Executive Summary

Customer churn prevention is a major part of the Customer Relationship Management (CRM) in any business. Same is the norm in telecom industry where, due to immense competition among the telecom carriers, there is a dire need for churn management for the operators to retain their current subscribers. Churn describes the subscribers who terminate their relationship with the service provider and move their subscription to the competitor.

In this project we are analyzing a Customer Churn dataset of a major wireless telecom operator in South Asia, South Asian Telecom Operator (SATO).

Analysis of this Churn data set for SATO has provided us the following insights: -

1. Maximum number of users of the network use the 3G spectrum for their connectivity.
2. Distribution of Churned and Active customers are the same in the provided data set.
3. The number of complaints made by the Active Customers is more than the complaints received by the Churned customers, pointing to the fact that complains might not be the sole cause for a customer to leave the network.
4. The revenue from Calls made outside the network for Active & Churned customers seems to be have a correlation among each other more than half the time.
5. Among the active users, the 3G spectrum users drive the revenue from SMSs and have the highest number of calls made.
6. Among the churned users, the 3G spectrum users drive the revenue from SMSs & Data as well and have the highest number of calls made.
7. Revenue received from SMS and revenue received from Data do not seem to share a trend among themselves.

The areas that we have tried to explore in our dataset are based on the following requirements of SATO:

1. SATO wants to investigate if the number of complaints in case of their Churned Customers are higher than their Active Customers, suggesting that it could be a reason for the Customers to leave the network.
2. SATO wants to investigate whether there is a relationship between a User's Status & the corresponding Revenue received from Calls made within and outside of their network. Based on the findings, SATO will devise some new strategies for their marketing campaign.
3. SATO wants to find the average calls made, average revenue from SMS and Data services for their Active customers based on the different user types.
4. SATO wants to find the average calls made, average revenue from SMS and Data services for their Churned customers based on the different user types.
5. SATO wants to investigate whether there is a trend between the revenue received from SMS & the revenue received from Data.

Data Definition

Data Source: https://www.kaggle.com/mahreen/sato2015

South Asian Telecom Operator (SATO) data set is a real-life data collected from a major wireless telecom operator in South Asia. Most of the attributes in the data sets are associated with call detail records (CDR), billing information. It contains 2000 subscribers. All these subscribers were not contract based and had a monthly based subscription. The subscriber data was extracted from the time interval of month i.e. August 2015.

Please find below the fields included in this dataset:

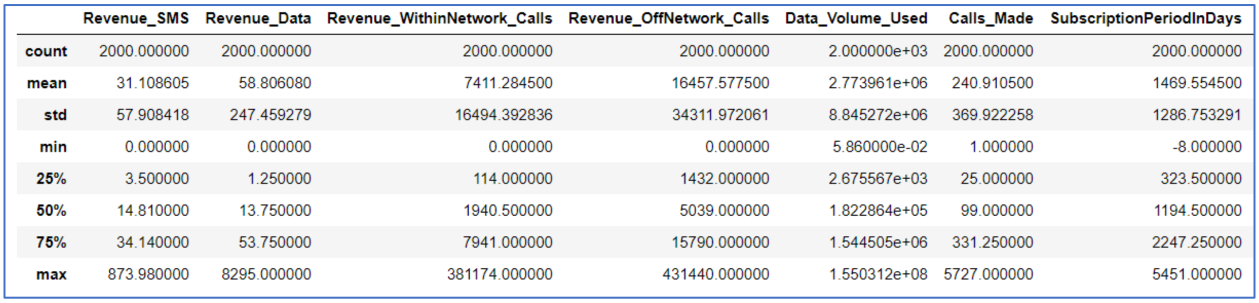
Fields	Description
UserID	Unique identifier for each user
Revenue_SMS	The revenue generated from SMS services
Revenue_Data	The revenue generated from data usage
Revenue_WithinNetwork_Calls	The revenue generated from calls made within the SATO network
Revenue_OffNetwork_Calls	The revenue generated from calls made outside the SATO network
Data_Volume_Used	Data volume used by the user
Calls_Made	Number of calls made by the user
SubscriptionPeriodInDays	No. of days the user is associated with SATO network
ComplaintCount	The number of complaints filed by the user.
UserType	The Spectrum used by the user (2G/3G/Other)
Status	The current status of the user with the network (Active/Churned)

Data Processing Results

1. Missing values was only found for the field: UserType; there were 245 missing values. The missing values were replaced with the value: “Others”.
2. Added a new field to the data set: Network_Spectrum based on the values for the field, “UserType”.
3. We have binned the continuous variable: “Data_VolumeUsed_Binned” with the following 5 equal bins:
    a. (2151812.82, 155031212.8]
    b. (465830.54, 2151812.82]
    c. (59281.5, 465830.54]
    d. (888.33, 59281.5]
    e. (0.048999999999999995, 888.33]

Data Exploration Results

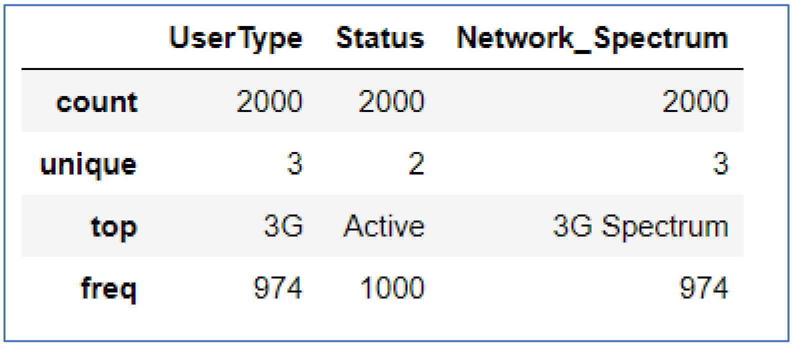
Quantitative Variables:



	Revenue_SMS	Revenue_Data	Revenue_WithinNetwork_Calls	Revenue_OffNetwork_Calls	Data_Volume_Used	Calls_Made	SubscriptionPeriodInDays
count	2000.000000	2000.000000	2000.000000	2000.000000	2.000000e+03	2000.000000	2000.000000
mean	31.108605	58.806080	7411.284500	16457.577500	2.773961e+06	240.910500	1469.554500
std	57.908418	247.459279	16494.392836	34311.972061	8.845272e+06	369.922258	1286.753291
min	0.000000	0.000000	0.000000	0.000000	5.860000e-02	1.000000	-8.000000
25%	3.500000	1.250000	114.000000	1432.000000	2.675567e+03	25.000000	323.500000
50%	14.810000	13.750000	1940.500000	5039.000000	1.822864e+05	99.000000	1194.500000
75%	34.140000	53.750000	7941.000000	15790.000000	1.544505e+06	331.250000	2247.250000
max	873.980000	8295.000000	381174.000000	431440.000000	1.550312e+08	5727.000000	5451.000000

- The monthly average revenue generated for SMS services usage is $31.
- The monthly average revenue generated for Data services usage is $58.
- The monthly average revenue generated for calls made within the network is $7411.
- The monthly average revenue generated for calls made outside the network is $16458.
- The monthly average Data Volume used is 2773961 bytes of data.
- The monthly average number of calls made is 241.
- The monthly average association of a customer with the network is 1470 days, i.e., 4.02 years

Categorical Variables:



	UserType	Status	Network_Spectrum
count	2000	2000	2000
unique	3	2	3
top	3G	Active	3G Spectrum
freq	974	1000	974

- Among the pool of active customers, the users opting the 3G spectrum seem to dominate in numbers in comparison to users of other spectrums.

Conclusion

Analysis of this Churn data set for SATO has provided us the following insights: -

1. Maximum number of users of the network use the 3G spectrum for their connectivity.
2. Distribution of Churned and Active customers are the same in the provided data set.
3. The number of complaints made by the Active Customers is more than the complaints received by the Churned customers, pointing to the fact that complains might not be the sole cause for a customer to leave the network.
4. The revenue from Calls made outside the network for Active & Churned customers seems to be have a correlation among each other more than half the time.
5. Among the active users, the 3G spectrum users drive the revenue from SMSs and have the highest number of calls made.
6. Among the churned users, the 3G spectrum users drive the revenue from SMSs & Data as well and have the highest number of calls made.
7. Revenue received from SMS and revenue received from Data do not seem to share a trend among themselves.

Appendix A: Part One & Two Code

Part One:

```
import pandas as pd
import numpy as np
from pandas import Series, DataFrame
from io import StringIO
import os
 import matplotlib.pyplot as plt

# Changing the directory to current director
os.chdir('C:\\Users\\ayans\\Documents')

#Reading the first worksheet of the data excel file
dfUserRevenues = pd.read_excel('SATO2015.xlsx',sheet_name='UserRevenues')
dfUserRevenues.head()

#Reading the second worksheet of the data excel file
dfUserComplaints = pd.read_excel('SATO2015.xlsx',sheet_name='UserComplaints')
dfUserComplaints.head()

#Reading the third worksheet of the data excel file
dfUserStatus = pd.read_excel('SATO2015.xlsx',sheet_name='UserStatus')
dfUserStatus.head()

#Merging all the three data frames to one final data frame
dfSATO = pd.merge(dfUserRevenues,dfUserComplaints, on='UserID',how='inner')
dfSATO = pd.merge(dfSATO,dfUserStatus, on='UserID',how='inner')
dfSATO.head()






#Data Processing

#Looking for missing values

#Field: Revenue_SMS
dfSATO[dfSATO['Revenue_SMS'].isnull()].Revenue_SMS   #No missing values found for this
field

#Field: Revenue_Data
dfSATO[dfSATO['Revenue_Data'].isnull()].Revenue_Data   #No missing values found for this
field

#Field: Revenue_WithinNetwork_Calls
dfSATO[dfSATO['Revenue_WithinNetwork_Calls'].isnull()].Revenue_WithinNetwork_Calls   #No
missing values found for this field

#Field: Revenue_OffNetwork_Calls
dfSATO[dfSATO['Revenue_OffNetwork_Calls'].isnull()].Revenue_OffNetwork_Calls   #No
missing values found for this field

#Field: Data_Volume_Used
```

```
dfSATO[dfSATO['Data_Volume_Used'].isnull()].Data_Volume_Used   #No missing values found
for this field

#Field: Calls_Made
dfSATO[dfSATO['Calls_Made'].isnull()].Calls_Made   #No missing values found for this
field

#Field: SubscriptionPeriodInDays
dfSATO[dfSATO['SubscriptionPeriodInDays'].isnull()].SubscriptionPeriodInDays   #No
missing values found for this field

#Field: ComplaintCount
dfSATO[dfSATO['ComplaintCount'].isnull()].ComplaintCount   #No missing values found for
this field

#Field: UserType
dfSATO[dfSATO['UserType'].isnull()].UserType   #245 missing values found for this field

# Replacing the missing values for the field: UserType with value:"Other"
dfSATO['UserType']=dfSATO['UserType'].fillna('Other')
#Re-verification of missing values for the field after treatment of missing values
dfSATO[dfSATO['UserType'].isnull()].UserType

#Field: Status
dfSATO[dfSATO['Status'].isnull()].UserType   #No missing values found for this field

# Adding a new detail to the dataframe by recoding another field
dfSATO['Network_Spectrum']=np.where(dfSATO['UserType']=='2G','2G Spectrum',
                                    np.where(dfSATO['UserType']=='3G','3G
Spectrum','Unknown'))
dfSATO.head()

# Binning the Field: Data_Volume_Used
Data_VolumeUsed_Binned = pd.qcut(dfSATO['Data_Volume_Used'],5,precision=2)
pd.value_counts(Data_VolumeUsed_Binned)

#Data Exploration

#Creating basic summaries for the quantitative fields
dfSATO.columns

#Creating basic summaries for the Quantitative variables

dfSATO[['Revenue_SMS','Revenue_Data','Revenue_WithinNetwork_Calls','Revenue_OffNetwork_Ca
lls','Data_Volume_Used','Calls_Made','SubscriptionPeriodInDays','ComplaintCount']].descri
be()

#Creating basic summaries for the Categorical variables

dfSATO[['UserType','Status','Network_Spectrum']].describe()

#Value Counts for UserType
dfSATO.UserType.value_counts()

#Bar Plot
UserTypeCnts = dfSATO.UserType.value_counts()
y=UserTypeCnts.values
n=len(y)
x=np.arange(n)
plt.bar(x,y,width=.75, color = 'blue')
plt.ylabel('Counts')
plt.xticks(x,UserTypeCnts.index)
plt.xticks(x,UserTypeCnts.index,color='black',rotation='vertical')
plt.title('User Type Disribution in data')
plt.show()
```

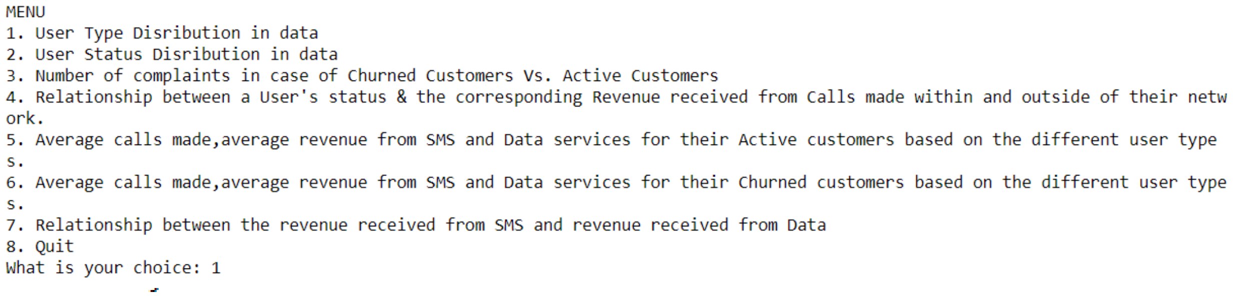
```
#Unique values for UserType
dfSATO.UserType.unique()

#Value Counts for Status
dfSATO.Status.value_counts()

#Bar Plot
StatusCnts = dfSATO.Status.value_counts()
y=StatusCnts.values
n=len(y)
x=np.arange(n)
plt.bar(x,y,width=.75, color = 'blue')
plt.ylabel('Counts')
plt.xticks(x,StatusCnts.index)
plt.xticks(x,StatusCnts.index,color='black',rotation='vertical')
plt.title('User Status Disribution in data')
plt.show()

#Unique values for Status
dfSATO.Status.unique()

#Value Counts for Network Spectrum
dfSATO.Network_Spectrum.value_counts()

#Unique values for Network Spectrum
dfSATO.Network_Spectrum.unique()

#Number of complaints in case of Churned Customers Vs. Active Customers
dfSATO.Calls_Made.groupby(dfSATO.Status).sum()

# Relationship between a User's status & the corresponding Revenue received from Calls
made within and outside of their network.

#Scatterplot with 2 series
x1 = dfSATO.Revenue_WithinNetwork_Calls[dfSATO.Status=='Active']
y1 = dfSATO.Revenue_OffNetwork_Calls[dfSATO.Status=='Active']
OneCorr = round(np.corrcoef(x1,y1)[0,1],3)
x2 = dfSATO.Revenue_WithinNetwork_Calls[dfSATO.Status=='Churned']
y2 = dfSATO.Revenue_OffNetwork_Calls[dfSATO.Status=='Churned']
TwoCorr = round(np.corrcoef(x2,y2)[0,1],3)
plt.scatter(x1,y1,color='red',label='Revenues from Active Customers-Corr:'+str(OneCorr))
plt.scatter(x2,y2,color='blue',label='Revenues from Churned Customers-
Corr:'+str(TwoCorr))
plt.title('Relationship between a User status & the corresponding Revenue received from
Calls')
plt.xlabel('Number Of Calls')
plt.ylabel('Revenue')
plt.legend()
plt.grid(True)
plt.show()

#Average calls made, average revenue from SMS and Data services for their Active
customers based on the different user types.
dfSATO[['Calls_Made','Revenue_SMS','Revenue_Data']][dfSATO.Status=='Active'].groupby(dfSA
TO.UserType).mean()

#Average calls made, average revenue from SMS and Data services for their Churned
customers based on the different user types.
dfSATO[['Calls_Made','Revenue_SMS','Revenue_Data']][dfSATO.Status=='Churned'].groupby(dfS
ATO.UserType).mean()

#Relationship between the revenue received from SMS and revenue received from Data
# Scatterplot with 1 series
x1 = dfSATO.Revenue_SMS
y1 = dfSATO.Revenue_Data
OneCorr = round(np.corrcoef(x1,y1)[0,1],3)
```

```
plt.scatter(x1,y1,color='blue',label='Correlation b/w revenue from SMS and 
Data:'+str(OneCorr))
plt.title('Relationship between the revenue received from SMS and revenue received from 
Data')
plt.xlabel('Revenue from SMS')
plt.ylabel('Revenue from Data')
plt.legend()
plt.grid(True)
plt.show()

#Writing the final dataframe to a .csv file for Part 2 of the project
dfSATO.to_csv('Part2_InputDataFrame.csv')
```

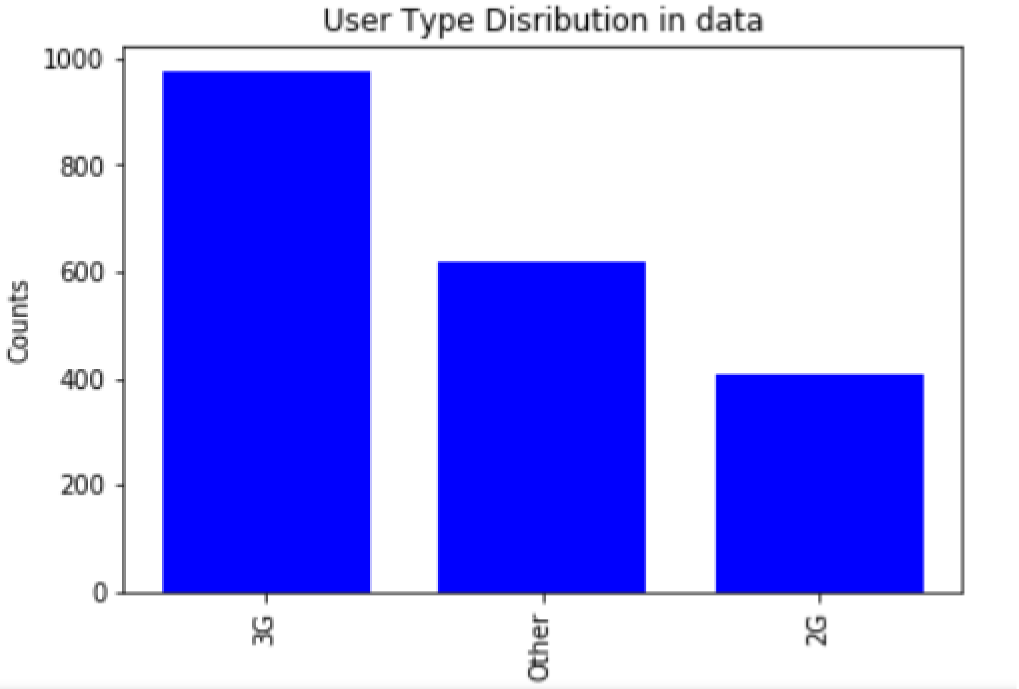
Part Two:

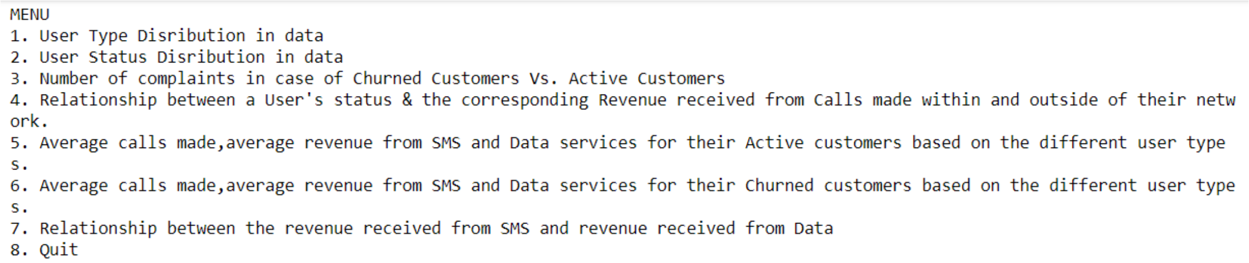
```
import pandas as pd
import numpy as np
from pandas import Series, DataFrame
from io import StringIO
import os
 import matplotlib.pyplot as plt

#Reading the data input from Part one of the code
dfSATO = pd.read_csv('Part2_InputDataFrame.csv')
dfSATO=dfSATO.drop('Unnamed: 0',axis=1)
dfSATO.head()

#Creating a Menu System
def runMenu(df):
    #Menu system
    quit = False
    while quit == False:
        print("\nMENU")
        print("1. User Type Disribution in data")
        print("2. User Status Disribution in data")
        print("3. Number of complaints in case of Churned Customers Vs. Active 
Customers")
        print("4. Relationship between a User's status & the corresponding Revenue 
received from Calls made within and outside of their network.")
        print("5. Average calls made,average revenue from SMS and Data services for their 
Active customers based on the different user types.")
        print("6. Average calls made,average revenue from SMS and Data services for their 
Churned customers based on the different user types.")
        print("7. Relationship between the revenue received from SMS and revenue received 
from Data")
        print("8. Quit")
        menu_choice = input("What is your choice: ")
        try:
            menu_choice = int(menu_choice) #convert to integer
        except: #If the user enters text or a symbol
            print("ERROR: Please enter 1, 2, 3, 4, 5, 6, 7 or 8")
            continue #returns to top of loop
        if menu_choice not in [1, 2, 3, 4,5,6,7,8]:
            print("ERROR: Please enter 1, 2, 3, 4, 5, 6, 7 or 8")
        else:
            if menu_choice == 1:
                UserTypeCnts = dfSATO.UserType.value_counts()
                y=UserTypeCnts.values
                n=len(y)
                x=np.arange(n)
                plt.bar(x,y,width=.75, color = 'blue')
                plt.ylabel('Counts')
                plt.xticks(x,UserTypeCnts.index)
                plt.xticks(x,UserTypeCnts.index,color='black',rotation='vertical')
                plt.title('User Type Disribution in data')
                plt.show()
            if menu_choice == 2:
                StatusCnts = dfSATO.Status.value_counts()
                y=StatusCnts.values
                n=len(y)
                x=np.arange(n)
                plt.bar(x,y,width=.75, color = 'blue')
                plt.ylabel('Counts')
                plt.xticks(x,StatusCnts.index)
                plt.xticks(x,StatusCnts.index,color='black',rotation='vertical')
                plt.title('User Status Disribution in data')
                plt.show()
            if menu_choice == 3:
                print("\nNumber of complaints in case of Churned Customers Vs. Active
Customers")
                print(dfSATO.Calls_Made.groupby(dfSATO.Status).sum())
            if menu_choice == 4:
                x1 = dfSATO.Revenue_WithinNetwork_Calls[dfSATO.Status=='Active']
                y1 = dfSATO.Revenue_OffNetwork_Calls[dfSATO.Status=='Active']
                OneCorr = round(np.corrcoef(x1,y1)[0,1],3)
                x2 = dfSATO.Revenue_WithinNetwork_Calls[dfSATO.Status=='Churned']
                y2 = dfSATO.Revenue_OffNetwork_Calls[dfSATO.Status=='Churned']
                TwoCorr = round(np.corrcoef(x2,y2)[0,1],3)
                plt.scatter(x1,y1,color='red',label='Revenues from Active Customers-
Corr:'+str(OneCorr))
                plt.scatter(x2,y2,color='blue',label='Revenues from Churned Customers-
Corr:'+str(TwoCorr))
                plt.title('Relationship between a User status & the corresponding Revenue
received from Calls')
                plt.xlabel('Number Of Calls')
                plt.ylabel('Revenue')
                plt.legend()
                plt.grid(True)
                plt.show()
            if menu_choice == 5:
                print("\nAverage calls made,average revenue from SMS and Data services
for their Active customers based on the different user types.")

print(dfSATO[['Calls_Made','Revenue_SMS','Revenue_Data']][dfSATO.Status=='Active'].groupb
y(dfSATO.UserType).mean())
            if menu_choice == 6:
                print("\nAverage calls made,average revenue from SMS and Data services
for their Churned customers based on the different user types.")

print(dfSATO[['Calls_Made','Revenue_SMS','Revenue_Data']][dfSATO.Status=='Churned'].group
by(dfSATO.UserType).mean())
            if menu_choice == 7:
                x1 = dfSATO.Revenue_SMS
                y1 = dfSATO.Revenue_Data
                OneCorr = round(np.corrcoef(x1,y1)[0,1],3)
                plt.scatter(x1,y1,color='blue',label='Correlation b/w revenue from SMS
and Data:'+str(OneCorr))
                plt.title('Relationship between the revenue received from SMS and revenue
received from Data')
                plt.xlabel('Revenue from SMS')
                plt.ylabel('Revenue from Data')
                plt.legend()
                plt.grid(True)
                plt.show()
            if menu_choice == 8:
                quit = True
runMenu(dfSATO)
```

Appendix A: Screenshots of Part Two Code

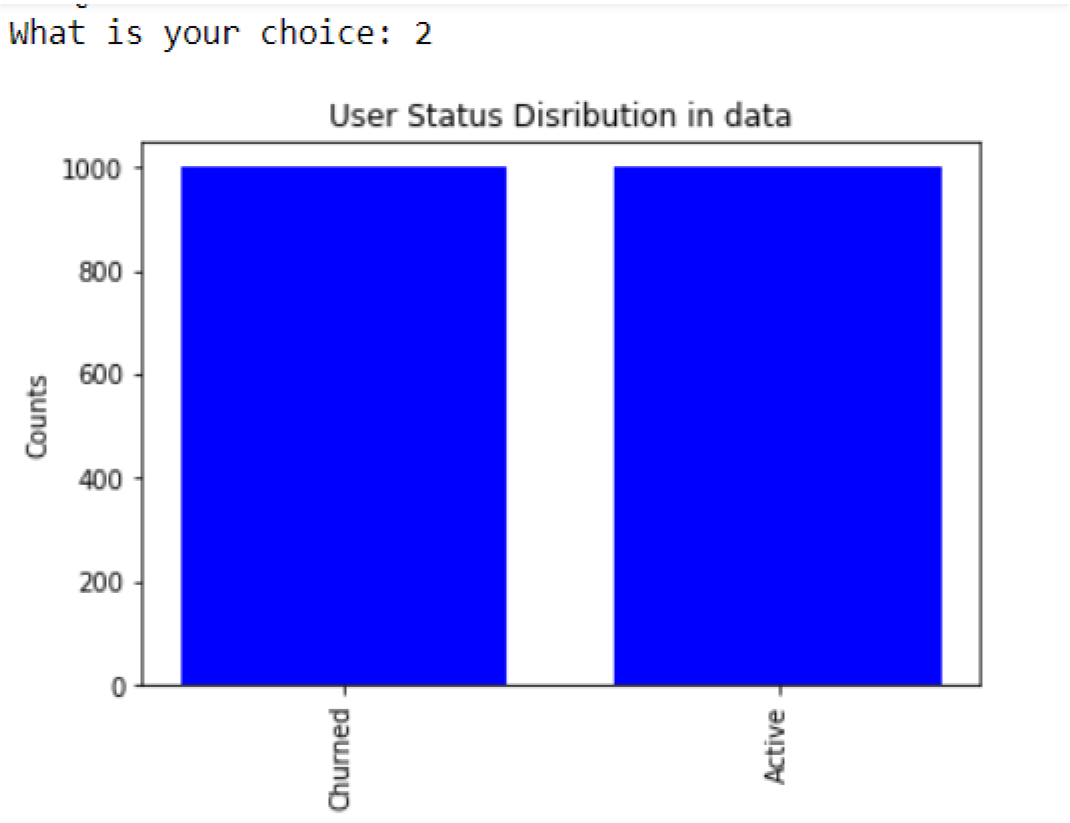


```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
What is your choice: 1
```

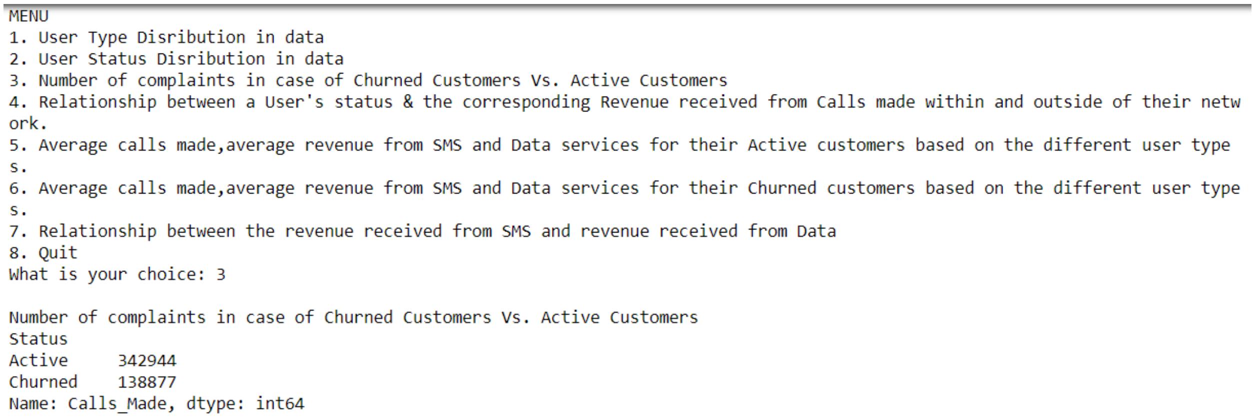




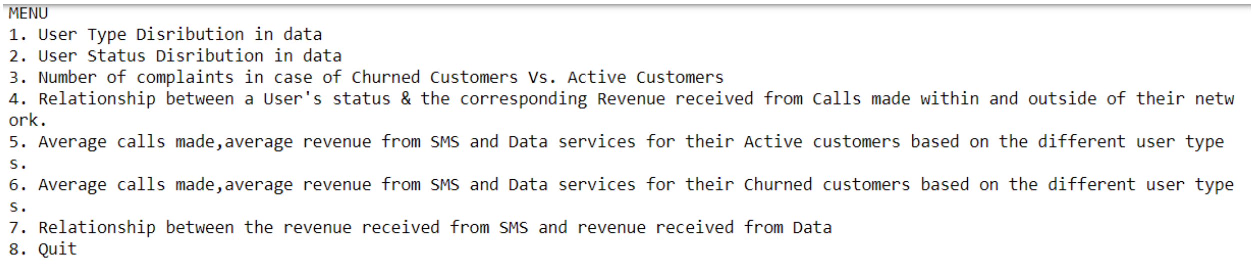
```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
```



```
What is your choice: 2
```

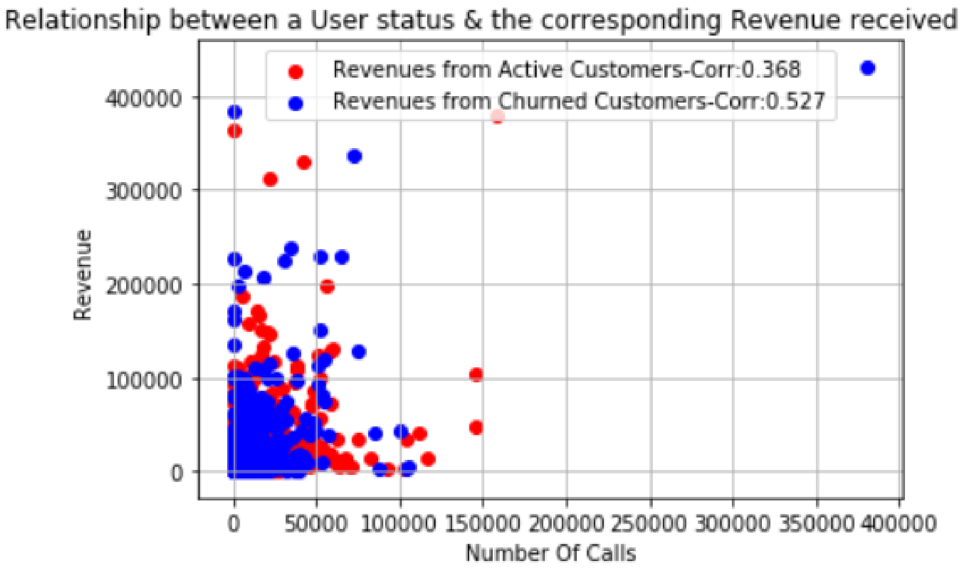


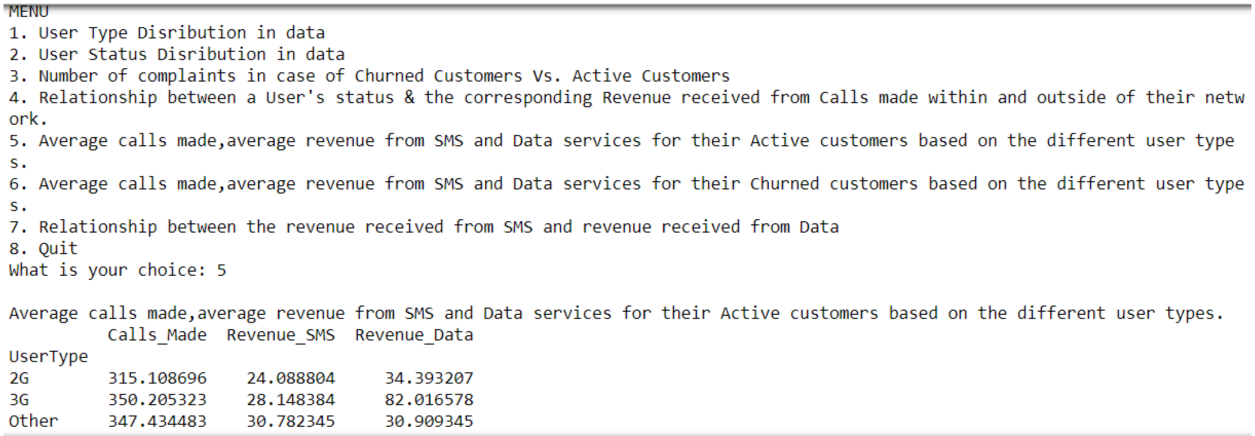
```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
What is your choice: 3

Number of complaints in case of Churned Customers Vs. Active Customers
Status
Active      342944
Churned     138877
Name: Calls_Made, dtype: int64
```



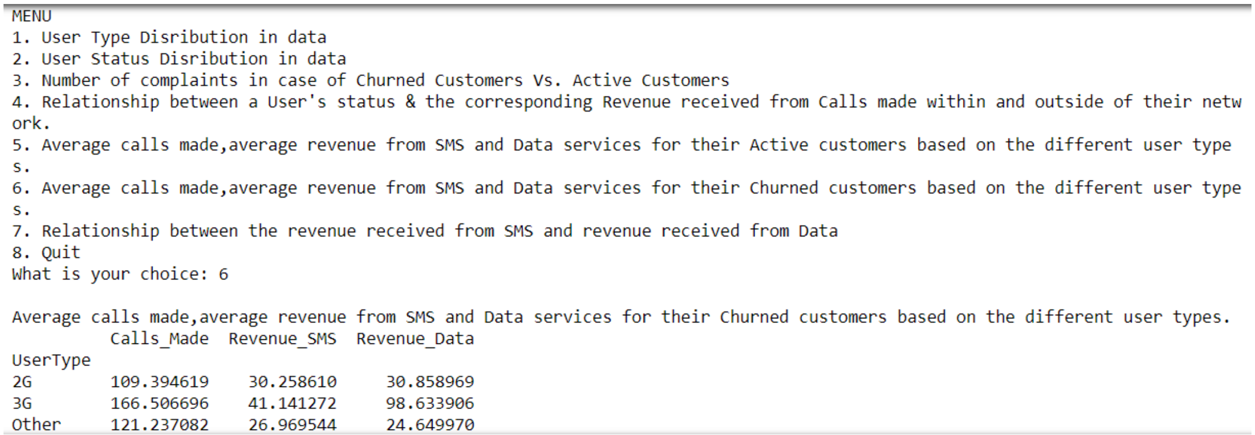
```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
```

```
What is your choice: 4
```

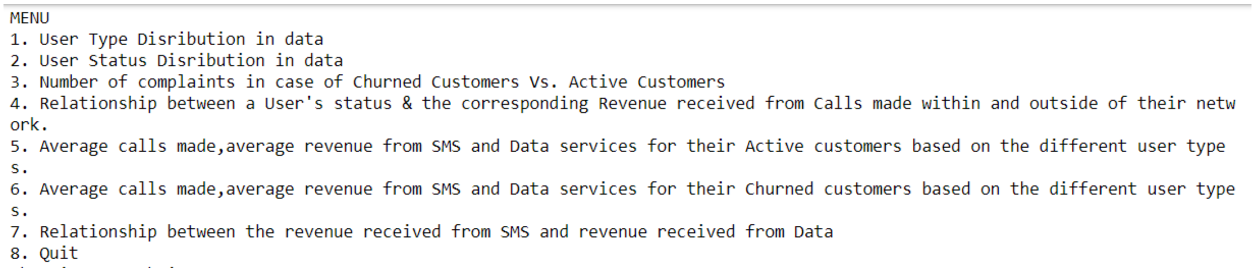




```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
What is your choice: 5

Average calls made,average revenue from SMS and Data services for their Active customers based on the different user types.
          Calls_Made  Revenue_SMS  Revenue_Data
UserType
2G        315.108696    24.088804     34.393207
3G        350.205323    28.148384     82.016578
Other     347.434483    30.782345     30.909345
```

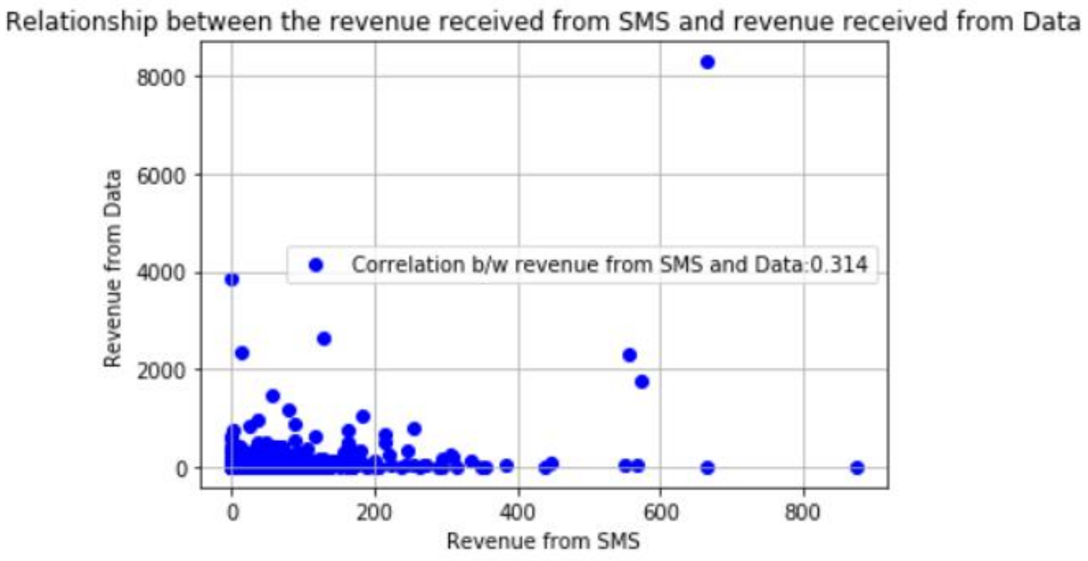


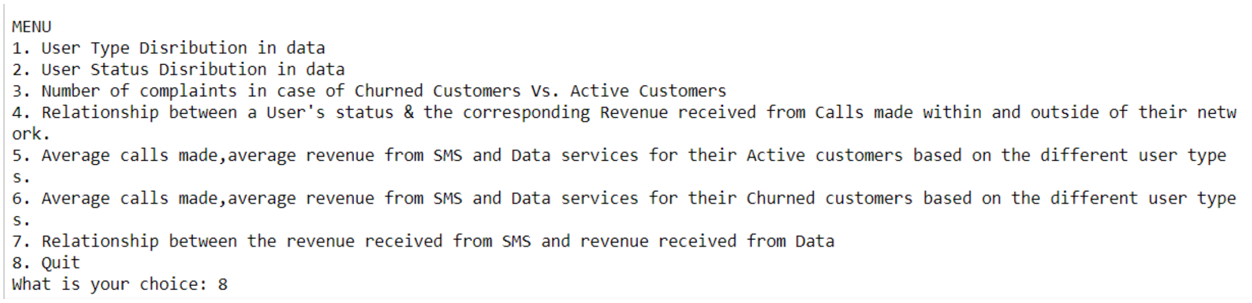
```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
What is your choice: 6

Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user types.
          Calls_Made  Revenue_SMS  Revenue_Data
UserType
2G        109.394619    30.258610     30.858969
3G        166.506696    41.141272     98.633906
Other     121.237082    26.969544     24.649970
```



```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
```

```
What is your choice: 7
```





```
MENU
1. User Type Disribution in data
2. User Status Disribution in data
3. Number of complaints in case of Churned Customers Vs. Active Customers
4. Relationship between a User's status & the corresponding Revenue received from Calls made within and outside of their netw
ork.
5. Average calls made,average revenue from SMS and Data services for their Active customers based on the different user type
s.
6. Average calls made,average revenue from SMS and Data services for their Churned customers based on the different user type
s.
7. Relationship between the revenue received from SMS and revenue received from Data
8. Quit
What is your choice: 8
```