Data Analysis Challenge 05 - Tip Analysis

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams["figure.facecolor"] = "white"

url = "https://raw.githubusercontent.com/guipsamora/pandas\_exercises/master/07\_Visualization/Tips/tips.csv"
df = pd.read\_csv(url)

df.head(10)

|  | Unnamed: 0 | total\_bill | tip | sex | smoker | day | time | size |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 16.99 | 1.01 | Female | No | Sun | Dinner | 2 |
| 1 | 1 | 10.34 | 1.66 | Male | No | Sun | Dinner | 3 |
| 2 | 2 | 21.01 | 3.50 | Male | No | Sun | Dinner | 3 |
| 3 | 3 | 23.68 | 3.31 | Male | No | Sun | Dinner | 2 |
| 4 | 4 | 24.59 | 3.61 | Female | No | Sun | Dinner | 4 |
| 5 | 5 | 25.29 | 4.71 | Male | No | Sun | Dinner | 4 |
| 6 | 6 | 8.77 | 2.00 | Male | No | Sun | Dinner | 2 |
| 7 | 7 | 26.88 | 3.12 | Male | No | Sun | Dinner | 4 |
| 8 | 8 | 15.04 | 1.96 | Male | No | Sun | Dinner | 2 |
| 9 | 9 | 14.78 | 3.23 | Male | No | Sun | Dinner | 2 |

### 1. Delete the Unnamed 0 column

df = df.drop(columns=["Unnamed: 0"])

### 2. Plot the total\_bill column histogram

total\_bill\_values = df["total\_bill"]
fig, ax = plt.subplots(nrows=1,ncols=1,figsize=(10,5))
ax.hist(total\_bill\_values, bins = 10, density=False)
ax.set\_title("Histogram of total bills")
ax.set\_xlabel("Total bill")
ax.set\_ylabel("Frequency")
plt.show()

|  |
| --- |
| Figure 1: A histogram with frequency of total bill values |

### 3. Create a scatter plot presenting the relationship between total\_bill and tip

tips = df["tip"]
fig, ax = plt.subplots(nrows=1,ncols=1,figsize=(10,5))
ax.scatter(x=total\_bill\_values,y=tips)
ax.set\_title("Relationship between total bills and tips")
ax.set\_xlabel("Total bill")
ax.set\_ylabel("Tips")
plt.show()

|  |
| --- |
| Figure 2: A scatter plot to display relationship between total bill values and tips |

### 4. Create one image with the relationship of total\_bill, tip and size.

fig, ax = plt.subplots(nrows=1,ncols=1,figsize=(10,5))
ax.scatter(x=total\_bill\_values,y=tips, c=df["size"])
ax.set\_title("Relation ship bewteen total bill, tips and size")
ax.set\_xlabel("Total bill")
ax.set\_ylabel("Tips")

plt.show()

|  |
| --- |
| Figure 3: A scatter plot to display relationship between total bill, tips adn size |

### 5. Present the relationship between days and total\_bill value.

sunday\_data = np.array(df[df["day"] == "Sun"]["total\_bill"])
saturday\_data = np.array(df[df["day"] == "Sat"]["total\_bill"])
Thursday\_data = np.array(df[df["day"] == "Thur"]["total\_bill"])
Friday\_data = np.array(df[df["day"] == "Fri"]["total\_bill"])
total\_bill\_days = [sunday\_data, saturday\_data, Thursday\_data, Friday\_data]
labels = df["day"].unique()
fig, ax = plt.subplots(nrows=1,ncols=1,figsize=(10,5))
ax.boxplot(x=total\_bill\_days,labels=labels)
ax.set\_title("Relationship between days and total bill")
ax.set\_xlabel("Days")
ax.set\_ylabel("Total bill")
plt.show()

|  |
| --- |
| Figure 4: Displaying relationship between total bill and days |

### 6. Create a scatter plot with the day as the y-axis and tip as the x-axis, differ the dots by sex

fig, ax = plt.subplots(nrows=1,ncols=1,figsize=(10,5))
ax.scatter(y=df["day"],x=df["tip"])
ax.set\_title("Tips on given day")
ax.set\_xlabel("Tips")
ax.set\_ylabel("Days")
plt.show()

|  |
| --- |
| Figure 5: Scatter plot with tips and days |

### 7. Create a box plot presenting the total\_bill per day differentiation the time (Dinner or Lunch)

#Prepare data for box plot
dinner\_data = np.array(df[df["time"] == "Dinner"]["total\_bill"])
lunch\_data = np.array(df[df["time"] == "Lunch"]["total\_bill"])
total\_bill\_time = [dinner\_data, lunch\_data]
time\_labels = df["time"].unique()

fig, ax = plt.subplots(nrows=1,ncols=1, figsize=(10,5))
ax.boxplot(x=total\_bill\_time,labels=time\_labels)
ax.set\_title("Relation ship between time and total bill")
ax.set\_xlabel("Time")
ax.set\_ylabel("Total bill")
plt.show()

|  |
| --- |
| Figure 6: Displaying relationship between total bill and time |

### 8. Create two histograms of the tip value based for Dinner and Lunch. They must be side by side

dinner\_tip\_data = np.array(df[df["time"] == "Dinner"]["tip"])
lunch\_tip\_data = np.array(df[df["time"] == "Lunch"]["tip"])
total\_bill\_time = [dinner\_data, lunch\_data]

fig, ax = plt.subplots(nrows=1,ncols=2, figsize=(15, 5))

ax\_lunch = ax[0].hist(x=lunch\_data, bins=10)
ax[0].set\_title("Tips during lunch")
ax[0].set\_xlabel("Tips")
ax[0].set\_ylabel("Frequency")

ax\_dinner = ax[1].hist(x=dinner\_data,bins=10)
ax[1].set\_title("Tips during dinner")
ax[1].set\_xlabel("Tips")
ax[1].set\_ylabel("Frequency")
plt.show()

|  |
| --- |
| Figure 7: Displaying values of tips during lunch and dinner |

### 9. Create two scatterplots graphs, one for Male and another for Female, presenting the total\_bill value and tip relationship, differing by smoker or no smoker

#Prepare data for scatter plot
male\_bill\_tip = df[df["sex"]=="Male"]
female\_bill\_tip = df[df["sex"]=="Female"]
labels\_sex = df["sex"].unique()

fig, (ax\_male, ax\_female) = plt.subplots(nrows=1, ncols=2, figsize=(15, 5))

# Male scatter plot with trendline and R value
x\_male = male\_bill\_tip["total\_bill"]
y\_male = male\_bill\_tip["tip"]
ax\_male.scatter(x\_male, y\_male)
ax\_male.set\_title("Relationship between total bill and tips in males")
ax\_male.set\_xlabel("Total bill")
ax\_male.set\_ylabel("Tips")
z\_male = np.polyfit(x\_male, y\_male, 1)
p\_male = np.poly1d(z\_male)
ax\_male.plot(x\_male, p\_male(x\_male), "r-")
r\_male = np.corrcoef(x\_male, y\_male)[0, 1]
ax\_male.text(0.05, 0.95, f"R = {r\_male:.2f}", transform=ax\_male.transAxes, ha='left', va='top')

# Female scatter plot with trendline and R value
x\_female = female\_bill\_tip["total\_bill"]
y\_female = female\_bill\_tip["tip"]
ax\_female.scatter(x\_female, y\_female)
ax\_female.set\_title("Relationship between total bill and tips in females")
ax\_female.set\_xlabel("Total bill")
ax\_female.set\_ylabel("Tips")
z\_female = np.polyfit(x\_female, y\_female, 1)
p\_female = np.poly1d(z\_female)
ax\_female.plot(x\_female, p\_female(x\_female), "r-")
r\_female = np.corrcoef(x\_female, y\_female)[0, 1]
ax\_female.text(0.05, 0.95, f"R = {r\_female:.2f}", transform=ax\_female.transAxes, ha='left', va='top')

plt.show()

|  |
| --- |
| Figure 8: Exploring correlation between total bills and tips based on sex |