**Assignment-4**

**Name:** Dheeraj Reddy Pandham **UCM-ID:**700747151

1. Data Manipulation:

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv('data.csv') # reading CSV file

df.head()

df.describe() # describing the Statistical data

df.fillna(df.mean(), inplace=True) # replacing the NULL values with mean values

# selecting two columns and aggregating the data

df[['Calories', 'Maxpulse']].agg(['min', 'max', 'count', 'mean'])

# selecting rows between 500 and 1000

df[(df['Calories'] >= 500) & (df['Calories'] <= 1000)]

# selecting rows greater than 500 and less than 1000

df[(df['Calories'] > 500) & (df['Maxpulse'] < 100)]

#contains all columns except Maxpulse

df\_modified = df.drop('Maxpulse', axis=1)

#deleting 'Maxpulse Column from main dataframe

df = df.drop('Maxpulse', axis=1)

print(df)

# converting calories to int datatype

df['Calories'] = df['Calories'].astype(int)

#plotting the data using the matplotlib

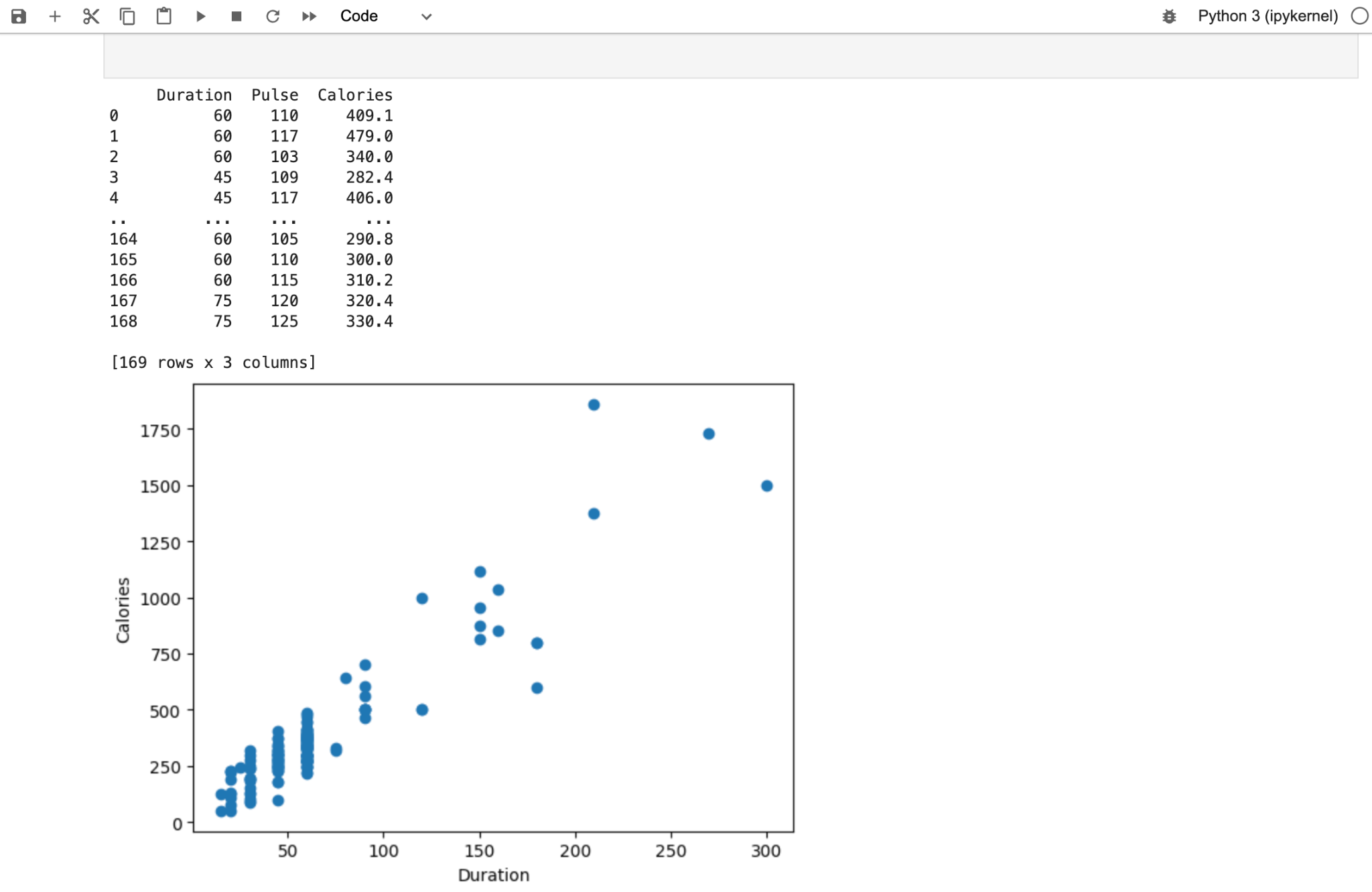
plt.scatter(df['Duration'], df['Calories'])

plt.xlabel("Duration") #xaxis

plt.ylabel("Calories") #yaxis

plt.show() # displaying the Scatter Plot

Output:



2. Linear Regression:

import pandas as pd

from sklearn.metrics import mean\_squared\_error

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

salariesData = pd.read\_csv('Salary\_Data.csv') #importing data from the CSV file

#splitting the data in to training and testing

X = salariesData.iloc[:, :-1].values

Y= salariesData.iloc[:, 1].values

#splitting 1/3 of the data

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size = 1/3, random\_state = 0)

# Fitting Simple Linear Regression to the training set

reg = LinearRegression()

reg.fit(X\_train, Y\_train)

# Predicting the Test set result

pred = reg.predict(X\_test)

# Calculating the Mean\_squared\_error

mse = mean\_squared\_error(Y\_test, pred)

#Visualising the Training set results and Test set results

plt.scatter(X\_train, Y\_train, color = 'blue')

plt.scatter(X\_test, Y\_test, color = 'red')

plt.title('Salary Data')

plt.xlabel('Experience (Years)')

plt.ylabel('Salary')

plt.show()

Output:

