

https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb

Neural Network Deep Learning

1. Data Manipulation

- Read the provided CSV file 'data.csv'.
- <https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>
- Show the basic statistical description about the data.
- Check if the data has null values. i. Replace the null values with the mean
- Select at least two columns and aggregate the data using: min, max, count, mean.
- Filter the dataframe to select the rows with calories values between 500 and 1000.
- Filter the dataframe to select the rows with calories values > 500 and pulse < 100 .
- Create a new "df_modified" dataframe that contains all the columns from df except for "Maxpulse".
- Delete the "Maxpulse" column from the main df dataframe
- Convert the datatype of Calories column to int datatype.
- Using pandas create a scatter plot for the two columns (Duration and Calories).

The screenshot shows a Jupyter Notebook titled 'NN_Assignment4' with the following content:

```
[27]: #QUESTION 1
#read the provided CSV file
import pandas as pd
data=pd.read_csv("data.csv")
data.head()
```

The output of the first cell is a table showing the first 5 rows of the data:

	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

The second cell contains the command `data.describe()`, and its output is a summary statistics table:

	Duration	Pulse	Maxpulse	Calories
count	169.000000	169.000000	169.000000	164.000000
mean	63.846154	107.461538	134.047337	375.790244
std	42.299949	14.510259	16.450434	266.379919
min	15.000000	80.000000	100.000000	50.300000
25%	45.000000	100.000000	124.000000	250.925000
50%	60.000000	105.000000	131.000000	318.600000
75%	60.000000	111.000000	141.000000	387.600000
max	300.000000	159.000000	184.000000	1860.400000

https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb

Jupyter NN_Assignment4 Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help Trusted

JupyterLab Python 3 (ipykernel)

```
[3]: #check if the data has null values
data.isnull().any()

[3]: Duration    False
Pulse          False
Maxpulse       False
Calories       True
dtype: bool

[4]: #replace them with the mean
data.fillna(data.mean(),inplace=True)
data.isnull().any()

[4]: Duration    False
Pulse          False
Maxpulse       False
Calories       False
dtype: bool

[5]: #select at least 2 columns and aggregate the data
data.agg({'Duration':['min','max','count','mean'], 'Pulse':['min','max','count','mean']})

[5]:
```

	Duration	Pulse
min	15.000000	80.000000
max	300.000000	159.000000
count	169.000000	169.000000
mean	63.846154	107.461538

```
[6]: #filter the dataframe to select the rows btwn 500 and 1000
data.loc[(data['Calories']>500)&(data['Calories']<1000)]

[6]:
```

	Duration	Pulse	Maxpulse	Calories
51	80	123	146	643.1

Jupyter NN_Assignment4 Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help Trusted

JupyterLab Python 3 (ipykernel)

```
[6]: #filter the dataframe to select the rows btwn 500 and 1000
data.loc[(data['Calories']>500)&(data['Calories']<1000)]

[6]:
```

	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
90	180	101	127	600.1
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb

JupyterLab NN_Assignment4 Last Checkpoint: 7 minutes ago Trusted

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
[7]: #filter the dataframe to select the rows greater than 500 and less than 100
data.loc[(data['Calories']>500)&(data['Pulse']<100)]
```

```
[7]:
```

	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

```
[8]: #create a new dataframe that contains all columns except Maxpulse
df_modified=data[['Duration','Pulse','Calories']]
df_modified.head()
```

```
[8]:
```

	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

JupyterLab Trusted

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
[9]: #delete the maxpulse column from the main df
del data['Maxpulse']
```

```
[10]: data.head()
```

```
[10]:
```

	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

```
[11]: #convert the calories column to int
data.dtypes
```

```
[11]:
```

```
Duration      int64
Pulse         int64
Calories      float64
dtype: object
```

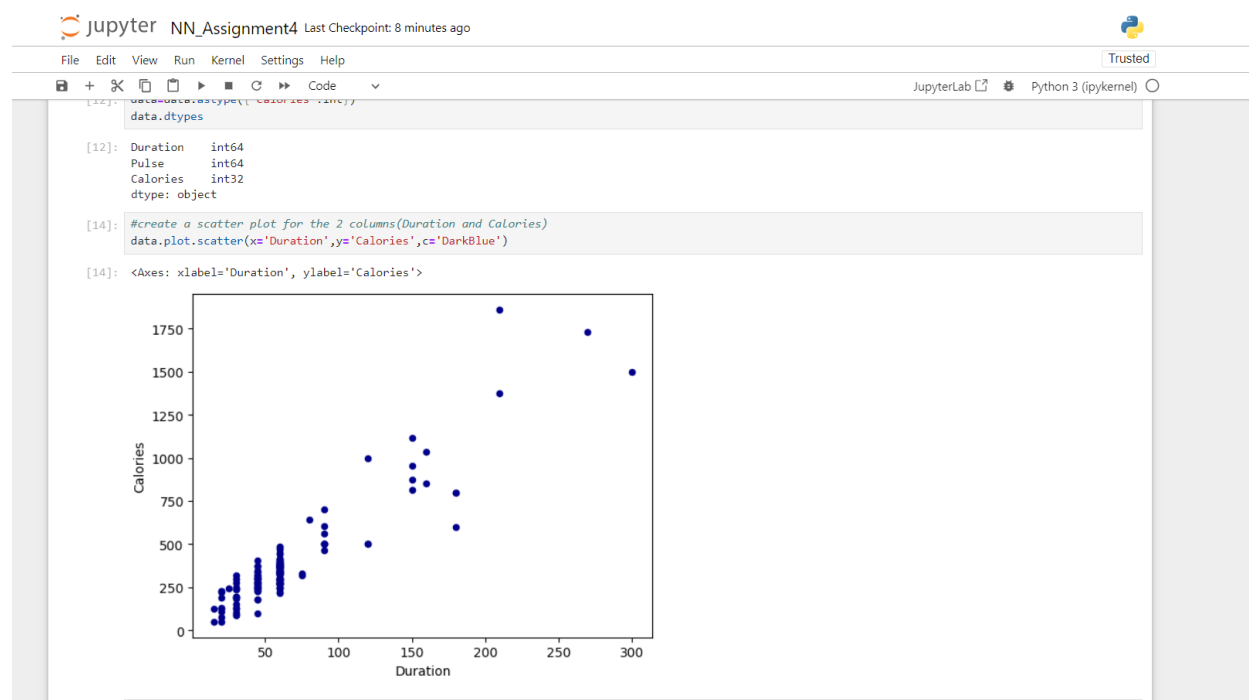
```
[12]: data=data.astype({'Calories':int})
data.dtypes
```

```
[12]:
```

```
Duration      int64
Pulse         int64
Calories      int32
dtype: object
```

```
[14]: #create a scatter plot for the 2 columns(Duration and Calories)
```

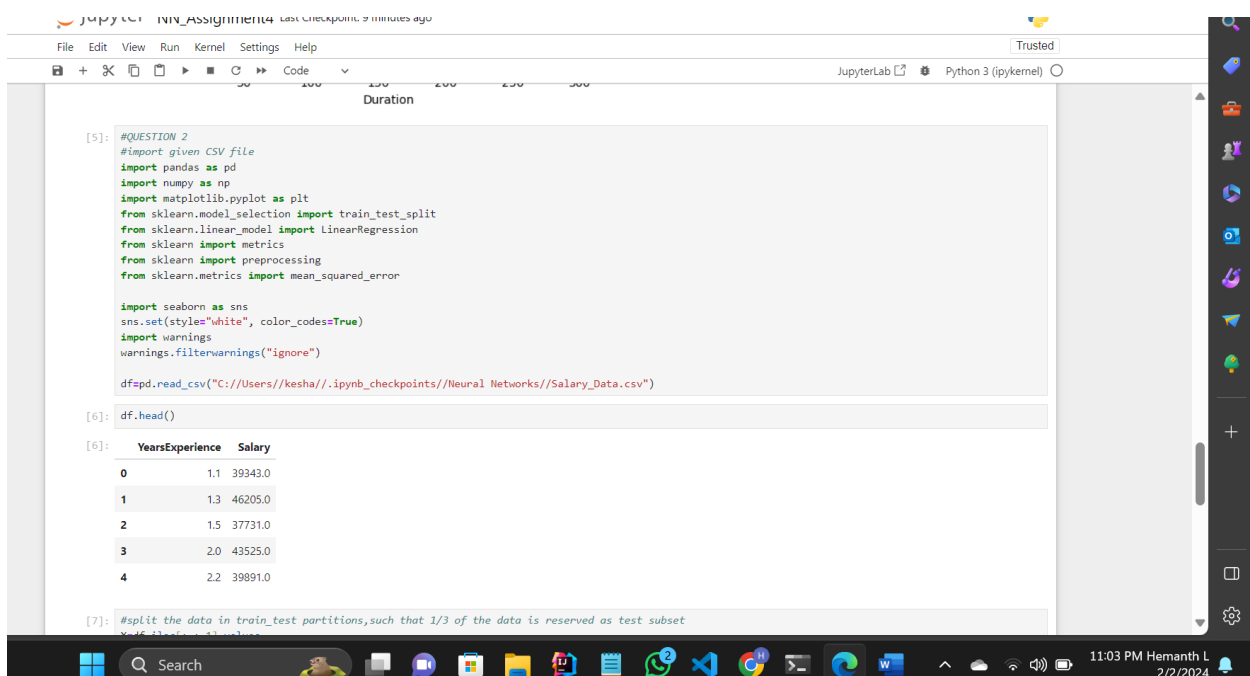
https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb



https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb

2. Linear Regression

- Import the given “Salary_Data.csv”
- Split the data in train_test partitions, such that 1/3 of the data is reserved as test subset.
- Train and predict the model.
- Calculate the mean_squared error
- Visualize both train and test data using scatter plot.



```
[5]: #QUESTION 2
#Import given CSV file
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from sklearn import preprocessing
from sklearn.metrics import mean_squared_error

import seaborn as sns
sns.set(style="white", color_codes=True)
import warnings
warnings.filterwarnings("ignore")

df=pd.read_csv("C://Users//keshal/.ipynb_checkpoints//Neural Networks//Salary_Data.csv")

[6]: df.head()
```

	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

```
[7]: #split the data in train_test partitions,such that 1/3 of the data is reserved as test subset
```



```
[7]: #split the data in train_test partitions,such that 1/3 of the data is reserved as test subset
X=df.iloc[:, :-1].values
Y=df.iloc[:, 1].values
X_Train,X_Test,Y_Train,Y_Test = train_test_split(X,Y,test_size=1/3,random_state=0)

[8]: #train and predict the model
regressor=LinearRegression()
regressor.fit(X_Train,Y_Train)

Y_Pred=regressor.predict(X_Test)

[9]: #calculate the mean squared error
mean_squared_error(Y_Test,Y_Pred)
```

[9]: 21026037.329511296

https://github.com/HemanthLakkimsetti76/NeuralNetworks_Assign4/blob/main/NN_Assignment4.ipynb

[9]: 21026037.329511296

```
[10]: #visualize both train and test data using scatter plot
plt.title("Training Data")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.scatter(X_Train,Y_Train)
plt.show
```

[10]: <function matplotlib.pyplot.show(close=None, block=None)>

