GROUP MEMBERS

- **Ayush Patel** **447**
- **Shashank Patokar** **449**
- **Sudhanshu Pendor** **452**
- **Sameer Sabale** **456**

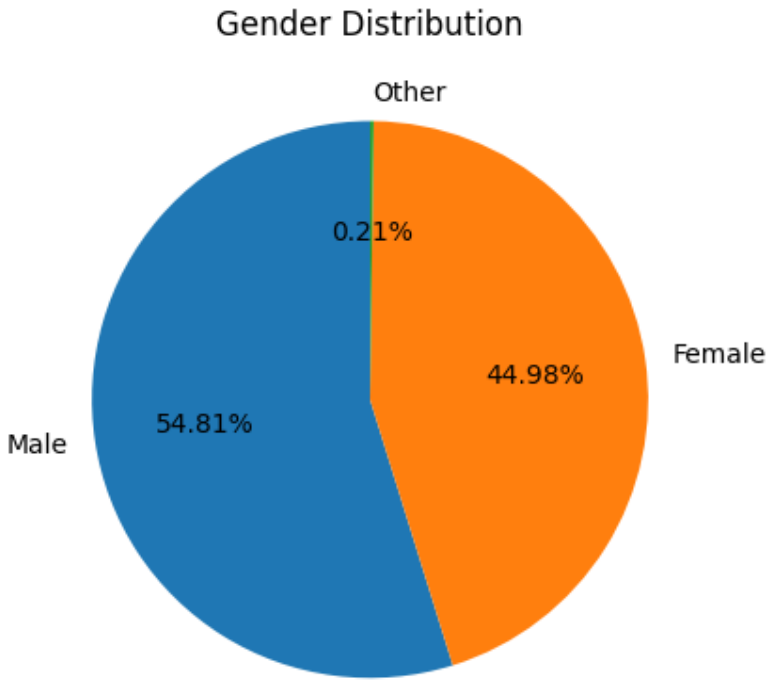
Project Code

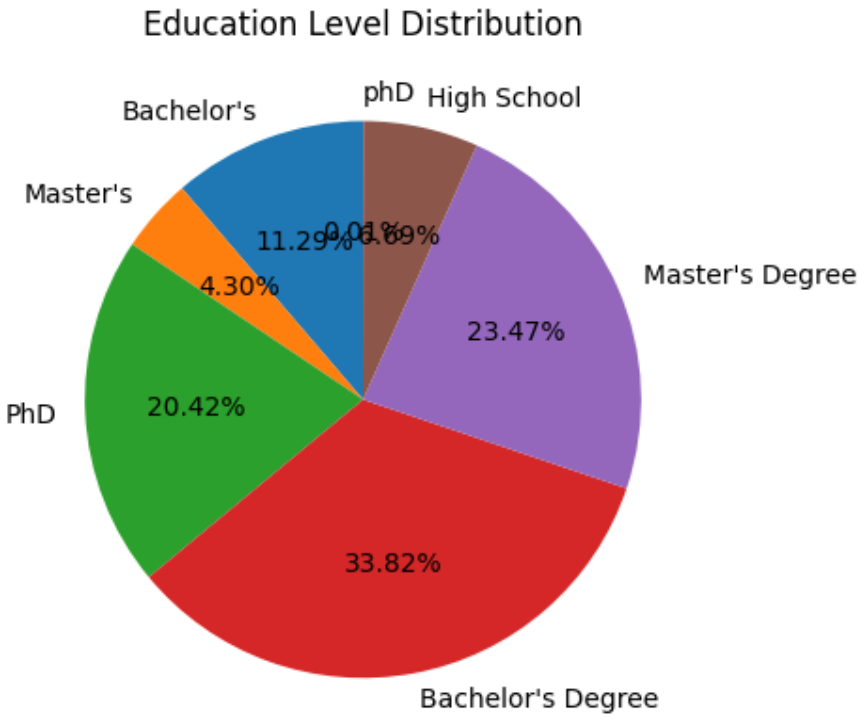
```
import pandas as pd
import matplotlib.pyplot as plt

# Read CSV file
df = pd.read_csv("Salary_Data.csv")

#-----Cleaning the data---------#

# Deletes all the rows whose columns have na value or are empty
df.dropna(subset=["Age", "Gender", "Education Level", "Job Title", "Years of
Experience", "Salary"], inplace= True)

# Calculating mean for replacing NaN values or missing values
age_mean = df["Age"].mean()
years_mean = df["Years of Experience"].mean()
salary_mean = df["Salary"].mean()

# Replacing NaN or missing values with the mean value calculated above
df.fillna({"Age":age_mean, "Years of Experience":years_mean,
"Salary":salary_mean}, inplace=True)

#-----Plotting the graph-----------#

#distribution based on gender
label = df['Gender'].unique()
values = [df[df['Gender'] == 'Male'].shape[0], df[df['Gender'] ==
'Female'].shape[0], df[df['Gender'] == 'Other'].shape[0]]
plt.pie(values, labels=label, startangle= 90, autopct='%1.2f%%')
plt.title('Gender Distribution')
plt.show()

#distribution based on education level
label = df['Education Level'].unique()
values = [df[df['Education Level']== "Bachelor's"].shape[0], df[df['Education
Level']== "Master's"].shape[0], df[df['Education Level']==
"PhD"].shape[0],df[df['Education Level']== "Bachelor's Degree"].shape[0],
df[df['Education Level']== "Master's Degree"].shape[0], df[df['Education
Level']== "High School"].shape[0], df[df['Education Level']== "phD"].shape[0]]
```

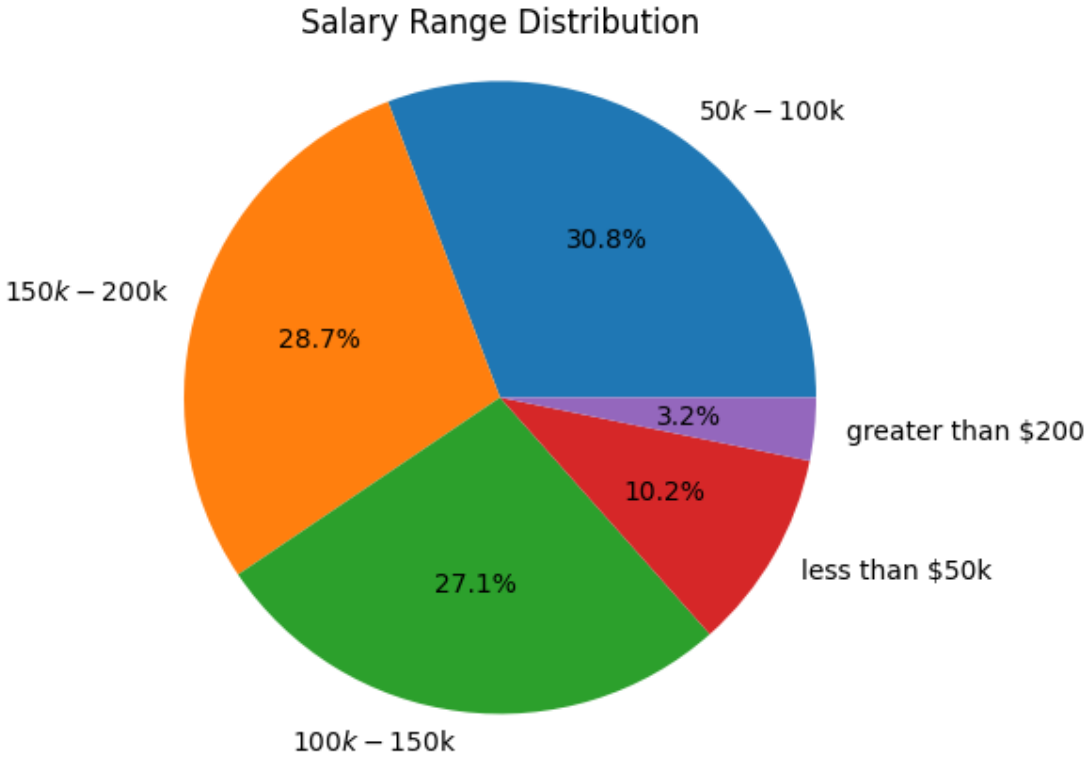
```
plt.pie(values,labels=label, startangle= 90, autopct='%1.2f%%')
plt.title('Education Level Distribution')
plt.show()

#distribution based on salary
ranges = [0, 50000, 100000, 150000, 200000, float('inf')]
labels = ['less than $50k', '$50k - $100k', '$100k - $150k', '$150k - $200k',
'greater than $200k']
salary_counts = pd.cut(df['Salary'], ranges, labels=labels,
right=False).value_counts()
plt.pie(salary_counts, labels=salary_counts.index, autopct='%1.1f%%')
plt.title('Salary Range Distribution')
plt.axis('equal')
plt.show()

#distribution based on years of expericene

ranges = [0, 2, 5, 10, 15, float('inf')]
labels = ['less than 2 years', '2 - 5 years', '5 - 10 years', '10 - 15 years',
'greater than 15 years']
exp_counts = pd.cut(df['Years of Experience'], ranges, labels=labels,
right=False).value_counts()
plt.pie(exp_counts, labels=exp_counts.index, autopct='%1.1f%%')
plt.title('Years of Experience Distribution')
plt.axis('equal')
plt.show()

#distributon basedo on age
ranges = [20, 30, 40, 50, 60, 70]
labels = ['20-29', '30-39', '40-49', '50-59', '60-69']
age_counts = pd.cut(df['Age'], ranges, labels=labels,
right=False).value_counts()
plt.pie(age_counts, labels=age_counts.index, autopct='%1.1f%%')
plt.title('Age Distribution')
plt.axis('equal')
plt.show()

# Age vs. Salary (Scatter plot)
plt.scatter(df['Age'], df['Salary'])
plt.xlabel('Age')
plt.ylabel('Salary')
plt.title('Age vs. Salary')
plt.show()

# Education vs. Salary (Bar plot)
df.groupby('Education Level')['Salary'].mean().plot(kind='bar')
plt.xlabel('Education Level')
plt.ylabel('Salary')
```

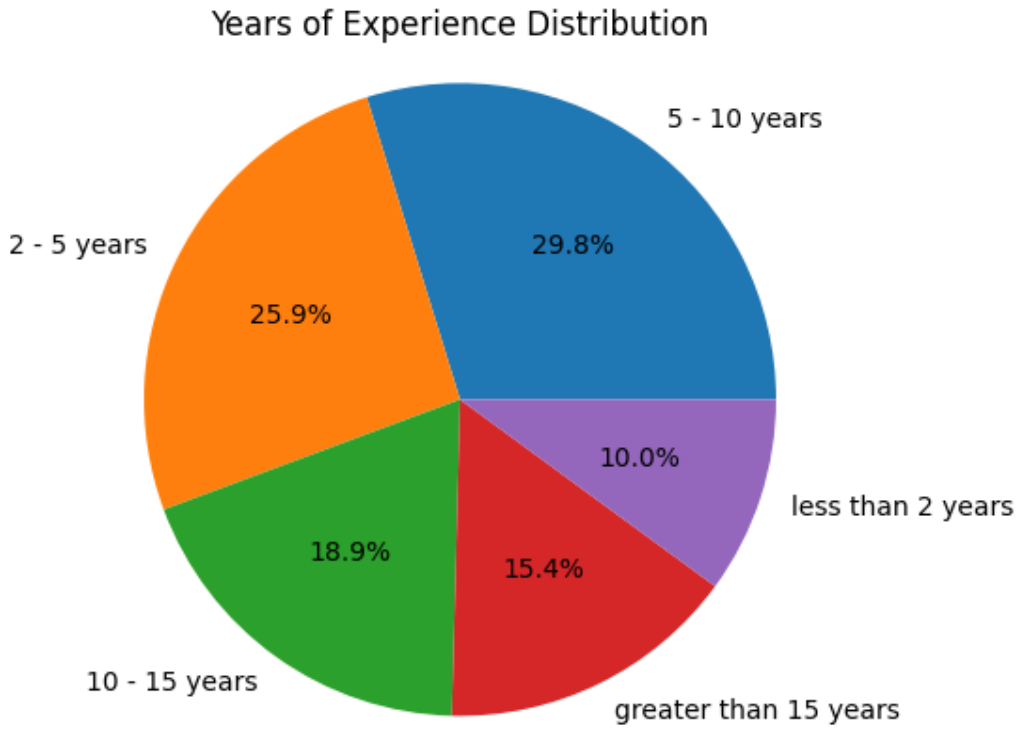
```
plt.title('Education Level vs. Salary')
plt.show()

# Years of Experience vs. Salary (Line plot)
df.groupby('Years of Experience')['Salary'].mean().plot(kind='line',
marker='o')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.title('Years of Experience vs. Salary')
plt.show()
```

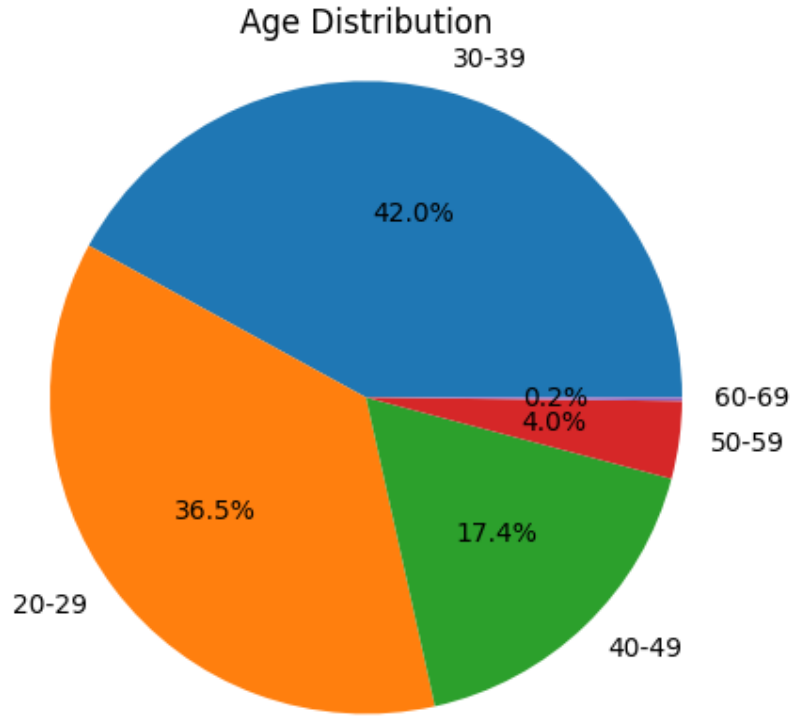
OUTPUT GRAPHS

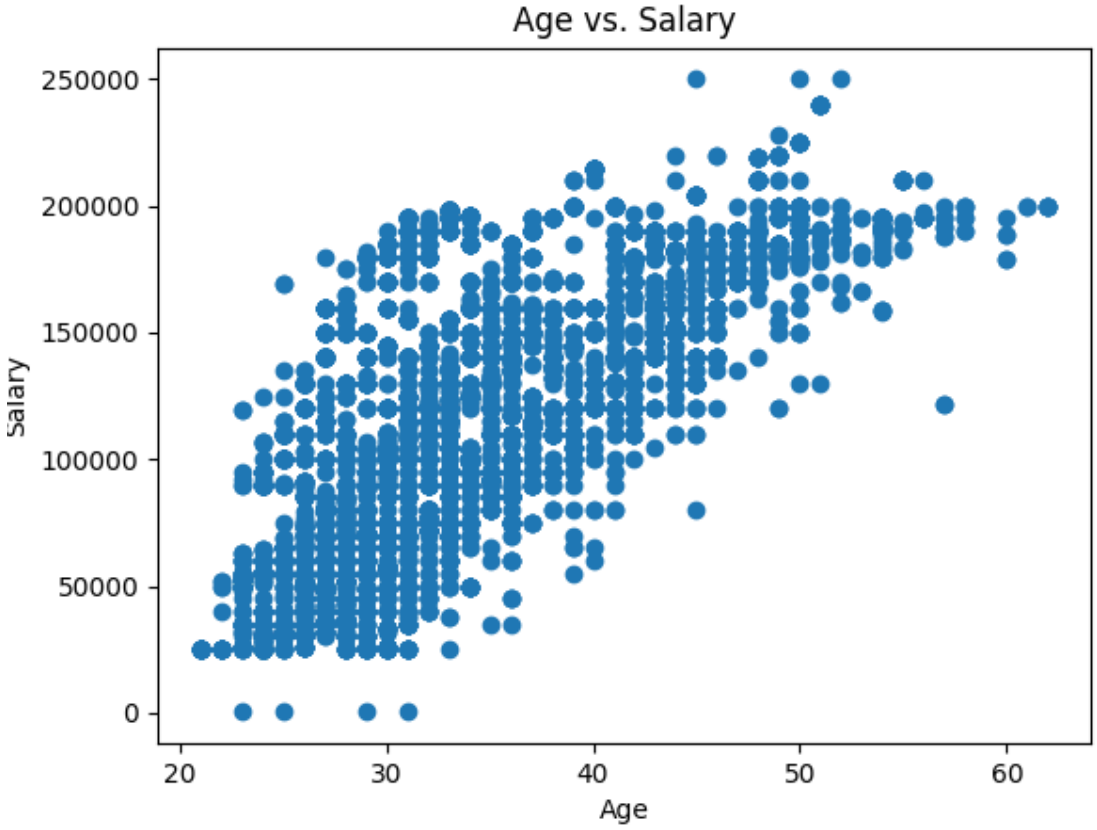


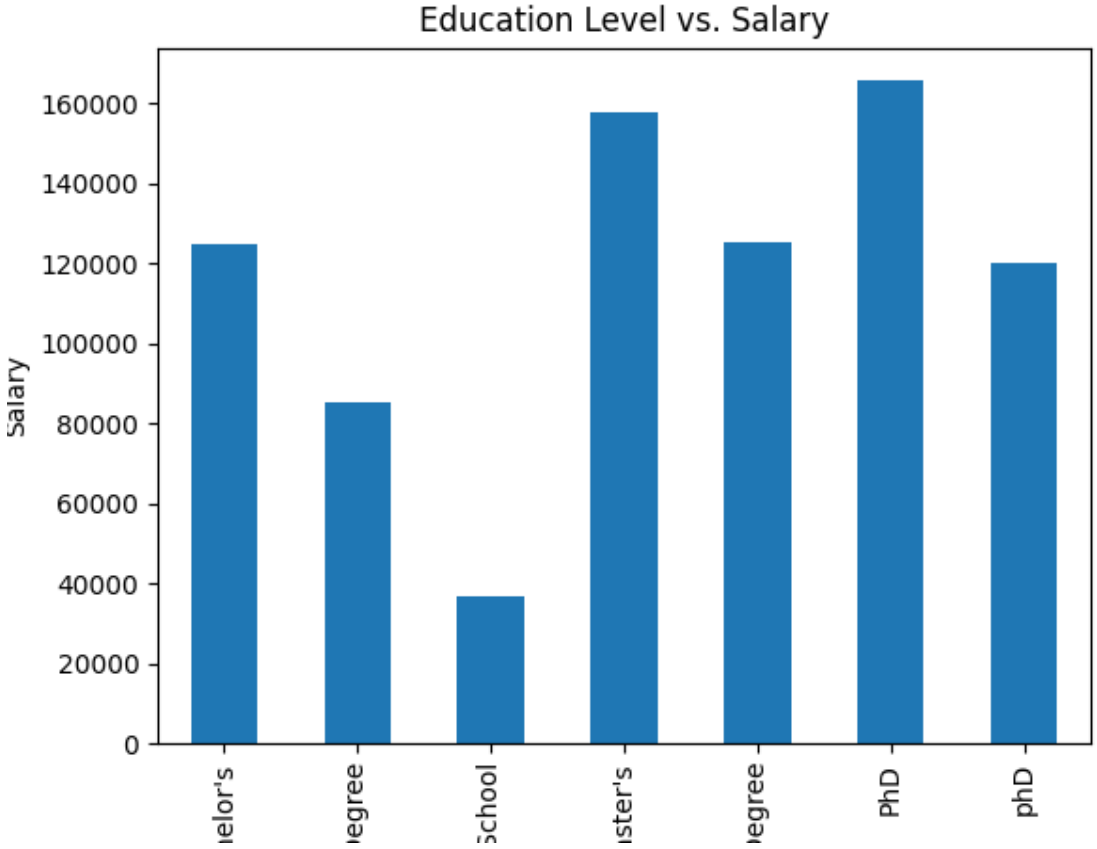

Education Level Distribution
Bachelor's
phD
High School
Master's
11.29%
0.01%
6.69%
4.30%
Master's Degree
23.47%
PhD
20.42%
33.82%
Bachelor's Degree


Salary Range Distribution
$50k - 100k$
30.8%
$150k - 200k$
28.7%
3.2%
greater than $200
10.2%
less than $50k
27.1%
$100k - 150k$


Years of Experience Distribution
5 - 10 years
29.8%
2 - 5 years
25.9%
10.0%
less than 2 years
18.9%
15.4%
10 - 15 years
greater than 15 years


Age Distribution
30-39
42.0%
0.2%
60-69
4.0%
50-59
36.5%
17.4%
20-29
40-49


Age vs. Salary
Salary
250000
200000
150000
100000
50000
0
20
30
40
50
60
Age


Education Level vs. Salary
Salary
160000
140000
120000
100000
80000
60000
40000
20000
0
PhD
phD


Years of Experience vs. Salary
Salary
Years of Experience
200000
175000
150000
125000
100000
75000
50000
25000
0
5
10
15
20
25
30
35