Big Data Tools for Managers

Dept. of MBA, Siddaganga Institute of Technology-Tumkur

Working with Credit Card Churn Dataset

About Dataset A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate if one could predict for them who is gonna get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction Download Credit Card Churn Dataset

- **CLIENTNUM**: Client number. Unique identifier for the customer holding the account
- **Attrition_Flag**: Internal event/Churn (customer activity) variable – if the account is closed then 1 else 0
- **Customer_Age**: Demographic variable – Customer's Age in Years
- **Gender**: Demographic variable – M=Male, F=Female
- **Dependent_count**: Demographic variable – Number of dependents
- **Education_Level**: Educational Qualification of the account holder (example: high school, college graduate, etc.)
- **Marital_Status**: Demographic variable – Married, Single, Divorced, Unknown
- **Income_Category**: Demographic variable – Annual Income Category of the account holder
- **Card_Category**: Product Variable – Type of Card (Blue, Silver, Gold, Platinum)
- **Months_on_book**: Period of relationship with bank
- **Total_Relationship_Count**: Total no. of products held by the customer
- **Months_Inactive_12_mon**: No. of months inactive in the last 12 months
- **Contacts_Count_12_mon** No. of Contacts in the last 12 months
- **Credit_Limit**: Credit Limit on the Credit Card
- **Total_Revolving_Bal**: Total Revolving Balance on the Credit Card
- **Avg_Open_To_Buy**: Open to Buy Credit Line (Average of last 12 months)
- **Total_Amt_Chng_Q4_Q1**: Change in Transaction Amount (Q4 over Q1)
- **Total_Trans_Amt**: Total Transaction Amount (Last 12 months)
- **Total_Trans_Ct**: Total Transaction Count (Last 12 months)
- **Total_Ct_Chng_Q4_Q1**: Change in Transaction Count (Q4 over Q1)
- **Avg_Utilization_Ratio** Average Card Utilization Ratio

Write R Code for below questions.

1. Read Credit Card Churn dataset in R
2. Get the dimension of Credit Card dataset
3. Display column names of dataset
4. View data in Excel like screen
5. Get Quick summary for all the columns
6. Check how many customers are churned and existing customer in dataset
7. What percentage of customers are churned and existing customer in dataset
8. Display in Pie chart, Check how many customers are churned and existing customer in dataset
9. Analyze customers based on their Marital Status
10. Look the customers by Matrial status who has canceled the credit card
11. What are the income level for Married customers who has canceled the credit card
12. Look the Credit card cancellation by Income category
13. Look for the credit card type bank is providing and customer holding, display the count as well as show in barplot
14. Total Number of transcation in last 12 month for Active(Existing) customer in histogram
15. Split the dataset by Attition_Flag
16. Display the structure of Spitted Data
17. Display Customer data based on their Age (Smallest to Highest), and select Customer Age, Gender, Education Level, Marital_Status, Income_Category
18. Display top 10 records from the dataset
19. Display Unique Education_Level for the customers Data
20. Count the Frequency for each Education Level

Note : This analysis is based on very few variables, you may have to practice yourself with remaing variables to get more hands-on with R & Data

Note : Statements start with #(hash) are comments, Feel free to ignore comments while writing R code

```
In [ ]:   # 1. Read Credit Card Churn dataset in R
          data <- read.csv("BankChurners.csv")

In [ ]:   # 2. Get the dimension of Credit Card dataset
          dim(data)

          # We have 10127 rows & 23 variables(colums)

In [ ]:   # 3. Display column names of dataset
          colnames(data)

In [ ]:   # 4. View data in Excel like screen
          View(data)

In [ ]:   # 5. Get Quick summary for all the columns
          summary(data)

In [ ]:   # 6. Check how many customers are churned and existing customer in dataset
          table(data$Attrition_Flag)

In [ ]:   #7. What percentage of customers are churned and existing customer in dataset

          res <- table(data$Attrition_Flag)

          prop.table(res) * 100

          # There are 16% percentage of customers are churned customers and 84% are existing customers for the Bank

In [ ]:   #8. Display in Pie chart, Check how many customers are churned and existing customer in dataset
          res <- table(data$Attrition_Flag)
          pie(res,
              main="Attrition Flag in Dataset",
              col=c("red","green"))

In [ ]:   # 9. Analyze customers based on their Marital Status
          res <- table(data$Marital_Status)
```

```r
prop.table(res) * 100
# 46% of customers are Married and those are the most benificial customers to the bank since they normally spend good
```

In [ ]:
```r
# 10. Look the customers by Matrial status who has canceled the credit card
churned_customer = subset(data, Attrition_Flag=="Attrited Customer")

table(churned_customer$Marital_Status)
# There are 709 Married who has canceled the credit card
```

In [ ]:
```r
# 11. what are the income level for Married customers who has canceled the credit card
married_customers = subset(churned_customer, Marital_Status =='Married')
res = table(married_customers$Income_Category)
prop.table(res) * 100
# ~37% married customer are earning < $40,000, obviously this categories of customers looks for more offers and discou
# This ~37% customers usualy churn more if Bank is not offering good offers
```

In [ ]:
```r
# 12. Look the Credit card cancellation by Income category
table(data$Income_Category, data$Attrition_Flag)

# We can see < $40,000 income category having more customers compare woth other income group,
# and since < $40,000 has more customers and churn is also high since they normally look for more offers and discounts
```

In [ ]:
```r
# 13. Look for the credit card type bank is providing and customer holding, display the count as well as show in barpl
res = table(data$Card_Category)

barplot(res)
print(res)
# Mostly Bank has issued Blue category of Credit card to the customers
```

In [ ]:
```r
#14. Total Number of transcation in last 12 month for Active(Existing) customer in histogram
active_customer = subset(data, Attrition_Flag=='Existing Customer')

hist(active_customer$Total_Trans_Ct,
    main="Transaction Count For Last 12 Months",
    col="green")
```

In [ ]:
```r
#15. Split the dataset by Attition_Flag
splitted_data = split(data, data$Attrition_Flag)

# splitted_data will be in List data structure
```

```
In [ ]:  #16. Display the structure of Spitted Data
         str(splitted_data)
```

```
In [ ]:  #17. Display Customer data based on their Age (Smallest to Highest), and select Customer Age, Gender, Education Level,

         data[ order(data$Customer_Age),
              c("Customer_Age","Gender","Education_Level", "Marital_Status", "Income_Category")
            ]
```

```
In [ ]:  # 18 Display top 10 records from the dataset
         head(data,10)
```

```
In [ ]:  #19. Display Unique Education Level for the customer data
         unique(data$Education_Level)
```

```
In [ ]:  #20. Count the Frequency for each Education Level
         table(data$Education_Level)
         # Mostly Bank having Educatated customers and who has issued the credit card from the bank
```