

MACHINE LEARNING –Assignment-1

Student Name: Likhitha Parvathi Tadikonda

Student Id: 700752941

Github Link: <https://github.com/LikhithaTadikonda/Machine-Learning-ICPs/tree/master/Assignment-1>

Content | 19680719 | Article_ML_Summer | Assignment-1 - Google | WhatsApp | ML Assign-1.ipynb | LikhithaTadikonda/ | +

colab.research.google.com/drive/1LfnY_OT4jwtdfdT09hKVq_bumz1khQGV

Google | YouTube | MY SLU | UCM Login | ML Article | My UCM | United Health Care | UBI | Travel Card Canara... | Canara Bank | UCM Blackboard | Transfer Out SLU

ML Assign-1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:45 AM

+ Code + Text

Connect + Gemini

Name: Likhitha Parvathi Tadikonda
700752941
Assignment-1

```
[ ] # 1. Read the provided csv file 'data.csv'.
import pandas as pd
file_path = 'data.csv'
df = pd.read_csv(file_path)

# 2. Show the basic statistical description about the data.
stat_description = df.describe()
print(stat_description)
```

	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

```
# 3. Check if the data has null values
null_values = df.isnull().sum()
print("Null values before replacement:")
print(null_values)
```

Content | 19680719 | Article_ML_Summer | Assignment-1 - Google | WhatsApp | ML Assign-1.ipynb | LikhithaTadikonda/ | +

colab.research.google.com/drive/1LfnY_OT4jwtdfdT09hKVq_bumz1khQGV

Google | YouTube | MY SLU | UCM Login | ML Article | My UCM | United Health Care | UBI | Travel Card Canara... | Canara Bank | UCM Blackboard | Transfer Out SLU

ML Assign-1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:45 AM

+ Code + Text

Connect + Gemini

```
# 3. Check if the data has null values
null_values = df.isnull().sum()
print("Null values before replacement:")
print(null_values)

# a. Replace the null values with the mean
df.fillna(df.mean(), inplace=True)
null_values_after = df.isnull().sum()
print("\nNull values after replacement:")
print(null_values_after)
```

```
Null values before replacement:
Duration    0
Pulse       0
Maxpulse    0
Calories    5
dtype: int64

Null values after replacement:
Duration    0
Pulse       0
Maxpulse    0
Calories    0
dtype: int64
```

```
[ ] # 4. Select at least two columns & aggregate the data using: min, max, count, mean.
selected_columns = ['Pulse', 'Calories']
aggregated_data = df[selected_columns].agg(['min', 'max', 'count', 'mean'])
print(aggregated_data)
```

```
# 'Pulse' and 'Calories' columns for aggregation
# Perform aggregation
```

Content | 19680719 | Article_ML_Summer | Assignment-1 - Google | (107) WhatsApp | ML Assign-1.ipynb | LikhithaTadikonda/ | +

colab.research.google.com/drive/1LfnY_OT4jwtdfdT09hKVq_bumz1khQGV

Google | YouTube | MY SLU | UCM Login | ML Article | My UCM | United Health Care | UBI | Travel Card Canara... | Canara Bank | UCM Blackboard | Transfer Out SLU

ML Assign-1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:45 AM

+ Code + Text

Connect + Gemini

```
[ ] # 4. Select at least two columns & aggregate the data using: min, max, count, mean.
selected_columns = ['Pulse', 'Calories']
aggregated_data = df[selected_columns].agg(['min', 'max', 'count', 'mean'])
print(aggregated_data)
```

#Pulse' and 'Calories' columns for aggregation
Perform aggregation

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

```
[ ] # 5. 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_df)
```

	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

Content | 19680719 | Article_ML_Summer | Assignment-1 - Google | (107) WhatsApp | ML Assign-1.ipynb | LikhithaTadikonda/ | +

colab.research.google.com/drive/1LfnY_OT4jwtdfdT09hKVq_bumz1khQGV

Google | YouTube | MY SLU | UCM Login | ML Article | My UCM | United Health Care | UBI | Travel Card Canara... | Canara Bank | UCM Blackboard | Transfer Out SLU

ML Assign-1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:45 AM

+ Code + Text

Connect + Gemini

```
[ ] # 6. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.
filtered_df = df[(df['Calories'] > 500) & (df['Pulse'] < 100)]
print(filtered_df)
```

	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

```
[ ] # 7. Create a new "df_modified" dataframe that contains all the columns from df except for "Maxpulse".
import pandas as pd
df_modified = df.drop(columns=['Maxpulse'])
print(df_modified.head())
```

	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

```
[ ] # 8. Delete the "Maxpulse" column from the main df dataframe
df.drop(columns=['Maxpulse'], inplace=True)
```

#Assume df is your main DataFrame

Content 19680719 Article_ML_Summer Assignment-1 - Google (107) WhatsApp ML Assign-1.ipynb LikhithaTadikonda/...

colab.research.google.com/drive/1LfnY_OT4jwtdfdT09hKVq_bumz1khQGV

Google YouTube MY SLU UCM Login ML Article My UCM United Health Care UBI Travel Card Canara... Canara Bank UCM Blackboard Transfer Out SLU All Bookmarks

ML Assign-1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:45 AM

Comment Share

+ Code + Text

Connect + Gemini

```
[ ] # 9. Convert the datatype of Calories column to int datatype.
df['Calories'] = df['Calories'].astype(int)

[ ] df.head()
```

	Duration	Pulse	Calories
0	60	110	409
1	60	117	479
2	60	103	340
3	45	109	282
4	45	117	406

Windows Taskbar: 10:59 11-06-2024