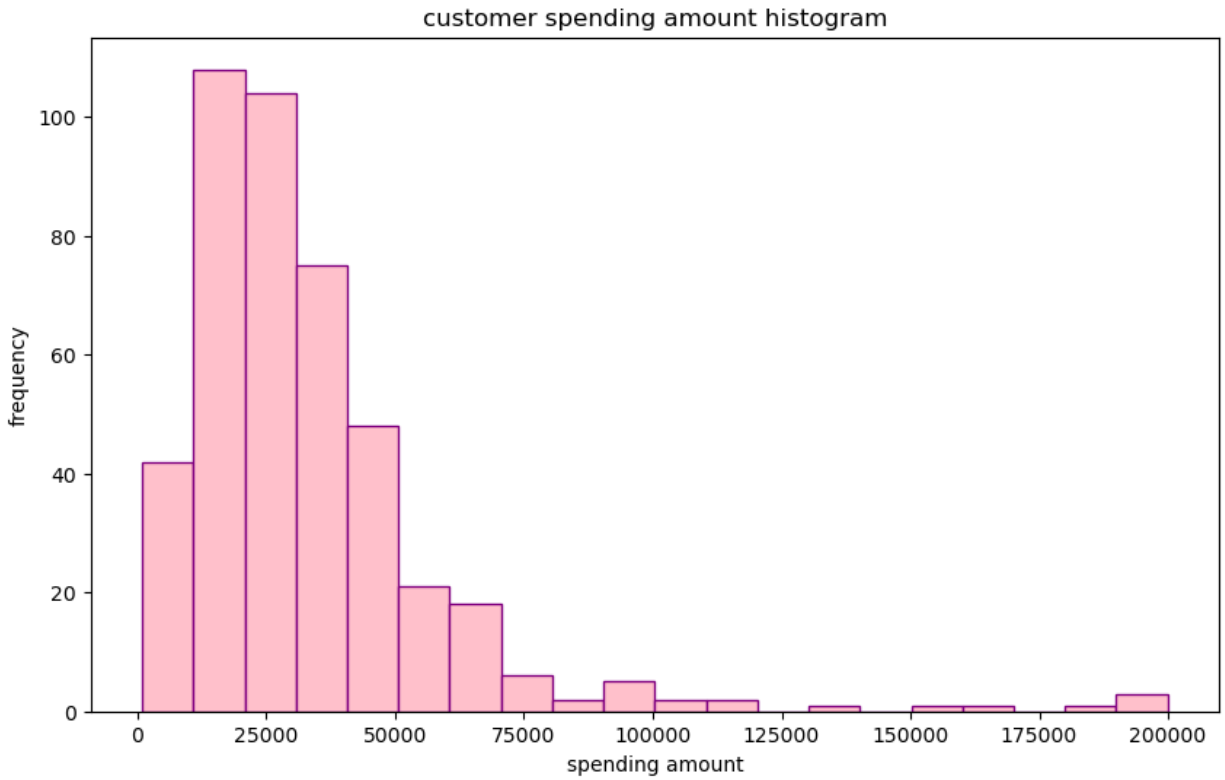
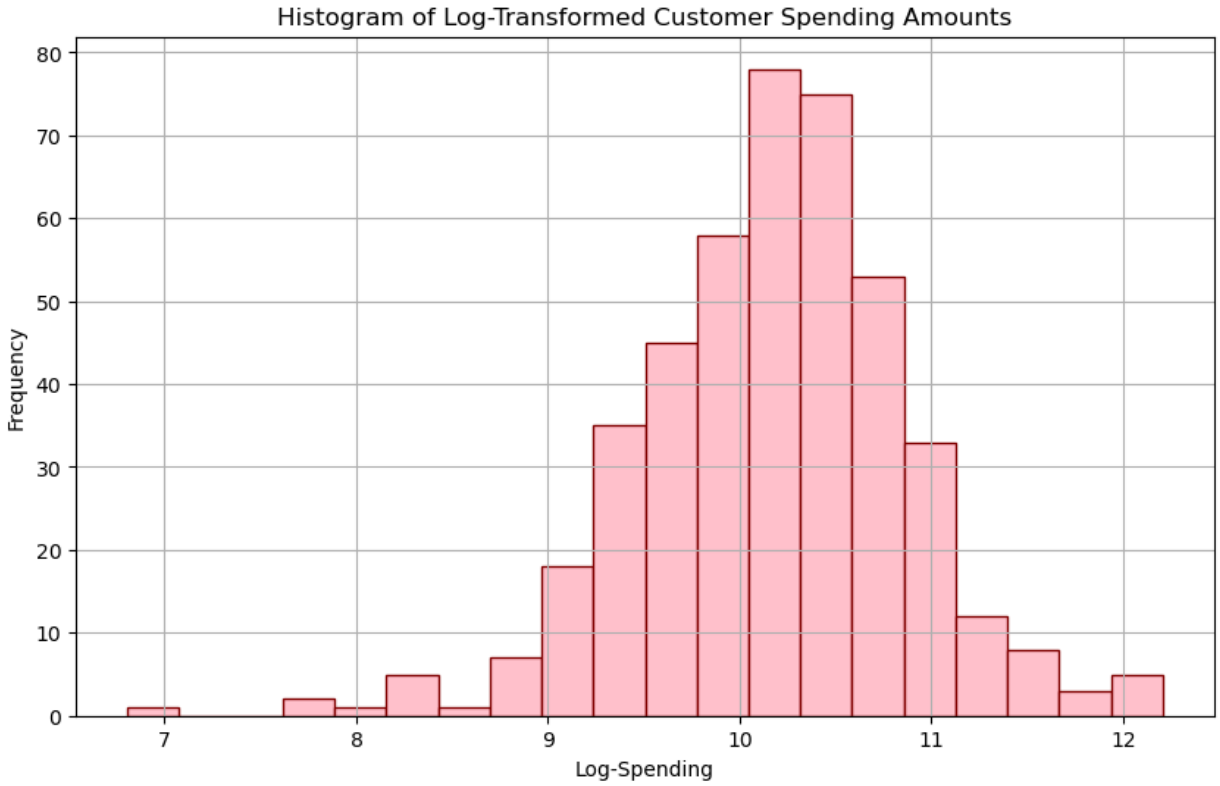
In [2]:
```
#For this problem, you will be using the data set customer spending.csv. This data set
#cusstomer spending.csv

#Make a histogram of the customer spending amounts

#load pd, plt
import pandas as pd
import matplotlib.pyplot as plt

customer_data= pd.read_csv('cusstomer spending.csv')

#ann_spendinig
plt.figure(figsize=(10, 6))
plt.hist(customer_data['ann_spending'], bins=20, color='pink', edgecolor='purple')

#plt table
plt.xlabel('spending amount')
plt.ylabel('frequency')
plt.title('customer spending amount histogram')
plt.show()
```



In [6]:
```
#Make a new data set that is a log transformation of the customer spending amounts.
#Load np, pd
import pandas as pd
import numpy as np

#cusstomer spending.csv
customer_data = pd.read_csv("cusstomer spending.csv")

#column with log transformed spending amounts
#ann_spending
customer_data['Log_Spending'] = np.log(customer_data['ann_spending'])
customer_data.to_csv("customer_spending_log_transformed.csv", index=False)
```

In [13]:
```
#Make a histogram of the log transformed dataset.

#load pd, np, plt
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#log pull
#customer_spending_log_transformed.csv
log_transformed_data = pd.read_csv("customer_spending_log_transformed.csv")

#hitogram plot
plt.figure(figsize=(10, 6))
plt.hist(log_transformed_data['Log_Spending'], bins=20, color='pink', edgecolor='maroo
plt.title('Histogram of Log-Transformed Customer Spending Amounts')
plt.xlabel('Log-Spending')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



Compare the two histograms. Discuss why it might be useful to apply a log transformation tothis data for modeling purposes.

The histogram for the original customer spending amounts may showed a skewed distibution which is common in financial data. In some cases, the data may show a long tail, meaning there are few customers with high spending amounts, and using a histogram may make it challenging to model certain statistical methods that require symmetric distributions. The histogram of the log transformed customer spending amounts usually show the more symmetric distribution, by compressing the range of values, that gives it the more balanced distribution appearance.

In [ ]:

In [ ]:

In [ ]:

In [ ]: