Practice Project - Matplotlib based : CSV based Data Visualization

Table of Contents


```
Problem statement
1. Introduction
2. Algorithm
3. Python implementation
4. Improvements
```


Problem Statement:

```
Problem using Polynomial Regression
```

- Activity: Students are provided with this Dataset 'students.csv'
  ```
  In this data set you need to visualize whether we have more male or female students in school.<br>
  ```
- Activity: Get Correlation Matrix for 'students.csv' data.

Introduction

1. Data Visualization is the process of presenting data in the form of graphs or charts.
2. It helps to understand large and complex amounts of data very easily.
3. It allows us to identify new trends and patterns very easily.

Matploptib is a low-level library of Python which is used for data visualization, This library is built on the top of NumPy arrays and consist of several plots like line chart, bar chart, histogram, etc.These various plots we can be visualized using Pyplot

Importing Matplotlib

In [ ]:
```
import matplotlib.pyplot as plt
```

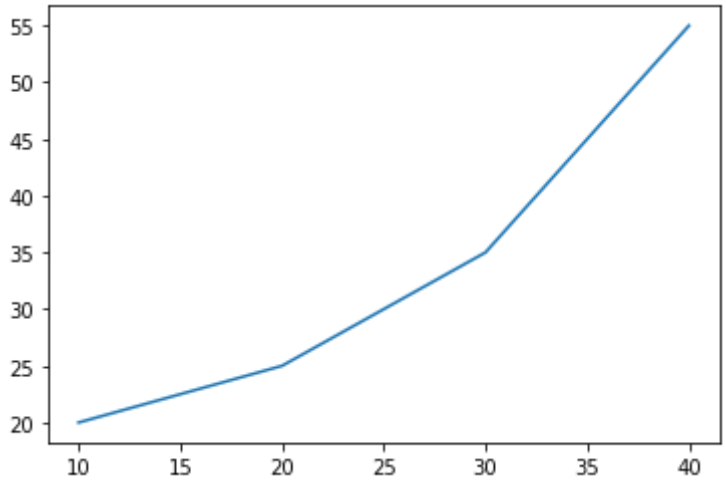
Initializing the data

In [ ]:
```
x = [10, 20, 30, 40]
y = [20, 25, 35, 55]
```

plotting the data

In [ ]:
```
plt.plot(x, y)

plt.show()
```

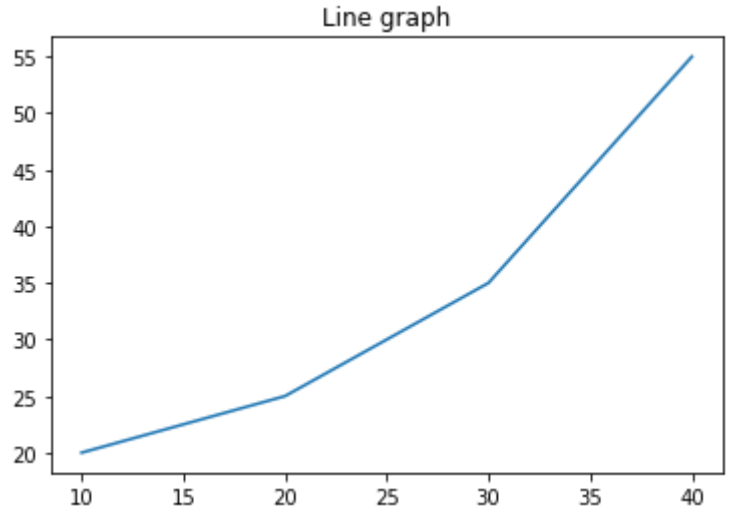


Adding title to the plot

In [ ]:
```
plt.plot(x, y)

plt.title("Line graph")

plt.show()
```



Adding label on the X - Axis and Y - Axis

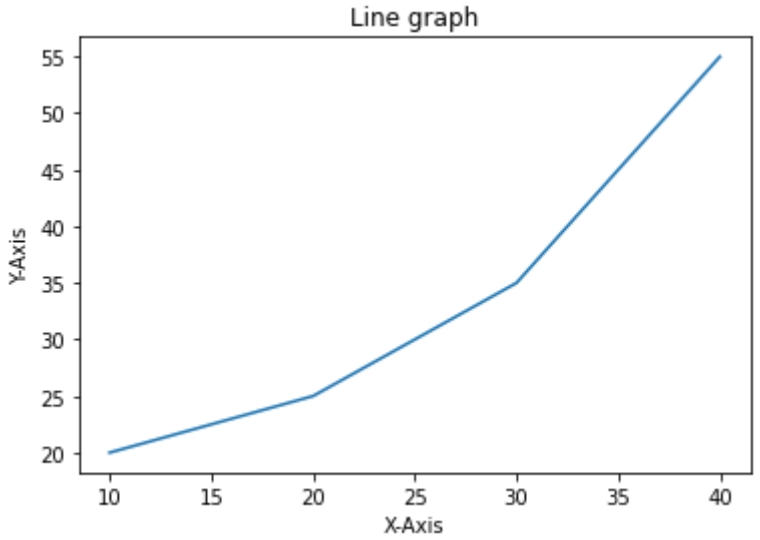
In [ ]:
```
plt.plot(x, y)

plt.title("Line graph")

plt.xlabel('X-Axis')

plt.ylabel('Y-Axis')
```

Out[ ]:
```
Text(0, 0.5, 'Y-Axis')
```



Activity

Create Two Variables Age and Height Start with initial age 5yrs plot a line graph of your age with Height Give it a Title and add X AND Y Labels

In [ ]:
```
age = [5,6,7,8,9,10,11,12,13]
Height = [3.4,4,4.2,4.5,4.8,5.1,5.1,5.2,5.4]
```

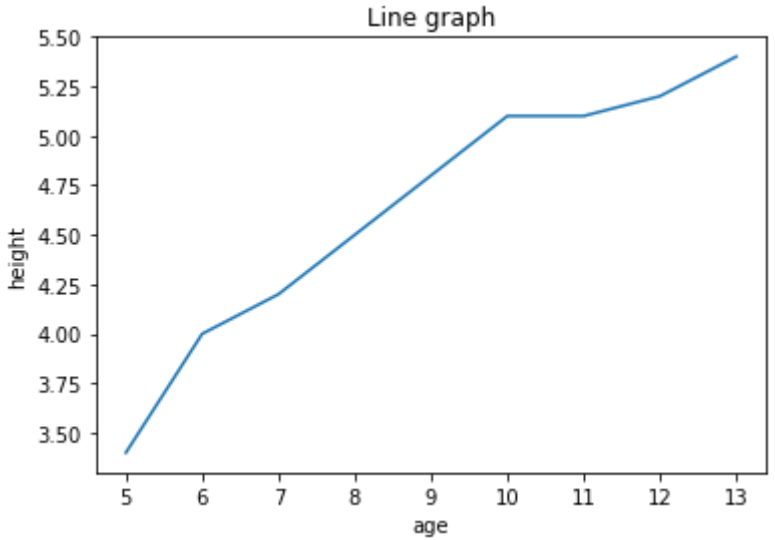
In [ ]:
```
plt.plot(age, Height)

plt.title("Line graph")

plt.xlabel('age')

plt.ylabel('height')
```

Out[ ]:
```
Text(0, 0.5, 'height')
```



Working on Actual Data

In [ ]:
```
import pandas as pd
```

In [ ]:
```
data = pd.read_csv('tips.csv')
```

In [ ]:
```
data.head()
```

Out[ ]:

	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

Understanding Day Wise Revenue

In [ ]:
```
x = data['day']
y = data['total_bill']
```

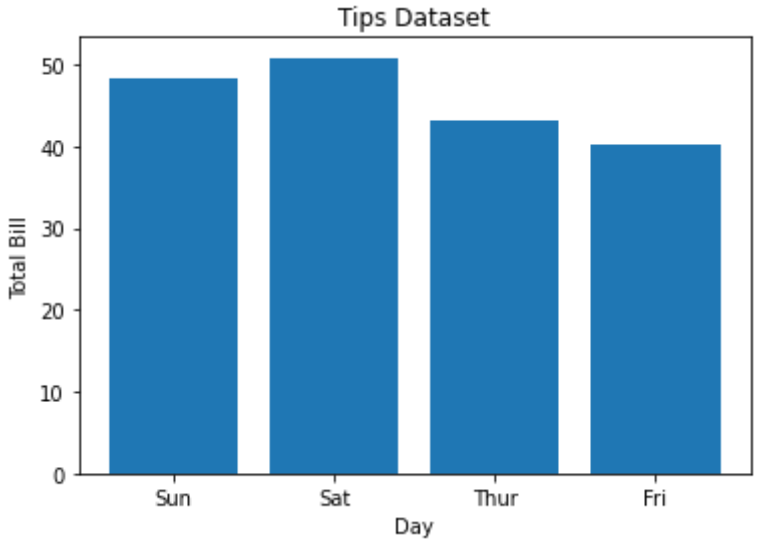
In [ ]:
```
# plotting the data
plt.bar(x, y)

# Adding title to the plot
plt.title("Tips Dataset")

# Adding label on the y-axis
plt.ylabel('Total Bill')

# Adding label on the x-axis
plt.xlabel('Day')

plt.show()
```

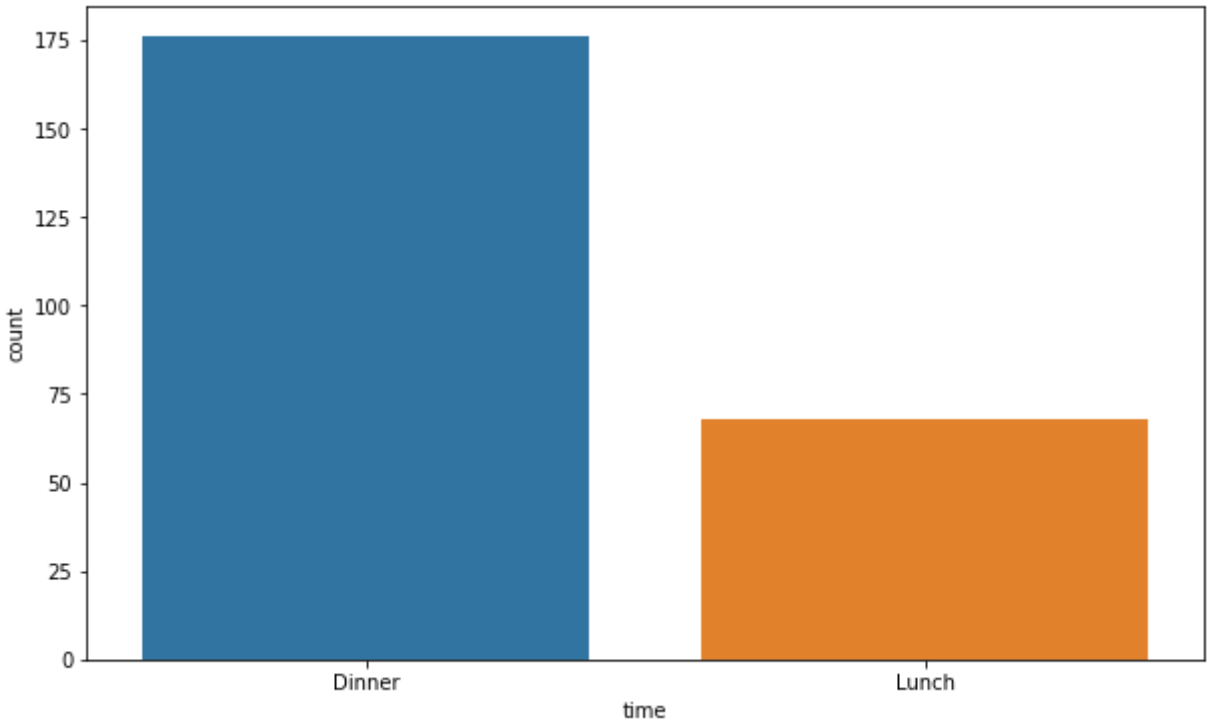


Understanding the data and analyse whether More customer visit for lunch or Dinner

In [ ]:
```
import seaborn as sns
fig = plt.figure(figsize=(10,6))
```

```
sns.countplot(data['time'],data=data)
plt.show()
```

```
C:\Users\khatr\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required fo
r this version of SciPy (detected version 1.23.1
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
C:\Users\khatr\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg:
x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```



In [ ]:

```
head_count=data['time'].value_counts()
print(head_count)
```

```
Dinner    176
Lunch      68
Name: time, dtype: int64
```

Activity: Now Students You Have this Dataset students.csv, in this data set you need to visualize whether we have more male or female students in school.

In [ ]:
```
df = pd.read_csv('students.csv')
```

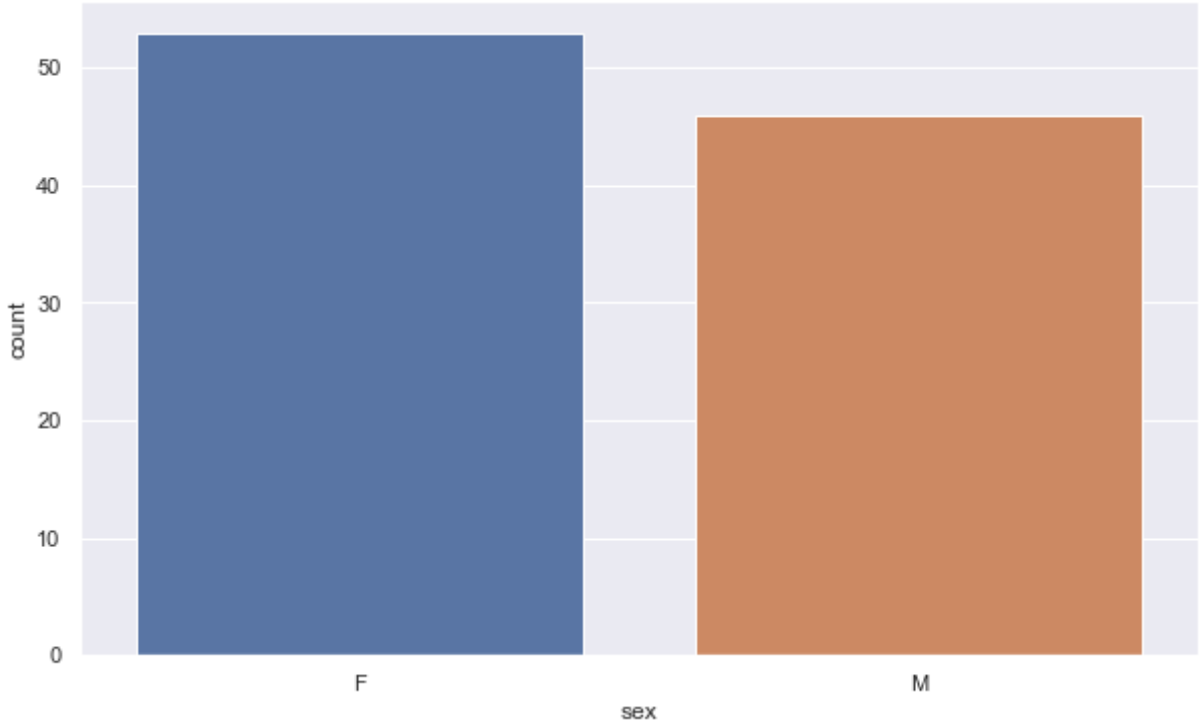
In [ ]:
```
df.head()
```

Out[ ]:

	Gr.No	school	sex	age	guardian
0	1000	GP	F	18	mother
1	1001	GP	F	17	father
2	1002	GP	F	15	mother
3	1003	GP	F	15	mother
4	1004	GP	F	16	father

In [ ]:
```
import seaborn as sns
fig = plt.figure(figsize=(10,6))
sns.countplot(df['sex'],data=df)
plt.show()
```

```
C:\Users\khatr\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg:
x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```

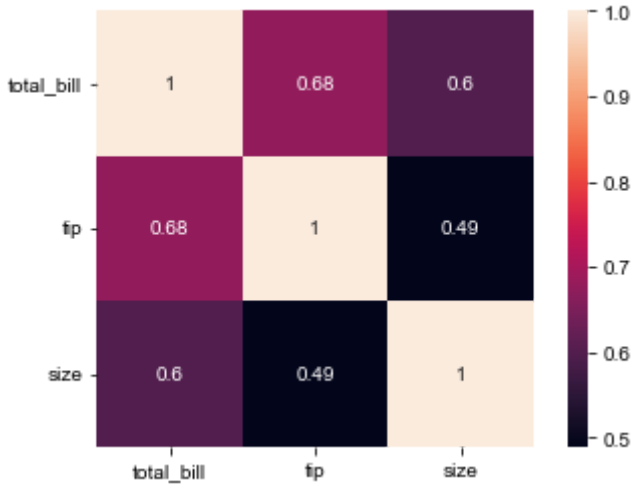


In [ ]:
```
head_count=df['sex'].value_counts()
print(head_count)
```

```
F    53
M    46
Name: sex, dtype: int64
```

Understanding correlation

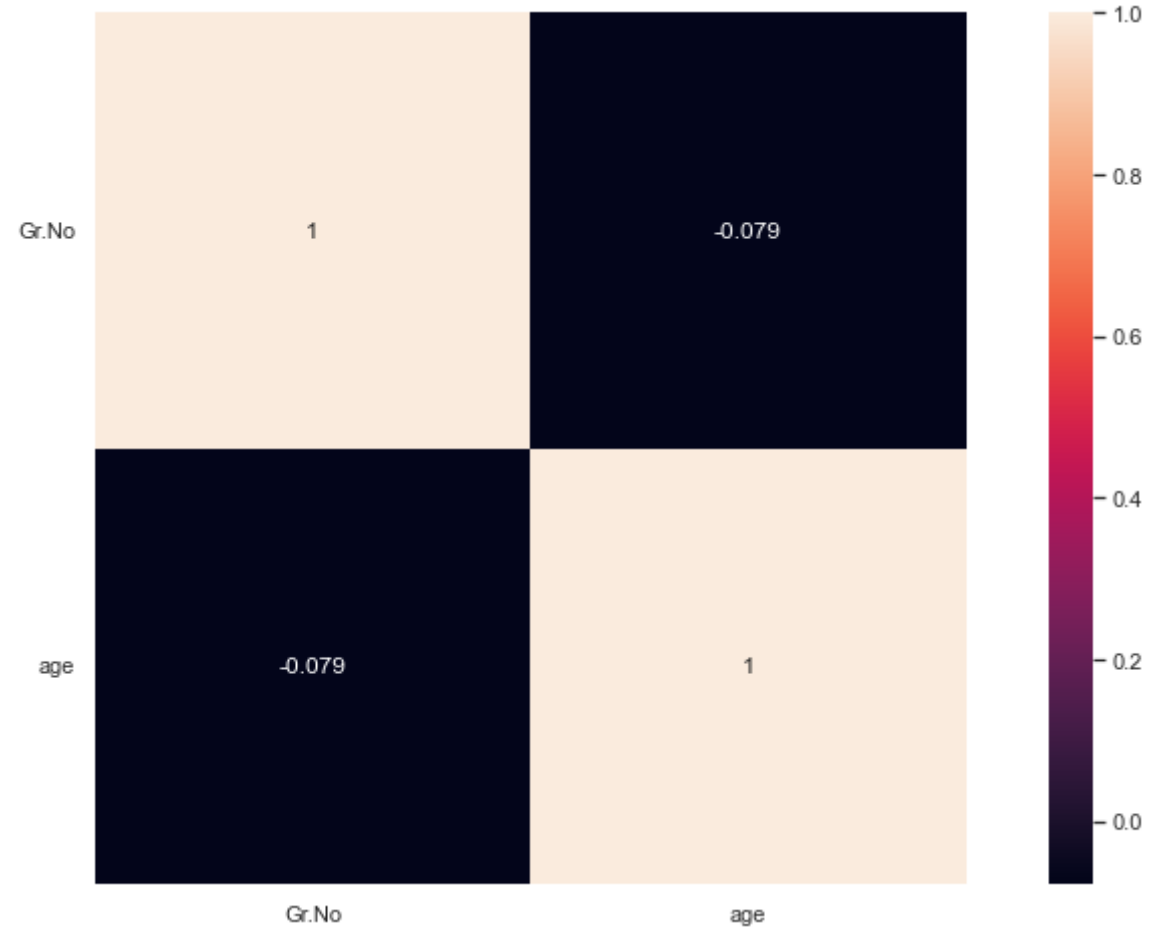
In [ ]:
```
import seaborn as sns
corr = data.corr()
sns.heatmap(corr, annot=True, square=True)
plt.yticks(rotation=0)
sns.set(rc = {'figure.figsize':(150,80)}) #run commands
plt.show()
```



Activity: get Correlation Matrix for students.csv data

In [ ]:

```
import seaborn as sns
corr = df.corr()
sns.heatmap(corr, annot=True, square=True)
plt.yticks(rotation=0)
sns.set(rc = {'figure.figsize':(100,60)})
plt.show()
```


Gr.No
age
1
-0.079
-0.079
1
Gr.No
age
1.0
0.8
0.6
0.4
0.2
0.0