```
import os

# Create folders for outputs
os.makedirs('results', exist_ok=True)
os.makedirs('data/processed', exist_ok=True)
```

```
# Install dependencies from requirements.txt
!pip install -r https://raw.githubusercontent.com/sonakshijais1401/research1_workflow_project/main/requirements.txt

# Optional: mount Google Drive if you use large datasets
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

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

Choose Files project.csv

- **project.csv**(text/csv) - 342 bytes, last modified: 4/18/2025 - 100% done

```
Saving project.csv to project (1).csv
```

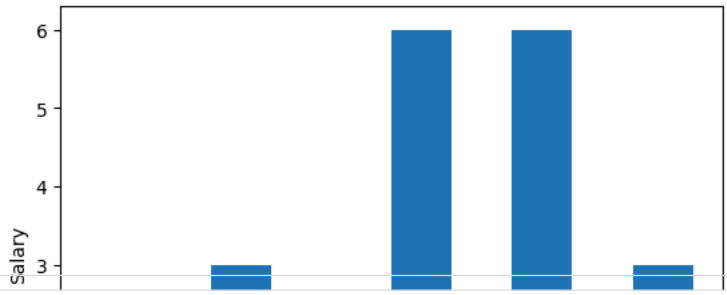
```
import pandas as pd

df = pd.read_csv('project.csv')  # use the exact uploaded filename
df.head()
```

	S.No.	Salary	P. Rating	Department
0	1	35000	9	IT
1	2	50000	7	ECE
2	3	67000	5	MCA
3	4	42000	6	CSE
4	5	34000	8	MBA

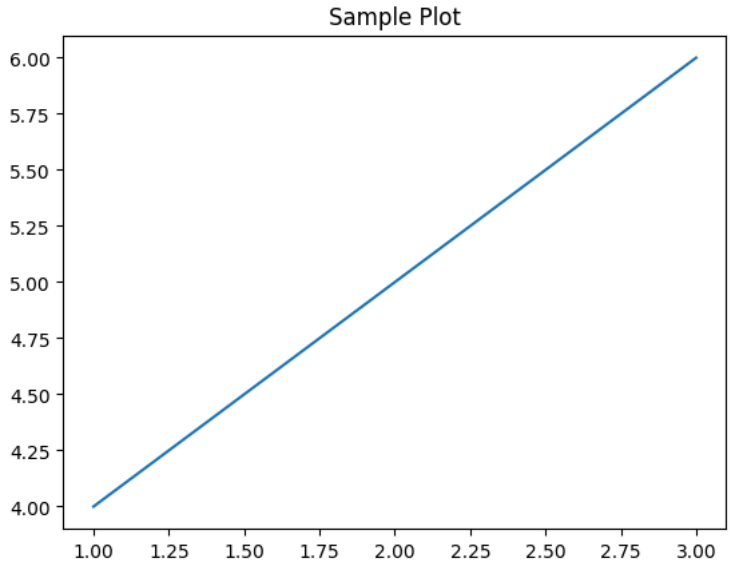
Next steps: Generate code with df | View recommended plots | New interactive sheet

```
# Check for missing values
df.isnull().sum()

# Drop rows with missing data (if needed)
df = df.dropna()

# Reset index after dropping rows
df = df.reset_index(drop=True)
```

```
# Describe data
df.describe()

# Simple plot
import matplotlib.pyplot as plt

plt.hist(df['P. Rating'])
plt.xlabel('Department')
plt.ylabel('Salary')
plt.show()
```




```
import matplotlib.pyplot as plt

# Example plot
plt.plot([1, 2, 3], [4, 5, 6])
plt.title("Sample Plot")

# Save it to Colab's results folder
plt.savefig('results/sample_plot.png')
plt.show()
```




```
import pandas as pd

# Example DataFrame
df = pd.DataFrame({'A':[1,2,3], 'B':[4,5,6]})

# Save it to Colab's data/processed folder
df.to_csv('data/processed/cleaned_data.csv', index=False)
```