Introduction to Matplotlib

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: October 22, 2024
Section: CPE21S4	**Date Submitted**: October 22, 2024
Name: Don Eleazar T. Fernandez	**Instructor**: Maria Rizette Sayo

Intended Learning Outcomes (ILO):

By the end of this laboratory session, learners will be able to:

1. Utilize Matplotlib's pyplot interface to create a variety of visualizations, including line plots, scatter plots, histograms, and box plots, demonstrating an understanding of the library's syntax and functionality.

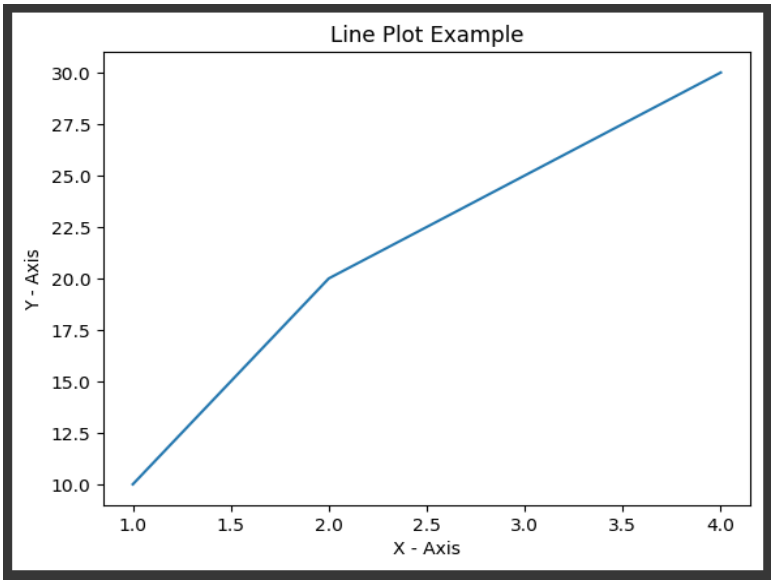
2. Customize visual elements such as titles, labels, and legends to enhance the clarity and aesthetics of their plots, applying best practices in data visualization.

3. Analyze and interpret visual data representations to extract meaningful insights, effectively communicating findings through well-structured graphical presentations.

Part 1: Perform the following codes, and understand the difference between line plot, scatter plot, histogram, bar chart, box plot, and pie chart using matplotlib's pyplot sub-module. **(Provide a screenshot of your output.)**

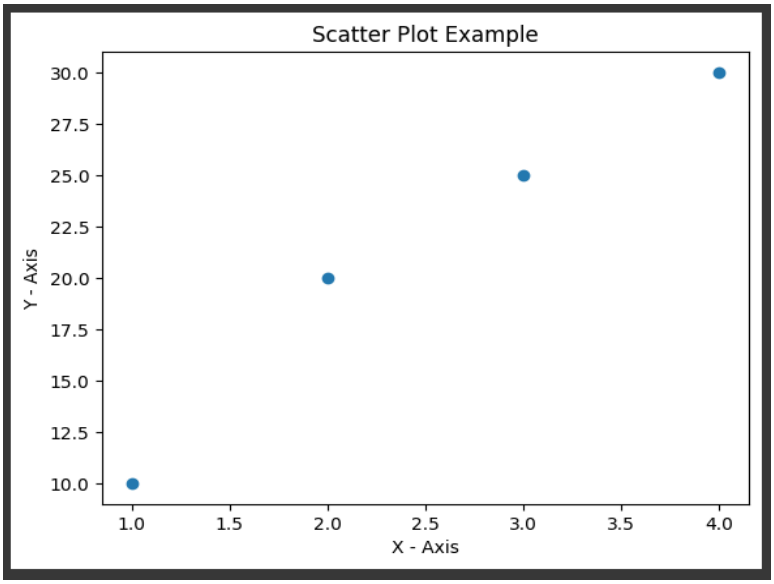
1. Line Plot

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.plot(x, y)
plt.title("Line Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```



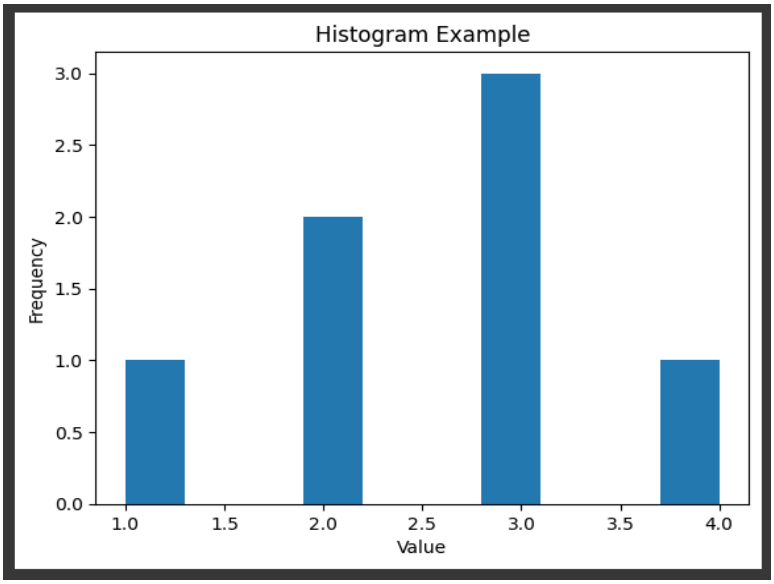
2. Scatter Plot

```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```



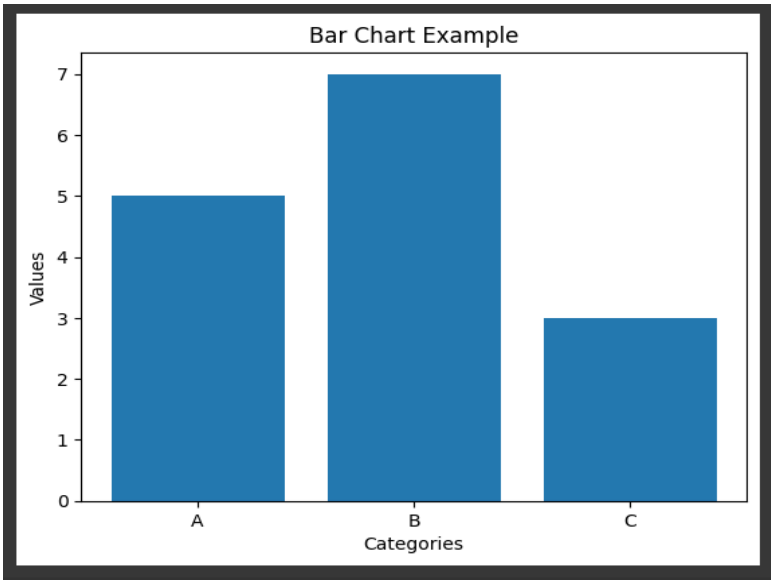
3. Histogram

```
import matplotlib.pyplot as plt

data = [1, 2, 2, 3, 3, 3, 4]
plt.hist(data)
plt.title("Histogram Example")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```



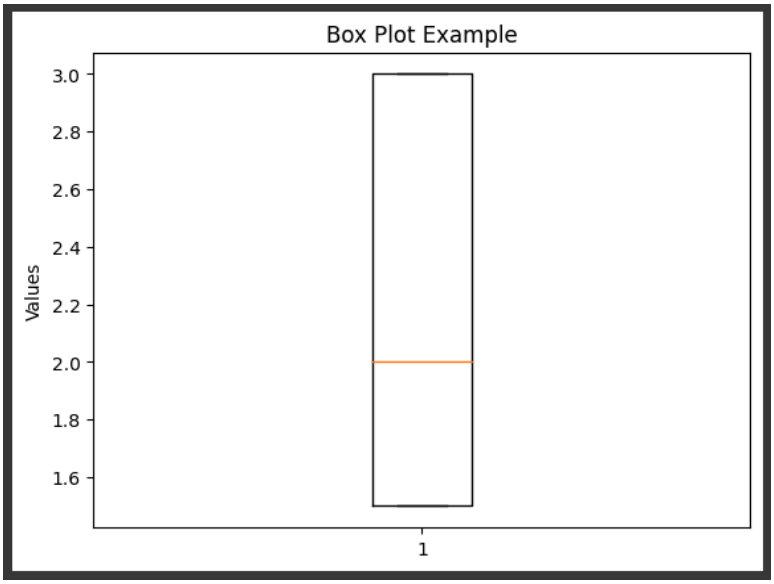
4. Bar Chart

```
import matplotlib.pyplot as plt

categories = ['A', 'B', 'C']
values = [5, 7, 3]
plt.bar(categories, values)
plt.title("Bar Chart Example")
plt.xlabel("Categories")
plt.ylabel("Values")
plt.show()
```



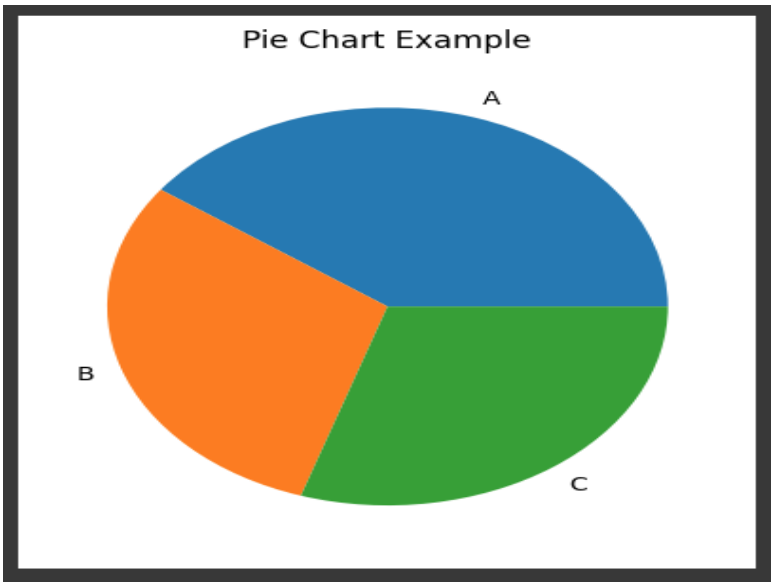
5. Box plot

```
import matplotlib.pyplot as plt

data = [[1.5]*10 + [2]*10 + [3]*10]

plt.boxplot(data)

plt.title("Box Plot Example")
plt.ylabel("Values")
plt.show()
```



6. Pie chart

```
import matplotlib.pyplot as plt

labels = ['A', 'B', 'C']
sizes = [40, 30, 30]

plt.pie(sizes, labels=labels)
plt.title("Pie Chart Example")
plt.show()
```



Part 2: Refer to the instructions below.

1. **Find a dataset for this activity**: Please visit Kaggle and look for a new dataset that would allow you to perform visualization and analysis using matplotlib.

```
] #Find a dataset for this activity
  import pandas as pd

  from google.colab import files
  uploaded = files.upload()
```

```
Choose Files  ai_job_mar... Insights.csv
• ai_job_market_insights - AI Job Market Insights.csv(text/csv) - 44723 bytes, last modified: 10/22/2024 - 100% done
Saving ai_job_market_insights - AI Job Market Insights.csv to ai_job_market_insights - AI Job Market Insights (1).csv
```

2. **Creating a dataframe from your CSV file:** Once you have successfully loaded your dataset, you need to create a dataframe from your uploaded CSV file

```
#Creating a dataframe from your CSV file
data = pd.read_csv('ai_job_market_insights - AI Job Market Insights.csv')
df = pd.DataFrame(data)
df
```

	Job_Title	Industry	Company_Size	Location	AI_Adoption_Level	Aut
0	Cybersecurity Analyst	Entertainment	Small	Dubai	Medium	
1	Marketing Specialist	Technology	Large	Singapore	Medium	
2	AI Researcher	Technology	Large	Singapore	Medium	
3	Sales Manager	Retail	Small	Berlin	Low	
4	Cybersecurity Analyst	Entertainment	Small	Tokyo	Low	
...	...	...	...	...	...	

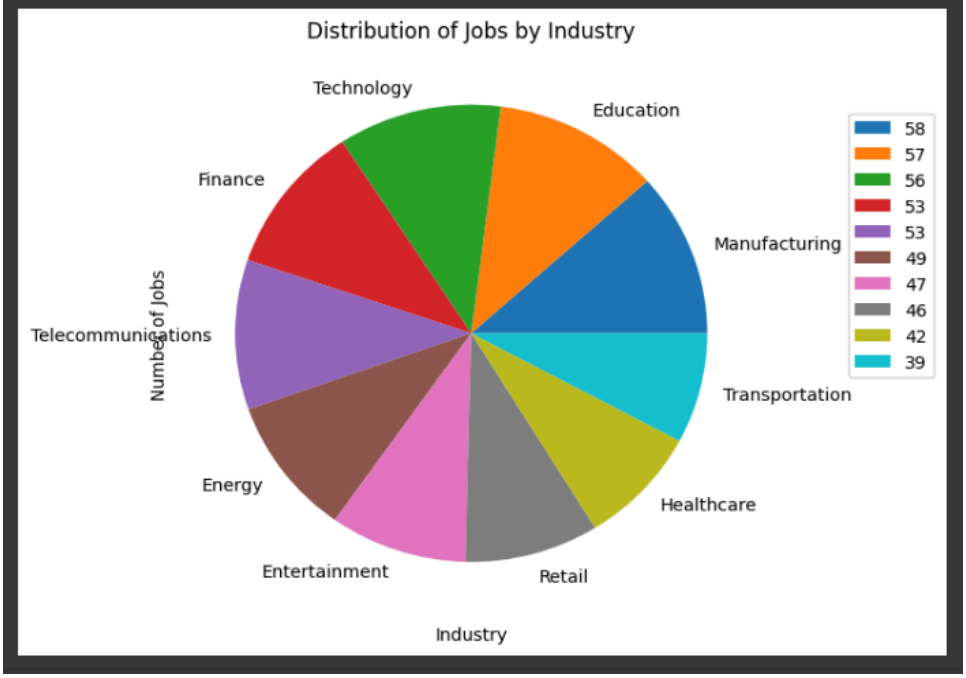
3. **Import the matplotlib.pyplot**

```
#Import the matplotlib.pyplot
import matplotlib.pyplot as plt
```

4. **Based on your chosen dataset, you will develop three questions that you will answer using pyplot visualizations. This means that you will need to produce at least three pyplot visualizations. You are also required to make certain customizations on your data vizes.**

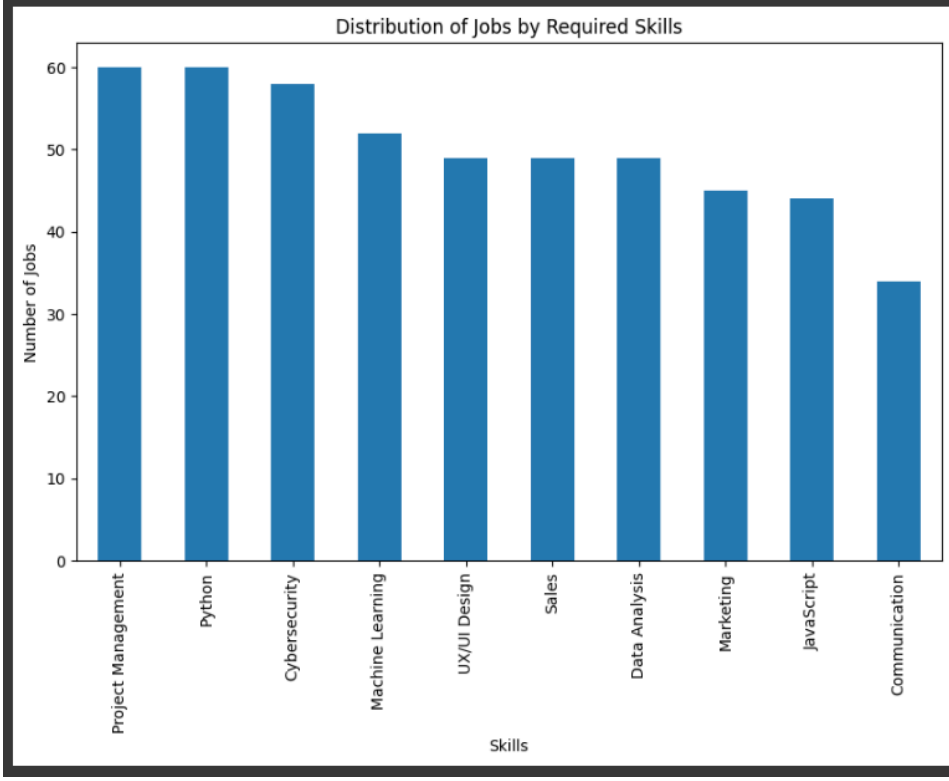
Question: What is the foremost job by industry in the diagram?

```
import matplotlib.pyplot as plt

industry = df['Industry'].value_counts()
plt.figure(figsize = (10, 6))
industry.plot(kind = 'pie')
plt.title('Distribution of Jobs by Industry')
plt.legend(labels = industry, loc = "upper right", bbox_to_anchor = (1.3, 0.9))
plt.xlabel('Industry')
plt.ylabel('Number of Jobs')
plt.xticks(rotation = 90)
plt.show()
```



Question: What is the foremost skill required for the job in the diagram?

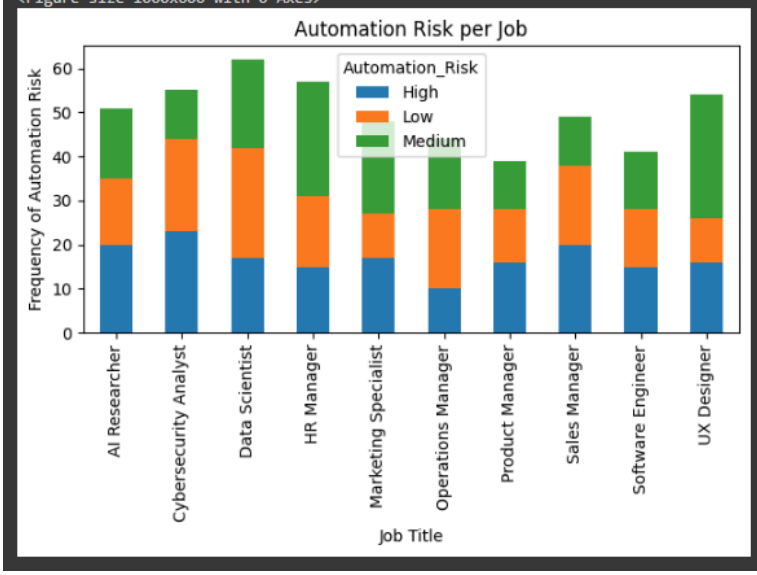
```
import matplotlib.pyplot as plt

skills = df['Required_Skills'].value_counts()
plt.figure(figsize = (10, 6))
skills.plot(kind = 'bar')
plt.title('Distribution of Jobs by Required Skills')
plt.xlabel('Skills')
plt.ylabel('Number of Jobs')
plt.xticks(rotation = 90)
plt.show()
```



Question: Which of the job has the highest “high” automation risk in the diagram?

```
import matplotlib.pyplot as plt

risk = df[['Job_Title', 'Automation_Risk']]
plt.figure(figsize = (10, 6))
risk.groupby('Job_Title')['Automation_Risk'].value_counts().unstack().plot(kind = 'bar', stacked = True)
plt.title('Automation Risk per Job')
plt.xlabel('Job Title')
plt.ylabel('Frequency of Automation Risk')
plt.xticks(rotation = 90)
plt.tight_layout()
plt.show()
```

```
<Figure size 1000x600 with 0 Axes>
```



5. Provide observations for each of your data viz, then **produce one insight not longer than five sentences given your three observations**. Your output shall follow this outline:
    a. Introduction (Describe your dataset)
    b. Questions
    c. Visualization and Observation
    d. Insight

    - The dataset that was utilized in this activity was about a job in each artificial intelligence industry. Each of the category in the data set contain information about the job, such as the job, industry, company size, location, ai adoption level, and much more. The questions that were formed are “What is the foremost job by industry in the diagram?”, “What is the foremost skill required for the job in the diagram?”, and “Which of the job has the highest “high” automation risk in the diagram?”.The observation for the visualization are manufacturing has the lead for job by industry, project management and python is the foremost skill required in the overall industry, and cybersecurity analyst has the highest automation risk. The insight I gained from the analysis of the data is that project management and python is the most required skill for the competitive job in different industries.

6. Your grade will depend on the quality of the question, difficulty/complexity of the visualization, and value-add of the insight that you will generate.