14-09-2023

Project Title:

Analysis & visualization of product categories, across their market and region using Python matplot library.

Name: T.Jaya Krishna

Branch: CSE-B

Year: B.tech 4th year

Project Guide

Mr. Lokesh B. Sir

Problem Statement

An organization called product based company publishing their problem statement in 2011 hackthon. Now the organization wants tabular data into visualization format, as a data science Engineer.I have to provide solution for following Task.

Tools:

1. Jupyter Notebook.
2. Python Programing Language.
3. Numpy (Python Libery_)
4. Google Colab.

Project Title : Analisis & visualization of product categories, across their market and region using python matplot library.

Organization Name : IBM research pvt ltd

Problem Statement:A Organization called product based company publishing their problem statement in 2011 hackthon. Now the organization wants tabular data into visualization format, as a data science engineer.I have to provide solution for following task.

Task-1:-

Find out what is the maximum & minimum profit value across sales.It is mandatory to be in Bar graph'd only.

Task-2:-

Find out maximum & minimum sales for profit across different categories all categories must be in different colors.

Task-3:-

In which month sales give highest Employee visualization must be Histogram.

Task-4:-

What is the highest age of employee.Visualization must be Histogram.

Visualisation in Python - Matplotlib

You will be working with the sales dataset for an online retailer. The data is collected over a period of three years: 2012 to 2015. It contains the information of sales made by the company.

The products captured belong to three categories:

- Furniture
- Office Supplies
- Technology

Also, the company caters to five different markets:

- USCA
- LATAM
- ASPAC
- EUR
- AFR

Let's get started with the plots. We will be using the '**pyplot**' package of the Matplotlib library.

```
# importing numpy and the pyplot package of matplotlib
import matplotlib.pyplot as plt
import numpy as np
```

```
# Creating an array with product categories
product_categories=np.array(["Furniture","Office Supplies","Technology"])
```

```
# Creating an array with the sales amount
# Furniture: 4110451.90
# Technology: 4744557.50
# Office Supplies: 3787492.52
sales=np.array([4110451.90,4744557.50,3787492.52])
```

It is not necessary that you are provided with the aggregated values every time. In such cases, you first need to calculate the values and then build the graphs.

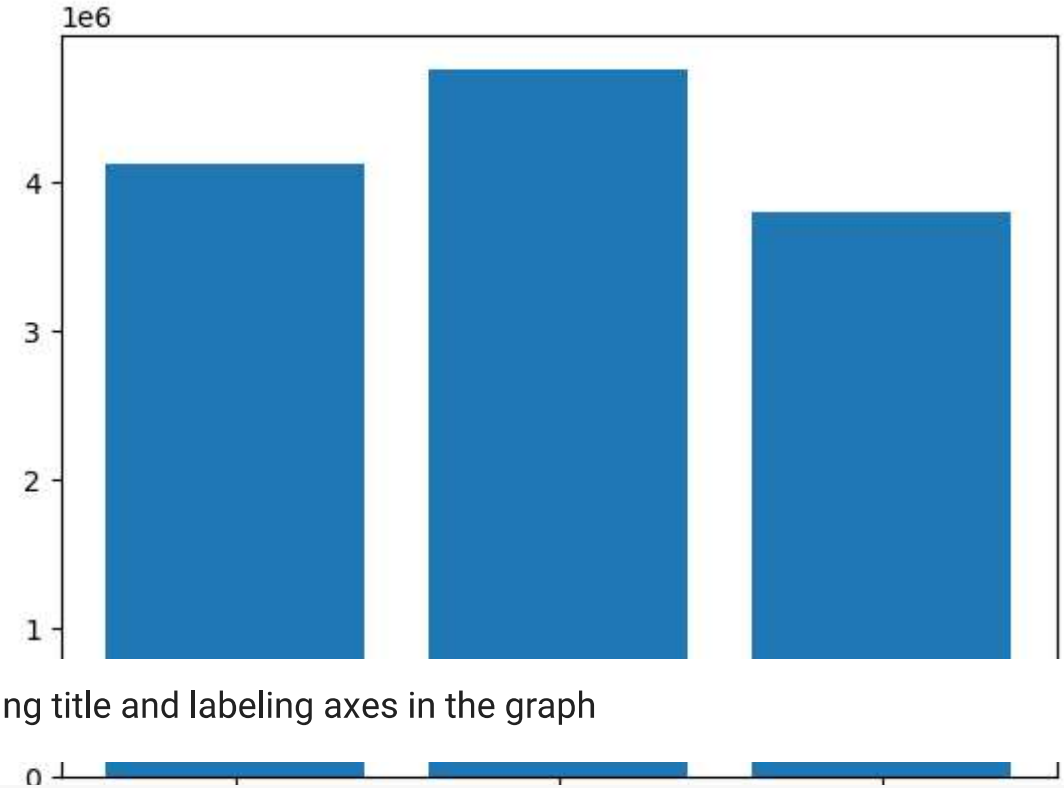
Let's see how to plot a bar graph for the provided values.

Bar Graph: Plotting sales across each product category

- A bar chart uses bars to show comparisons between categories of data.
- One axis will generally have numerical values or measures,
- The other will describe the types of categories being compared or dimensions.

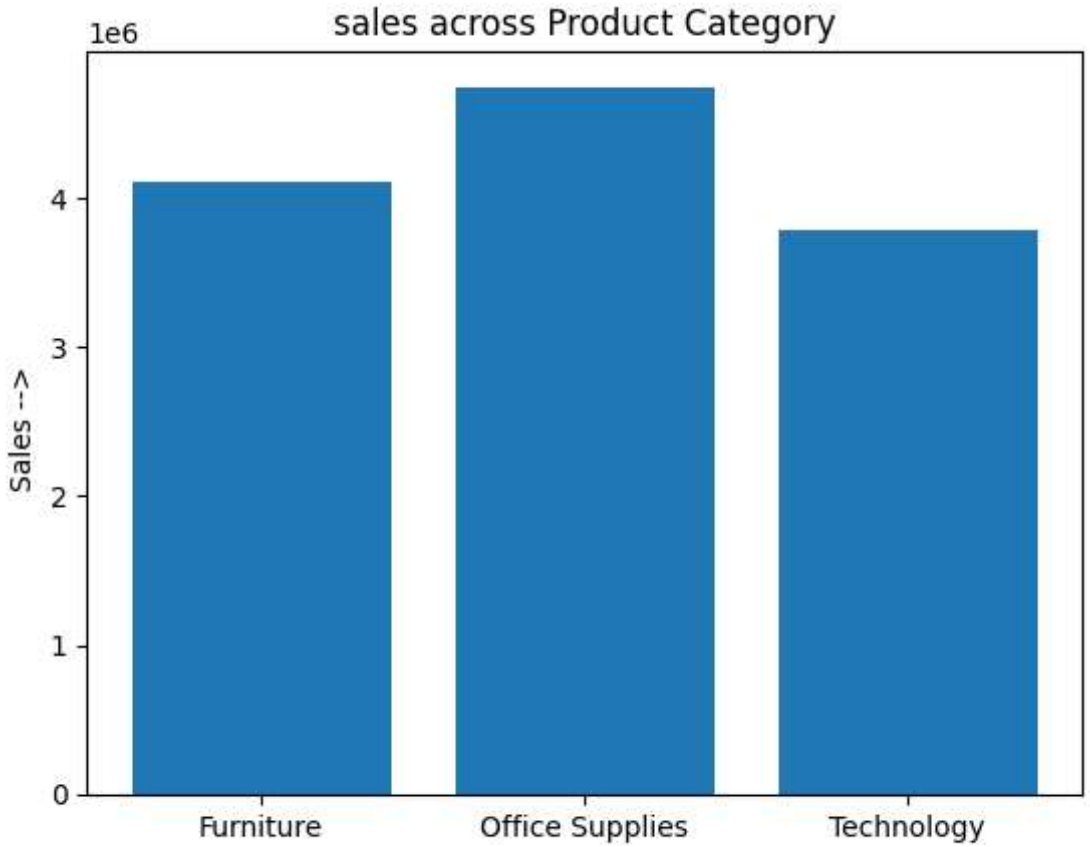
Let's start with plotting a bar graph representing the sales across different categories over the period.

```
# plotting the bar graph with product categories on x-axis and sales amount of y-axis
plt.bar(product_categories, sales)

# necessary command to display the created graph
plt.show()
```



Adding title and labeling axes in the graph

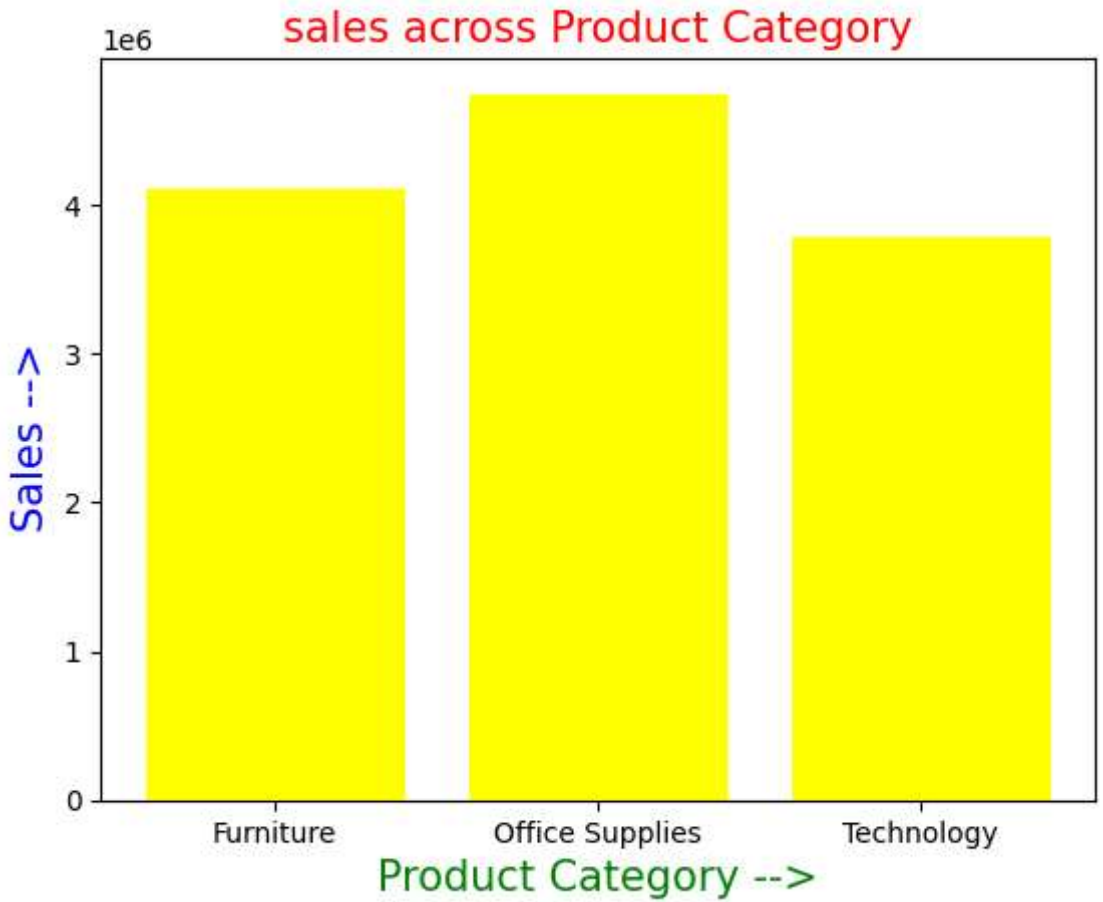
```
# plotting the bar graph with product categories on x-axis and sales amount of y-axis
plt.bar(product_categories,sales)

# adding title to the graph

plt.title("sales across Product Category")
# labeling axes
plt.xlabel("Product Category -->")
plt.ylabel("Sales --> ")

# necessary command to display the created graph
plt.show()
```



▾ Modifying the bars in the graph

```
# changing color of the bars in the bar graph
# plotting the bar graph with product categories on x-axis and sales amount of y-axis
plt.bar(product_categories,sales,color="Yellow")

# adding title to the graph
plt.title("sales across Product Category",{"fontsize":15,"fontweight":7,"color":"Red"})

# labeling axes
plt.xlabel("Product Category -->",{"fontsize":15,"fontweight":7,"color":"Green"})
plt.ylabel("Sales --> ",{"fontsize":15,"fontweight":7,"color":"Blue"})
```

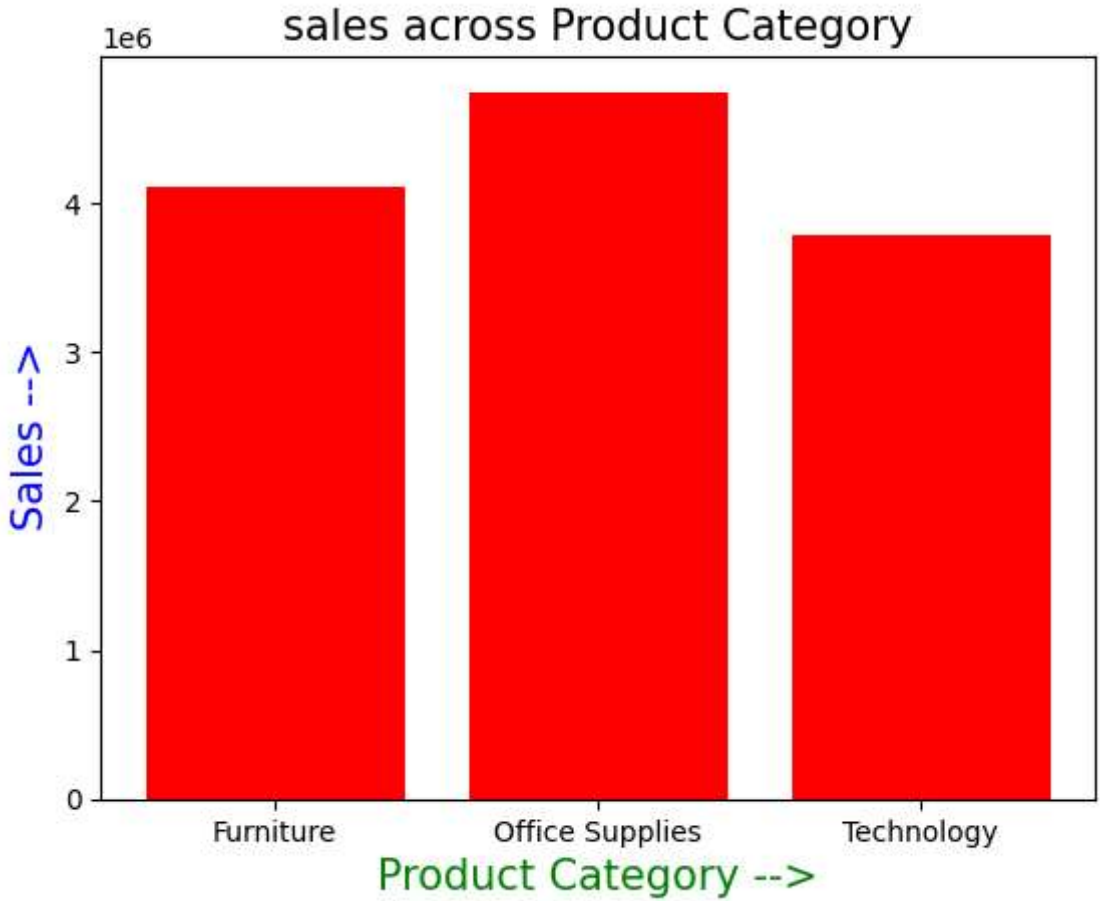
```
# necessary command to display the created graph
plt.show()
```



Adjusting tick values and the value labels

```
# plotting the bar graph with product categories on x-axis and sales amount of y-axis
product_categories=np.array(["Furniture","Office Supplies","Technology"])

sales=np.array([4110451,4744557,3787492])
```

```
plt.bar(product_categories,sales,color="Red")
# adding title to the graph
plt.title("sales across Product Category",{"fontsize":15,"fontweight":7,"color":"Black"})

# labeling axes
plt.xlabel("Product Category -->",{"fontsize":15,"fontweight":7,"color":"Green"})
plt.ylabel("Sales --> ",{"fontsize":15,"fontweight":7,"color":"Blue"})

# Modifying the ticks to show information in (lakhs)
tick_values =(0,50000,10000)
tick_labels =("0","10","20","30","40")

plt.ylabel=(tick_values,tick_labels)
# necessary command to display the created graph
plt.show()
```


sales across Product Category
1e6
4
3
2
1
0
Sales -->
Furniture
Office Supplies
Technology
Product Category -->

Scatter Chart: Plotting Sales vs Profits

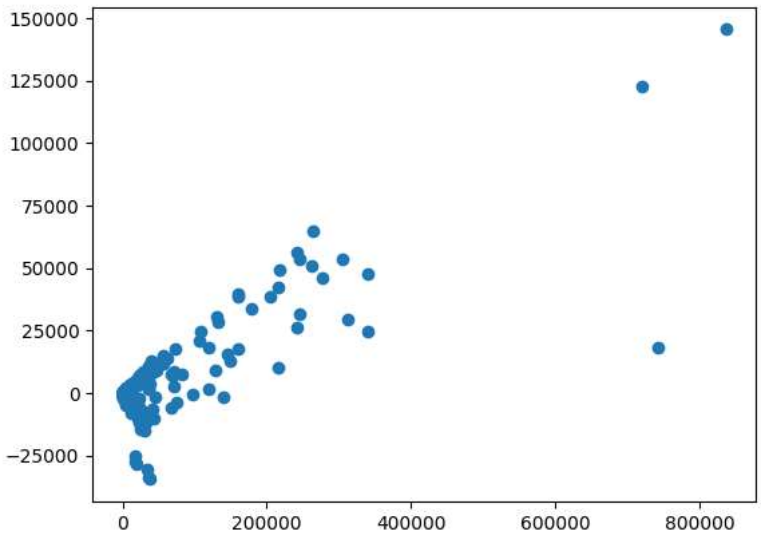
- Scatter plots are used when you want to show the relationship between two facts or measures.

Now, you have the sales and profit data of different product categories across different countries. Let's try to build scatterplots to visualise the data at hand.

```
# importing numpy and the pyplot package of matplotlib
import matplotlib.pyplot as plt
import numpy as np
# Sales and Profit data for different product categories across different countries
sales = np.array ([1013.14, 8298.48, 875.51, 22320.83, 9251.6, 4516.86, 585.16, 836154.03, 216748.48, 174.2, 27557.79, 563.25, 558.1
profit = np.array([-1213.46, 1814.13, -1485.7, -2286.73, -2872.12, 946.8, 198.48, 145454.95, 49476.1, -245.56, 5980.77, -790.47, -89
```

Plotting a scatterplot

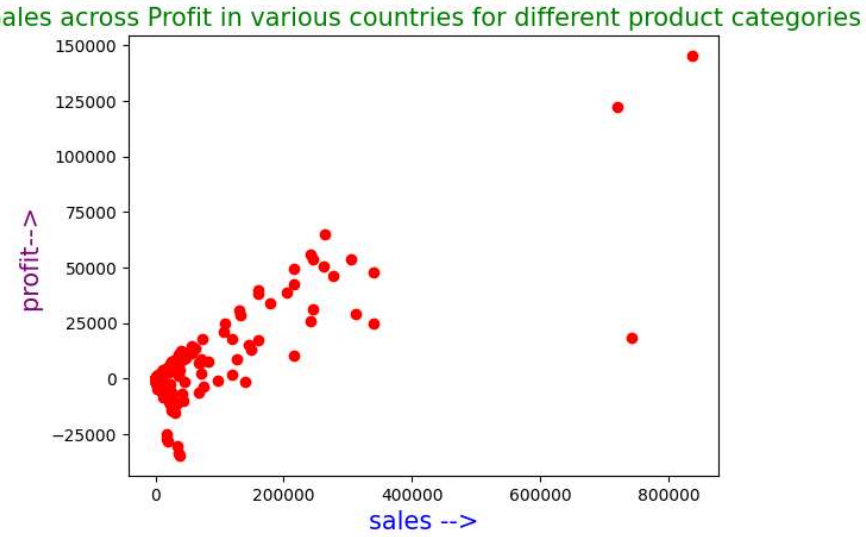
```
# plotting scatterplot
plt.scatter(sales, profit)

# necessary command to display graph
plt.show()
```




```
# Sales and Profit data for different product categories across different countries
sales = np.array ([1013.14, 8298.48, 875.51, 22320.83, 9251.6, 4516.86, 585.16, 836154.03, 216748.48, 174.2, 27557.79, 563.25, 558.1
profit = np.array([-1213.46, 1814.13, -1485.7, -2286.73, -2872.12, 946.8, 198.48, 145454.95, 49476.1, -245.56, 5980.77, -790.47, -89

# corresponding category and country value to the above arrays
product_category = np.array(['Technology', 'Technology', 'Technology', 'Technology', 'Technology', 'Technology', 'Technology', 'Tech
country = np.array(['Zimbabwe', 'Zambia', 'Yemen', 'Vietnam', 'Venezuela', 'Uzbekistan', 'Uruguay', 'United States', 'United Kingdom
```

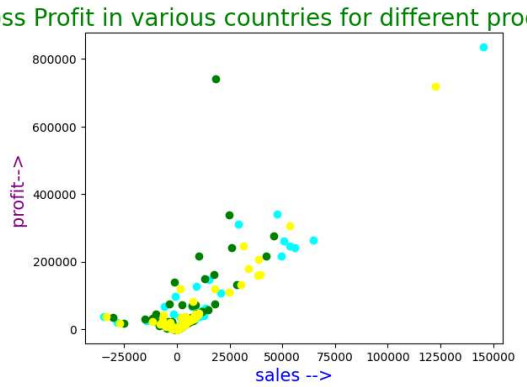
Adding title and labeling axes

```
# plotting scatter chart
plt.scatter(sales, profit,color="red")

# Adding and formatting title
plt.title("Sales across Profit in various countries for different product categories",{"fontsize":15,"fontweight":5,"color":"green"}

# Labeling Axes
plt.xlabel("sales -->",{"fontsize":15,"fontweight":5,"color":"blue"})
plt.ylabel("profit--> ",{"fontsize":15,"fontweight":5,"color":"purple"})

plt.show()
```



Representing product categories using different colors

```
product_categories = np.array(["Technology", "Furniture", "Office Supplies"])
colors = np.array(["cyan", "green", "yellow"])

# plotting the scatterplot with color coding the points belonging to different categories
for color,category in zip (colors, product_categories):
  sales_category=sales[product_category == category]
  profit_category=profit[product_category == category]

  plt.scatter(profit_category,sales_category, c=color , label = category)

# Adding and formatting title
plt.title("Sales across Profit in various countries for different product categories",{"fontsize":20,"fontweight":5,"color":"green"}

# Labeling Axes
plt.xlabel("sales -->",{"fontsize":15,"fontweight":5,"color":"blue"})
plt.ylabel("profit--> ",{"fontsize":15,"fontweight":5,"color":"purple"})

# Adding legend for interpretation of points
plt.legend

plt.show()
```



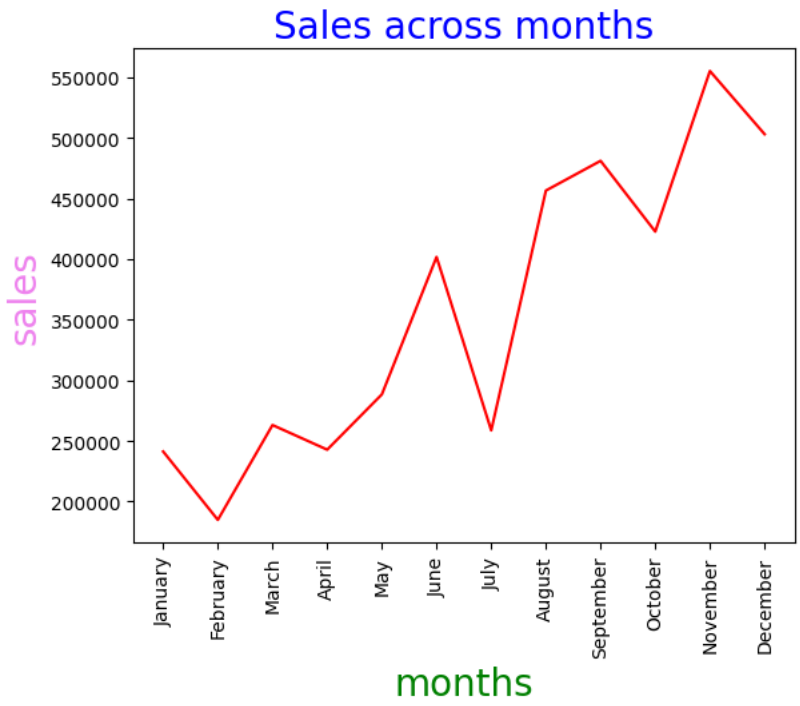
✓ 0s completed at 12:38 PM

Visualisation in Python - Matplotlib

Line Chart: Trend of sales over the 12 months

- Can be used to present the trend with time variable on the x-axis
- In some cases, can be used as an alternative to scatterplot to understand the relationship between 2 variables

```
# importing the required libraries
import matplotlib.pyplot as plt
import numpy as np

# Sales data across months
months = np.array(['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'De
sales = np.array([241268.56, 184837.36, 263100.77, 242771.86, 288401.05, 401814.06, 258705.68, 456619.94, 481157.24, 422766.63, 5552


# plotting a line chart
plt.plot(months, sales,color="red")

# adding title to the chart
plt.title("Sales across months",{"fontsize":20,"color":"blue","fontweight":7})

# labeling the axes
plt.xlabel("months",{"fontsize":20,"color":"green","fontweight":7})
plt.ylabel("sales",{"fontsize":20,"color":"violet","fontweight":7})

# rotating the tick values of x-axis
plt.xticks(rotation=90)

# displating the created plot
plt.show()
```



✓ 0s completed at 2:09 PM

Visualisation in Python - Matplotlib

Histogram: Distibution of employees across different age groups

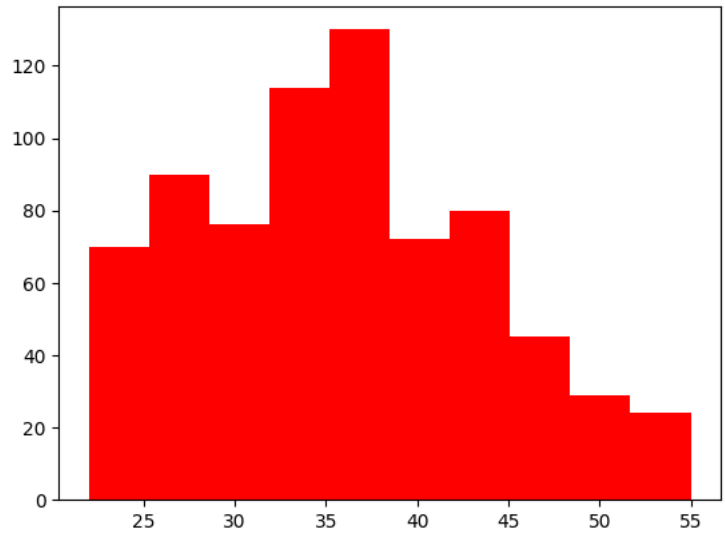
- Useful in checking the distribution of data range
- Builds a bar corresponding to each element in the data range showing its frequency

```
# importing the required libraries - numpy, matplotlib.pyplot
import matplotlib.pyplot as plt
import numpy as np

# data corresponding to age of the employees in the company
age = np.array([23, 22, 24, 24, 23, 23, 22, 23, 24, 24, 24, 22, 24, 23, 24, 23, 22, 24, 23, 23, 22, 23, 23, 24, 23, 24, 23, 22, 24,
```

```
# Checking the number of employees
len(age)
```

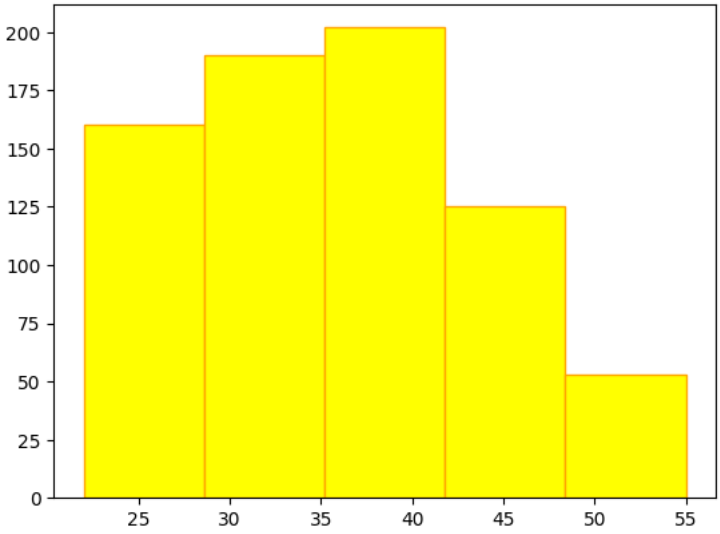
```
730
```

```
# plotting a histogram
plt.hist(age,color="red")

plt.show()
```



Plotting a histogram with fixed number of bins

```
# plotting a histogram
plt.hist(age,bins=5,color="yellow" , edgecolor="orange")

plt.show()
```



✓ 0s completed at 2:23 PM

Visualisation in Python - Matplotlib

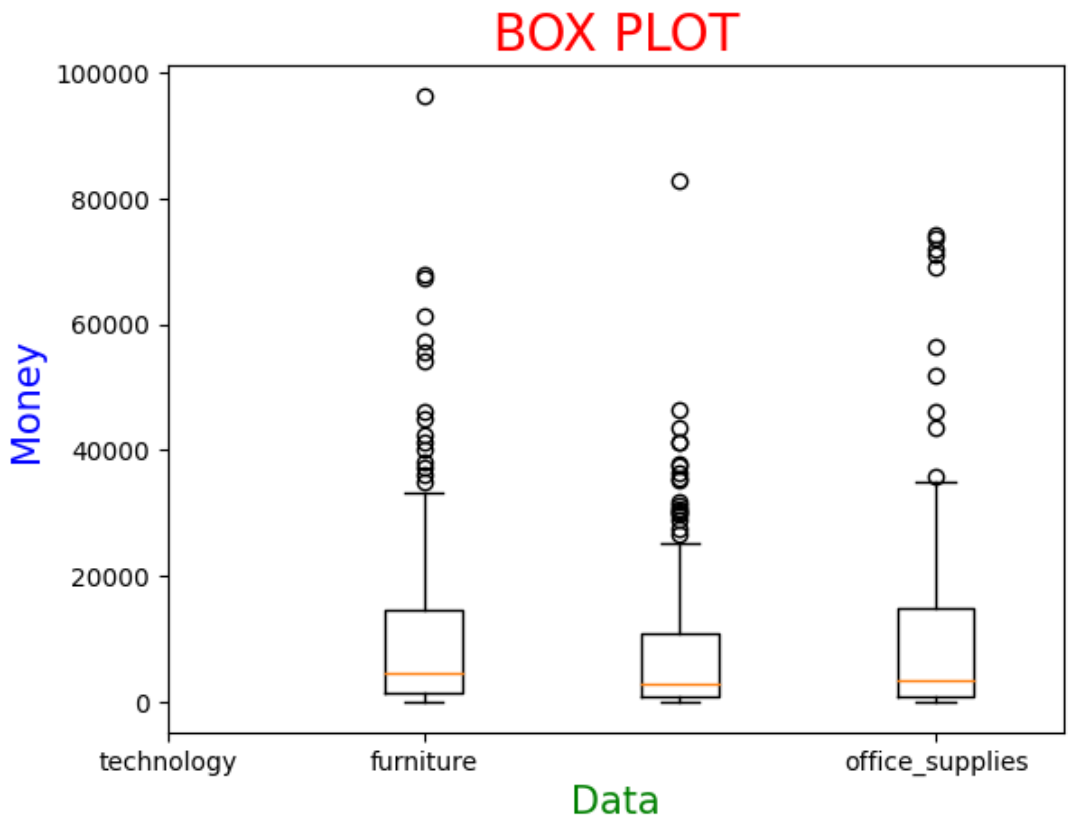
Box plot: Understanding the spread of sales across different countries

- Useful in understanding the spread of the data
- Divides the data based on the percentile values
- Helps identify the presence of outliers

```
# importing numpy and the pyplot package of matplotlib
import matplotlib.pyplot as plt
import numpy as np

# Creating arrays with sales in different countries across each category: 'Furniture', 'Technology' and 'Office Supplies'
sales_technology = np.array ([1013.14, 8298.48, 875.51, 22320.83, 9251.6, 4516.86, 585.16, 174.2, 27557.79, 563.25, 558.11, 37117.45, 357.
sales_office_supplies = np.array ([1770.13, 7527.18, 1433.65, 423.3, 21601.72, 10035.72, 2378.49, 3062.38, 345.17, 30345.78, 300.71, 940.8
sales_furniture = np.array ([981.84, 10209.84, 156.56, 243.06, 21287.52, 7300.51, 434.52, 6065.0, 224.75, 28953.6, 757.98, 528.15, 34922.4
```

```
# plotting box plot for each category
zxcvbn=[sales_technology,sales_office_supplies,sales_furniture]
plt.boxplot(zxcvbn)
#plt.boxplot(sales_technology)
#plt.boxplot(sales_office_supplies)
#plt.boxplot(sales_furniture)

# adding title to the graph
plt.title("BOX PLOT",{"fontsize":20,"color":"red","fontweight":7})

# labeling the axes
plt.xlabel("Data",{"fontsize":15,"color":"green","fontweight":7})
plt.ylabel("Money",{"fontsize":15,"color":"blue","fontweight":7})

# Replacing the x ticks with respective category
a_label=(0,3,1)
```

```
b_values=["technology","office_supplies","furniture"])
plt.xticks(a_label,b_values)

plt.show()
```



Conclusion :

Bar Graph :

Bar graphs are a convenient method to represent different sets of data. Apart from this, these types of graphs are easy to describe the maximum and minimum profits values across sales.

Scatterplot :

Scatterplots can visually show the maximum and minimum profits sales across different categories.whether there is a positive or negative association between the variables.All categories must be in different colors.

Line Graph :

Line Chart is used to represent trend of sales over the 12 months with sales across the months with maximum and minimum values by representing the line value with hight peak vlaue.

Histogram :

Here Histogram is used to represent the distibution of employees across different age groups checking the distribution of data range using maximum and minumum of with fixed number of bins.

Box Plot :

Boxplot is used to Understanding the spread of sales across different countries.Based on these data it seems that divides the data based on the percentile values.It helps to identify the presence of outliers over the graph.