

Task 3 - Customer Churn Prediction

```
# Importing Necessary Library import pandas as pd import numpy as np from sklearn.model_selection import train_test_split from sklearn.;
```

```
import pandas as pd
```

```
# Replace 'your_data_file.csv' with the actual path to your data file.  
telecom_customer_churn = pd.read_csv('/content/telecom_customer_churn.csv')
```

```
# Print the first few rows of the DataFrame.  
print(telecom_customer_churn.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]

```
print(telecom_customer_churn.tail())
```

	Customer ID	Gender	Age	Married	Number of Dependents	City	\
7038	9987-LUTYD	Female	20	No	0	La Mesa	
7039	9992-RRAMN	Male	40	Yes	0	Riverbank	
7040	9992-UJOEL	Male	22	No	0	Elk	
7041	9993-LHIEB	Male	21	Yes	0	Solana Beach	
7042	9995-HOTOH	Male	36	Yes	0	Sierra City	

	Zip Code	Latitude	Longitude	Number of Referrals	...	\
7038	91941	32.759327	-116.997260	0	...	
7039	95367	37.734971	-120.954271	1	...	
7040	95432	39.108252	-123.645121	0	...	
7041	92075	33.001813	-117.263628	5	...	
7042	96125	39.600599	-120.636358	1	...	

	Payment Method	Monthly Charge	Total Charges	Total Refunds	\
7038	Credit Card	55.15	742.90	0.0	
7039	Bank Withdrawal	85.10	1873.70	0.0	
7040	Credit Card	50.30	92.75	0.0	
7041	Credit Card	67.85	4627.65	0.0	
7042	Bank Withdrawal	59.00	3707.60	0.0	

	Total Extra Data Charges	Total Long Distance Charges	Total Revenue	\
7038	0	606.84	1349.74	
7039	0	356.40	2230.10	
7040	0	37.24	129.99	
7041	0	142.04	4769.69	
7042	0	0.00	3707.60	

	Customer Status	Churn Category	Churn Reason
7038	Stayed	NaN	NaN
7039	Churned	Dissatisfaction	Product dissatisfaction
7040	Joined	NaN	NaN

7041	Stayed	NaN	NaN
7042	Stayed	NaN	NaN

[5 rows x 38 columns]

```
print(telecom_customer_churn.shape)
```

(7043, 38)

```
print(telecom_customer_churn.columns)
```

```
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')
```

```
column_to_check = 'Customer ID'
```

```
is_unique = telecom_customer_churn[column_to_check].is_unique
```

```
print(f"Are values in '{column_to_check}' unique? {is_unique}")
```

Are values in 'Customer ID' unique? True

```
print(telecom_customer_churn.nunique())
```

Customer ID	7043
Gender	2
Age	62
Married	2
Number of Dependents	10
City	1106
Zip Code	1626
Latitude	1626
Longitude	1625
Number of Referrals	12
Tenure in Months	72
Offer	6
Phone Service	2
Avg Monthly Long Distance Charges	3583
Multiple Lines	2
Internet Service	2
Internet Type	3
Avg Monthly GB Download	49
Online Security	2
Online Backup	2
Device Protection Plan	2
Premium Tech Support	2
Streaming TV	2
Streaming Movies	2
Streaming Music	2
Unlimited Data	2
Contract	3
Paperless Billing	2
Payment Method	3
Monthly Charge	1591
Total Charges	6540
Total Refunds	500
Total Extra Data Charges	16
Total Long Distance Charges	6068
Total Revenue	6975
Customer Status	3
Churn Category	5
Churn Reason	20

dtype: int64

```
print(telecom_customer_churn.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
None

```

```

column_to_drop = 'Age'
data = telecom_customer_churn.drop(column_to_drop, axis=1)

```

```

rows_to_drop = [0, 1, 2]
data = telecom_customer_churn.drop(rows_to_drop)

```

```

print(telecom_customer_churn.describe())

```

	Age	Number of Dependents	Zip Code	Latitude \
count	7043.000000	7043.000000	7043.000000	7043.000000
mean	46.509726	0.468692	93486.070567	36.197455
std	16.750352	0.962802	1856.767505	2.468929
min	19.000000	0.000000	90001.000000	32.555828
25%	32.000000	0.000000	92101.000000	33.990646
50%	46.000000	0.000000	93518.000000	36.205465
75%	60.000000	0.000000	95329.000000	38.161321
max	80.000000	9.000000	96150.000000	41.962127

	Longitude	Number of Referrals	Tenure in Months \
count	7043.000000	7043.000000	7043.000000
mean	-119.756684	1.951867	32.386767
std	2.154425	3.001199	24.542061
min	-124.301372	0.000000	1.000000
25%	-121.788090	0.000000	9.000000
50%	-119.595293	0.000000	29.000000
75%	-117.969795	3.000000	55.000000
max	-114.192901	11.000000	72.000000

	Avg Monthly Long Distance Charges	Avg Monthly GB Download \
count	6361.000000	5517.000000
mean	25.420517	26.189958
std	14.200374	19.586585
min	1.010000	2.000000
25%	13.050000	13.000000
50%	25.690000	21.000000
75%	37.680000	30.000000
max	49.990000	85.000000

	Monthly Charge	Total Charges	Total Refunds	Total Extra Data Charges \
count	7043.000000	7043.000000	7043.000000	7043.000000
mean	63.596131	2280.381264	1.962182	6.860713
std	31.204743	2266.220462	7.902614	25.104978
min	-10.000000	18.800000	0.000000	0.000000
25%	30.400000	400.150000	0.000000	0.000000
50%	70.050000	1394.550000	0.000000	0.000000
75%	89.750000	3786.600000	0.000000	0.000000
max	118.750000	8684.800000	49.790000	150.000000

	Total Long Distance Charges	Total Revenue
count	7043.000000	7043.000000
mean	749.099262	3034.379056

std	846.660055	2865.204542
min	0.000000	21.360000
25%	70.545000	605.610000
50%	401.440000	2108.640000
75%	1191.100000	4801.145000
max	3564.720000	11979.340000

telecom_customer_churn.dtypes

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

print(telecom_customer_churn.isnull())

	Customer ID	Gender	Age	Married	Number of Dependents	City	\
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	
7038	False	False	False	False	False	False	
7039	False	False	False	False	False	False	
7040	False	False	False	False	False	False	
7041	False	False	False	False	False	False	
7042	False	False	False	False	False	False	

	Zip Code	Latitude	Longitude	Number of Referrals	...	Payment Method	\
0	False	False	False	False	...	False	
1	False	False	False	False	...	False	
2	False	False	False	False	...	False	
3	False	False	False	False	...	False	
4	False	False	False	False	...	False	
...	
7038	False	False	False	False	...	False	
7039	False	False	False	False	...	False	
7040	False	False	False	False	...	False	
7041	False	False	False	False	...	False	
7042	False	False	False	False	...	False	

	Monthly Charge	Total Charges	Total Refunds	Total Extra Data Charges	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
7038	False	False	False	False	
7039	False	False	False	False	
7040	False	False	False	False	
7041	False	False	False	False	

7042 False False False False

	Total Long Distance Charges	Total Revenue	Customer Status	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
...	
7038	False	False	False	
7039	False	False	False	
7040	False	False	False	
7041	False	False	False	
7042	False	False	False	

	Churn Category	Churn Reason
0	True	True
1	True	True
2	False	False
3	False	False
4	False	False

```
print(telecom_customer_churn.isnull().sum())
```

```
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  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
```

```
# Check for duplicate rows duplicates = data[telecom_customer_churn.duplicated()] # Display duplicate rows print(duplicates)
```

```
# Remove duplicate rows data_no_duplicates = telecom_customer_churn.drop_duplicates()
```

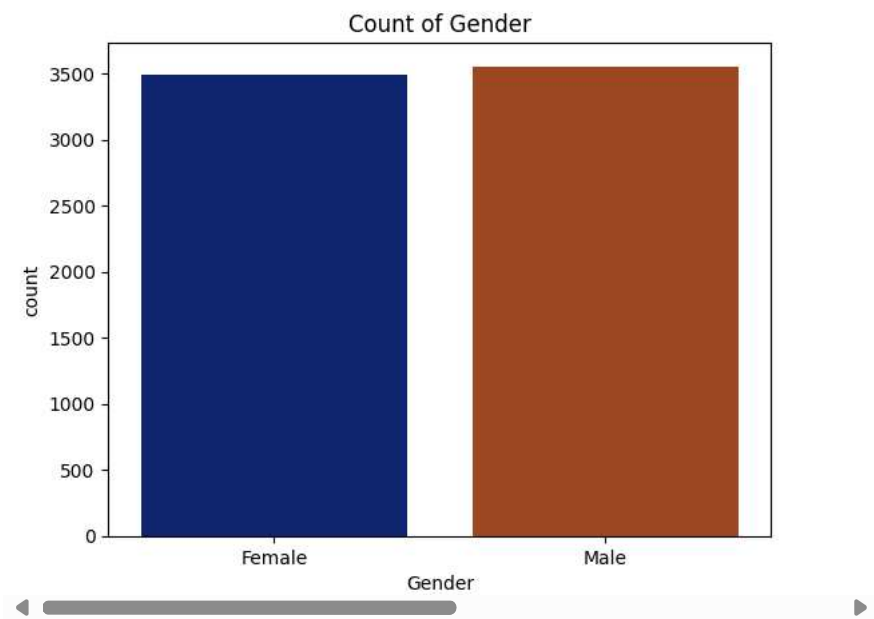
```
print(telecom_customer_churn.duplicated().sum())
```

0

```
# Countplot Analysis
import seaborn as sns
import matplotlib.pyplot as plt
```

```
sns.countplot(x='Gender', data=telecom_customer_churn, palette='dark')
plt.title('Count of Gender')
plt.show()
```

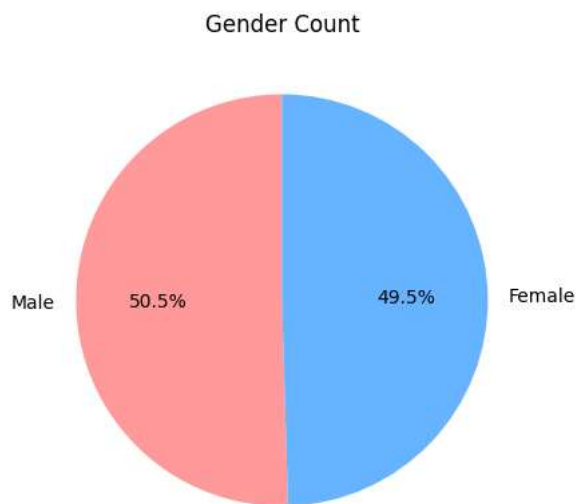
```
<ipython-input-76-e87acb6d9a0e>:5: FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.
sns.countplot(x='Gender', data=telecom_customer_churn, palette='dark')
```



```
import pandas as pd
import matplotlib.pyplot as plt

# Pie chart based on the counts
churn_counts = telecom_customer_churn['Gender'].value_counts()

plt.figure(figsize=(5, 5))
plt.pie(churn_counts, labels=churn_counts.index, autopct='%1.1f%%', startangle=90, colors=['#ff9999', '#66b3ff'])
plt.title('Gender Count')
plt.show()
```



```
# Heatmap Analysis
import seaborn as sns
import matplotlib.pyplot as plt

correlation_matrix = telecom_customer_churn.corr()
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='viridis', fmt=".2f", linewidths=0.5)
plt.title('Correlation Heatmap', fontsize=8)
plt.show()
```

```
<ipython-input-92-9154a62379be>:5: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future ver correlation_matrix = telecom_customer_churn.corr()
```

