Plotting Graphs using Matplotlib

What is Matplotlib?

- Matplotlib is a low level graph plotting library in python that serves as a visualization utility.
- Pandas uses the `plot()` method to create diagrams.
- We can use Pyplot, a submodule of the Matplotlib library to visualize the diagram on the screen.

1.Importing the Pandas and Matplotlib:

```
import pandas as pd
import matplotlib.pyplot as plt
```

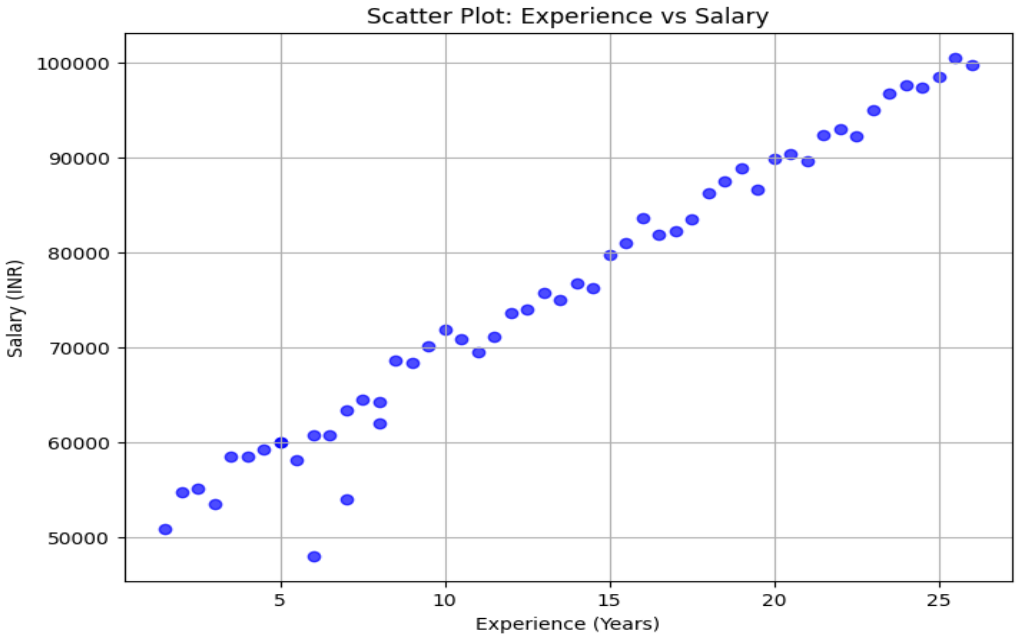
2.Reading CSV file:

```
# Load the CSV file
data =
pd.read_csv('/content/updated_salary_dataset.csv')
```

3.Visualization of data using different graphs:

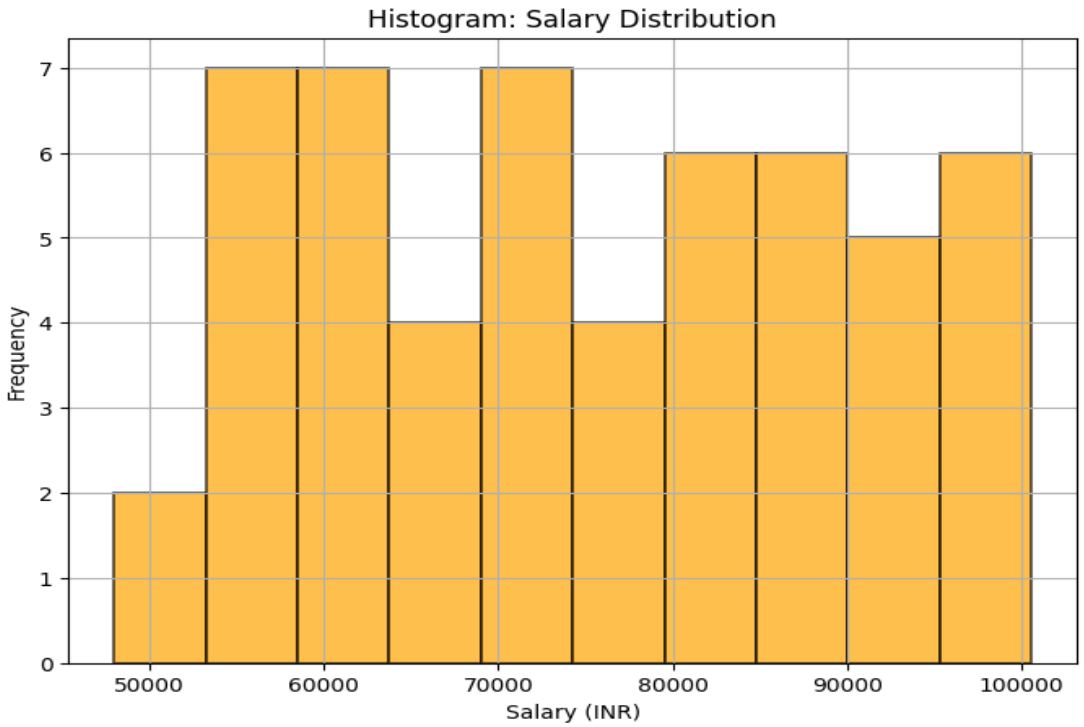
a) Scatter Plot:

```
# Scatter plot: Experience vs Salary
plt.figure(figsize=(8,6))
plt.scatter(data['Experience'], data['Salary'],
color='b', alpha=0.7)
plt.title('Scatter Plot: Experience vs Salary')
plt.xlabel('Experience (Years)')
plt.ylabel('Salary (INR)')
plt.grid(True)
plt.show()
```



b) Histogram:

```
# Histogram: Salary distribution
plt.figure(figsize=(8,6))
plt.hist(data['Salary'], bins=10,
color='orange', alpha=0.7,
edgecolor='black', linewidth=1.2)
plt.title('Histogram: Salary Distribution')
plt.xlabel('Salary (INR)')
plt.ylabel('Frequency')
plt.grid(True)
```

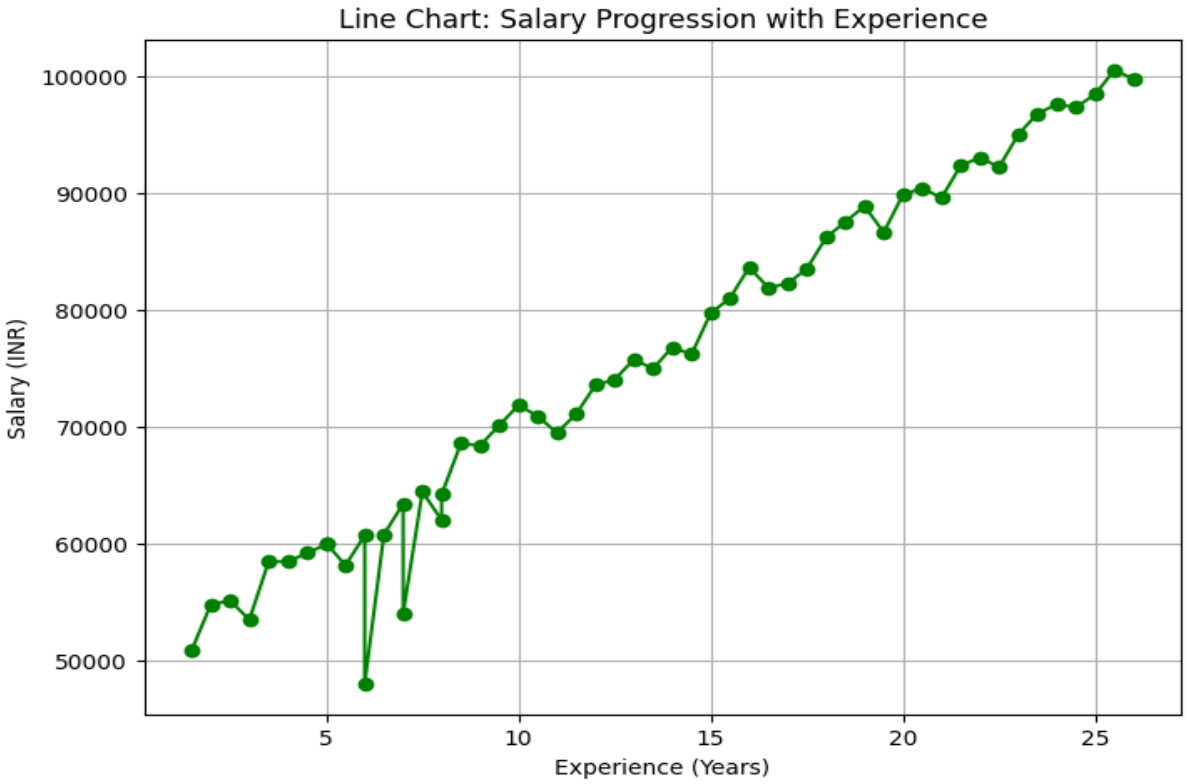



```
plt.show()
```

c) Line Chart:

```
# Line Chart: Salary progression based on Experience
data_sorted = data.sort_values('Experience')
plt.figure(figsize=(8,6))
plt.plot(data_sorted['Experience'],
data_sorted['Salary'], marker='o', linestyle='-',
color='g')
plt.title('Line Chart: Salary Progression with
Experience')
plt.xlabel('Experience (Years)')
plt.ylabel('Salary (INR)')
plt.grid(True)
```

```
plt.show()
```



d) Bar Graph:

df_groups = df.groupby(['group_var'])['values_var'].sum()/mean()

```
# Bar Graph: Average Salary by Role
avg_salary_by_role = data.groupby('Role')['Salary'].mean()
plt.figure(figsize=(8,6))
avg_salary_by_role.plot(kind='bar', color='purple', alpha=0.7)
plt.title('Bar Graph: Average Salary by Role')
plt.xlabel('Role')
plt.ylabel('Average Salary (INR)')
plt.grid(True)
plt.show()
```

```
# importing packages
import seaborn as sb
# load dataset
df = sb.load_dataset('tips')
# perform groupby
df = df.groupby(['size',
'day']).agg(mean_total_bill=("total_bill", 'mean'))
df = df.reset_index()
# plot barplot
sb.barplot(x="size",
    y="mean_total_bill",
    hue="day",
    data=df)
```