ICP4

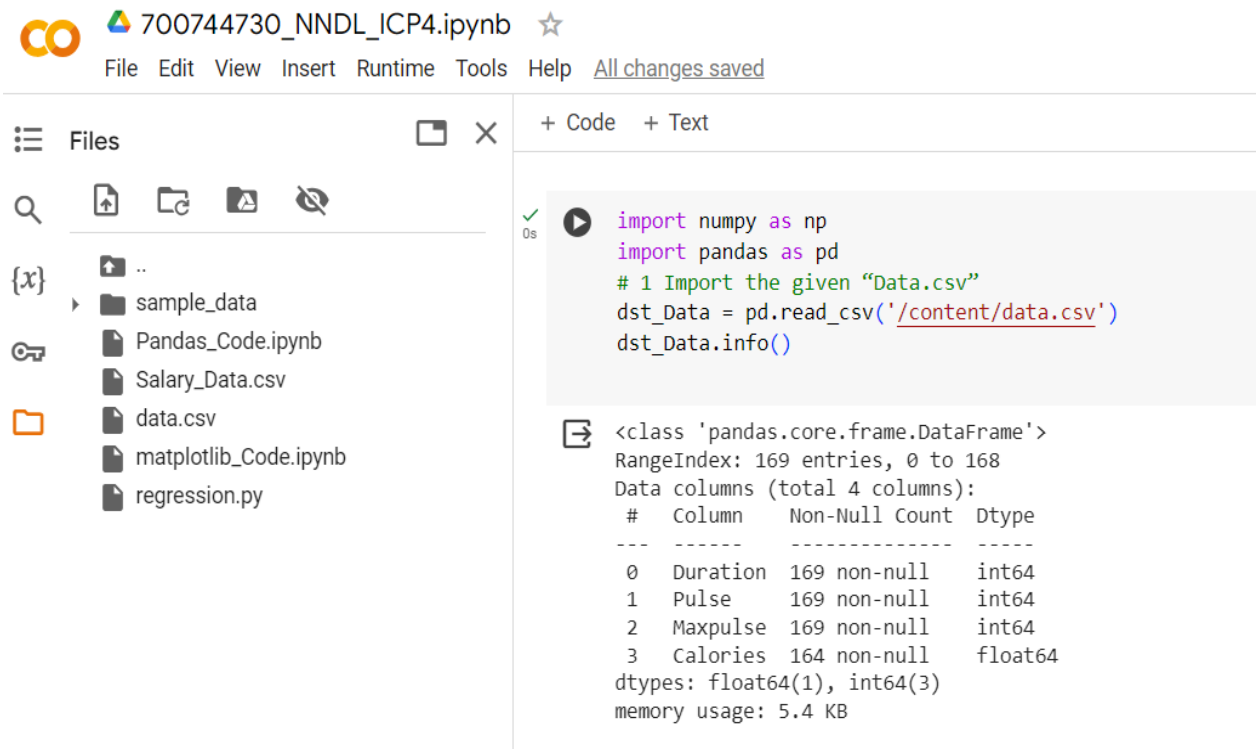
Vamsi Krishna Remala
Student id: 700744730

GitHub Link: https://github.com/vamsikrishnaremala/700744730_NNDL_ICP4

Video Link:
https://drive.google.com/file/d/1OHXr3Rl8r2THtC6UIHvxU7ujB-0RRF2P/view?usp=sharing

1. Data Manipulation

a. Read the provided CSV file 'data.csv'.



b. Show the basic statistical description about the data.

c. Check if the data has null values.

i. Replace the null values with the mean.

700744730_NNDL_ICP4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

..
sample_data
Pandas_Code.ipynb
Salary_Data.csv
data.csv
matplotlib_Code.ipynb
regression.py

Disk 81.46 GB available

+ Code + Text

[2] # Show the basic statistical description about the data.
dst_Data.head()

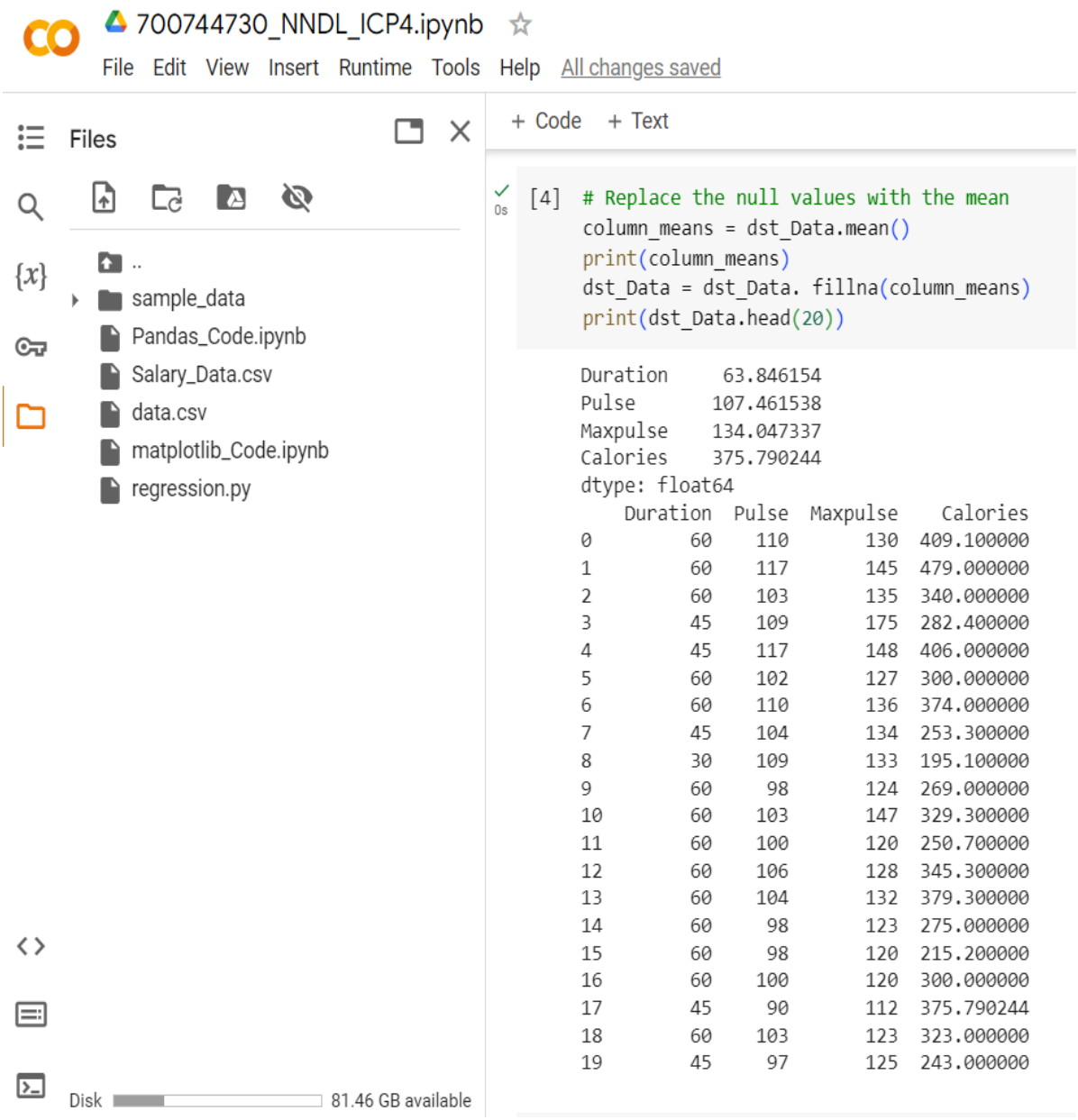
	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0

[ ] # Check if the data has null values.
dst_Data.isnull().any()

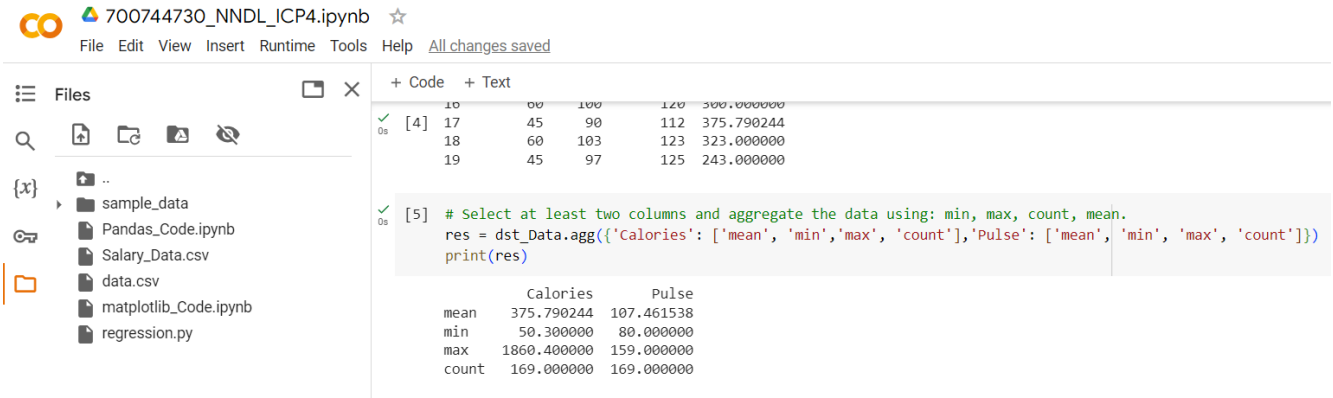
```
Duration    False
Pulse       False
Maxpulse    False
Calories     True
dtype: bool
```

[3] dst_Data.fillna(dst_Data.mean(), inplace=True)
dst_Data.isnull().any()

```
Duration    False
Pulse       False
Maxpulse    False
Calories    False
dtype: bool
```

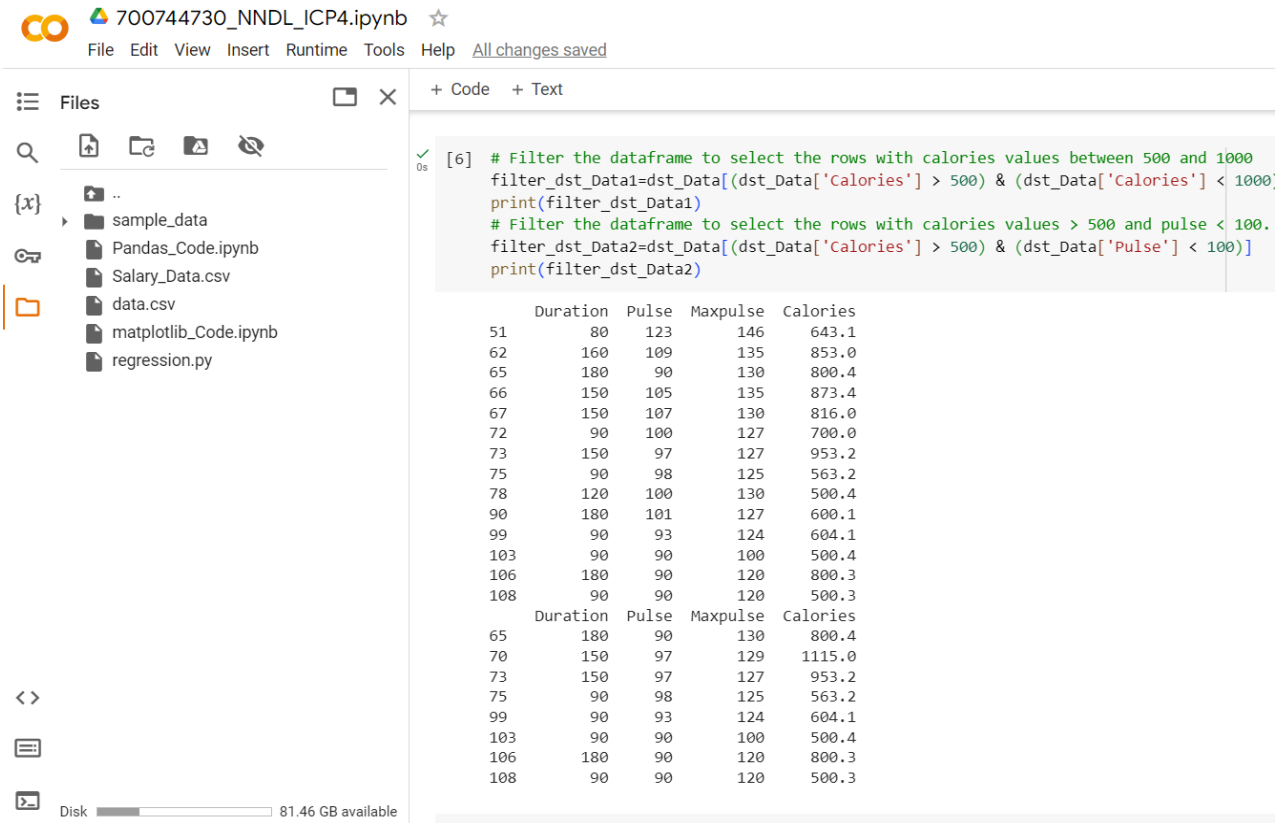


d. Select at least two columns and aggregate the data using: min, max, count, mean.



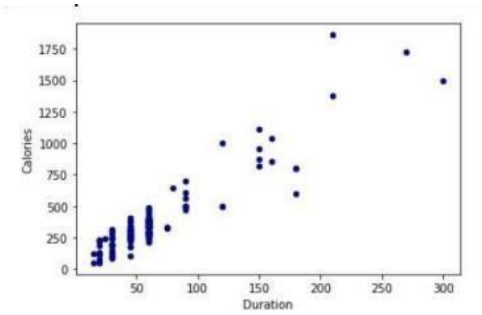
e. Filter the dataframe to select the rows with calories values between 500 and1000.

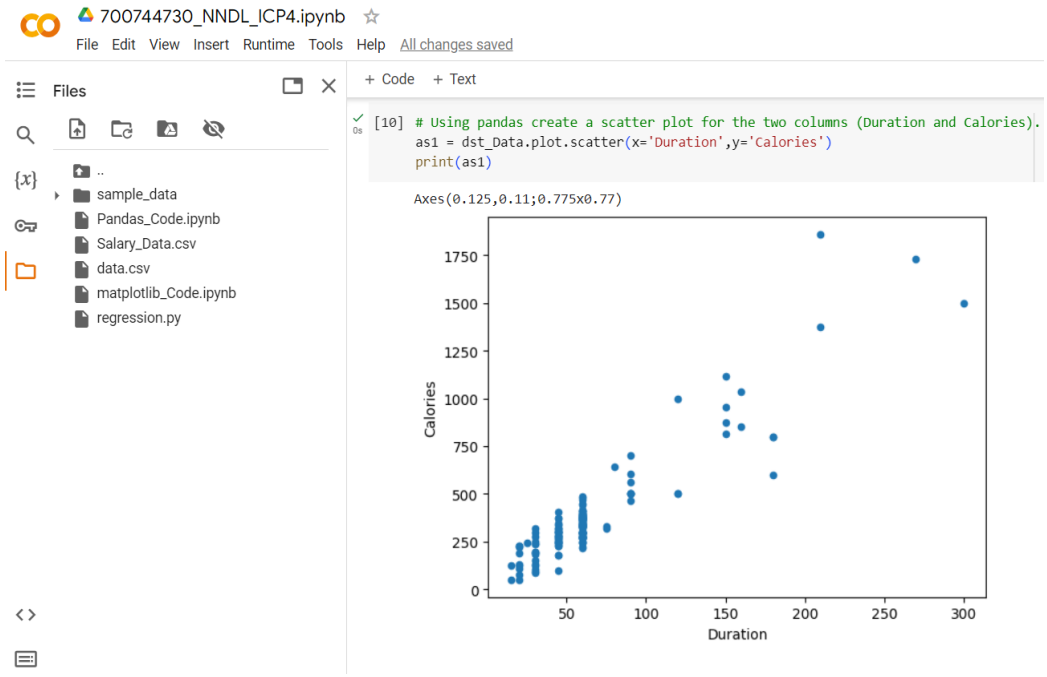
f. Filter the dataframe to select the rows with calories values > 500 and pulse.



k. Using pandas create a scatter plot for the two columns (Duration and Calories).

Example





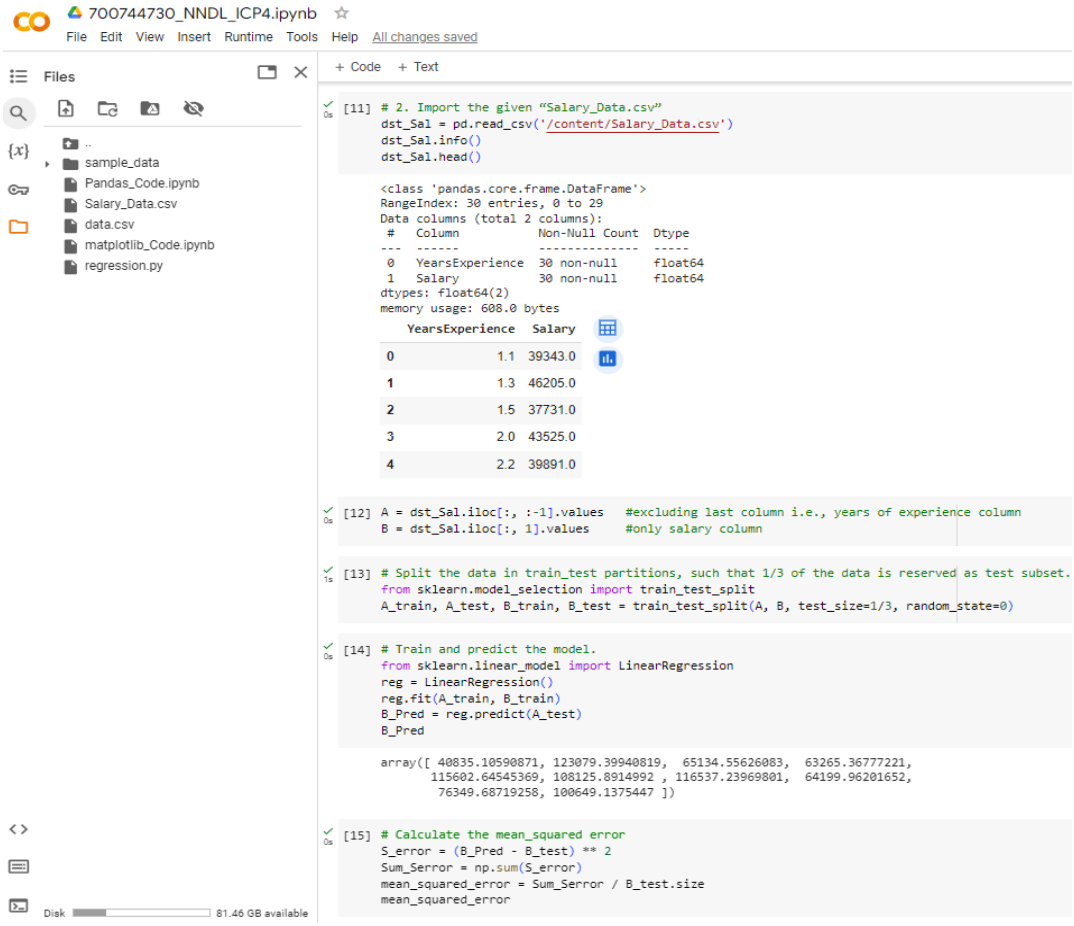
2. Linear Regression

a) Import the given “Salary_Data.csv”

b) Split the data in train_test partitions, such that 1/3 of the data is reserved as test subset.

c) Train and predict the model.

d) Calculate the mean_squared error



e) Visualize both train and test data using scatter plot.

700744730_NNDL_ICP4.ipynb

File Edit View Insert Runtime Tools Help

Files

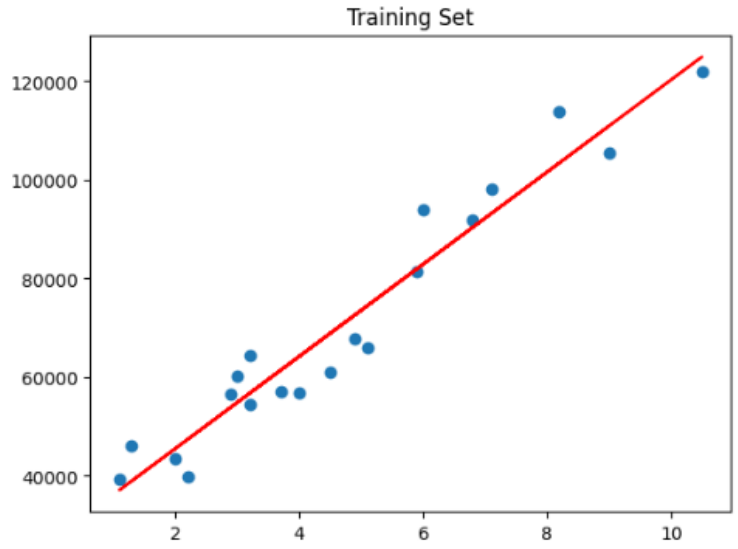
..
sample_data
Pandas_Code.ipynb
Salary_Data.csv
data.csv
matplotlib_Code.ipynb
regression.py

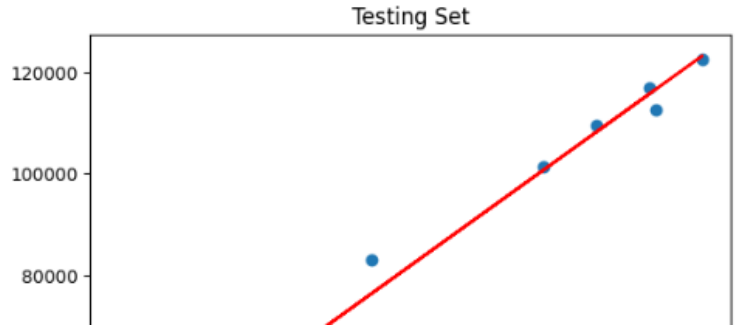
Disk 81.46 GB available

+ Code + Text

```
# Visualize both train and test data using scatter plot.
import matplotlib.pyplot as plt
# Training Data set
plt.scatter(A_train, B_train)
plt.plot(A_train, reg.predict(A_train), color='red')
plt.title('Training Set')
plt.show()

# Testing Data set
plt.scatter(A_test, B_test)
plt.plot(A_test, reg.predict(A_test), color='red')
plt.title('Testing Set')
plt.show()
```





700744730_NNDL_ICP4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

..
sample_data
Pandas_Code.ipynb
Salary_Data.csv
data.csv
matplotlib_Code.ipynb
regression.py

Disk 81.46 GB available

+ Code + Text

```
plt.title('Testing Set')
plt.show()
```