Task 1

Using the Pandas library, load a CSV file and perform basic data analysis tasks, such as calculating the average of a selected column. Additionally, use Matplotlib to create visualizations, including bar charts, scatter plots, and heatmaps, to analyze the data. Provide insights and observations based on the analysis and visualizations.

Colab Link:

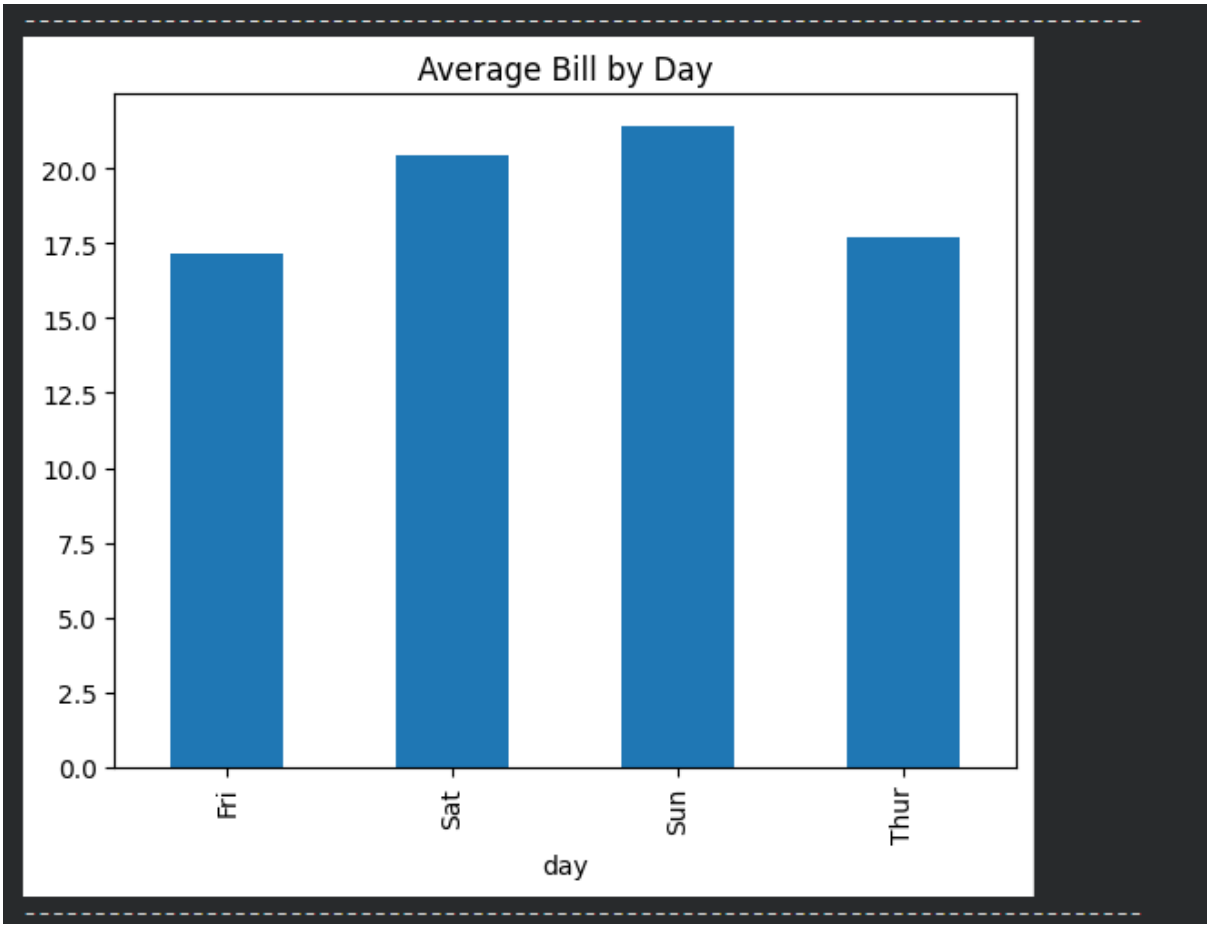
https://colab.research.google.com/drive/1jSWFFsLOARrgN4TvCciu-3Qbdc7ZCGQs?usp=sharing

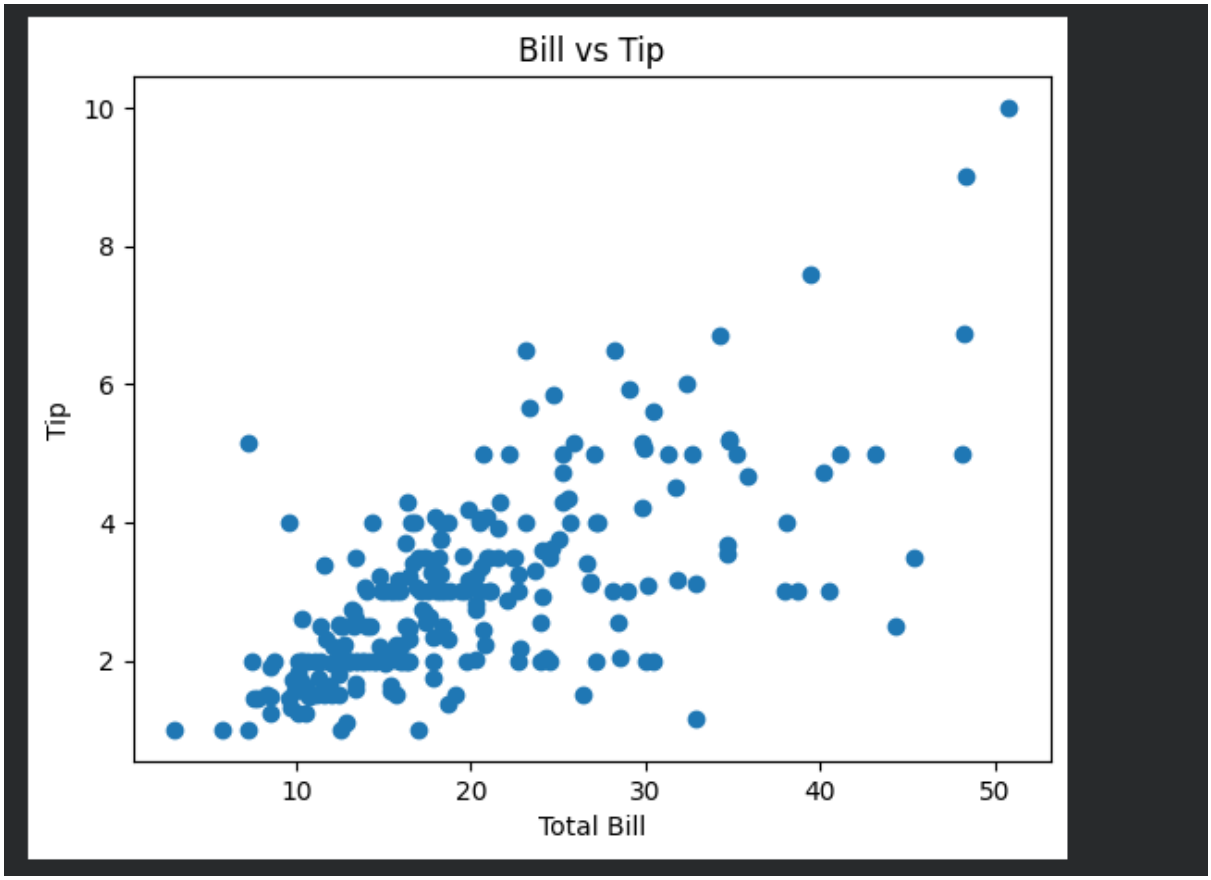
Code:

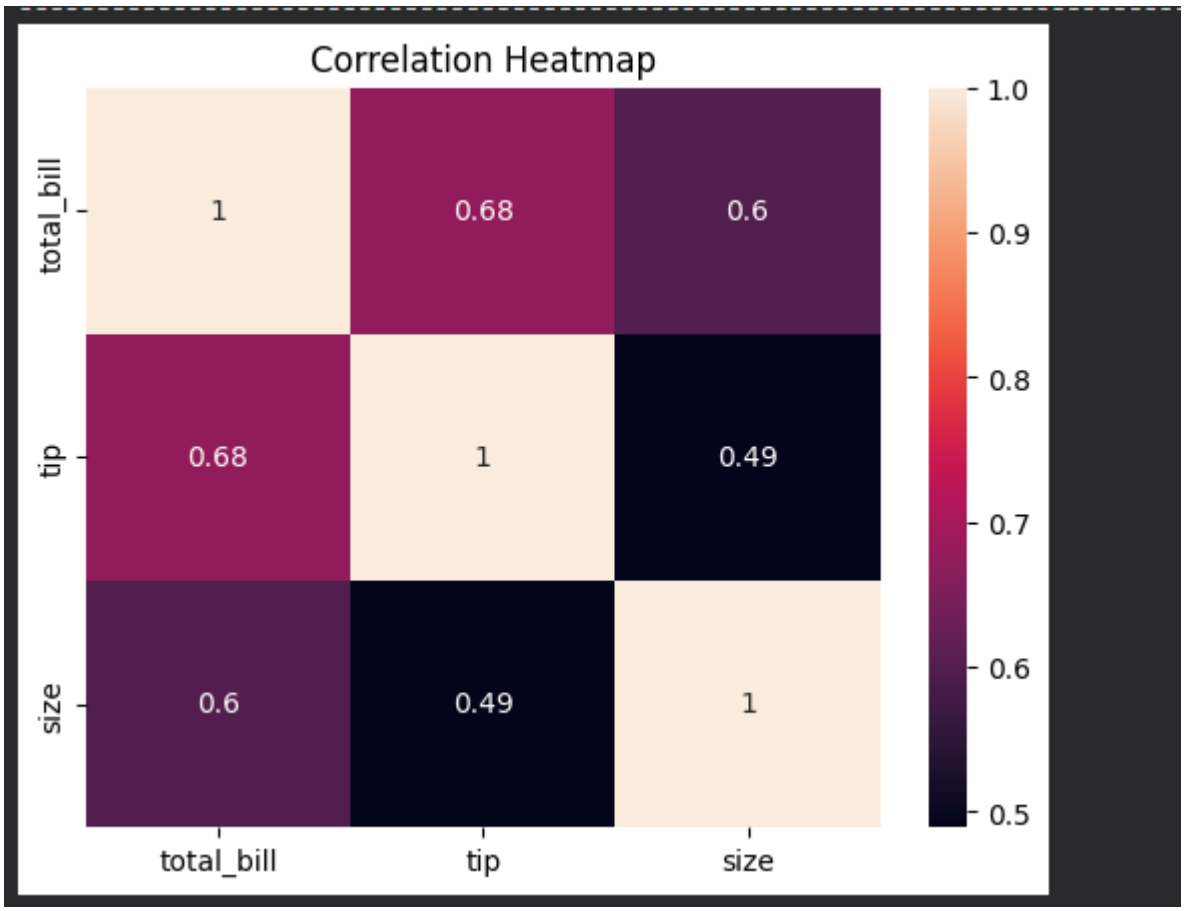
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_excel("/content/Task1_DataAnalysis-data.xlsx")
print('-' * 80)
print(df.head())
print("\nAverage total bill:", df["total_bill"].mean())

print('-' * 80)

# Bar chart
df.groupby("day")["total_bill"].mean().plot(kind="bar")
plt.title("Average Bill by Day")
plt.show()

print('-' * 80)

# Scatter plot
plt.scatter(df["total_bill"], df["tip"])
plt.xlabel("Total Bill")
plt.ylabel("Tip")
plt.title("Bill vs Tip")
plt.show()

print('-' * 80)

# Heatmap
sns.heatmap(df.corr(numeric_only=True), annot=True)
plt.title("Correlation Heatmap")
plt.show()
```

Output:

```
   total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4

Average total bill: 19.78594262295082
```




Bill vs Tip
Tip
Total Bill
10
8
6
4
2
10
20
30
40
50


Correlation Heatmap
total_bill
tip
size
1
0.68
0.6
0.68
1
0.49
0.6
0.49
1
1.0
0.9
0.8
0.7
0.6
0.5