ASSIGNMENT 5

NAME : **Kamlesh Sunil Pawar**

PRN NO. : **202201070138**

ROLL NO : **379**

DIV : **C-4**

Dataset:

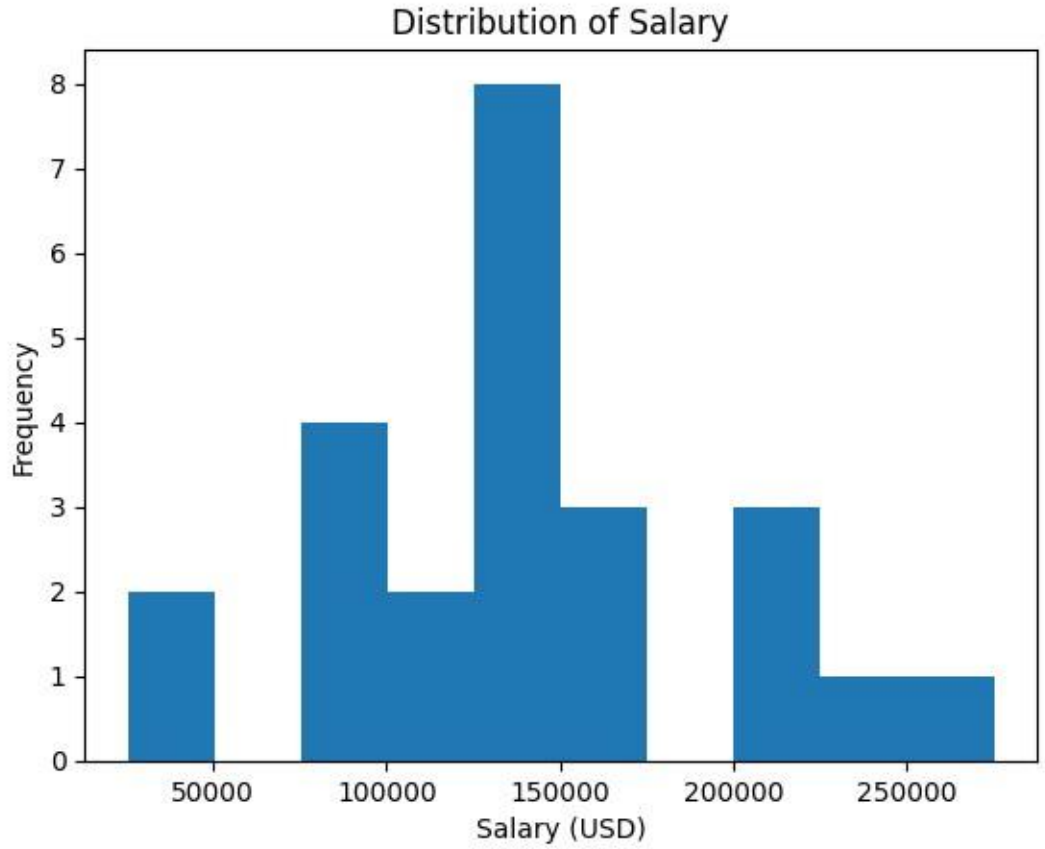
https://docs.google.com/spreadsheets/d/1E3dYzlc6bIebo4rydsWuEpdxFPVx6Moy2IjAG03xTM/edit?usp=drive_link

```
[19] import pandas as pd

     # Read the CSV file
     data = pd.read_csv('/content/company1.csv')

     # Display the data
     print(data.head())
```

```
   work_year experiNewce_level employment_type                 job_title  \
0       2023           below10            Full  Principal Data SciNewtist
1       2021           above10            Part               ML Newgineer
2       2023           above10            Part               ML Newgineer
3       2023           below10            Full            Data SciNewtist
4       2023           below10            Full            Data SciNewtist

   salary Rs  salary_in_usd employee_residNewce company_location company_size
0      80000          85847                  ES                ES            L
1      30000          30000                  US                US            S
2      25500          25500                  US                US            S
3     175000         175000                  CA                CA            M
4     120000         120000                  CA                CA            M
```

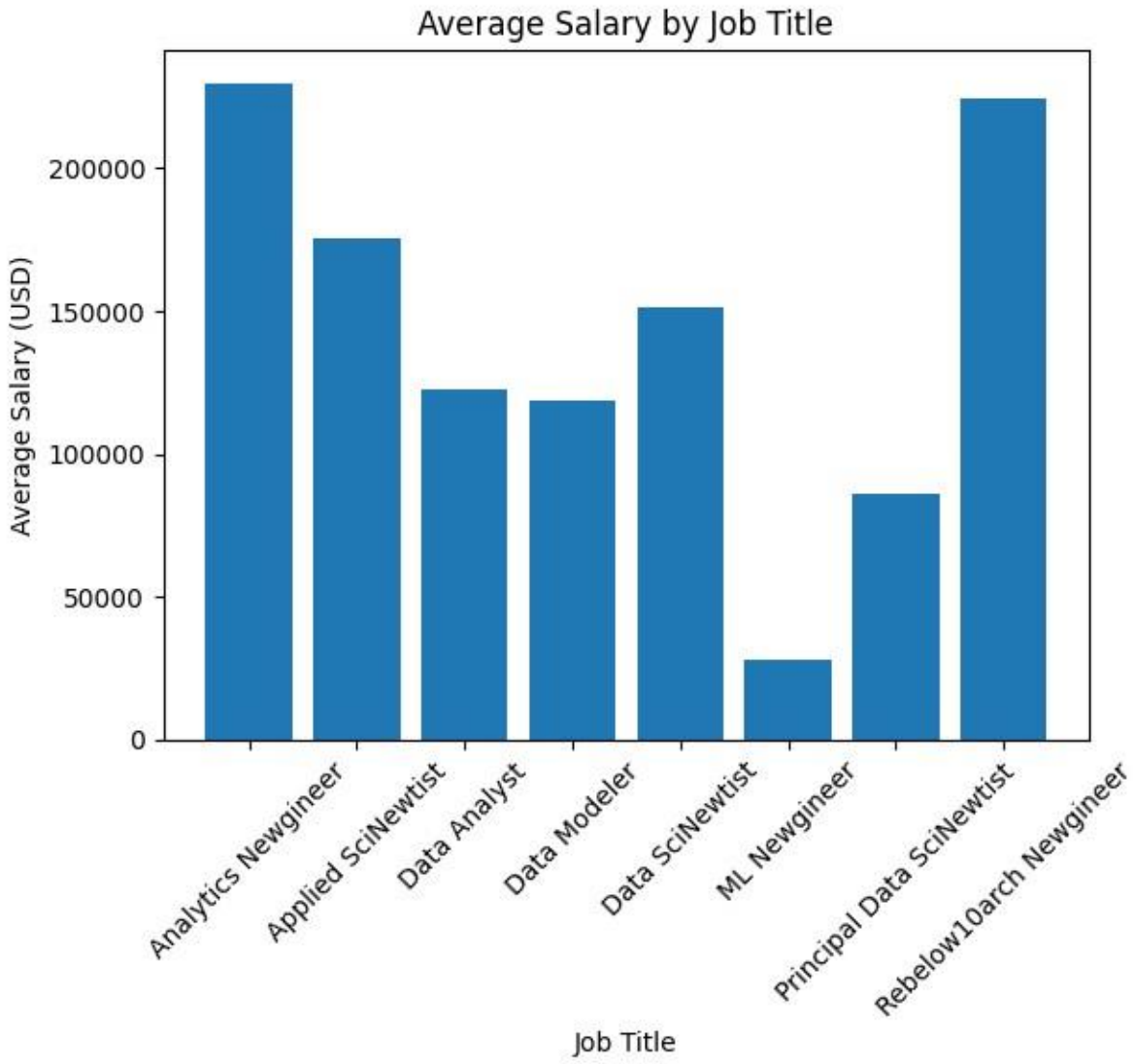
Problem 1: Distribution of Salary

```
import matplotlib.pyplot as plt

plt.hist(data['salary_in_usd'])
plt.title('Distribution of Salary')
plt.xlabel('Salary (USD)')
plt.ylabel('Frequency')
plt.show()
```



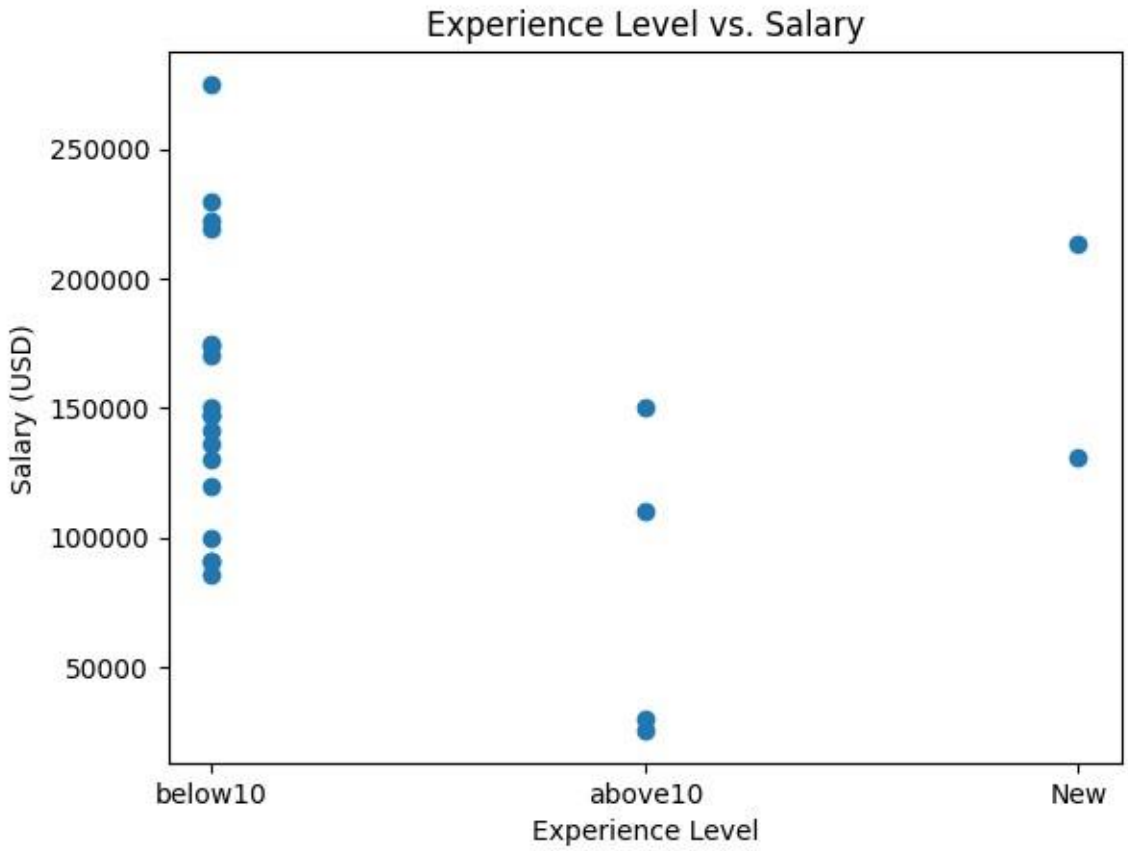
Problem 2: Comparison of Salary by Job Title

```
import matplotlib.pyplot as plt

avg_salary_by_job = data.groupby('job_title')['salary_in_usd'].mean()
plt.bar(avg_salary_by_job.index, avg_salary_by_job.values)
plt.title('Average Salary by Job Title')
plt.xlabel('Job Title')
plt.ylabel('Average Salary (USD)')
plt.xticks(rotation=45)
plt.show()
```



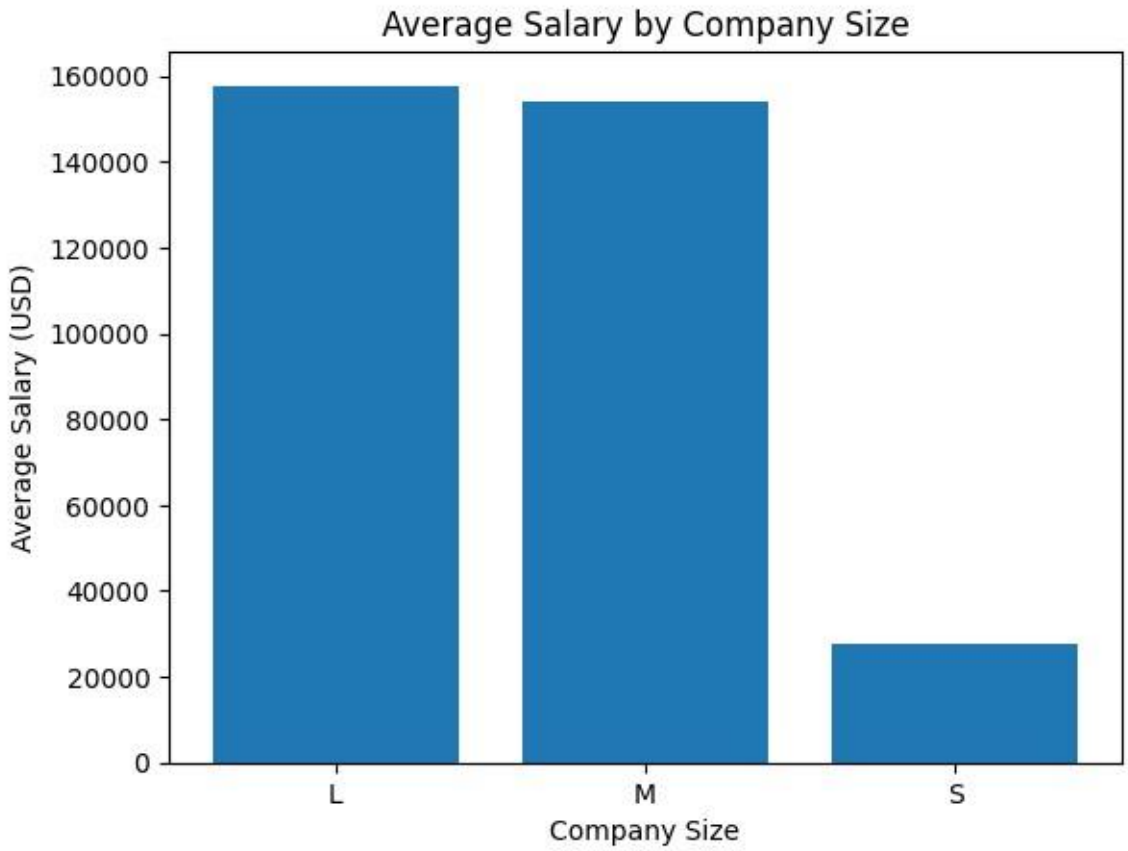
Problem 3: Experience Level vs. Salary

```
import matplotlib.pyplot as plt

plt.scatter(data['experiNewce_level'], data['salary_in_usd'])
plt.title('Experience Level vs. Salary')
plt.xlabel('Experience Level')
plt.ylabel('Salary (USD)')
plt.show()
```



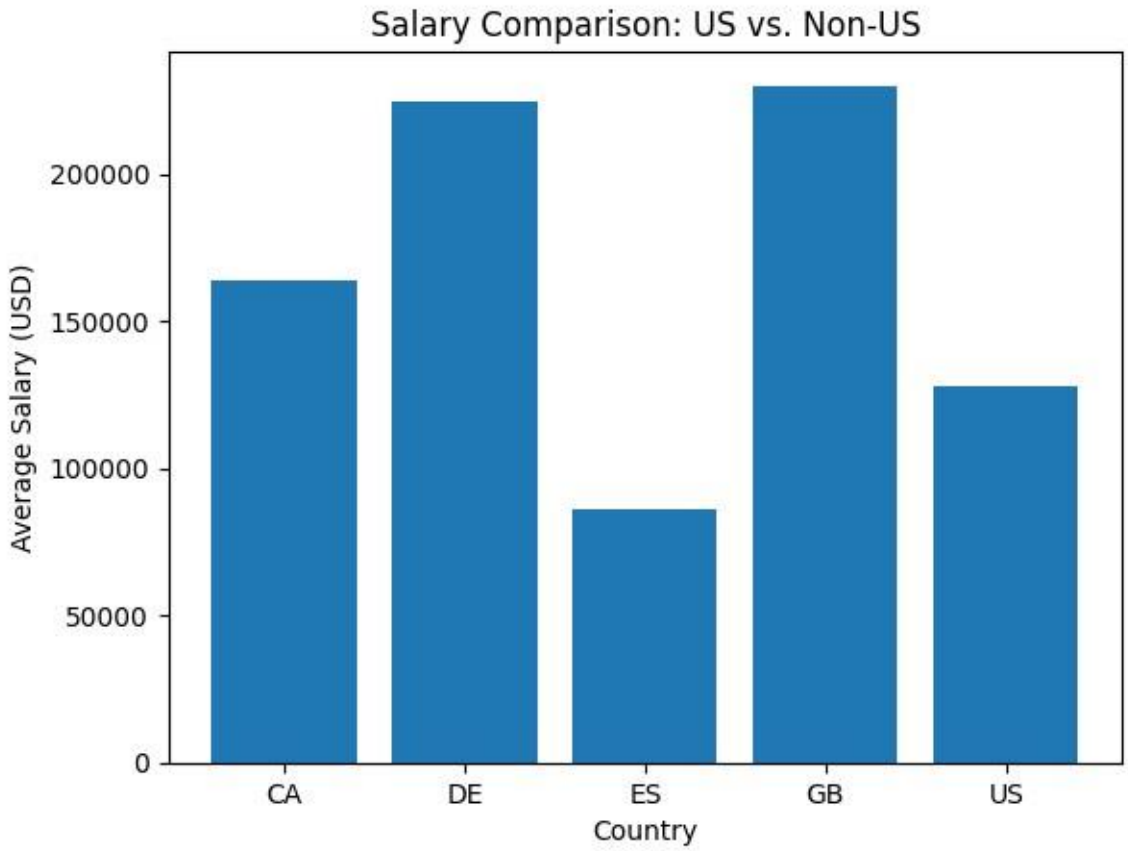
Problem 4: Salary by Company size

```
import matplotlib.pyplot as plt

avg_salary_by_size = data.groupby('company_size')['salary_in_usd'].mean()
plt.bar(avg_salary_by_size.index, avg_salary_by_size.values)
plt.title('Average Salary by Company Size')
plt.xlabel('Company Size')
plt.ylabel('Average Salary (USD)')
plt.show()
```



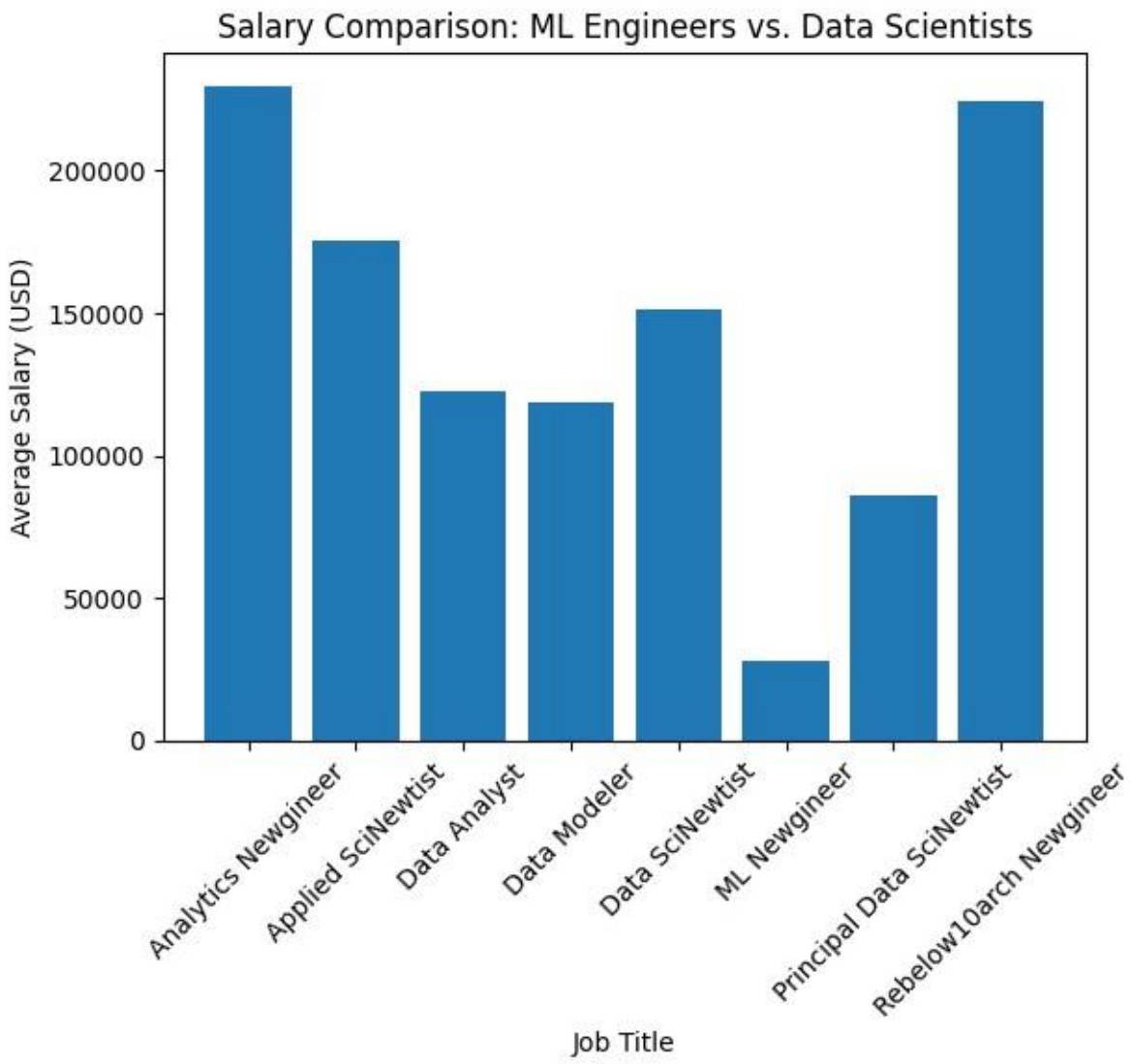
Problem 5: Salary Comparison between US and Non-US **Employees**

```
import matplotlib.pyplot as plt

avg_salary_by_country = data.groupby('company_location')['salary_in_usd'].mean()
plt.bar(avg_salary_by_country.index, avg_salary_by_country.values)
plt.title('Salary Comparison: US vs. Non-US')
plt.xlabel('Country')
plt.ylabel('Average Salary (USD)')
plt.show()
```



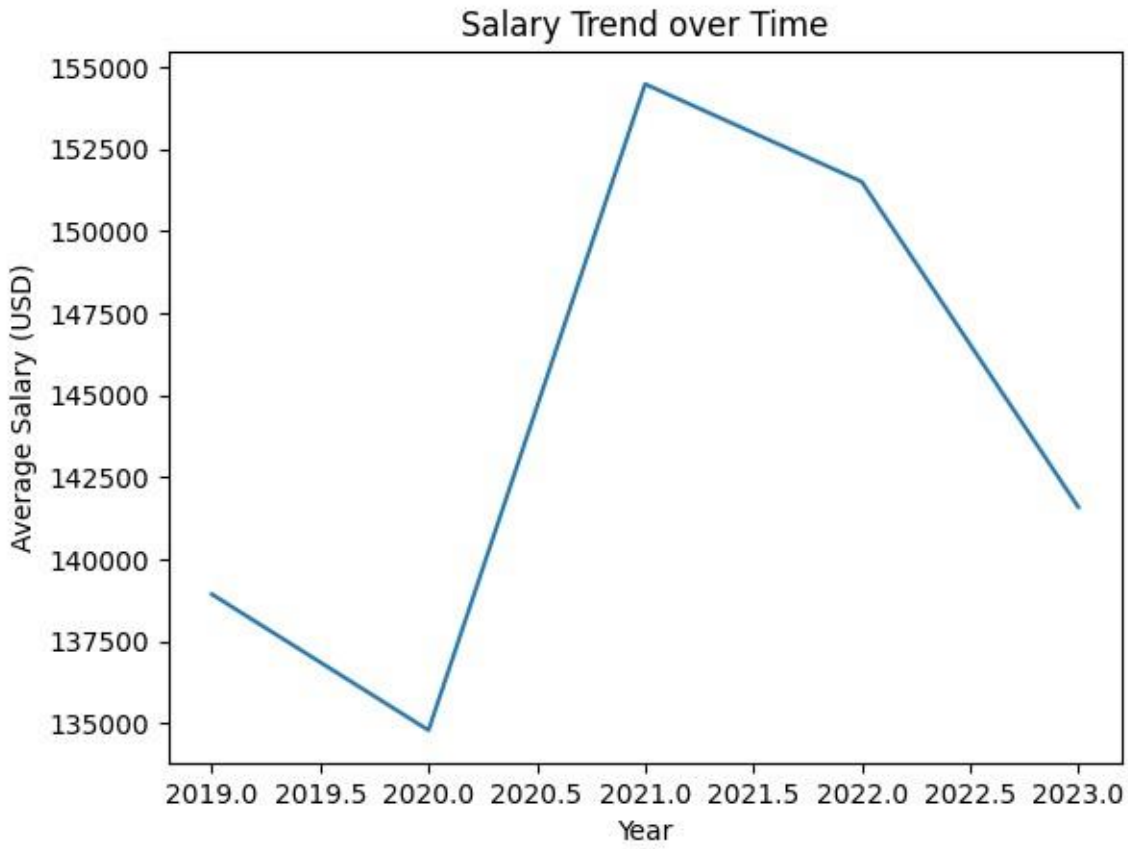
Problem 6: Salary Comparison between ML Engineers and Data Scientists

```
import matplotlib.pyplot as plt

avg_salary_by_role = data.groupby('job_title')['salary_in_usd'].mean()
plt.bar(avg_salary_by_role.index, avg_salary_by_role.values)
plt.title('Salary Comparison: ML Engineers vs. Data Scientists')
plt.xlabel('Job Title')
plt.ylabel('Average Salary (USD)')
plt.xticks(rotation=45)
plt.show()
```



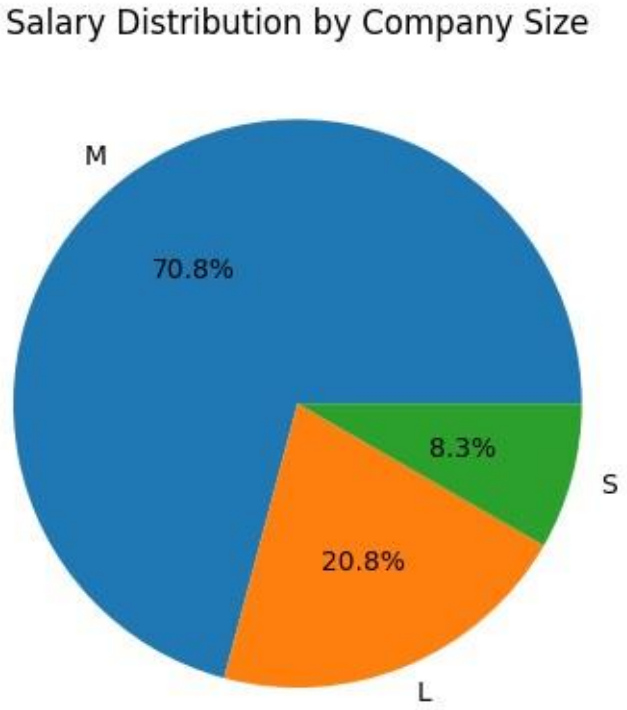
Problem 7: Salary Trend over Time

```
[26] import matplotlib.pyplot as plt

     avg_salary_over_time = data.groupby('work_year')['salary_in_usd'].mean()
     plt.plot(avg_salary_over_time.index, avg_salary_over_time.values)
     plt.title('Salary Trend over Time')
     plt.xlabel('Year')
     plt.ylabel('Average Salary (USD)')
     plt.show()
```



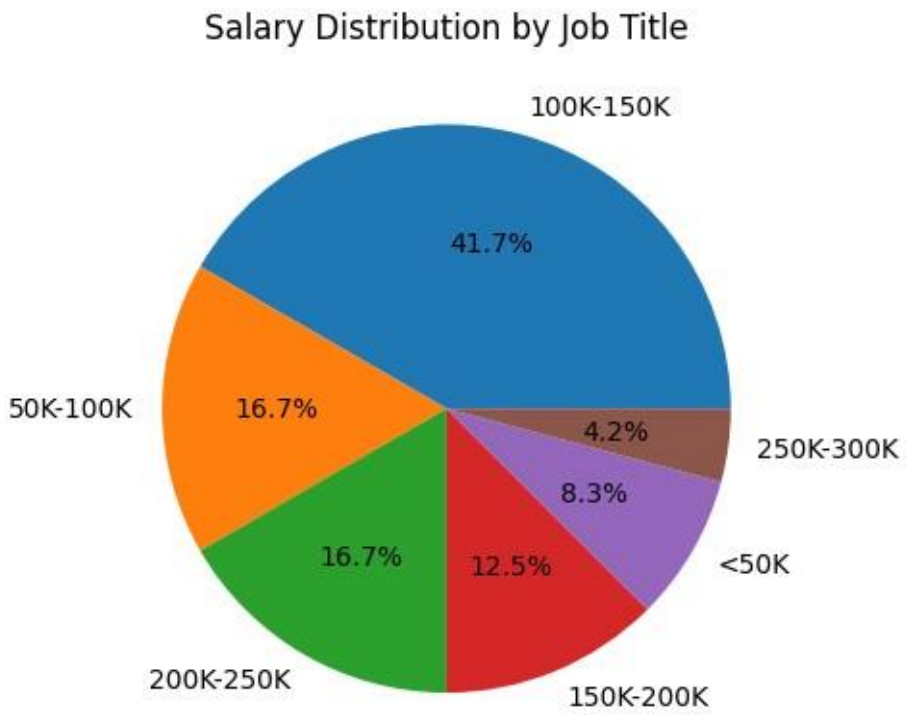
Problem 8: Salary Distribution by Company Size

```
[27] import matplotlib.pyplot as plt

     salary_counts_by_company_size = data['company_size'].value_counts()
     plt.pie(salary_counts_by_company_size, labels=salary_counts_by_company_size.index, autopct='%1.1f%%')
     plt.title('Salary Distribution by Company Size')
     plt.show()
```



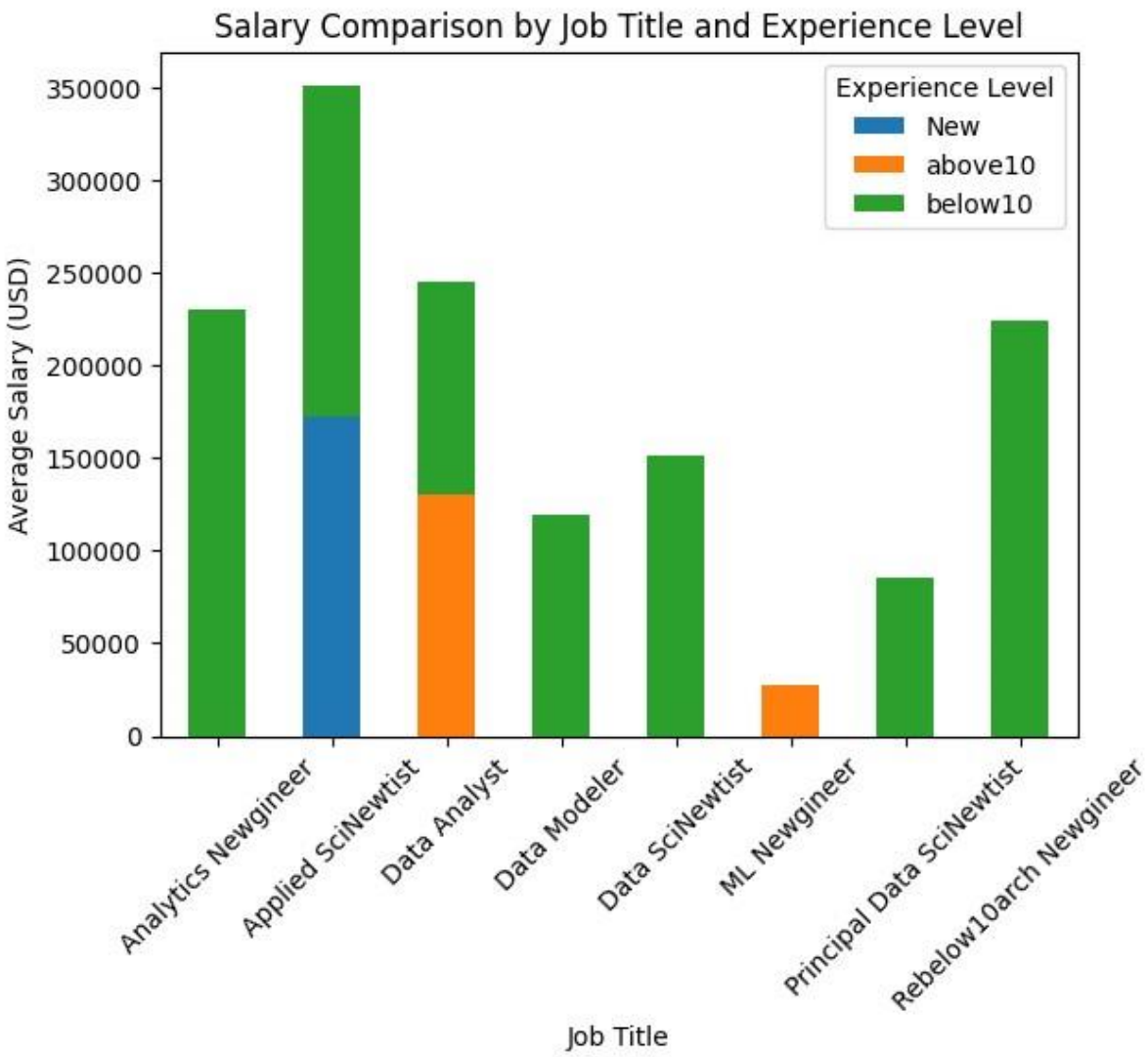
Problem 9: Salary Distribution by Job Title

```
import matplotlib.pyplot as plt

salary_bins = [0, 50000, 100000, 150000, 200000, 250000, 300000]
labels = ['<50K', '50K-100K', '100K-150K', '150K-200K', '200K-250K', '250K-300K']

data['salary_range'] = pd.cut(data['salary_in_usd'], bins=salary_bins, labels=labels)
salary_distribution = data['salary_range'].value_counts()

plt.pie(salary_distribution, labels=salary_distribution.index, autopct='%1.1f%%')
plt.title('Salary Distribution by Job Title')
plt.show()
```



Problem 10: Salary Comparison by Job Title and Experience Level

```
import matplotlib.pyplot as plt

avg_salary_by_job_exp = data.groupby(['job_title', 'experiNewce_level'])['salary_in_usd'].mean().unstack()
avg_salary_by_job_exp.plot(kind='bar', stacked=True)
plt.title('Salary Comparison by Job Title and Experience Level')
plt.xlabel('Job Title')
plt.ylabel('Average Salary (USD)')
plt.xticks(rotation=45)
plt.legend(title='Experience Level')
plt.show()
```


Salary Comparison by Job Title and Experience Level
Experience Level
New
above10
below10
Average Salary (USD)
350000
300000
250000
200000
150000
100000
50000
0
Analytics Newgineer
Applied SciNewtist
Data Analyst
Data Modeler
Data SciNewtist
ML Newgineer
Principal Data SciNewtist
Rebelow10arch Newgineer
Job Title