Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
plt.rcParams['figure.figsize']=[15,8]
```

Set Options

```
pd.options.display.max_columns = None
pd.options.display.max_rows = None

pd.options.display.float_format = '{:.6f}'.format
```

Read data

```
df = pd.read_csv('cell2celltrain.csv')
df.head()
```

```
   CustomerID Churn  MonthlyRevenue  MonthlyMinutes
TotalRecurringCharge  \
0     3000002   Yes       24.000000      219.000000
22.000000
1     3000010   Yes       16.990000       10.000000
17.000000
2     3000014    No       38.000000        8.000000
38.000000
3     3000022    No       82.280000     1312.000000
75.000000
4     3000026   Yes       17.140000        0.000000
17.000000

   DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
0               0.250000        0.000000      0.000000         -
157.000000
1               0.000000        0.000000      0.000000          -
4.000000
2               0.000000        0.000000      0.000000          -
2.000000
3               1.240000        0.000000      0.000000
157.000000
4               0.000000        0.000000      0.000000
0.000000

   PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls  \
```

```
0          -19.000000      0.700000      0.700000         6.300000
1            0.000000      0.300000      0.000000         2.700000
2            0.000000      0.000000      0.000000         0.000000
3            8.100000     52.000000      7.700000        76.000000
4           -0.200000      0.000000      0.000000         0.000000

   CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls  \
0           0.000000       0.000000      97.200000       0.000000
1           0.000000       0.000000       0.000000       0.000000
2           0.000000       0.000000       0.400000       0.300000
3           4.300000       1.300000     200.300000     370.300000
4           0.000000       0.000000       0.000000       0.000000

   InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
0      0.000000       58.000000          24.000000
1.300000
1      0.000000        5.000000           1.000000
0.300000
2      0.000000        1.300000           3.700000
0.000000
3    147.000000      555.700000         303.700000
59.700000
4      0.000000        0.000000           0.000000
0.000000

   CallForwardingCalls  CallWaitingCalls  MonthsInService  UniqueSubs
\
0             0.000000          0.300000               61           2

1             0.000000          0.000000               58           1

2             0.000000          0.000000               60           1

3             0.000000         22.700000               59           2

4             0.000000          0.000000               53           2


   ActiveSubs ServiceArea  Handsets  HandsetModels
CurrentEquipmentDays  \
0           1   SEAPOR503  2.000000       2.000000
361.000000
1           1   PITHOM412  2.000000       1.000000
1504.000000
2           1   MILMIL414  1.000000       1.000000
1812.000000
3           2   PITHOM412  9.000000       4.000000
458.000000
4           2   OKCTUL918  4.000000       3.000000
```

```
852.000000

     AgeHH1    AgeHH2 ChildrenInHH HandsetRefurbished
HandsetWebCapable  \
0 62.000000  0.000000           No                 No
Yes
1 40.000000 42.000000          Yes                 No
No
2 26.000000 26.000000          Yes                 No
No
3 30.000000  0.000000           No                 No
Yes
4 46.000000 54.000000           No                 No
No

  TruckOwner RVOwner Homeownership BuysViaMailOrder
RespondsToMailOffers  \
0         No      No         Known               Yes
Yes
1         No      No         Known               Yes
Yes
2         No      No       Unknown                No
No
3         No      No         Known               Yes
Yes
4         No      No         Known               Yes
Yes

  OptOutMailings NonUSTravel OwnsComputer HasCreditCard
RetentionCalls  \
0             No           No          Yes            Yes
1
1             No           No          Yes            Yes
0
2             No           No           No            Yes
0
3             No           No           No            Yes
0
4             No           No          Yes            Yes
0

   RetentionOffersAccepted NewCellphoneUser NotNewCellphoneUser  \
0                        0               No                  No
1                        0              Yes                  No
2                        0              Yes                  No
3                        0              Yes                  No
4                        0               No                 Yes

   ReferralsMadeBySubscriber  IncomeGroup OwnsMotorcycle  \
0                          0            4             No
```

```
1                             0            5              No
2                             0            6              No
3                             0            6              No
4                             0            9              No

   AdjustmentsToCreditRating HandsetPrice MadeCallToRetentionTeam  \
0                             0           30                     Yes
1                             0           30                      No
2                             0      Unknown                      No
3                             0           10                      No
4                             1           10                      No

  CreditRating PrizmCode    Occupation MaritalStatus
0    1-Highest  Suburban  Professional            No
1     4-Medium  Suburban  Professional           Yes
2       3-Good      Town        Crafts           Yes
3     4-Medium     Other         Other            No
4    1-Highest     Other  Professional           Yes
```

Data Analysis and Preparation

Data Dimension

```
df.shape
```

```
(51047, 58)
```

- There are 51047 rows and 58 columns in the dataset.

Check the datatypes

```
df.dtypes
```

```
CustomerID                    int64
Churn                        object
MonthlyRevenue              float64
MonthlyMinutes              float64
TotalRecurringCharge        float64
DirectorAssistedCalls       float64
OverageMinutes              float64
RoamingCalls                float64
PercChangeMinutes           float64
PercChangeRevenues          float64
DroppedCalls                float64
BlockedCalls                float64
UnansweredCalls             float64
CustomerCareCalls           float64
ThreewayCalls               float64
ReceivedCalls               float64
OutboundCalls               float64
```

```
InboundCalls                  float64
PeakCallsInOut                float64
OffPeakCallsInOut             float64
DroppedBlockedCalls           float64
CallForwardingCalls           float64
CallWaitingCalls              float64
MonthsInService                 int64
UniqueSubs                      int64
ActiveSubs                      int64
ServiceArea                    object
Handsets                      float64
HandsetModels                 float64
CurrentEquipmentDays          float64
AgeHH1                        float64
AgeHH2                        float64
ChildrenInHH                   object
HandsetRefurbished             object
HandsetWebCapable              object
TruckOwner                     object
RVOwner                        object
Homeownership                  object
BuysViaMailOrder               object
RespondsToMailOffers           object
OptOutMailings                 object
NonUSTravel                    object
OwnsComputer                   object
HasCreditCard                  object
RetentionCalls                  int64
RetentionOffersAccepted         int64
NewCellphoneUser               object
NotNewCellphoneUser            object
ReferralsMadeBySubscriber       int64
IncomeGroup                     int64
OwnsMotorcycle                 object
AdjustmentsToCreditRating       int64
HandsetPrice                   object
MadeCallToRetentionTeam        object
CreditRating                   object
PrizmCode                      object
Occupation                     object
MaritalStatus                  object
dtype: object
```

- All the datatypes are assigned correctly except the first one. CustomerID should be a categorical column since no mathematical operation can be done.

```
df.CustomerID = df.CustomerID.astype('object')
```

Checking the info of the dataframe

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51047 entries, 0 to 51046
Data columns (total 58 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   CustomerID                  51047 non-null  object 
 1   Churn                       51047 non-null  object 
 2   MonthlyRevenue              50891 non-null  float64
 3   MonthlyMinutes              50891 non-null  float64
 4   TotalRecurringCharge        50891 non-null  float64
 5   DirectorAssistedCalls       50891 non-null  float64
 6   OverageMinutes              50891 non-null  float64
 7   RoamingCalls                50891 non-null  float64
 8   PercChangeMinutes           50680 non-null  float64
 9   PercChangeRevenues          50680 non-null  float64
 10  DroppedCalls                51047 non-null  float64
 11  BlockedCalls                51047 non-null  float64
 12  UnansweredCalls             51047 non-null  float64
 13  CustomerCareCalls           51047 non-null  float64
 14  ThreewayCalls               51047 non-null  float64
 15  ReceivedCalls               51047 non-null  float64
 16  OutboundCalls               51047 non-null  float64
 17  InboundCalls                51047 non-null  float64
 18  PeakCallsInOut              51047 non-null  float64
 19  OffPeakCallsInOut           51047 non-null  float64
 20  DroppedBlockedCalls         51047 non-null  float64
 21  CallForwardingCalls         51047 non-null  float64
 22  CallWaitingCalls            51047 non-null  float64
 23  MonthsInService             51047 non-null  int64  
 24  UniqueSubs                  51047 non-null  int64  
 25  ActiveSubs                  51047 non-null  int64  
 26  ServiceArea                 51023 non-null  object 
 27  Handsets                    51046 non-null  float64
 28  HandsetModels               51046 non-null  float64
 29  CurrentEquipmentDays        51046 non-null  float64
 30  AgeHH1                      50138 non-null  float64
 31  AgeHH2                      50138 non-null  float64
 32  ChildrenInHH                51047 non-null  object 
 33  HandsetRefurbished          51047 non-null  object 
 34  HandsetWebCapable           51047 non-null  object 
 35  TruckOwner                  51047 non-null  object 
 36  RVOwner                     51047 non-null  object 
 37  Homeownership               51047 non-null  object 
 38  BuysViaMailOrder            51047 non-null  object 
 39  RespondsToMailOffers        51047 non-null  object 
 40  OptOutMailings              51047 non-null  object 
```

```
 41  NonUSTravel                 51047 non-null  object
 42  OwnsComputer                51047 non-null  object
 43  HasCreditCard               51047 non-null  object
 44  RetentionCalls              51047 non-null  int64
 45  RetentionOffersAccepted     51047 non-null  int64
 46  NewCellphoneUser            51047 non-null  object
 47  NotNewCellphoneUser         51047 non-null  object
 48  ReferralsMadeBySubscriber   51047 non-null  int64
 49  IncomeGroup                 51047 non-null  int64
 50  OwnsMotorcycle              51047 non-null  object
 51  AdjustmentsToCreditRating   51047 non-null  int64
 52  HandsetPrice                51047 non-null  object
 53  MadeCallToRetentionTeam     51047 non-null  object
 54  CreditRating                51047 non-null  object
 55  PrizmCode                   51047 non-null  object
 56  Occupation                  51047 non-null  object
 57  MaritalStatus               51047 non-null  object
dtypes: float64(26), int64(8), object(24)
memory usage: 22.6+ MB
```

Summary Statistics

```
df.describe()
```

```
       MonthlyRevenue  MonthlyMinutes  TotalRecurringCharge  \
count    50891.000000    50891.000000          50891.000000
mean        58.834492      525.653416             46.830088
std         44.507336      529.871063             23.848871
min         -6.170000        0.000000            -11.000000
25%         33.610000      158.000000             30.000000
50%         48.460000      366.000000             45.000000
75%         71.065000      723.000000             60.000000
max       1223.380000     7359.000000            400.000000

       DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
count           50891.000000    50891.000000  50891.000000
50680.000000
mean                0.895229       40.027785      1.236244          -
11.547908
std                 2.228546       96.588076      9.818294
257.514772
min                 0.000000        0.000000      0.000000       -
3875.000000
25%                 0.000000        0.000000      0.000000         -
83.000000
50%                 0.250000        3.000000      0.000000          -
5.000000
75%                 0.990000       41.000000      0.300000
66.000000
```

```
max                159.390000     4321.000000   1112.400000
5192.000000

       PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls
\
count        50680.000000  51047.000000  51047.000000     51047.000000

mean            -1.191985      6.011489      4.085672        28.288981

std             39.574915      9.043955     10.946905        38.876194

min          -1107.700000      0.000000      0.000000         0.000000

25%             -7.100000      0.700000      0.000000         5.300000

50%             -0.300000      3.000000      1.000000        16.300000

75%              1.600000      7.700000      3.700000        36.300000

max           2483.500000    221.700000    384.300000       848.700000


       CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls
\
count       51047.000000   51047.000000   51047.000000   51047.000000

mean            1.868999       0.298838     114.800121      25.377715

std             5.096138       1.168277     166.485896      35.209147

min             0.000000       0.000000       0.000000       0.000000

25%             0.000000       0.000000       8.300000       3.300000

50%             0.000000       0.000000      52.800000      13.700000

75%             1.700000       0.300000     153.500000      34.000000

max           327.300000      66.000000    2692.400000     644.300000


       InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
count  51047.000000    51047.000000       51047.000000
51047.000000
mean       8.178104       90.549515          67.650790
10.158003
std       16.665878      104.947470          92.752699
15.555284
min        0.000000        0.000000           0.000000
0.000000
```

```
25%        0.000000       23.000000          11.000000
1.700000
50%        2.000000       62.000000          35.700000
5.300000
75%        9.300000      121.300000          88.700000
12.300000
max      519.300000     2090.700000        1474.700000
411.700000

       CallForwardingCalls  CallWaitingCalls  MonthsInService
UniqueSubs  \
count          51047.000000      51047.000000     51047.000000
51047.000000
mean               0.012277          1.840504        18.756264
1.532157
std                0.594168          5.585129         9.800138
1.223384
min                0.000000          0.000000         6.000000
1.000000
25%                0.000000          0.000000        11.000000
1.000000
50%                0.000000          0.300000        16.000000
1.000000
75%                0.000000          1.300000        24.000000
2.000000
max               81.300000        212.700000        61.000000
196.000000

        ActiveSubs     Handsets  HandsetModels
CurrentEquipmentDays  \
count 51047.000000 51046.000000   51046.000000          51046.000000

mean      1.354340     1.805646       1.558751            380.545841

std       0.675477     1.331173       0.905932            253.801982

min       0.000000     1.000000       1.000000             -5.000000

25%       1.000000     1.000000       1.000000            205.000000

50%       1.000000     1.000000       1.000000            329.000000

75%       2.000000     2.000000       2.000000            515.000000

max      53.000000    24.000000      15.000000           1812.000000


            AgeHH1       AgeHH2  RetentionCalls
RetentionOffersAccepted  \
count 50138.000000 50138.000000    51047.000000
```

```
51047.000000
mean     31.338127    21.144142        0.037201
0.018277
std      22.094635    23.931368        0.206483
0.142458
min       0.000000     0.000000        0.000000
0.000000
25%       0.000000     0.000000        0.000000
0.000000
50%      36.000000     0.000000        0.000000
0.000000
75%      48.000000    42.000000        0.000000
0.000000
max      99.000000    99.000000        4.000000
3.000000

       ReferralsMadeBySubscriber  IncomeGroup
AdjustmentsToCreditRating
count                51047.000000 51047.000000
51047.000000
mean                     0.052070     4.324524
0.053911
std                      0.307592     3.138236
0.383147
min                      0.000000     0.000000
0.000000
25%                      0.000000     0.000000
0.000000
50%                      0.000000     5.000000
0.000000
75%                      0.000000     7.000000
0.000000
max                     35.000000     9.000000
25.000000
```

Inference:

- We are having the negative values in the dataset, Monthly revenue is in Negative which cannot be true.
- There a huge difference between our mean value and the maximum value, which means there are potential outliers in the data.
- In PercChangeMinutes column, we have our mean value also in negative terms.

```
df.describe(include = object)

        CustomerID  Churn ServiceArea ChildrenInHH HandsetRefurbished
\
count        51047  51047       51023        51047                51047

unique       51047      2         747            2                    2
```

```
top         3000002     No    NYCBRO917           No                   No

freq              1  36336         1684         38679                43956


       HandsetWebCapable TruckOwner RVOwner Homeownership
BuysViaMailOrder  \
count              51047      51047   51047          51047
51047
unique                 2          2       2              2
2
top                  Yes         No      No          Known
No
freq               46046      41524   46894          33987
32596

       RespondsToMailOffers OptOutMailings NonUSTravel OwnsComputer  \
count                 51047          51047       51047        51047
unique                    2              2           2            2
top                      No             No          No           No
freq                  31821          50295       48168        41583

       HasCreditCard NewCellphoneUser NotNewCellphoneUser
OwnsMotorcycle  \
count          51047             51047                51047
51047
unique             2                 2                    2
2
top              Yes               No                   No
No
freq           34503             41223                44012
50368

       HandsetPrice MadeCallToRetentionTeam CreditRating PrizmCode
Occupation  \
count         51047                   51047        51047     51047
51047
unique           16                       2            7         4
8
top         Unknown                      No       2-High     Other
Other
freq          28982                   49302        18993     24655
37637

       MaritalStatus
count          51047
unique             3
top          Unknown
freq           19700
```

Inference:

- ServiceArea - NYCBRO917, has is the most repeated value in that column.
- All the other variables except HandsetWebCapable, HasCreditCard has the highest frequency as no.
- The same can be visualized below using the relationship between the target variable and the categorical variable.

Duplicates

```
df[df.duplicated()]
```

```
Empty DataFrame
Columns: [CustomerID, Churn, MonthlyRevenue, MonthlyMinutes,
TotalRecurringCharge, DirectorAssistedCalls, OverageMinutes,
RoamingCalls, PercChangeMinutes, PercChangeRevenues, DroppedCalls,
BlockedCalls, UnansweredCalls, CustomerCareCalls, ThreewayCalls,
ReceivedCalls, OutboundCalls, InboundCalls, PeakCallsInOut,
OffPeakCallsInOut, DroppedBlockedCalls, CallForwardingCalls,
CallWaitingCalls, MonthsInService, UniqueSubs, ActiveSubs,
ServiceArea, Handsets, HandsetModels, CurrentEquipmentDays, AgeHH1,
AgeHH2, ChildrenInHH, HandsetRefurbished, HandsetWebCapable,
TruckOwner, RVOwner, Homeownership, BuysViaMailOrder,
RespondsToMailOffers, OptOutMailings, NonUSTravel, OwnsComputer,
HasCreditCard, RetentionCalls, RetentionOffersAccepted,
NewCellphoneUser, NotNewCellphoneUser, ReferralsMadeBySubscriber,
IncomeGroup, OwnsMotorcycle, AdjustmentsToCreditRating,
MadeCallToRetentionTeam, CreditRating, PrizmCode, Occupation,
MaritalStatus]
Index: []
```

```
# There are no duplicate values in the data.
```

Missing Values

```
Total = df.isnull().sum().sort_values(ascending=False)
Percent =
(df.isnull().sum()*100/df.isnull().count()).sort_values(ascending=Fals
e)
missing_data = pd.concat([Total, Percent], axis = 1, keys = ['Total',
'Percentage of Missing Values'])

missing_data
```

	Total	Percentage of Missing Values
AgeHH2	909	1.780712
AgeHH1	909	1.780712
PercChangeMinutes	367	0.718945
PercChangeRevenues	367	0.718945
MonthlyRevenue	156	0.305601
MonthlyMinutes	156	0.305601

TotalRecurringCharge	156	0.305601
DirectorAssistedCalls	156	0.305601
OverageMinutes	156	0.305601
RoamingCalls	156	0.305601
ServiceArea	24	0.047015
CurrentEquipmentDays	1	0.001959
Handsets	1	0.001959
HandsetModels	1	0.001959
OwnsComputer	0	0.000000
NonUSTravel	0	0.000000
BuysViaMailOrder	0	0.000000
OptOutMailings	0	0.000000
RespondsToMailOffers	0	0.000000
RetentionCalls	0	0.000000
HasCreditCard	0	0.000000
IncomeGroup	0	0.000000
RetentionOffersAccepted	0	0.000000
NewCellphoneUser	0	0.000000
NotNewCellphoneUser	0	0.000000
ReferralsMadeBySubscriber	0	0.000000
RVOwner	0	0.000000
OwnsMotorcycle	0	0.000000
AdjustmentsToCreditRating	0	0.000000
HandsetPrice	0	0.000000
MadeCallToRetentionTeam	0	0.000000
CreditRating	0	0.000000
PrizmCode	0	0.000000
Occupation	0	0.000000
Homeownership	0	0.000000
CustomerID	0	0.000000
TruckOwner	0	0.000000
HandsetWebCapable	0	0.000000
DroppedCalls	0	0.000000
BlockedCalls	0	0.000000
UnansweredCalls	0	0.000000
CustomerCareCalls	0	0.000000
ThreewayCalls	0	0.000000
ReceivedCalls	0	0.000000
OutboundCalls	0	0.000000
InboundCalls	0	0.000000
PeakCallsInOut	0	0.000000
OffPeakCallsInOut	0	0.000000
DroppedBlockedCalls	0	0.000000
CallForwardingCalls	0	0.000000
CallWaitingCalls	0	0.000000
MonthsInService	0	0.000000
UniqueSubs	0	0.000000
ActiveSubs	0	0.000000
Churn	0	0.000000

```
ChildrenInHH                    0                      0.000000
HandsetRefurbished              0                      0.000000
MaritalStatus                   0                      0.000000
```

```
mis_index = missing_data[missing_data['Total']>0].index
mis_index
```

```
Index(['AgeHH2', 'AgeHH1', 'PercChangeMinutes', 'PercChangeRevenues',
       'MonthlyRevenue', 'MonthlyMinutes', 'TotalRecurringCharge',
       'DirectorAssistedCalls', 'OverageMinutes', 'RoamingCalls',
       'ServiceArea', 'CurrentEquipmentDays', 'Handsets',
'HandsetModels'],
      dtype='object')
```

- In the above columns we have missing data. There are only very few null values in the dataset. Either we can drop the null values or replace them with mean/median/mode.
- Almost all the columns are numerical except the `ServiceArea`. We can replace them median value and the ServiceArea column with mode value.

```
# Replacing the missing from the numerical columns

df[['AgeHH2', 'AgeHH1', 'PercChangeMinutes', 'PercChangeRevenues',
   'MonthlyRevenue', 'MonthlyMinutes', 'TotalRecurringCharge',
   'DirectorAssistedCalls', 'OverageMinutes', 'RoamingCalls',
'Handsets', 'HandsetModels','CurrentEquipmentDays']] = df[['AgeHH2',
'AgeHH1', 'PercChangeMinutes', 'PercChangeRevenues',
   'MonthlyRevenue', 'MonthlyMinutes', 'TotalRecurringCharge',
   'DirectorAssistedCalls', 'OverageMinutes', 'RoamingCalls',
'Handsets',
'HandsetModels','CurrentEquipmentDays']].fillna(df[['AgeHH2',
'AgeHH1', 'PercChangeMinutes', 'PercChangeRevenues',
   'MonthlyRevenue', 'MonthlyMinutes', 'TotalRecurringCharge',
   'DirectorAssistedCalls', 'OverageMinutes', 'RoamingCalls',
'Handsets', 'HandsetModels','CurrentEquipmentDays']].median())

# Replacing the null values from categorical columns.

df['ServiceArea'] =
df['ServiceArea'].fillna(df['ServiceArea'].ffill())

Total = df.isnull().sum().sort_values(ascending=False)
Percent =
(df.isnull().sum()*100/df.isnull().count()).sort_values(ascending=Fals
e)
missing_data = pd.concat([Total, Percent], axis = 1, keys = ['Total',
'Percentage of Missing Values'])

missing_data
```

```
                           Total  Percentage of Missing Values
CustomerID                     0                      0.000000
```

HasCreditCard	0	0.000000
AgeHH2	0	0.000000
ChildrenInHH	0	0.000000
HandsetRefurbished	0	0.000000
HandsetWebCapable	0	0.000000
TruckOwner	0	0.000000
RVOwner	0	0.000000
Homeownership	0	0.000000
BuysViaMailOrder	0	0.000000
RespondsToMailOffers	0	0.000000
OptOutMailings	0	0.000000
NonUSTravel	0	0.000000
OwnsComputer	0	0.000000
RetentionCalls	0	0.000000
Churn	0	0.000000
RetentionOffersAccepted	0	0.000000
NewCellphoneUser	0	0.000000
NotNewCellphoneUser	0	0.000000
ReferralsMadeBySubscriber	0	0.000000
IncomeGroup	0	0.000000
OwnsMotorcycle	0	0.000000
AdjustmentsToCreditRating	0	0.000000
HandsetPrice	0	0.000000
MadeCallToRetentionTeam	0	0.000000
CreditRating	0	0.000000
PrizmCode	0	0.000000
Occupation	0	0.000000
AgeHH1	0	0.000000
CurrentEquipmentDays	0	0.000000
HandsetModels	0	0.000000
Handsets	0	0.000000
MonthlyRevenue	0	0.000000
MonthlyMinutes	0	0.000000
TotalRecurringCharge	0	0.000000
DirectorAssistedCalls	0	0.000000
OverageMinutes	0	0.000000
RoamingCalls	0	0.000000
PercChangeMinutes	0	0.000000
PercChangeRevenues	0	0.000000
DroppedCalls	0	0.000000
BlockedCalls	0	0.000000
UnansweredCalls	0	0.000000
CustomerCareCalls	0	0.000000
ThreewayCalls	0	0.000000
ReceivedCalls	0	0.000000
OutboundCalls	0	0.000000
InboundCalls	0	0.000000
PeakCallsInOut	0	0.000000
OffPeakCallsInOut	0	0.000000

```
DroppedBlockedCalls            0                      0.000000
CallForwardingCalls            0                      0.000000
CallWaitingCalls               0                      0.000000
MonthsInService                0                      0.000000
UniqueSubs                     0                      0.000000
ActiveSubs                     0                      0.000000
ServiceArea                    0                      0.000000
MaritalStatus                  0                      0.000000

#Hereby all the missing values has been replaced.
```

- Lets divide the data into numerical and categorical datasets, so that it becomes easy to sort the numerical and categorical values separately.

```
df_num = df.select_dtypes(include = [np.number])
df_num.head()

   MonthlyRevenue  MonthlyMinutes  TotalRecurringCharge  \
0       24.000000      219.000000             22.000000
1       16.990000       10.000000             17.000000
2       38.000000        8.000000             38.000000
3       82.280000     1312.000000             75.000000
4       17.140000        0.000000             17.000000

   DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
0               0.250000        0.000000      0.000000         -
157.000000
1               0.000000        0.000000      0.000000           -
4.000000
2               0.000000        0.000000      0.000000           -
2.000000
3               1.240000        0.000000      0.000000
157.000000
4               0.000000        0.000000      0.000000
0.000000

   PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls  \
0          -19.000000      0.700000      0.700000         6.300000
1            0.000000      0.300000      0.000000         2.700000
2            0.000000      0.000000      0.000000         0.000000
3            8.100000     52.000000      7.700000        76.000000
4           -0.200000      0.000000      0.000000         0.000000

   CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls  \
0           0.000000       0.000000      97.200000       0.000000
1           0.000000       0.000000       0.000000       0.000000
2           0.000000       0.000000       0.400000       0.300000
3           4.300000       1.300000     200.300000     370.300000
4           0.000000       0.000000       0.000000       0.000000
```

```
   InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
0      0.000000       58.000000          24.000000
1.300000
1      0.000000        5.000000           1.000000
0.300000
2      0.000000        1.300000           3.700000
0.000000
3    147.000000      555.700000         303.700000
59.700000
4      0.000000        0.000000           0.000000
0.000000

   CallForwardingCalls  CallWaitingCalls  MonthsInService  UniqueSubs
\
0             0.000000          0.300000               61           2

1             0.000000          0.000000               58           1

2             0.000000          0.000000               60           1

3             0.000000         22.700000               59           2

4             0.000000          0.000000               53           2


   ActiveSubs  Handsets  HandsetModels  CurrentEquipmentDays    AgeHH1
\
0           1  2.000000       2.000000            361.000000 62.000000

1           1  2.000000       1.000000           1504.000000 40.000000

2           1  1.000000       1.000000           1812.000000 26.000000

3           2  9.000000       4.000000            458.000000 30.000000

4           2  4.000000       3.000000            852.000000 46.000000


     AgeHH2  RetentionCalls  RetentionOffersAccepted  \
0  0.000000               1                        0
1 42.000000               0                        0
2 26.000000               0                        0
3  0.000000               0                        0
4 54.000000               0                        0

   ReferralsMadeBySubscriber  IncomeGroup  AdjustmentsToCreditRating
0                          0            4                          0
1                          0            5                          0
2                          0            6                          0
```

```
3                          0           6                          0
4                          0           9                          1

df_num.columns

Index(['MonthlyRevenue', 'MonthlyMinutes', 'TotalRecurringCharge',
       'DirectorAssistedCalls', 'OverageMinutes', 'RoamingCalls',
       'PercChangeMinutes', 'PercChangeRevenues', 'DroppedCalls',
       'BlockedCalls', 'UnansweredCalls', 'CustomerCareCalls',
'ThreewayCalls',
       'ReceivedCalls', 'OutboundCalls', 'InboundCalls',
'PeakCallsInOut',
       'OffPeakCallsInOut', 'DroppedBlockedCalls',
'CallForwardingCalls',
       'CallWaitingCalls', 'MonthsInService', 'UniqueSubs',
'ActiveSubs',
       'Handsets', 'HandsetModels', 'CurrentEquipmentDays', 'AgeHH1',
'AgeHH2',
       'RetentionCalls', 'RetentionOffersAccepted',
       'ReferralsMadeBySubscriber', 'IncomeGroup',
       'AdjustmentsToCreditRating'],
      dtype='object')

df_cat = df.select_dtypes(include = 'O')
df_cat.head()

  CustomerID Churn ServiceArea ChildrenInHH HandsetRefurbished  \
0    3000002   Yes   SEAPOR503           No                 No
1    3000010   Yes   PITHOM412          Yes                 No
2    3000014    No   MILMIL414          Yes                 No
3    3000022    No   PITHOM412           No                 No
4    3000026   Yes   OKCTUL918           No                 No

  HandsetWebCapable TruckOwner RVOwner Homeownership BuysViaMailOrder
\
0               Yes         No      No         Known               Yes

1                No         No      No         Known               Yes

2                No         No      No       Unknown                No

3               Yes         No      No         Known               Yes

4                No         No      No         Known               Yes


  RespondsToMailOffers OptOutMailings NonUSTravel OwnsComputer
HasCreditCard  \
0                  Yes             No          No          Yes
Yes
1                  Yes             No          No          Yes
```

```
Yes
2                   No              No           No            No
Yes
3                  Yes              No           No            No
Yes
4                  Yes              No           No           Yes
Yes

  NewCellphoneUser NotNewCellphoneUser OwnsMotorcycle HandsetPrice  \
0               No                  No             No           30
1              Yes                  No             No           30
2              Yes                  No             No      Unknown
3              Yes                  No             No           10
4               No                 Yes             No           10

  MadeCallToRetentionTeam CreditRating PrizmCode    Occupation
MaritalStatus
0                     Yes    1-Highest  Suburban  Professional
No
1                      No     4-Medium  Suburban  Professional
Yes
2                      No       3-Good      Town        Crafts
Yes
3                      No     4-Medium     Other         Other
No
4                      No    1-Highest     Other  Professional
Yes
```

Correlation

```
df.corr()

                          MonthlyRevenue  MonthlyMinutes  \
MonthlyRevenue                  1.000000        0.712968
MonthlyMinutes                  0.712968        1.000000
TotalRecurringCharge            0.629835        0.584014
DirectorAssistedCalls           0.407372        0.392377
OverageMinutes                  0.787865        0.571582
RoamingCalls                    0.303638        0.112722
PercChangeMinutes              -0.027392       -0.045125
PercChangeRevenues             -0.016557       -0.040702
DroppedCalls                    0.429911        0.593126
BlockedCalls                    0.232162        0.318009
UnansweredCalls                 0.449610        0.645437
CustomerCareCalls               0.189039        0.375658
ThreewayCalls                   0.200670        0.288409
ReceivedCalls                   0.604635        0.828031
OutboundCalls                   0.494493        0.693354
InboundCalls                    0.373571        0.557527
PeakCallsInOut                  0.640922        0.770326
```

OffPeakCallsInOut	0.471562	0.763910
DroppedBlockedCalls	0.416785	0.572790
CallForwardingCalls	0.011413	0.018664
CallWaitingCalls	0.468730	0.614813
MonthsInService	-0.002305	-0.068506
UniqueSubs	-0.014309	-0.029803
ActiveSubs	-0.041631	-0.065828
Handsets	0.242457	0.301769
HandsetModels	0.235636	0.291961
CurrentEquipmentDays	-0.217373	-0.310101
AgeHH1	-0.105784	-0.160494
AgeHH2	-0.106391	-0.142221
RetentionCalls	0.011142	0.009560
RetentionOffersAccepted	0.015402	0.014874
ReferralsMadeBySubscriber	0.018933	0.067452
IncomeGroup	-0.081593	-0.144348
AdjustmentsToCreditRating	0.034164	0.045164

	TotalRecurringCharge	DirectorAssistedCalls
MonthlyRevenue	0.629835	0.407372
MonthlyMinutes	0.584014	0.392377
TotalRecurringCharge	1.000000	0.311077
DirectorAssistedCalls	0.311077	1.000000
OverageMinutes	0.202951	0.287366
RoamingCalls	0.073767	0.070091
PercChangeMinutes	-0.016129	-0.026073
PercChangeRevenues	-0.025502	-0.016037
DroppedCalls	0.352390	0.263450
BlockedCalls	0.178886	0.139936
UnansweredCalls	0.357709	0.304613
CustomerCareCalls	0.152470	0.093301
ThreewayCalls	0.148592	0.132401
ReceivedCalls	0.480432	0.280689
OutboundCalls	0.408444	0.315140
InboundCalls	0.306259	0.186031

```
PeakCallsInOut                      0.550819              0.418529
OffPeakCallsInOut                   0.367319              0.273295
DroppedBlockedCalls                 0.333693              0.252728
CallForwardingCalls                 0.014085              0.002323
CallWaitingCalls                    0.354574              0.305112
MonthsInService                    -0.047954              0.017861
UniqueSubs                         -0.020661             -0.002697
ActiveSubs                         -0.056717             -0.014810
Handsets                            0.233702              0.184442
HandsetModels                       0.225111              0.172056
CurrentEquipmentDays               -0.247995             -0.121587
AgeHH1                             -0.103210             -0.057514
AgeHH2                             -0.101194             -0.058179
RetentionCalls                     -0.022028              0.008477
RetentionOffersAccepted            -0.002306              0.013311
ReferralsMadeBySubscriber           0.036715             -0.001083
IncomeGroup                        -0.077023             -0.017408
AdjustmentsToCreditRating           0.043396              0.022424


                          OverageMinutes  RoamingCalls
PercChangeMinutes  \
MonthlyRevenue                  0.787865      0.303638          -
0.027392
MonthlyMinutes                  0.571582      0.112722          -
0.045125
TotalRecurringCharge            0.202951      0.073767          -
0.016129
DirectorAssistedCalls           0.287366      0.070091          -
0.026073
OverageMinutes                  1.000000      0.064085          -
0.028412
RoamingCalls                    0.064085      1.000000          -
```

```
0.053848
PercChangeMinutes                -0.028412     -0.053848
1.000000
PercChangeRevenues               -0.018636     -0.090721
0.609519
DroppedCalls                      0.331685      0.068502          -
0.081111
BlockedCalls                      0.190500      0.029203          -
0.055366
UnansweredCalls                   0.362670      0.039327          -
0.086126
CustomerCareCalls                 0.137771      0.020460          -
0.039963
ThreewayCalls                     0.163859      0.024516          -
0.028904
ReceivedCalls                     0.515630      0.045897          -
0.106330
OutboundCalls                     0.405672      0.045332          -
0.063543
InboundCalls                      0.319870      0.019837          -
0.070678
PeakCallsInOut                    0.517877      0.062661          -
0.109553
OffPeakCallsInOut                 0.390641      0.031633          -
0.089024
DroppedBlockedCalls               0.329256      0.060832          -
0.087144
CallForwardingCalls               0.003702      0.003243          -
0.004799
CallWaitingCalls                  0.456640      0.035981          -
0.129588
MonthsInService                   0.001236     -0.010866
0.004069
UniqueSubs                       -0.002879     -0.003860
0.001805
ActiveSubs                       -0.012533     -0.003190
0.008515
Handsets                          0.142617      0.024343          -
0.012494
HandsetModels                     0.143260      0.022872          -
0.012229
CurrentEquipmentDays             -0.124260     -0.029030          -
0.006167
AgeHH1                           -0.063302     -0.010196
0.014070
AgeHH2                           -0.063191     -0.013209
0.003614
RetentionCalls                    0.014811     -0.002067          -
0.021247
```

```
RetentionOffersAccepted           0.009279     -0.000816           -
0.008209
ReferralsMadeBySubscriber         0.010081     -0.006843           -
0.001978
IncomeGroup                      -0.049999     -0.015228
0.011412
AdjustmentsToCreditRating         0.020464      0.004398           -
0.004303

                            PercChangeRevenues  DroppedCalls
BlockedCalls  \
MonthlyRevenue                       -0.016557      0.429911
0.232162
MonthlyMinutes                       -0.040702      0.593126
0.318009
TotalRecurringCharge                 -0.025502      0.352390
0.178886
DirectorAssistedCalls                -0.016037      0.263450
0.139936
OverageMinutes                       -0.018636      0.331685
0.190500
RoamingCalls                         -0.090721      0.068502
0.029203
PercChangeMinutes                     0.609519     -0.081111      -
0.055366
PercChangeRevenues                    1.000000     -0.037290      -
0.033033
DroppedCalls                         -0.037290      1.000000
0.185124
BlockedCalls                         -0.033033      0.185124
1.000000
UnansweredCalls                      -0.055766      0.543263
0.255713
CustomerCareCalls                    -0.032508      0.294747
0.199638
ThreewayCalls                         0.013153      0.264223
0.262207
ReceivedCalls                        -0.079477      0.517854
0.260401
OutboundCalls                        -0.041632      0.564465
0.230012
InboundCalls                         -0.047392      0.394289
0.188910
PeakCallsInOut                       -0.080690      0.574093
0.301752
OffPeakCallsInOut                    -0.052543      0.601072
0.330684
DroppedBlockedCalls                  -0.045726      0.715420
0.816328
```

```
CallForwardingCalls                  -0.002364      0.004301
0.015123
CallWaitingCalls                     -0.093941      0.396783
0.286784
MonthsInService                      -0.006921     -0.046175     -
0.072684
UniqueSubs                            0.002532     -0.021223     -
0.016170
ActiveSubs                            0.005129     -0.049663     -
0.026820
Handsets                             -0.018096      0.225186
0.088458
HandsetModels                        -0.017873      0.217869
0.085308
CurrentEquipmentDays                  0.005798     -0.216536     -
0.124673
AgeHH1                                0.006732     -0.117797     -
0.046762
AgeHH2                                0.002391     -0.107388     -
0.033732
RetentionCalls                       -0.019110      0.020308
0.010841
RetentionOffersAccepted              -0.012163      0.014858
0.008350
ReferralsMadeBySubscriber            -0.001785      0.029546
0.023848
IncomeGroup                           0.005717     -0.096316     -
0.065375
AdjustmentsToCreditRating            -0.005524      0.031039
0.004167

                           UnansweredCalls  CustomerCareCalls
ThreewayCalls  \
MonthlyRevenue                    0.449610           0.189039
0.200670
MonthlyMinutes                    0.645437           0.375658
0.288409
TotalRecurringCharge              0.357709           0.152470
0.148592
DirectorAssistedCalls             0.304613           0.093301
0.132401
OverageMinutes                    0.362670           0.137771
0.163859
RoamingCalls                      0.039327           0.020460
0.024516
PercChangeMinutes                -0.086126          -0.039963      -
0.028904
PercChangeRevenues               -0.055766          -0.032508
0.013153
```

```
DroppedCalls                  0.543263          0.294747
0.264223
BlockedCalls                  0.255713          0.199638
0.262207
UnansweredCalls               1.000000          0.404060
0.301705
CustomerCareCalls             0.404060          1.000000
0.242993
ThreewayCalls                 0.301705          0.242993
1.000000
ReceivedCalls                 0.550330          0.298607
0.243663
OutboundCalls                 0.575704          0.285395
0.220351
InboundCalls                  0.460582          0.215370
0.165698
PeakCallsInOut                0.674665          0.280366
0.256693
OffPeakCallsInOut             0.719626          0.404053
0.315158
DroppedBlockedCalls           0.498485          0.321728
0.340332
CallForwardingCalls           0.015151          0.013506
0.001580
CallWaitingCalls              0.488569          0.212042
0.221297
MonthsInService              -0.065783         -0.107552      -
0.057165
UniqueSubs                   -0.019853         -0.051419      -
0.018429
ActiveSubs                   -0.044963         -0.086749      -
0.031942
Handsets                      0.252150          0.108153
0.093039
HandsetModels                 0.243701          0.097099
0.084583
CurrentEquipmentDays         -0.243193         -0.169502      -
0.111433
AgeHH1                       -0.119593         -0.099774      -
0.034702
AgeHH2                       -0.101761         -0.082794      -
0.042406
RetentionCalls                0.030330          0.026069
0.007382
RetentionOffersAccepted       0.025852          0.022238
0.006319
ReferralsMadeBySubscriber     0.047777          0.044990
0.017090
IncomeGroup                  -0.129093         -0.119494      -
```

```
0.046028
AdjustmentsToCreditRating          0.036239           0.019600
0.004713

                            ReceivedCalls  OutboundCalls  InboundCalls
\
MonthlyRevenue                   0.604635       0.494493      0.373571

MonthlyMinutes                   0.828031       0.693354      0.557527

TotalRecurringCharge             0.480432       0.408444      0.306259

DirectorAssistedCalls            0.280689       0.315140      0.186031

OverageMinutes                   0.515630       0.405672      0.319870

RoamingCalls                     0.045897       0.045332      0.019837

PercChangeMinutes               -0.106330      -0.063543     -0.070678

PercChangeRevenues              -0.079477      -0.041632     -0.047392

DroppedCalls                     0.517854       0.564465      0.394289

BlockedCalls                     0.260401       0.230012      0.188910

UnansweredCalls                  0.550330       0.575704      0.460582

CustomerCareCalls                0.298607       0.285395      0.215370

ThreewayCalls                    0.243663       0.220351      0.165698

ReceivedCalls                    1.000000       0.653062      0.619394

OutboundCalls                    0.653062       1.000000      0.724416

InboundCalls                     0.619394       0.724416      1.000000

PeakCallsInOut                   0.746892       0.710487      0.597737

OffPeakCallsInOut                0.738185       0.742311      0.654717

DroppedBlockedCalls              0.486951       0.493178      0.363934

CallForwardingCalls              0.012649       0.010597      0.013402

CallWaitingCalls                 0.650456       0.508989      0.546966

MonthsInService                 -0.023353      -0.026049     -0.015009

UniqueSubs                      -0.020814       0.015099      0.046173
```

```
ActiveSubs                    -0.046933       0.010791       0.068431
Handsets                       0.278439       0.265567       0.247948
HandsetModels                  0.272337       0.266455       0.248529
CurrentEquipmentDays          -0.248619      -0.239835      -0.201829
AgeHH1                        -0.139881      -0.131116      -0.105657
AgeHH2                        -0.122924      -0.112166      -0.078370
RetentionCalls                 0.006999       0.009353       0.004186
RetentionOffersAccepted        0.012184       0.008459       0.008839
ReferralsMadeBySubscriber      0.048008       0.060229       0.060459
IncomeGroup                   -0.130570      -0.103130      -0.101608
AdjustmentsToCreditRating      0.040914       0.042118       0.038283


                          PeakCallsInOut  OffPeakCallsInOut  \
MonthlyRevenue                  0.640922           0.471562
MonthlyMinutes                  0.770326           0.763910
TotalRecurringCharge            0.550819           0.367319
DirectorAssistedCalls           0.418529           0.273295
OverageMinutes                  0.517877           0.390641
RoamingCalls                    0.062661           0.031633
PercChangeMinutes              -0.109553          -0.089024
PercChangeRevenues             -0.080690          -0.052543
DroppedCalls                    0.574093           0.601072
BlockedCalls                    0.301752           0.330684
UnansweredCalls                 0.674665           0.719626
CustomerCareCalls               0.280366           0.404053
ThreewayCalls                   0.256693           0.315158
ReceivedCalls                   0.746892           0.738185
OutboundCalls                   0.710487           0.742311
InboundCalls                    0.597737           0.654717
PeakCallsInOut                  1.000000           0.698905
OffPeakCallsInOut               0.698905           1.000000
DroppedBlockedCalls             0.548645           0.585426
CallForwardingCalls             0.023314           0.023212
CallWaitingCalls                0.667853           0.663871
MonthsInService                 0.042537          -0.097256
UniqueSubs                      0.005609          -0.035030
ActiveSubs                     -0.006871          -0.064676
Handsets                        0.344345           0.270012
```

```
HandsetModels                    0.334877          0.262844
CurrentEquipmentDays            -0.240670         -0.285003
AgeHH1                          -0.107956         -0.138641
AgeHH2                          -0.097155         -0.114393
RetentionCalls                   0.021051          0.016210
RetentionOffersAccepted          0.020961          0.013736
ReferralsMadeBySubscriber        0.034816          0.054278
IncomeGroup                     -0.088100         -0.142229
AdjustmentsToCreditRating        0.068910          0.019426

                          DroppedBlockedCalls  CallForwardingCalls  \
MonthlyRevenue                       0.416785             0.011413
MonthlyMinutes                       0.572790             0.018664
TotalRecurringCharge                 0.333693             0.014085
DirectorAssistedCalls                0.252728             0.002323
OverageMinutes                       0.329256             0.003702
RoamingCalls                         0.060832             0.003243
PercChangeMinutes                   -0.087144            -0.004799
PercChangeRevenues                  -0.045726            -0.002364
DroppedCalls                         0.715420             0.004301
BlockedCalls                         0.816328             0.015123
UnansweredCalls                      0.498485             0.015151
CustomerCareCalls                    0.321728             0.013506
ThreewayCalls                        0.340332             0.001580
ReceivedCalls                        0.486951             0.012649
OutboundCalls                        0.493178             0.010597
InboundCalls                         0.363934             0.013402
PeakCallsInOut                       0.548645             0.023314
OffPeakCallsInOut                    0.585426             0.023212
DroppedBlockedCalls                  1.000000             0.013287
CallForwardingCalls                  0.013287             1.000000
CallWaitingCalls                     0.434301             0.024426
MonthsInService                     -0.079104             0.001858
UniqueSubs                          -0.024384             0.003218
ActiveSubs                          -0.048683            -0.000574
Handsets                             0.194916             0.014349
HandsetModels                        0.188486             0.013627
CurrentEquipmentDays                -0.216128            -0.009194
AgeHH1                              -0.102348            -0.005688
AgeHH2                              -0.086835            -0.006902
RetentionCalls                       0.019479             0.005171
RetentionOffersAccepted              0.014559            -0.001100
ReferralsMadeBySubscriber            0.034197             0.003394
IncomeGroup                         -0.103028            -0.009438
AdjustmentsToCreditRating            0.021112             0.001060

                          CallWaitingCalls  MonthsInService
UniqueSubs  \
MonthlyRevenue                    0.468730        -0.002305   -
```

```
0.014309
MonthlyMinutes                  0.614813        -0.068506   -
0.029803
TotalRecurringCharge            0.354574        -0.047954   -
0.020661
DirectorAssistedCalls           0.305112         0.017861   -
0.002697
OverageMinutes                  0.456640         0.001236   -
0.002879
RoamingCalls                    0.035981        -0.010866   -
0.003860
PercChangeMinutes              -0.129588         0.004069
0.001805
PercChangeRevenues             -0.093941        -0.006921
0.002532
DroppedCalls                    0.396783        -0.046175   -
0.021223
BlockedCalls                    0.286784        -0.072684   -
0.016170
UnansweredCalls                 0.488569        -0.065783   -
0.019853
CustomerCareCalls               0.212042        -0.107552   -
0.051419
ThreewayCalls                   0.221297        -0.057165   -
0.018429
ReceivedCalls                   0.650456        -0.023353   -
0.020814
OutboundCalls                   0.508989        -0.026049
0.015099
InboundCalls                    0.546966        -0.015009
0.046173
PeakCallsInOut                  0.667853         0.042537
0.005609
OffPeakCallsInOut               0.663871        -0.097256   -
0.035030
DroppedBlockedCalls             0.434301        -0.079104   -
0.024384
CallForwardingCalls             0.024426         0.001858
0.003218
CallWaitingCalls                1.000000        -0.003784   -
0.006666
MonthsInService                -0.003784         1.000000
0.016668
UniqueSubs                     -0.006666         0.016668
1.000000
ActiveSubs                     -0.017167         0.031670
0.775776
Handsets                        0.235632         0.388851
0.033972
HandsetModels                   0.225479         0.403806
```

```
0.041341
CurrentEquipmentDays               -0.166953          0.455842   -
0.013297
AgeHH1                             -0.065372          0.122317
0.028450
AgeHH2                             -0.049798          0.084733
0.038512
RetentionCalls                      0.010610          0.069383   -
0.008652
RetentionOffersAccepted             0.010994          0.054548   -
0.000618
ReferralsMadeBySubscriber           0.047291         -0.016852   -
0.023087
IncomeGroup                        -0.066715          0.139341
0.048088
AdjustmentsToCreditRating           0.022840          0.222697   -
0.002110

                           ActiveSubs  Handsets  HandsetModels  \
MonthlyRevenue              -0.041631  0.242457       0.235636
MonthlyMinutes              -0.065828  0.301769       0.291961
TotalRecurringCharge        -0.056717  0.233702       0.225111
DirectorAssistedCalls       -0.014810  0.184442       0.172056
OverageMinutes              -0.012533  0.142617       0.143260
RoamingCalls                -0.003190  0.024343       0.022872
PercChangeMinutes            0.008515 -0.012494      -0.012229
PercChangeRevenues           0.005129 -0.018096      -0.017873
DroppedCalls                -0.049663  0.225186       0.217869
BlockedCalls                -0.026820  0.088458       0.085308
UnansweredCalls             -0.044963  0.252150       0.243701
CustomerCareCalls           -0.086749  0.108153       0.097099
ThreewayCalls               -0.031942  0.093039       0.084583
ReceivedCalls               -0.046933  0.278439       0.272337
OutboundCalls                0.010791  0.265567       0.266455
InboundCalls                 0.068431  0.247948       0.248529
PeakCallsInOut              -0.006871  0.344345       0.334877
OffPeakCallsInOut           -0.064676  0.270012       0.262844
DroppedBlockedCalls         -0.048683  0.194916       0.188486
CallForwardingCalls         -0.000574  0.014349       0.013627
CallWaitingCalls            -0.017167  0.235632       0.225479
MonthsInService              0.031670  0.388851       0.403806
UniqueSubs                   0.775776  0.033972       0.041341
ActiveSubs                   1.000000  0.023699       0.033118
Handsets                     0.023699  1.000000       0.887839
HandsetModels                0.033118  0.887839       1.000000
CurrentEquipmentDays         0.006024 -0.352019      -0.378004
AgeHH1                       0.081397 -0.023929      -0.026384
AgeHH2                       0.088561 -0.034679      -0.034895
RetentionCalls              -0.033273  0.102426       0.101267
```

```
RetentionOffersAccepted     -0.014169  0.099518       0.106210
ReferralsMadeBySubscriber   -0.031382  0.036727       0.039852
IncomeGroup                  0.105815 -0.026566      -0.023305
AdjustmentsToCreditRating   -0.004098  0.191469       0.169111

                          CurrentEquipmentDays    AgeHH1    AgeHH2  \
MonthlyRevenue                       -0.217373 -0.105784 -0.106391
MonthlyMinutes                       -0.310101 -0.160494 -0.142221
TotalRecurringCharge                 -0.247995 -0.103210 -0.101194
DirectorAssistedCalls                -0.121587 -0.057514 -0.058179
OverageMinutes                       -0.124260 -0.063302 -0.063191
RoamingCalls                         -0.029030 -0.010196 -0.013209
PercChangeMinutes                    -0.006167  0.014070  0.003614
PercChangeRevenues                    0.005798  0.006732  0.002391
DroppedCalls                         -0.216536 -0.117797 -0.107388
BlockedCalls                         -0.124673 -0.046762 -0.033732
UnansweredCalls                      -0.243193 -0.119593 -0.101761
CustomerCareCalls                    -0.169502 -0.099774 -0.082794
ThreewayCalls                        -0.111433 -0.034702 -0.042406
ReceivedCalls                        -0.248619 -0.139881 -0.122924
OutboundCalls                        -0.239835 -0.131116 -0.112166
InboundCalls                         -0.201829 -0.105657 -0.078370
PeakCallsInOut                       -0.240670 -0.107956 -0.097155
OffPeakCallsInOut                    -0.285003 -0.138641 -0.114393
DroppedBlockedCalls                  -0.216128 -0.102348 -0.086835
CallForwardingCalls                  -0.009194 -0.005688 -0.006902
CallWaitingCalls                     -0.166953 -0.065372 -0.049798
MonthsInService                       0.455842  0.122317  0.084733
UniqueSubs                           -0.013297  0.028450  0.038512
ActiveSubs                            0.006024  0.081397  0.088561
Handsets                             -0.352019 -0.023929 -0.034679
HandsetModels                        -0.378004 -0.026384 -0.034895
CurrentEquipmentDays                  1.000000  0.124982  0.105688
AgeHH1                                0.124982  1.000000  0.666340
AgeHH2                                0.105688  0.666340  1.000000
RetentionCalls                       -0.025441 -0.010450 -0.010455
RetentionOffersAccepted              -0.038754 -0.003854 -0.004665
ReferralsMadeBySubscriber            -0.028883 -0.017699 -0.011138
IncomeGroup                           0.137730  0.625019  0.480547
AdjustmentsToCreditRating             0.039338  0.018222  0.006212

                          RetentionCalls  RetentionOffersAccepted  \
MonthlyRevenue                  0.011142                 0.015402
MonthlyMinutes                  0.009560                 0.014874
TotalRecurringCharge           -0.022028                -0.002306
DirectorAssistedCalls           0.008477                 0.013311
OverageMinutes                  0.014811                 0.009279
RoamingCalls                   -0.002067                -0.000816
PercChangeMinutes              -0.021247                -0.008209
```

PercChangeRevenues	-0.019110	-0.012163
DroppedCalls	0.020308	0.014858
BlockedCalls	0.010841	0.008350
UnansweredCalls	0.030330	0.025852
CustomerCareCalls	0.026069	0.022238
ThreewayCalls	0.007382	0.006319
ReceivedCalls	0.006999	0.012184
OutboundCalls	0.009353	0.008459
InboundCalls	0.004186	0.008839
PeakCallsInOut	0.021051	0.020961
OffPeakCallsInOut	0.016210	0.013736
DroppedBlockedCalls	0.019479	0.014559
CallForwardingCalls	0.005171	-0.001100
CallWaitingCalls	0.010610	0.010994
MonthsInService	0.069383	0.054548
UniqueSubs	-0.008652	-0.000618
ActiveSubs	-0.033273	-0.014169
Handsets	0.102426	0.099518
HandsetModels	0.101267	0.106210
CurrentEquipmentDays	-0.025441	-0.038754
AgeHH1	-0.010450	-0.003854
AgeHH2	-0.010455	-0.004665
RetentionCalls	1.000000	0.734113
RetentionOffersAccepted	0.734113	1.000000
ReferralsMadeBySubscriber	0.006206	0.007340
IncomeGroup	-0.017875	-0.006782
AdjustmentsToCreditRating	0.025907	0.023940

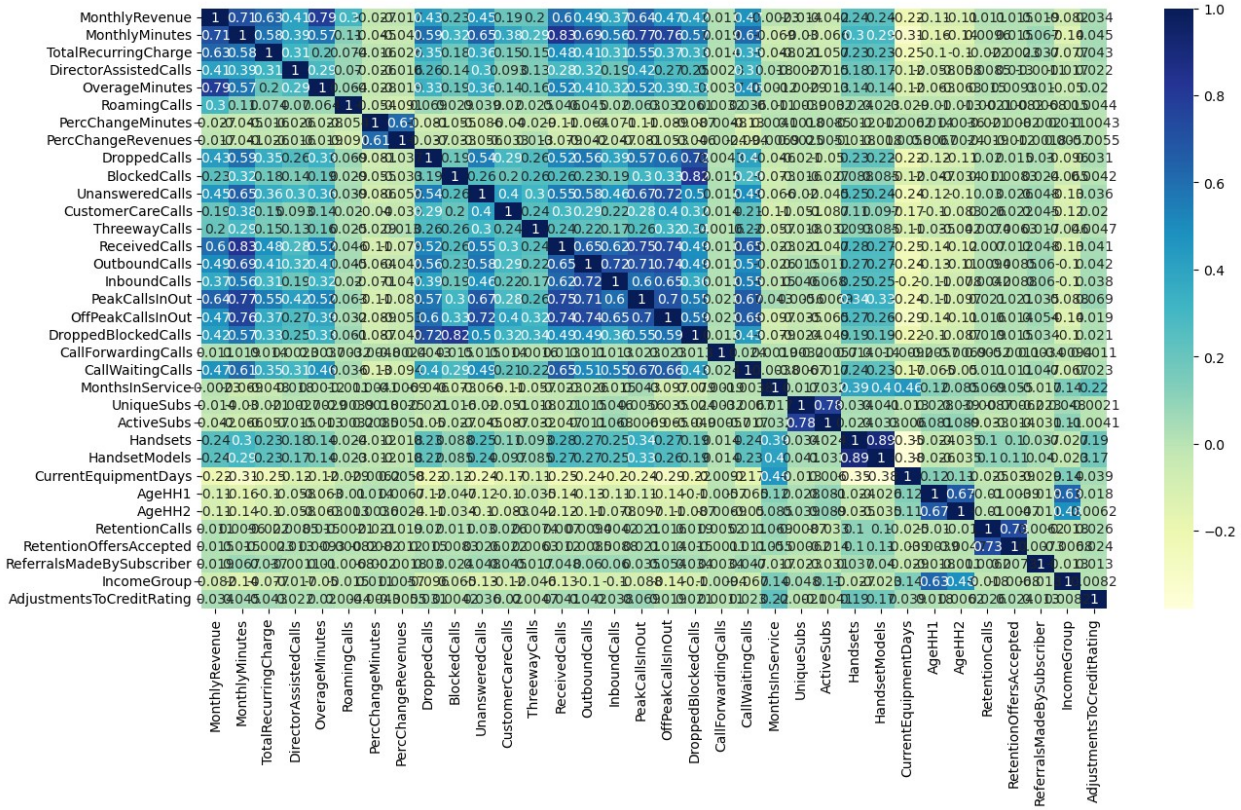
	ReferralsMadeBySubscriber	IncomeGroup	\
MonthlyRevenue	0.018933	-0.081593	
MonthlyMinutes	0.067452	-0.144348	
TotalRecurringCharge	0.036715	-0.077023	
DirectorAssistedCalls	-0.001083	-0.017408	
OverageMinutes	0.010081	-0.049999	
RoamingCalls	-0.006843	-0.015228	
PercChangeMinutes	-0.001978	0.011412	
PercChangeRevenues	-0.001785	0.005717	
DroppedCalls	0.029546	-0.096316	
BlockedCalls	0.023848	-0.065375	
UnansweredCalls	0.047777	-0.129093	
CustomerCareCalls	0.044990	-0.119494	
ThreewayCalls	0.017090	-0.046028	
ReceivedCalls	0.048008	-0.130570	
OutboundCalls	0.060229	-0.103130	
InboundCalls	0.060459	-0.101608	
PeakCallsInOut	0.034816	-0.088100	
OffPeakCallsInOut	0.054278	-0.142229	
DroppedBlockedCalls	0.034197	-0.103028	
CallForwardingCalls	0.003394	-0.009438	

CallWaitingCalls	0.047291	-0.066715
MonthsInService	-0.016852	0.139341
UniqueSubs	-0.023087	0.048088
ActiveSubs	-0.031382	0.105815
Handsets	0.036727	-0.026566
HandsetModels	0.039852	-0.023305
CurrentEquipmentDays	-0.028883	0.137730
AgeHH1	-0.017699	0.625019
AgeHH2	-0.011138	0.480547
RetentionCalls	0.006206	-0.017875
RetentionOffersAccepted	0.007340	-0.006782
ReferralsMadeBySubscriber	1.000000	-0.012899
IncomeGroup	-0.012899	1.000000
AdjustmentsToCreditRating	0.012750	0.008177

	AdjustmentsToCreditRating
MonthlyRevenue	0.034164
MonthlyMinutes	0.045164
TotalRecurringCharge	0.043396
DirectorAssistedCalls	0.022424
OverageMinutes	0.020464
RoamingCalls	0.004398
PercChangeMinutes	-0.004303
PercChangeRevenues	-0.005524
DroppedCalls	0.031039
BlockedCalls	0.004167
UnansweredCalls	0.036239
CustomerCareCalls	0.019600
ThreewayCalls	0.004713
ReceivedCalls	0.040914
OutboundCalls	0.042118
InboundCalls	0.038283
PeakCallsInOut	0.068910
OffPeakCallsInOut	0.019426
DroppedBlockedCalls	0.021112
CallForwardingCalls	0.001060
CallWaitingCalls	0.022840
MonthsInService	0.222697
UniqueSubs	-0.002110
ActiveSubs	-0.004098
Handsets	0.191469
HandsetModels	0.169111
CurrentEquipmentDays	0.039338
AgeHH1	0.018222
AgeHH2	0.006212
RetentionCalls	0.025907
RetentionOffersAccepted	0.023940
ReferralsMadeBySubscriber	0.012750

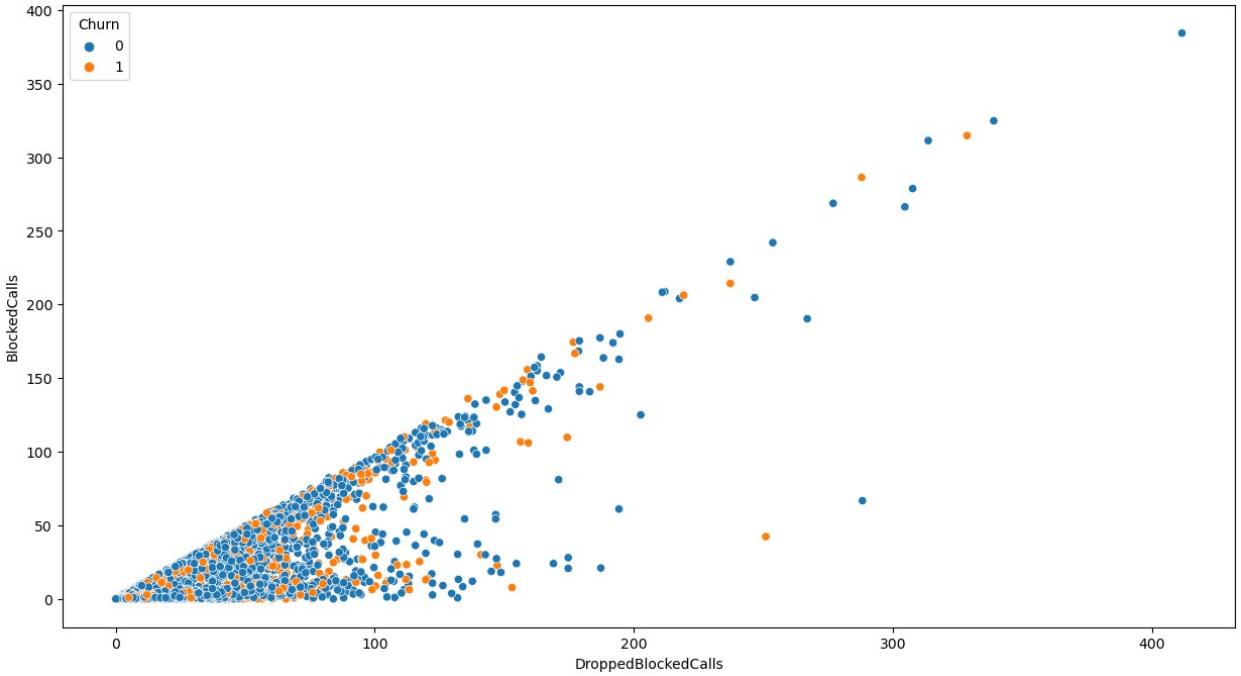
```
IncomeGroup                                0.008177
AdjustmentsToCreditRating                  1.000000
```

```
sns.heatmap(df.corr(), annot = True, cmap='YlGnBu')
plt.show()
```



- We can see that few columns are correlated with the other columns which means there is `Multicollinearity` in the data.
- `Droppedblockedcalls` and `Blockedcalls` have the maximum correlation.

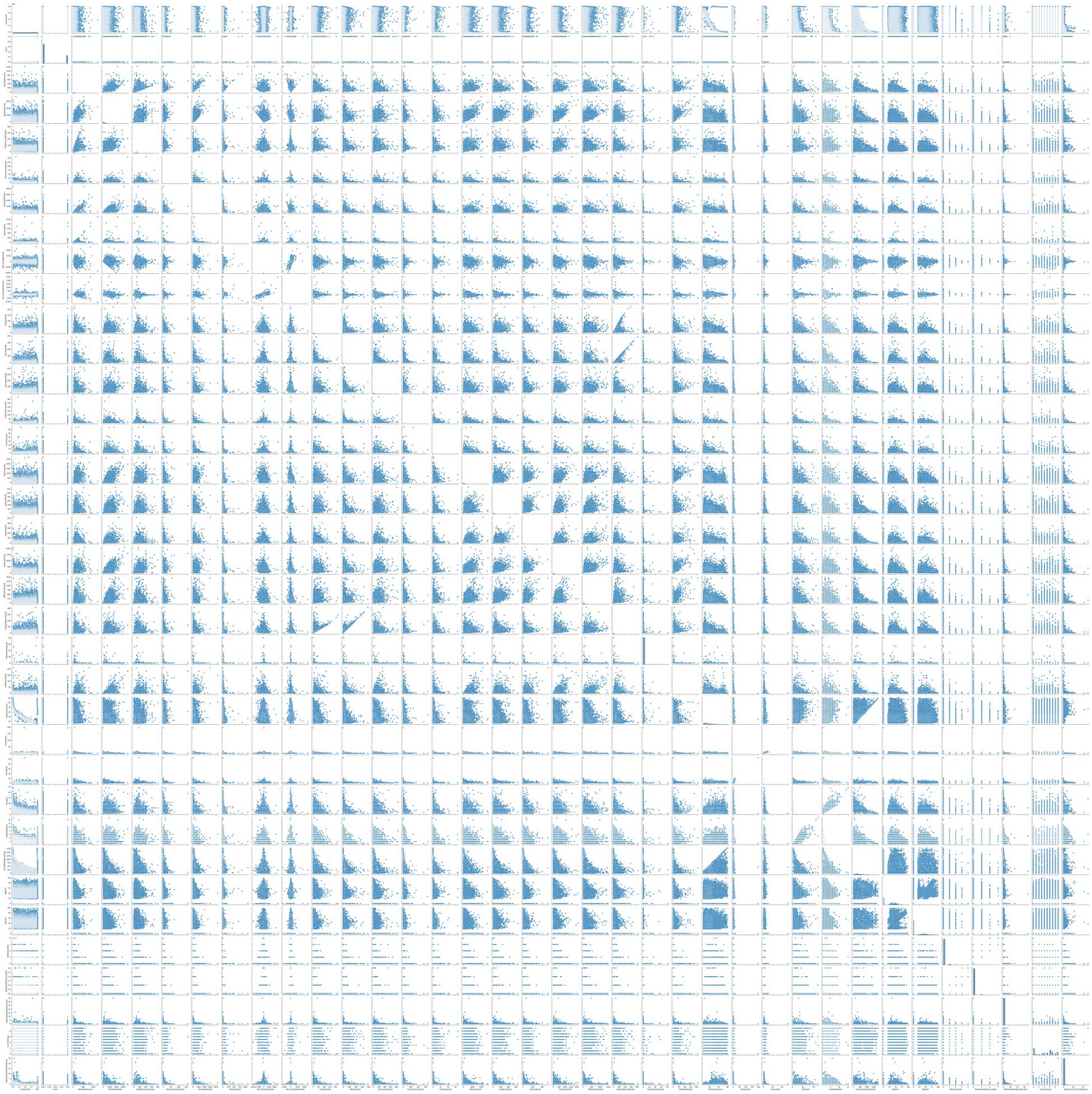
```
sns.scatterplot(data = df, x = 'DroppedBlockedCalls', y =
'BlockedCalls', hue = 'Churn')
plt.show()
```



- Just to view a positive correlation between DroppedBlockedCalls and BlockedCalls with hue as 'Churn'.

Pairplot

```
sns.pairplot(df)
plt.show()
```



- A mixed of positive correlation, no correlation and negative correlations values are present.

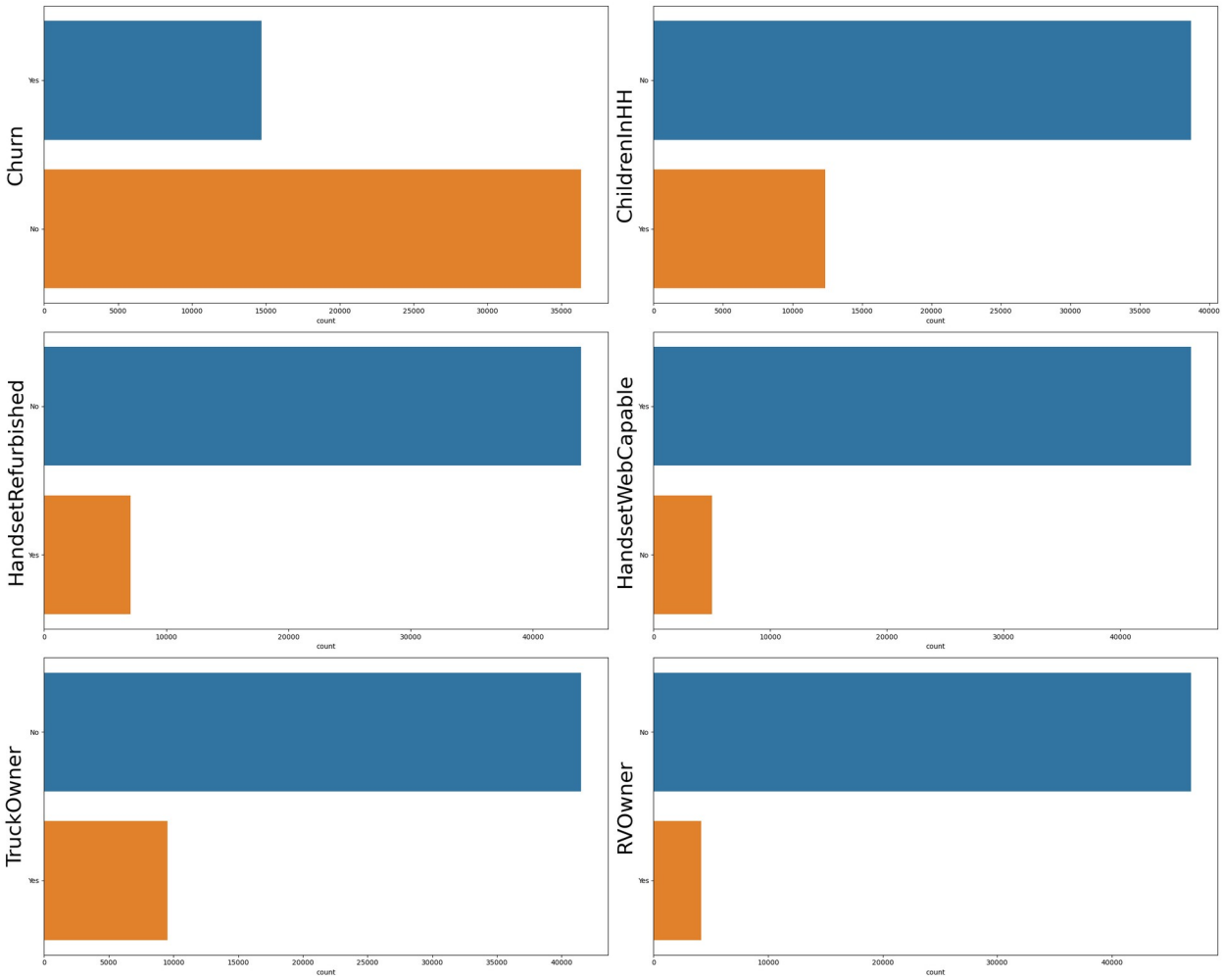
Analyze Categorical Variables

```
df_cat.describe()
```

```
        CustomerID  Churn ServiceArea ChildrenInHH HandsetRefurbished
\
count        51047  51047        51047         51047                51047

unique       51047      2          747             2                    2
```

```
top        3000002     No   NYCBRO917           No                   No

freq             1  36336        1685        38679               43956


       HandsetWebCapable TruckOwner RVOwner Homeownership
BuysViaMailOrder  \
count              51047      51047   51047         51047
51047
unique                 2          2       2             2
2
top                  Yes         No      No         Known
No
freq               46046      41524   46894         33987
32596

       RespondsToMailOffers OptOutMailings NonUSTravel OwnsComputer  \
count                 51047          51047       51047        51047
unique                    2              2           2            2
top                      No             No          No           No
freq                  31821          50295       48168        41583

       HasCreditCard NewCellphoneUser NotNewCellphoneUser
OwnsMotorcycle  \
count          51047            51047               51047
51047
unique             2                2                   2
2
top              Yes               No                  No
No
freq           34503            41223               44012
50368

       HandsetPrice MadeCallToRetentionTeam CreditRating PrizmCode
Occupation  \
count         51047                   51047        51047     51047
51047
unique           16                       2            7         4
8
top         Unknown                      No       2-High     Other
Other
freq          28982                   49302        18993     24655
37637

       MaritalStatus
count          51047
unique             3
top          Unknown
freq           19700
```

```
df_cat_features = df_cat.drop(['CustomerID', 'ServiceArea'], axis=1)

fig, ax = plt.subplots(3, 2, figsize=(25, 20))

for variable, subplot in zip(df_cat_features, ax.flatten()):

    countplot = sns.countplot(y=df_cat_features[variable],
ax=subplot )
    countplot.set_ylabel(variable, fontsize = 30)

plt.tight_layout()

# display the plot
plt.show()
```



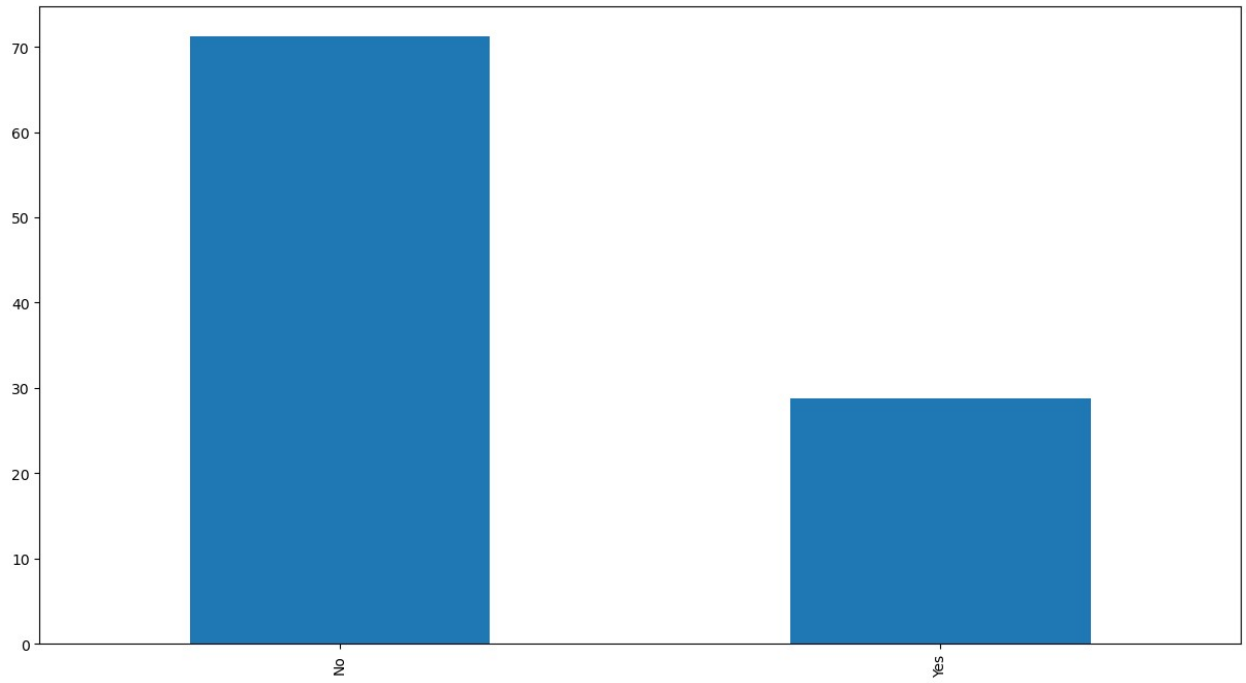
- From the above Visualization, we can infer that there are more `No` than `Yes`, and that's why our target variable(churn) also has `No` as the majority class.

Lets Analyze the target variable

```
df.Churn.value_counts()/df.Churn.count()*100
```

```
No     71.181460
Yes    28.818540
Name: Churn, dtype: float64
```

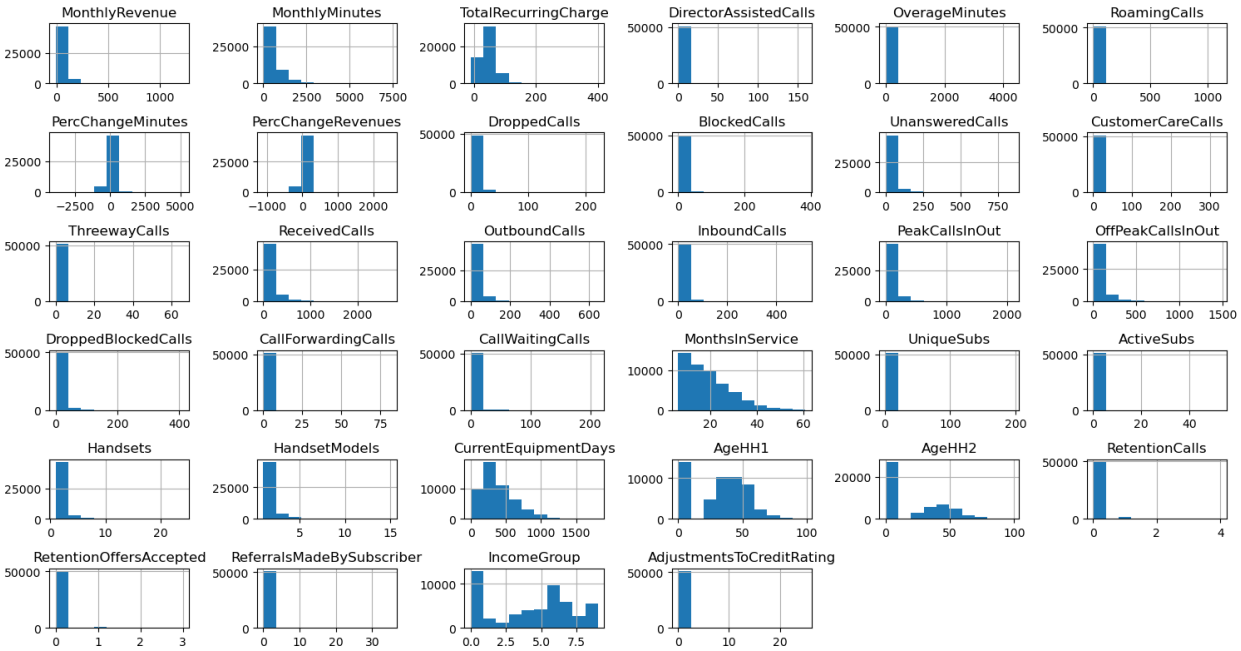
```
(df.Churn.value_counts()/df.Churn.count()*100).plot(kind = 'bar')
plt.show()
```




```
# There are more number of customers who dont churn than the customers
who actually churn.
```

Univariate Analysis

```
#Lets check the distribution of the variables.

df.hist()
plt.tight_layout()
```




```
# We can see that most of them are right skewed.

df.skew()
```

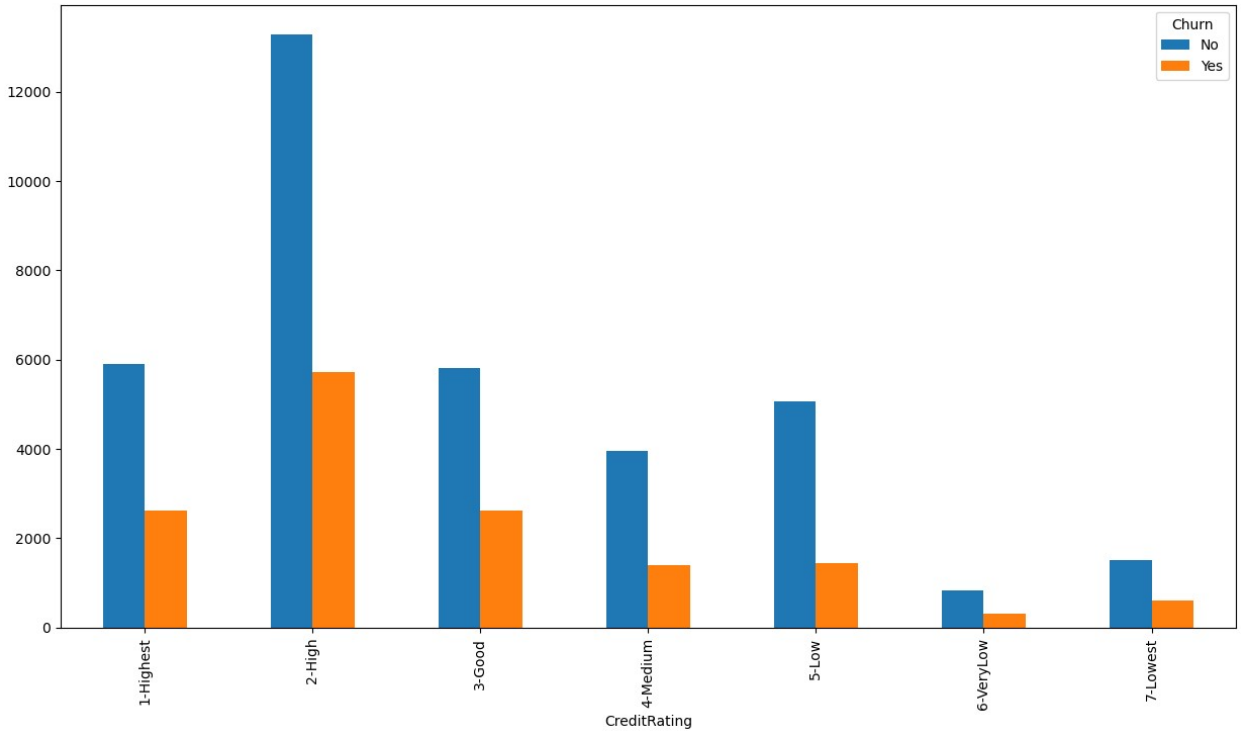
```
CustomerID                 -0.009563
MonthlyRevenue              4.189224
MonthlyMinutes              2.199096
TotalRecurringCharge        1.632499
DirectorAssistedCalls      13.586024
OverageMinutes              8.124212
RoamingCalls               57.967352
PercChangeMinutes          -0.325975
PercChangeRevenues          7.913268
DroppedCalls                4.549222
BlockedCalls                9.790454
UnansweredCalls             4.381230
CustomerCareCalls          14.235913
ThreewayCalls              17.553060
ReceivedCalls               3.115673
OutboundCalls               3.525679
InboundCalls                5.927980
PeakCallsInOut              3.327854
OffPeakCallsInOut           3.493146
DroppedBlockedCalls         5.522108
CallForwardingCalls        91.634791
CallWaitingCalls           11.121429
MonthsInService             1.056937
UniqueSubs                 79.635758
ActiveSubs                 10.647697
Handsets                    3.288712
```

```
HandsetModels                 2.405276
CurrentEquipmentDays          1.086127
AgeHH1                       -0.258175
AgeHH2                        0.577634
RetentionCalls                6.296663
RetentionOffersAccepted       8.699968
ReferralsMadeBySubscriber    36.739725
IncomeGroup                  -0.175756
AdjustmentsToCreditRating    18.623149
dtype: float64

# UniqueSubs and Callforwardingcalls has the highest skewness.
# Only PerchangeMinutes follows a little bit of Normal Distribution.
```

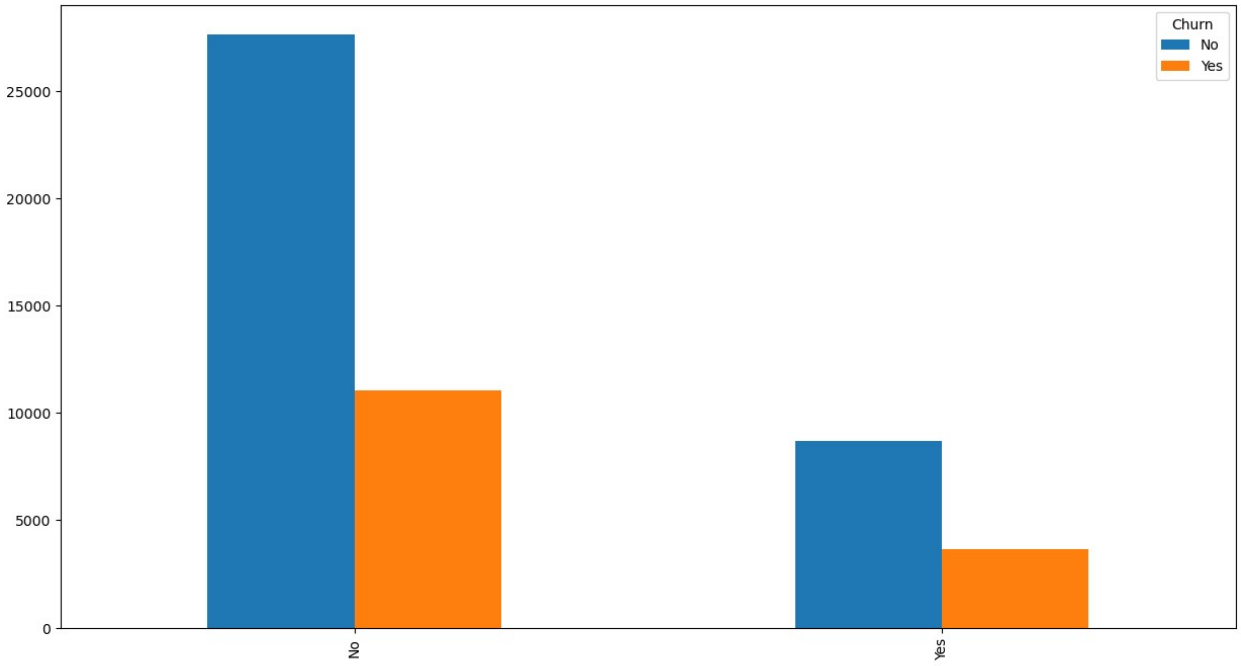
BiVariate Analysis

Analyze relationship between target and categorical variables

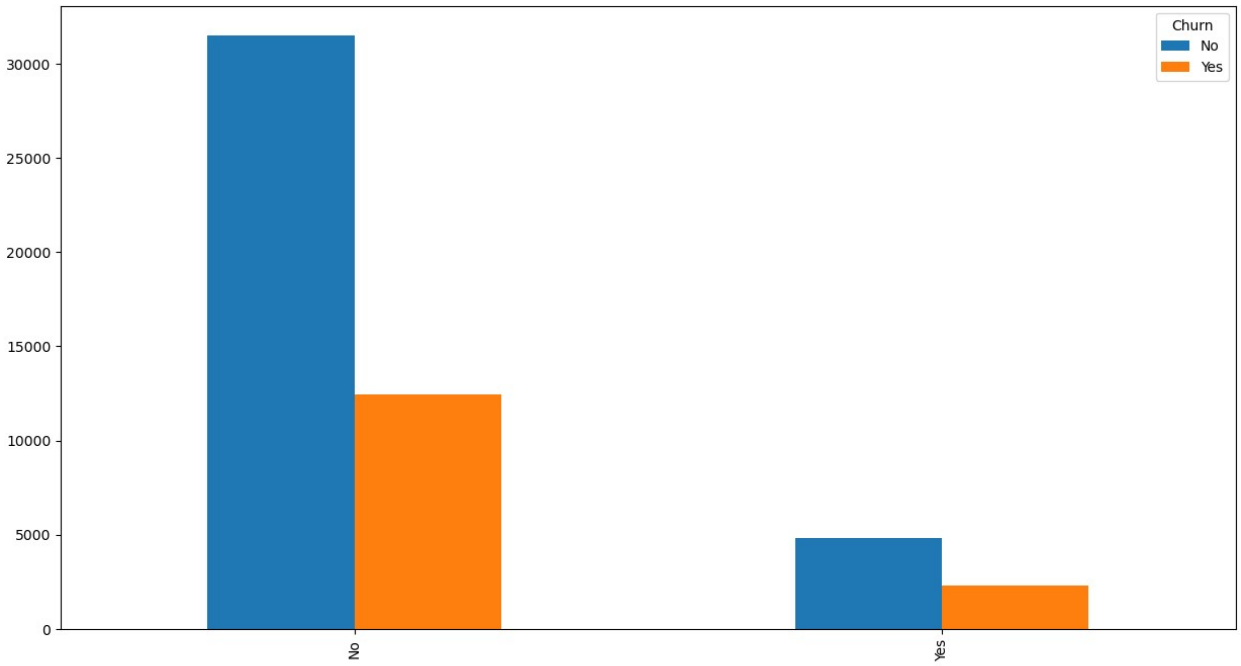
```
df.CreditRating.value_counts()

2-High       18993
1-Highest     8522
3-Good        8410
5-Low         6499
4-Medium      5357
7-Lowest      2114
6-VeryLow     1152
Name: CreditRating, dtype: int64

df.MaritalStatus.value_counts()

Unknown    19700
Yes        18651
No         12696
Name: MaritalStatus, dtype: int64

pd.crosstab(df.CreditRating, df.Churn).plot(kind = 'bar')

<Axes: xlabel='CreditRating'>
```



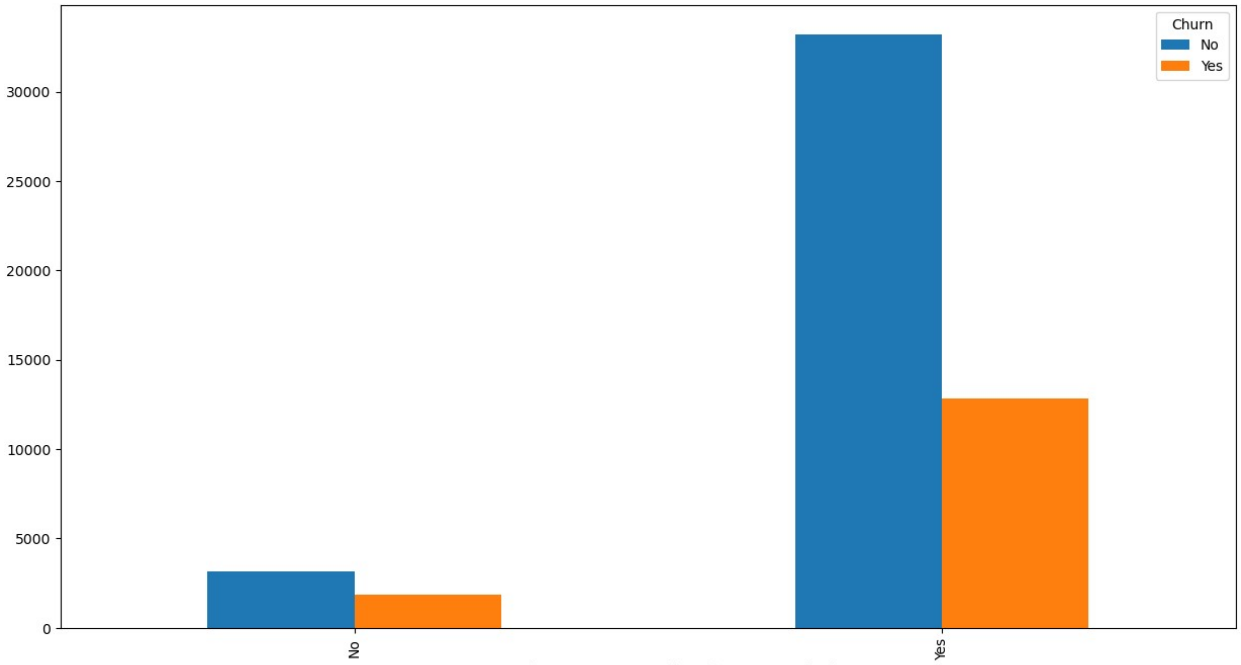

```
for variable in df_cat_features.drop('Churn', axis =1):
    crosstab = pd.crosstab(df[variable], df.Churn).plot(kind = 'bar')
    crosstab.set_xlabel(variable, fontsize = 30)

plt.tight_layout()

plt.show()
```



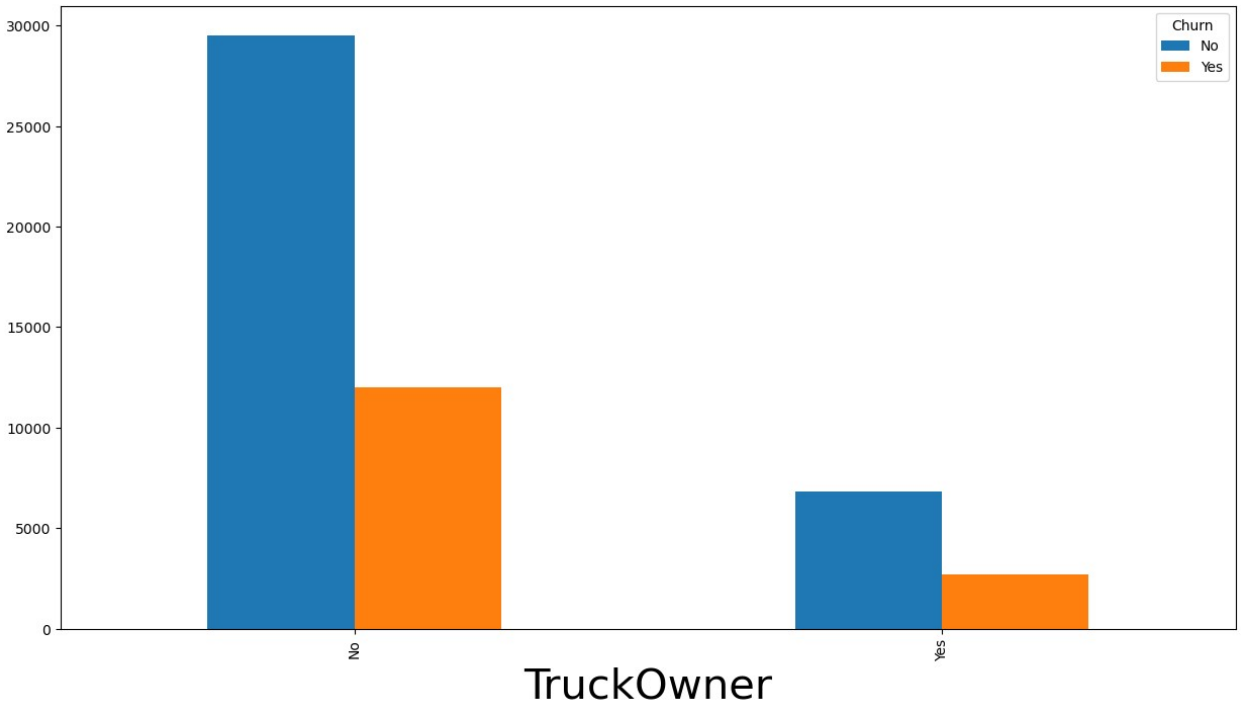
ChildrenInHH



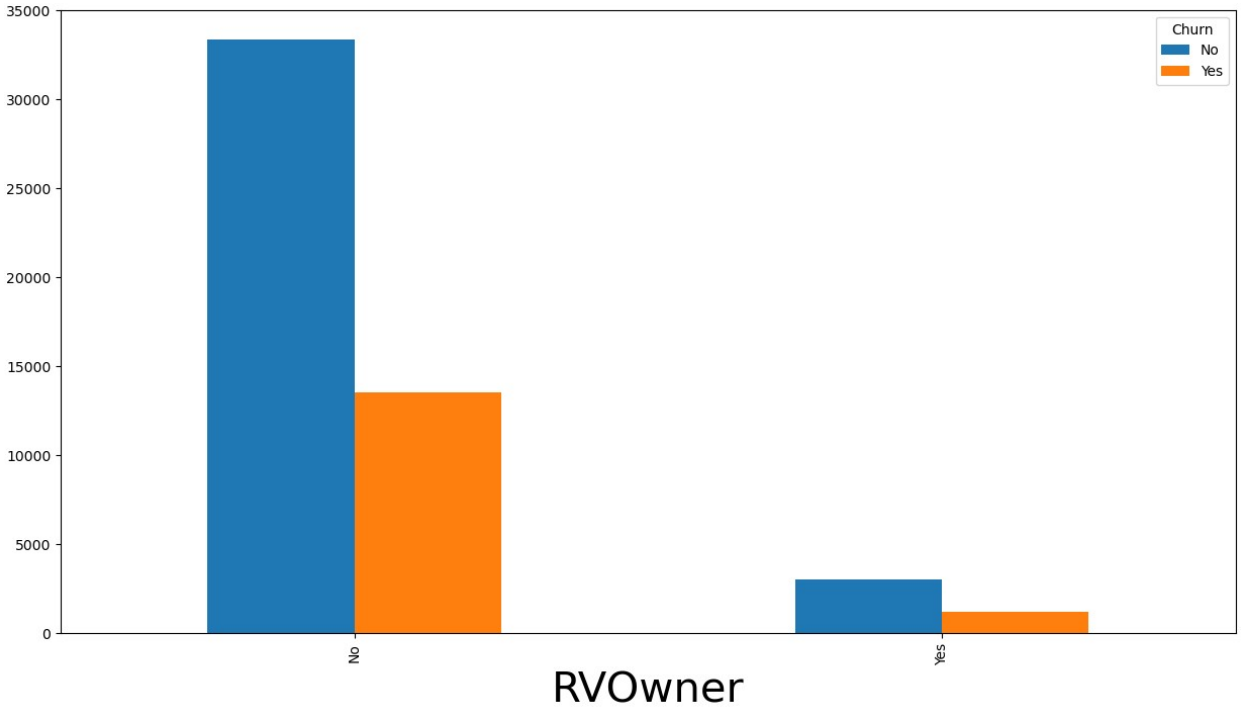
HandsetRefurbished



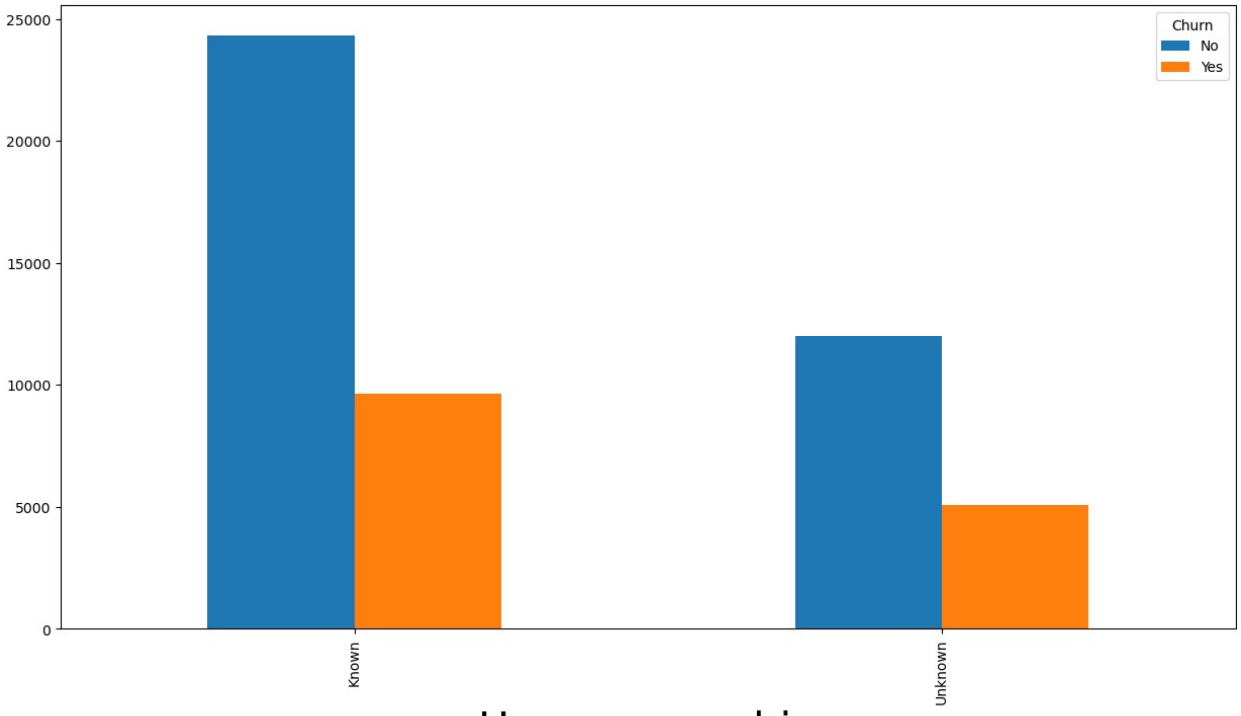
HandsetWebCapable



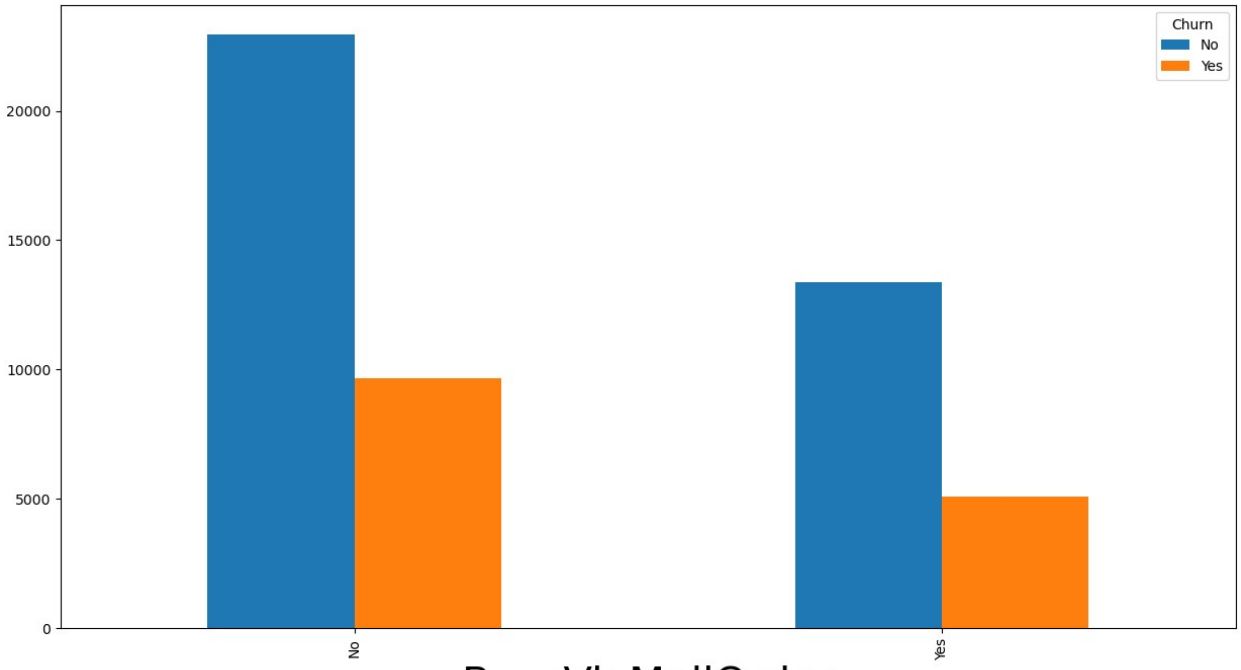
TruckOwner



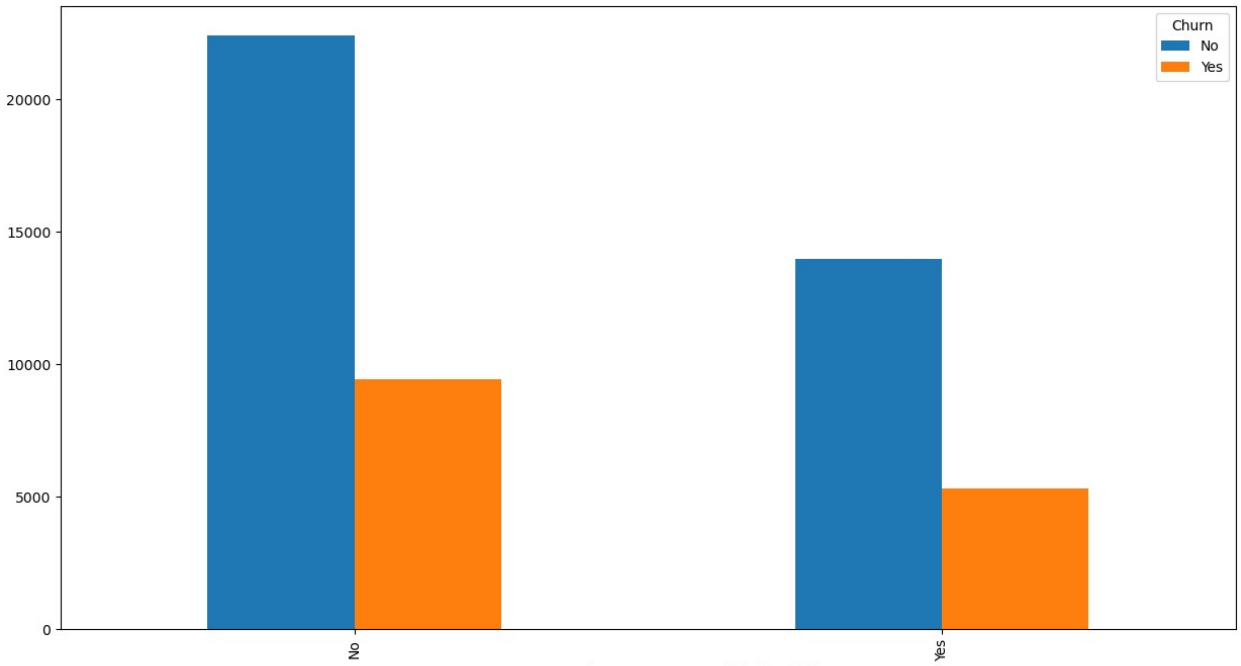
RVOwner



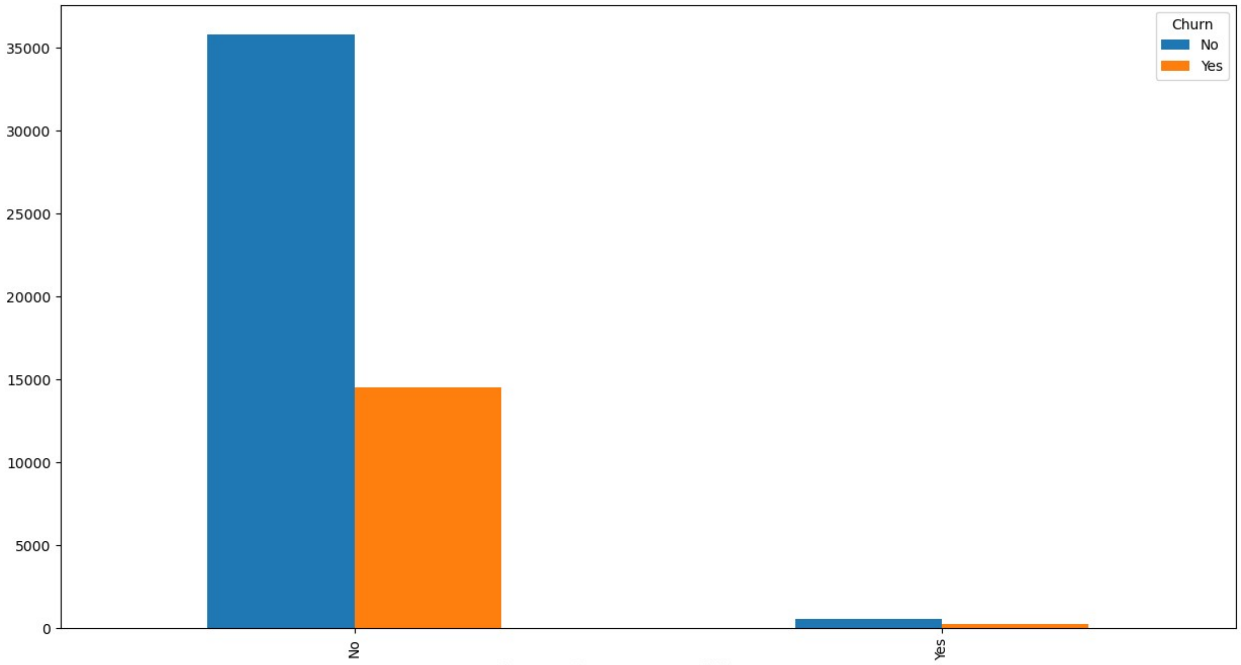
Homeownership



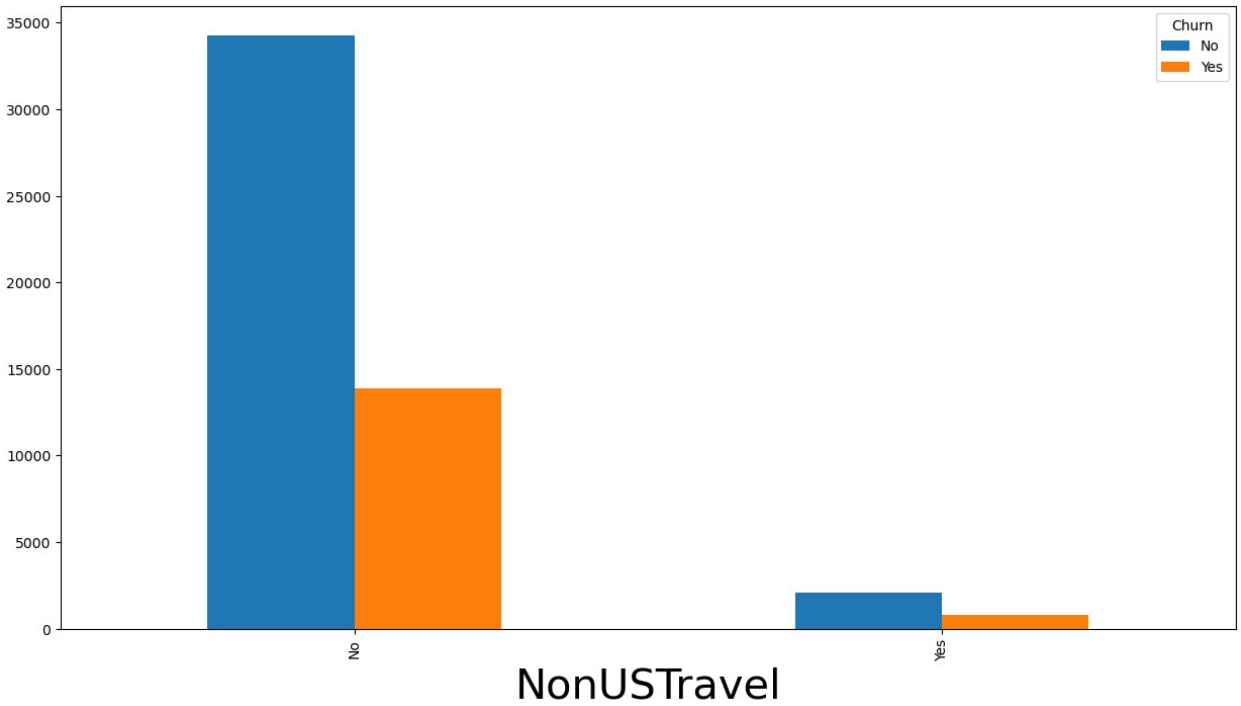
BuysViaMailOrder



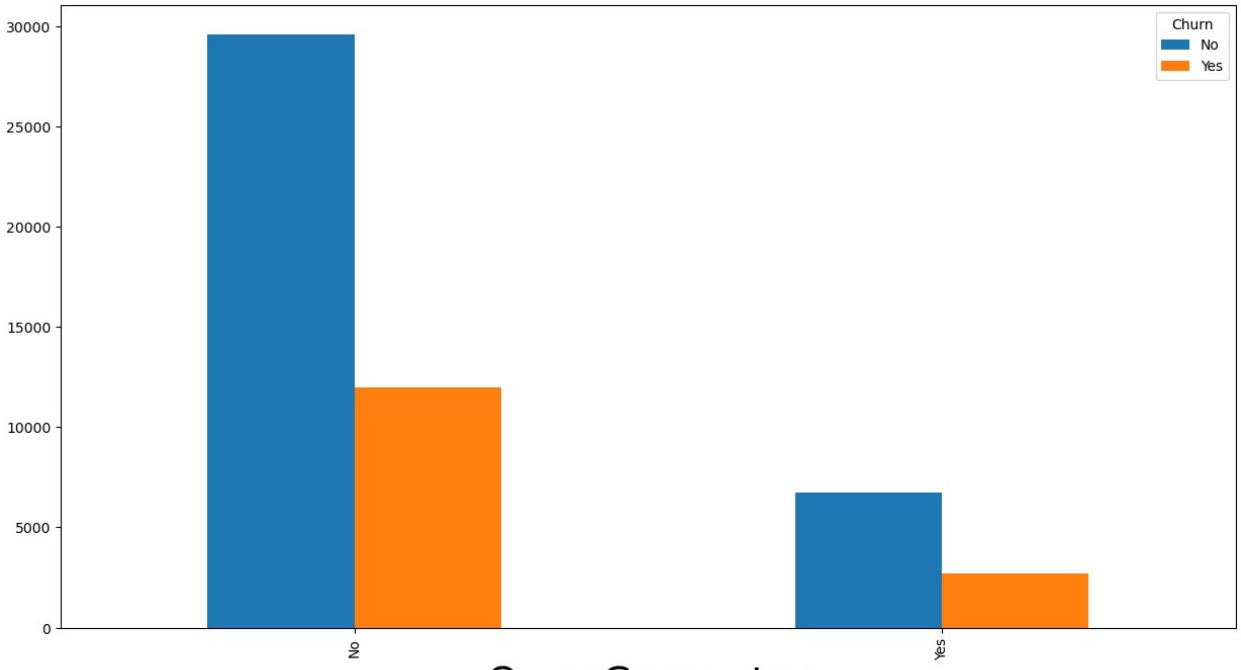
RespondsToMailOffers



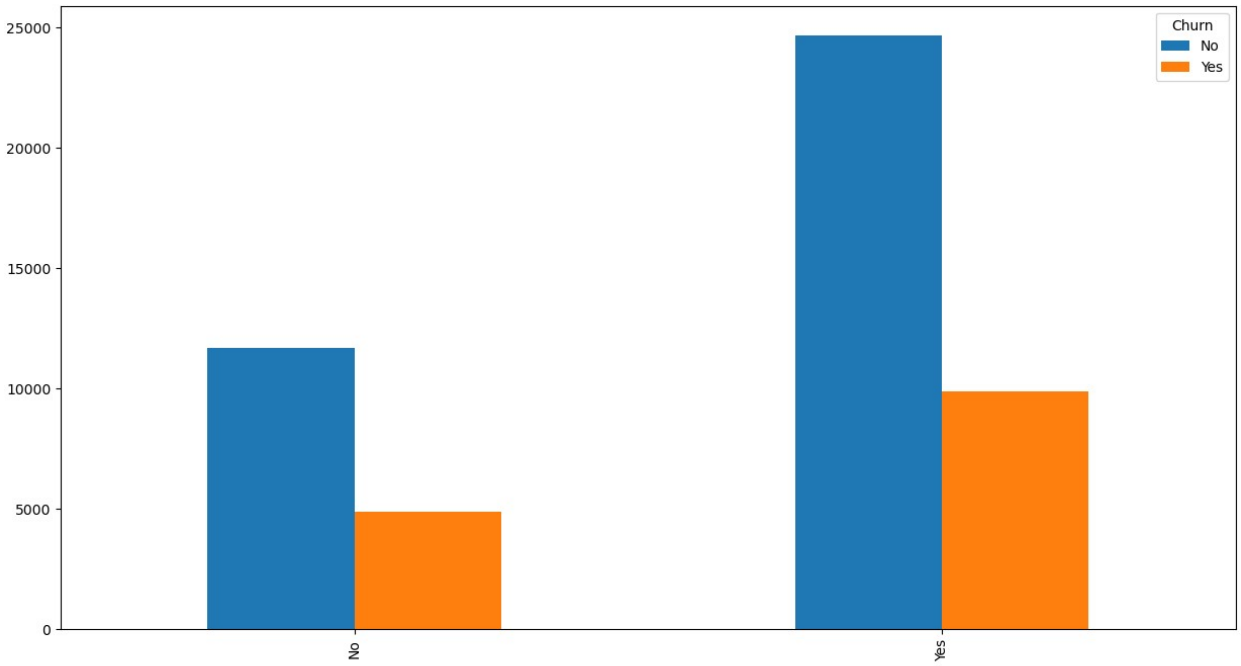
OptOutMailings



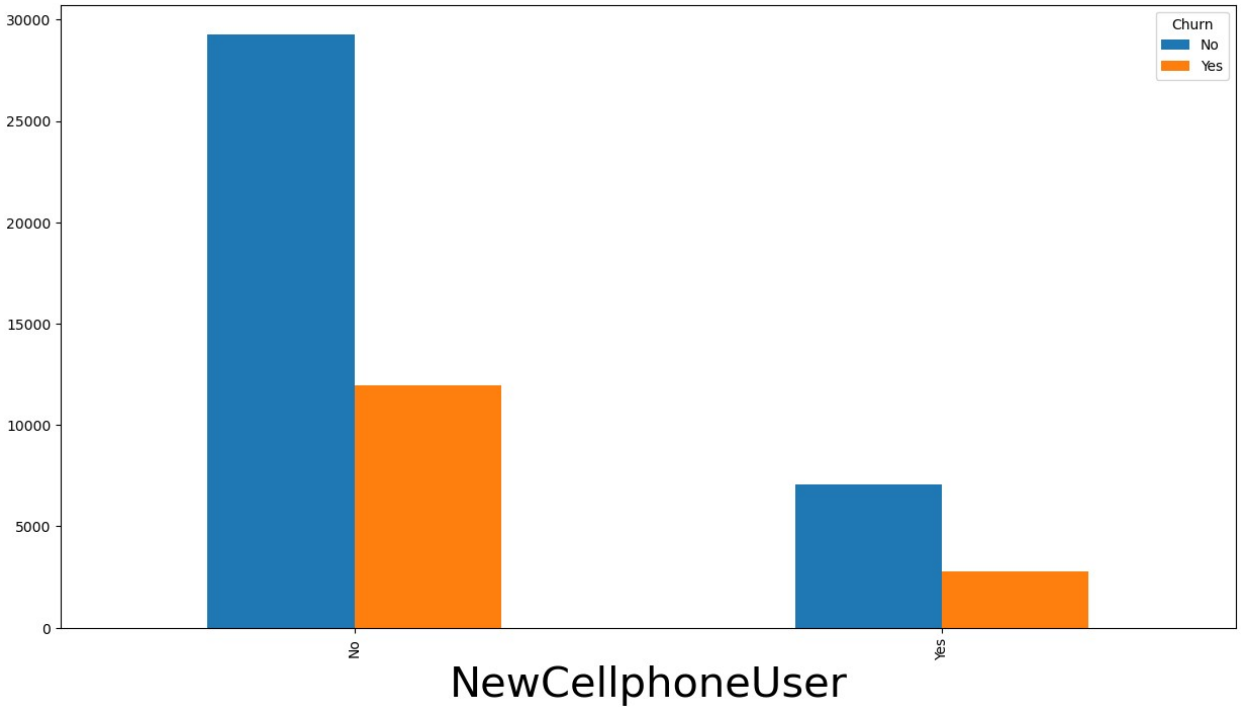
NonUSTravel

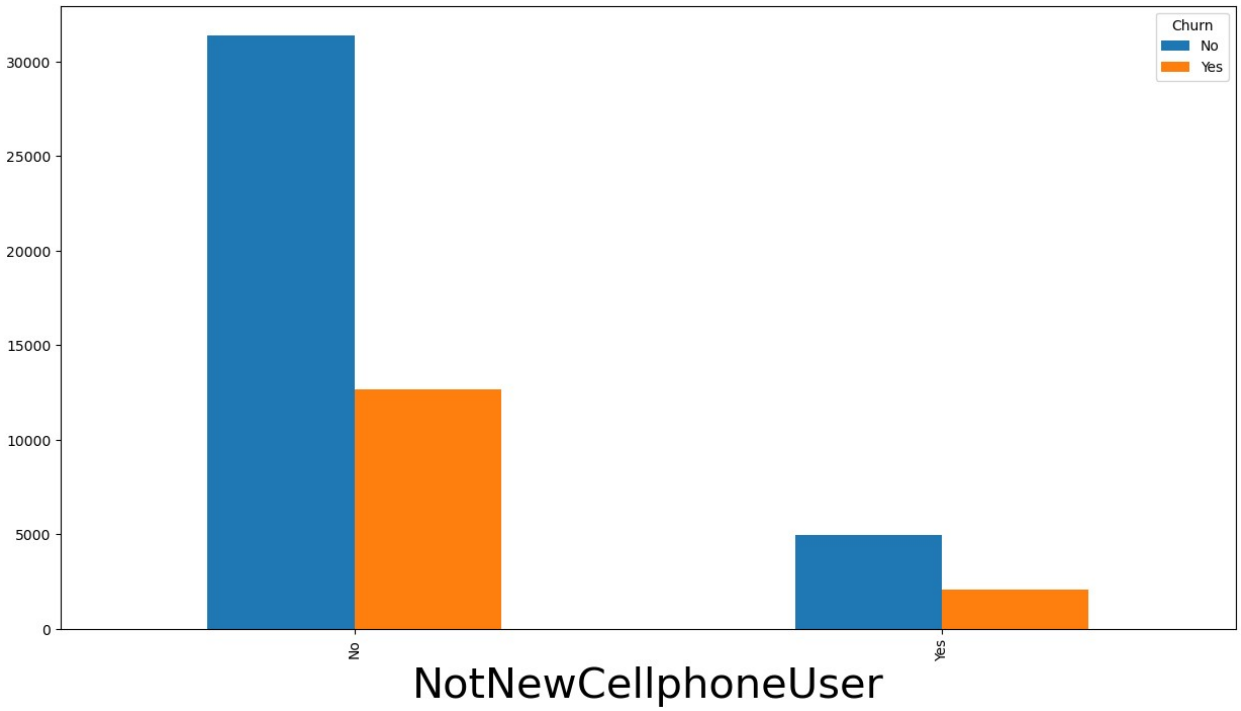


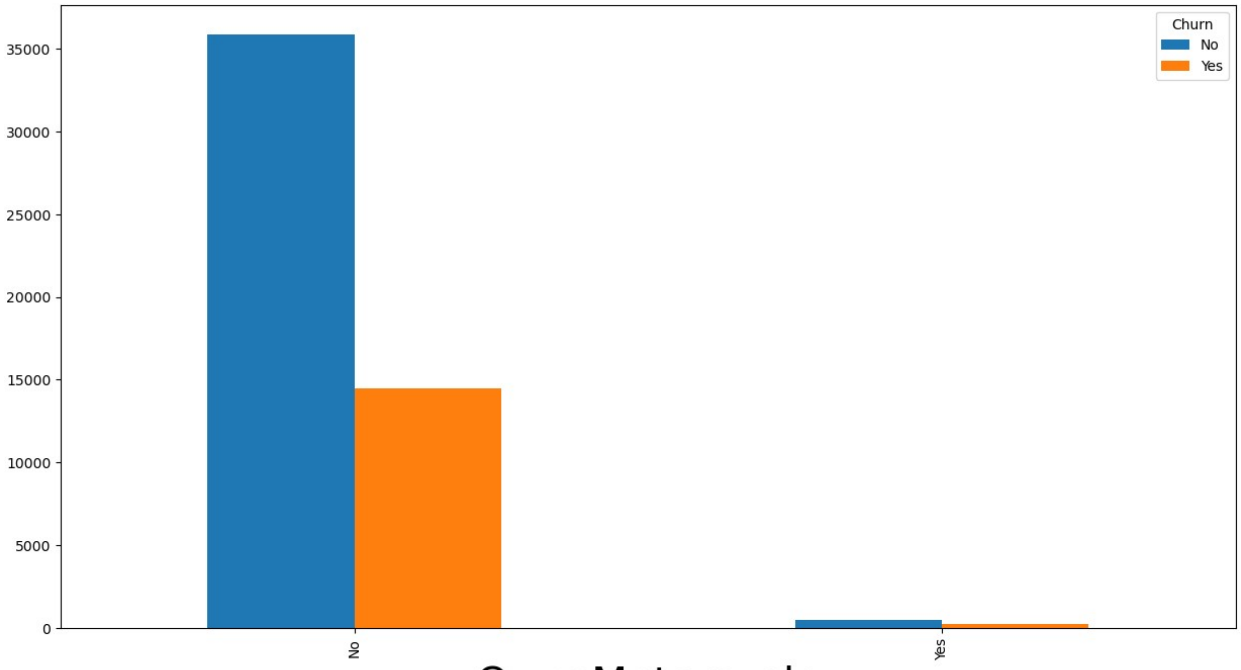
OwnsComputer



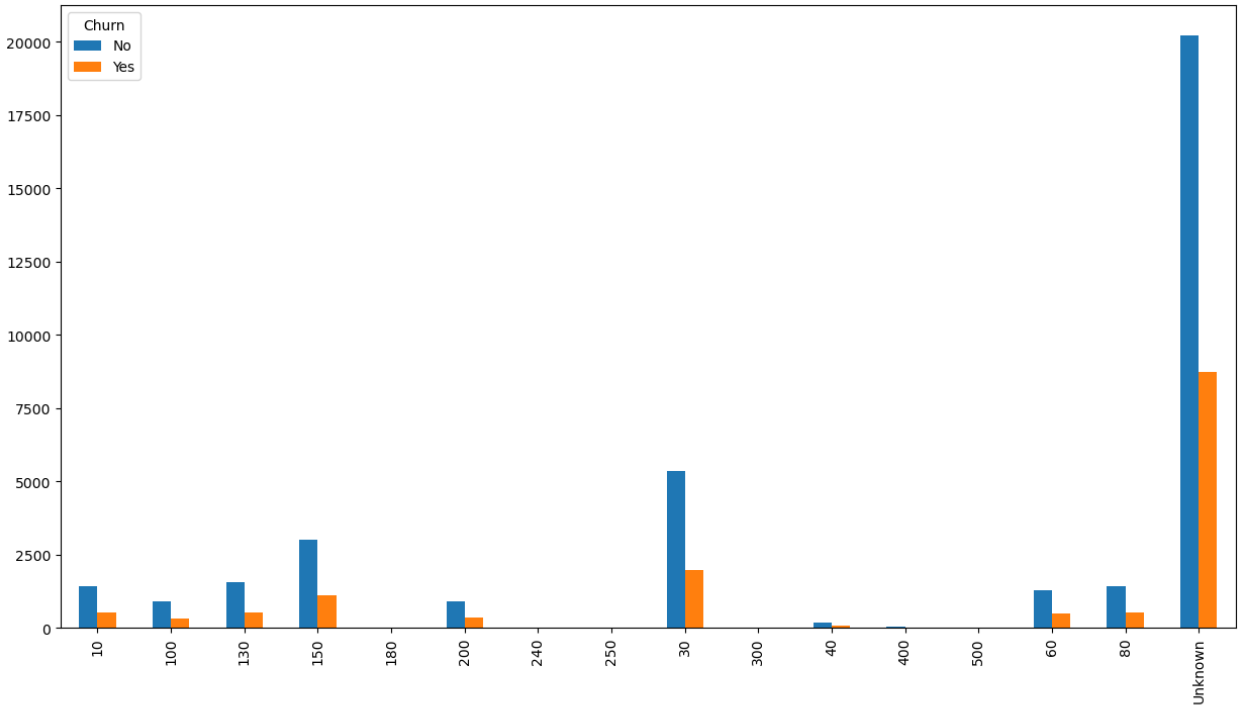
HasCreditCard


Churn
No
Yes
30000
25000
20000
15000
10000
5000
0
No
Yes
NewCellphoneUser

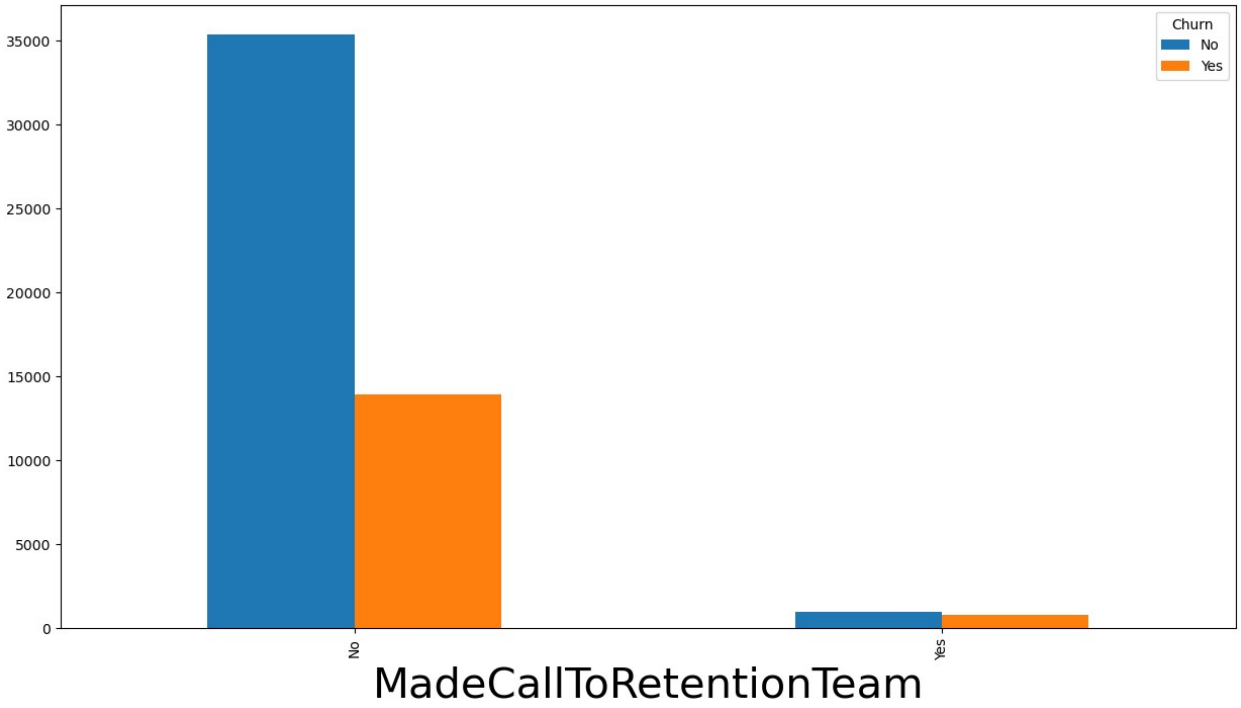

Churn
No
Yes
30000
25000
20000
15000
10000
5000
0
No
Yes
NotNewCellphoneUser

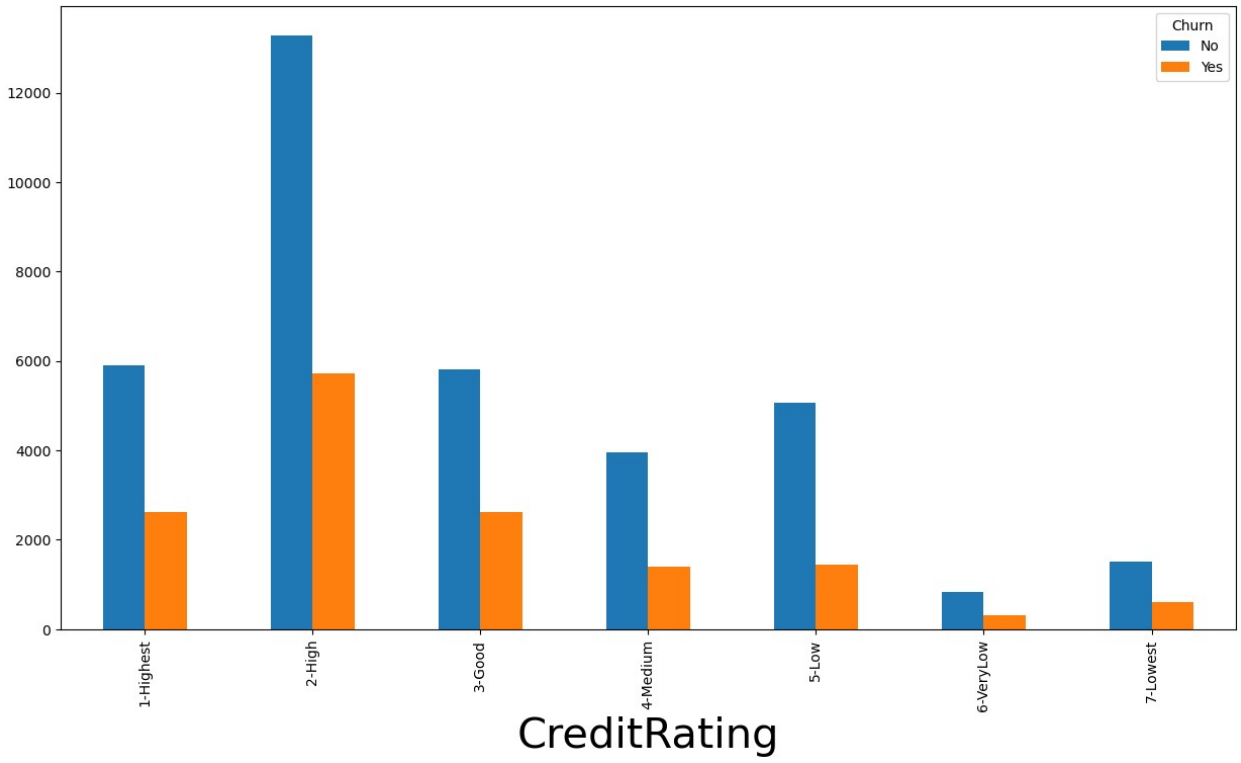


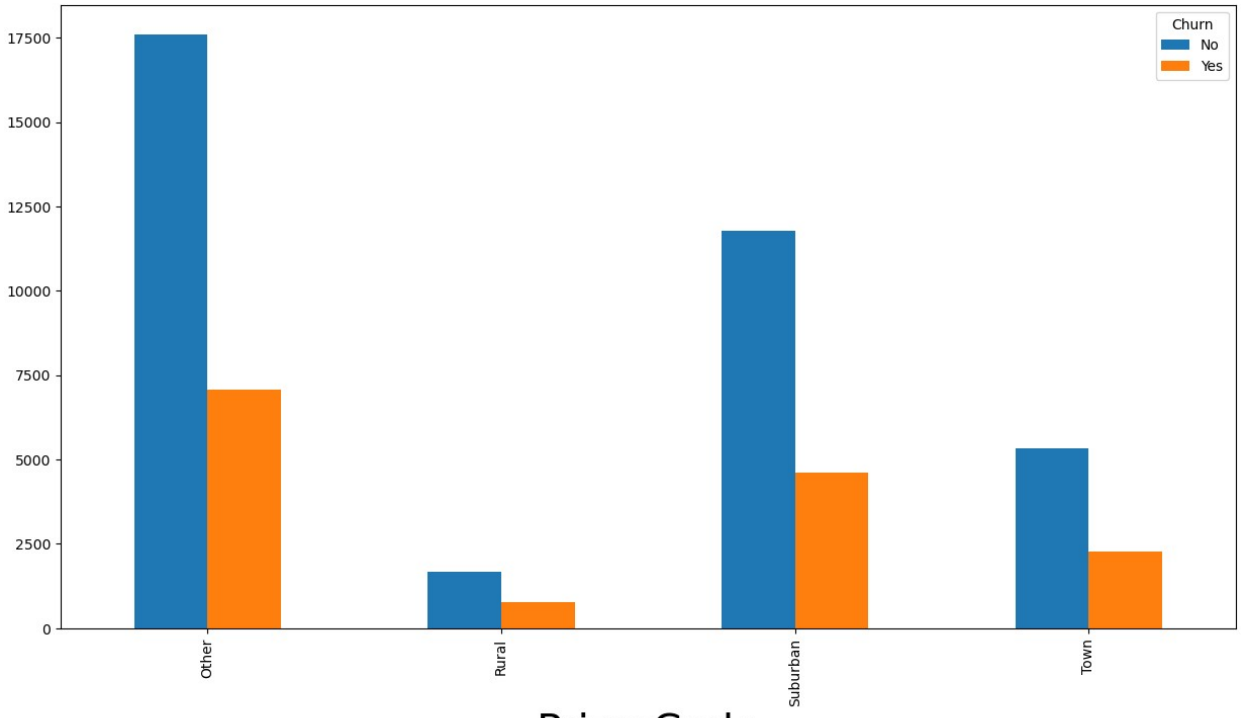
OwnsMotorcycle



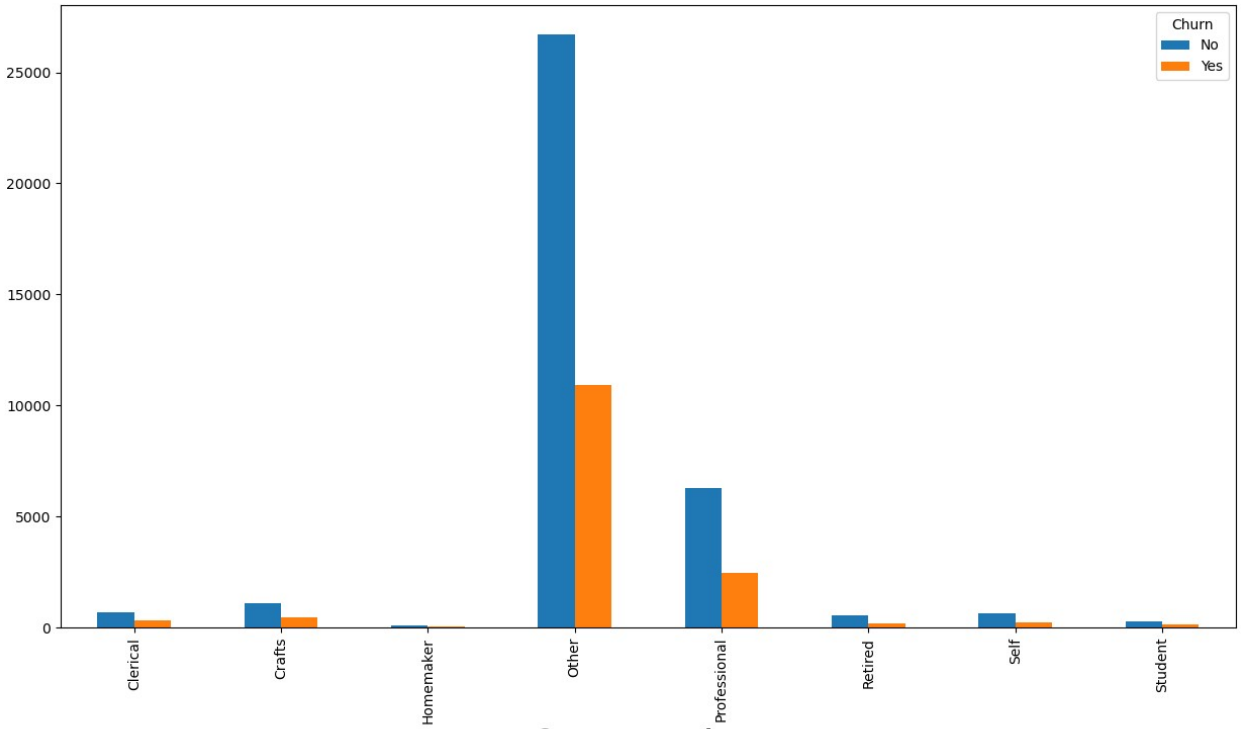
HandsetPrice


Churn
No
Yes
35000
30000
25000
20000
15000
10000
5000
0
No
Yes
MadeCallToRetentionTeam

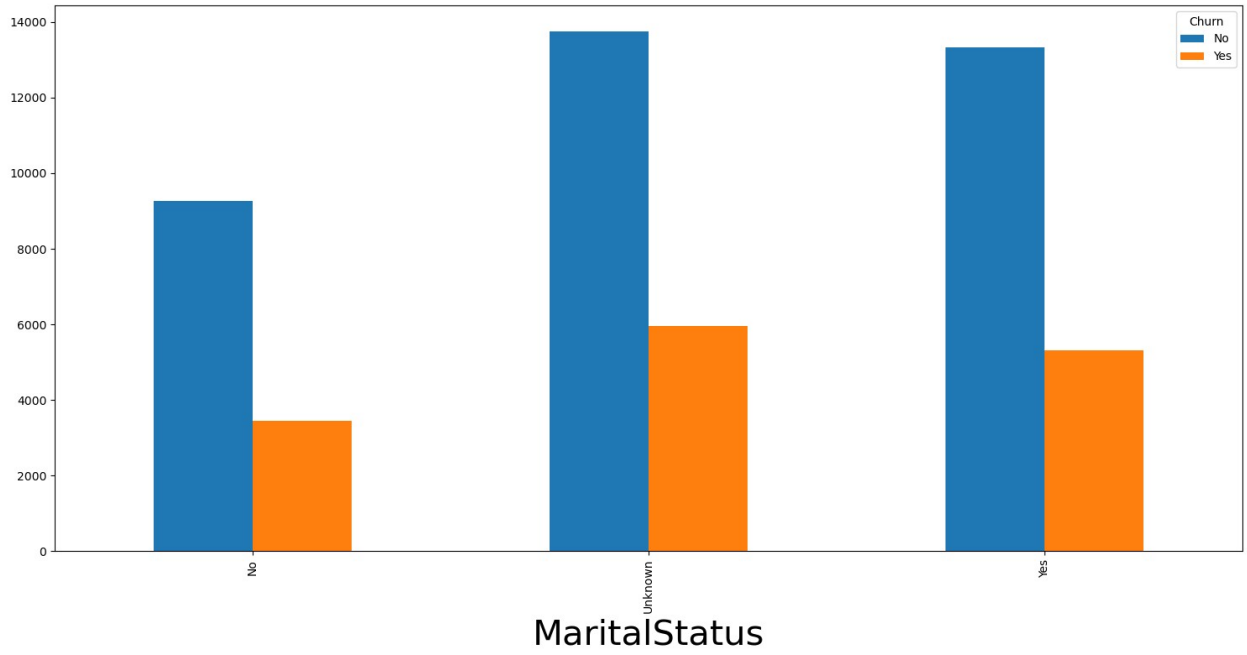

Churn
No
Yes
12000
10000
8000
6000
4000
2000
0
1-Highest
2-High
3-Good
4-Medium
5-Low
6-VeryLow
7-Lowest
CreditRating



PrizmCode



Occupation



- Here we found some misclassification in the data. HandsetPrice has a value known as `Unknown` and also some numerical variables. It has been categorized as object, but in realtime its numeric. Its not a good practice to replace all these Unknown values with the mean/median value of the remaining variables. Instead, we can find out the distribution of these `Unknown` values and prefer to drop them.

```
df.HandsetPrice.value_counts()

Unknown    28982
30          7328
150         4115
130         2105
80          1960
10          1928
60          1776
200         1266
100         1235
40           249
400           46
250           20
300           13
180           10
500            8
240            6
Name: HandsetPrice, dtype: int64

#Almost 57% values are unknown, so we can drop them.

df.HandsetPrice.value_counts()[0]/df.shape[0]

0.5677512880286795
```

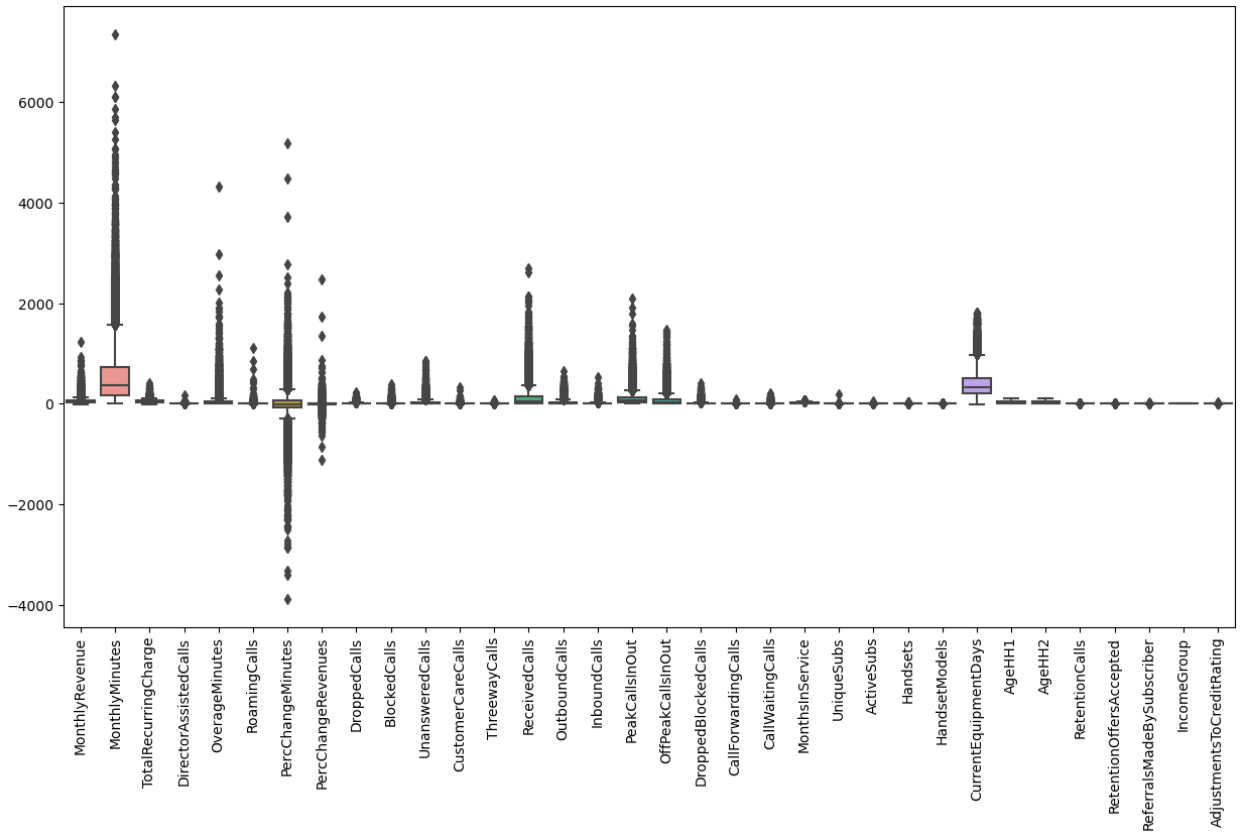
```
df.drop('HandsetPrice', axis =1, inplace = True)
```

Interpretation:

- When credit rating is high, there are chances of churning/not churning.
- Occupation - others, Prizmcode - others, Marital status - Unknown have the highest churn value.
- Also, We are going to drop this column HandsetPrice.

Outliers

```
sns.boxplot(df_num)
plt.xticks(rotation = 90)
plt.show()
```



Inference:

- There are many outliers in the dataset.
- We are having a series of negative values in PerChangeMinutes and PercChangeRevenues, lets check whether these values are correct.

```
df.describe()
```

```
       MonthlyRevenue  MonthlyMinutes  TotalRecurringCharge  \
count    51047.000000    51047.000000          51047.000000
mean        58.802788      525.165514             46.824495
std         44.442964      529.134170             23.812615
min         -6.170000        0.000000            -11.000000
25%         33.660000      159.000000             30.000000
50%         48.460000      366.000000             45.000000
75%         70.960000      722.000000             60.000000
max       1223.380000     7359.000000            400.000000

       DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
count           51047.000000    51047.000000  51047.000000
51047.000000
mean                0.893257       39.914628      1.232466          -
11.500833
std                 2.225423       96.462028      9.803517
256.587986
min                 0.000000        0.000000      0.000000       -
3875.000000
25%                 0.000000        0.000000      0.000000          -
82.000000
50%                 0.250000        3.000000      0.000000           -
5.000000
75%                 0.990000       40.000000      0.200000
65.000000
max               159.390000     4321.000000   1112.400000
5192.000000

       PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls
\
count        51047.000000  51047.000000  51047.000000     51047.000000

mean            -1.185572      6.011489      4.085672        28.288981

std             39.432467      9.043955     10.946905        38.876194

min          -1107.700000      0.000000      0.000000         0.000000

25%             -6.900000      0.700000      0.000000         5.300000

50%             -0.300000      3.000000      1.000000        16.300000

75%              1.550000      7.700000      3.700000        36.300000

max           2483.500000    221.700000    384.300000       848.700000


       CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls
\
count       51047.000000   51047.000000   51047.000000   51047.000000
```

```
mean             1.868999       0.298838     114.800121      25.377715

std              5.096138       1.168277     166.485896      35.209147

min              0.000000       0.000000       0.000000       0.000000

25%              0.000000       0.000000       8.300000       3.300000

50%              0.000000       0.000000      52.800000      13.700000

75%              1.700000       0.300000     153.500000      34.000000

max            327.300000      66.000000    2692.400000     644.300000


       InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
count  51047.000000    51047.000000       51047.000000
51047.000000
mean       8.178104       90.549515          67.650790
10.158003
std       16.665878      104.947470          92.752699
15.555284
min        0.000000        0.000000           0.000000
0.000000
25%        0.000000       23.000000          11.000000
1.700000
50%        2.000000       62.000000          35.700000
5.300000
75%        9.300000      121.300000          88.700000
12.300000
max      519.300000     2090.700000        1474.700000
411.700000

       CallForwardingCalls  CallWaitingCalls  MonthsInService
UniqueSubs  \
count         51047.000000      51047.000000     51047.000000
51047.000000
mean              0.012277          1.840504        18.756264
1.532157
std               0.594168          5.585129         9.800138
1.223384
min               0.000000          0.000000         6.000000
1.000000
25%               0.000000          0.000000        11.000000
1.000000
50%               0.000000          0.300000        16.000000
1.000000
75%               0.000000          1.300000        24.000000
```

```
2.000000
max              81.300000        212.700000        61.000000
196.000000

        ActiveSubs     Handsets  HandsetModels
CurrentEquipmentDays  \
count 51047.000000 51047.000000   51047.000000          51047.000000

mean      1.354340     1.805630       1.558740            380.544831

std       0.675477     1.331165       0.905927            253.799599

min       0.000000     1.000000       1.000000             -5.000000

25%       1.000000     1.000000       1.000000            205.000000

50%       1.000000     1.000000       1.000000            329.000000

75%       2.000000     2.000000       2.000000            515.000000

max      53.000000    24.000000      15.000000           1812.000000


            AgeHH1       AgeHH2  RetentionCalls
RetentionOffersAccepted  \
count 51047.000000 51047.000000    51047.000000
51047.000000
mean     31.421141    20.767626        0.037201
0.018277
std      21.905705    23.881611        0.206483
0.142458
min       0.000000     0.000000        0.000000
0.000000
25%       0.000000     0.000000        0.000000
0.000000
50%      36.000000     0.000000        0.000000
0.000000
75%      48.000000    42.000000        0.000000
0.000000
max      99.000000    99.000000        4.000000
3.000000

       ReferralsMadeBySubscriber  IncomeGroup
AdjustmentsToCreditRating
count               51047.000000 51047.000000
51047.000000
mean                    0.052070     4.324524
0.053911
std                     0.307592     3.138236
0.383147
min                     0.000000     0.000000
```

```
0.000000
25%                      0.000000     0.000000
0.000000
50%                      0.000000     5.000000
0.000000
75%                      0.000000     7.000000
0.000000
max                     35.000000     9.000000
25.000000

df[df.PercChangeMinutes<0].shape

(27183, 57)
```

- There are almost 27183 values negative in PercChangeMinutes, Lets not drop this and continue our analysis, check our results and if something goes wrong then we can remove these negative values.

```
df[df.MonthlyRevenue<0]

      CustomerID Churn  MonthlyRevenue  MonthlyMinutes
TotalRecurringCharge  \
26596    3210322    No       -2.520000      211.000000
0.000000
33352    3265738    No       -5.860000        0.000000              -
5.000000
48038    3378298    No       -6.170000        0.000000              -
6.000000

       DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
26596                0.330000        0.000000      0.000000           -
5.000000
33352                0.000000        0.000000      0.000000
0.000000
48038                0.000000        0.000000      0.000000
0.000000

       PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls
\
26596           -0.300000      0.000000      0.000000          7.300000

33352            5.900000      0.000000      0.000000          0.000000

48038            6.200000      0.000000      0.000000          0.000000


       CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls
\
26596           1.300000       0.000000       1.700000       8.000000
```

```
33352           0.000000       0.000000       0.000000       0.000000

48038           0.000000       0.000000       0.000000       0.000000


       InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
26596      1.300000       12.700000          17.000000
0.000000
33352      0.000000        0.000000           0.000000
0.000000
48038      0.000000        0.000000           0.000000
0.000000

       CallForwardingCalls  CallWaitingCalls  MonthsInService
UniqueSubs  \
26596             0.000000          0.000000               18
2
33352             0.000000          0.000000               15
3
48038             0.000000          0.000000                7
1

       ActiveSubs ServiceArea  Handsets  HandsetModels
CurrentEquipmentDays  \
26596           2   KCYKCK913  2.000000       2.000000
281.000000
33352           3   NEVLVS702  1.000000       1.000000
452.000000
48038           1   NYCSUF516  1.000000       1.000000
203.000000

         AgeHH1   AgeHH2 ChildrenInHH HandsetRefurbished
HandsetWebCapable  \
26596  0.000000 0.000000           No                Yes
Yes
33352 34.000000 0.000000           No                 No
Yes
48038  0.000000 0.000000           No                Yes
Yes

      TruckOwner RVOwner Homeownership BuysViaMailOrder
RespondsToMailOffers  \
26596         No      No       Unknown               No
No
33352         No      No         Known               No
No
48038         No      No         Known               No
No
```

```
      OptOutMailings NonUSTravel OwnsComputer HasCreditCard
RetentionCalls  \
26596             No          No            No             No
0
33352             No          No            No             No
0
48038             No          No            No             No
0

       RetentionOffersAccepted NewCellphoneUser NotNewCellphoneUser  \
26596                        0               No                  No
33352                        0               No                  No
48038                        0              Yes                  No

       ReferralsMadeBySubscriber  IncomeGroup OwnsMotorcycle  \
26596                          0            0             No
33352                          0            6             No
48038                          0            6             No

       AdjustmentsToCreditRating MadeCallToRetentionTeam CreditRating
\
26596                          0                      No       3-Good

33352                          0                      No       3-Good

48038                          0                      No     7-Lowest


      PrizmCode Occupation MaritalStatus
26596  Suburban      Other       Unknown
33352  Suburban      Other           Yes
48038  Suburban      Other            No

df[df.CurrentEquipmentDays<0].head(10)

     CustomerID Churn  MonthlyRevenue  MonthlyMinutes
TotalRecurringCharge  \
107     3000762   Yes      180.220000     3559.000000
150.000000
424     3003242    No       36.340000      247.000000
30.000000
2107    3016538    No       95.350000     1254.000000
85.000000
2145    3016766    No       88.980000        1.000000
50.000000
3883    3030326    No       62.170000      839.000000
81.000000
3923    3030654    No       71.600000      560.000000
51.000000
4075    3031906    No       45.640000      296.000000
```

```
45.000000
4318    3033854    No       36.520000      148.000000
45.000000
4319    3033858    No      125.190000      592.000000
87.000000
4732    3037134    No       77.410000      543.000000
75.000000

      DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
107               11.380000       99.000000      0.000000         -
149.000000
424                2.480000        7.000000      1.000000
154.000000
2107               0.000000       29.000000      0.000000         -
380.000000
2145               0.000000        0.000000      0.000000
0.000000
3883               0.250000        0.000000      0.000000
138.000000
3923               0.500000       18.000000      4.000000
287.000000
4075               9.900000        0.000000      0.000000
84.000000
4318               1.240000        0.000000      0.300000         -
125.000000
4319               0.000000        8.000000      0.000000
0.000000
4732               2.230000        0.000000      0.000000         -
364.000000

      PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls
\
107           -11.400000      3.700000     31.000000       120.700000

424            14.100000      2.700000      0.700000        21.300000

2107          -14.100000     14.300000      0.000000        25.300000

2145            0.000000      0.000000      0.000000         0.000000

3883           -3.500000      4.700000      5.000000        46.000000

3923           56.900000      3.700000      0.000000        15.300000

4075           12.400000      1.300000      3.000000         4.300000

4318           -0.400000      0.000000      0.000000         8.300000

4319            0.000000      0.000000      0.000000         0.000000
```

```
4732            0.300000      9.300000      0.700000        66.700000


      CustomerCareCalls  ThreewayCalls  ReceivedCalls
OutboundCalls  \
107             0.000000       0.300000    1543.700000       26.700000

424             0.000000       0.000000      14.000000       22.700000

2107            3.300000       0.000000     660.400000       32.300000

2145            0.000000       0.000000       0.000000        0.000000

3883            3.000000       0.300000     245.400000       21.700000

3923            0.000000       0.000000     160.300000       31.300000

4075            0.000000       0.000000      20.200000        4.700000

4318            1.700000       0.000000      17.600000        4.300000

4319            0.000000       0.000000       0.000000        0.000000

4732            0.000000       0.000000      80.200000       26.700000


      InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
107       6.700000      725.300000         468.700000
34.700000
424       3.700000       68.000000          25.000000
3.300000
2107     14.700000      316.000000         121.700000
14.300000
2145      0.000000        0.000000           0.000000
0.000000
3883      4.300000      245.000000          86.300000
9.700000
3923      3.000000       81.000000          48.700000
3.700000
4075      2.300000       53.700000          11.000000
4.300000
4318      1.000000       39.700000          20.300000
0.000000
4319      0.000000        0.000000           0.000000
0.000000
4732      7.000000      198.300000          70.300000
10.000000

      CallForwardingCalls  CallWaitingCalls  MonthsInService
```

```
UniqueSubs  \
107               0.000000         60.700000               55
3
424               0.000000          0.000000               52
1
2107              0.000000          3.000000               40
2
2145              0.000000          0.000000               40
1
3883              0.000000          9.000000               37
2
3923              0.000000          0.000000               37
2
4075              0.000000          0.000000               36
1
4318              0.000000          0.300000               36
1
4319              0.000000          0.000000               36
1
4732              0.000000          2.000000               35
1

      ActiveSubs ServiceArea  Handsets  HandsetModels
CurrentEquipmentDays  \
107            2   OKCLRK501 10.000000       5.000000             -
3.000000
424            1   SEASEA206  4.000000       4.000000             -
1.000000
2107           2   NYCSUF516  5.000000       4.000000             -
3.000000
2145           1   SANCRP512  4.000000       3.000000             -
1.000000
3883           1   MIANDA305  8.000000       7.000000             -
3.000000
3923           1   SFRSFR415  3.000000       3.000000             -
2.000000
4075           1   HOUHOU281  7.000000       4.000000             -
5.000000
4318           1   OHICOL614  5.000000       4.000000             -
2.000000
4319           1   HOUHOU281  2.000000       2.000000             -
1.000000
4732           1   MILMIL414  2.000000       2.000000             -
3.000000

        AgeHH1    AgeHH2 ChildrenInHH HandsetRefurbished
HandsetWebCapable  \
107   0.000000  0.000000           No                 Yes
Yes
```

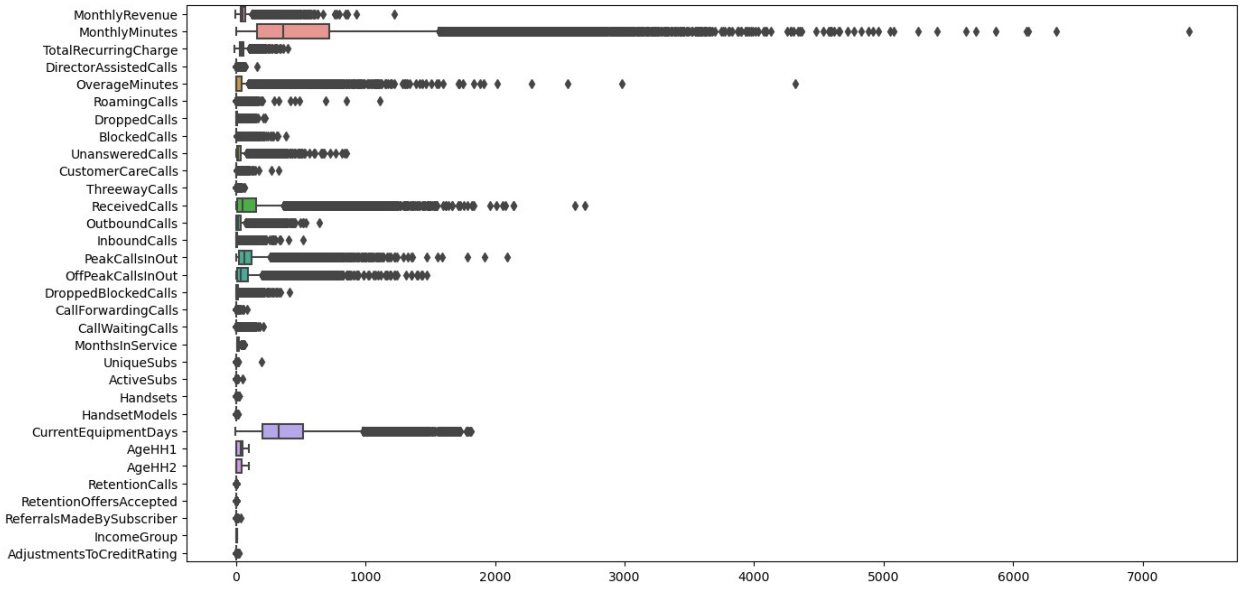
```
424  62.000000 54.000000           No                  No
Yes
2107 46.000000 42.000000          Yes                  No
Yes
2145 60.000000 30.000000          Yes                  No
Yes
3883  0.000000  0.000000           No                  No
Yes
3923 50.000000  0.000000           No                  No
Yes
4075 40.000000 56.000000           No                  No
Yes
4318 28.000000 26.000000           No                 Yes
Yes
4319 32.000000 22.000000           No                  No
Yes
4732 54.000000  0.000000           No                  No
Yes

     TruckOwner RVOwner Homeownership BuysViaMailOrder
RespondsToMailOffers  \
107          No      No       Unknown                No
No
424          No      No         Known               Yes
Yes
2107        Yes     Yes         Known               Yes
Yes
2145        Yes     Yes         Known               Yes
Yes
3883         No      No       Unknown                No
No
3923         No      No         Known                No
No
4075        Yes      No         Known                No
No
4318        Yes     Yes         Known                No
No
4319         No      No         Known                No
No
4732         No      No         Known               Yes
Yes

     OptOutMailings NonUSTravel OwnsComputer HasCreditCard
RetentionCalls  \
107              No          No            No             No
0
424              No          No           Yes            Yes
1
2107             No          No           Yes            Yes
```

```
0
2145              No           No            No            Yes
0
3883              No           No            No             No
0
3923              No           No            No            Yes
0
4075              No           No            No            Yes
0
4318              No           No            No            Yes
0
4319              No           No            No            Yes
0
4732              No           No            No            Yes
0

      RetentionOffersAccepted NewCellphoneUser NotNewCellphoneUser  \
107                         0               No                  Yes
424                         1               No                   No
2107                        0               No                   No
2145                        0               No                   No
3883                        0              Yes                   No
3923                        0              Yes                   No
4075                        0              Yes                   No
4318                        0              Yes                   No
4319                        0               No                   No
4732                        0              Yes                   No

      ReferralsMadeBySubscriber  IncomeGroup OwnsMotorcycle  \
107                           0            0             No
424                           0            6             No
2107                          0            6             No
2145                          0            6             No
3883                          0            0             No
3923                          0            5             No
4075                          0            5             No
4318                          0            6             No
4319                          0            3             No
4732                          0            8             No

      AdjustmentsToCreditRating MadeCallToRetentionTeam
CreditRating  \
107                           0                      No       3-Good

424                           0                     Yes    1-Highest

2107                          0                      No    1-Highest

2145                          2                      No    6-VeryLow
```

```
3883                              2                        No    1-Highest

3923                              0                        No    1-Highest

4075                              0                        No    1-Highest

4318                              0                        No     4-Medium

4319                              1                        No    6-VeryLow

4732                              0                        No    1-Highest
```

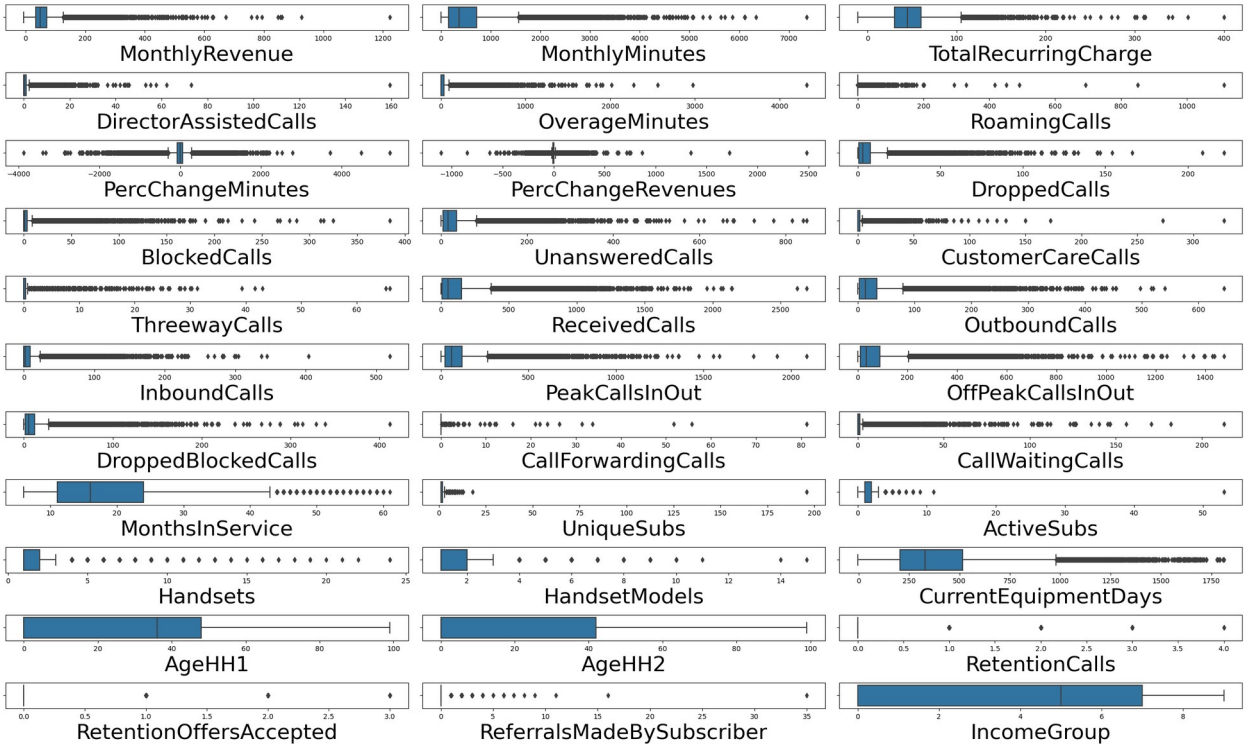
```
     PrizmCode    Occupation MaritalStatus
107      Other         Other       Unknown
424       Town          Self           Yes
2107     Other          Self           Yes
2145     Other  Professional       Unknown
3883     Other         Other       Unknown
3923     Other         Other       Unknown
4075     Other         Other            No
4318      Town         Other           Yes
4319  Suburban         Other           Yes
4732     Other  Professional           Yes
```

```
# Basically checking the values which are less than 0(Negative values)

sns.boxplot(df_num.drop(['PercChangeMinutes', 'PercChangeRevenues'],
axis =1), orient = 'h')
#plt.xticks(rotation = 90)
plt.show()
```



- There are many outliers in the data, We cannot simple drop the outliers using IQR or Z score method, we have to analyze each variable and check whether dropping the outliers make sense.

```
fig, ax = plt.subplots(11, 3, figsize=(25, 15))

for variable, subplot in zip(df_num, ax.flatten()):
    boxplt = sns.boxplot(x=variable,data=df_num, ax=subplot)
    boxplt.set_xlabel(variable, fontsize = 30)

plt.tight_layout()
plt.show()
```




```
# The above boxplots gives us a better understanding of the outliers.

#Lets convert the Churn values to 0 and 1.
df['Churn'].replace(to_replace='Yes', value=1, inplace=True)
df['Churn'].replace(to_replace='No',  value=0, inplace=True)

df.head()
```

```
  CustomerID  Churn  MonthlyRevenue  MonthlyMinutes
TotalRecurringCharge  \
0    3000002      1       24.000000      219.000000
22.000000
1    3000010      1       16.990000       10.000000
17.000000
2    3000014      0       38.000000        8.000000
```

```
38.000000
3    3000022      0       82.280000     1312.000000
75.000000
4    3000026      1       17.140000        0.000000
17.000000

   DirectorAssistedCalls  OverageMinutes  RoamingCalls
PercChangeMinutes  \
0               0.250000        0.000000      0.000000         -
157.000000
1               0.000000        0.000000      0.000000           -
4.000000
2               0.000000        0.000000      0.000000           -
2.000000
3               1.240000        0.000000      0.000000
157.000000
4               0.000000        0.000000      0.000000
0.000000

   PercChangeRevenues  DroppedCalls  BlockedCalls  UnansweredCalls  \
0          -19.000000      0.700000      0.700000         6.300000
1            0.000000      0.300000      0.000000         2.700000
2            0.000000      0.000000      0.000000         0.000000
3            8.100000     52.000000      7.700000        76.000000
4           -0.200000      0.000000      0.000000         0.000000

   CustomerCareCalls  ThreewayCalls  ReceivedCalls  OutboundCalls  \
0           0.000000       0.000000      97.200000       0.000000
1           0.000000       0.000000       0.000000       0.000000
2           0.000000       0.000000       0.400000       0.300000
3           4.300000       1.300000     200.300000     370.300000
4           0.000000       0.000000       0.000000       0.000000

   InboundCalls  PeakCallsInOut  OffPeakCallsInOut
DroppedBlockedCalls  \
0      0.000000       58.000000          24.000000
1.300000
1      0.000000        5.000000           1.000000
0.300000
2      0.000000        1.300000           3.700000
0.000000
3    147.000000      555.700000         303.700000
59.700000
4      0.000000        0.000000           0.000000
0.000000

   CallForwardingCalls  CallWaitingCalls  MonthsInService  UniqueSubs
\
0             0.000000          0.300000               61           2
```

```
1             0.000000          0.000000                58            1

2             0.000000          0.000000                60            1

3             0.000000         22.700000                59            2

4             0.000000          0.000000                53            2


   ActiveSubs ServiceArea  Handsets  HandsetModels
CurrentEquipmentDays  \
0           1   SEAPOR503  2.000000       2.000000
361.000000
1           1   PITHOM412  2.000000       1.000000
1504.000000
2           1   MILMIL414  1.000000       1.000000
1812.000000
3           2   PITHOM412  9.000000       4.000000
458.000000
4           2   OKCTUL918  4.000000       3.000000
852.000000

     AgeHH1    AgeHH2 ChildrenInHH HandsetRefurbished
HandsetWebCapable  \
0 62.000000  0.000000           No                 No
Yes
1 40.000000 42.000000          Yes                 No
No
2 26.000000 26.000000          Yes                 No
No
3 30.000000  0.000000           No                 No
Yes
4 46.000000 54.000000           No                 No
No

  TruckOwner RVOwner Homeownership BuysViaMailOrder
RespondsToMailOffers  \
0         No      No         Known               Yes
Yes
1         No      No         Known               Yes
Yes
2         No      No       Unknown                No
No
3         No      No         Known               Yes
Yes
4         No      No         Known               Yes
Yes

  OptOutMailings NonUSTravel OwnsComputer HasCreditCard
RetentionCalls  \
```

```
0              No           No           Yes            Yes
1
1              No           No           Yes            Yes
0
2              No           No            No            Yes
0
3              No           No            No            Yes
0
4              No           No           Yes            Yes
0

   RetentionOffersAccepted NewCellphoneUser NotNewCellphoneUser  \
0                        0               No                  No
1                        0              Yes                  No
2                        0              Yes                  No
3                        0              Yes                  No
4                        0               No                 Yes

   ReferralsMadeBySubscriber  IncomeGroup OwnsMotorcycle  \
0                          0            4             No
1                          0            5             No
2                          0            6             No
3                          0            6             No
4                          0            9             No

   AdjustmentsToCreditRating MadeCallToRetentionTeam CreditRating
PrizmCode  \
0                          0                     Yes    1-Highest
Suburban
1                          0                      No     4-Medium
Suburban
2                          0                      No       3-Good
Town
3                          0                      No     4-Medium
Other
4                          1                      No    1-Highest
Other

     Occupation MaritalStatus
0  Professional            No
1  Professional           Yes
2        Crafts           Yes
3         Other            No
4  Professional           Yes

df.Churn = df.Churn.astype(int)

df.Churn.dtypes

dtype('int64')
```

Summary:

- These are the basic preprocessing steps that are done.
- In the next phase, we will be removing the outliers by analyzing all the data variables correctly.
- Doing feature engineering.
- Also other processing steps like encoding the categorical variables and scaling the numerical values.
- Since this is a classification problem, we will be building a Logistic Regression as the base model and then building a Random forest or any other high end algorithm.
- In this module, we have understood the relationship between the target variable and the categorical variables and also the distribution of Boxplot etc..
- In the early stage, where we have to replace the null values, we have used the median value, we can also drop the rows which are having null values, since a very small number of missing values are present in the data.