Mini Project

Team Members : Nikita Shelar (557) , Dnyaneshwari Modhave (561) , Ayush Tilekar (565)

```
import pandas as pd

data = pd.read_csv('C:\\Users\\admin\\Downloads\\salary_data.csv')

print(data.head())
```

```
   YearsExperience  Salary
0              1.1   39343
1              1.3   46205
2              1.5   37731
3              2.0   43525
4              2.2   39891
```

```
mean_salary = data['Salary'].mean()

print("Mean Salary:", mean_salary)
```

```
Mean Salary: 76003.0
```

```
max_experience = data['YearsExperience'].max()

print("Max Experience:", max_experience)
```

```
Max Experience: 10.5
```

```
threshold = 60000

filtered_data = data[data['Salary'] > threshold]

print(filtered_data)
```

```
    YearsExperience  Salary
6               3.0   60150
8               3.2   64445
10              3.9   63218
14              4.5   61111
15              4.9   67938
16              5.1   66029
17              5.3   83088
18              5.9   81363
```

```
19              6.0   93940
20              6.8   91738
21              7.1   98273
22              7.9  101302
23              8.2  113812
24              8.7  109431
25              9.0  105582
26              9.5  116969
27              9.6  112635
28             10.3  122391
29             10.5  121872
```

data['SalarySquared'] = data['Salary'] ** 2

print(data.head())

```
YearsExperience  Salary  SalarySquared
0              1.1   39343     1547871649
1              1.3   46205     2134902025
2              1.5   37731     1423628361
3              2.0   43525     1894425625
4              2.2   39891     1591291881
```

grouped_data = data.groupby('YearsExperience')['Salary'].mean()

print(grouped_data)

```
YearsExperience
1.1      39343.0
1.3      46205.0
1.5      37731.0
2.0      43525.0
2.2      39891.0
2.9      56642.0
3.0      60150.0
3.2      59445.0
3.7      57189.0
3.9      63218.0
4.0      56375.5
4.1      57081.0
4.5      61111.0
4.9      67938.0
5.1      66029.0
5.3      83088.0
5.9      81363.0
6.0      93940.0
6.8      91738.0
7.1      98273.0
7.9     101302.0
```

```
8.2     113812.0
8.7     109431.0
9.0     105582.0
9.5     116969.0
9.6     112635.0
10.3    122391.0
10.5    121872.0
Name: Salary, dtype: float64
```

import matplotlib.pyplot as plt

plt.scatter(data['YearsExperience'], data['Salary'])

plt.xlabel('Years of Experience')

plt.ylabel('Salary')

plt.title('Salary vs. Years of Experience')

plt.show()

```
from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error


# Split the data into training and testing sets

X = data[['YearsExperience']]

y = data['Salary']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


# Train the linear regression model

model = LinearRegression()

model.fit(X_train, y_train)


# Make predictions on the test set

y_pred = model.predict(X_test)


# Evaluate the model using mean squared error

mse = mean_squared_error(y_test, y_pred)

print("Mean Squared Error:", mse)
```

```
Mean Squared Error: 49830096.85590839
```