Module 3: Data Visualization

Demo I

Demo I

1. You are given Hurricanes dataset, which is present in the LMS, containing the number of hurricanes occurring in the United States along the coast of the Atlantic. Load the data from the dataset into your program and plot a Bar Graph of the data, taking the Year as the x-axis and the number of hurricanes occurring as the Y-axis.

2. The dataset given, records data of city temperatures over the years 2014 and 2015. Plot the histogram of the temperatures over this period for the cities of San Francisco and Moscow.

3. Plot a pie-chart of the number of models released by every manufacturer, recorded in the data provide. Also mention the name of the manufacture with the largest releases.

4. Perform the following steps on the Sales dataset:
   Phase 1 - Read the data as Pandas dataframe
   Phase 2 - Describe the data on the **unit price**
   Phase 3 - Filter the data by creating a new dataframe that have columns 'name', 'net_price', 'date' and group all the records according to name
   Phase 4 - Plot the graph after calculating the total sales by each Sales Representative. Representative name should be on x axis and total sales in y axis.

5. Let the x axis data points and y axis data points are
   X = [1,2,3,4]
   y = [20, 21, 20.5, 20.8]
   5.1: Draw a Simple plot
   5.2: Configure the line and markers in simple plot
   5.3: Configure the axes
   5.4: Give title of Graph & labels of x axis and y axis
   5.5: Give error bar if y_error = [0.12, 0.13, 0.2, 0.1]
   5.6: Define width, height as figsize= (4,5) DPI and adjust plot dpi=100
   5.7: Give a font size of 14
   5.8: Draw a scatter graph of any 50 random values of x and y axis

5.9: Create a dataframe from following data

'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze'],
'female': [0, 1, 1, 0, 1],
'age': [42, 52, 36, 24, 73],
'preTestScore': [4, 24, 31, 2, 3],
'postTestScore': [25, 94, 57, 62, 70]

Draw a Scatterplot of preTestScore and postTestScore, with the size of each point determined by age

5.10: Draw a Scatterplot from the data in question 9 of preTestScore and postTestScore with the size = 300 and the color determined by sex

Demo I Solution

1. You are given Hurricanes dataset, which is present in the LMS, containing the number of hurricanes occurring in the United States along the coast of the Atlantic. Load the data from the dataset into your program and plot a Bar Graph of the data, taking the Year as the x-axis and the number of hurricanes occurring as the Y-axis.

Solution

```
import matplotlib.pyplot as plt, pandas
df = pandas.read_csv('Hurricanes.csv', delimiter=',')
print(df.head())
plt.bar(df["Year"], df["Hurricanes"])
plt.show()
```

2. The dataset given, records data of city temperatures over the years 2014 and 2015. Plot the histogram of the temperatures over this period for the cities of San Francisco and Moscow.

Solution

```
import matplotlib.pyplot as plt, pandas
df = pandas.read_csv('City_temperatures.csv', delimiter=',')
df1=df.loc[:, ['Melbourne', 'San Francisco']]
df1.hist(density=True, histtype='bar', color="blue")
plt.show()
```

3. Plot a pie-chart of the number of models released by every manufacturer, recorded in the data provide. Also mention the name of the manufacture with the largest releases.

Solution

```
import matplotlib.pyplot as plt, pandas
car_data = pandas.read_csv('Car_manufactures.csv', delimiter=',')
manufacturer = car_data['Make'].value_counts()
manufacturer.plot.pie(figsize=(10,15),autopct='%1.1f%%')
plt.show()
```

4. Perform the following steps on the Sales dataset:

Solution

```
#Phase 1 - Read the data as Pandas dataframe

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
sales=pd.read_csv("Sample_sales.csv",parse_dates=['date'])
print(sales.head())

#Phase 2 - Describe the data on the unit price

print(sales['unit price'].describe())

#Phase 3 - Filter the data by creating a new dataframe that have columns
#'name', 'net_price', 'date' and group all the records according to name

customers = sales[['name','net_price','date']]
customer_group = customers.groupby('name')
print(customer_group.size())

#Phase 4 - Plot the graph after calculating the total sales by each Sales
#Representative.  Representative name should be on x axis and total sales in y
axis.

sales_totals = customer_group.size()
my_plot = sales_totals.plot(kind='bar')
plt.show()
```

5. Let the x axis data points and y axis data points are
   X = [1,2,3,4]
   y = [20, 21, 20.5, 20.8]

Solution

#5.1: Draw a Simple plot

```
import pandas as pd
import matplotlib.pyplot as plt
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#plot data
plt.plot(x, y)
#show plot
plt.show()
```

#5.2: Configure the line and markers in simple plot


```
import matplotlib.pyplot as plt
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#show plot
plt.show()
```

#5.3: Configure the axes

```
import matplotlib.pyplot as plt
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#configure  X axes
plt.xlim(0.5,4.5)
plt.xticks([1,2,3,4])
#configure  Y axes
plt.ylim(19.8,21.2)
```

```
plt.yticks([20, 21, 20.5, 20.8])
#show plot
plt.show()
```

#5.4: Give title of Graph & labels of x axis and y axis

```
import matplotlib.pyplot as plt
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#configure  X axes
plt.xlim(0.5,4.5)
plt.xticks([1,2,3,4])
#configure  Y axes
plt.ylim(19.8,21.2)
plt.yticks([20, 21, 20.5, 20.8])
#labels
plt.xlabel("this is X")
plt.ylabel("this is Y")
#title
plt.title("Simple plot")
#show plot
plt.show()
```


#5.5: Give error bar if y_error = [0.12, 0.13, 0.2, 0.1]

```
import matplotlib.pyplot as plt
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#error data
y_error = [0.12, 0.13, 0.2, 0.1]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#plot only errorbars
plt.errorbar(x, y, yerr=y_error, linestyle="None", marker="None",
color="green")
#configure  X axes
plt.xlim(0.5,4.5)
plt.xticks([1,2,3,4])
```

```
#configure  Y axes
plt.ylim(19.8,21.2)
plt.yticks([20, 21, 20.5, 20.8])
#labels
plt.xlabel("this is X")
plt.ylabel("this is Y")
#title
plt.title("Simple plot")
#show plot
plt.show()
```

#5.6: Define width, height as figsize= (4,5) DPI and adjust plot dpi=100

```
import matplotlib.pyplot as plt
#define plot size in inches (width, height) & resolution(DPI)
fig = plt.figure(figsize=(4, 5), dpi=100)
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#error data
y_error = [0.12, 0.13, 0.2, 0.1]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#plot only errorbars
plt.errorbar(x, y, yerr=y_error, linestyle="None", marker="None",
color="green")
#configure  X axes
plt.xlim(0.5,4.5)
plt.xticks([1,2,3,4])
#configure  Y axes
plt.ylim(19.8,21.2)
plt.yticks([20, 21, 20.5, 20.8])
#labels
plt.xlabel("this is X")
plt.ylabel("this is Y")
#title
plt.title("Simple plot")
#adjust plot
plt.subplots_adjust(left=0.18)
#show plot
plt.show()
```

5.7: Give a font size of 14

```
import matplotlib.pyplot as plt
#define plot size in inches (width, height) & resolution(DPI)
fig = plt.figure(figsize=(4, 5), dpi=100)
#define font size
plt.rc("font", size=14)
#define some data
x = [1,2,3,4]
y = [20, 21, 20.5, 20.8]
#error data
y_error = [0.12, 0.13, 0.2, 0.1]
#plot data
plt.plot(x, y, linestyle="dashed", marker="o", color="green")
#plot only errorbars
plt.errorbar(x, y, yerr=y_error, linestyle="None", marker="None",
color="green")
#configure  X axes
plt.xlim(0.5,4.5)
plt.xticks([1,2,3,4])
#configure  Y axes
plt.ylim(19.8,21.2)
plt.yticks([20, 21, 20.5, 20.8])
#labels
plt.xlabel("this is X")
plt.ylabel("this is Y")
#title
plt.title("Simple plot")
#adjust plot
plt.subplots_adjust(left=0.19)
#show plot
plt.show()
```

#5.8: Draw a scatter graph of any 50 random values of x and y axis

```
import numpy as np
import matplotlib.pyplot as plt
N = 50
x = np.random.rand(N)
y = np.random.rand(N)
colors = np.random.rand(N)
```

```
area = np.pi * (15 * np.random.rand(N))**2  # 0 to 15 point radii
plt.scatter(x, y, s=area, c=colors, alpha=0.5)
plt.show()
```

#5.9: Create a dataframe from following data

```
        '''first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
         'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze'],
         'female': [0, 1, 1, 0, 1],
         'age': [42, 52, 36, 24, 73],
        'preTestScore': [4, 24, 31, 2, 3],
        'postTestScore': [25, 94, 57, 62, 70] '''
```

#Draw a Scatterplot of preTestScore and postTestScore, with the size of each point determined by age

```
raw_data = {'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
    'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze'],
    'female': [0, 1, 1, 0, 1],
    'age': [42, 52, 36, 24, 73],
    'preTestScore': [4, 24, 31, 2, 3],
    'postTestScore': [25, 94, 57, 62, 70]}
df = pd.DataFrame(raw_data, columns = ['first_name', 'last_name', 'age',
'female', 'preTestScore', 'postTestScore'])
print(df.head(3))
plt.scatter(df.preTestScore, df.postTestScore, s=df.age)
```

#5.10: Draw a Scatterplot from the data in question 9 of preTestScore and postTestScore with the size = 300 and the color determined by sex

```
plt.scatter(df.preTestScore, df.postTestScore, s=300, c=df.female)
```