <chr> | <dbl> | <chr> |
| accuracy | binary | 0.8997407 | Preprocessor1_Model1 |
| roc_auc | binary | 0.9447723 | Preprocessor1_Model1 |

2 rows

```
t_pred <- tree_last_fit %>% collect_predictions()
```

ROC Curve and Confusion Matrix

```
#ROC Curve
roc_curve(t_pred, truth  = customer_status, estimate = .pred_closed_account) %>%
          autoplot()
```

```
#roc-AUC
roc_auc(t_pred, truth = customer_status, .pred_closed_account)
```

.metric	.estimator	.estimate
<chr>	<chr>	<dbl>
roc_auc	binary	0.9447723

1 row

```
#Confusion Matrix

conf_mat(t_pred, truth = customer_status, estimate = .pred_class)
```

```
##                     Truth
## Prediction      closed_account active
##    closed_account          481     74
##    active                   42    560
```

Summary of Results

Write a summary of your overall findings and recommendations to the executives at the bank. Think of this section as your closing remarks of a presentation, where you summarize your key findings, model performance, and make recommendations to improve customer retention and service at the bank.

Your executive summary must be written in a professional tone (https://www.universalclass.com/articles/writing/business-writing/appropriate-tone-in-business-communications.htm), with minimal grammatical errors, and should include the following sections:

1. An introduction where you explain the business problem and goals of your data analysis

    ○ What problem(s) is this company trying to solve? Why are they important to their future success?

    ○ What was the goal of your analysis? What questions were you trying to answer and why do they matter?

2. Highlights and key findings from your Exploratory Data Analysis section
    ○ What were the interesting findings from your analysis and **why are they important for the business**?

    ○ This section is meant to **establish the need for your recommendations** in the following section

3. Your "best" classification model and an analysis of its performance
    ○ In this section you should talk about the expected error of your model on future data
        ▪ To estimate future performance, you can use your model performance results on the **test data**
    ○ You should discuss at least one performance metric, such as an F1, sensitivity, specificity, or ROC AUC for your model. However, you must explain the results in an **intuitive, non-technical manner**. Your audience in this case are executives at a bank with limited knowledge of machine learning.

4. Your recommendations to the bank on how to reduce the number of customers closing their credit card accounts

    ○ Each recommendation must be supported by your data analysis results

    ○ You must clearly explain why you are making each recommendation and which results from your data analysis support this recommendation

    ○ You must also describe the potential business impact of your recommendation:

        ▪ Why is this a good recommendation?

        ▪ What benefits will the business achieve?

Summary

Add your summary here. Please do not place your text within R code chunks.

• The US Bank has experienced record levels of customers closing their credit card service in the past couple of months and this is leading to large financial losses. The bank is looking to find out the factors that lead to customers canceling their credit card account and whether it can predict if a customer will cancel their account in the future. It is important to find the problem of why customers are canceling their credit card services. • The goal of my analysis is to identify the reason behind why customers are canceling their credit card account and try to minimize the risk of canceling the credit card.

Many elements such as customer age, education level, marital status, employment_status, income, and card_type are taken into considerations to predict the reason behind why the customer is canceling their credit card.

-Is there a link between a customer's decision to cancel their credit card and their marital status? -Credit card cancellation versus yearly income? -Is there a link between education and credit card cancellation? -With customer status, compare transactions to total spend last year. Is there a connection? -Which credit card type accounts for the majority of credit card cancellations? -When comparing customer status, total accounts, and months inactive last year, are there any conclusions?

Below are the findings from the exploratory evaluation:

• The majority of single persons canceled their accounts. The number of single persons who stopped their accounts was 47 percent, while the number of married people who deleted their accounts was 43.6 percent. 44.9 percent of divorced persons had their accounts canceled. • The majority of 2092 individuals who have canceled their credit accounts had an average salary of 61601.51. There are 2535 persons with active credit accounts, with an average salary of 62842.83. • We also discovered that the percentage of persons who closed their accounts with a doctoral degree is high at 7%, or 53.4 percent, compared to 46.6 percent for active accounts. The percentage of closed accounts is lower than the percentage of active accounts in the remaining academic levels. • Last year, 85 percent of people's transactions were fewer than 60, indicating that they had discontinued their credit card service. Those who have an active credit card account have 32 percent more transactions than the previous year, while 58.6% of blue cardholders have canceled their accounts. Silver and gold cards have a lower percentage of credit card cancellation than active credit card services when compared to another card type.* • People with 2 and 3 credit cards have closed their accounts. People who had not used their credit cards for three months were also more likely to terminate their accounts.

MODELS Three classification models Logistic Regression, KNN Analysis, and Decision Tree are implemented to predict the target variable (cancel service in this case). Based on the ROC AUC, Logistic Regression and Linear Discriminant Analysis models have the best performance when compared with Decision Trees. The ROC AUC is 0.868 for Logistic Regression, 0.866 for Linear Discriminant Analysis, and 0.76 for Decision Tree. AUC tells how much the model is fit for recognizing classes and ROC is the probability curve. Higher the AUC, the better the model performance.

The model limitations: • The models always perform better when the data set is larger. Since the dataset taken has lesser data, the performance is higher. • Also, the factors for the model are limited to the target variable that might get affected if more factor variables are added.

The customer churn rate is affected in a different way depending on the factors. The solutions can be easily found once the problem is found. Below are a few proposals: • Months with the company, monthly charges, and average call minutes are the most significant factors that the cancellation of service depends on. So, if the company plans on changing these factors in a more customer-reliable way, there is a high chance that the customers might stay with the company. • Digital service is preferable compared to fiber-optic service, so if the company provides more digital services, it will attract more users. If the company introduces longer contracts for a lesser price, the users tend to stay longer with the company.

Three classification models Logistic Regression, KNN Analysis, and Decision Tree are implemented to predict the target variable (cancel service in this case). Based on the ROC AUC, Logistic Regression and KNN Analysis models have the best performance when compared with Decision Trees. The ROC AUC is 0.942 for Logistic Regression, 0.920 for KNN Analysis, and 0.945 for Decision Tree. AUC tells how much the model is fit for recognizing classes and ROC is the probability curve. Higher the AUC, the better the model performance. We can see that the decision tree has the highest ROC value and it is the better model.

The model limitations: • The models always perform better when the data sets are larger. Since the data set taken has less data, the performance is higher. • Also, the factors for the model are limited to the target variable that might get affected if more factor variables are added.

The customer credit card cancellation is affected in different ways depending on the factors. The solutions can be easily found once the problem is found. Below are a few proposals: • Income, Total Accounts, card type, employment status, and utilization ratio are the most significant factors that the cancellation of credit card service depends on. So, if the company plans on changing these factors in a more customer-reliable way, there is a high chance that the customers might stay with the credit card account. • Silver credit card account is preferable compared to other credit cards, as the number of active cards is high, it will attract more users by adding some more offers. If the company introduces contracts for a lesser rate of interest, the users tend to stay longer with the credit account.

• From all the observations, it is important for the business to maintain customers be in their service. We need to decrease monthly credit charges and if possible increase the credit card limit. Implement some offers for every 1 year and 2-year so that customers can stay for a long time and give offers on all the credit card services equally. Bank can decrease the interest rate to attract the customer. Also, we can give offers such as when we refer a friend and he applies for a credit card from that bank we can give rewards to both the customers. We can also increase the rate of cash back for online shopping and also we can give points on credit cards depending on their travel rate where the count of miles is considered. This way we can increase the number of customers to incerase the bank profits.