

EDS Assignment 5

Monali Babde (205)
Swarup Divekar (217)
Pranav Falke (219)

```
1 import matplotlib.pyplot as plt
2 import pandas as pd
3 df = pd.read_csv('/content/salary.csv')
4 print(df)
5
```

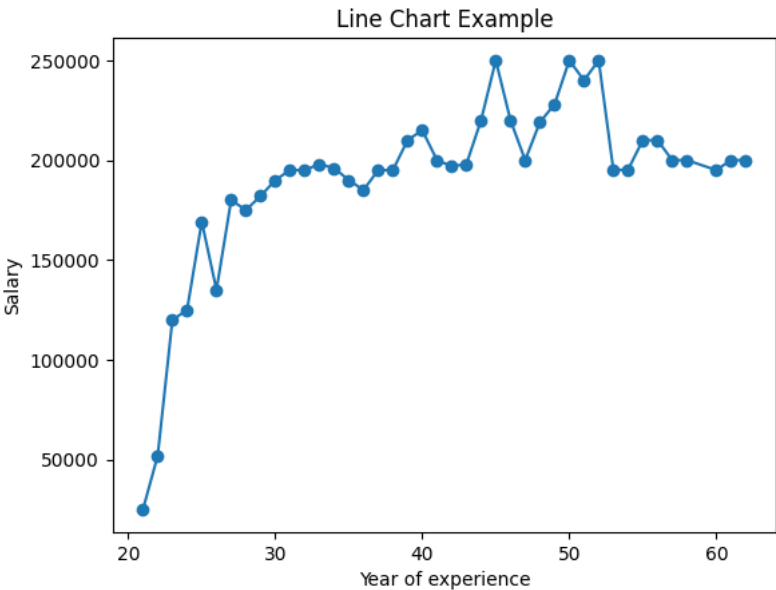
	Age	Gender	Education Level	Job Title \
0	32.0	Male	Bachelor's	Software Engineer
1	28.0	Female	Master's	Data Analyst
2	45.0	Male	PhD	Senior Manager
3	36.0	Female	Bachelor's	Sales Associate
4	52.0	Male	Master's	Director
...
6699	49.0	Female	PhD	Director of Marketing
6700	32.0	Male	High School	Sales Associate
6701	30.0	Female	Bachelor's Degree	Financial Manager
6702	46.0	Male	Master's Degree	Marketing Manager
6703	26.0	Female	High School	Sales Executive

	Years of Experience	Salary
0	5.0	90000.0
1	3.0	65000.0
2	15.0	150000.0
3	7.0	60000.0
4	20.0	200000.0
...
6699	20.0	200000.0
6700	3.0	50000.0
6701	4.0	55000.0
6702	14.0	140000.0
6703	1.0	35000.0

[6704 rows x 6 columns]

```
1 df1 = df.groupby('Age').max()
2
3 plt.plot(df1.index, df1['Salary'], marker='o')
4
5 # Customize the chart
6
7 plt.title("Line Chart Example")
8 plt.xlabel("Year of experience")
9 plt.ylabel("Salary")
10
11 # Display the chart
12 plt.show()
```

```
<ipython-input-60-4a420bcc2724>:1: FutureWarning: Dropping invalid columns in DataFrame
df1 = df.groupby('Age').max()
```

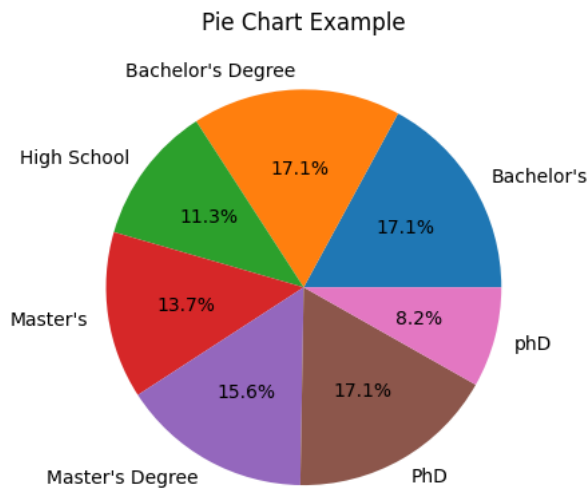


```
1 import matplotlib.pyplot as plt
2
3 # Example data
4
5
6 df1 = df.groupby('Education Level').max()
7
```

```

8 # Plotting the pie chart
9 plt.pie(df1['Salary'],labels=df1.index, autopct='%1.1f%%')
10
11 # Adding a title
12 plt.title('Pie Chart Example')
13
14 # Displaying the pie chart
15 plt.show()

```



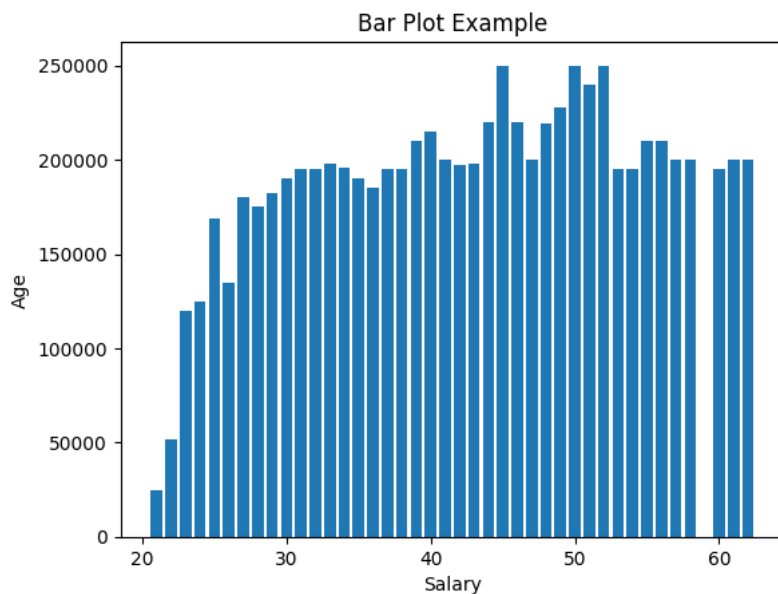
```

1 import matplotlib.pyplot as plt
2
3 # plotting the line 1 points
4 df1 = df.groupby('Age').max()
5 plt.plot(df1.index, df1['Salary'], label = "line 1", marker='o')
6
7 df2 = df.groupby('Education Level').min()
8
9 # plotting the line 2 points
10 plt.plot(df2.index, df2['Salary'], label = "line 1", marker='o')
11
12 # naming the x axis
13 plt.xlabel('Salary')
14 # naming the y axis
15 plt.ylabel('Education Level')
16 # giving a title to my graph
17 plt.title('Two lines on same graph!')
18
19 # show a legend on the plot
20 plt.legend()
21
22 # function to show the plot
23 plt.show()

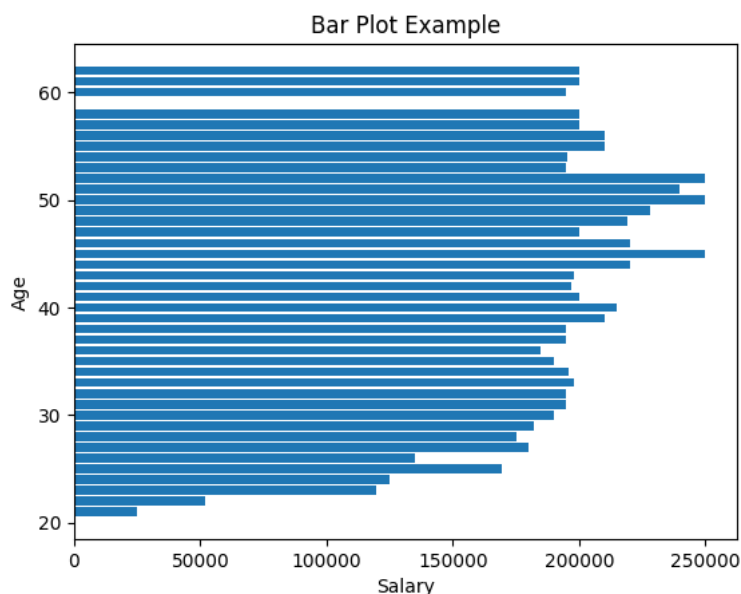
```

```
<ipython-input-7-df6bfb4fe275>:4: FutureWarning: Dropping invalid columns in DataFrame
df1 = df.groupby('Age').max()
```

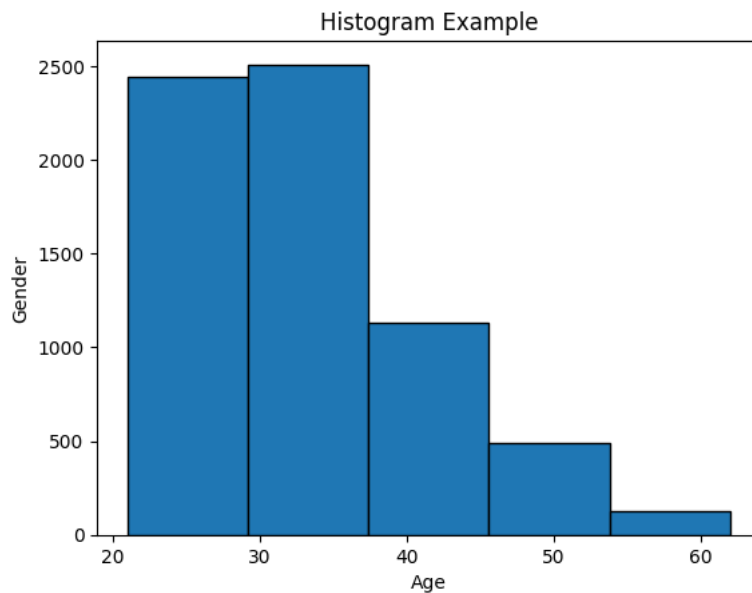
```
1 import matplotlib.pyplot as plt
2 x = df['Age']
3 y = df['Salary']
4 # Create a bar plot
5 plt.bar(x,y)
6 # plt.bar(Salary, Age) # Customize the plot
7 plt.title("Bar Plot Example")
8
9 plt.xlabel("Salary")
10
11 plt.ylabel("Age")
12
13 # Display the plot
14 plt.show()
```



```
1 import matplotlib.pyplot as plt
2 x = df['Age']
3 y = df['Salary']
4 # Create a bar plot
5 plt.barh(x,y)
6 # plt.bar(Salary, Age) # Customize the plot
7 plt.title("Bar Plot Example")
8
9 plt.xlabel("Salary")
10
11 plt.ylabel("Age")
12
13 # Display the plot
14 plt.show()
```



```
1 import matplotlib.pyplot as plt
2
3 # Example data
4
5 b=df["Age"]
6 #Plotting the histogram
7 plt.hist(b, bins=5, edgecolor="black")
8
9 # Adding labels and title
10 plt.xlabel('Age')
11 plt.ylabel('Gender')
12 plt.title('Histogram Example')
13
14 # Displaying the histogram
15 plt.show()
```

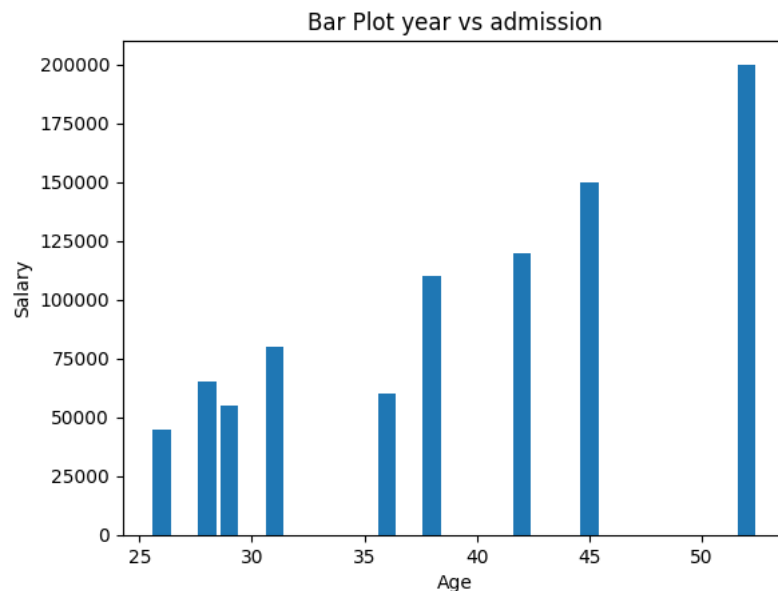


```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 df1 = df.groupby('Age').max()
4 plt.plot(df1.index, df1['Salary'], marker='o')
5
6
7 plt.title("Salary V/s Age ")
8 plt.xlabel("Salary")
9 plt.ylabel("Age")
10 plt.grid()
11
12 plt.show()
```

```

1 a=df["Age"]
2 b=df["Salary"]
3 c=a.iloc[1:10]
4 d=b.iloc[1:10]
5 plt.bar(c,d)
6
7 # Customize the plot
8 plt.title("Bar Plot year vs admission")
9 plt.xlabel("Age")
10 plt.ylabel("Salary")
11
12 # Display the plot
13 plt.show()
14

```



```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv('/content/salary.csv')
4 df = pd.DataFrame(data)
5
6 age = df['Age']
7 overall = df['Salary']
8 # Create scatter plot
9 plt.scatter(age, overall, alpha=0.5)
10 # Set axis labels
11 plt.xlabel('Age')
12 plt.ylabel('Salary')
13 # Set plot title
14 plt.title('Salary : Age vs Salary')
15 # Show the plot
16 plt.show()

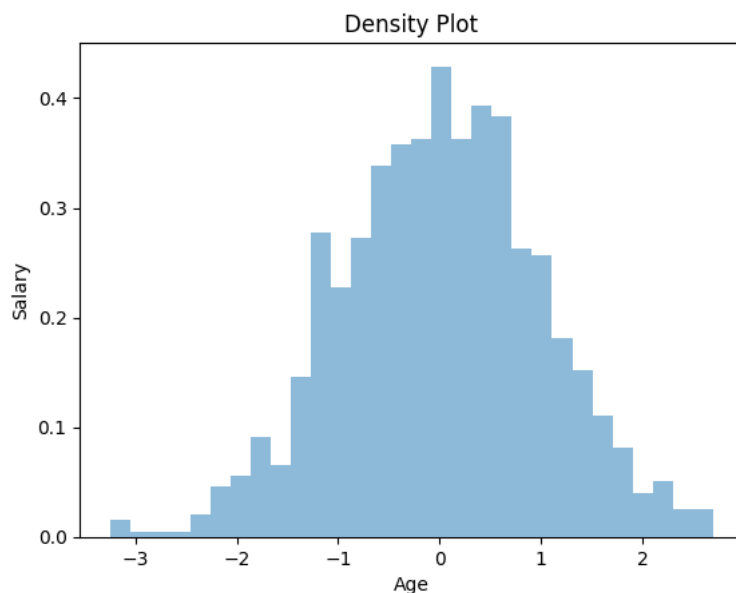
```

Salary · Age vs Salary

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3
4
5 # Create a sample DataFrame
6 data = pd.read_csv('/content/salary.csv')
7 df = pd.DataFrame(data)
8
9 # Generate some random data
10 data = np.random.randn(1000)
11
12 # Create a density plot
13
14 plt.hist(data, density=True, bins=30, alpha=0.5)
15
16 # # Create a density plot
17 # df.plot.kde()
18
19 # Add labels and title
20 plt.xlabel('Age')
21 plt.ylabel('Salary')
22 plt.title('Density Plot')
23
24 # Show the plot
25 plt.show()

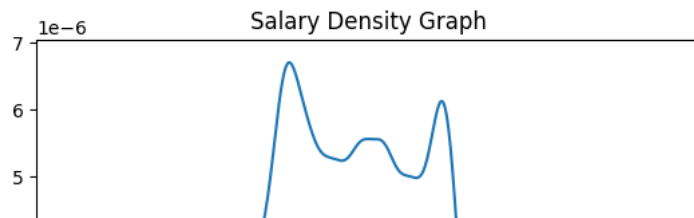
```



```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 data = pd.read_csv('/content/salary.csv')
5 df['Salary'].plot.kde()
6
7 plt.xlabel('Salary')
8 plt.ylabel('Density')
9 plt.title('Salary Density Graph')
10
11 plt.show()
12

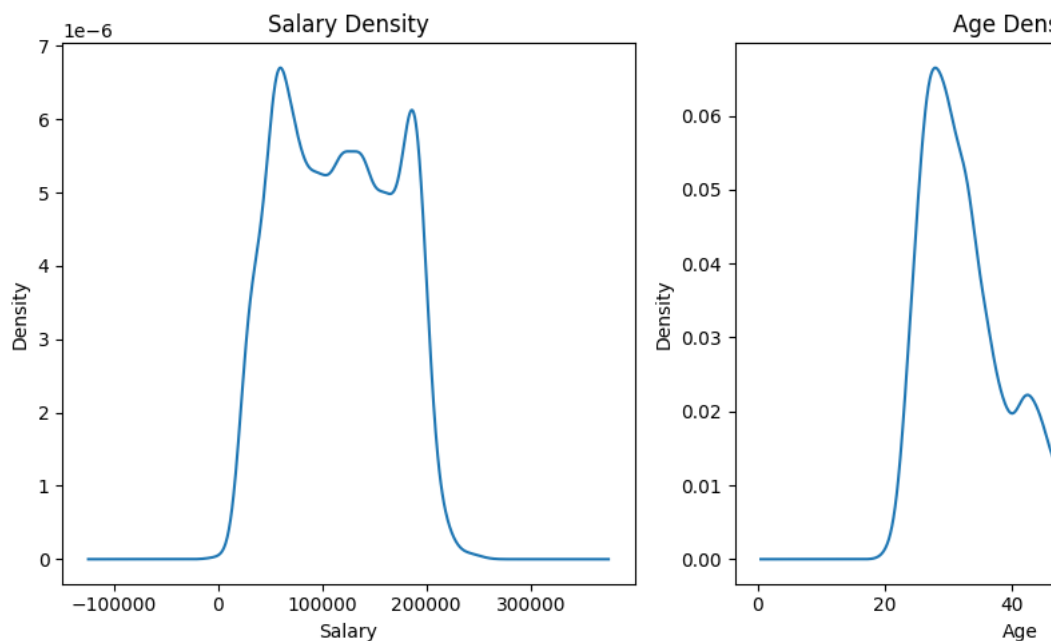
```



```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the salary dataset into a DataFrame
5 data = pd.read_csv('/content/salary.csv')
6
7 # Create a figure and two subplots
8 fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(10, 5))
9
10 # Plot the density graph for the first subplot
11 axes[0].set_title('Salary Density')
12 df['Salary'].plot.kde(ax=axes[0])
13 axes[0].set_xlabel('Salary')
14 axes[0].set_ylabel('Density')
15
16 # Plot the density graph for the second subplot
17 axes[1].set_title('Age Density')
18 df['Age'].plot.kde(ax=axes[1])
19 axes[1].set_xlabel('Age')
20 axes[1].set_ylabel('')
21
22 # Adjust the spacing between subplots
23 plt.tight_layout()
24
25 # Display the graph
26 plt.show()
27

```



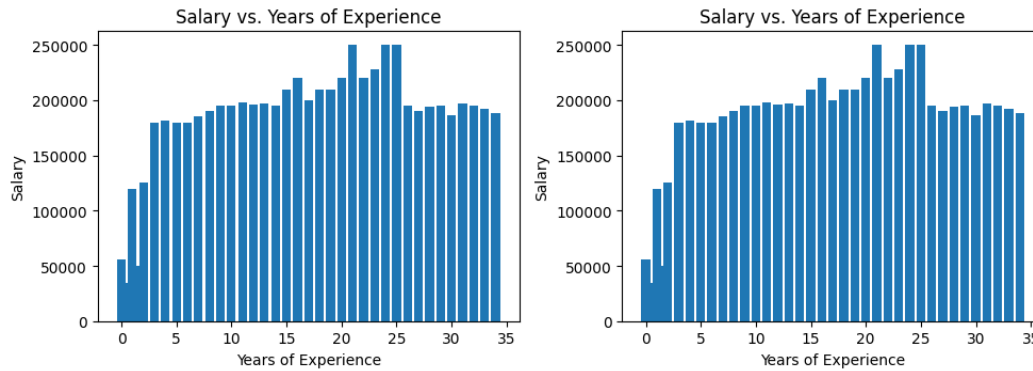
```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data = pd.read_csv('/content/salary.csv')
4 fig, axes = plt.subplots(1, 2, figsize=(10, 4))
5 # First subplot
6 axes[0].bar(data['Years of Experience'], data['Salary'])
7 axes[0].set_xlabel('Years of Experience')
8 axes[0].set_ylabel('Salary')
9 axes[0].set_title('Salary vs. Years of Experience')
10
11 # Second subplot
12 axes[1].bar(data['Years of Experience'], data['Salary'])
13 axes[1].set_xlabel('Years of Experience')
14 axes[1].set_ylabel('Salary')
15 axes[1].set_title('Salary vs. Years of Experience')
16 fig.suptitle('Salary Analysis')
17 plt.tight_layout()

```

```
18 plt.show()
19
```

Salary Analysis



```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the dataset
5 data = pd.read_csv('/content/salary.csv')
6
7 # Create a panel of subplots
8 fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(10, 8))
9
10 # Subplot 1: Histogram of salaries
11 axes[0, 0].hist(data['Salary'], bins=10)
12 axes[0, 0].set_xlabel('Salary')
13 axes[0, 0].set_ylabel('Frequency')
14 axes[0, 0].set_title('Salary Distribution')
15
16 # Subplot 2: Scatter plot of salary vs. years of experience
17 axes[0, 1].scatter(data['Years of Experience'], data['Salary'])
18 axes[0, 1].set_xlabel('Years of Experience')
19 axes[0, 1].set_ylabel('Salary')
20 axes[0, 1].set_title('Salary vs. Years of Experience')
21
22 # Subplot 3: Box plot of salaries by education level
23 data.boxplot(column='Salary', by='Education Level', ax=axes[1, 0])
24 axes[1, 0].set_xlabel('Education Level')
25 axes[1, 0].set_ylabel('Salary')
26 axes[1, 0].set_title('Salary by Education Level')
27
28 # Subplot 4: Line plot of average salary by years of experience
29 avg_salary_by_exp = data.groupby('Years of Experience')['Salary'].mean()
30 axes[1, 1].plot(avg_salary_by_exp.index, avg_salary_by_exp.values)
31 axes[1, 1].set_xlabel('Years of Experience')
32 axes[1, 1].set_ylabel('Average Salary')
33 axes[1, 1].set_title('Average Salary by Years of Experience')
34
35 # Adjust the layout
36 plt.tight_layout()
37
38 # Display the plot
39 plt.show()
40
```