Neural Networks & Deep Learning Assignment-4

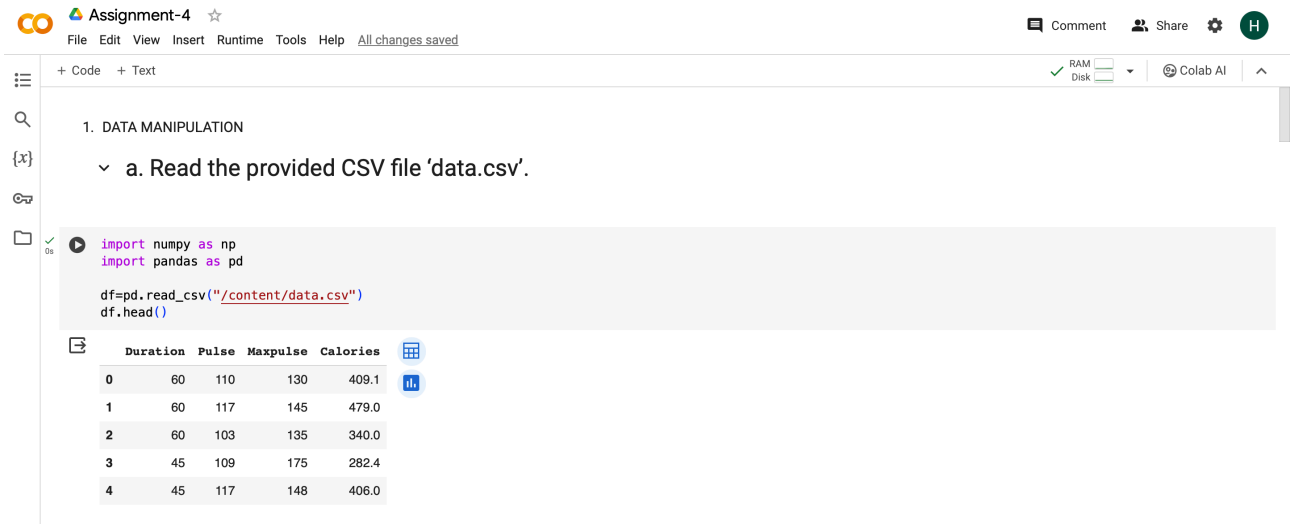
Harshavardhan Reddy Pattepur
700754833

Repository Link :
https://github.com/harshavardhanreddy27/Neural-Networks-Deep-Learning-Assignment--4-

Video Link:
https://drive.google.com/file/d/1DAUUHKpuywp42joMfdc50JthahtWqA8_/view?usp=share_link

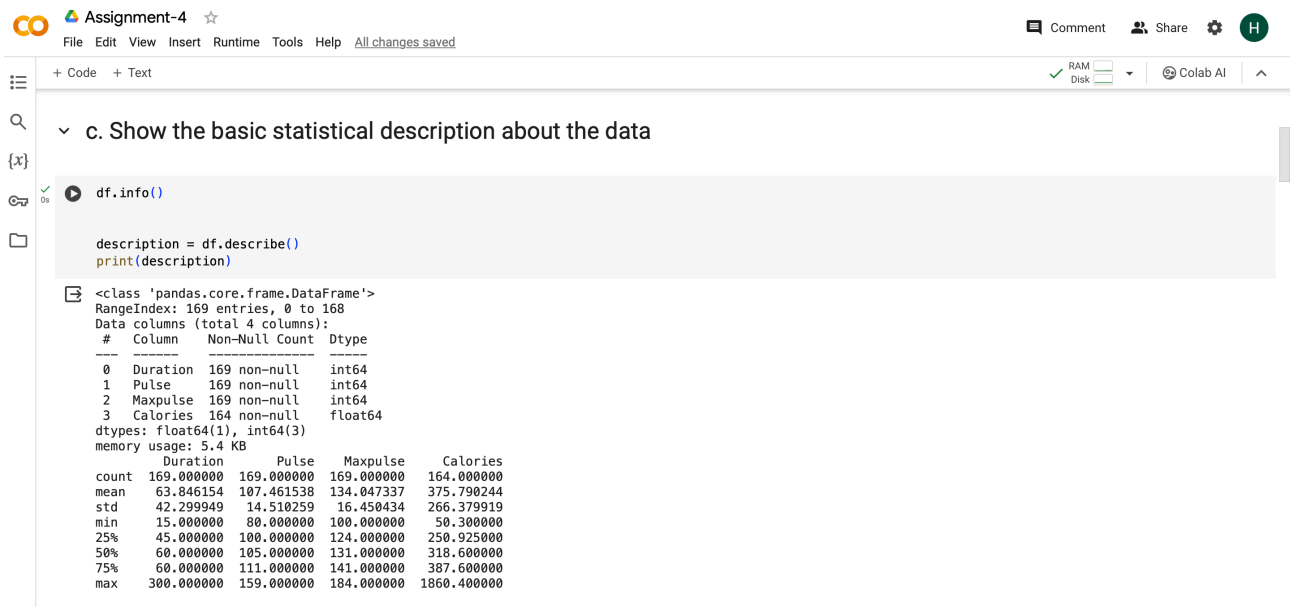
Code Screenshots:

Question:1A)



	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

C)



D)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

max 300.000000 159.000000 184.000000 1860.400000

d. Check if the data has null values.

```
null_values = df.isnull().sum()
print(null_values)
```

```
Duration    0
Pulse       0
Maxpulse    0
Calories    5
dtype: int64
```

i. Replace the null values with the mean

```
[22] df.fillna(df.mean(), inplace=True)
     null_values_after = df.isnull().sum()
     print(null_values_after)
```

```
Duration    0
Pulse       0
Maxpulse    0
Calories    0
dtype: int64
```

E)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

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

```
[23] selected_columns = ['Pulse', 'Calories']

     aggregated_data = df[selected_columns].agg(['min', 'max', 'count', 'mean'])

     print(aggregated_data)
```

```
              Pulse     Calories
min      80.000000    50.300000
max     159.000000  1860.400000
count   169.000000   169.000000
mean    107.461538   375.790244
```

F)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

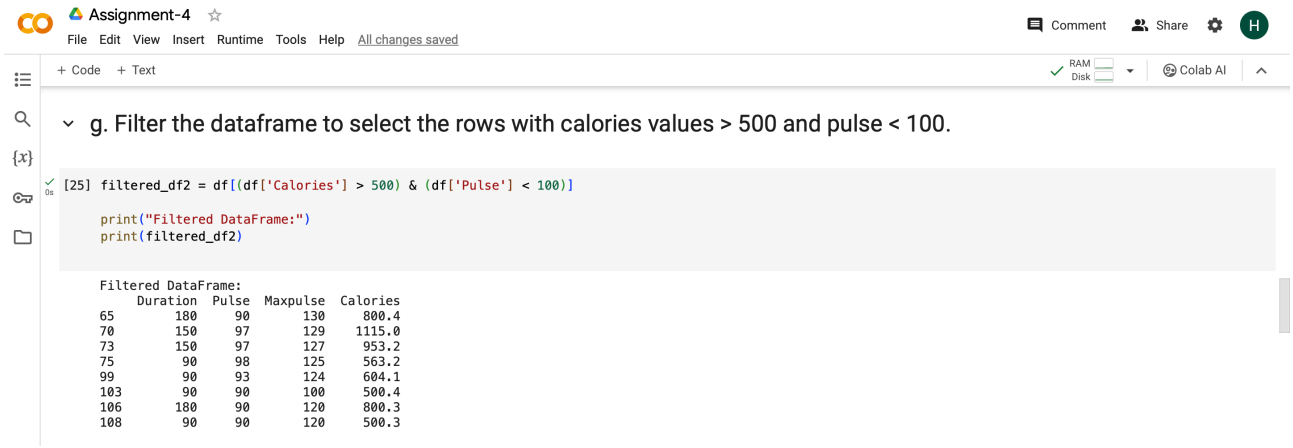
mean 107.461538 375.790244

f. Filter the dataframe to select the rows with calories values between 500 and 1000.

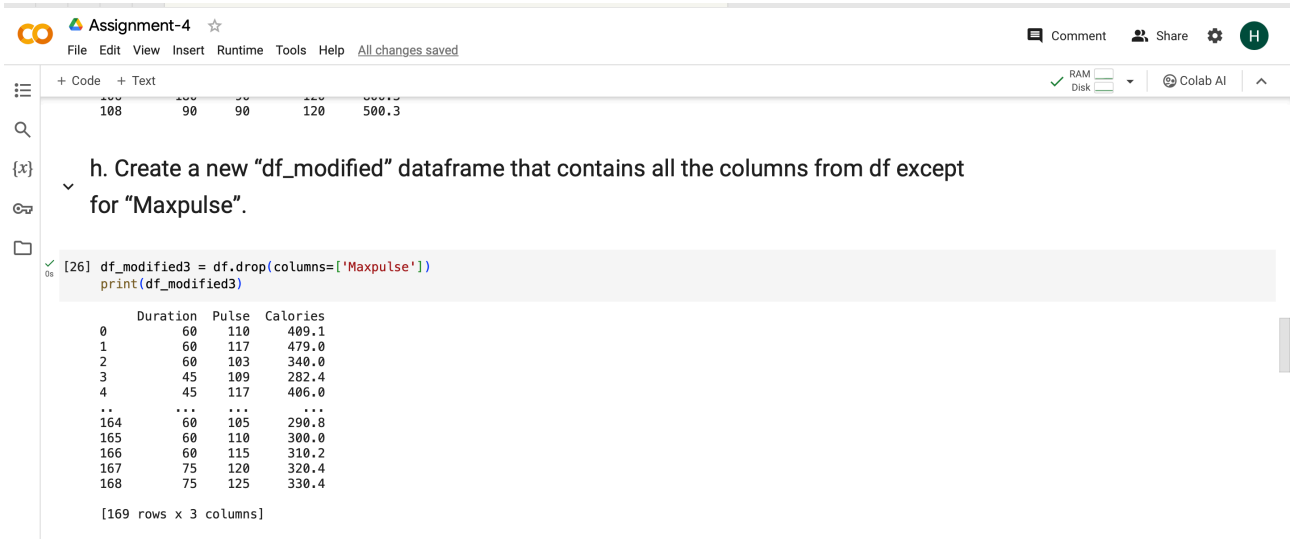
```
filtered_df = df[(df['Calories'] >= 500) & (df['Calories'] <= 1000)]

print("Filtered DataFrame:")
print(filtered_df)
```

```
Filtered DataFrame:
     Duration  Pulse  Maxpulse  Calories
51         80    123       146     643.1
62        160    109       135     853.0
65        180     90       130     800.4
66        150    105       135     873.4
67        150    107       130     816.0
72         90    100       127     700.0
73        150     97       127     953.2
75         90     98       125     563.2
78        120    100       130     500.4
83        120    100       130     500.0
90        180    101       127     600.1
99         90     93       124     604.1
101        90     90       110     500.0
102        90     90       100     500.0
103        90     90       100     500.4
106       180     90       120     800.3
108        90     90       120     500.3
```

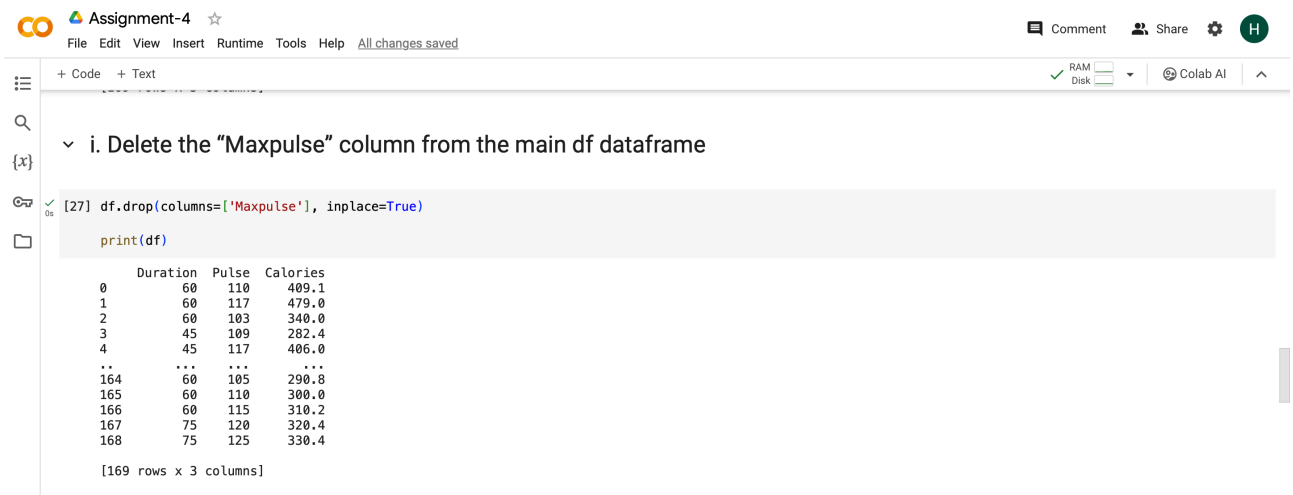
G)


Assignment-4
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
H
+ Code + Text
RAM
Disk
Colab AI
g. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.
[25] filtered_df2 = df[(df['Calories'] > 500) & (df['Pulse'] < 100)]
print("Filtered DataFrame:")
print(filtered_df2)
Filtered DataFrame:
Duration Pulse Maxpulse Calories
65 180 90 130 800.4
70 150 97 129 1115.0
73 150 97 127 953.2
75 90 98 125 563.2
99 90 93 124 604.1
103 90 90 100 500.4
106 180 90 120 800.3
108 90 90 120 500.3

H)


Assignment-4
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
H
+ Code + Text
RAM
Disk
Colab AI
108 90 90 120 500.3
h. Create a new “df_modified” dataframe that contains all the columns from df except for “Maxpulse”.
[26] df_modified3 = df.drop(columns=['Maxpulse'])
print(df_modified3)
Duration Pulse Calories
0 60 110 409.1
1 60 117 479.0
2 60 103 340.0
3 45 109 282.4
4 45 117 406.0
.. ..
164 60 105 290.8
165 60 110 300.0
166 60 115 310.2
167 75 120 320.4
168 75 125 330.4
[169 rows x 3 columns]

I)


Assignment-4
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
H
+ Code + Text
RAM
Disk
Colab AI
i. Delete the “Maxpulse” column from the main df dataframe
[27] df.drop(columns=['Maxpulse'], inplace=True)
print(df)
Duration Pulse Calories
0 60 110 409.1
1 60 117 479.0
2 60 103 340.0
3 45 109 282.4
4 45 117 406.0
.. ..
164 60 105 290.8
165 60 110 300.0
166 60 115 310.2
167 75 120 320.4
168 75 125 330.4
[169 rows x 3 columns]

J)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

j. Convert the datatype of Calories column to int datatype

```
df['Calories'] = df['Calories'].round().astype(int)
print(df)
df.info()
```

```
     Duration  Pulse  Calories
0          60    110       409
1          60    117       479
2          60    103       340
3          45    109       282
4          45    117       406
..        ...    ...       ...
164        60    105       291
165        60    110       300
166        60    115       310
167        75    120       320
168        75    125       330

[169 rows x 3 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 169 entries, 0 to 168
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Duration  169 non-null    int64
 1   Pulse     169 non-null    int64
 2   Calories  169 non-null    int64
dtypes: int64(3)
memory usage: 4.1 KB
```

K)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

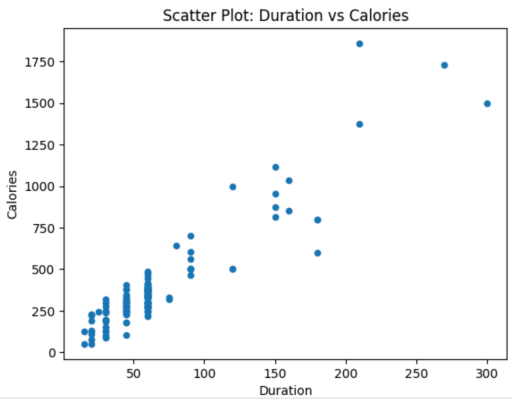
Comment Share H

+ Code + Text

RAM Disk Colab AI

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

```
import matplotlib.pyplot as plt

df = pd.DataFrame(df)

df.plot.scatter(x='Duration', y='Calories', title='Scatter Plot: Duration vs Calories')
plt.xlabel('Duration')
plt.ylabel('Calories')
plt.show()
```



0s completed at 5:47 PM

Question:2

A)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

2) Linear Regression

a) Import the given "Salary_Data.csv"

```
df2=pd.read_csv("/content/Salary_Data (2).csv")
print(df2)
```

```
    YearsExperience    Salary
0               1.1   39343.0
1               1.3   46205.0
2               1.5   37731.0
3               2.0   43525.0
4               2.2   39891.0
5               2.9   56642.0
6               3.0   60150.0
7               3.2   54445.0
8               3.2   64445.0
9               3.7   57189.0
10              3.9   63218.0
11              4.0   55794.0
12              4.0   56957.0
13              4.1   57081.0
14              4.5   61111.0
15              4.9   67938.0
16              5.1   66029.0
17              5.3   83088.0
18              5.9   81363.0
19              6.0   93940.0
20              6.8   91738.0
21              7.1   98273.0
22              7.9  101302.0
23              8.2  113812.0
24              8.7  109431.0
25              9.0  105582.0
26              9.5  116969.0
27              9.6  112635.0
```

0s completed at 5:47 PM

B ,C)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

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

```
[32] from sklearn.model_selection import train_test_split

     X = df2[['YearsExperience']]
     y = df2['Salary']
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=1/3, random_state=42)
```

c) Train and predict the model.

```
[33] from sklearn.linear_model import LinearRegression

     model = LinearRegression()
     model.fit(X_train, y_train)
     y_pred_train = model.predict(X_train)
     y_pred_test = model.predict(X_test)
```

0s completed at 5:47 PM

D)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

d)Calculate the mean_squared error

```
[36] from sklearn.metrics import mean_squared_error

     mse_train = mean_squared_error(y_train, y_pred_train)

     mse_test = mean_squared_error(y_test, y_pred_test)

     print("Mean Squared Error (Train):", mse_train)
     print("Mean Squared Error (Test):", mse_test)
```

```
Mean Squared Error (Train): 29793161.082422983
Mean Squared Error (Test): 35301898.887134895
```

E)

Assignment-4

File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

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

```
plt.figure(figsize=(10, 6))
plt.scatter(X_train, y_train, color='blue', label='Training Data')
plt.scatter(X_test, y_test, color='red', label='Test Data')
plt.plot(X_train, y_pred_train, color='green', linewidth=2, label='Regression Line')

plt.title('Linear Regression - Salary Prediction')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.legend()
plt.show()
```

Assignment-4

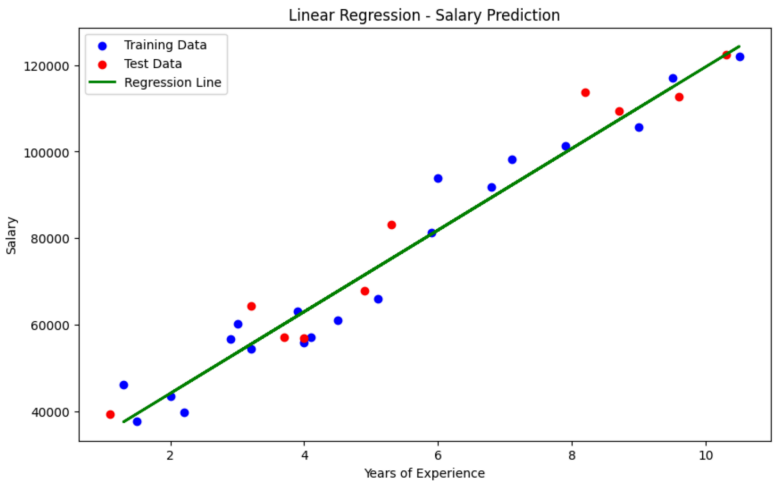
File Edit View Insert Runtime Tools Help All changes saved

Comment Share H

+ Code + Text

RAM Disk Colab AI

```
plt.legend()
plt.show()
```



0s completed at 5:47 PM