In [52]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
data=pd.read_csv(r"C:\Users\lenovo\Desktop\data for project\code veda\Churn Prdicti
data
```

Out[52]:

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Tot ev minut
0	LA	117	408	No	No	0	184.5	97	31.37	351
1	IN	65	415	No	No	0	129.1	137	21.95	228
2	NY	161	415	No	No	0	332.9	67	56.59	317
3	SC	111	415	No	No	0	110.4	103	18.77	137
4	HI	49	510	No	No	0	119.3	117	20.28	215
...	...	...	...	...	...	...	...	...	...	
662	WI	114	415	No	Yes	26	137.1	88	23.31	155
663	AL	106	408	No	Yes	29	83.6	131	14.21	203
664	VT	60	415	No	No	0	193.9	118	32.96	85
665	WV	159	415	No	No	0	169.8	114	28.87	197
666	CT	184	510	Yes	No	0	213.8	105	36.35	159

667 rows × 20 columns

In [53]:
```
data.isnull().sum()
```

Out[53]:
```
State                     0
Account length            0
Area code                 0
International plan        0
Voice mail plan           0
Number vmail messages     0
Total day minutes         0
Total day calls           0
Total day charge          0
Total eve minutes         0
Total eve calls           0
Total eve charge          0
Total night minutes       0
Total night calls         0
Total night charge        0
Total intl minutes        0
Total intl calls          0
Total intl charge         0
Customer service calls    0
Churn                     0
dtype: int64
```

In [54]:
```
data.describe()
```

Out[54]:

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes
count	667.000000	667.000000	667.000000	667.000000	667.000000	667.000000	667.000000
mean	102.841079	436.157421	8.407796	180.948126	100.937031	30.761769	203.355322
std	40.819480	41.783305	13.994480	55.508628	20.396790	9.436463	49.719268
min	1.000000	408.000000	0.000000	25.900000	30.000000	4.400000	48.100000
25%	76.000000	408.000000	0.000000	146.250000	87.500000	24.860000	171.050000
50%	102.000000	415.000000	0.000000	178.300000	101.000000	30.310000	203.700000
75%	128.000000	415.000000	20.000000	220.700000	115.000000	37.520000	236.450000
max	232.000000	510.000000	51.000000	334.300000	165.000000	56.830000	361.800000

In [55]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 667 entries, 0 to 666
Data columns (total 20 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   State                   667 non-null    object 
 1   Account length          667 non-null    int64  
 2   Area code               667 non-null    int64  
 3   International plan      667 non-null    object 
 4   Voice mail plan         667 non-null    object 
 5   Number vmail messages   667 non-null    int64  
 6   Total day minutes       667 non-null    float64
 7   Total day calls         667 non-null    int64  
 8   Total day charge        667 non-null    float64
 9   Total eve minutes       667 non-null    float64
 10  Total eve calls         667 non-null    int64  
 11  Total eve charge        667 non-null    float64
 12  Total night minutes     667 non-null    float64
 13  Total night calls       667 non-null    int64  
 14  Total night charge      667 non-null    float64
 15  Total intl minutes      667 non-null    float64
 16  Total intl calls        667 non-null    int64  
 17  Total intl charge       667 non-null    float64
 18  Customer service calls  667 non-null    int64  
 19  Churn                   667 non-null    bool   
dtypes: bool(1), float64(8), int64(8), object(3)
memory usage: 99.8+ KB
```

In [56]:
```
data.shape
```

Out[56]:
```
(667, 20)
```

In [57]:
```
data.dtypes
```

Out[57]:
```
State                    object
Account length            int64
Area code                 int64
International plan       object
Voice mail plan          object
Number vmail messages     int64
Total day minutes       float64
Total day calls           int64
Total day charge        float64
Total eve minutes       float64
Total eve calls           int64
Total eve charge        float64
Total night minutes     float64
Total night calls         int64
Total night charge      float64
Total intl minutes      float64
Total intl calls          int64
Total intl charge       float64
Customer service calls    int64
Churn                      bool
dtype: object
```

In [58]:
```
data.duplicated().sum()
```

Out[58]:
```
np.int64(0)
```

In [59]:
```
list(data.columns)
```

Out[59]:
```
['State',
 'Account length',
 'Area code',
 'International plan',
 'Voice mail plan',
 'Number vmail messages',
 'Total day minutes',
 'Total day calls',
 'Total day charge',
 'Total eve minutes',
 'Total eve calls',
 'Total eve charge',
 'Total night minutes',
 'Total night calls',
 'Total night charge',
 'Total intl minutes',
 'Total intl calls',
 'Total intl charge',
 'Customer service calls',
 'Churn']
```

In [60]:
```
data["Churn"].value_counts()
```

Out[60]:
```
Churn
False    572
True      95
Name: count, dtype: int64
```

In [61]:
```
data["International plan"].value_counts()
```

Out[61]:
```
International plan
No     614
Yes     53
Name: count, dtype: int64
```

In [62]:
```
data["Voice mail plan"].value_counts()
```

Out[62]:
```
Voice mail plan
No     478
Yes    189
Name: count, dtype: int64
```

In [69]:
```
data["State"].nunique()
```

Out[69]:
```
51
```

In [64]:
```
data[["Total day minutes","Total eve minutes","Total night minutes"]].value_counts(
```

Out[64]:
```
Total day minutes  Total eve minutes  Total night minutes
334.3              192.1              191.0                 1
25.9               206.5              228.1                 1
35.1               180.8              251.6                 1
40.9               133.4              264.2                 1
44.9               134.2              168.4                 1
                                                           ..
83.6               203.9              229.5                 1
83.8               240.2              158.6                 1
85.9               193.9              231.5                 1
87.2               169.3              166.7                 1
87.6               262.0              184.6                 1
Name: count, Length: 667, dtype: int64
```

Customers who churn usually make more customer service calls.

In [65]:
```
data.groupby("Churn")["Customer service calls"].mean()
```

Out[65]:
```
Churn
False    1.437063
True     2.326316
Name: Customer service calls, dtype: float64
```

Customers with international plans show higher churn.

In [66]:
```
pd.crosstab(data["International plan"],data["Churn"])
```

Out[66]:

Churn	False	True
International plan		
No	538	76
Yes	34	19

In [70]:
```
data.groupby("Churn")["Total day charge"].mean()
```

Out[70]:
```
Churn
False    29.828829
True     36.379053
Name: Total day charge, dtype: float64
```

In [71]:
```
data.corr(numeric_only=True)
```

Out[71]:

	Account length	Area code	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Tota
Account length	1.000000	-0.026327	-0.011993	0.017833	0.035703	0.017839	0.027043	0.02
Area code	-0.026327	1.000000	-0.006907	0.051507	-0.008972	0.051492	0.017160	0.01
Number vmail messages	-0.011993	-0.006907	1.000000	-0.069172	-0.009952	-0.069187	0.040865	-0.05
Total day minutes	0.017833	0.051507	-0.069172	1.000000	-0.032306	1.000000	0.017987	0.04
Total day calls	0.035703	-0.008972	-0.009952	-0.032306	1.000000	-0.032319	-0.004688	0.00
Total day charge	0.017839	0.051492	-0.069187	1.000000	-0.032319	1.000000	0.017983	0.04
Total eve minutes	0.027043	0.017160	0.040865	0.017987	-0.004688	0.017983	1.000000	-0.02
Total eve calls	0.021237	0.017783	-0.051951	0.043219	0.005851	0.043231	-0.029077	1.00
Total eve charge	0.027051	0.017182	0.040876	0.017945	-0.004664	0.017941	1.000000	-0.02
Total night minutes	-0.007527	-0.016832	0.039751	-0.031600	0.079536	-0.031613	-0.007705	-0.00
Total night calls	0.027228	0.036421	0.003367	0.052761	-0.030074	0.052748	0.001938	0.03
Total night charge	-0.007528	-0.016818	0.039680	-0.031603	0.079529	-0.031616	-0.007603	-0.00
Total intl minutes	0.002362	-0.037980	0.029949	-0.006725	-0.015319	-0.006720	-0.027855	-0.00
Total intl calls	0.031279	-0.010530	-0.036847	0.016597	-0.005155	0.016582	0.002929	0.07
Total intl charge	0.002456	-0.038044	0.029999	-0.006841	-0.015201	-0.006836	-0.027887	-0.00
Customer service calls	-0.027677	0.000103	0.007859	0.029291	-0.045953	0.029290	-0.012213	0.00
Churn	0.012315	0.027129	-0.102381	0.242781	0.019360	0.242777	0.175614	0.05

In [80]:
```
data['total_minutes'] = (data['Total day minutes'] + data['Total eve minutes'] + da
(data['total_minutes'])
```

Out[80]:
```
0      751.9
1      566.4
2      811.3
3      437.3
4      513.1
       ...
662    540.4
663    517.0
664    489.0
665    561.2
666    512.6
Name: total_minutes, Length: 667, dtype: float64
```

In [ ]:
```
# Customers with international plans churn more
# More customer service calls = higher churn
# Higher usage & charges increase churn
# Long-term customers churn less
```

Conclusion

In [ ]:
```
# Exploratory Data Analysis of the churn dataset reveals that customer service call
# international plan usage, and higher charges are major factors contributing to cu
# These insights can help businesses take preventive actions to improve customer re
```

I performed EDA on the churn dataset by cleaning data, analyzing customer behavior,

identifying churn patterns, and extracting business insights.