DATA COLLECTION

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
#Load the dataset
df  = pd.read_csv('telecom_customer_churn.csv')
df
```

Out[2]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude	Number of Referrals	...	Payment Method	Monthly Charge
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073	2	...	Credit Card	65.60
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869	0	...	Credit Card	-4.00
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613	0	...	Bank Withdrawal	73.90
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432	1	...	Bank Withdrawal	98.00
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903	3	...	Credit Card	83.90
...	...	...	...	...	...	...	...	...	...	...	...	...	...
7038	9987-LUTYD	Female	20	No	0	La Mesa	91941	32.759327	-116.997260	0	...	Credit Card	55.15
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	95367	37.734971	-120.954271	1	...	Bank Withdrawal	85.10
7040	9992-UJOEL	Male	22	No	0	Elk	95432	39.108252	-123.645121	0	...	Credit Card	50.30
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	92075	33.001813	-117.263628	5	...	Credit Card	67.85
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	96125	39.600599	-120.636358	1	...	Bank Withdrawal	59.00

7043 rows × 38 columns

In [3]:
```
#Understanding the Dataset
print(df.head())
```

```
  Customer ID  Gender  Age Married  Number of Dependents          City  \
0  0002-ORFBO  Female   37     Yes                     0  Frazier Park
1  0003-MKNFE    Male   46      No                     0      Glendale
2  0004-TLHLJ    Male   50      No                     0    Costa Mesa
3  0011-IGKFF    Male   78     Yes                     0      Martinez
4  0013-EXCHZ  Female   75     Yes                     0     Camarillo

   Zip Code   Latitude   Longitude  Number of Referrals  ...   Payment Method  \
0     93225  34.827662 -118.999073                    2  ...      Credit Card
1     91206  34.162515 -118.203869                    0  ...      Credit Card
2     92627  33.645672 -117.922613                    0  ...  Bank Withdrawal
3     94553  38.014457 -122.115432                    1  ...  Bank Withdrawal
4     93010  34.227846 -119.079903                    3  ...      Credit Card

  Monthly Charge Total Charges  Total Refunds Total Extra Data Charges  \
0           65.6        593.30           0.00                        0
1           -4.0        542.40          38.33                       10
2           73.9        280.85           0.00                        0
3           98.0       1237.85           0.00                        0
4           83.9        267.40           0.00                        0

  Total Long Distance Charges Total Revenue  Customer Status   Churn Category  \
0                      381.51        974.81           Stayed              NaN
1                       96.21        610.28           Stayed              NaN
2                      134.60        415.45          Churned       Competitor
3                      361.66       1599.51          Churned  Dissatisfaction
4                       22.14        289.54          Churned  Dissatisfaction

                   Churn Reason
0                           NaN
1                           NaN
2  Competitor had better devices
3        Product dissatisfaction
4            Network reliability

[5 rows x 38 columns]
```

In [4]:
```
df.tail()
```

Out[4]:

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude	Number of Referrals	...	Payment Method	Monthly Charge
7038	9987-LUTYD	Female	20	No	0	La Mesa	91941	32.759327	-116.997260	0	...	Credit Card	55.15
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	95367	37.734971	-120.954271	1	...	Bank Withdrawal	85.10
7040	9992-UJOEL	Male	22	No	0	Elk	95432	39.108252	-123.645121	0	...	Credit Card	50.30
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	92075	33.001813	-117.263628	5	...	Credit Card	67.85
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	96125	39.600599	-120.636358	1	...	Bank Withdrawal	59.00

5 rows × 38 columns

DATA PRE-PROCESSING

In [5]:
```
# sanity check of data
df.shape
```

Out[5]: (7043, 38)

In [6]:
```
df.dtypes
```

Out[6]:
```
Customer ID                         object
Gender                              object
Age                                  int64
Married                             object
Number of Dependents                 int64
City                                object
Zip Code                             int64
Latitude                          float64
Longitude                         float64
Number of Referrals                  int64
Tenure in Months                     int64
Offer                               object
Phone Service                       object
Avg Monthly Long Distance Charges  float64
Multiple Lines                      object
Internet Service                    object
Internet Type                       object
Avg Monthly GB Download           float64
Online Security                     object
Online Backup                       object
Device Protection Plan              object
Premium Tech Support                object
Streaming TV                        object
Streaming Movies                    object
Streaming Music                     object
Unlimited Data                      object
Contract                            object
Paperless Billing                   object
Payment Method                      object
Monthly Charge                    float64
Total Charges                     float64
Total Refunds                     float64
Total Extra Data Charges             int64
Total Long Distance Charges       float64
Total Revenue                     float64
Customer Status                     object
Churn Category                      object
Churn Reason                        object
dtype: object
```

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Customer ID', 'Gender', 'Age', 'Married', 'Number of Dependents',
       'City', 'Zip Code', 'Latitude', 'Longitude', 'Number of Referrals',
       'Tenure in Months', 'Offer', 'Phone Service',
       'Avg Monthly Long Distance Charges', 'Multiple Lines',
       'Internet Service', 'Internet Type', 'Avg Monthly GB Download',
       'Online Security', 'Online Backup', 'Device Protection Plan',
       'Premium Tech Support', 'Streaming TV', 'Streaming Movies',
       'Streaming Music', 'Unlimited Data', 'Contract', 'Paperless Billing',
       'Payment Method', 'Monthly Charge', 'Total Charges', 'Total Refunds',
       'Total Extra Data Charges', 'Total Long Distance Charges',
       'Total Revenue', 'Customer Status', 'Churn Category', 'Churn Reason'],
      dtype='object')
```

In [8]:
```
df.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                              3166 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
```

In [9]:
```
#Numerical values
df.describe()
```

Out[9]:

	Age	Number of Dependents	Zip Code	Latitude	Longitude	Number of Referrals	Tenure in Months	Avg Monthly Long Distance Charges	Avg Monthly GB Download	
count	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	7043.000000	6361.000000	5517.000000	70
mean	46.509726	0.468692	93486.070567	36.197455	-119.756684	1.951867	32.386767	25.420517	26.189958	
std	16.750352	0.962802	1856.767505	2.468929	2.154425	3.001199	24.542061	14.200374	19.586585	
min	19.000000	0.000000	90001.000000	32.555828	-124.301372	0.000000	1.000000	1.010000	2.000000	-
25%	32.000000	0.000000	92101.000000	33.990646	-121.788090	0.000000	9.000000	13.050000	13.000000	
50%	46.000000	0.000000	93518.000000	36.205465	-119.595293	0.000000	29.000000	25.690000	21.000000	
75%	60.000000	0.000000	95329.000000	38.161321	-117.969795	3.000000	55.000000	37.680000	30.000000	
max	80.000000	9.000000	96150.000000	41.962127	-114.192901	11.000000	72.000000	49.990000	85.000000	1

In [10]:
```
#Check for missing values
print("\nChecking for missing values:")
print(df.isnull().sum())
```

```
Checking for missing values:
Customer ID                            0
Gender                                 0
Age                                    0
Married                                0
Number of Dependents                   0
City                                   0
Zip Code                               0
Latitude                               0
Longitude                              0
Number of Referrals                    0
Tenure in Months                       0
Offer                               3877
Phone Service                          0
Avg Monthly Long Distance Charges    682
Multiple Lines                       682
Internet Service                       0
Internet Type                       1526
Avg Monthly GB Download             1526
Online Security                     1526
Online Backup                       1526
Device Protection Plan              1526
Premium Tech Support                1526
Streaming TV                        1526
Streaming Movies                    1526
Streaming Music                     1526
Unlimited Data                      1526
Contract                               0
Paperless Billing                      0
Payment Method                         0
Monthly Charge                         0
Total Charges                          0
Total Refunds                          0
Total Extra Data Charges               0
Total Long Distance Charges            0
Total Revenue                          0
Customer Status                        0
Churn Category                      5174
Churn Reason                        5174
dtype: int64
```

In [11]:
```
#finding duplicates
print("\nFinding duplicates values:")
print(df.duplicated().sum())
```

```
Finding duplicates values:
0
```

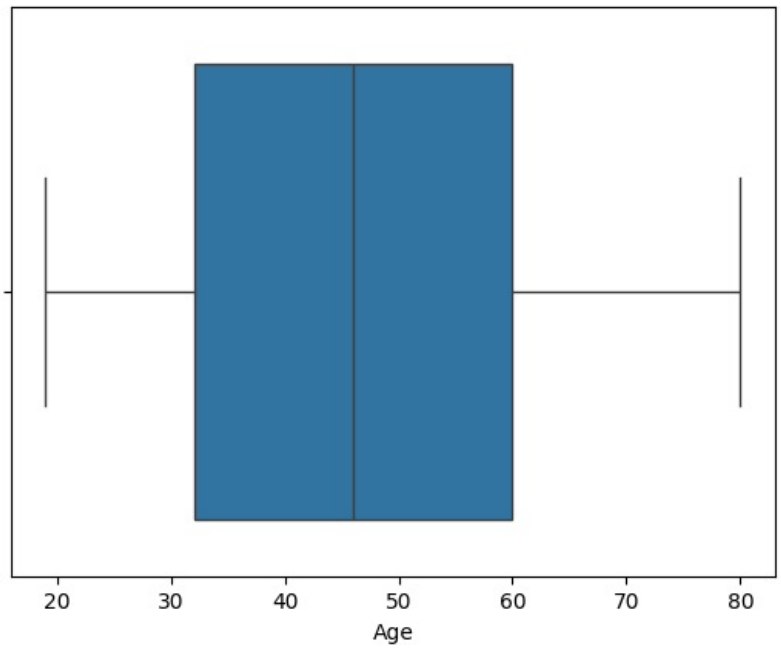
In [12]:
```
#identifying garbage values
for i in df.select_dtypes(include="object").columns:
    print(df[i].value_counts())
    print("***"*10)
```

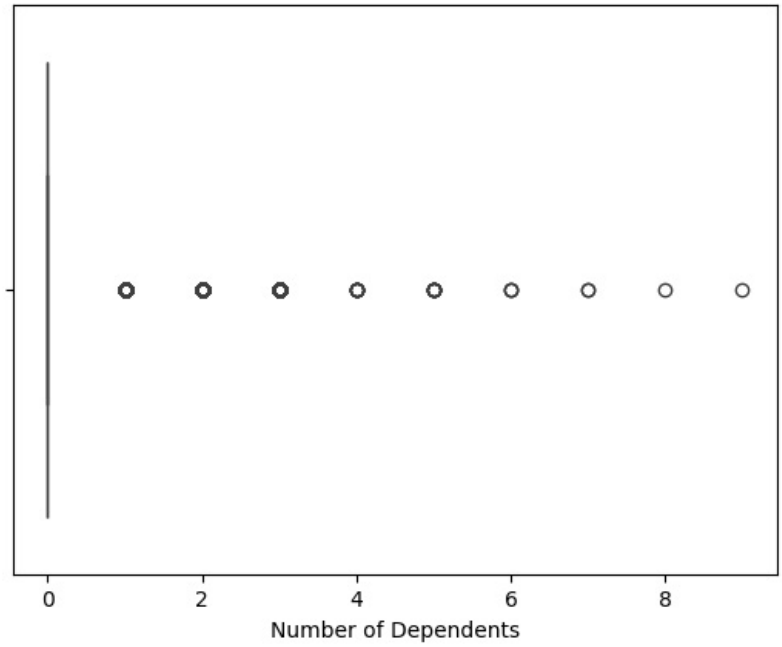
```
Customer ID
0002-ORFBO    1
6616-AALSR    1
6625-UTXEW    1
6625-IUTTT    1
6625-FLENO    1
             ..
3352-RICWQ    1
3352-ALMCK    1
3351-NQLDI    1
3351-NGXYI    1
9995-HOTOH    1
Name: count, Length: 7043, dtype: int64
******************************
Gender
Male      3555
Female    3488
Name: count, dtype: int64
******************************
Married
No     3641
Yes    3402
Name: count, dtype: int64
******************************
City
Los Angeles         293
San Diego           285
San Jose            112
Sacramento          108
San Francisco       104
                   ...
```

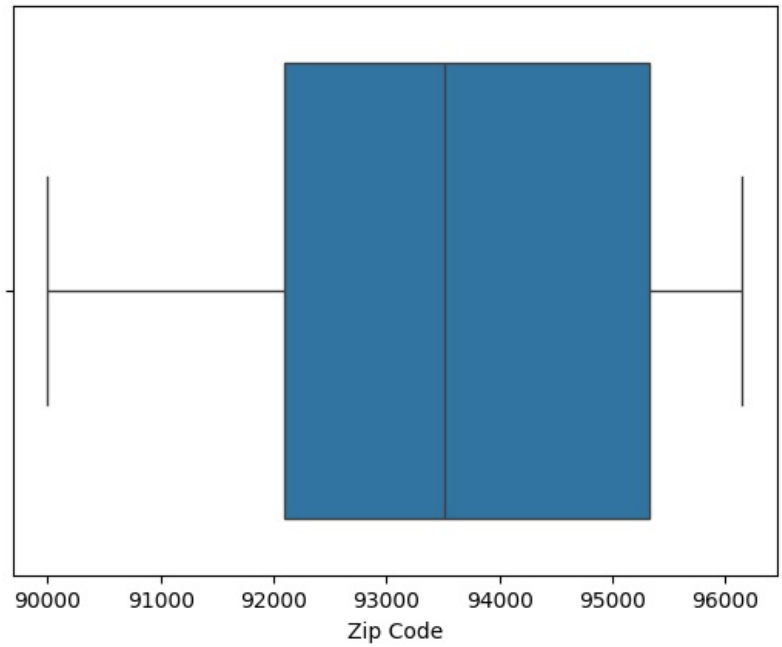
```
Johannesburg          2
South Lake Tahoe      2
Jacumba               2
Holtville             2
Eldridge              2
Name: count, Length: 1106, dtype: int64
******************************
Offer
Offer B    824
Offer E    805
Offer D    602
Offer A    520
Offer C    415
Name: count, dtype: int64
******************************
Phone Service
Yes    6361
No      682
Name: count, dtype: int64
******************************
Multiple Lines
No     3390
Yes    2971
Name: count, dtype: int64
******************************
Internet Service
Yes    5517
No     1526
Name: count, dtype: int64
******************************
Internet Type
Fiber Optic    3035
DSL            1652
Cable           830
Name: count, dtype: int64
******************************
Online Security
No     3498
Yes    2019
Name: count, dtype: int64
******************************
Online Backup
No     3088
Yes    2429
Name: count, dtype: int64
******************************
Device Protection Plan
No     3095
Yes    2422
Name: count, dtype: int64
******************************
Premium Tech Support
No     3473
Yes    2044
Name: count, dtype: int64
******************************
Streaming TV
No     2810
Yes    2707
Name: count, dtype: int64
******************************
Streaming Movies
No     2785
Yes    2732
Name: count, dtype: int64
******************************
Streaming Music
No     3029
Yes    2488
Name: count, dtype: int64
******************************
Unlimited Data
Yes    4745
No      772
Name: count, dtype: int64
******************************
Contract
Month-to-Month    3610
Two Year          1883
One Year          1550
Name: count, dtype: int64
******************************
Paperless Billing
```

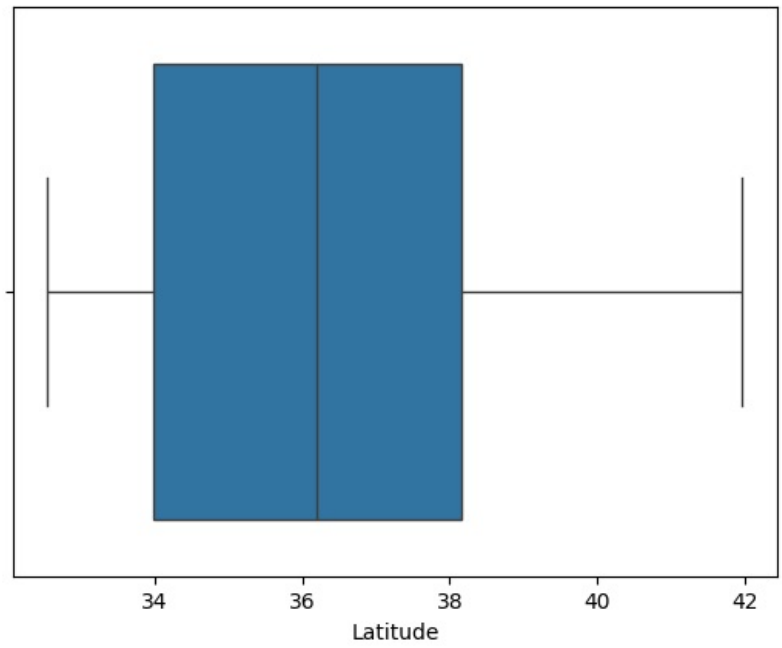
```
Yes    4171
No     2872
Name: count, dtype: int64
******************************
Payment Method
Bank Withdrawal    3909
Credit Card        2749
Mailed Check        385
Name: count, dtype: int64
******************************
Customer Status
Stayed     4720
Churned    1869
Joined      454
Name: count, dtype: int64
******************************
Churn Category
Competitor         841
Dissatisfaction    321
Attitude           314
Price              211
Other              182
Name: count, dtype: int64
******************************
Churn Reason
Competitor had better devices                313
Competitor made better offer                 311
Attitude of support person                   220
Don't know                                   130
Competitor offered more data                 117
Competitor offered higher download speeds    100
Attitude of service provider                  94
Price too high                                78
Product dissatisfaction                       77
Network reliability                           72
Long distance charges                         64
Service dissatisfaction                       63
Moved                                         46
Extra data charges                            39
Limited range of services                     37
Poor expertise of online support              31
Lack of affordable download/upload speed      30
Lack of self-service on Website               29
Poor expertise of phone support               12
Deceased                                       6
Name: count, dtype: int64
******************************
```

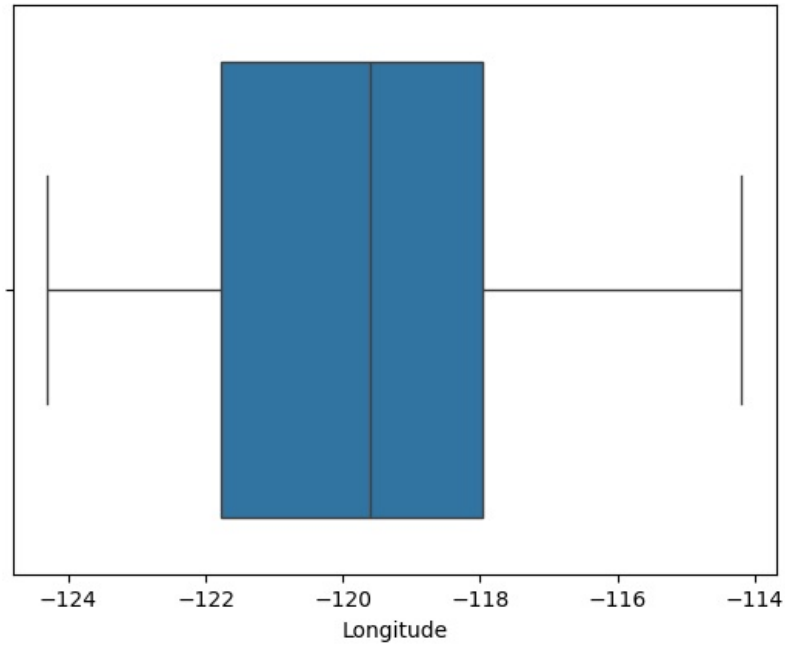
In [13]:
```
# Boxplot for outlier detection
import warnings
warnings.filterwarnings("ignore")
for i in df.select_dtypes(include="number").columns:
    sns.boxplot(data=df,x=i)
    plt.show()
```

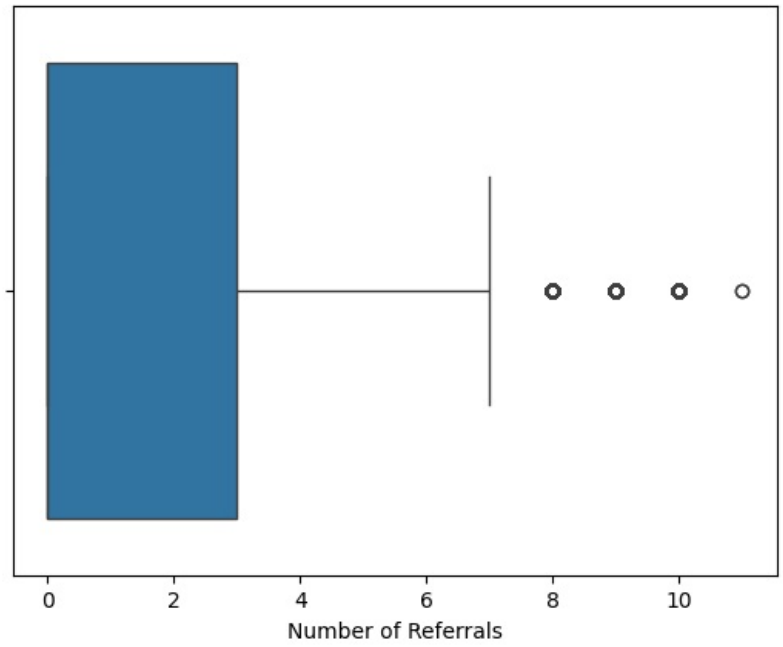


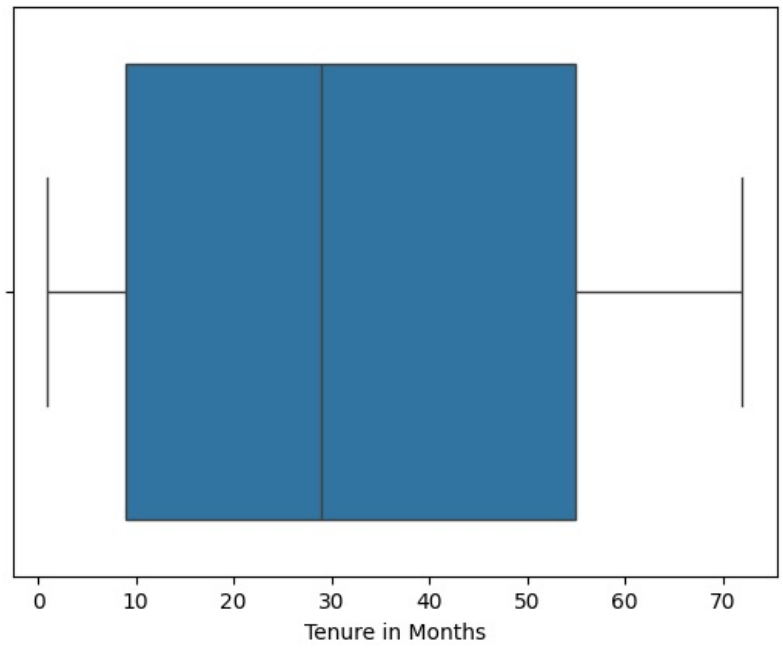

0
2
4
6
8
Number of Dependents

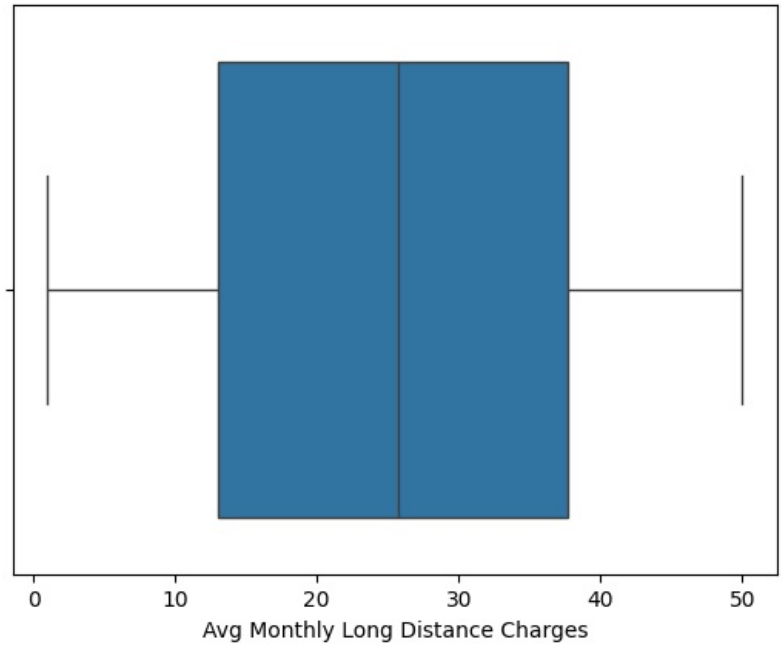

90000
91000
92000
93000
94000
95000
96000
Zip Code

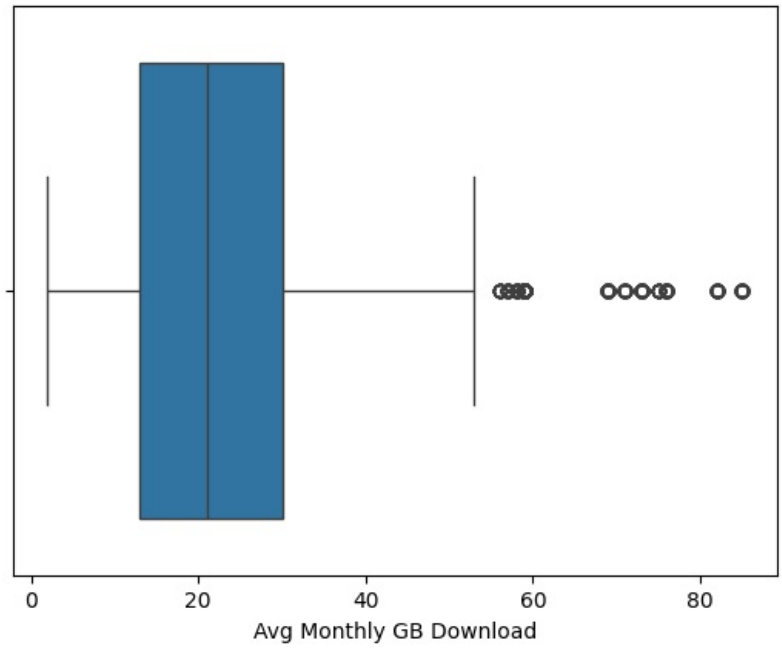

34
36
38
40
42
Latitude

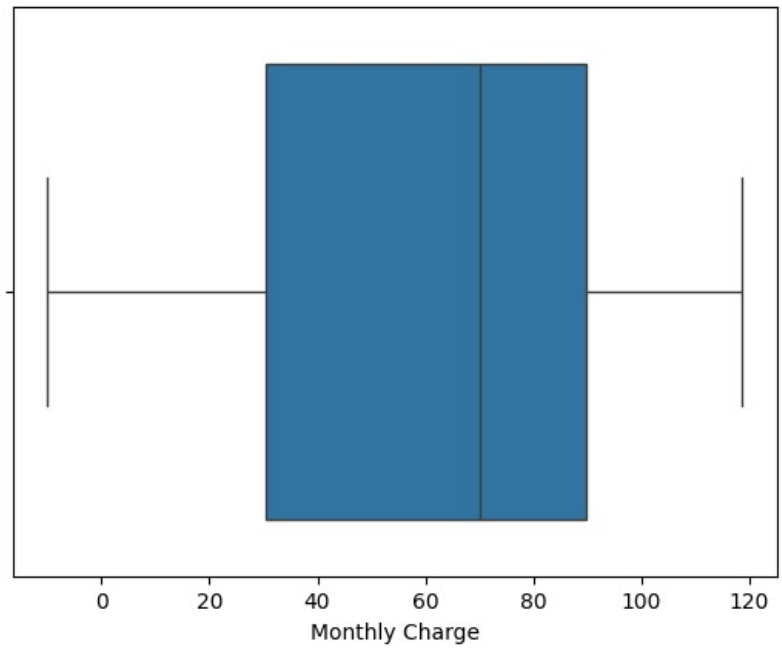

−124
−122
−120
−118
−116
−114
Longitude

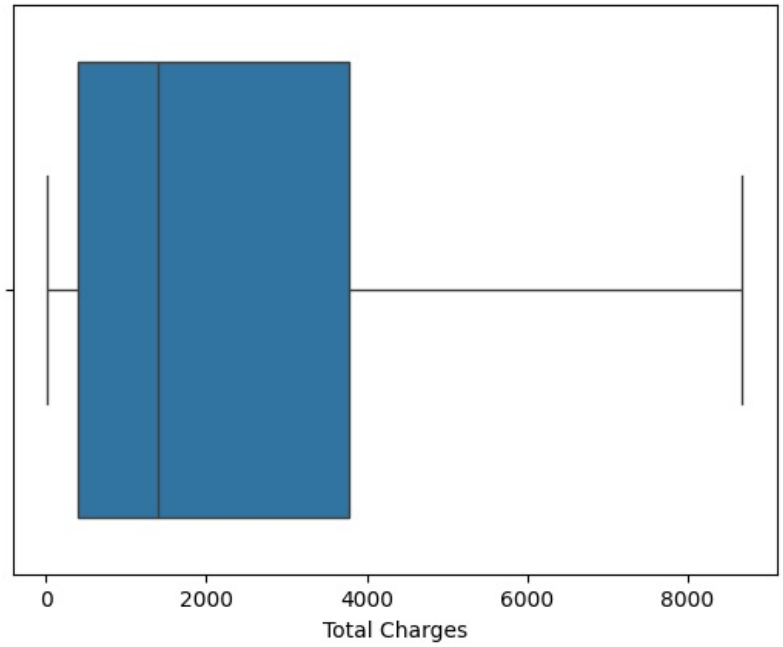

0
2
4
6
8
10
Number of Referrals

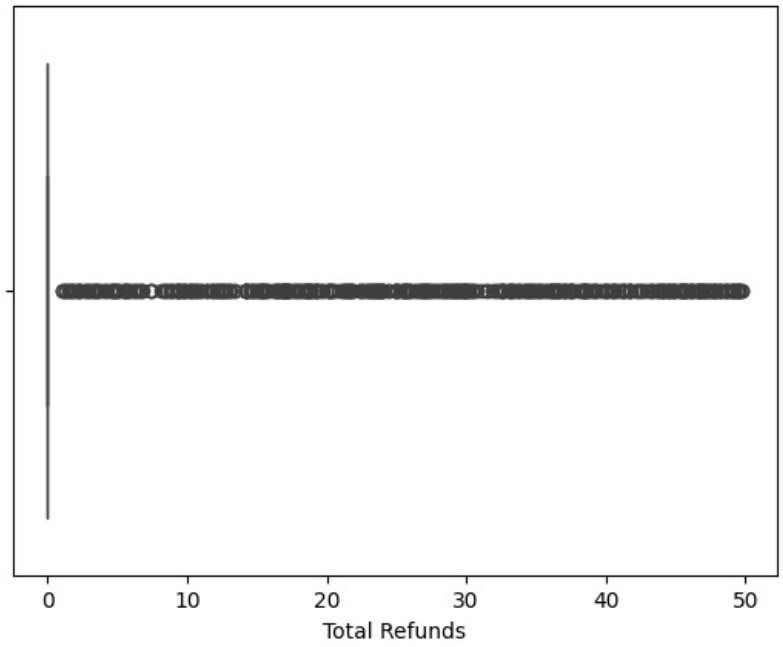

0
10
20
30
40
50
60
70
Tenure in Months

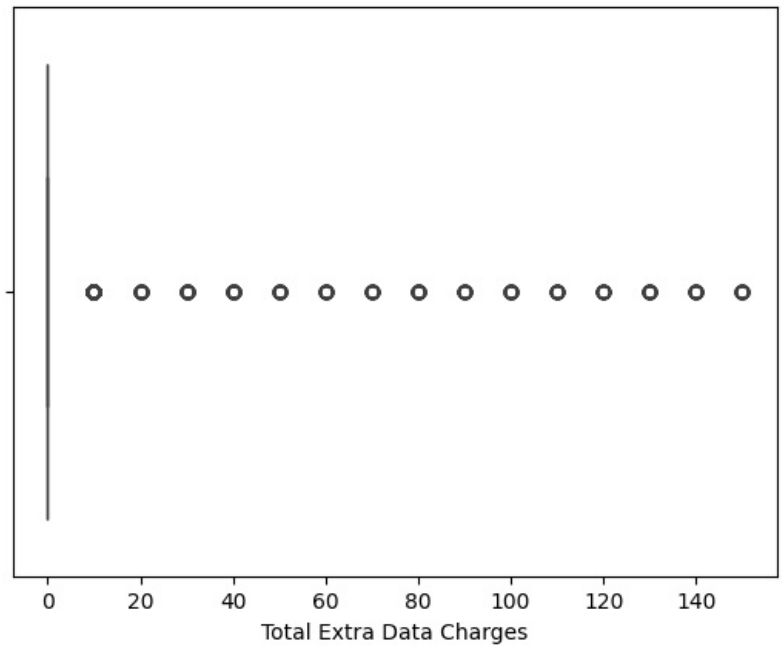

0
10
20
30
40
50
Avg Monthly Long Distance Charges

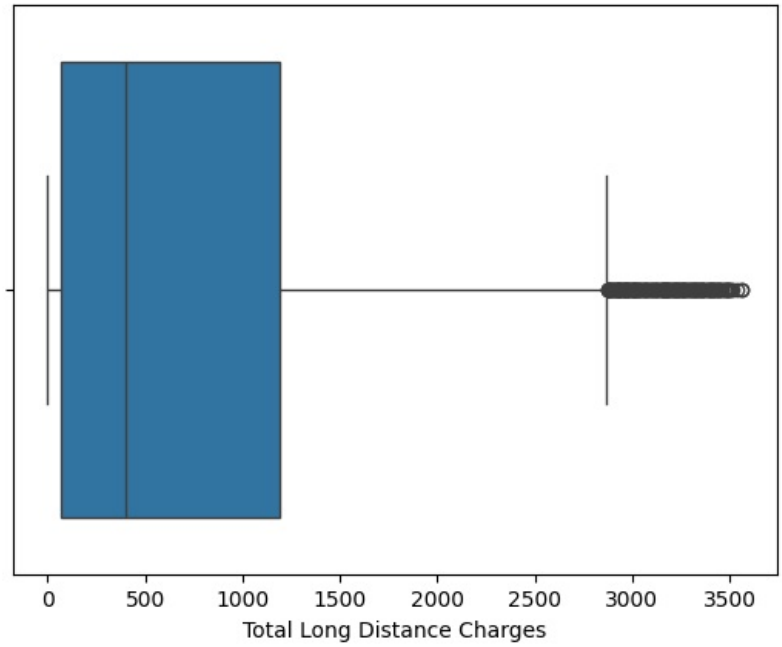

0
20
40
60
80
Avg Monthly GB Download

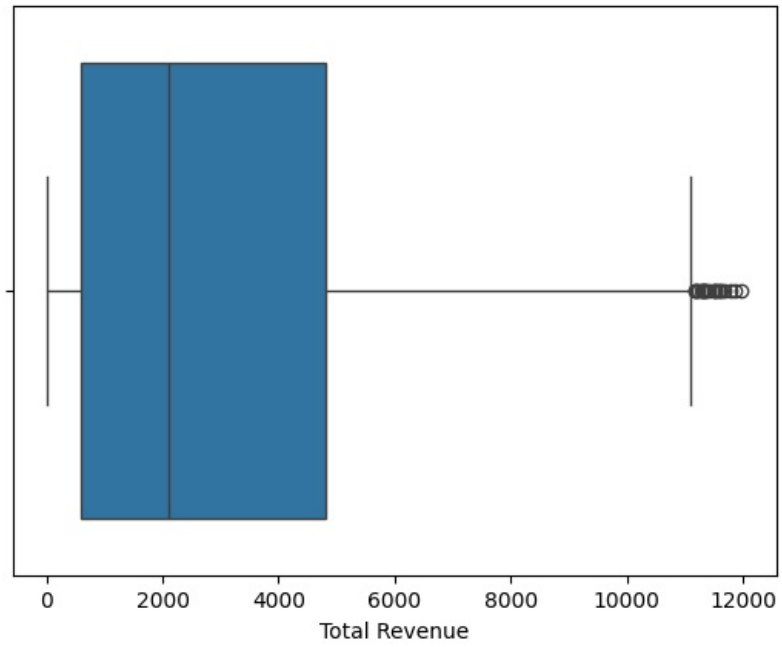

0
20
40
60
80
100
120
Monthly Charge


0
2000
4000
6000
8000
Total Charges

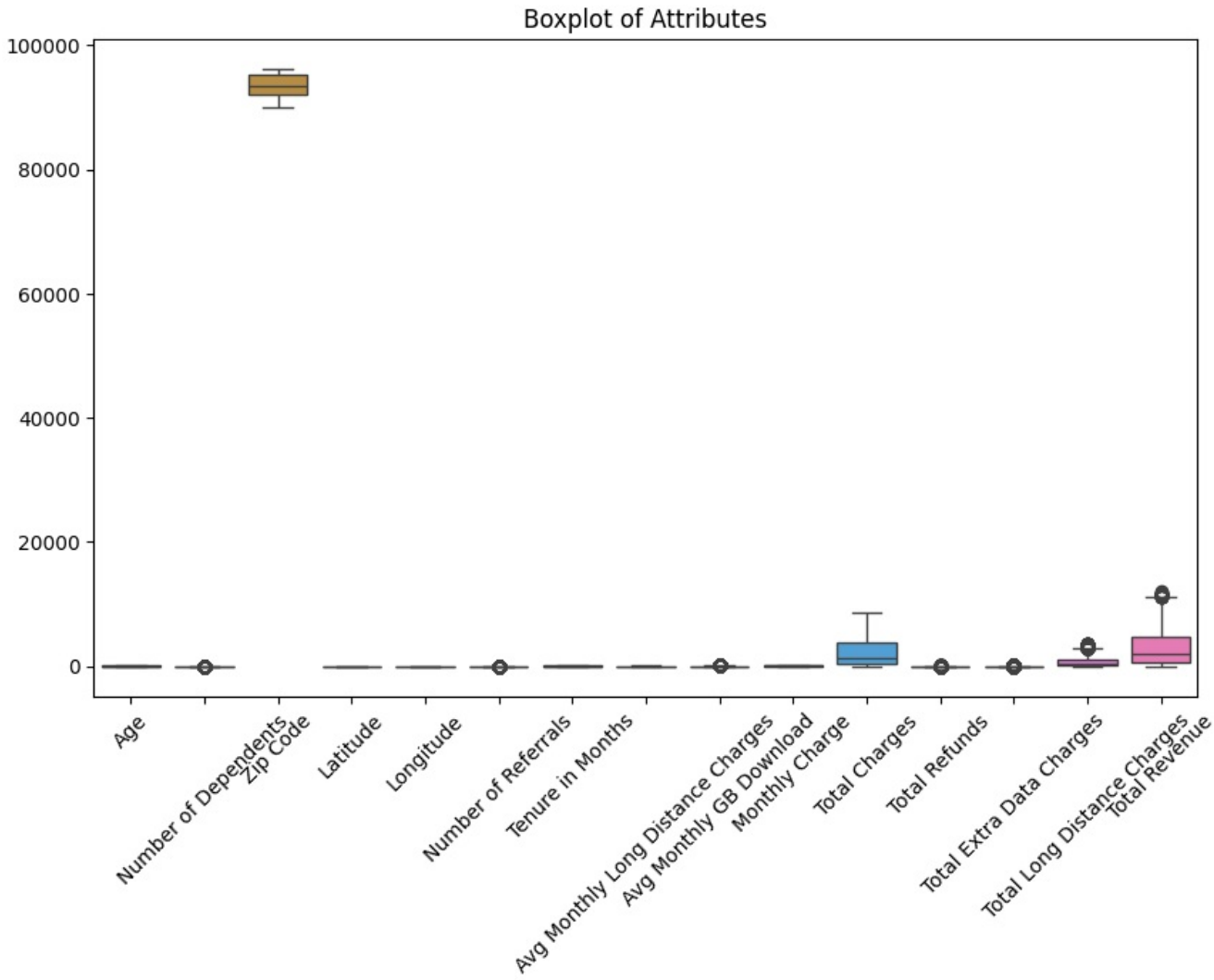

0
10
20
30
40
50
Total Refunds


0
20
40
60
80
100
120
140
Total Extra Data Charges





In [14]:
```
# Boxplot for outlier detection
plt.figure(figsize=(10, 6))
sns.boxplot(data=df)
plt.title('Boxplot of Attributes')
plt.xticks(rotation=45)
plt.show()
```



Data cleaning

In [15]:
```
#Missing Value treatments
for i in ["Avg Monthly Long Distance Charges","Avg Monthly GB Download"]:
    df[i].fillna(df[i].median(),inplace=True)
```

In [16]:
```
for i in ["Offer","Multiple Lines","Internet Type","Online Security","Online Backup","Device Protection Plan","
          "Streaming Movies","Streaming Music","Unlimited Data","Churn Category","Churn Reason"]:
    df[i].fillna(df[i].mode().iloc[0],inplace=True)
```

In [17]:
```
df.isnull().sum()
```

Out[17]:
```
Customer ID                          0
Gender                               0
Age                                  0
Married                              0
Number of Dependents                 0
City                                 0
Zip Code                             0
Latitude                             0
Longitude                            0
Number of Referrals                  0
Tenure in Months                     0
Offer                                0
Phone Service                        0
Avg Monthly Long Distance Charges    0
Multiple Lines                       0
Internet Service                     0
Internet Type                        0
Avg Monthly GB Download              0
Online Security                      0
Online Backup                        0
Device Protection Plan               0
Premium Tech Support                 0
Streaming TV                         0
Streaming Movies                     0
Streaming Music                      0
Unlimited Data                       0
Contract                             0
Paperless Billing                    0
Payment Method                       0
Monthly Charge                       0
Total Charges                        0
Total Refunds                        0
Total Extra Data Charges             0
Total Long Distance Charges          0
Total Revenue                        0
Customer Status                      0
Churn Category                       0
Churn Reason                         0
dtype: int64
```

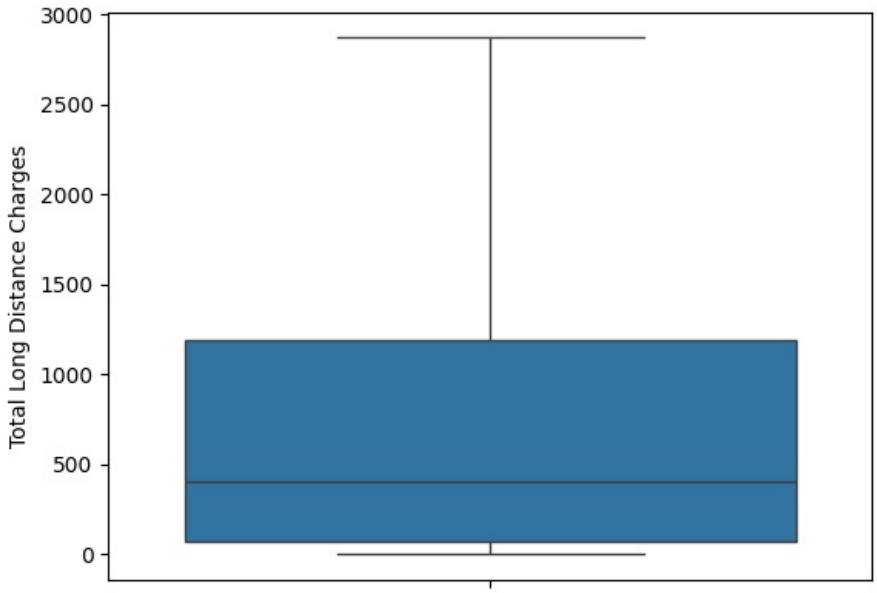
In [18]:
```
#Outliers treatments
def wisker(col):
    q1,q3=np.percentile(col,[25,75])
    iqr=q3-q1
    lw=q1-1.5*iqr
    uw=q3+1.5*iqr
    return lw,uw
```

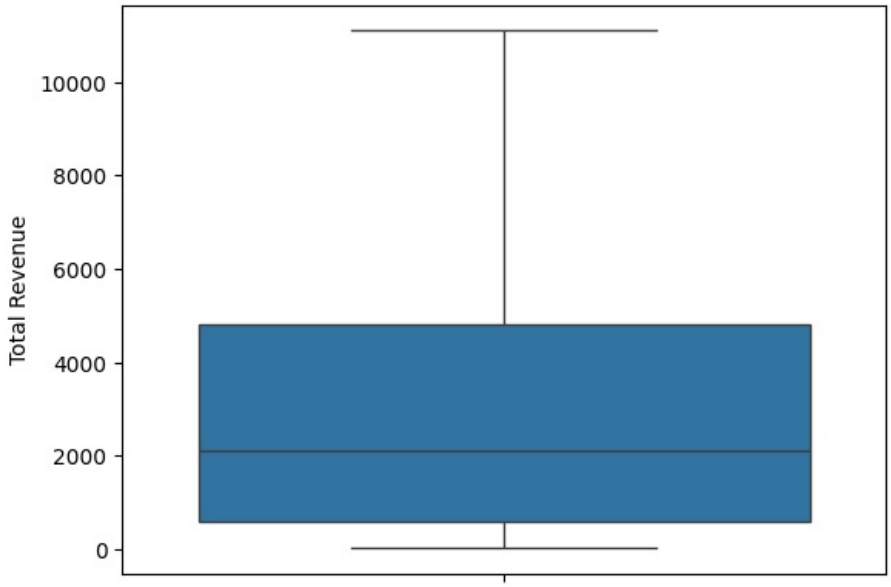
In [19]:
```
wisker(df['Number of Referrals'])
```

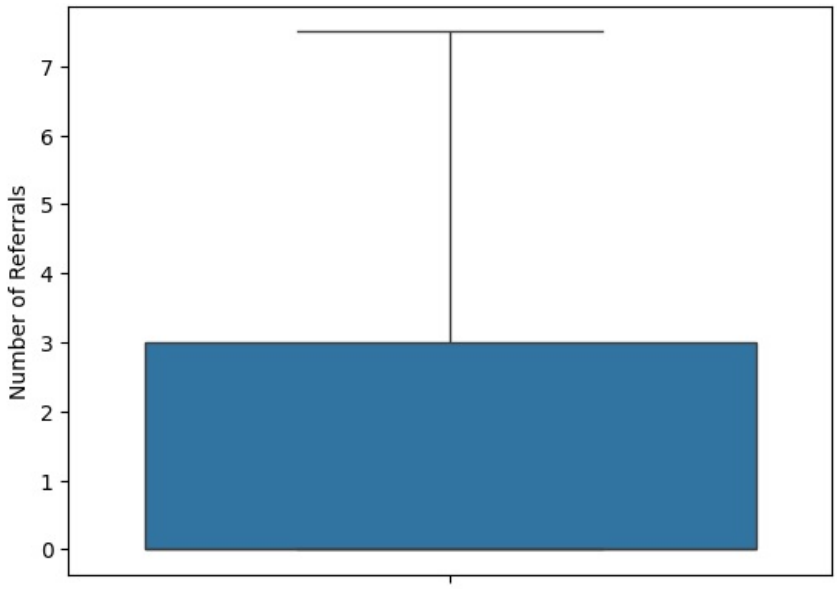
Out[19]:
```
(-4.5, 7.5)
```

In [20]:
```
for i in ['Total Long Distance Charges','Total Revenue','Number of Referrals']:
    lw,uw=wisker(df[i])
    df[i]=np.where(df[i]<lw,lw,df[i])
    df[i]=np.where(df[i]>uw,uw, df[i])
```

In [21]:
```
for i in ['Total Long Distance Charges','Total Revenue','Number of Referrals']:
    sns.boxplot(df[i])
    plt.show()
```







DATA ANALYSIS

In [ ]:
```
#Exploratory Data Analysis(EDA)
```

In [22]:
```
# (1)Show descriptive statistics of numerical column:
df.describe().T
```

Out[22]:

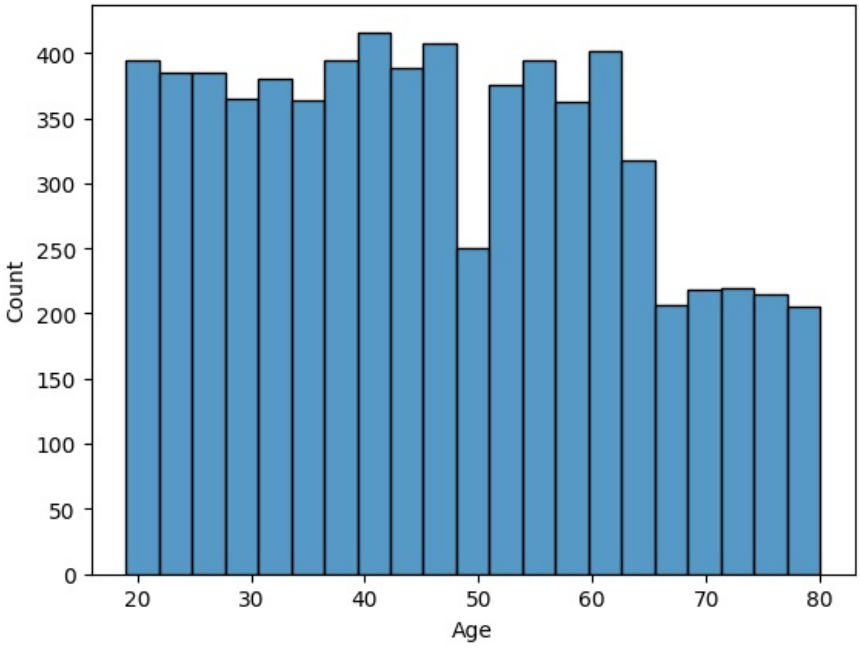
	count	mean	std	min	25%	50%	75%	max
Age	7043.0	46.509726	16.750352	19.000000	32.000000	46.000000	60.000000	80.000000
Number of Dependents	7043.0	0.468692	0.962802	0.000000	0.000000	0.000000	0.000000	9.000000
Zip Code	7043.0	93486.070567	1856.767505	90001.000000	92101.000000	93518.000000	95329.000000	96150.000000
Latitude	7043.0	36.197455	2.468929	32.555828	33.990646	36.205465	38.161321	41.962127
Longitude	7043.0	-119.756684	2.154425	-124.301372	-121.788090	-119.595293	-117.969795	-114.192901
Number of Referrals	7043.0	1.805907	2.661022	0.000000	0.000000	0.000000	3.000000	7.500000
Tenure in Months	7043.0	32.386767	24.542061	1.000000	9.000000	29.000000	55.000000	72.000000
Avg Monthly Long Distance Charges	7043.0	25.446612	13.495466	1.010000	14.455000	25.690000	36.395000	49.990000
Avg Monthly GB Download	7043.0	25.065455	17.466342	2.000000	15.000000	21.000000	27.000000	85.000000
Monthly Charge	7043.0	63.596131	31.204743	-10.000000	30.400000	70.050000	89.750000	118.750000
Total Charges	7043.0	2280.381264	2266.220462	18.800000	400.150000	1394.550000	3786.600000	8684.800000
Total Refunds	7043.0	1.962182	7.902614	0.000000	0.000000	0.000000	0.000000	49.790000
Total Extra Data Charges	7043.0	6.860713	25.104978	0.000000	0.000000	0.000000	0.000000	150.000000
Total Long Distance Charges	7043.0	740.864881	823.637706	0.000000	70.545000	401.440000	1191.100000	2871.932500
Total Revenue	7043.0	3033.269913	2861.983162	21.360000	605.610000	2108.640000	4801.145000	11094.447500

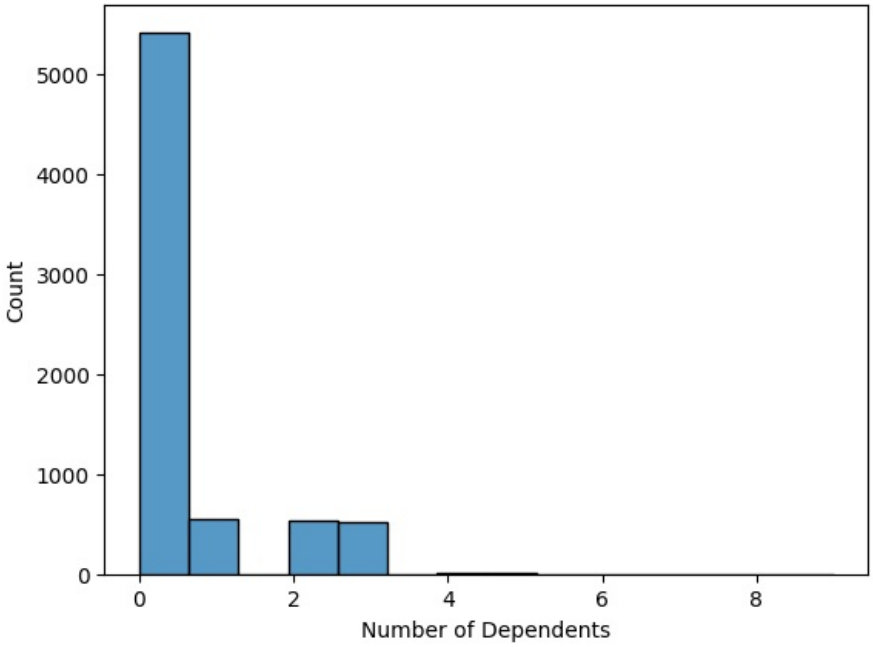
In [23]:
```
#(2)what all comes under descriptive statistics of object column?
df.describe(include="object").T
```

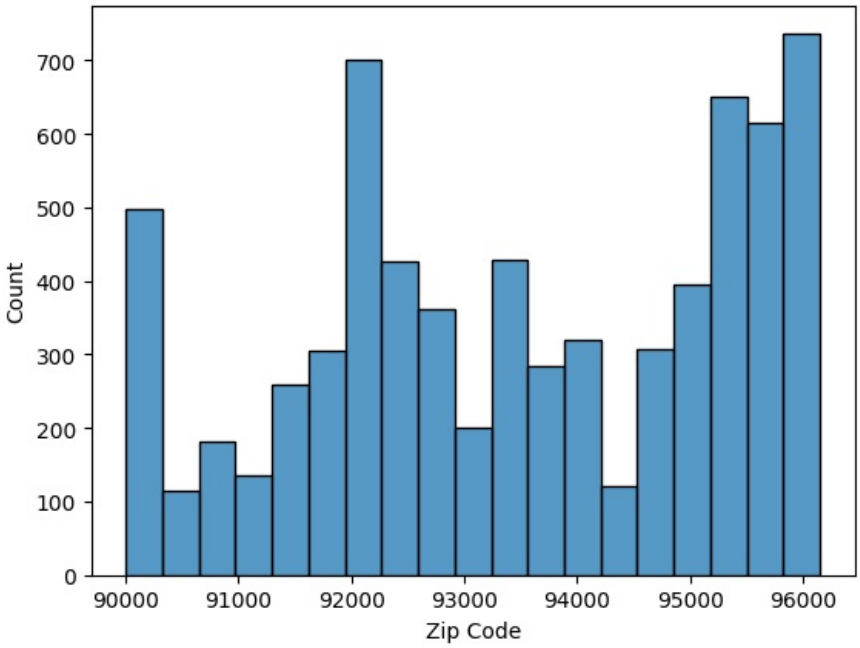
Out[23]:

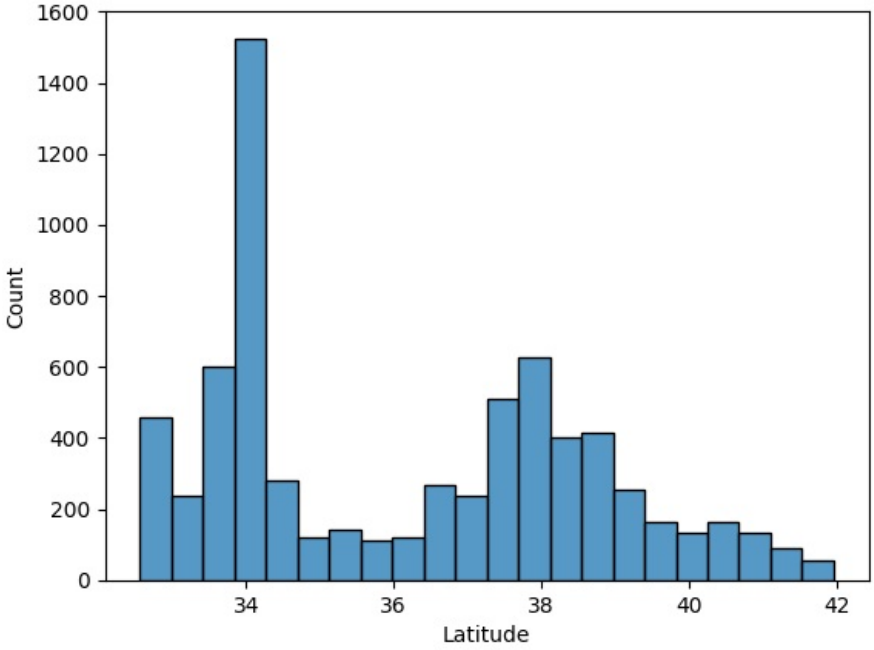
	count	unique	top	freq
Customer ID	7043	7043	0002-ORFBO	1
Gender	7043	2	Male	3555
Married	7043	2	No	3641
City	7043	1106	Los Angeles	293
Offer	7043	5	Offer B	4701
Phone Service	7043	2	Yes	6361
Multiple Lines	7043	2	No	4072
Internet Service	7043	2	Yes	5517
Internet Type	7043	3	Fiber Optic	4561
Online Security	7043	2	No	5024
Online Backup	7043	2	No	4614
Device Protection Plan	7043	2	No	4621
Premium Tech Support	7043	2	No	4999
Streaming TV	7043	2	No	4336
Streaming Movies	7043	2	No	4311
Streaming Music	7043	2	No	4555
Unlimited Data	7043	2	Yes	6271
Contract	7043	3	Month-to-Month	3610
Paperless Billing	7043	2	Yes	4171
Payment Method	7043	3	Bank Withdrawal	3909
Customer Status	7043	3	Stayed	4720
Churn Category	7043	5	Competitor	6015
Churn Reason	7043	20	Competitor had better devices	5487

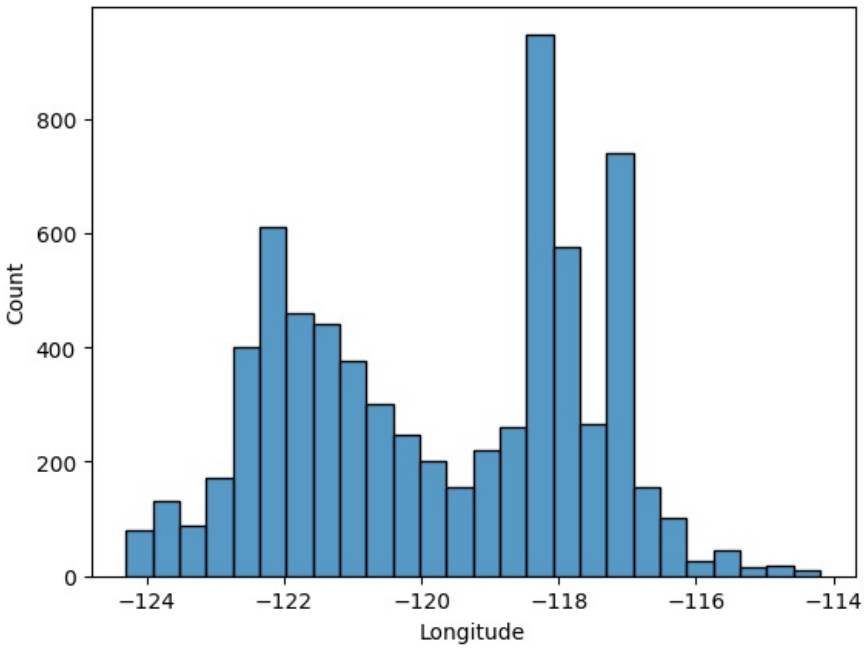
In [24]:
```
#(3)Plot histogram to show how Avg Monthly GB Download varies with all other numerical columns seperately?
import warnings
warnings.filterwarnings("ignore")
for i in df.select_dtypes(include="number").columns:
    sns.histplot(data=df,x=i)
    plt.show()
```

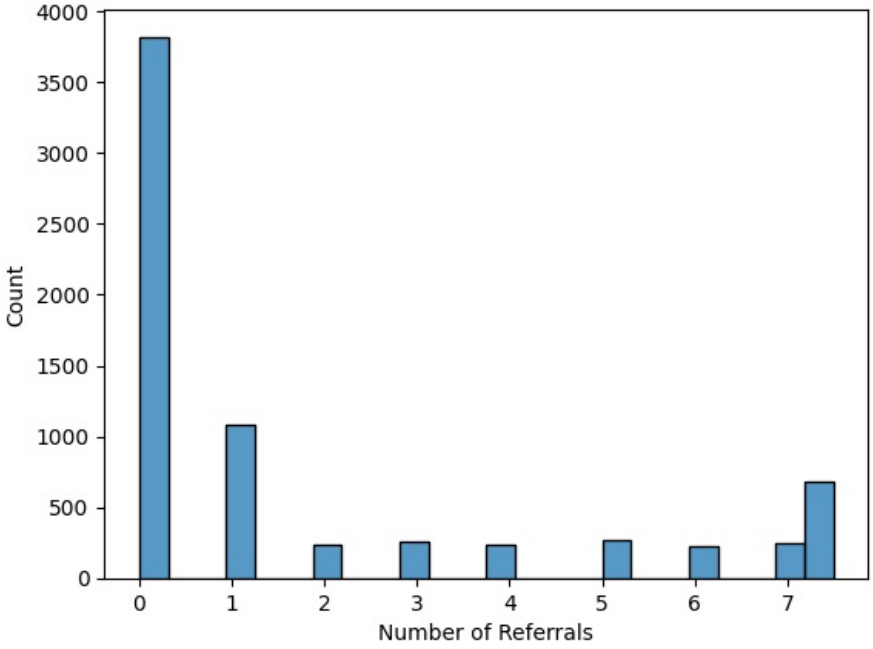

Count
Age
0
50
100
150
200
250
300
350
400
20
30
40
50
60
70
80

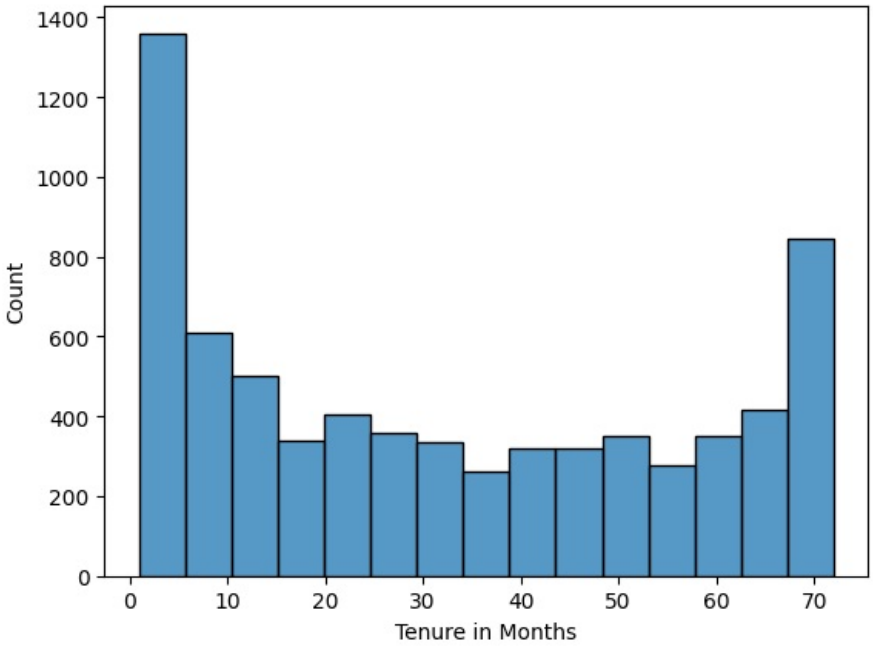

Count
Number of Dependents
0
1000
2000
3000
4000
5000
0
2
4
6
8

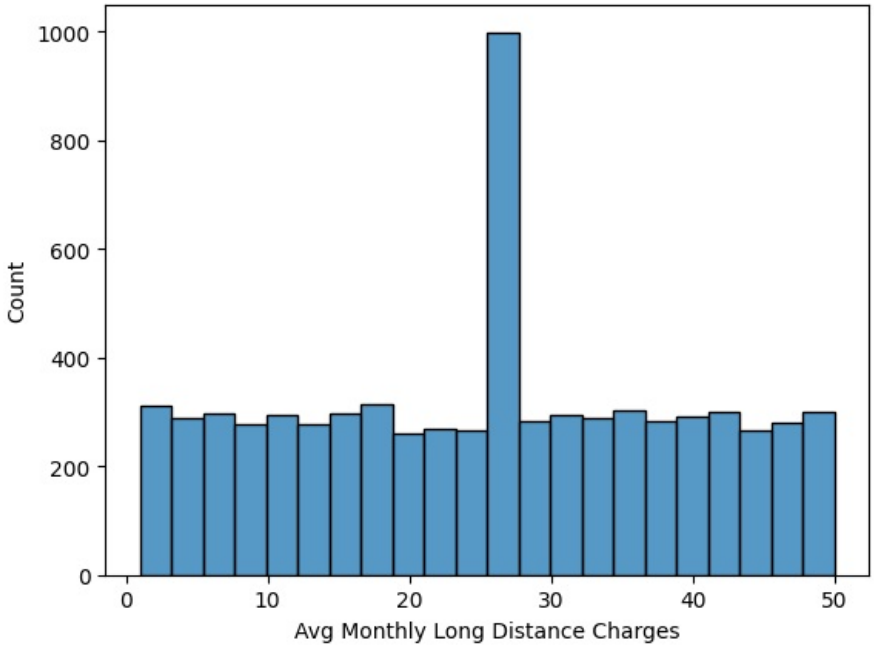

Count
Zip Code
0
100
200
300
400
500
600
700
90000
91000
92000
93000
94000
95000
96000

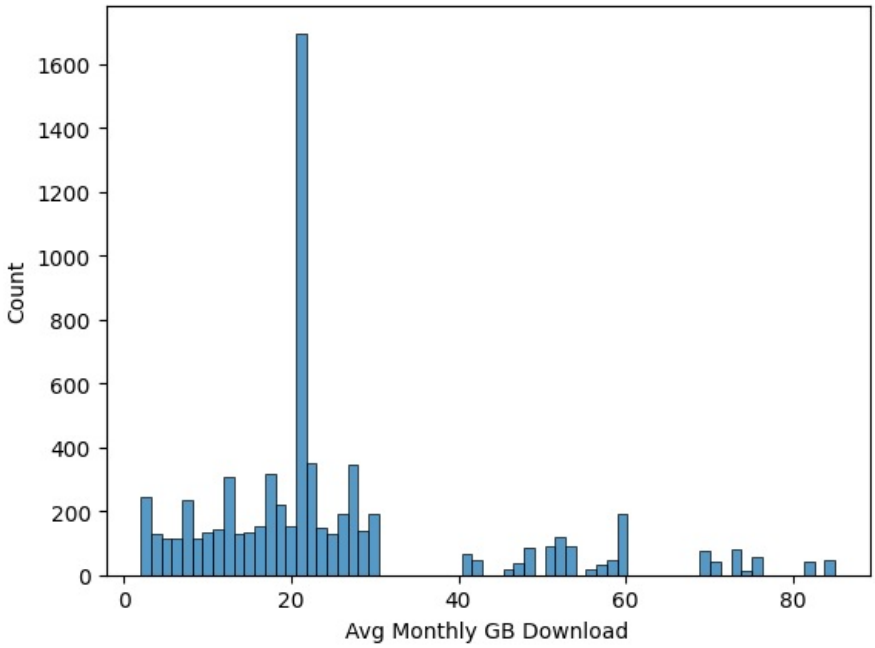

1600
1400
1200
1000
800
600
400
200
0
Count
34
36
38
40
42
Latitude

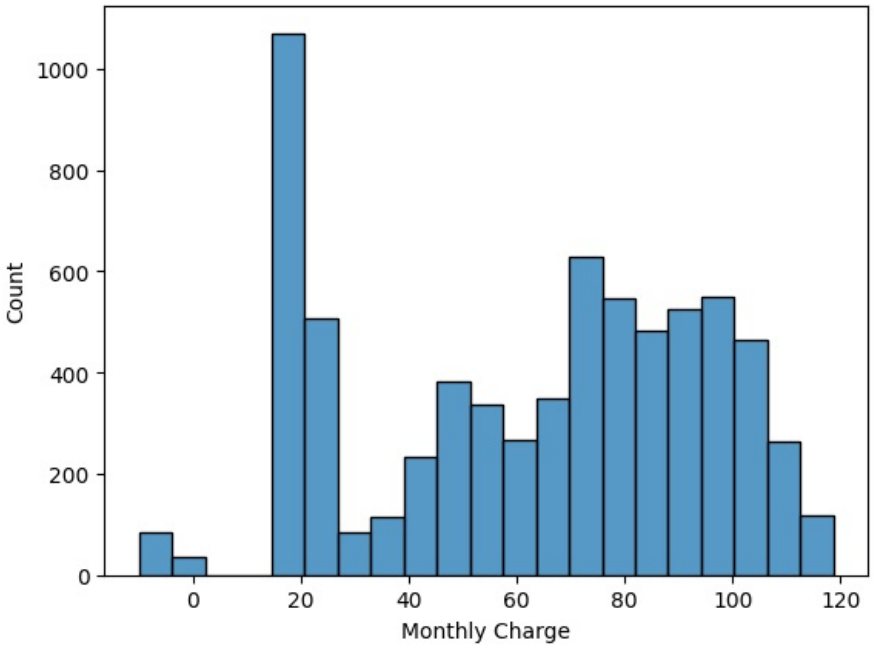

800
600
400
200
0
Count
−124
−122
−120
−118
−116
−114
Longitude

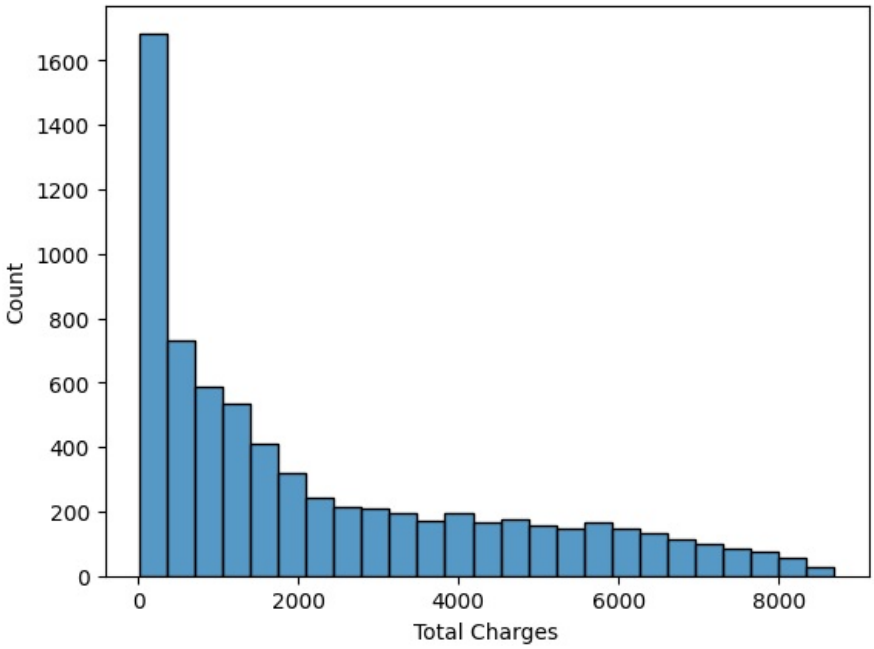

4000
3500
3000
2500
2000
1500
1000
500
0
Count
0
1
2
3
4
5
6
7
Number of Referrals

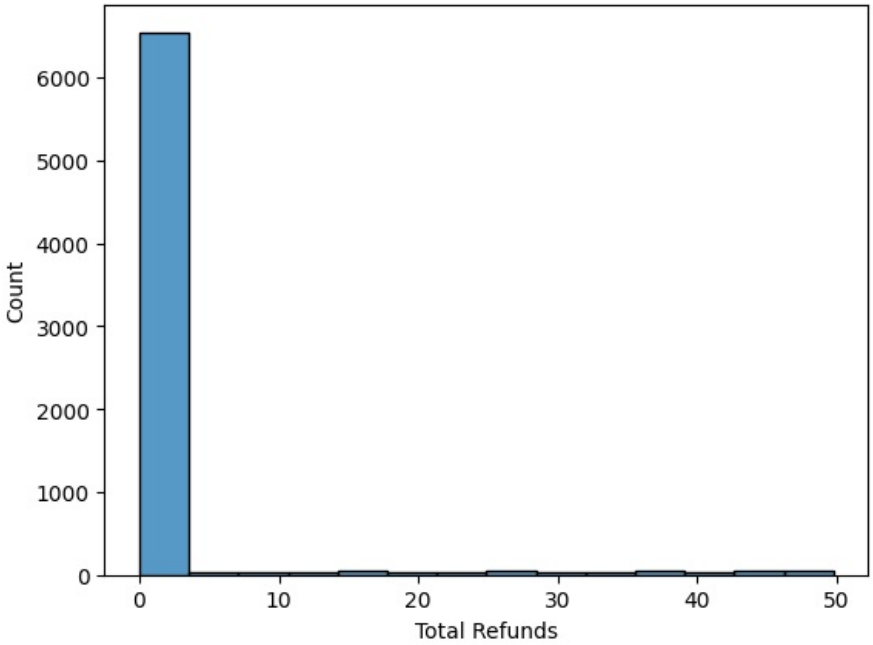

1400
1200
1000
800
600
400
200
0
Count
0
10
20
30
40
50
60
70
Tenure in Months

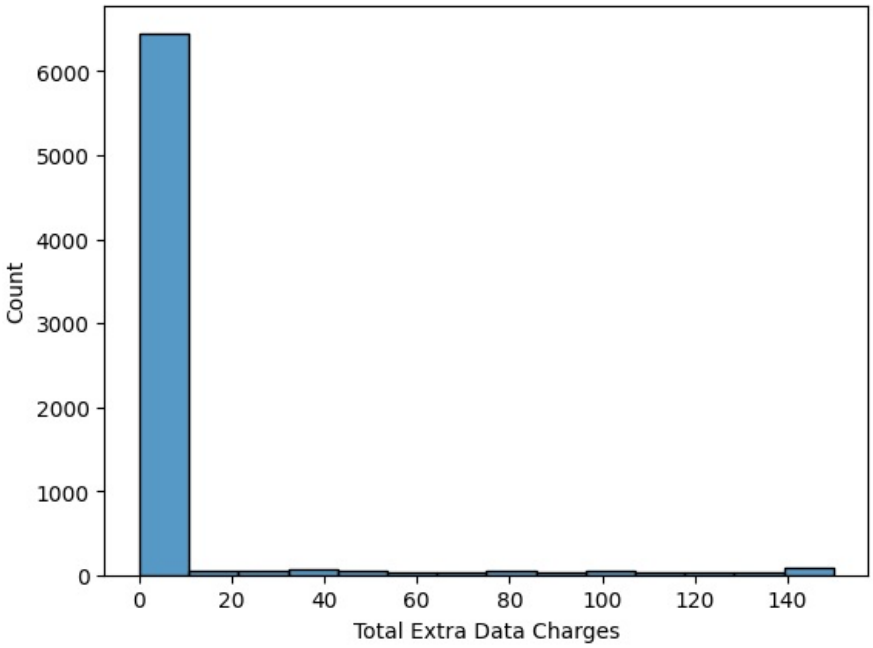

Count
1000
800
600
400
200
0
0
10
20
30
40
50
Avg Monthly Long Distance Charges

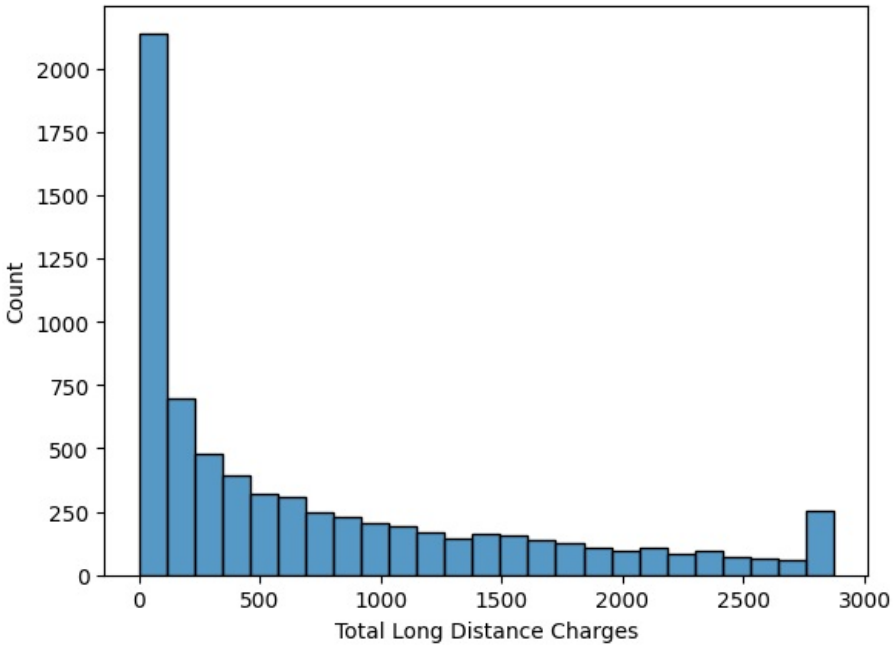

Count
1600
1400
1200
1000
800
600
400
200
0
0
20
40
60
80
Avg Monthly GB Download

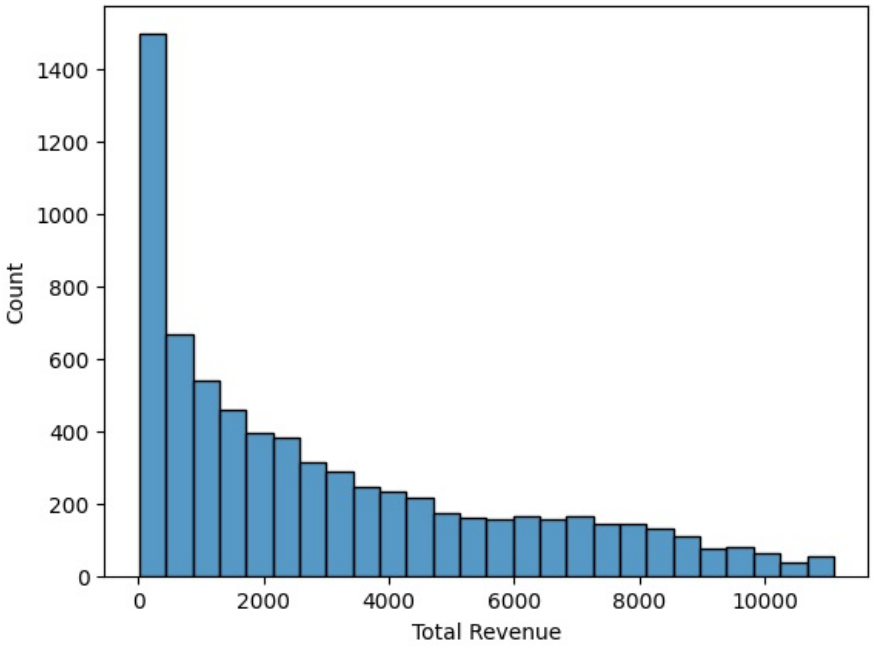

Count
1000
800
600
400
200
0
0
20
40
60
80
100
120
Monthly Charge


Count
1600
1400
1200
1000
800
600
400
200
0
0
2000
4000
6000
8000
Total Charges


6000
5000
4000
3000
2000
1000
0
Count
0
10
20
30
40
50
Total Refunds

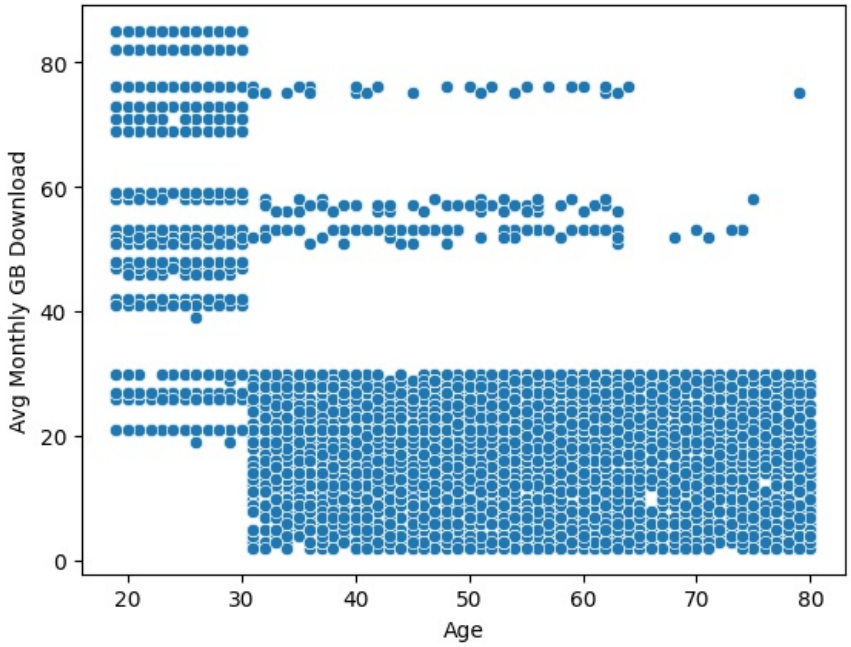

6000
5000
4000
3000
2000
1000
0
Count
0
20
40
60
80
100
120
140
Total Extra Data Charges

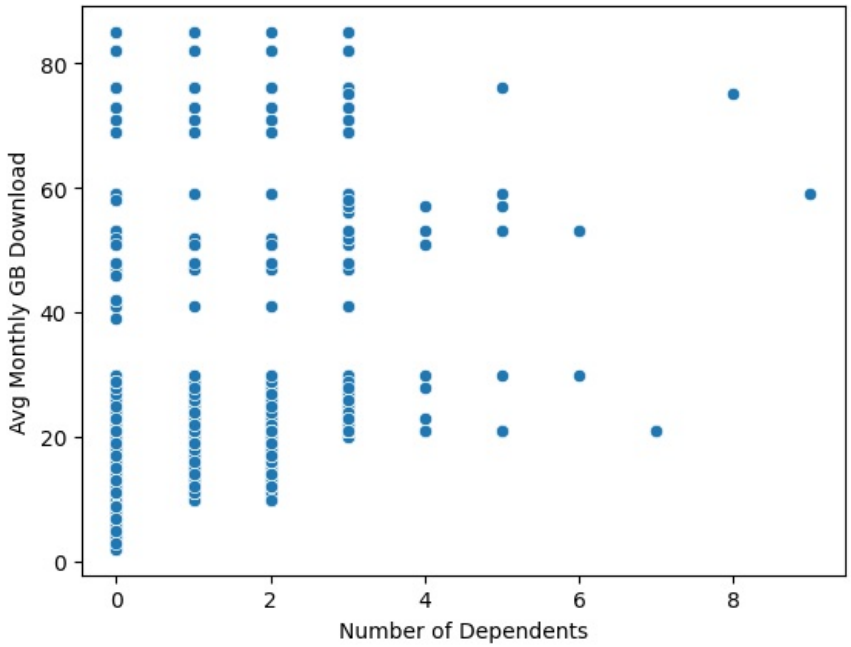


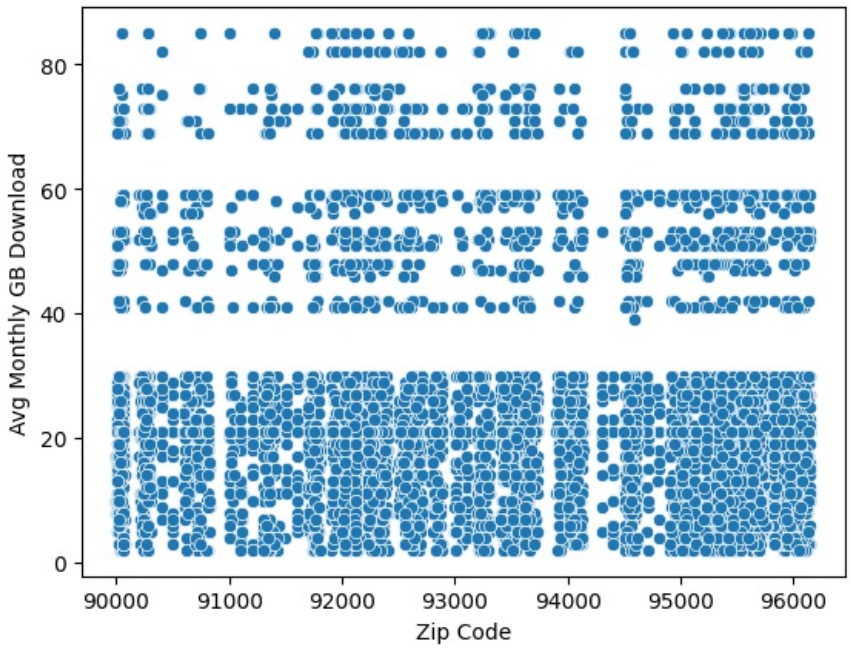


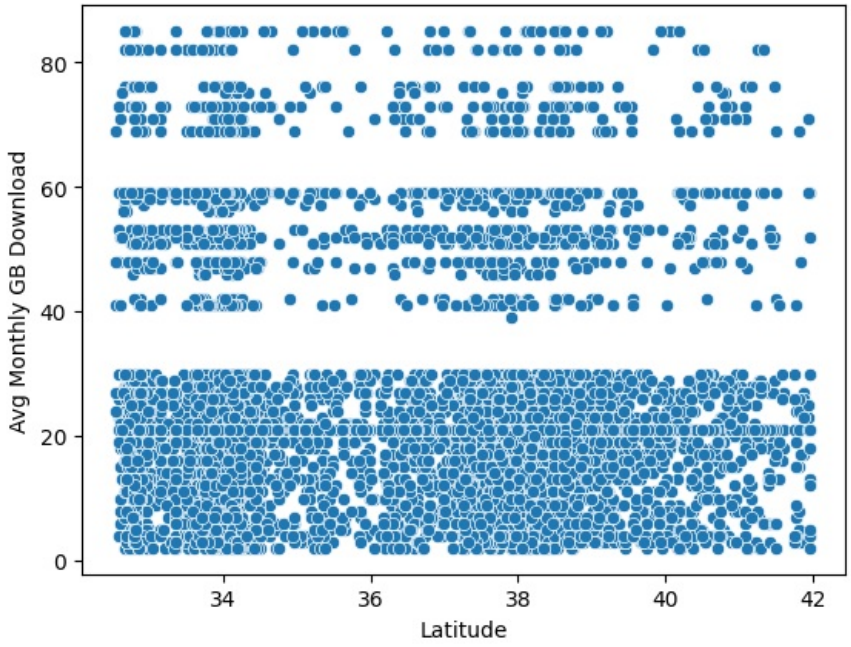
In [25]:
```
#(4)Show how the relationship is formed among columns using scatter plot?
for i in ['Age', 'Number of Dependents', 'Zip Code', 'Latitude', 'Longitude',
       'Number of Referrals', 'Tenure in Months',
       'Avg Monthly Long Distance Charges',
       'Monthly Charge', 'Total Charges', 'Total Refunds',
       'Total Extra Data Charges', 'Total Long Distance Charges',
```

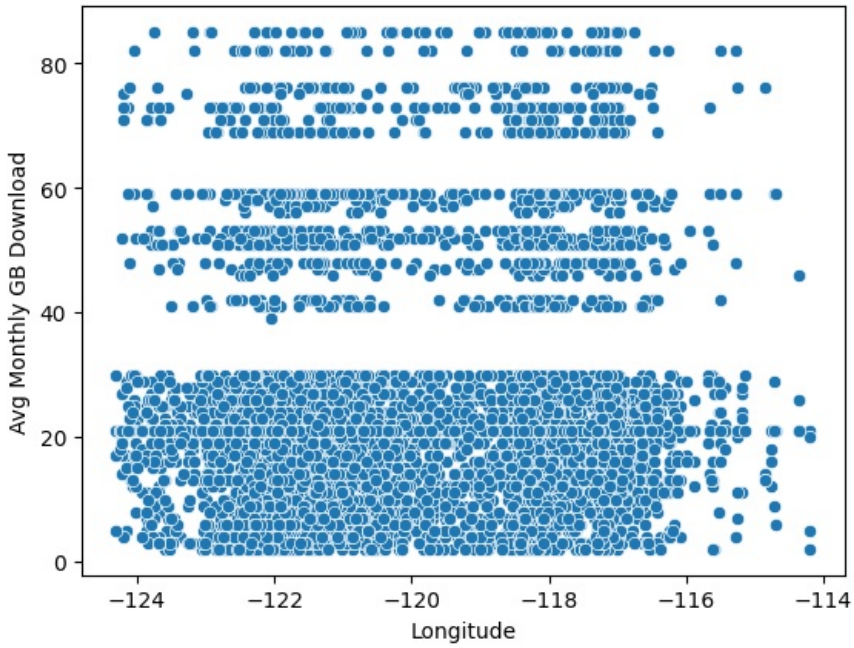
```
        'Total Revenue']:
    sns.scatterplot(data=df,x=i,y='Avg Monthly GB Download')
    plt.show()
```

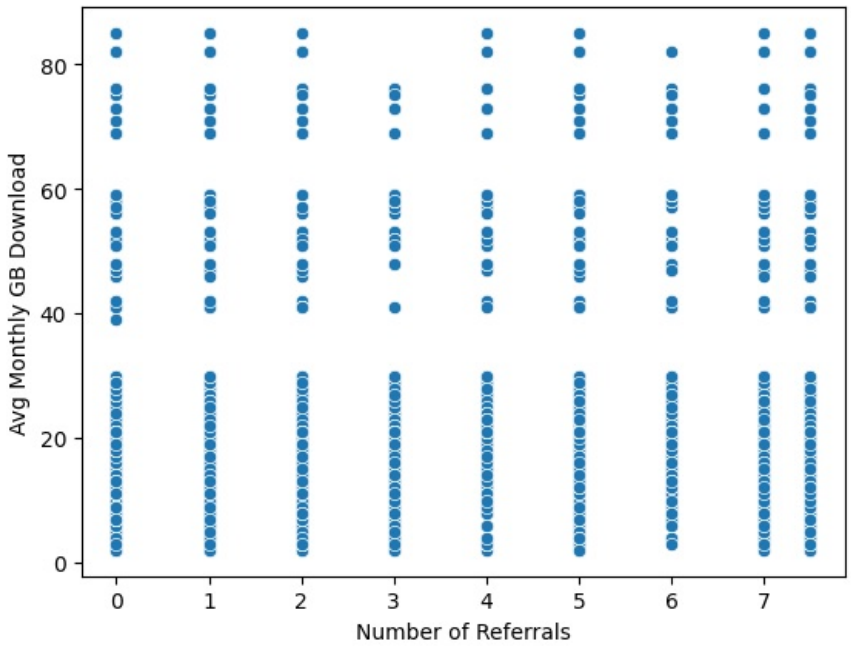


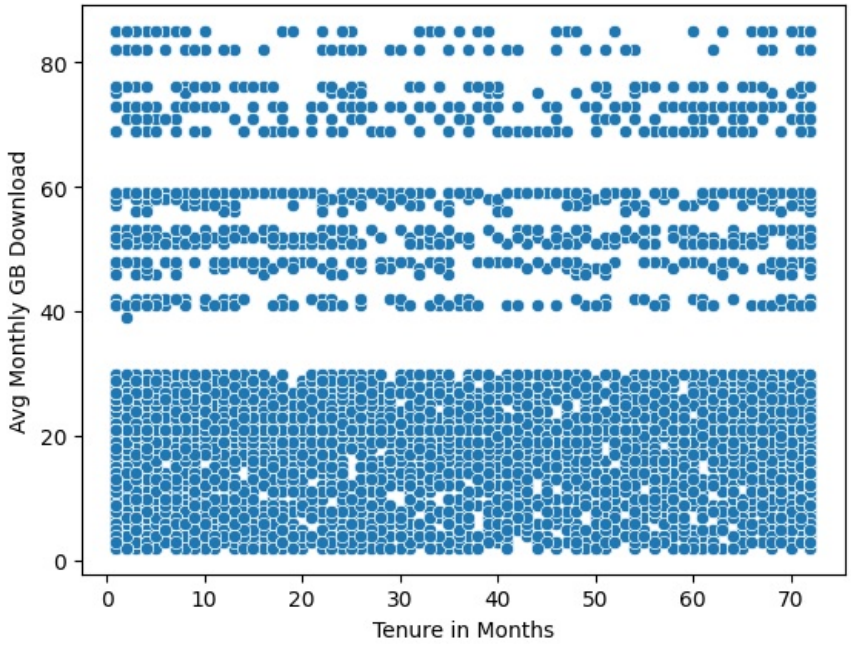


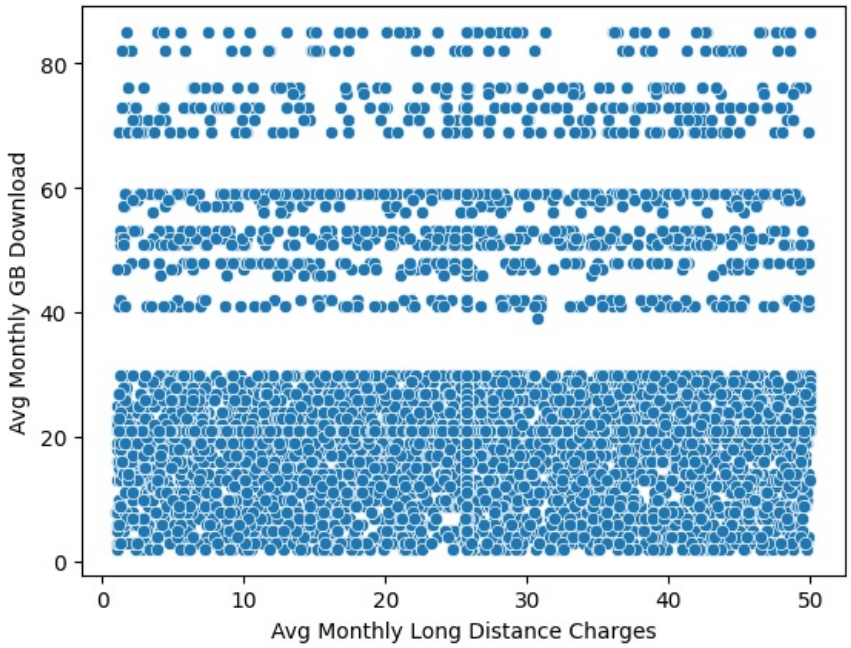


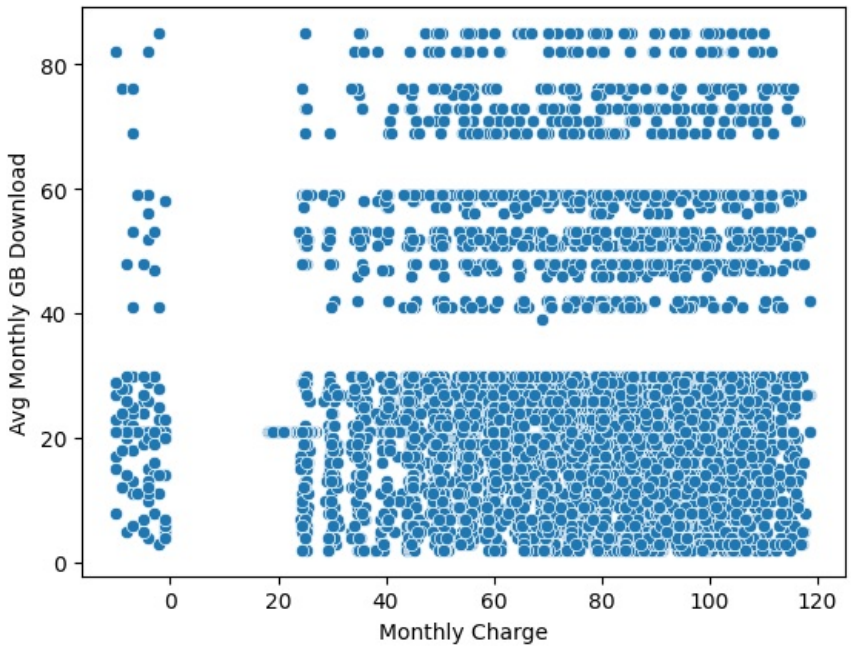

Avg Monthly GB Download
0
20
40
60
80
34
36
38
40
42
Latitude

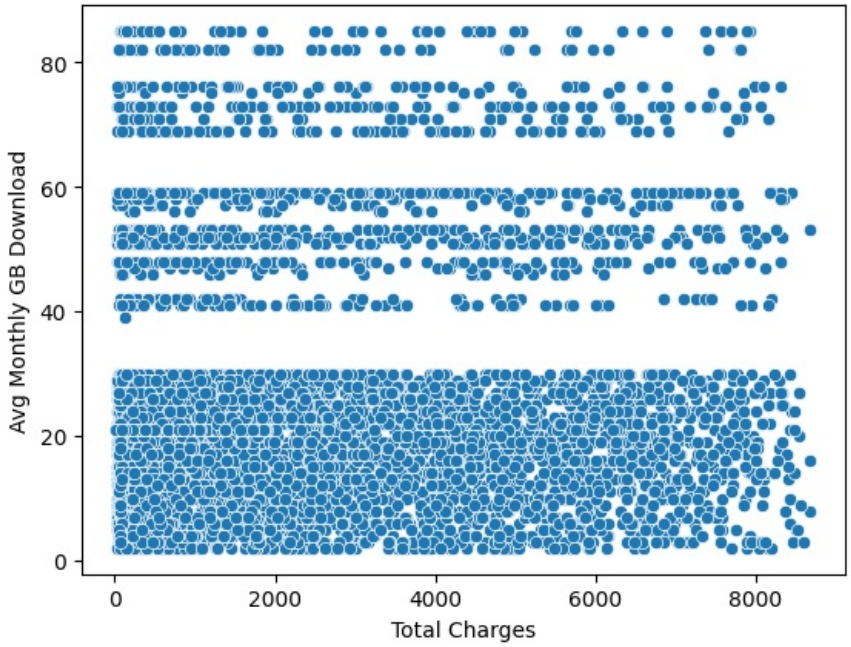

Avg Monthly GB Download
0
20
40
60
80
−124
−122
−120
−118
−116
−114
Longitude

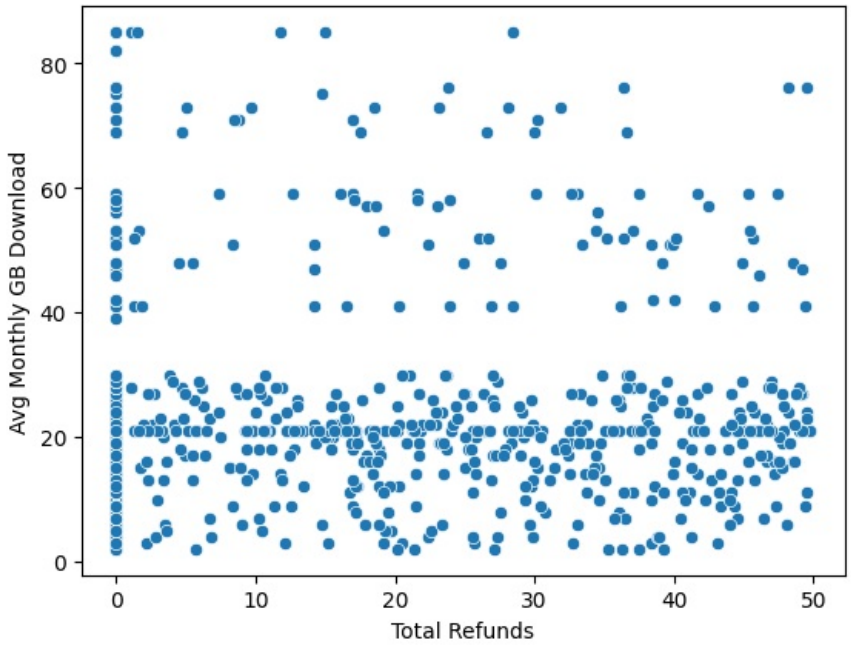

Avg Monthly GB Download
0
20
40
60
80
0
1
2
3
4
5
6
7
Number of Referrals

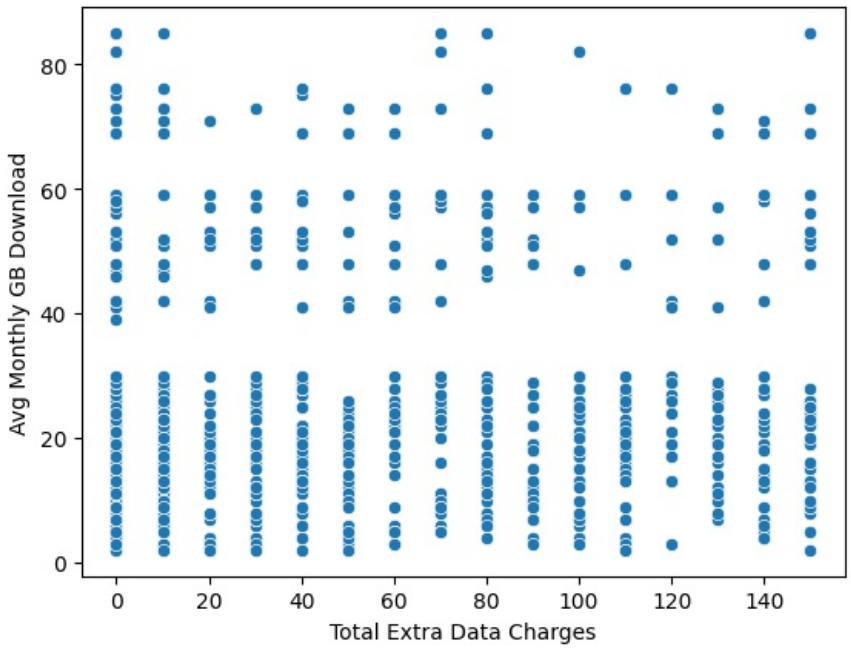

Avg Monthly GB Download
0
20
40
60
80
0
10
20
30
40
50
60
70
Tenure in Months

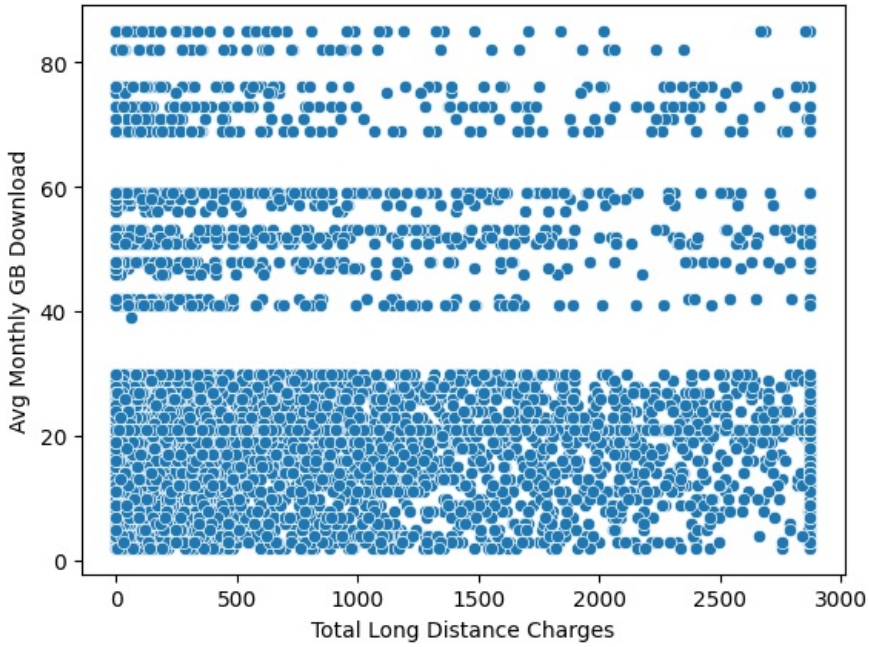

Avg Monthly GB Download
0
20
40
60
80
0
10
20
30
40
50
Avg Monthly Long Distance Charges

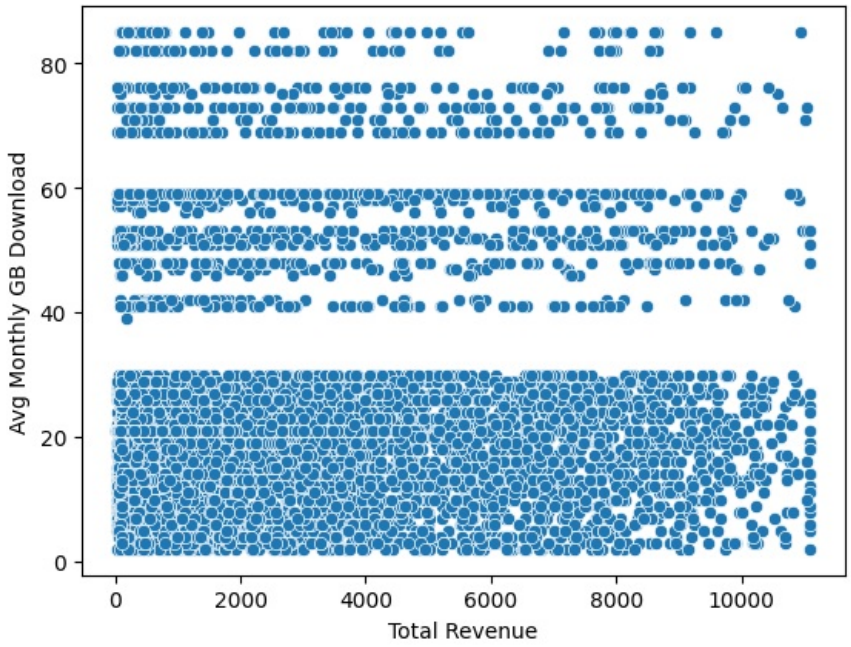

Avg Monthly GB Download
0
20
40
60
80
0
20
40
60
80
100
120
Monthly Charge


Avg Monthly GB Download
0
20
40
60
80
0
2000
4000
6000
8000
Total Charges

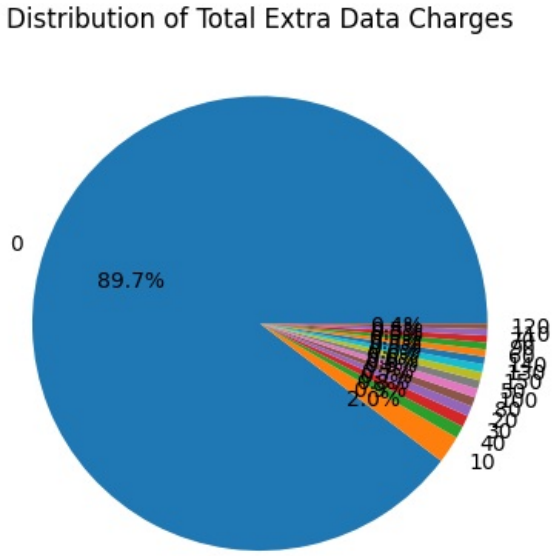

Avg Monthly GB Download
0
20
40
60
80
0
10
20
30
40
50
Total Refunds


Avg Monthly GB Download
0
20
40
60
80
0
20
40
60
80
100
120
140
Total Extra Data Charges





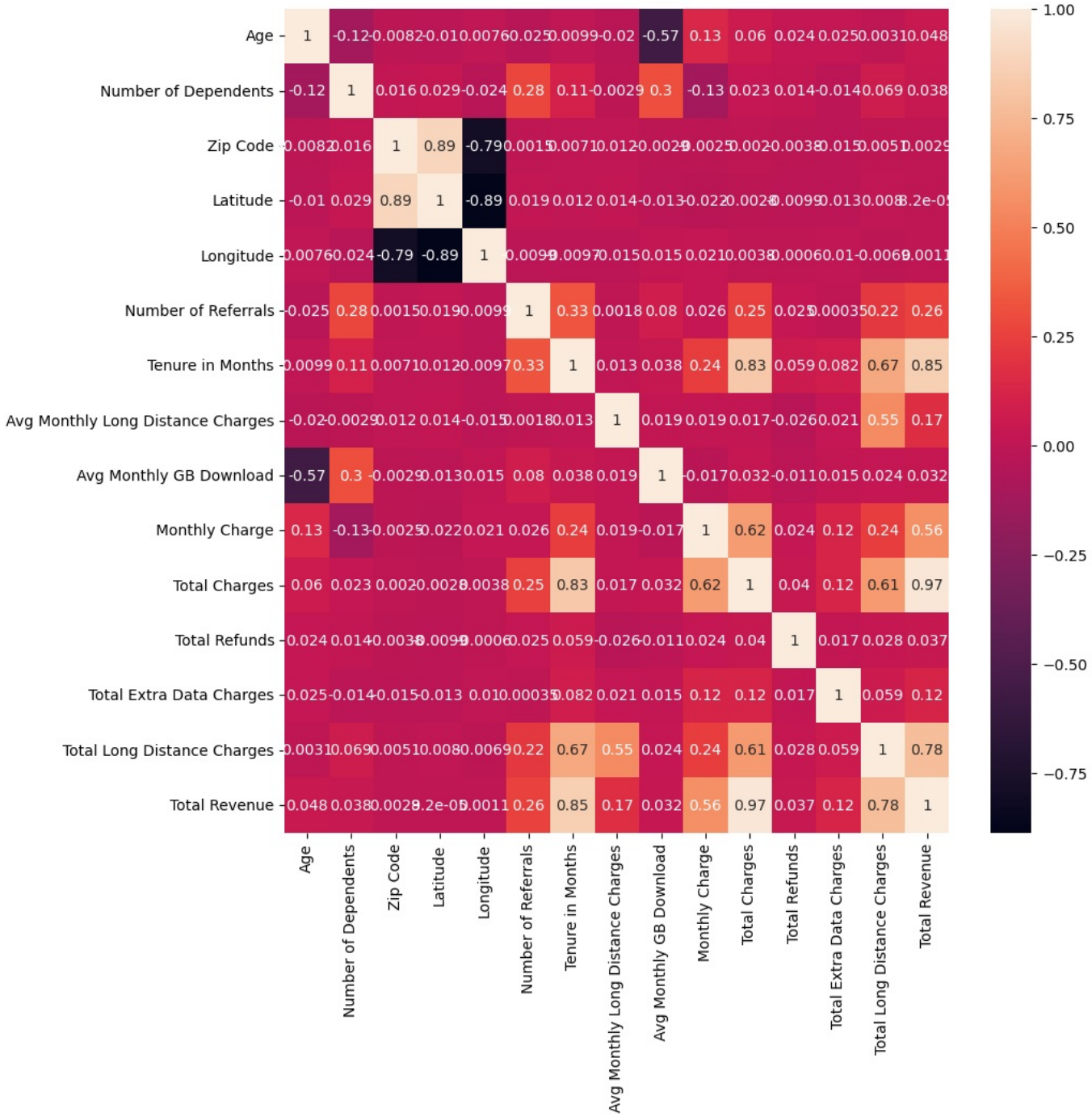
In [26]:
```python
#(5)What is the distribution of Total Extra Data Charges ?
import pandas as pd
df = pd.read_csv('telecom_customer_churn.csv')
import matplotlib.pyplot as plt

total_extra_data_charges  = df['Total Extra Data Charges'].value_counts()
plt.pie(total_extra_data_charges,labels=total_extra_data_charges.index, autopct='%1.1f%%')
plt.title('Distribution of Total Extra Data Charges')
plt.show()
```



In [27]:
```
#(6)correlation with heatmap to interpret the relation and multicolliniarity:
#Confusion matrix:
s=df.select_dtypes(include="number").corr()
```

In [31]:
```
plt.figure(figsize=(10,10))
sns.heatmap(s,annot=True)
```

Out[31]: <Axes: >



In [ ]: