DATA ANALYSIS - TELECOM COMPANY

Introduction:

Telecom is a fictional telecommunication company that provides internet and phone services to 7043 customers in California. Churn data is provided by the 'Maven Analytics'.

Problem Statement:

Company want to improve retention by identifying high value customers and churn risk. Stakeholders want a report showing trends and looking for some recommendations for increasing the revenue.

Data collection:

Data was provided by 'Maven Analytics' for a fictional company. Data includes 3 files with details about customer demographics, locations , service and current status. Csv file with 7043 rows and 38 columns was used for this EDA. It is Third party data.

Data Cleaning:

Data cleaning includes finding and removing blanks or any duplicate entries. Also verifying the data to be consistent. Before starting, required libraries were imported.Then data was imported and called up for doing further analysis. Data analysis was done by using 'Python' for this case study.

In [2]:
```
# importing the required Libraries
import pandas as pd
```

In [3]:
```
import numpy as np
```

In [4]:
```
# import & export CSV files
CustomerData = pd.read_csv('~/telecom_customer_churn.csv')
```

Table used for this analysis

In [5]:
```
#Printing the exported CSV
CustomerData
```

Out[5]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903
...	...	...	...	...	...	...	...	...	...
7038	9987-LUTYD	Female	20	No	0	La Mesa	91941	32.759327	-116.997260
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	95367	37.734971	-120.954271
7040	9992-UJOEL	Male	22	No	0	Elk	95432	39.108252	-123.645121
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	92075	33.001813	-117.263628
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	96125	39.600599	-120.636358

7043 rows × 38 columns

Exploring the data by using info() function , it shows us total rows and columns in dataframe

In [6]:
```
CustomerData.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 38 columns):
 #   Column                             Non-Null Count  Dtype  
---  ------                             --------------  -----  
 0   Customer ID                        7043 non-null   object 
 1   Gender                             7043 non-null   object 
 2   Age                                7043 non-null   int64  
 3   Married                            7043 non-null   object 
 4   Number of Dependents               7043 non-null   int64  
 5   City                               7043 non-null   object 
 6   Zip Code                           7043 non-null   int64  
 7   Latitude                           7043 non-null   float64
 8   Longitude                          7043 non-null   float64
 9   Number of Referrals                7043 non-null   int64  
 10  Tenure in Months                   7043 non-null   int64  
 11  Offer                              7043 non-null   object 
 12  Phone Service                      7043 non-null   object 
 13  Avg Monthly Long Distance Charges  6361 non-null   float64
 14  Multiple Lines                     6361 non-null   object 
 15  Internet Service                   7043 non-null   object 
 16  Internet Type                      5517 non-null   object 
 17  Avg Monthly GB Download            5517 non-null   float64
 18  Online Security                    5517 non-null   object 
 19  Online Backup                      5517 non-null   object 
 20  Device Protection Plan             5517 non-null   object 
 21  Premium Tech Support               5517 non-null   object 
 22  Streaming TV                       5517 non-null   object 
 23  Streaming Movies                   5517 non-null   object 
 24  Streaming Music                    5517 non-null   object 
 25  Unlimited Data                     5517 non-null   object 
 26  Contract                           7043 non-null   object 
 27  Paperless Billing                  7043 non-null   object 
 28  Payment Method                     7043 non-null   object 
 29  Monthly Charge                     7043 non-null   float64
 30  Total Charges                      7043 non-null   float64
 31  Total Refunds                      7043 non-null   float64
 32  Total Extra Data Charges           7043 non-null   int64  
 33  Total Long Distance Charges        7043 non-null   float64
 34  Total Revenue                      7043 non-null   float64
 35  Customer Status                    7043 non-null   object 
 36  Churn Category                     1869 non-null   object 
 37  Churn Reason                       1869 non-null   object 
dtypes: float64(9), int64(6), object(23)
memory usage: 2.0+ MB
```

To check Mean, standard deviation, min and max values, we can use describe() function. It helps to know more about data and get an idea about all statistics of the data frame

In [7]:
```
CustomerData.describe()
```

Out[7]:

	Age	Number of Dependents	Zip Code	Latitude	Longitude	Number of Referrals	Tenure i Montl
count	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.00000
mean	46.509726	0.468692	93486.070567	36.197455	-119.756684	1.951867	32.38676
std	16.750352	0.962802	1856.767505	2.468929	2.154425	3.001199	24.54206
min	19.000000	0.000000	90001.000000	32.555828	-124.301372	0.000000	1.00000
25%	32.000000	0.000000	92101.000000	33.990646	-121.788090	0.000000	9.00000
50%	46.000000	0.000000	93518.000000	36.205465	-119.595293	0.000000	29.00000
75%	60.000000	0.000000	95329.000000	38.161321	-117.969795	3.000000	55.00000
max	80.000000	9.000000	96150.000000	41.962127	-114.192901	11.000000	72.00000

Checking for Null Values

In [16]:
```
# Lets check null values and try to clean it
#replacing NaN with 0
CustomerData.fillna(0)
```

Out[16]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903
...	...	...	...	...	...	...	...	...	...
7038	9987-LUTYD	Female	20	No	0	La Mesa	91941	32.759327	-116.997260
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	95367	37.734971	-120.954271
7040	9992-UJOEL	Male	22	No	0	Elk	95432	39.108252	-123.645121
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	92075	33.001813	-117.263628
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	96125	39.600599	-120.636358

7043 rows × 38 columns

Dropping the blank values , Result showed 1586 rows left afte removing empty cells

In [17]:
```
CustomerData.dropna()
```

Out[17]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903
18	0022-TCJCI	Male	79	No	0	Daly City	94015	37.680844	-122.481310
21	0023-XUOPT	Female	26	Yes	0	Carnelian Bay	96140	39.227434	-120.091806
...	...	...	...	...	...	...	...	...	...
7013	9948-YPTDG	Male	44	Yes	0	San Rafael	94903	38.018065	-122.546024
7020	9961-JBNMK	Male	79	No	0	Desert Center	92239	33.889605	-115.257009
7023	9965-YOKZB	Male	73	No	0	Grizzly Flats	95636	38.636102	-120.522149
7036	9985-MWVIX	Female	53	No	0	Hume	93628	36.807595	-118.901544
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	95367	37.734971	-120.954271

1586 rows × 38 columns

Sorting and filtering

In [8]:
```
#Result of dropping shows that many rows have been removed as those were empty.
# Now for sorting on the basic of Total revenue from customers

TopCustomer = CustomerData.sort_values(by= ['Total Revenue'],ascending=False)
```

In [13]:
```
TopCustomer
```

Out[13]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude
95	0164-APGRB	Female	56	No	0	Calistoga	94515	38.629618	-122.593216
5879	8263-QMNTJ	Male	46	Yes	0	San Francisco	94104	37.791222	-122.402241
3879	5451-YHYPW	Female	75	Yes	0	San Francisco	94129	37.797526	-122.464531
2686	3810-DVDQQ	Female	28	Yes	1	Shasta Lake	96019	40.692523	-122.369876
5360	7569-NMZYQ	Female	33	Yes	3	Middletown	95461	38.787446	-122.586750
...	...	...	...	...	...	...	...	...	...
3589	5088-QZLRL	Male	50	No	0	Corona	92881	33.833686	-117.513063
2795	3976-NLDEZ	Male	55	No	0	Mira Loma	91752	33.999992	-117.535395
4690	6615-ZGEDR	Male	42	No	0	Smith River	95567	41.950683	-124.097094
4726	6654-QGBZZ	Female	21	No	0	Nevada City	95959	39.333737	-120.858667
5576	7853-WNZSY	Male	58	No	0	El Nido	95317	37.127386	-120.506422

7043 rows × 38 columns

Changing Data type for field required for analysis

In [30]:
```
TopCustomer['Number of Referrals'] = TopCustomer['Number of Referrals'].astype(int
```

In [50]:
```
type(TopCustomer['Number of Referrals'])
```

Out[50]:
```
pandas.core.series.Series
```

Data Analysis:

This is mainly about discovering some useful insights so that company can make data-driven decision for its growth. So, follwoing steps were taken for analysis.

Creating pivot table quickly help to summarize the important numerical field. Here , as my focus is on churn, created pivot table for Customer status(chruned/joined/stayed) to know the effect on Total revenue.

In [12]:
```
StatusVsRevenue = TopCustomer.pivot_table(index='Customer Status', values='Total R
StatusVsRevenue
```

Out[12]:

Customer Status	Total Revenue
Churned	3684459.82
Joined	54279.75
Stayed	17632392.12

In [30]:
```
StatusVsRevenue.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3 entries, Churned to Stayed
Data columns (total 1 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Total Revenue  3 non-null      float64
dtypes: float64(1)
memory usage: 48.0+ bytes
```

Another pivot table for checking number of referral by Churned Customers

In [5]:
```
# Also considering 'Referral' provided by each type of customer status
ReferralsVsRevenue = pd.pivot_table( CustomerData, index=["Customer Status"], valu
```

In [97]:
```
ReferralsVsRevenue
```

Out[97]:

Customer Status	Number of Referrals	Total Revenue
Churned	974	3684459.82
Joined	431	54279.75
Stayed	12342	17632392.12

In [13]:
```
a=(ReferralsVsRevenue['Total Revenue'])
print(a)
```

```
Customer Status
Churned     3684459.82
Joined        54279.75
Stayed     17632392.12
Name: Total Revenue, dtype: float64
```

Finding out the percentage of revenue came from

Customers who left the company. This will help us to know how much risk company can face from these customers.

In [106…
```
# finidng difference in revenue from stayed and churned customers
loss=(a[2]-a[0])
print(loss)
```

```
13947932.3
```

In [109…
```
Grand_Revenue= (a[0]+a[1]+a[2])
print(Grand_Revenue)
```

```
21371131.69
```

In [129…
```
# Showing loss in percentage,17% is occupied by customers who churned at the end o
percentage=(a[0]/Grand_Revenue)*100
print(percentage.round(2))
```

```
17.24
```

Also checking how many referrals were given by Churned customers

In [6]:
```
#Checking referrals by churned customers (in percentage)
b=(ReferralsVsRevenue['Number of Referrals'])
print(b)
total_ref=(b[0]+b[1]+b[2])
print(total_ref)
```

```
Customer Status
Churned      974
Joined       431
Stayed     12342
Name: Number of Referrals, dtype: int64
13747
```

In [11]:
```
ref_perc=(b[0]/total_ref)*100
print((ref_perc).round(2))
```

```
7.09
```

Analysing the reasons behind leaving Telecom. So it shows 20 reasons, company need to work on this for improving retention.

In [31]:
```
List_of_reasons=(CustomerData['Churn Reason'].unique())
print(List_of_reasons)
```

```
[nan 'Competitor had better devices' 'Product dissatisfaction'
 'Network reliability' 'Limited range of services'
 'Competitor made better offer' "Don't know" 'Long distance charges'
 'Attitude of service provider' 'Attitude of support person'
 'Competitor offered higher download speeds'
 'Competitor offered more data' 'Lack of affordable download/upload speed'
 'Deceased' 'Moved' 'Service dissatisfaction' 'Price too high'
 'Lack of self-service on Website' 'Poor expertise of online support'
 'Extra data charges' 'Poor expertise of phone support']
```

In [33]:
```
Reason_for_churn = CustomerData.groupby('Churn Reason')['Churn Reason'].count()
print(Reason_for_churn)
```

```
Churn Reason
Attitude of service provider                  94
Attitude of support person                   220
Competitor had better devices                313
Competitor made better offer                 311
Competitor offered higher download speeds    100
Competitor offered more data                 117
Deceased                                       6
Don't know                                   130
Extra data charges                            39
Lack of affordable download/upload speed      30
Lack of self-service on Website               29
Limited range of services                     37
Long distance charges                         64
Moved                                         46
Network reliability                           72
Poor expertise of online support              31
Poor expertise of phone support               12
Price too high                                78
Product dissatisfaction                       77
Service dissatisfaction                       63
Name: Churn Reason, dtype: int64
```

So, three most common reasons that most customer gave while leaving are:

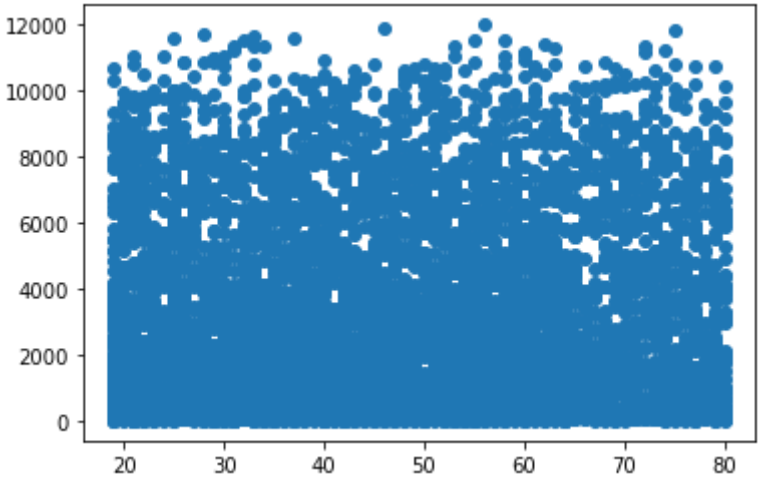
1. Competitor having better devices
2. Better offers from competitor company
3. Support person attitude issues.

Data Visualization:

First we need to import 'matplotlib' for creating visuals.

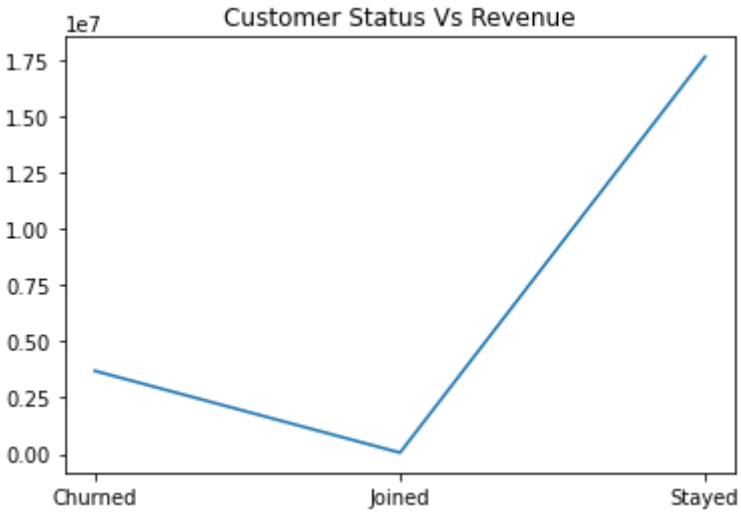
1. Scatter plot for Age and revenue: Scatter plot shows no relation between two and also the ages of customer is from all age groups.

In [29]:
```
import matplotlib.pyplot as plt


#Age Vs Revenue
x = TopCustomer['Age']
y = TopCustomer['Total Revenue']
plt.scatter(x, y)

plt.show()
```



2. Plot showing the trend for three categories of 'Customer status'

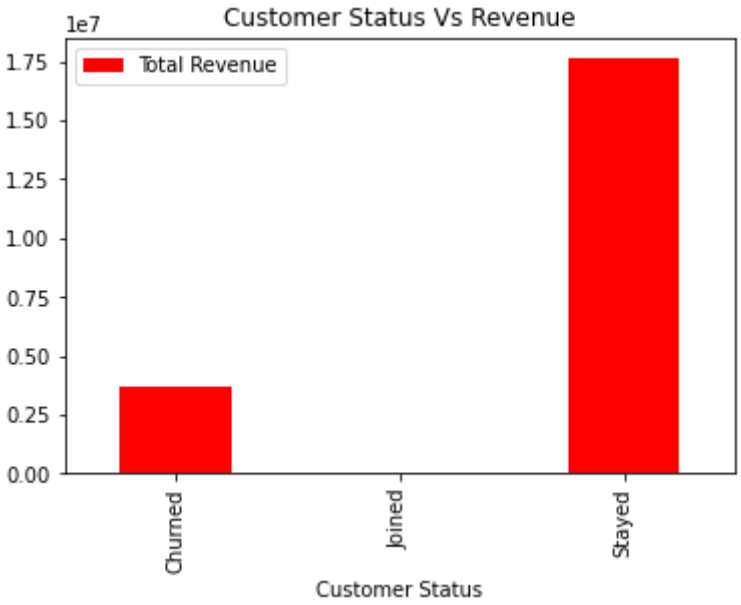
In [31]:
```
  # Creating a plot to see realation of Customer status with Revenue generated
plt.plot(StatusVsRevenue["Total Revenue"])

# Title to the plot
plt.title("Customer Status Vs Revenue")


plt.show()
```



3. Bar chart showing how much churned categort contributes to total revenue. Joined plays negligible role as per the data.

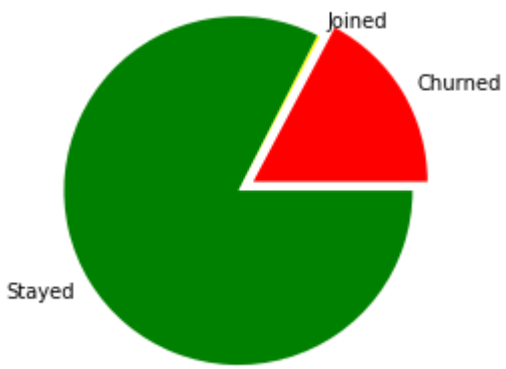
In [62]:
```
# Creating bar chart
StatusVsRevenue.plot(kind ="bar",title="Customer Status Vs Revenue" ,color='red')
```

Out[62]:
```
<AxesSubplot:title={'center':'Customer Status Vs Revenue'}, xlabel='Customer Statu
s'>
```



4. Pie Chart highlighting the part of revenue from Churned customer

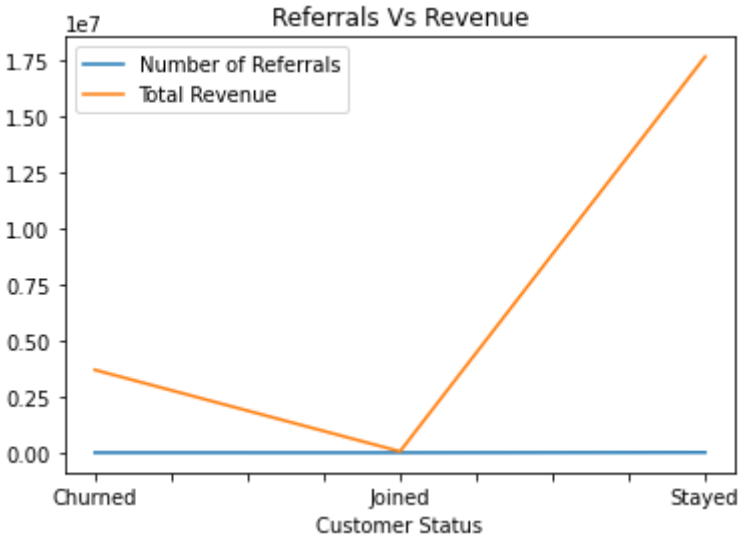
In [17]:
```
y = np.array([a[0],a[1],a[2]])
mylabels = ["Churned", "Joined", "Stayed"]
myexplode = [0.1, 0, 0, ]
mycolors = ["red", "yellow", "green"]
plt.pie(y, labels = mylabels, explode = myexplode, colors = mycolors )
plt.show()
```



5. Line graph showing both 'Total Revenue' and 'Referrals' from all Customers

In [65]:
```
ReferralsVsRevenue.plot(kind ="line",title="Referrals Vs Revenue")
# Number of Referrals don't play any critical role for revenue
```

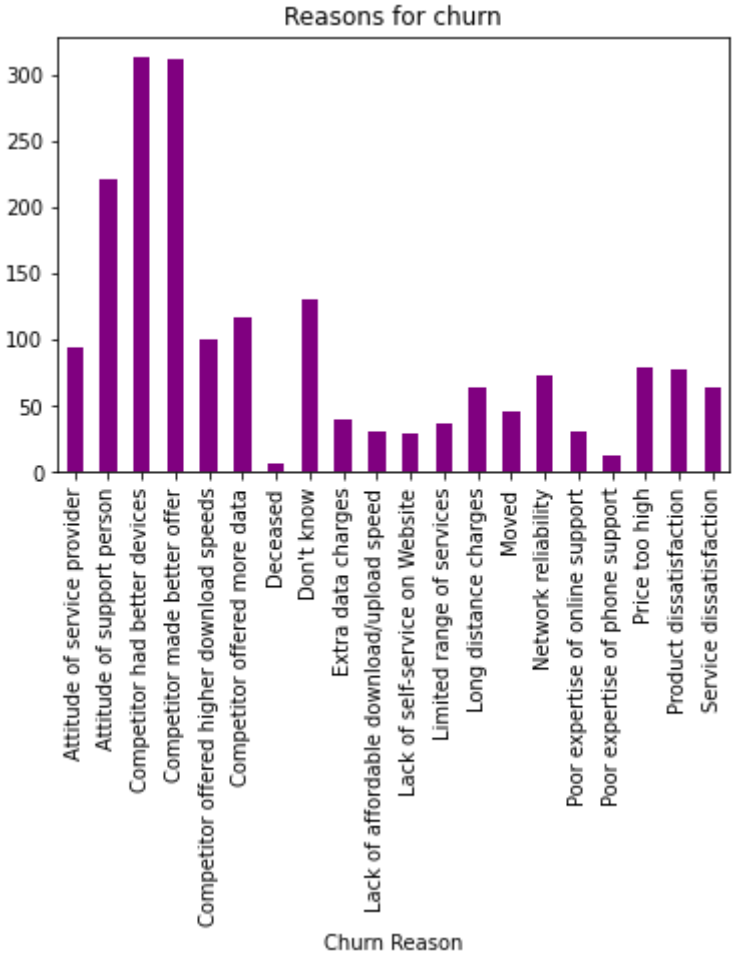
Out[65]:
```
<AxesSubplot:title={'center':'Referrals Vs Revenue'}, xlabel='Customer Status'>
```



6 . Bar graph showing the reason behind leaving company. Bar graph clearly shows 3 most common reason that many customer talked about while leaving company.

In [13]:
```
Reason_for_churn.plot(kind ="bar",title="Reasons for churn" ,color='purple')
```

Out[13]:
```
<AxesSubplot:title={'center':'Reasons for churn'}, xlabel='Churn Reason'>
```



Sharing the insights from data

1. 17% of the total revenue is generated by the customer who left the company which is great risk for company.

2. 7% of total referral are from churned Customer

3. Age of customer is not a factor that effecting revenue

4. Top 3 most common reasons that customer are leaving and joining competitor are showing that Support people are not good, Telecom offers and amount of data given by telecom is less compared to competitor company.

Act (Recommendations)

1. To improve the situation, training to 'support persons' on monthly basis focusing on 'assurance' and 'empthy' with customers. Also survey can be created for employees to know why attitude is one of the reason for customer leaving.

2. Rechecking of the offers given to customers and providing better offers to those who are loyal with company for many years.

3. Company can launch of scheme of 'gift hampers' after some time interval to improve customer-company relationship