Exploration Data Analysis

Objectives

- Understand the data
- Find out if it fits the task
- Gain insights for the model
- Find out what data cleaning and feature engineering is needed

Import Libraries

```
import pandas as pd
import numpy as np
import matplotlib as mpl
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
pd.set_option('display.max_columns', 50)
```

```
   Customer Gender Customer Status Churn Category Churn Reason  Age  Number of Referrals  Tenure in Months  Offer  Phone Service  Avg Monthly GB Download  Multiple Lines  Internet Service  Internet Type  Avg Monthly Long Distance Charges  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
0   0002-ORFBO  Female  Yes  Other  Competitor  37  0  9  None  Yes  42  No  Fiber Optic  Cable  79.65  No  No  No  No  No  No  No  No  Month-to-Month  Yes  Mailed Check  65.60  593.30  0.00  0  390.98  984.28
1   0004-TLHLJ  Male  No  NaN  NaN  51  1  6  None  No  NaN  No  DSL  DSL  NaN  No  No  No  No  No  No  No  No  Month-to-Month  No  Electronic Check  73.90  980.19  390.83  0  203.48  1174.03
2   0011-IGKFF  Male  No  NaN  NaN  70  1  63  Offer E  Yes  83  Yes  Fiber Optic  Fiber Optic  1024  No  No  No  No  No  No  No  No  One Year  Yes  Electronic Check  98.00  5957.9  0.00  0  4627.19  10585.1
3   0013-EXCHZ  Female  No  NaN  NaN  67  0  7  None  Yes  267  No  Fiber Optic  Cable  381.51  No  No  No  No  No  No  No  No  Month-to-Month  Yes  Mailed Check  83.90  587.45  0.00  0  2671.63  3259.08
4   0013-MHZWF  Female  No  NaN  NaN  67  0  23  Offer D  Yes  181  No  Fiber Optic  Cable  235.17  No  No  No  No  No  No  No  No  Month-to-Month  Yes  Credit Card  69.40  1587.3  0.00  0  5401.41  6988.71
5   0014-BMAQU  Female  No  NaN  NaN  63  2  17  None  Yes  304  No  Fiber Optic  Cable  265.10  No  No  No  No  No  No  No  No  Month-to-Month  No  Mailed Check  84.65  1837.5  0.00  0  4503.09  6340.59
6   0015-UOCOJ  Female  No  NaN  NaN  54  0  2  None  Yes  351  No  Fiber Optic  Cable  448.70  No  No  No  No  No  No  No  No  Month-to-Month  No  Electronic Check  90.35  108.15  0.00  0  897.31  1005.46
7   0016-QLJIS  Female  No  NaN  NaN  72  0  51  Offer A  Yes  526  No  Fiber Optic  Cable  1062.33  No  No  No  No  No  No  No  No  One Year  Yes  Mailed Check  94.15  4799.15  0.00  0  54064.3  58863.5
8   0017-DINOC  Male  No  NaN  NaN  70  0  60  None  No  NaN  No  DSL  DSL  NaN  No  No  No  No  No  No  No  No  Two Year  No  Mailed Check  48.20  2875.95  0.00  0  2888.65  5764.6
9   0018-SNJER  Female  No  NaN  NaN  78  0  9  None  Yes  20  No  Fiber Optic  Cable  16.43  No  No  No  No  No  No  No  No  Month-to-Month  No  Electronic Check  75.45  664.45  0.00  0  147.87  812.32
10  0019-EFAEP  Female  No  NaN  NaN  77  0  10  None  No  NaN  No  DSL  DSL  NaN  No  No  No  No  No  No  No  No  Month-to-Month  No  Mailed Check  46.75  452.1  0.00  0  467.5  919.6
11  0019-GFNxH  Female  No  NaN  NaN  67  0  4  None  Yes  8  No  Fiber Optic  Cable  13.43  No  No  No  No  No  No  No  No  Month-to-Month  No  Credit Card  75.10  289.75  0.00  0  53.72  343.47
12  0020-INWCK  Female  No  NaN  NaN  77  0  13  None  Yes  387  No  Fiber Optic  Cable  279.82  No  No  No  No  No  No  No  No  One Year  Yes  Electronic Check  95.45  1275.4  0.00  0  3639.66  4915.06
13  0020-TCJCI  Female  No  NaN  NaN  62  0  9  None  Yes  214  No  Fiber Optic  Cable  148.62  No  No  No  No  No  No  No  No  Month-to-Month  No  Mailed Check  85.30  759.05  0.00  0  1337.58  2096.63
14  0021-IUPVP  Female  No  NaN  NaN  85  0  2  None  Yes  1214  No  Fiber Optic  Cable  1214.00  No  No  No  No  No  No  No  No  Month-to-Month  No  Credit Card  75.05  150.1  0.00  0  2428.1  2578.2
15  0022-TCJIV  Female  No  NaN  NaN  67  0  27  Offer D  Yes  268  No  Fiber Optic  Cable  268.00  No  No  No  No  No  No  No  No  Month-to-Month  No  Credit Card  70.15  1894.05  0.00  0  7236.36  9130.41
16  0023-JHJKW  Female  No  NaN  NaN  50  0  12  None  Yes  1232  No  Fiber Optic  Cable  1232.00  No  No  No  No  No  No  No  No  Month-to-Month  No  Credit Card  70.15  841.8  0.00  0  14784.0  15625.8
17  0025-WKFRX  Male  No  NaN  NaN  78  0  19  None  No  NaN  No  DSL  DSL  NaN  No  No  No  No  No  No  No  No  Two Year  No  Mailed Check  45.05  855.95  0.00  0  855.95  1711.9
18  0026-XXZWD  Female  No  NaN  NaN  80  0  1  None  Yes  1  No  Fiber Optic  Cable  0.83  No  No  No  No  No  No  No  No  Month-to-Month  No  Credit Card  75.45  75.45  0.00  0  0.83  76.28
19  0027-NCNCB  Female  No  NaN  NaN  85  0  72  None  Yes  1036  No  Fiber Optic  Cable  1036.00  No  No  No  No  No  No  No  No  One Year  Yes  Credit Card  95.25  6886.1  0.00  0  74592.0  81478.1
```

```
(7043, 33)
```

```
Index(['Customer', 'Gender', 'Customer Status', 'Churn Category',
       'Churn Reason', 'Age', 'Number of Referrals', 'Tenure in Months',
       'Offer', 'Phone Service', 'Avg Monthly GB Download', 'Multiple Lines',
       'Internet Service', 'Internet Type', 'Avg Monthly Long Distance Charges',
       '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'],
      dtype=object)
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 33 columns):
 0   Customer                           7043 non-null   object
 1   Gender                             7043 non-null   object
 2   Customer Status                    7043 non-null   object
 3   Churn Category                     1869 non-null   object
 4   Churn Reason                       1869 non-null   object
 5   Age                                7043 non-null   int64
 6   Number of Referrals                7043 non-null   int64
 7   Tenure in Months                   7043 non-null   int64
 8   Offer                              3166 non-null   object
 9   Phone Service                      7043 non-null   object
 10  Avg Monthly GB Download            5517 non-null   int64
 11  Multiple Lines                     7043 non-null   object
 12  Internet Service                   7043 non-null   object
 13  Internet Type                      5517 non-null   object
 14  Avg Monthly Long Distance Charges  7043 non-null   float64
 15  Online Security                    5517 non-null   object
 16  Online Backup                      5517 non-null   object
 17  Device Protection Plan             5517 non-null   object
 18  Premium Tech Support               5517 non-null   object
 19  Streaming TV                       5517 non-null   object
 20  Streaming Movies                   5517 non-null   object
 21  Streaming Music                    5517 non-null   object
 22  Unlimited Data                     5517 non-null   object
 23  Contract                           7043 non-null   object
 24  Paperless Billing                  7043 non-null   object
 25  Payment Method                     7043 non-null   object
 26  Monthly Charge                     7043 non-null   float64
 27  Total Charges                      7043 non-null   float64
 28  Total Refunds                      7043 non-null   float64
 29  Total Extra Data Charges           7043 non-null   int64
 30  Total Long Distance Charges        7043 non-null   float64
 31  Total Revenue                      7043 non-null   float64
dtype: float64, int64, object
memory usage: 1.7MB
```

```
Customer                           0
Gender                             0
Customer Status                    0
Churn Category                     5174
Churn Reason                       5174
Age                                0
Number of Referrals                0
Tenure in Months                   0
Offer                              3877
Phone Service                      0
Avg Monthly GB Download            1526
Multiple Lines                     0
Internet Service                   0
Internet Type                      1526
Avg Monthly Long Distance Charges  0
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
dtype: int64
```

```
Customer                           0.000000
Gender                             0.000000
Customer Status                    0.000000
Churn Category                     0.734488
Churn Reason                       0.734488
Age                                0.000000
Number of Referrals                0.000000
Tenure in Months                   0.000000
Offer                              0.550615
Phone Service                      0.000000
Avg Monthly GB Download            0.216669
Multiple Lines                     0.000000
Internet Service                   0.000000
Internet Type                      0.216669
Avg Monthly Long Distance Charges  0.000000
Online Security                    0.216669
Online Backup                      0.216669
Device Protection Plan             0.216669
Premium Tech Support               0.216669
Streaming TV                       0.216669
Streaming Movies                   0.216669
Streaming Music                    0.216669
Unlimited Data                     0.216669
Contract                           0.000000
Paperless Billing                  0.000000
Payment Method                     0.000000
Monthly Charge                     0.000000
Total Charges                      0.000000
Total Refunds                      0.000000
Total Extra Data Charges           0.000000
Total Long Distance Charges        0.000000
Total Revenue                      0.000000
dtype: float64
```

Could be an issue if we have 17% of missing data on miles being driven.

Check correlation between the missing data in Churn Category and the other columns to see if there is a pattern in the missing data.

```
Churn Category
Customer Status    1.000000
Churn Reason       0.000000
Age                0.000000
Number of Referrals    0.000000
Tenure in Months       0.000000
Offer              0.000000
Phone Service      0.000000
Avg Monthly GB Download    0.000000
Multiple Lines     0.000000
Internet Service   0.000000
Internet Type      0.000000
Avg Monthly Long Distance Charges    0.000000
Online Security    0.000000
Online Backup      0.000000
Device Protection Plan    0.000000
Premium Tech Support    0.000000
Streaming TV       0.000000
Streaming Movies   0.000000
Streaming Music    0.000000
Unlimited Data     0.000000
Contract           0.000000
Paperless Billing  0.000000
Payment Method     0.000000
Monthly Charge     0.000000
Total Charges      0.000000
Total Refunds      0.000000
Total Extra Data Charges    0.000000
Total Long Distance Charges    0.000000
Total Revenue      0.000000
dtype: float64
```