Churn dataset EDA - includes chi-squared, count plot, crosstab, heatmap

In [1]:
```
import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import chi2_contingency
%matplotlib inline
import warnings
warnings.simplefilter(action="ignore", category=FutureWarning)
```

In [2]:
```
df = pd.read_csv('C:\WGU\D207\churn_clean.csv')
```

In [3]:
```
#chisquare
chichurn = pd.crosstab(df['Churn'], df['Area'])
chichurn
```

Out[3]:

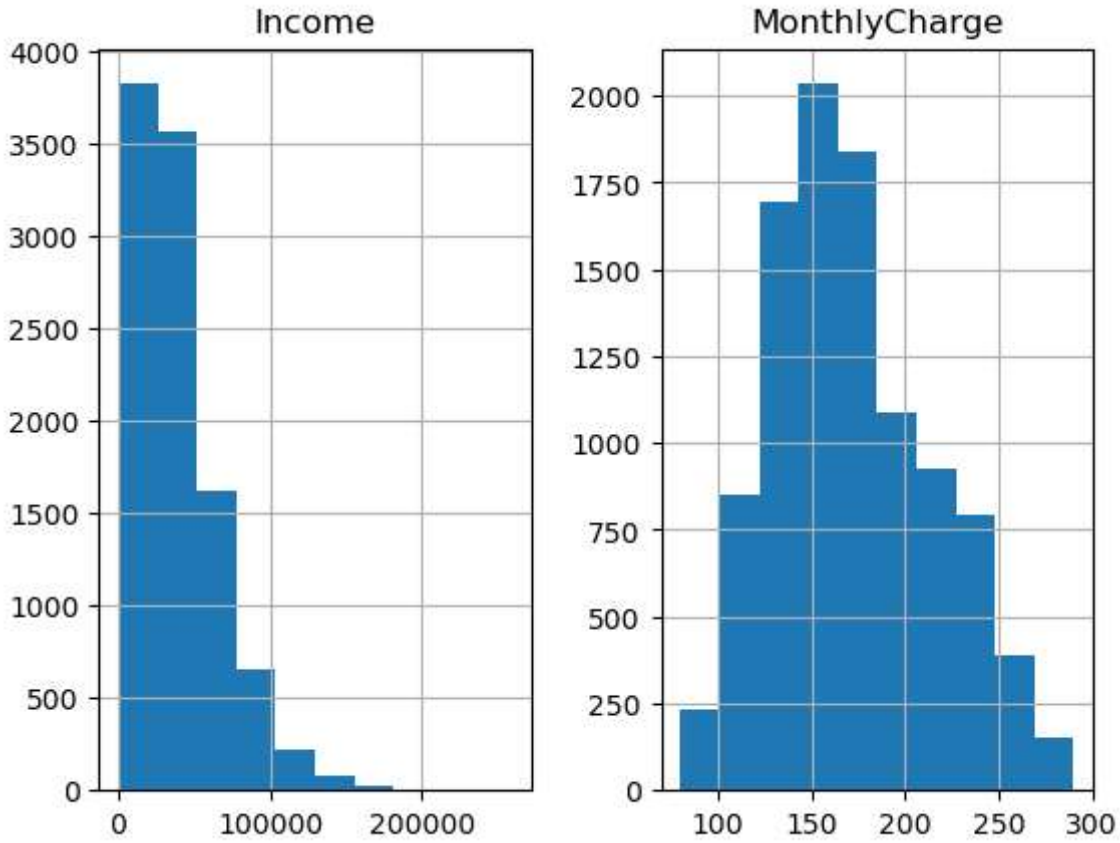
Area	Rural	Suburban	Urban
Churn			
No	2464	2473	2413
Yes	863	873	914

In [4]:
```
stats.chi2_contingency(chichurn)
```

Out[4]:
```
(2.4390738588073266,
 0.2953669109921032,
 2,
 array([[2445.345, 2459.31 , 2445.345],
        [ 881.655,  886.69 ,  881.655]]))
```

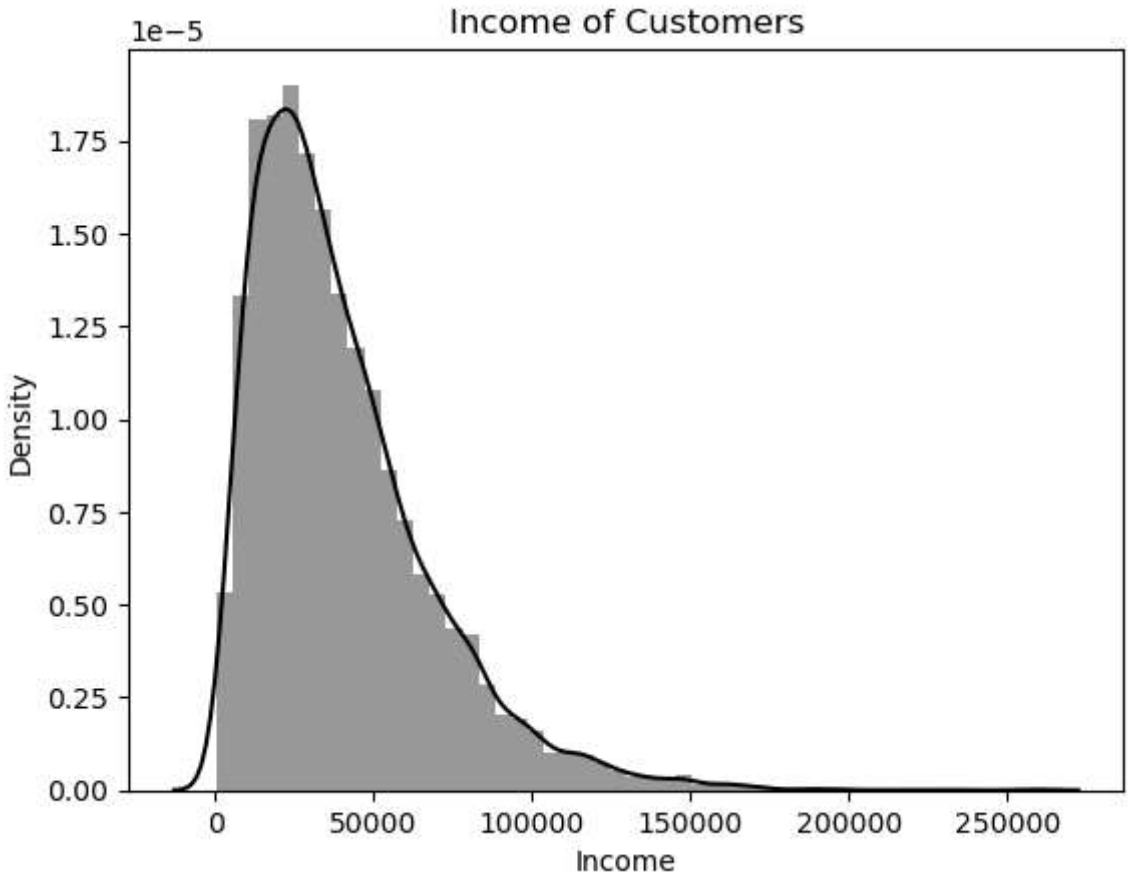
In [5]:
```
df[['Income', 'MonthlyCharge']].hist()
```

Out[5]:
```
array([[<AxesSubplot:title={'center':'Income'}>,
        <AxesSubplot:title={'center':'MonthlyCharge'}>]], dtype=object)
```

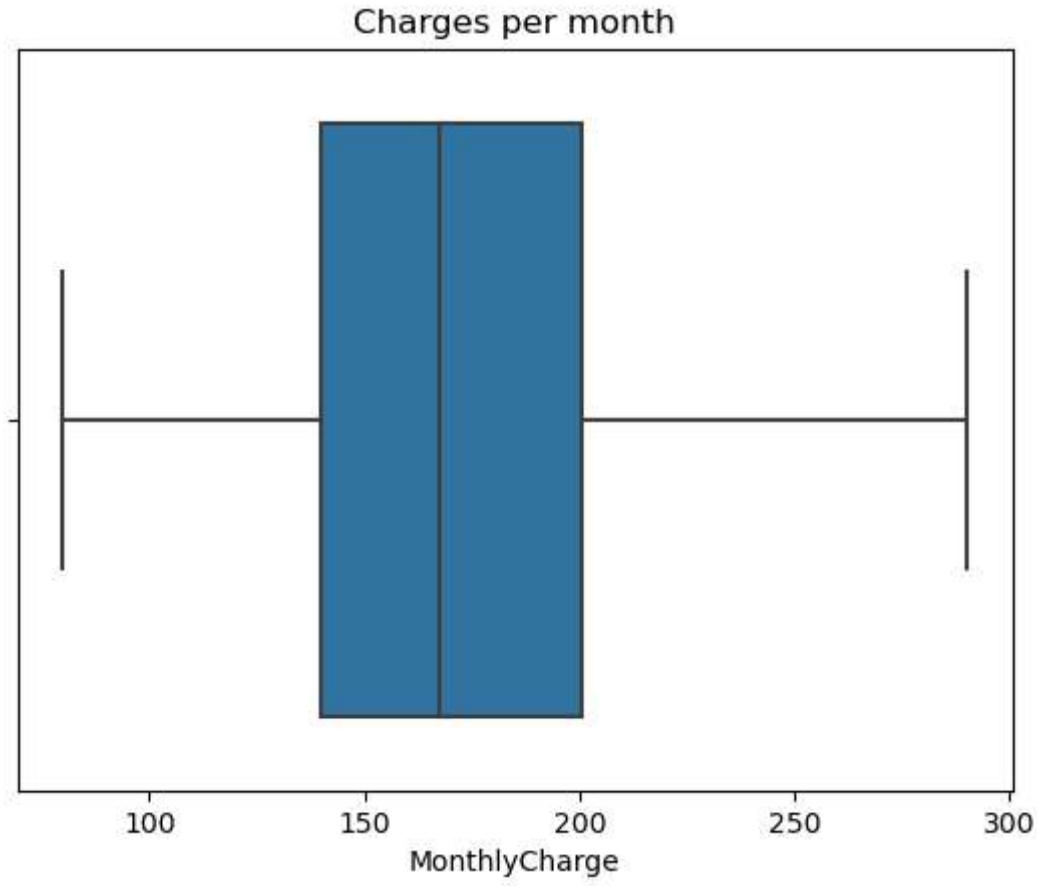


In [6]:
```
sns.distplot(df["Income"],color='black')
plt.title("Income of Customers")
```

Out[6]:
```
Text(0.5, 1.0, 'Income of Customers')
```

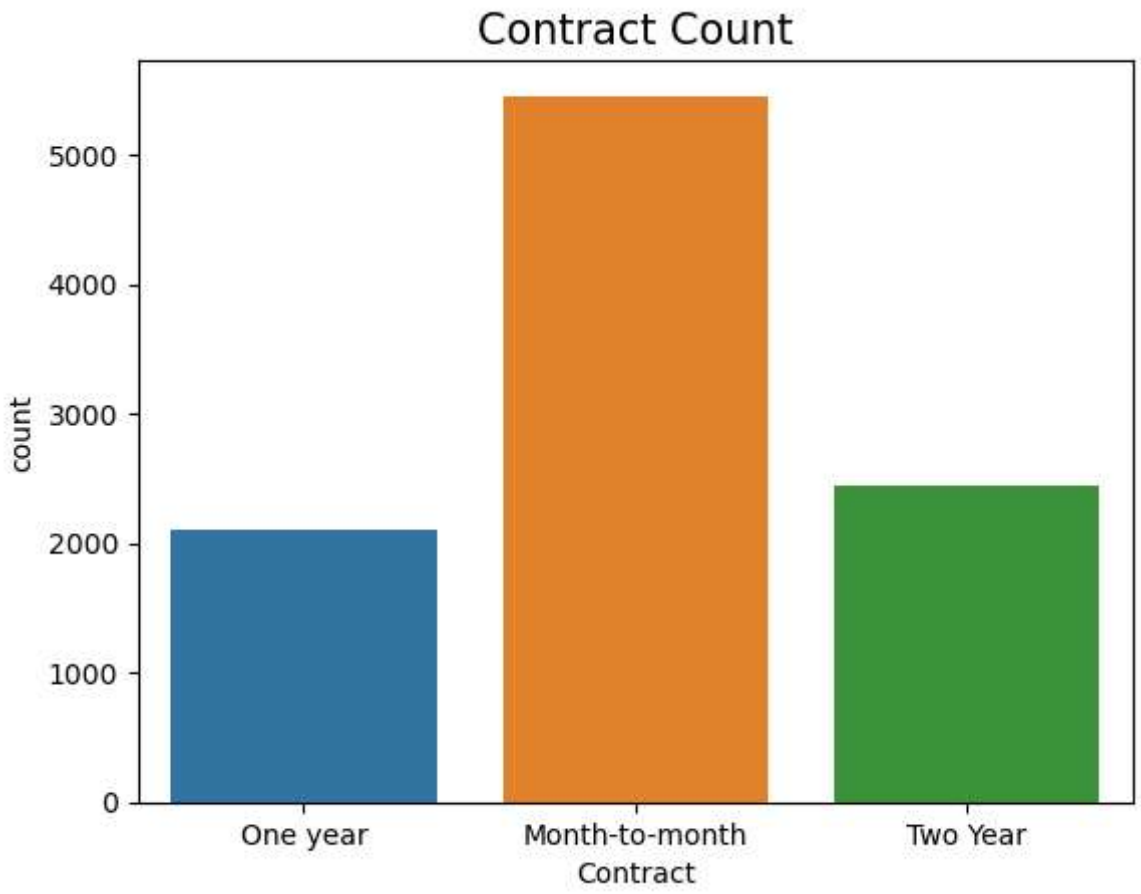


In [7]:
```
sns.boxplot('MonthlyCharge', data = df).set(title='Charges per month')
plt.show()
```



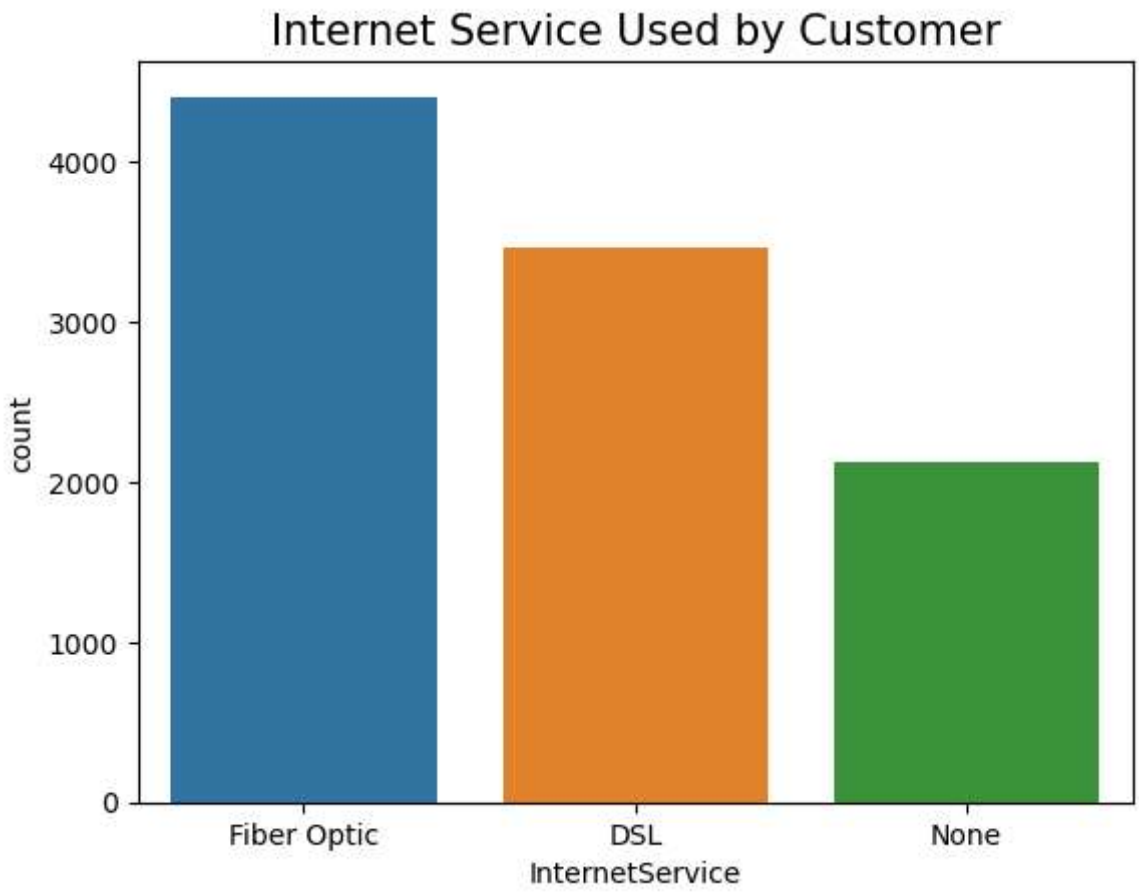
In [8]:
```
sns.countplot(df['Contract'])
plt.title("Contract Count", size=15)
```

Out[8]:
```
Text(0.5, 1.0, 'Contract Count')
```



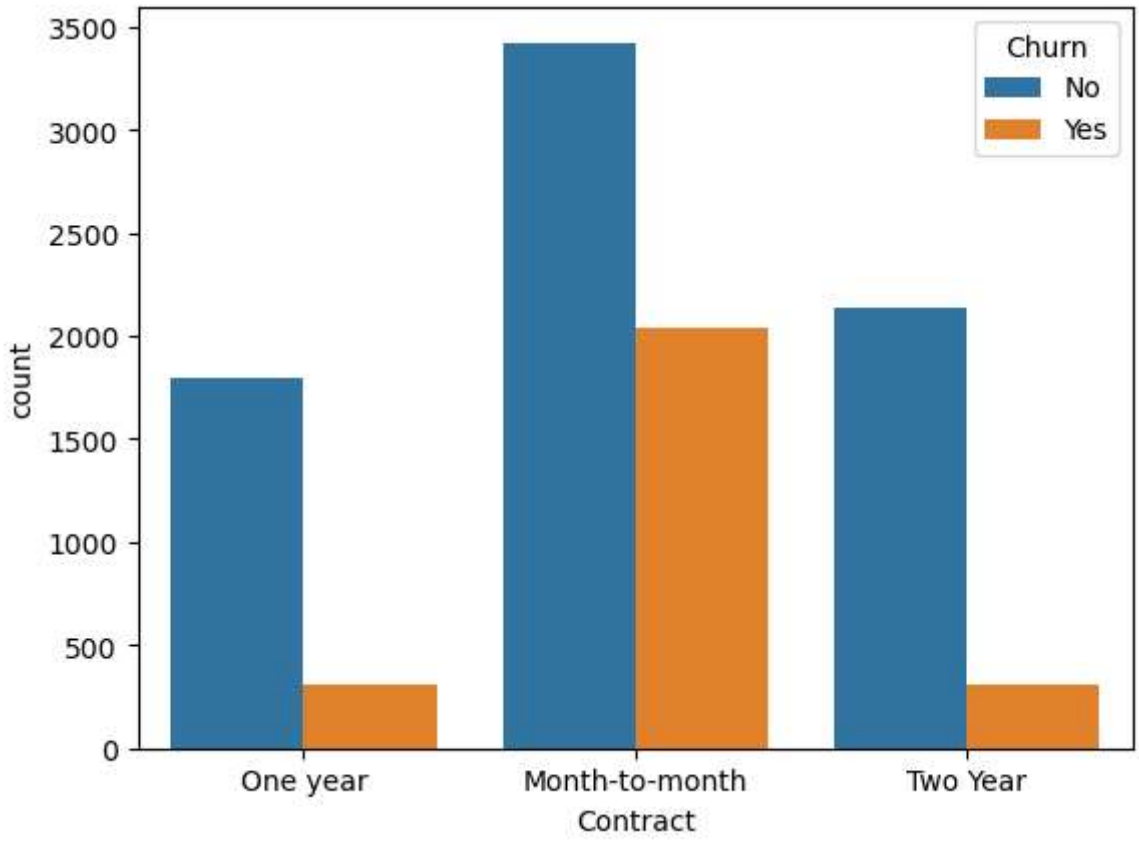
In [9]:
```
sns.countplot(df['InternetService'])
plt.title("Internet Service Used by Customer", size=15)
```

Out[9]:
```
Text(0.5, 1.0, 'Internet Service Used by Customer')
```



In [10]:
```
sns.countplot(x='Contract',hue='Churn',data=df)
```

Out[10]:
```
<AxesSubplot:xlabel='Contract', ylabel='count'>
```



In [11]:
```
sns.heatmap(df[['Age','Tenure',]].corr(), annot = True, fmt='.2g',cmap= 'coolwarm');
plt.show()
```