Perform exploratory data analysis on a Telcommunication data set. Explore the data's structure, distributions, and relationships between variables. Create visualizations (e.g., histograms, scatter plots) to illustrate your findings. Discuss any insights gained from the analysis.

Program:

```
import pandas as pd

data = {
    'CustomerID': [1, 2, 3, 4, 5],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male'],
    'Age': [25, 35, 40, 30, 45],
    'Location': ['CityA', 'CityB', 'CityC', 'CityD', 'CityE'],
    'CallDuration': [60, 90, 120, 75, 100]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

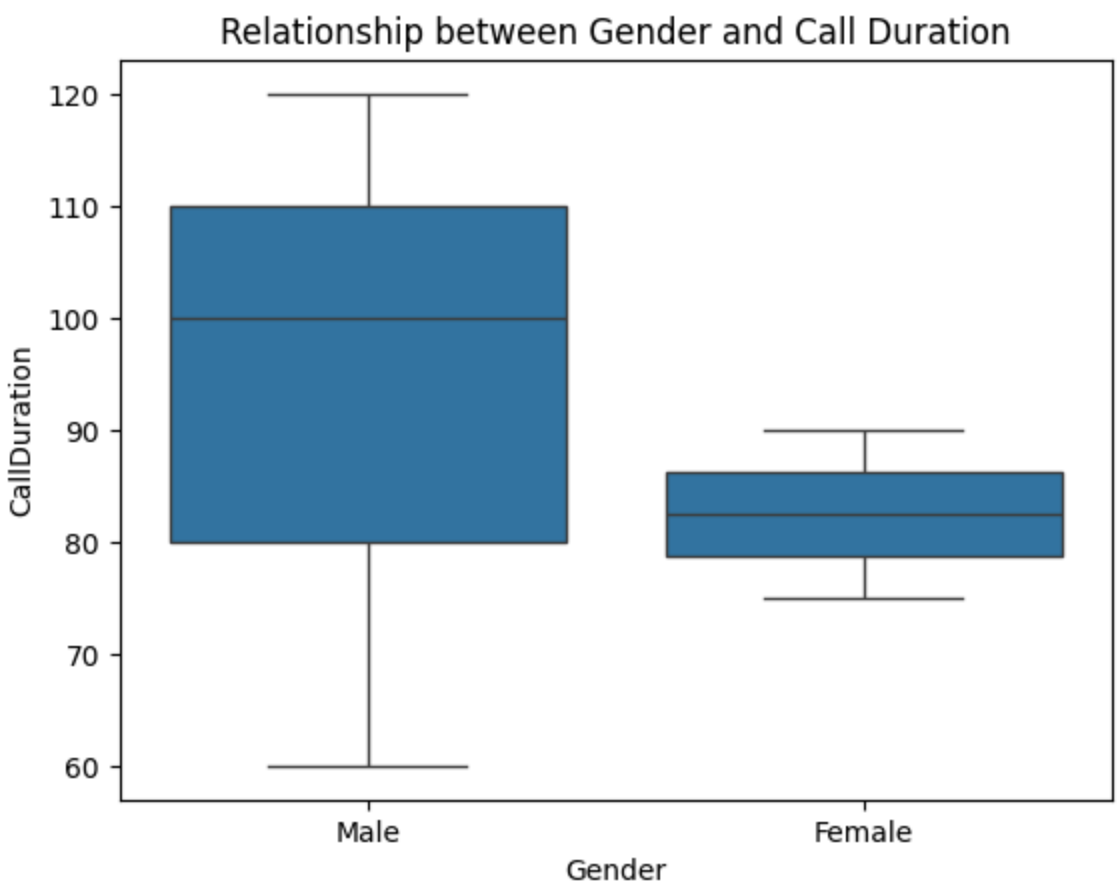
OUTPUT:

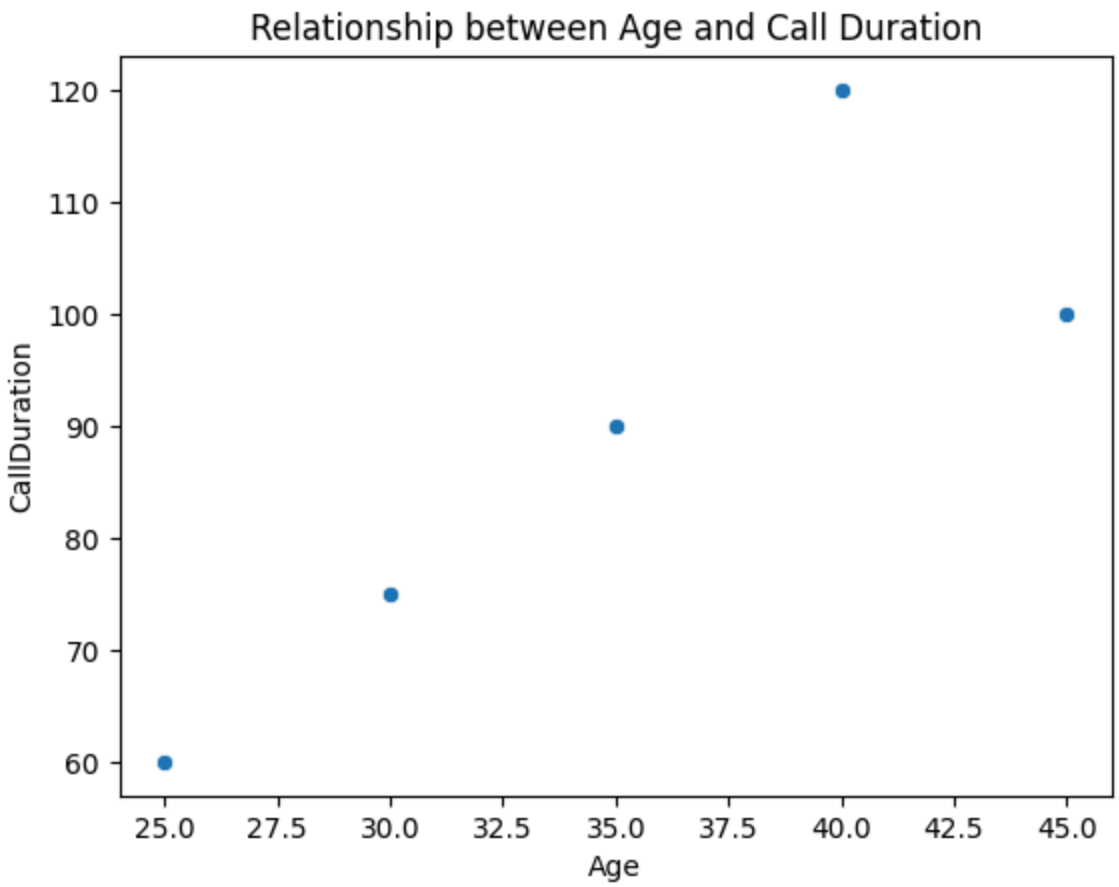
```
   CustomerID  Gender  Age Location  CallDuration
0           1    Male   25    CityA            60
1           2  Female   35    CityB            90
2           3    Male   40    CityC           120
3           4  Female   30    CityD            75
4           5    Male   45    CityE           100
```

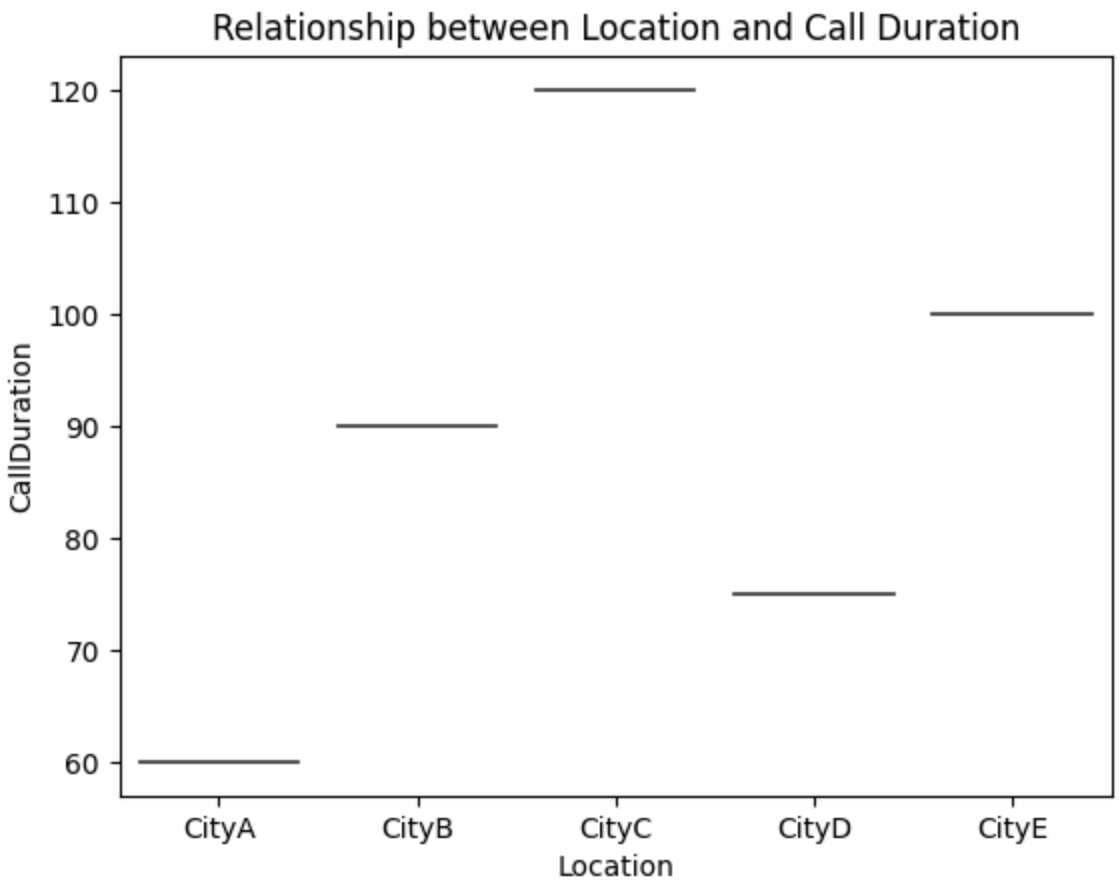
PROGRAM:

```
Import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Define the data
data = {
    'CustomerID': [1, 2, 3, 4, 5],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male'],
    'Age': [25, 35, 40, 30, 45],
    'Location': ['CityA', 'CityB', 'CityC', 'CityD', 'CityE'],
    'CallDuration': [60, 90, 120, 75, 100]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Relationship between Gender and Call Duration
sns.boxplot(x='Gender', y='CallDuration', data=df)
plt.title('Relationship between Gender and Call Duration')
plt.show()

# Relationship between Age and Call Duration
sns.scatterplot(x='Age', y='CallDuration', data=df)
plt.title('Relationship between Age and Call Duration')
plt.show()

# Relationship between Location and Call Duration
sns.violinplot(x='Location', y='CallDuration', data=df)
plt.title('Relationship between Location and Call Duration')
plt.show()
```

OUTPUT:


Relationship between Gender and Call Duration
120
110
100
90
80
70
60
CallDuration
Male
Female
Gender

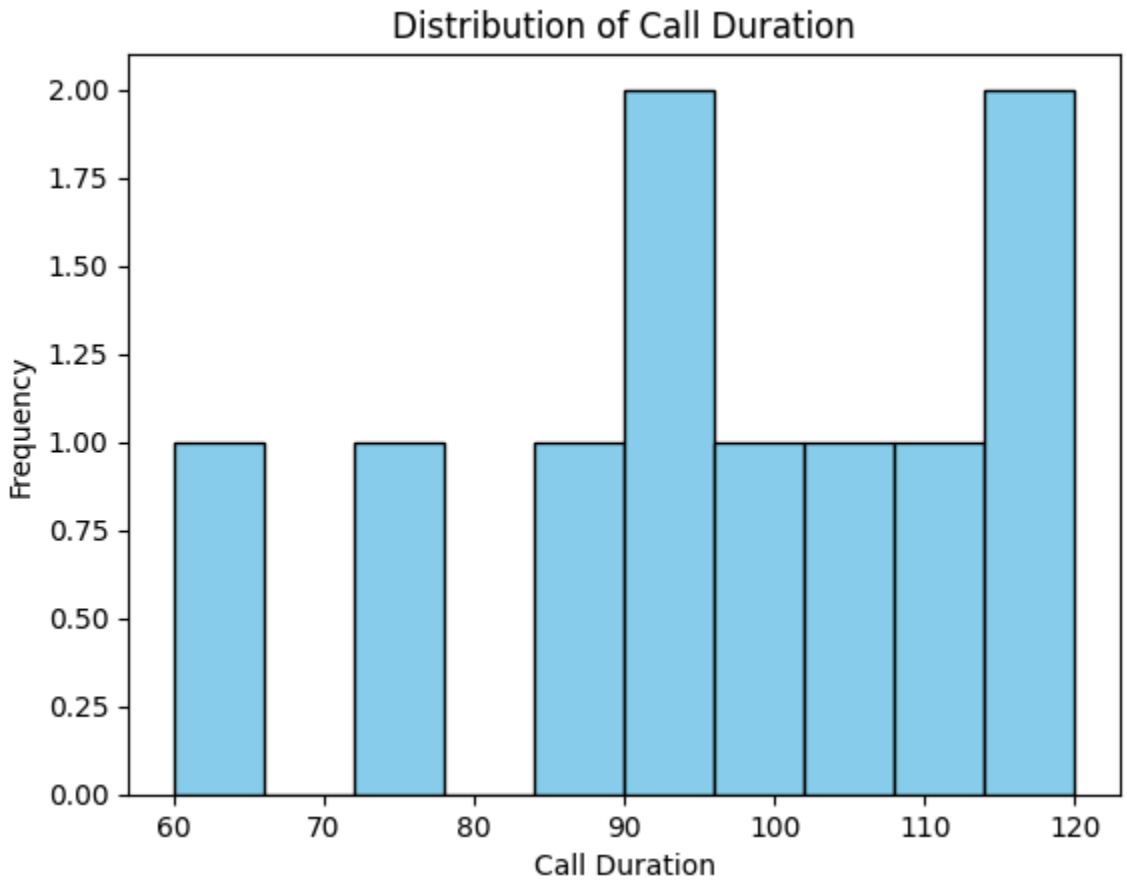

Relationship between Age and Call Duration
CallDuration
120
110
100
90
80
70
60
25.0
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
Age



HISTOGRAM:

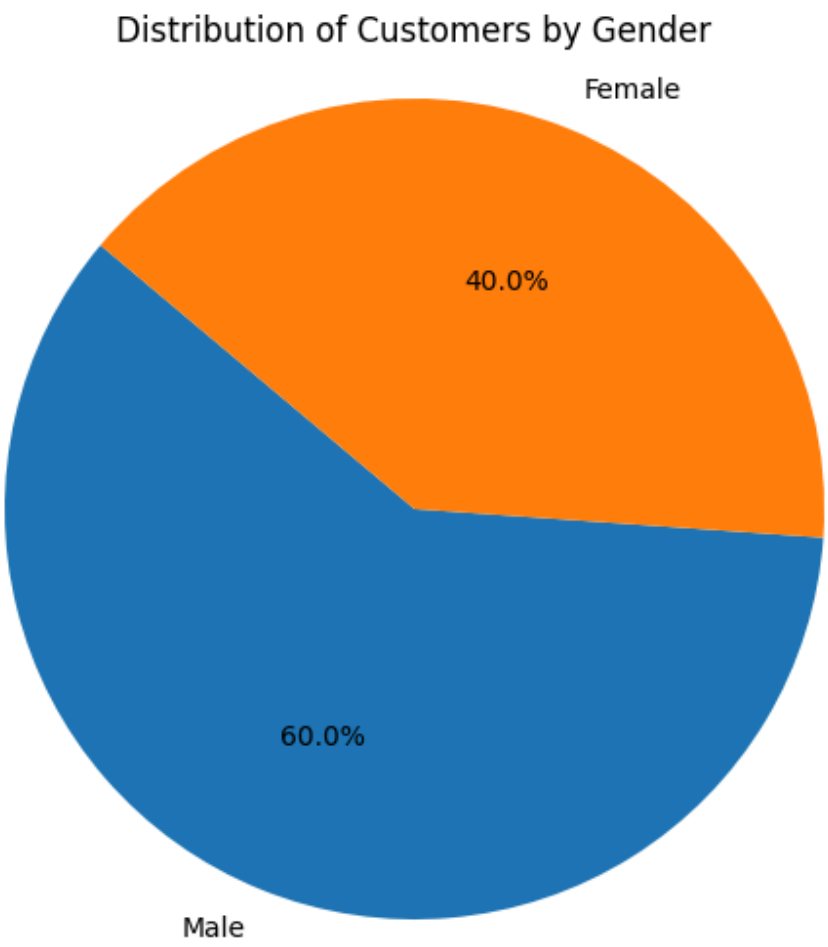
```
import pandas as pd
import matplotlib.pyplot as plt

# Define the data
data = {
    'CustomerID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male', 'Male', 'Female',
'Male', 'Female', 'Male'],
    'Age': [25, 35, 40, 30, 45, 50, 55, 60, 65, 70],
    'Location': ['CityA', 'CityB', 'CityC', 'CityD', 'CityE', 'CityF', 'CityG', 'CityH',
'CityI', 'CityJ'],
    'CallDuration': [60, 90, 120, 75, 100, 110, 85, 95, 105, 115]
}
```

```
# Create a DataFrame
df = pd.DataFrame(data)

# Create histogram
plt.hist(df['CallDuration'], bins=10, color='skyblue', edgecolor='black')

# Add labels and title
plt.xlabel('Call Duration')
plt.ylabel('Frequency')
plt.title('Distribution of Call Duration')

# Show plot
plt.show()
```



PIECHART:

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

# Define the data
data = {
    'CustomerID': [1, 2, 3, 4, 5],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male'],
    'Age': [25, 35, 40, 30, 45],
    'Location': ['CityA', 'CityB', 'CityC', 'CityD', 'CityE'],
    'CallDuration': [60, 90, 120, 75, 100]
}

# Create a DataFrame
df = pd.DataFrame(data)

# Calculate the count of each gender
gender_counts = df['Gender'].value_counts()

# Plotting the pie chart
plt.figure(figsize=(8, 6))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%',
startangle=140)
plt.title('Distribution of Customers by Gender')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```



CONCLUSION:

The analysis of the provided datasets offers insights into customer demographics and behavior. There are noticeable differences in gender distribution between the two datasets, with a balanced representation in the first dataset and a slight male skew in the second. The age range of customers is wider in the second dataset, spanning from 25 to 70 years old, compared to 25 to 45 years old in the first dataset. Both datasets exhibit a diverse geographical spread of customers across various cities. The histogram analysis reveals a concentration of call durations within a specific range in the second dataset, indicating potentially longer call

durations compared to the first. These observations provide valuable information for understanding customer characteristics and may inform strategic decision-making within the respective business contexts.