|  |  |
| --- | --- |
| **ITMD 513: Assignment 7** | |
| **Name: Sagar Ippili** | **CWID: A20417999** |

**Create Data Visualizations using Python Pandas and SeaBorn**

1. Source Code:

'''

ITMD 513: HW7 - Sagar Ippili

Summary: This program performs the data visualization using Seaborn

'''

#Import all the necessary Pandas libraries

import pathlib

import pandas as pd

import seaborn as sns; sns.set()

import matplotlib.pyplot as plt

#Function to create a scatterplot

def createScatterPlotA(df):

    #Plotting scatterplot with seaborn

    ax = sns.scatterplot(x='total\_bill', y='tip', data=df)

    #Setting a title for the plot

    plt.title('Scatter Plot: Total Bill Vs Tip')

    #Display the scatterplot

    plt.show()

def createScatterPlotB(df):

    #Plotting scatterplot with seaborn

    ax = sns.scatterplot(x='total\_bill', y='tip', hue='smoker', style='smoker', size='size', sizes=(10, 300), legend='full', data=df)

    #Setting a title for the plot

    plt.title('Scatter Plot: Total Bill Vs Tip Amongs Smokers')

    #Display the scatterplot

    plt.show()

def createBarPlot(df):

    #Plotting barplot with seaborn

    ax = sns.barplot(x='day', y='tip', data=df, ci=68)

    #Setting a title for the plot

    plt.title('Bar Plot: Daily Average Tip')

    #Display the scatterplot

    plt.show()

def createScatterPlotD(df):

    #Plotting scatterplot with seaborn

    ax = sns.scatterplot(x='total\_bill', y='tip', hue='time', style='time', size='size', sizes=(10, 300), legend='full', data=df)

    #Setting a title for the plot

    plt.title('Scatter Plot: Total Bill Vs Tip During Lunch & Dinner Time')

    #Display the scatterplot

    plt.show()

def createLinePlot(df):

    #Plotting scatterplot with seaborn

    ax = sns.lineplot(x='year', y='passengers', hue='month', estimator='mean', data=df)

    #Setting a title for the plot

    plt.title('Line Plot: Average Num of Passengers Per Year Per Month')

    #Display the scatterplot

    plt.show()

def createBarsPlotTitanic(df):

    #Plotting barplot with seaborn

    #countOdfObs = df['PassengerId'].sum()

    ax = sns.catplot(x="Sex", hue="Pclass", col="Survived", data=df, kind="count", height=4, aspect=.7)

    #Setting a title for the plot

    #plt.title('Bar Plot: Counts of Observation in each Categorical Bin', y=1.08)

    #plt.text(0.5, 1.08, 'Bar Plot: Counts of Observation in each Categorical Bin', horizontalalignment='center', fontsize=10, transform = ax.axes)

    #Display the scatterplot

    plt.show()

#Main function where the program starts execution

def main():

    #Files' paths and folder information

    fileDirectory = pathlib.Path(\_\_file\_\_).parent.absolute()

    workerstipsFile = str(fileDirectory)+'\\workerstips.csv'

    flightsDataFile = str(fileDirectory)+'\\flightsData.csv'

    titanicFile = str(fileDirectory)+'\\titanic.csv'

    #Create Data frames

    dfWorkersTips = pd.read\_csv(workerstipsFile)

    dfFlightsData = pd.read\_csv(flightsDataFile)

    dfTitanic = pd.read\_csv(titanicFile)

    #a) Create a scatterplot using workerstips.csv file

    createScatterPlotA(dfWorkersTips)

    #b) Create a scatterplot to differentiate between Smokers and Non Smokers

    createScatterPlotB(dfWorkersTips)

    #c) Create a bar plot

    createBarPlot(dfWorkersTips)

    #d) Create a scatterplot to differentiate between lunch and dinner time

    createScatterPlotD(dfWorkersTips)

    #e) Create a lineplot between avg number of passengers per year per month

    createLinePlot(dfFlightsData)

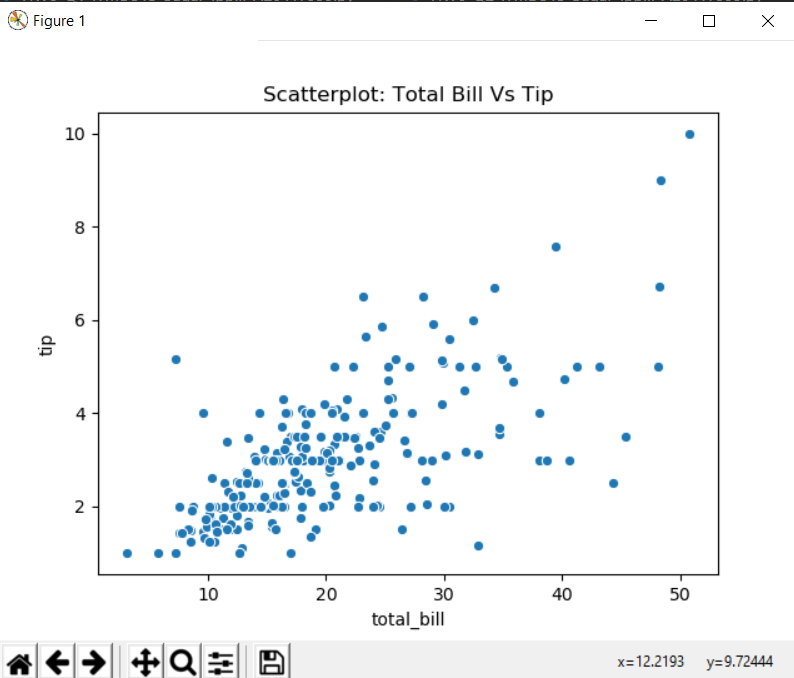
    #f) Create barsplot for titanic categorical counts

    createBarsPlotTitanic(dfTitanic)

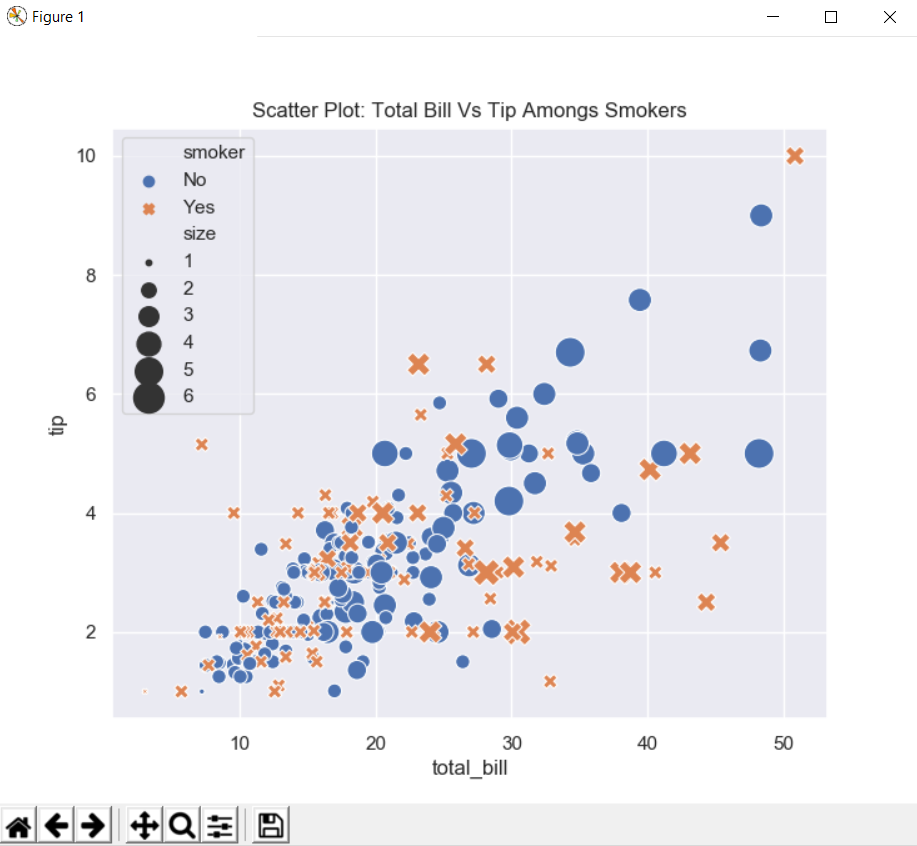
#Call the main() function

main()

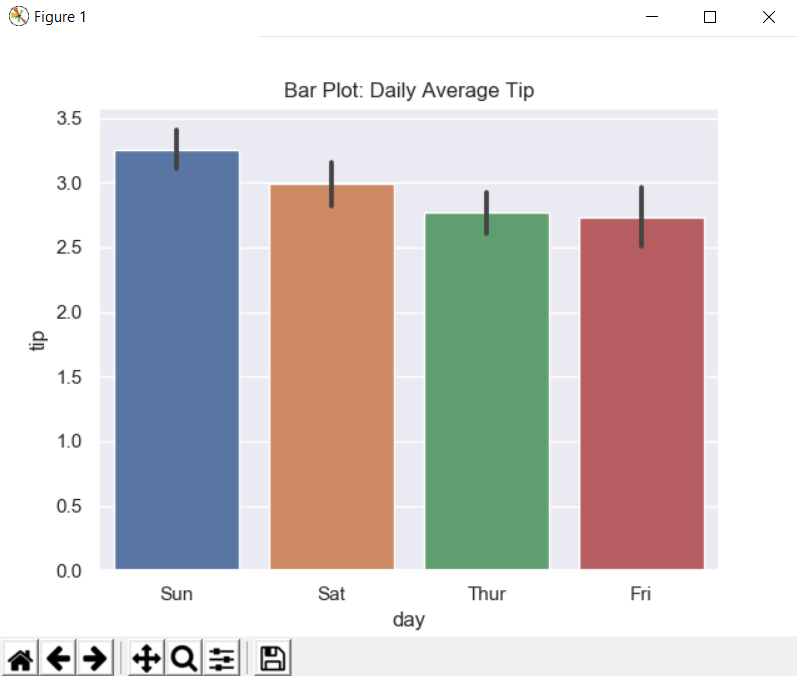
1. Output(s):
   1. ScatterPlot: Total Bill Vs Tips:



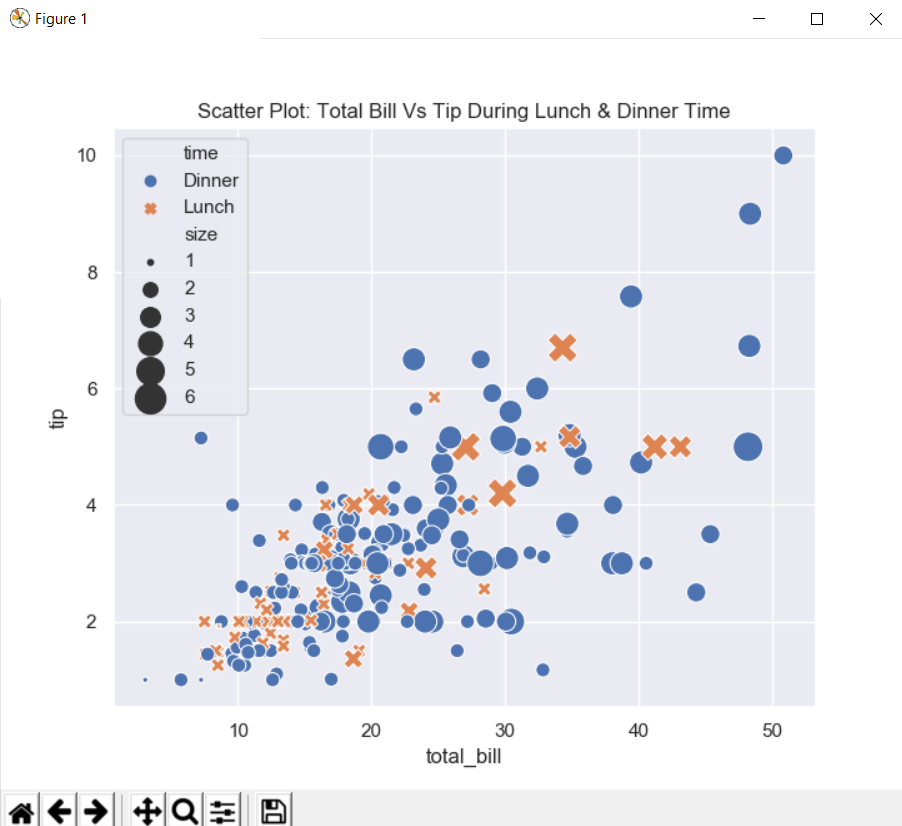
* 1. ScatterPlot to differentiate between smokers and non smokers:



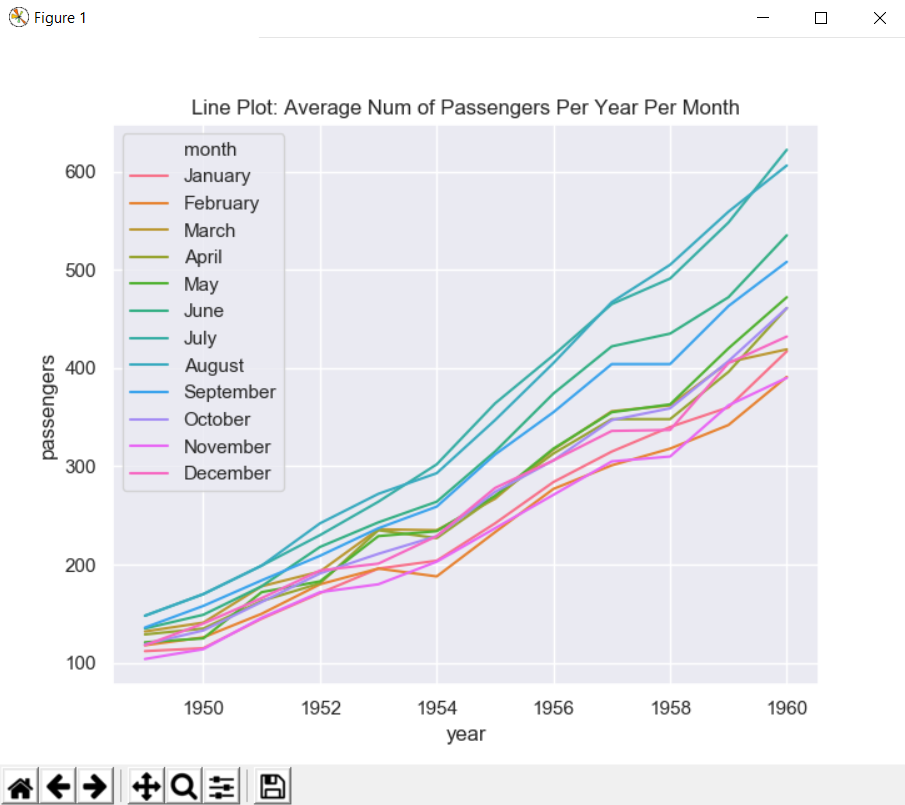
* 1. BarPlot to show average tips per day of the week:



* 1. ScatterPlot to show average tips per day of the week over the time:



* 1. LinePlot to show average number of passengers over time:



* 1. BarPlot to show counts of observations in each categorical bin:

