| Write-up | Correctness of Program | Documentation of Program | Viva | Timely Completion | Total | Dated Sign of Subject Teacher |
|---|---|---|---|---|---|---|
| 4 | 4 | 4 | 4 | 4 | 20 | |
| | | | | | | |

---

# Group A
# Assignment No: 2

---

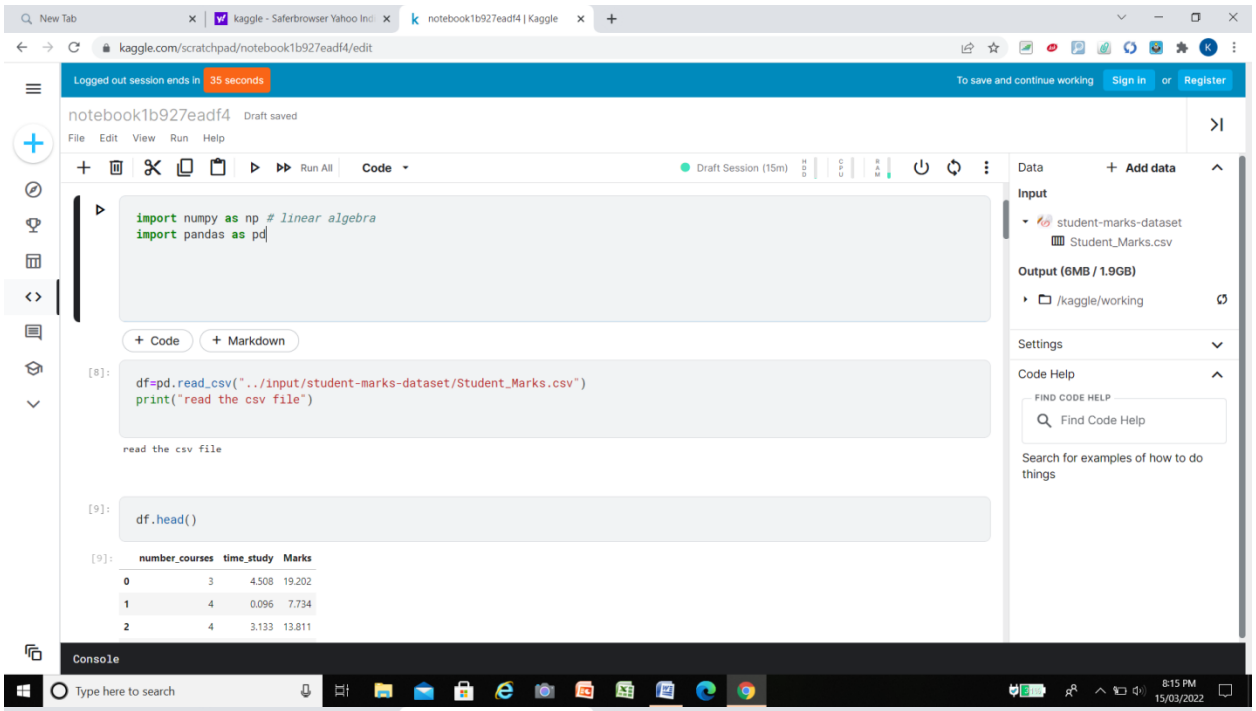
## Title of the Assignment: Data Wrangling, II

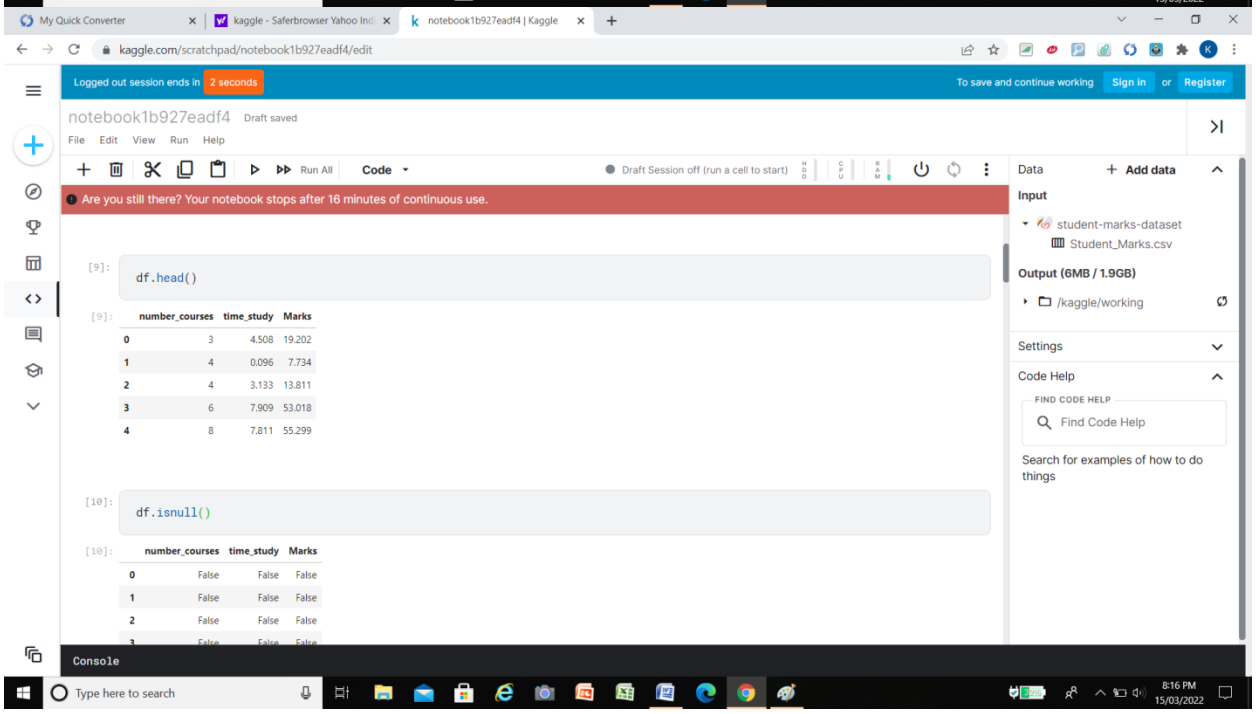
Create an "Academic performance" dataset of students and perform the following operations using Python.

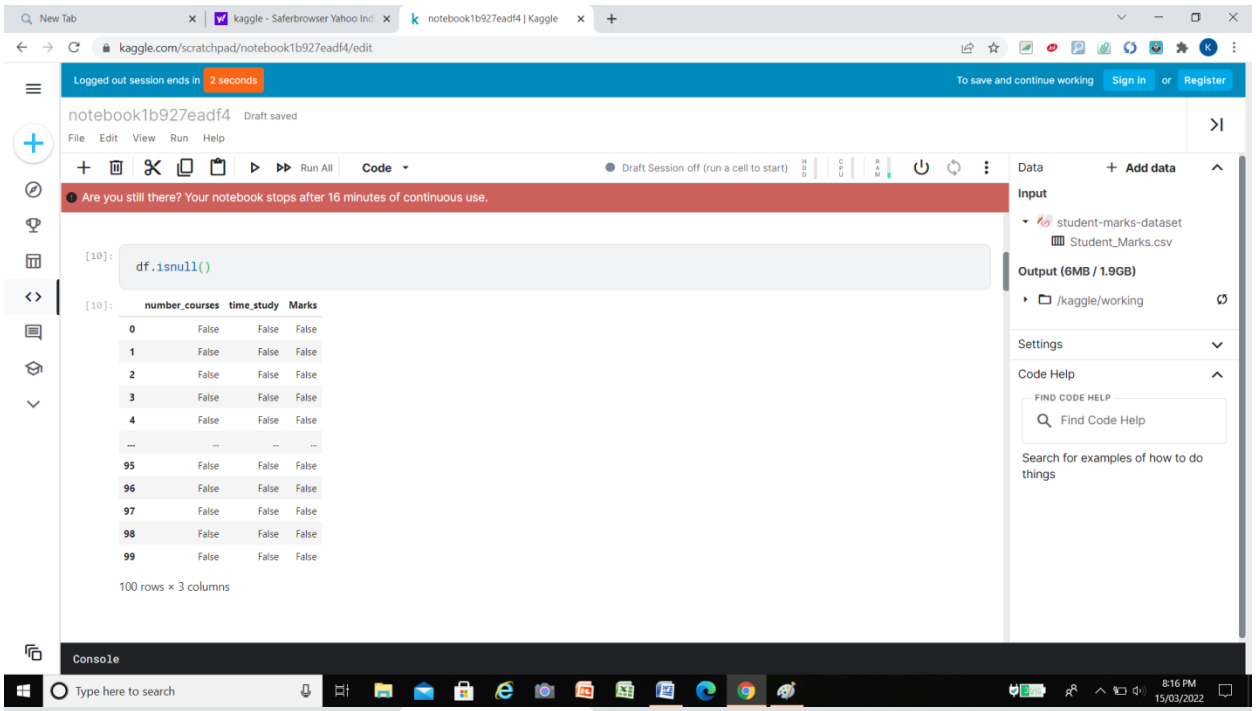
1. Scan all variables for missing values and inconsistencies. If there are missing values and/or inconsistencies, use any of the suitable techniques to deal with them.
2. Scan all numeric variables for outliers. If there are outliers, use any of the suitable techniques to deal with them.
3. Apply data transformations on at least one of the variables. The purpose of this transformation should be one of the following reasons: to change the scale for better understanding of the variable, to convert a non-linear relation into a linear one, or to decrease the skewness and convert the distribution into a normal distribution. Reason and document your approach properly.

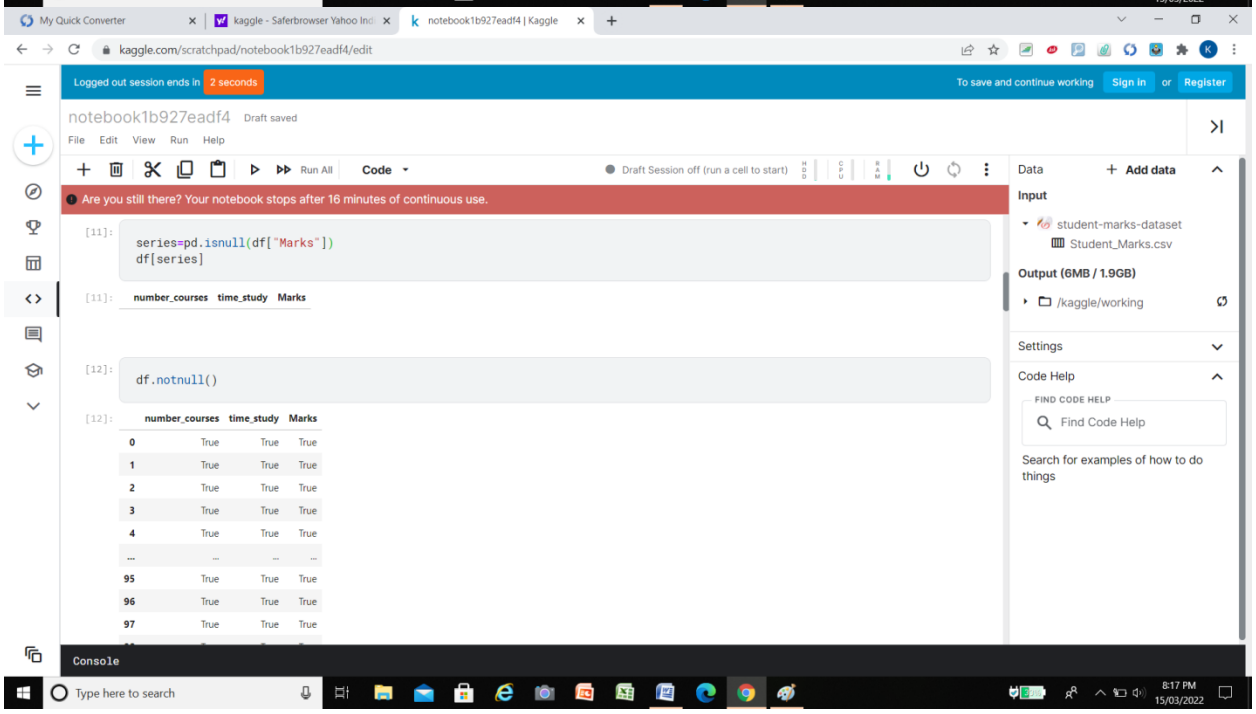
---

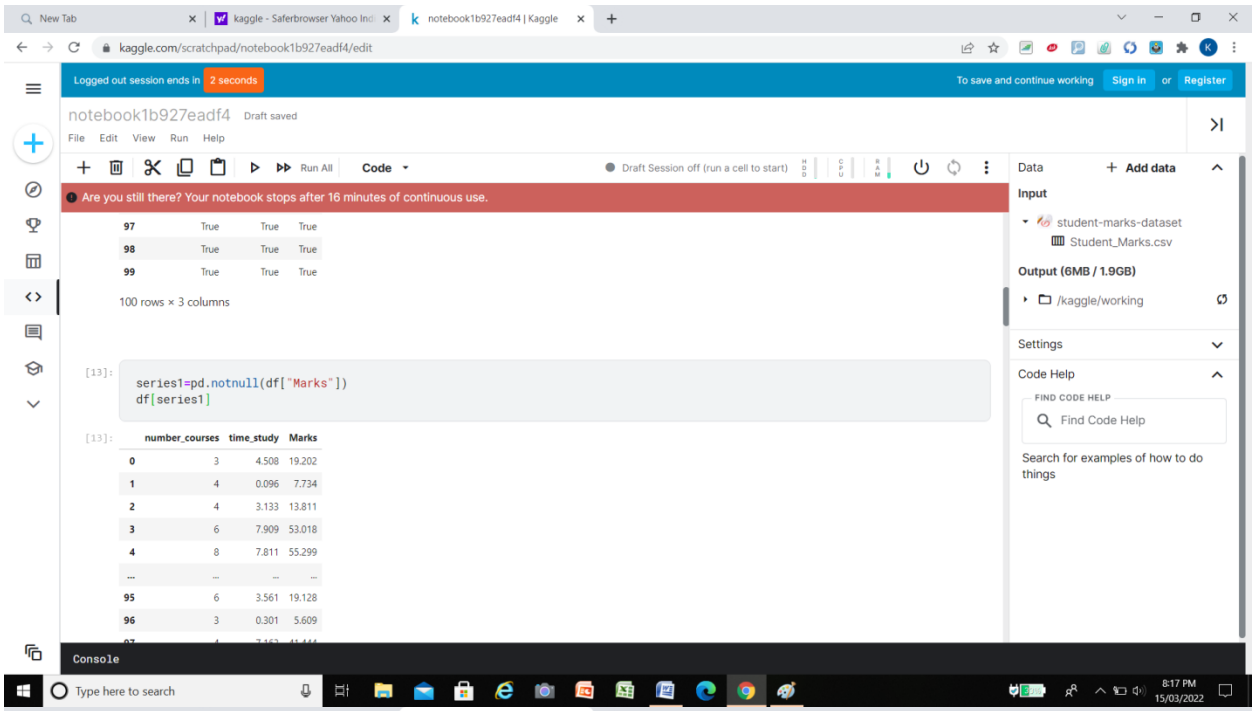
**Output:**

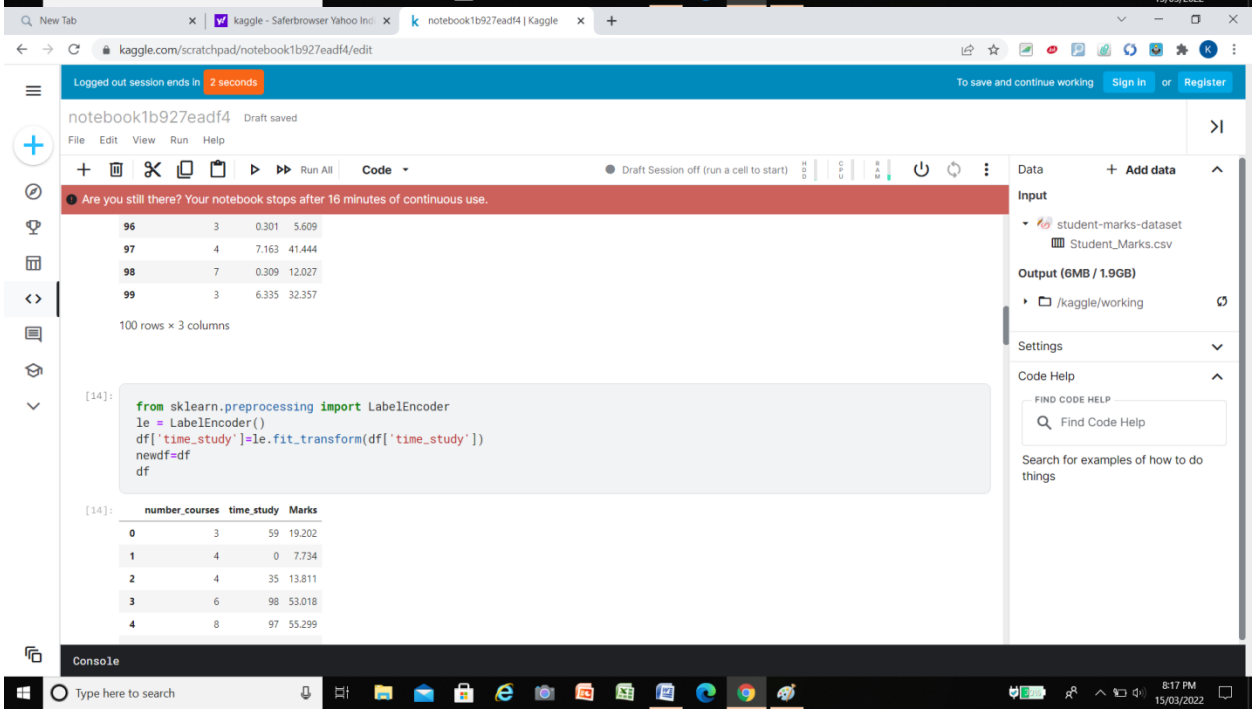

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 35 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (15m)
import numpy as np # linear algebra
import pandas as pd
+ Code + Markdown
[8]:
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
print("read the csv file")
read the csv file
[9]:
df.head()
[9]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
Data + Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:15 PM
15/03/2022

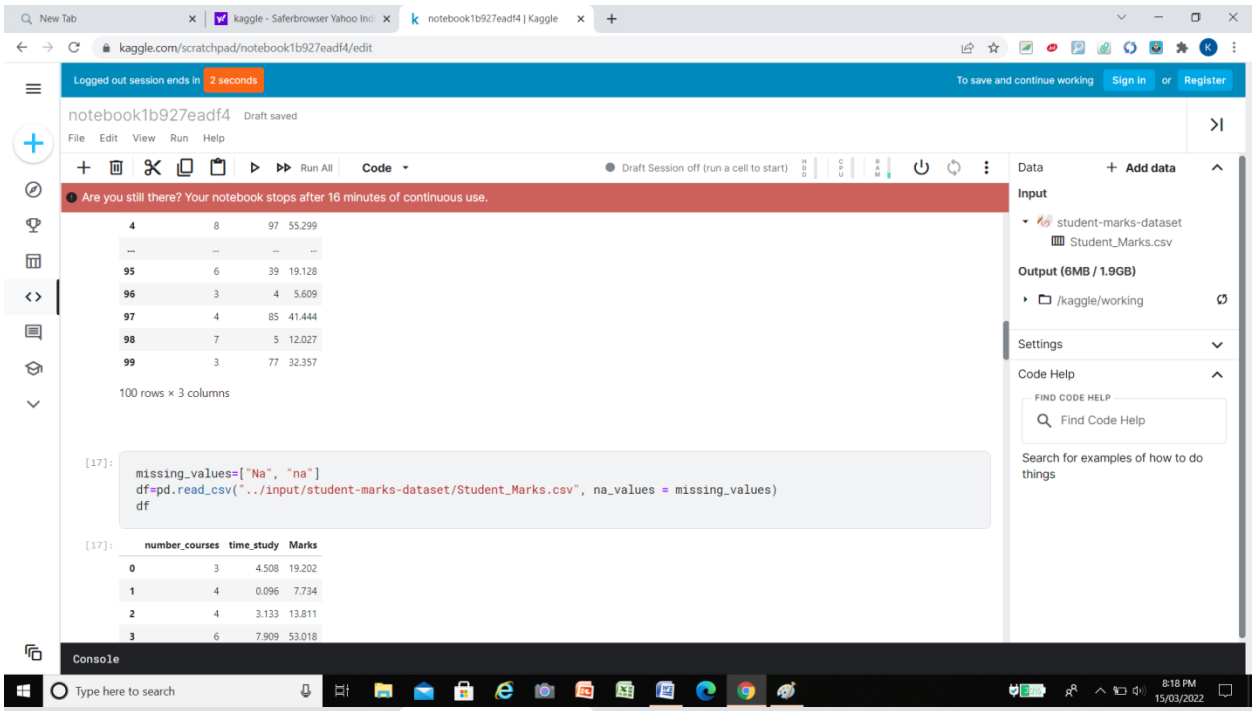

My Quick Converter
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[9]:
df.head()
[9]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
[10]:
df.isnull()
[10]:
number_courses time_study Marks
0 False False False
1 False False False
2 False False False
Data + Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:16 PM
15/03/2022

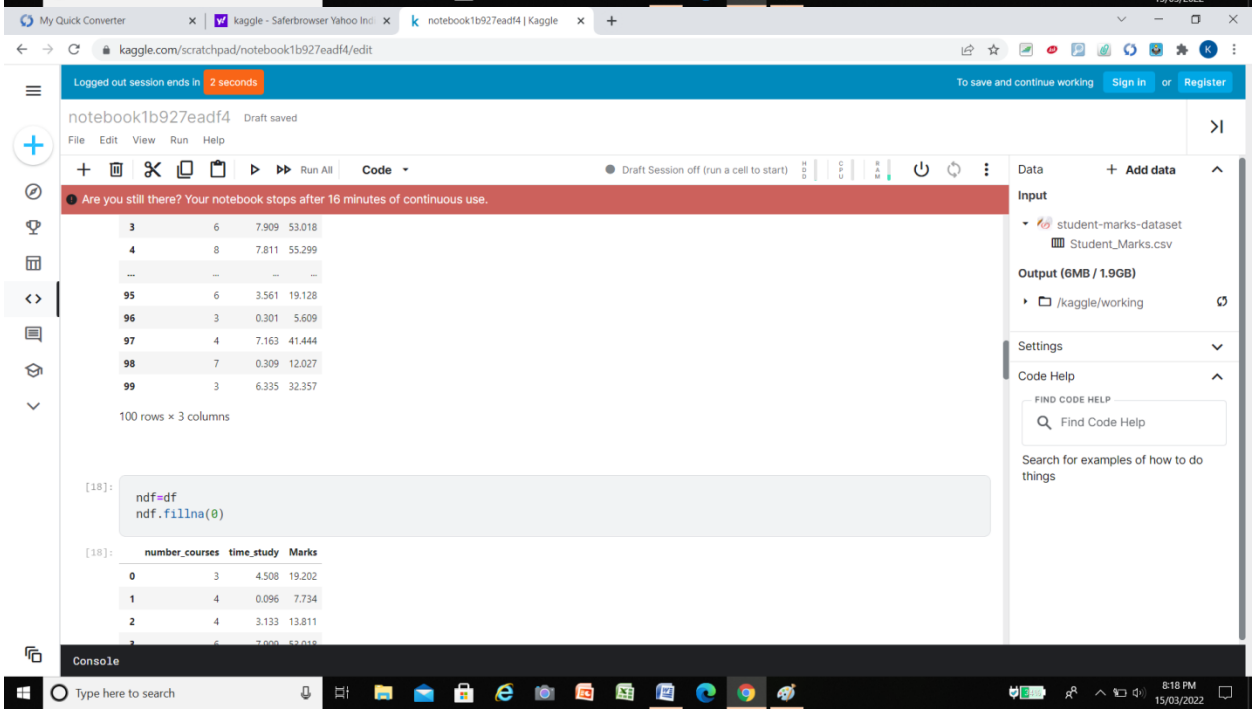

New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[10]: df.isnull()
[10]:
number_courses time_study Marks
0 False False False
1 False False False
2 False False False
3 False False False
4 False False False
... ... ... ...
95 False False False
96 False False False
97 False False False
98 False False False
99 False False False
100 rows × 3 columns
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:16 PM
15/03/2022

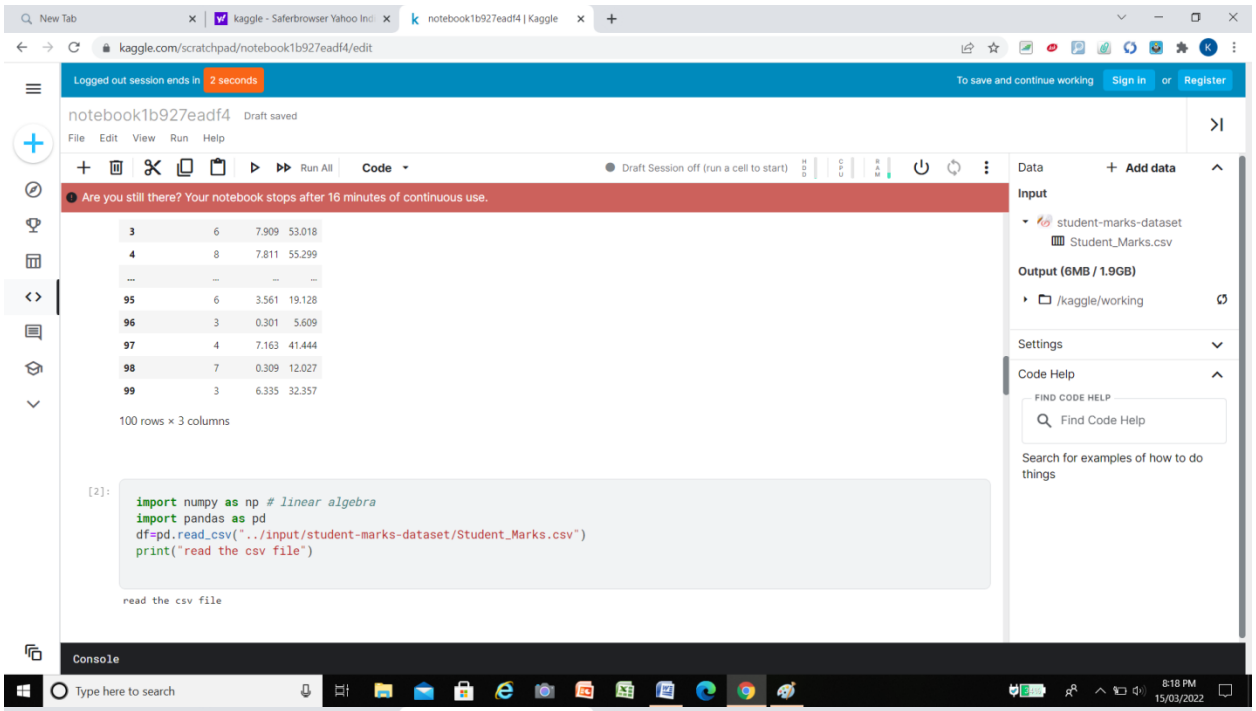

My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[11]: series=pd.isnull(df["Marks"])
df[series]
[11]:
number_courses time_study Marks
[12]: df.notnull()
[12]:
number_courses time_study Marks
0 True True True
1 True True True
2 True True True
3 True True True
4 True True True
... ... ... ...
95 True True True
96 True True True
97 True True True
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:17 PM
15/03/2022

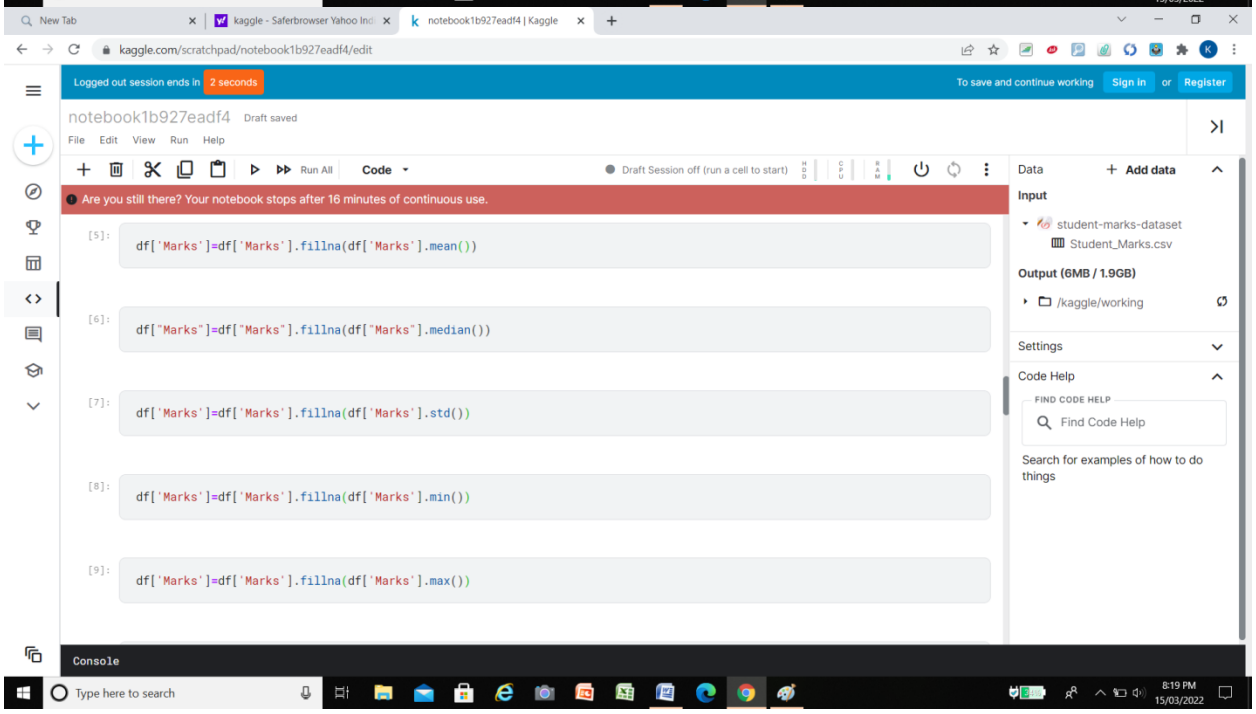

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
97 True True True
98 True True True
99 True True True
100 rows × 3 columns
[13]:
series1=pd.notnull(df["Marks"])
df[series1]
[13]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:17 PM
15/03/2022

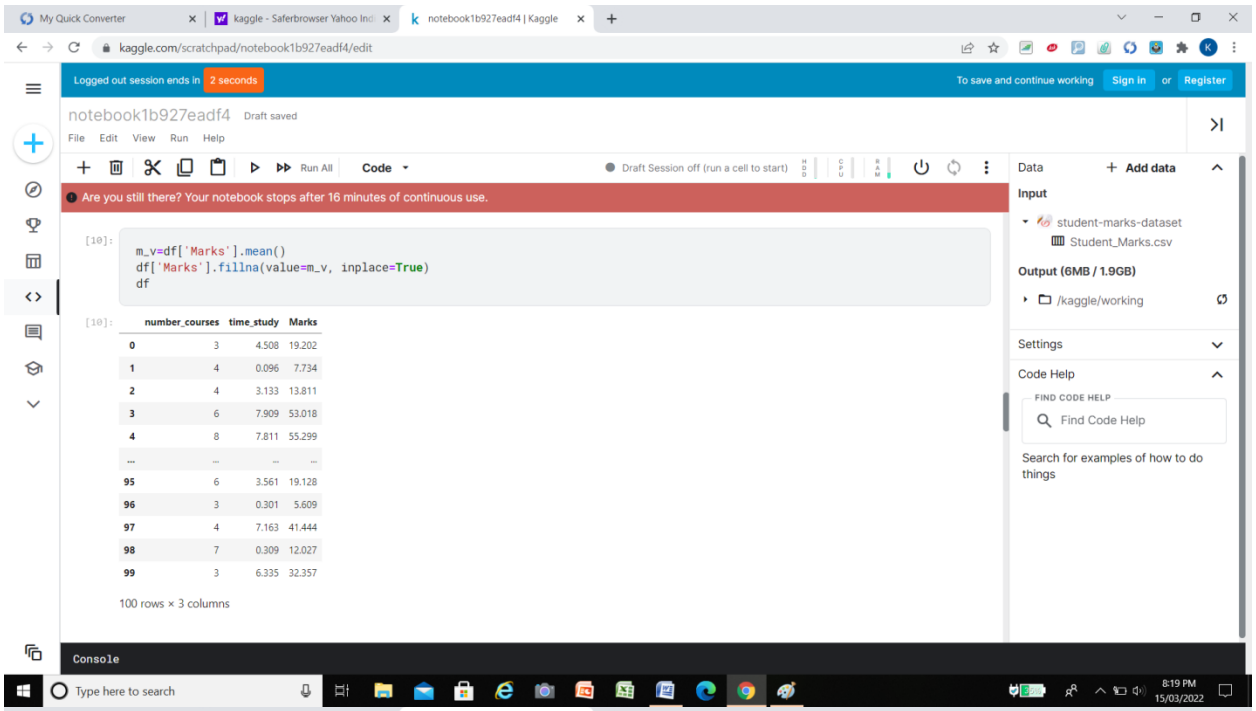

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
[14]:
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['time_study']=le.fit_transform(df['time_study'])
newdf=df
df
[14]:
number_courses time_study Marks
0 3 59 19.202
1 4 0 7.734
2 4 35 13.811
3 6 98 53.018
4 8 97 55.299
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:17 PM
15/03/2022

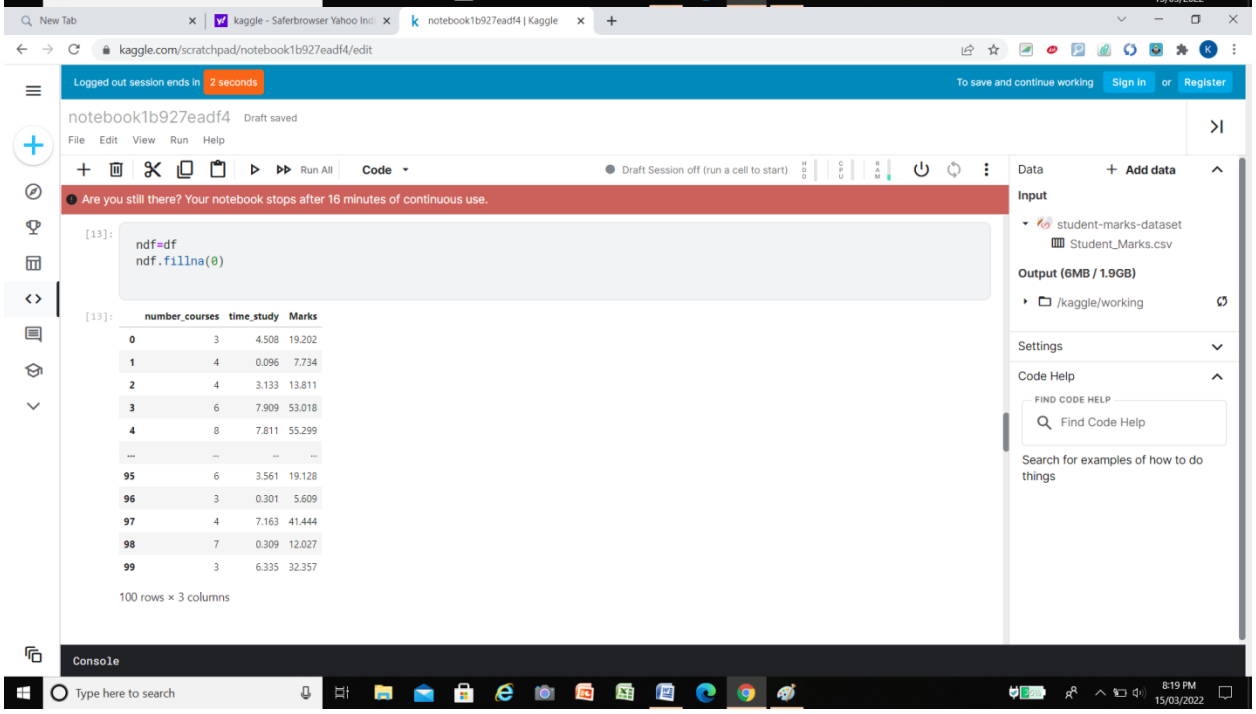

New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
4 8 97 55.299
... ... ... ...
95 6 39 19.128
96 3 4 5.609
97 4 85 41.444
98 7 5 12.027
99 3 77 32.357
100 rows × 3 columns
[17]:
missing_values=["Na", "na"]
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv", na_values = missing_values)
df
[17]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
Console
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:18 PM
15/03/2022

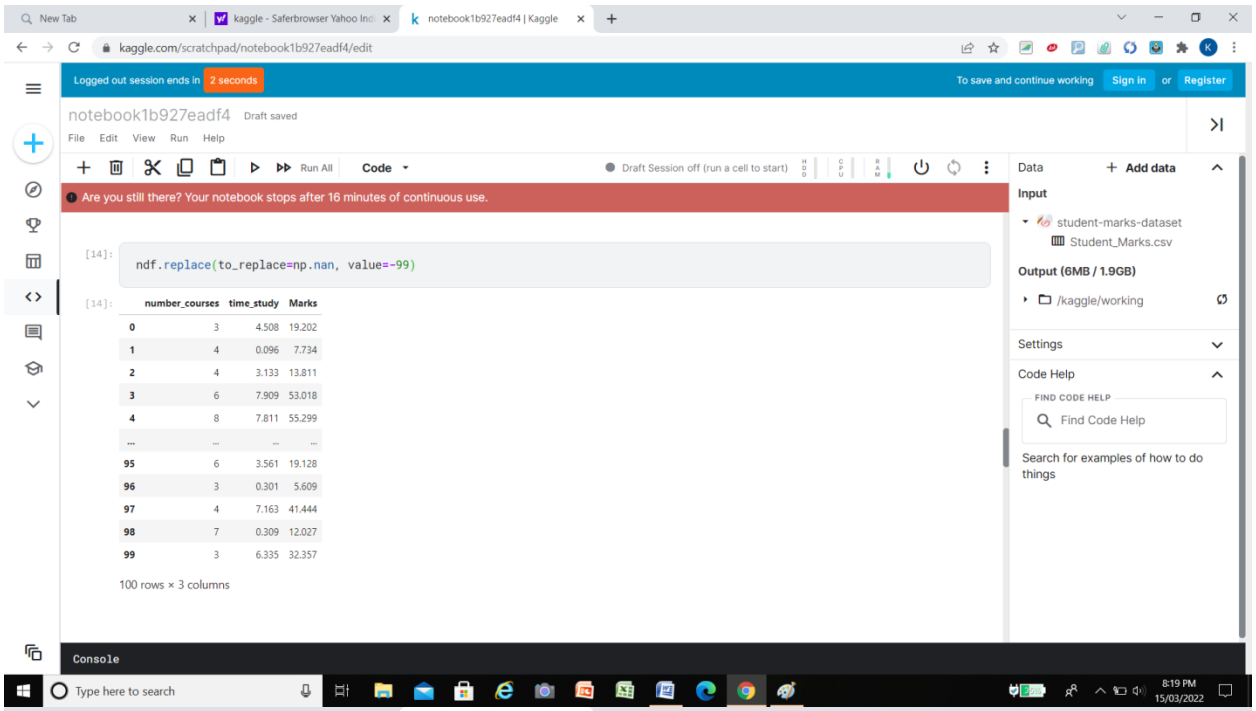

My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
[18]:
ndf=df
ndf.fillna(0)
[18]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
Console
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:18 PM
15/03/2022

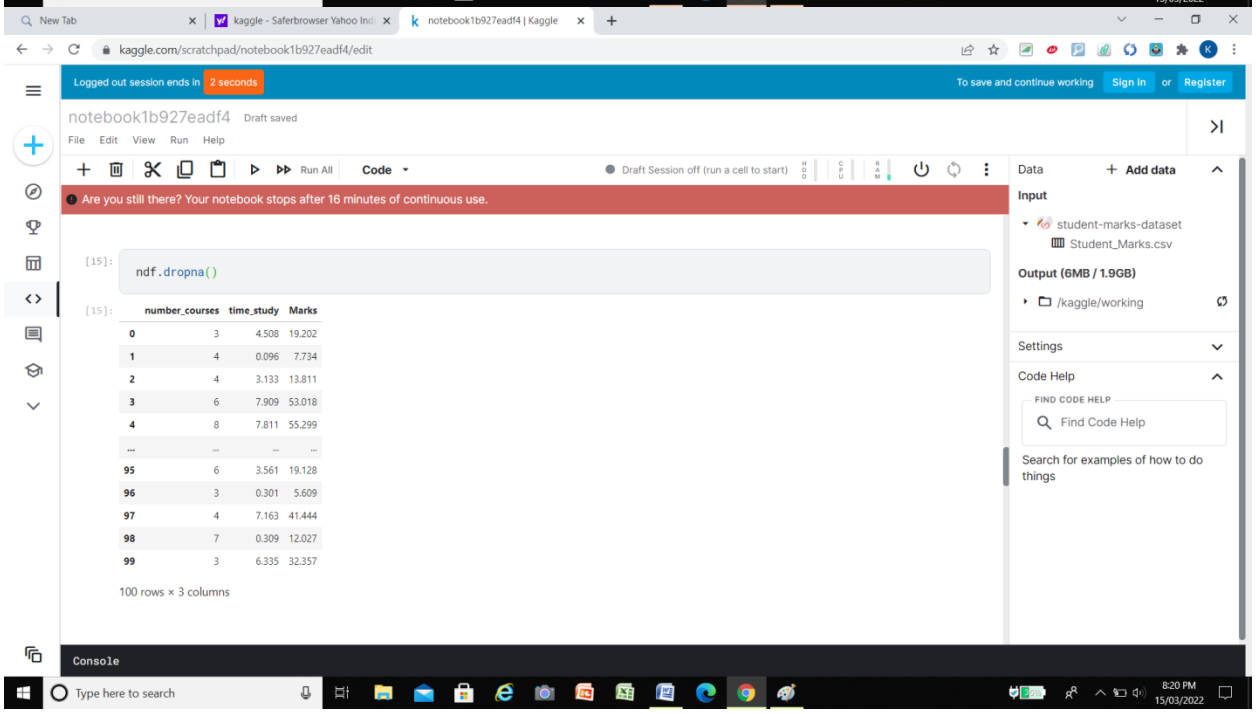

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
[2]:
import numpy as np # linear algebra
import pandas as pd
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
print("read the csv file")
read the csv file
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:18 PM
15/03/2022

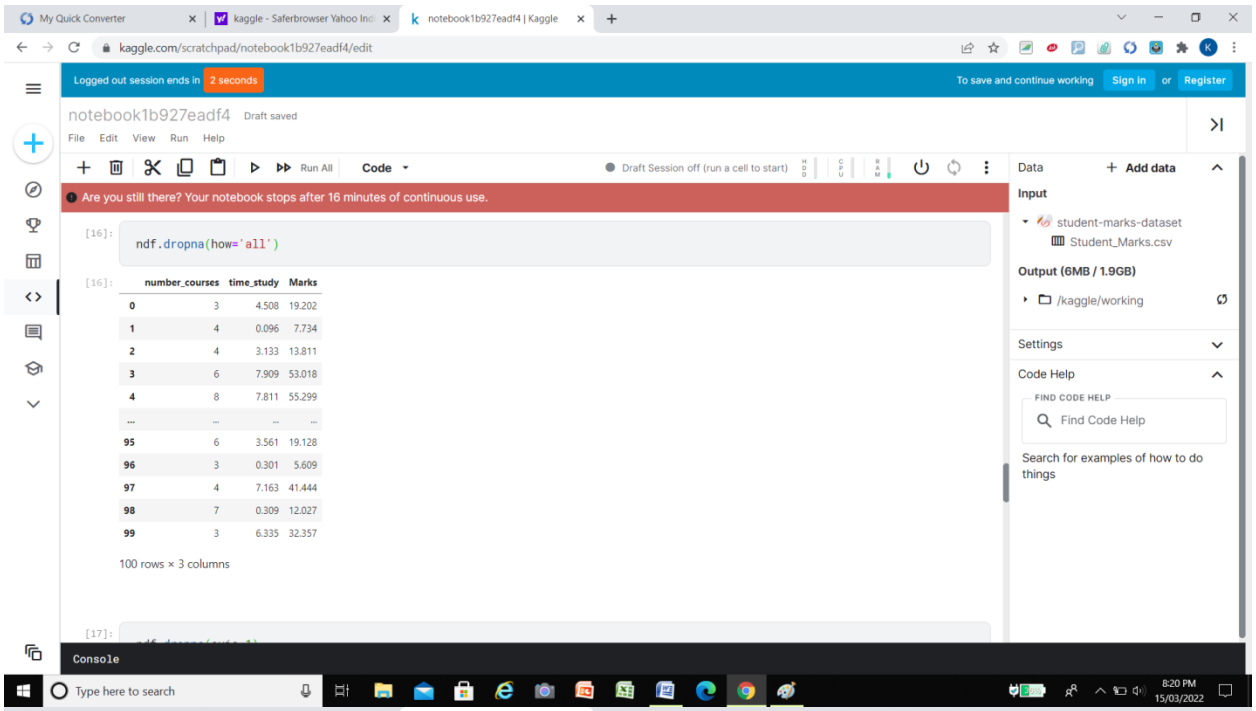

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[5]:
df['Marks']=df['Marks'].fillna(df['Marks'].mean())
[6]:
df["Marks"]=df["Marks"].fillna(df["Marks"].median())
[7]:
df['Marks']=df['Marks'].fillna(df['Marks'].std())
[8]:
df['Marks']=df['Marks'].fillna(df['Marks'].min())
[9]:
df['Marks']=df['Marks'].fillna(df['Marks'].max())
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:19 PM
15/03/2022

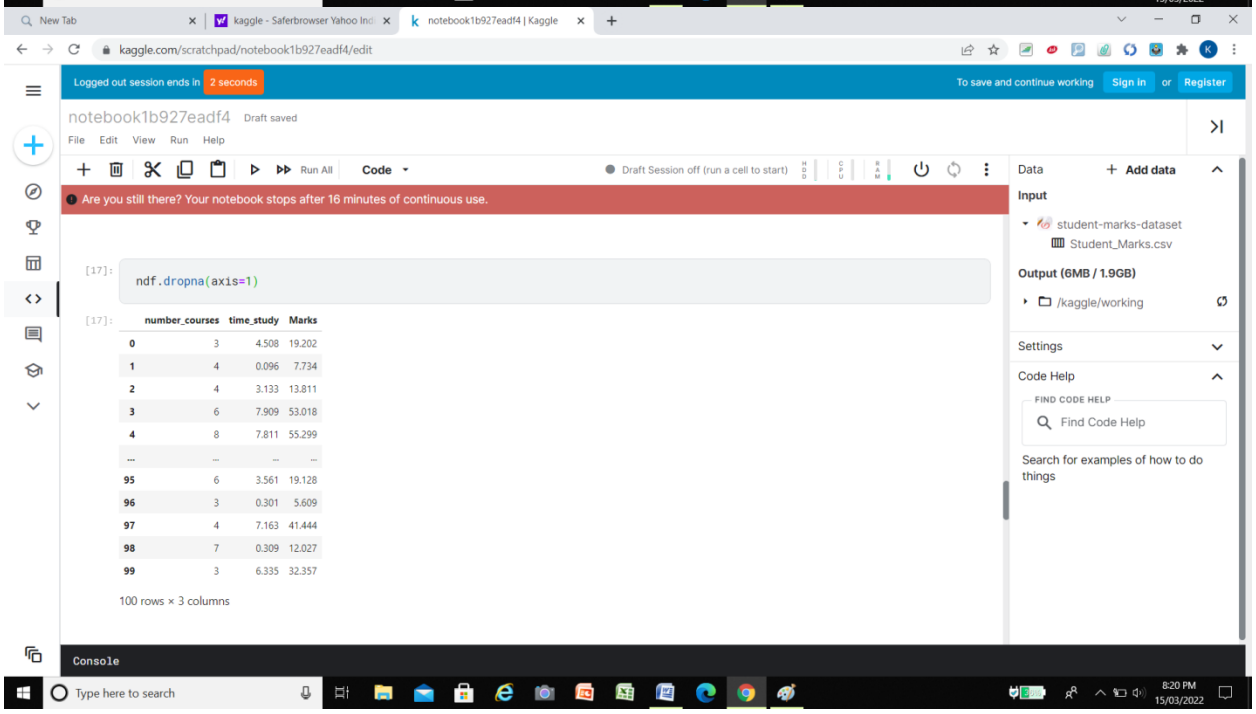

My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[10]:
m_v=df['Marks'].mean()
df['Marks'].fillna(value=m_v, inplace=True)
df
[10]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:19 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[13]:
ndf=df
ndf.fillna(0)
[13]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:19 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[14]: ndf.replace(to_replace=np.nan, value=-99)
[14]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:19 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[15]: ndf.dropna()
[15]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:20 PM
15/03/2022


My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[16]:
ndf.dropna(how='all')
[16]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
[17]:
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:20 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[17]:
ndf.dropna(axis=1)
[17]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
8:20 PM
15/03/2022

My Quick Converter | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session off (run a cell to start)

Are you still there? Your notebook stops after 16 minutes of continuous use.

[18]:
```
new_data=ndf.dropna(axis=0, how='any')
new_data
```

[18]:

| | number_courses | time_study | Marks |
|---|---|---|---|
| 0 | 3 | 4.508 | 19.202 |
| 1 | 4 | 0.096 | 7.734 |
| 2 | 4 | 3.133 | 13.811 |
| 3 | 6 | 7.909 | 53.018 |
| 4 | 8 | 7.811 | 55.299 |
| ... | ... | ... | ... |
| 95 | 6 | 3.561 | 19.128 |
| 96 | 3 | 0.301 | 5.609 |
| 97 | 4 | 7.163 | 41.444 |
| 98 | 7 | 0.309 | 12.027 |
| 99 | 3 | 6.335 | 32.357 |

100 rows × 3 columns

Data | Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Console

Type here to search — 8:21 PM 15/03/2022

My Quick Converter | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

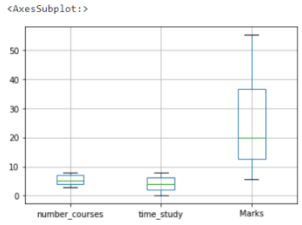
Run All | Code | Draft Session off (run a cell to start)

Are you still there? Your notebook stops after 16 minutes of continuous use.

100 rows × 3 columns

[19]:
```
col=['number_courses','time_study', 'Marks']
df.boxplot(col)
```

[19]:




[2]:
```
import numpy as np # linear algebra
```

Data | Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

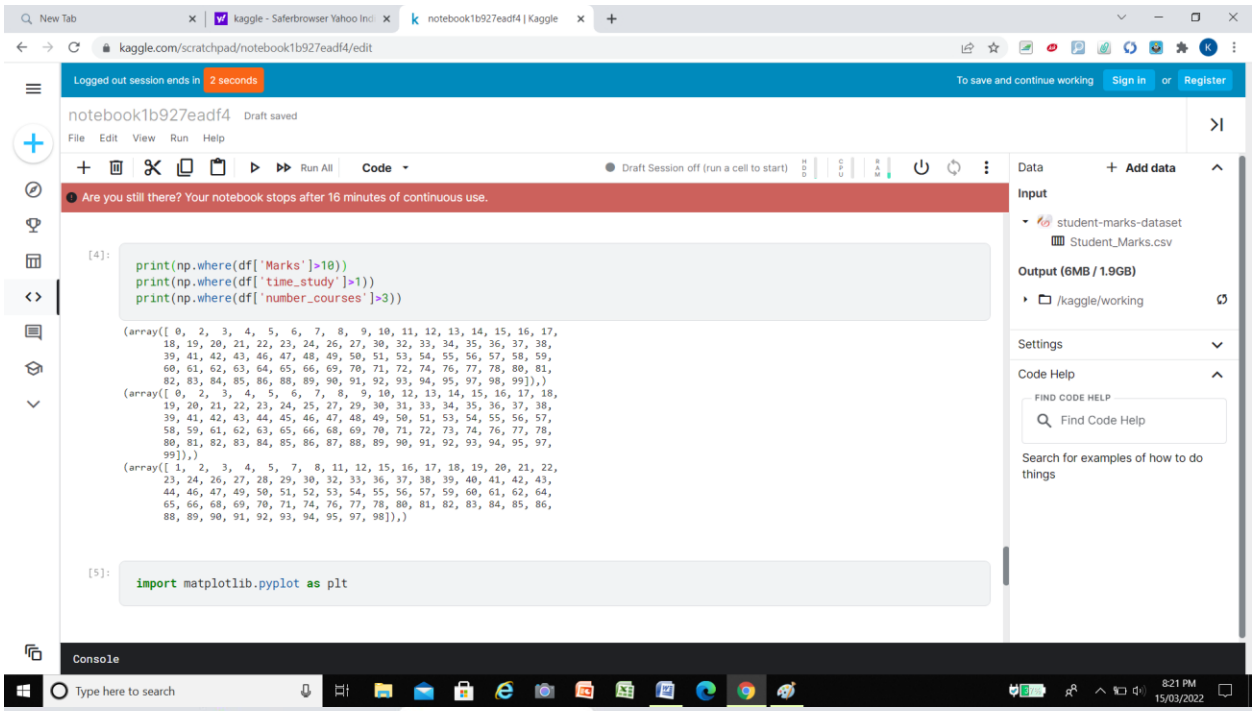
FIND CODE HELP

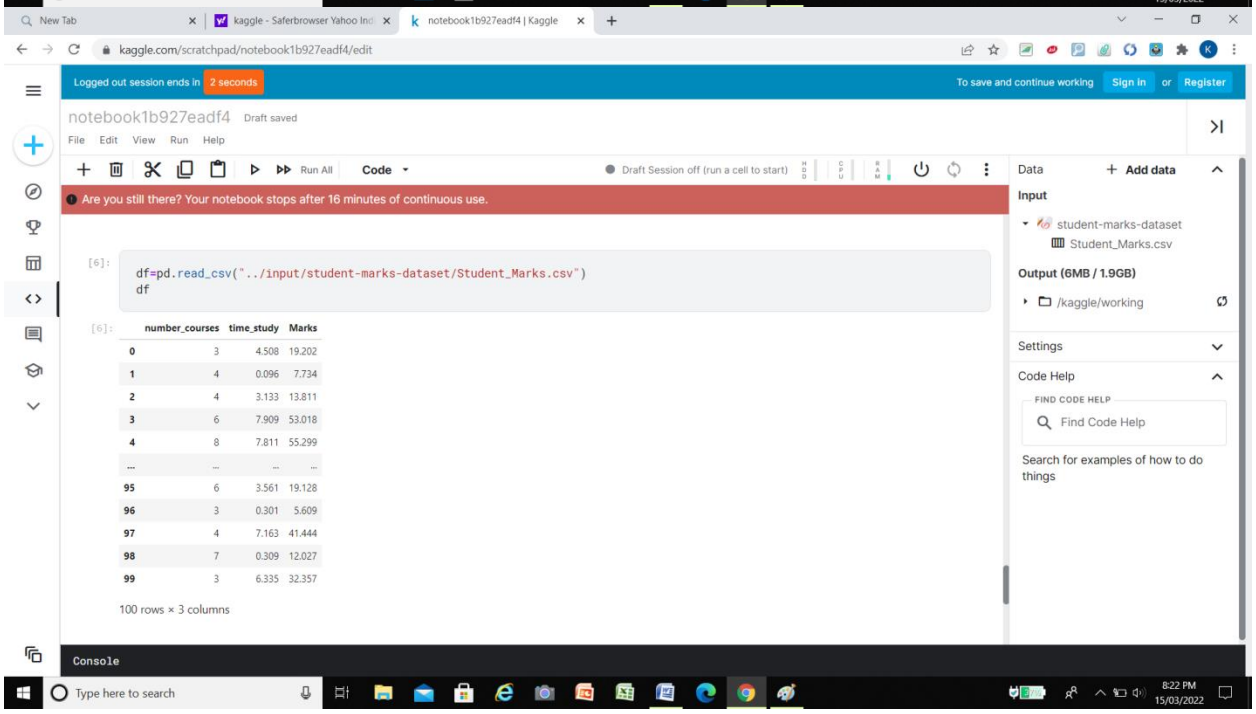
Find Code Help

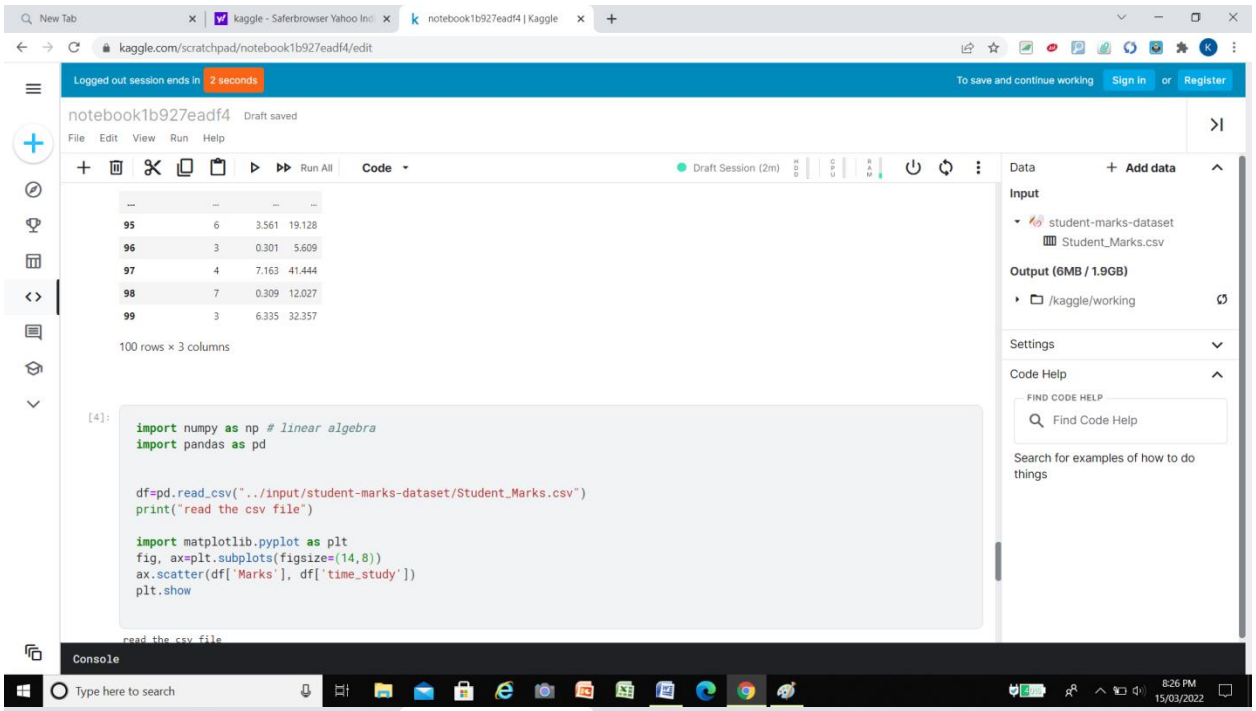
Search for examples of how to do things

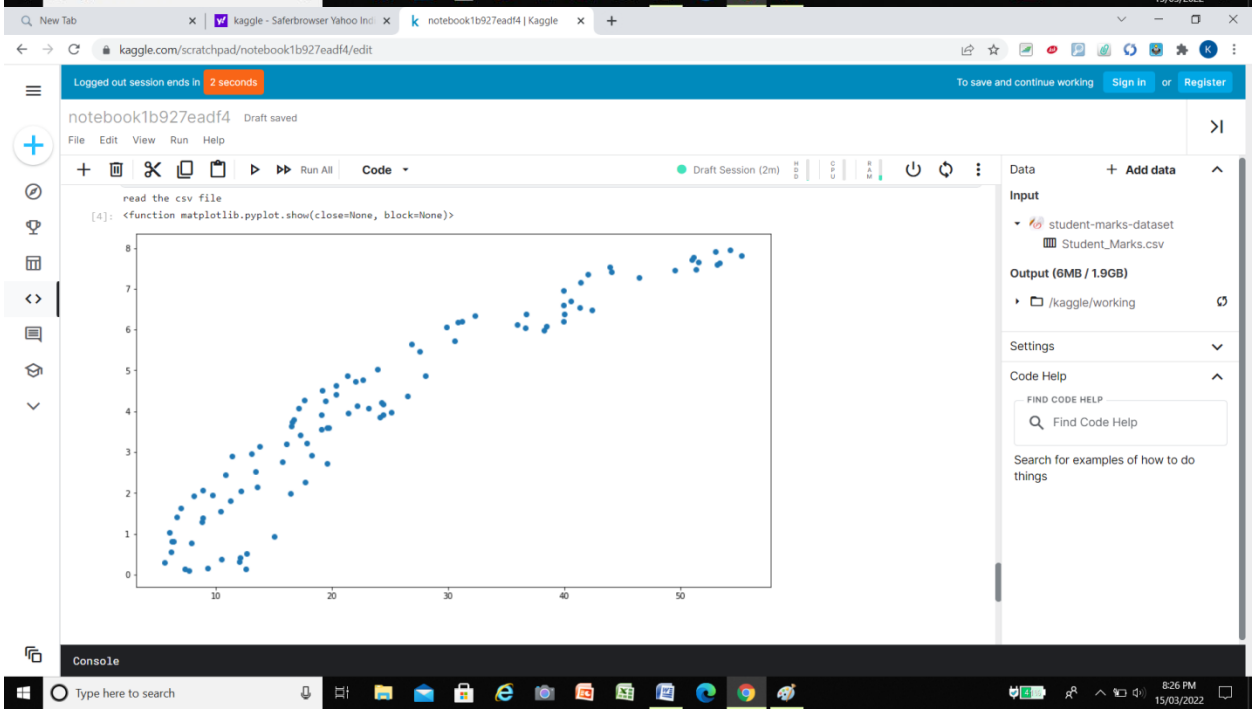
Console

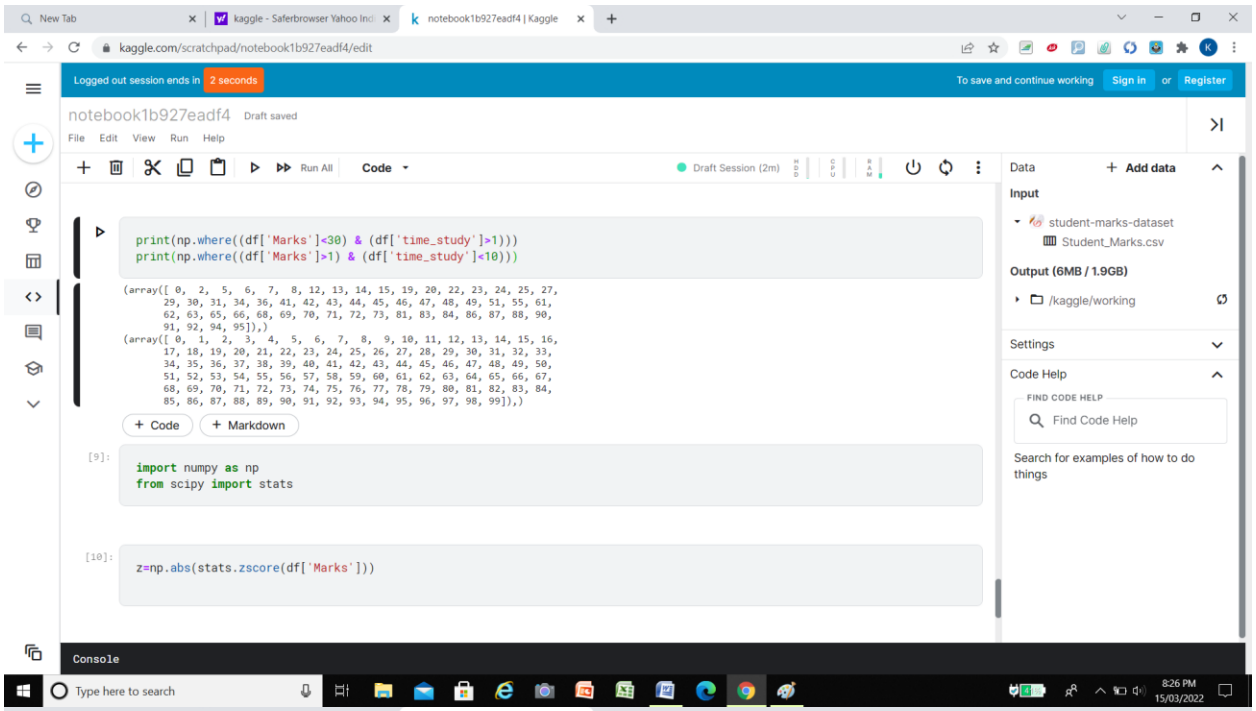
Type here to search — 8:21 PM 15/03/2022

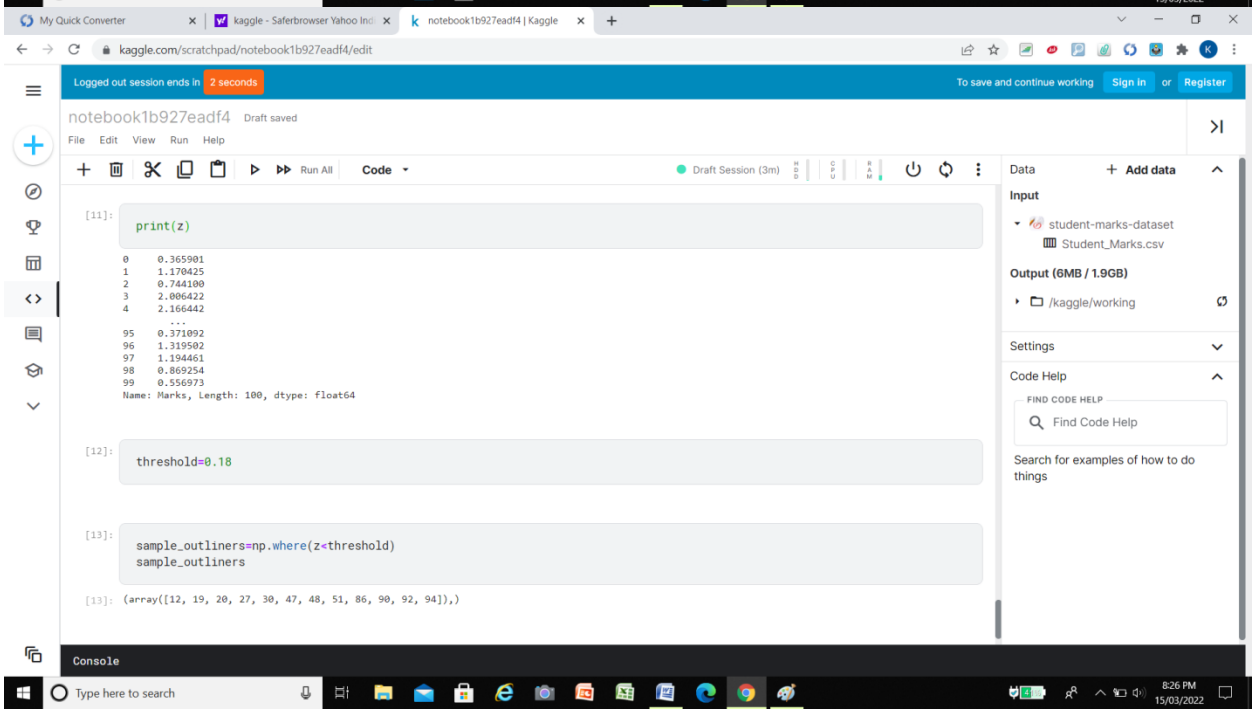

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[4]:
print(np.where(df['Marks']>10))
print(np.where(df['time_study']>1))
print(np.where(df['number_courses']>3))
(array([ 0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20, 21, 22, 23, 24, 26, 27, 30, 32, 33, 34, 35, 36, 37, 38,
39, 41, 42, 43, 46, 47, 48, 49, 50, 51, 53, 54, 55, 56, 57, 58, 59,
60, 61, 62, 63, 64, 65, 66, 69, 70, 71, 72, 74, 76, 77, 78, 80, 81,
82, 83, 84, 85, 86, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98, 99]),)
(array([ 0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18,
19, 20, 21, 22, 23, 24, 25, 27, 29, 30, 31, 33, 34, 35, 36, 37, 38,
39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 53, 54, 55, 56, 57,
58, 59, 61, 62, 63, 65, 66, 68, 69, 70, 71, 72, 73, 74, 76, 77, 78,
80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 97,
99]),)
(array([ 1, 2, 3, 4, 5, 7, 8, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22,
23, 24, 26, 27, 28, 29, 30, 32, 33, 36, 37, 38, 39, 40, 41, 42, 43,
44, 46, 47, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 62, 64,
65, 66, 68, 69, 70, 71, 74, 76, 77, 78, 80, 81, 82, 83, 84, 85, 86,
88, 89, 90, 91, 92, 93, 94, 95, 97, 98]),)
[5]:
import matplotlib.pyplot as plt
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:21 PM
15/03/2022

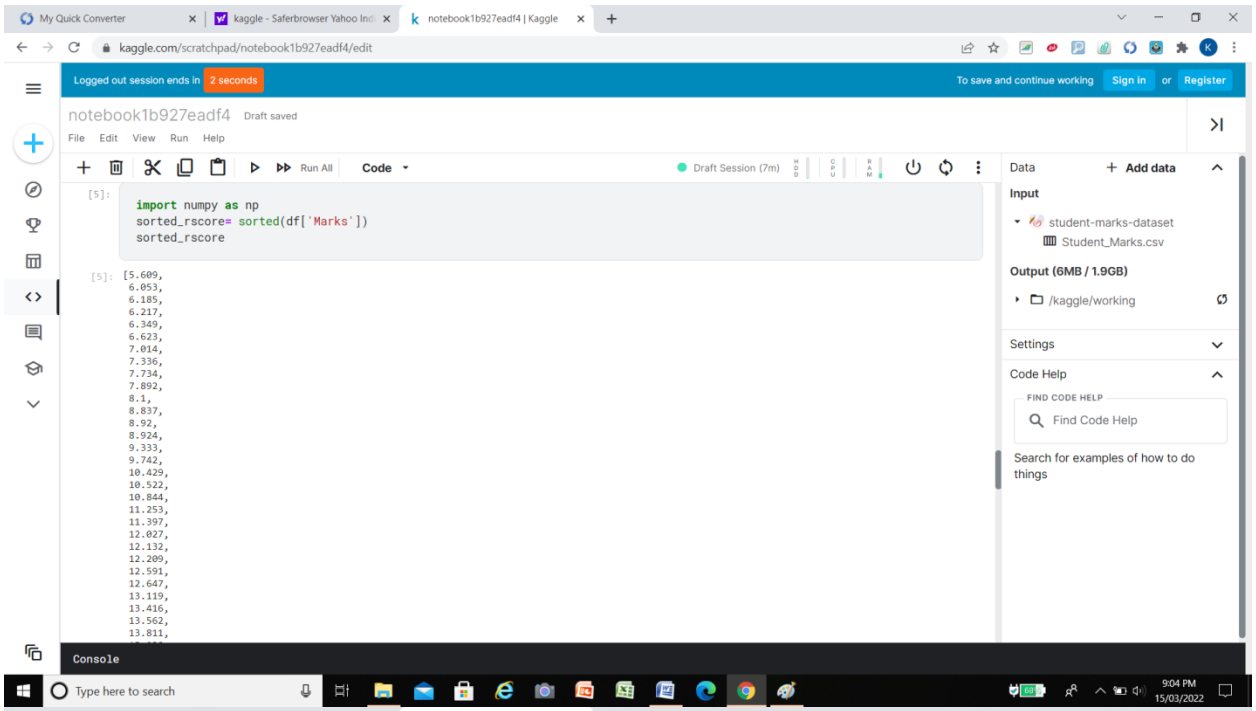

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code
Draft Session off (run a cell to start)
Are you still there? Your notebook stops after 16 minutes of continuous use.
[6]:
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
df
[6]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:22 PM
15/03/2022

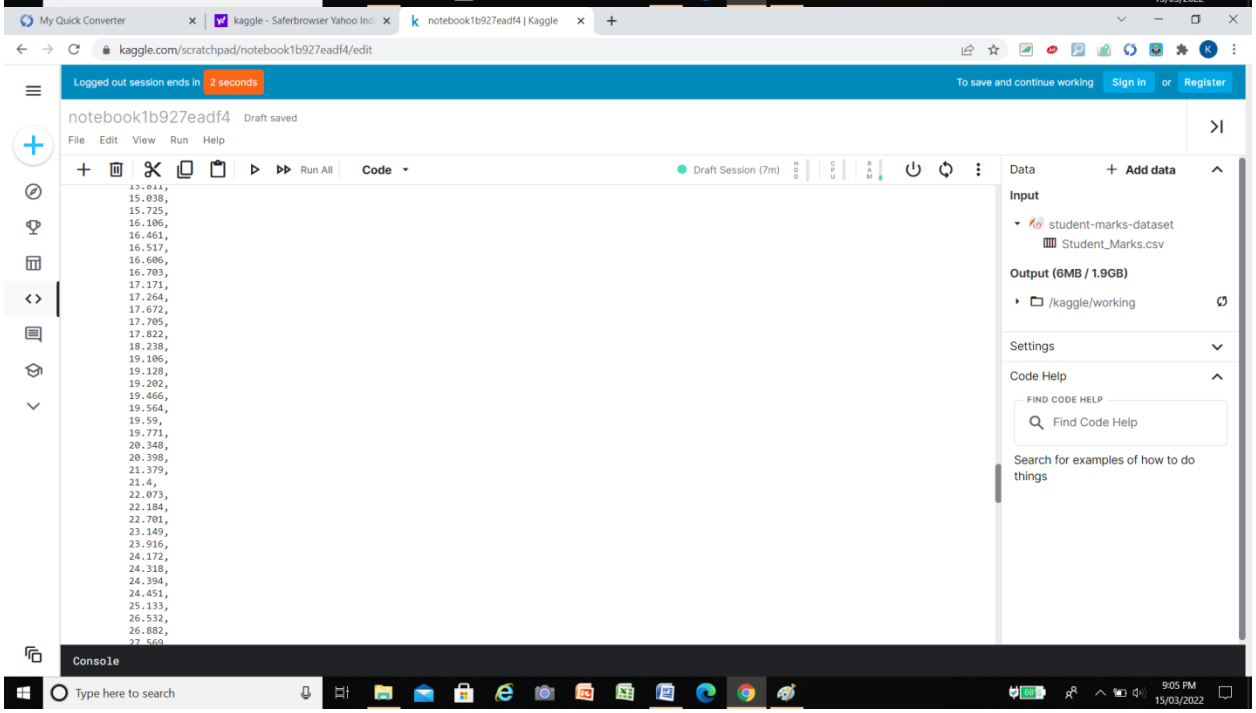

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session (2m)
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
[4]:
import numpy as np # linear algebra
import pandas as pd
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
print("read the csv file")
import matplotlib.pyplot as plt
fig, ax=plt.subplots(figsize=(14,8))
ax.scatter(df['Marks'], df['time_study'])
plt.show
read the csv file
Console
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:26 PM
15/03/2022

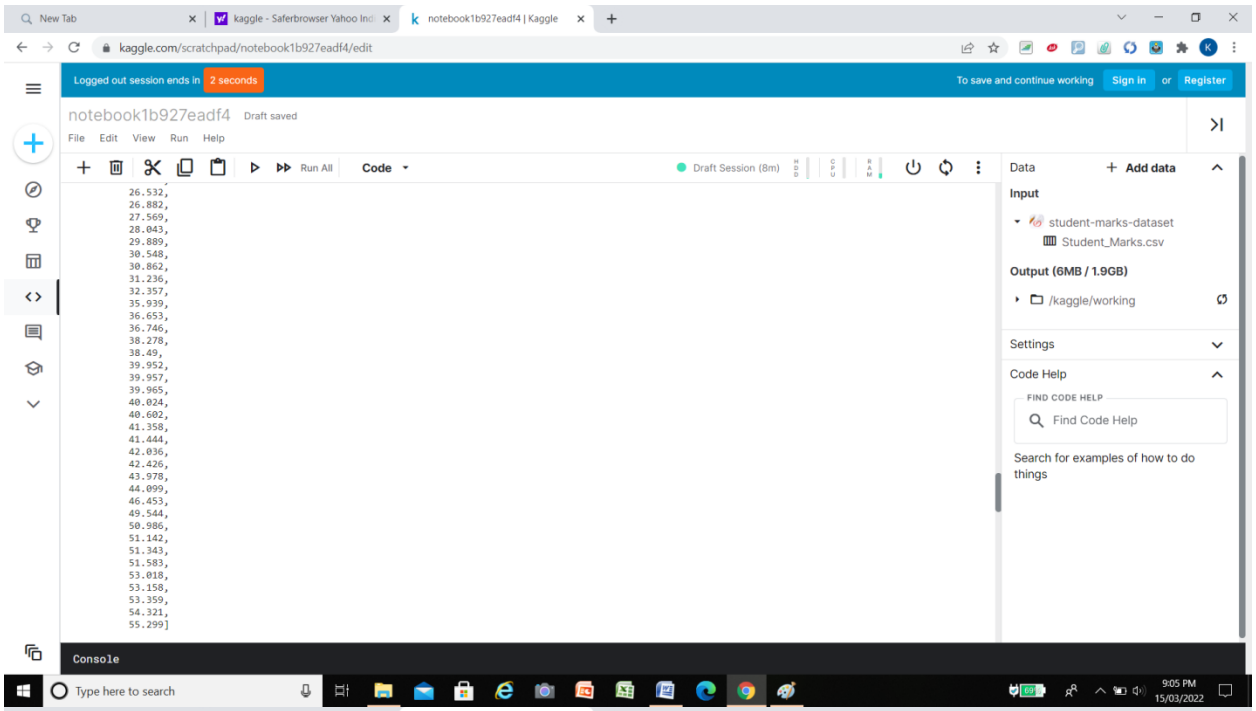

New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session (2m)
read the csv file
[4]: <function matplotlib.pyplot.show(close=None, block=None)>
0 1 2 3 4 5 6 7 8
10 20 30 40 50
Console
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:26 PM
15/03/2022

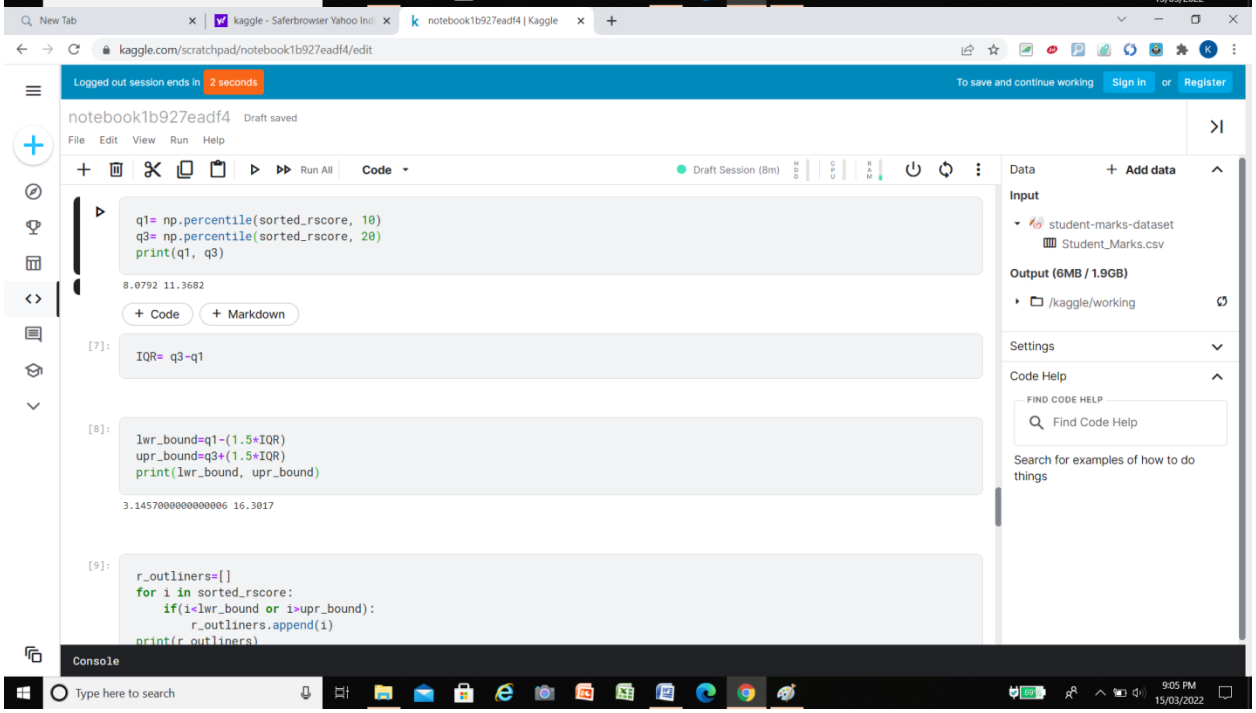

New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (2m)
print(np.where((df['Marks']<30) & (df['time_study']>1)))
print(np.where((df['Marks']>1) & (df['time_study']<10)))
(array([ 0, 2, 5, 6, 7, 8, 12, 13, 14, 15, 19, 20, 22, 23, 24, 25, 27,
29, 30, 31, 34, 36, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 55, 61,
62, 63, 65, 66, 68, 69, 70, 71, 72, 73, 81, 83, 84, 86, 87, 88, 90,
91, 92, 94, 95]),)
(array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67,
68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84,
85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]),)
+ Code + Markdown
[9]:
import numpy as np
from scipy import stats
[10]:
z=np.abs(stats.zscore(df['Marks']))
Console
Data + Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:26 PM
15/03/2022

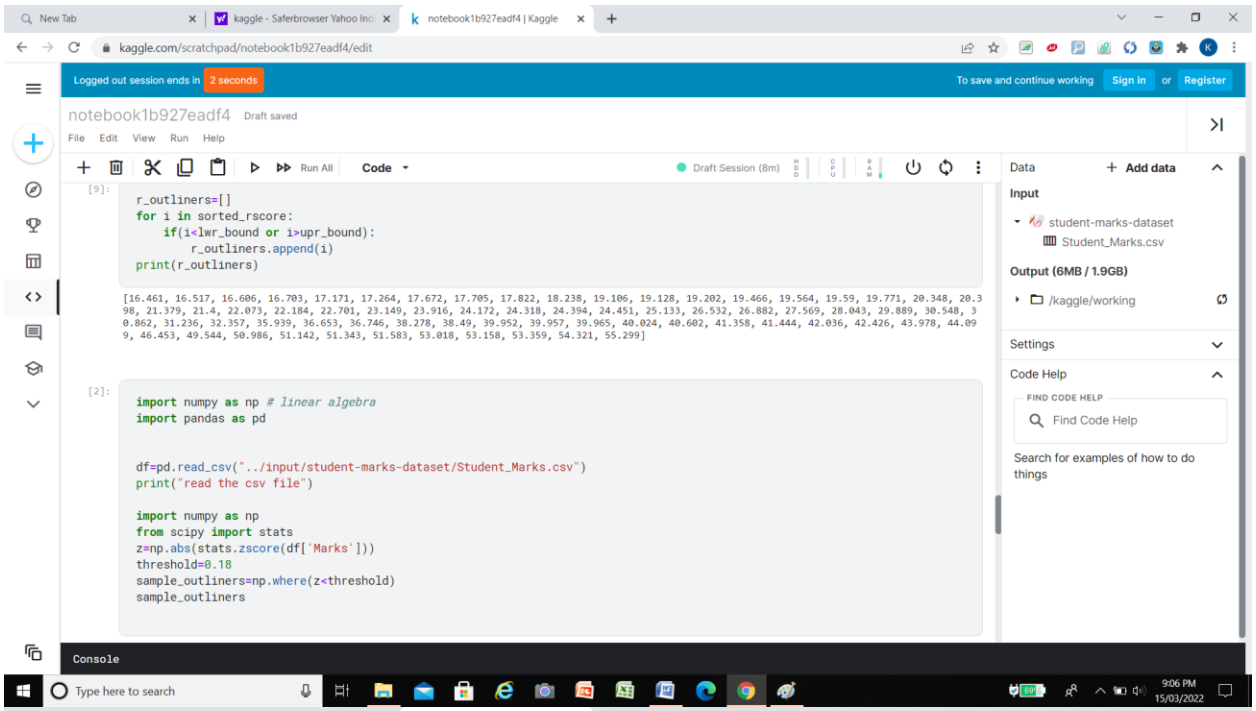

My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (3m)
[11]:
print(z)
0 0.365901
1 1.170425
2 0.744100
3 2.006422
4 2.166442
...
95 0.371092
96 1.319502
97 1.194461
98 0.869254
99 0.556973
Name: Marks, Length: 100, dtype: float64
[12]:
threshold=0.18
[13]:
sample_outliners=np.where(z<threshold)
sample_outliners
[13]: (array([12, 19, 20, 27, 30, 47, 48, 51, 86, 90, 92, 94]),)
Console
Data + Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
8:26 PM
15/03/2022

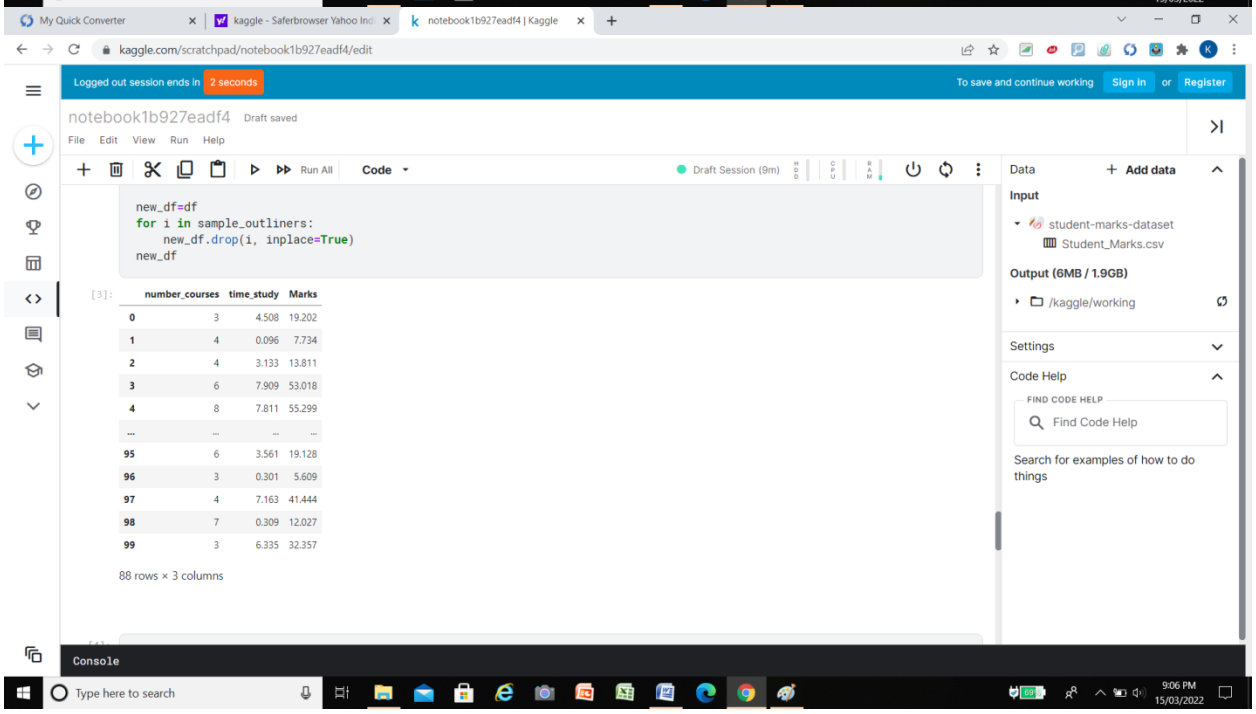

My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (7m)
[5]:
import numpy as np
sorted_rscore= sorted(df['Marks'])
sorted_rscore
[5]: [5.609,
6.053,
6.185,
6.217,
6.349,
6.623,
7.014,
7.336,
7.734,
7.892,
8.1,
8.837,
8.92,
8.924,
9.333,
9.742,
10.429,
10.522,
10.844,
11.253,
11.397,
12.027,
12.132,
12.209,
12.591,
12.647,
13.119,
13.416,
13.562,
13.811,
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
9:04 PM
15/03/2022


My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (7m)
15.038,
15.725,
16.106,
16.461,
16.517,
16.606,
16.703,
17.171,
17.264,
17.672,
17.705,
17.822,
18.238,
19.106,
19.128,
19.202,
19.466,
19.564,
19.59,
19.771,
20.348,
20.398,
21.379,
21.4,
22.073,
22.184,
22.701,
23.149,
23.916,
24.172,
24.318,
24.394,
24.451,
25.133,
26.532,
26.882,
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
9:05 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session (8m)
26.532,
26.882,
27.569,
28.043,
29.889,
30.548,
30.862,
31.236,
32.357,
35.939,
36.653,
36.746,
38.278,
38.49,
39.952,
39.957,
39.965,
40.024,
40.602,
41.358,
41.444,
42.036,
42.426,
43.978,
44.099,
46.453,
49.544,
50.986,
51.142,
51.343,
51.583,
53.018,
53.158,
53.359,
54.321,
55.299]
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
9:05 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session (8m)
q1= np.percentile(sorted_rscore, 10)
q3= np.percentile(sorted_rscore, 20)
print(q1, q3)
8.0792 11.3682
+ Code
+ Markdown
[7]:
IQR= q3-q1
[8]:
lwr_bound=q1-(1.5*IQR)
upr_bound=q3+(1.5*IQR)
print(lwr_bound, upr_bound)
3.1457000000000006 16.3017
[9]:
r_outliners=[]
for i in sorted_rscore:
if(i<lwr_bound or i>upr_bound):
r_outliners.append(i)
print(r_outliners)
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Console
Type here to search
9:05 PM
15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (8m)
[9]:
r_outliners=[]
for i in sorted_rscore:
if(i<lwr_bound or i>upr_bound):
r_outliners.append(i)
print(r_outliners)
[16.461, 16.517, 16.606, 16.703, 17.171, 17.264, 17.672, 17.705, 17.822, 18.238, 19.106, 19.128, 19.202, 19.466, 19.564, 19.59, 19.771, 20.348, 20.398, 21.379, 21.4, 22.073, 22.184, 22.701, 23.149, 23.916, 24.172, 24.318, 24.394, 24.451, 25.133, 26.532, 26.882, 27.569, 28.043, 29.889, 30.548, 30.862, 31.236, 32.357, 35.939, 36.653, 36.746, 38.278, 38.49, 39.952, 39.957, 39.965, 40.024, 40.602, 41.358, 41.444, 42.036, 42.426, 43.978, 44.099, 46.453, 49.544, 50.986, 51.142, 51.343, 51.583, 53.018, 53.158, 53.359, 54.321, 55.299]
[2]:
import numpy as np # linear algebra
import pandas as pd
df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
print("read the csv file")
import numpy as np
from scipy import stats
z=np.abs(stats.zscore(df['Marks']))
threshold=0.18
sample_outliners=np.where(z<threshold)
sample_outliners
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
9:06 PM
15/03/2022


My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (9m)
new_df=df
for i in sample_outliners:
new_df.drop(i, inplace=True)
new_df
[3]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
88 rows × 3 columns
Console
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
9:06 PM
15/03/2022

New Tab | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session (9m)

[4]:
```
df_stud=df
ninetieth_percentile=np.percentile(df_stud['Marks'],30)
b = np.where(df_stud['Marks']>ninetieth_percentile, ninetieth_percentile, df_stud['Marks'])
print("New array: " ,b)
```

```
New array:  [13.1487  7.734  13.1487 13.1487 13.1487 13.1487 13.1487 13.1487 13.1487
 13.1487 13.1487 12.132  13.1487 11.397  13.1487 13.1487 13.1487 13.1487
 13.1487 12.209  13.1487 13.1487  6.623  12.647   9.333   8.837   8.1
 13.1487 13.1487 13.1487 13.1487 13.119  13.1487 13.1487 13.1487  7.336
 13.1487 13.1487 10.429   9.742   8.924  13.1487 13.1487 13.1487  7.892
 13.1487 13.1487 13.1487 13.1487 13.1487 13.1487 13.1487 10.522  10.844
 13.1487 13.1487 12.591  13.1487 13.1487  6.185   8.92   13.1487 13.1487
 13.1487 13.1487  7.014  13.1487  6.217  13.1487 13.1487 13.1487  6.349
 13.1487 13.1487 13.1487 13.1487 13.1487 13.1487  6.053  11.253  13.1487
 13.1487 13.1487 13.1487  5.609  13.1487 12.027  13.1487]
```

[5]:
```
df_stud.insert(1,"time_study", b, True)
df_stud
```

[5]:

| | number_courses | time_study | time_study | Marks |
|---|---|---|---|---|
| 0 | 3 | 13.1487 | 4.508 | 19.202 |
| 1 | 4 | 7.7340 | 0.096 | 7.734 |
| 2 | 4 | 13.1487 | 3.133 | 13.811 |
| 3 | 6 | 13.1487 | 7.909 | 53.018 |

Console

Data | + Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Type here to search — 9:06 PM 15/03/2022

My Quick Converter | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session (10m)

| | | | | |
|---|---|---|---|---|
| 1 | 4 | 7.7340 | 0.096 | 7.734 |
| 2 | 4 | 13.1487 | 3.133 | 13.811 |
| 3 | 6 | 13.1487 | 7.909 | 53.018 |
| 4 | 8 | 13.1487 | 7.811 | 55.299 |
| ... | ... | ... | ... | ... |
| 95 | 6 | 13.1487 | 3.561 | 19.128 |
| 96 | 3 | 5.6090 | 0.301 | 5.609 |
| 97 | 4 | 13.1487 | 7.163 | 41.444 |
| 98 | 7 | 12.0270 | 0.309 | 12.027 |
| 99 | 3 | 13.1487 | 6.335 | 32.357 |

88 rows × 4 columns

[6]:
```
col=['number_courses']
df.boxplot(col)
```

[6]:
```
<AxesSubplot:>
```


Console

Data | + Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Type here to search — 9:07 PM 15/03/2022

My Quick Converter | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

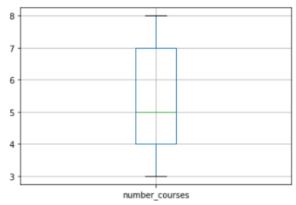
notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session (10m)

[6]:
```
col=['number_courses']
df.boxplot(col)
```

[6]: <AxesSubplot:>




[8]:
```
import numpy as np
sorted_rscore= sorted(df['Marks'])
sorted_rscore
median=np.median(sorted_rscore)
median
```

Data | Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Console

Type here to search

9:07 PM 15/03/2022

My Quick Converter | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session (10m)

[8]: 19.117

[3]:
```
import numpy as np # linear algebra
import pandas as pd


df=pd.read_csv("../input/student-marks-dataset/Student_Marks.csv")
print("read the csv file")
import numpy as np
sorted_rscore= sorted(df['Marks'])
sorted_rscore
median=np.median(sorted_rscore)
median
q1= np.percentile(sorted_rscore, 10)
q3= np.percentile(sorted_rscore, 20)
print(q1, q3)
IQR= q3-q1
lwr_bound=q1-(1.5*IQR)
upr_bound=q3+(1.5*IQR)
print(lwr_bound, upr_bound)
```

```
read the csv file
8.0792 11.3682
3.1457000000000006 16.3017
```

Data | Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

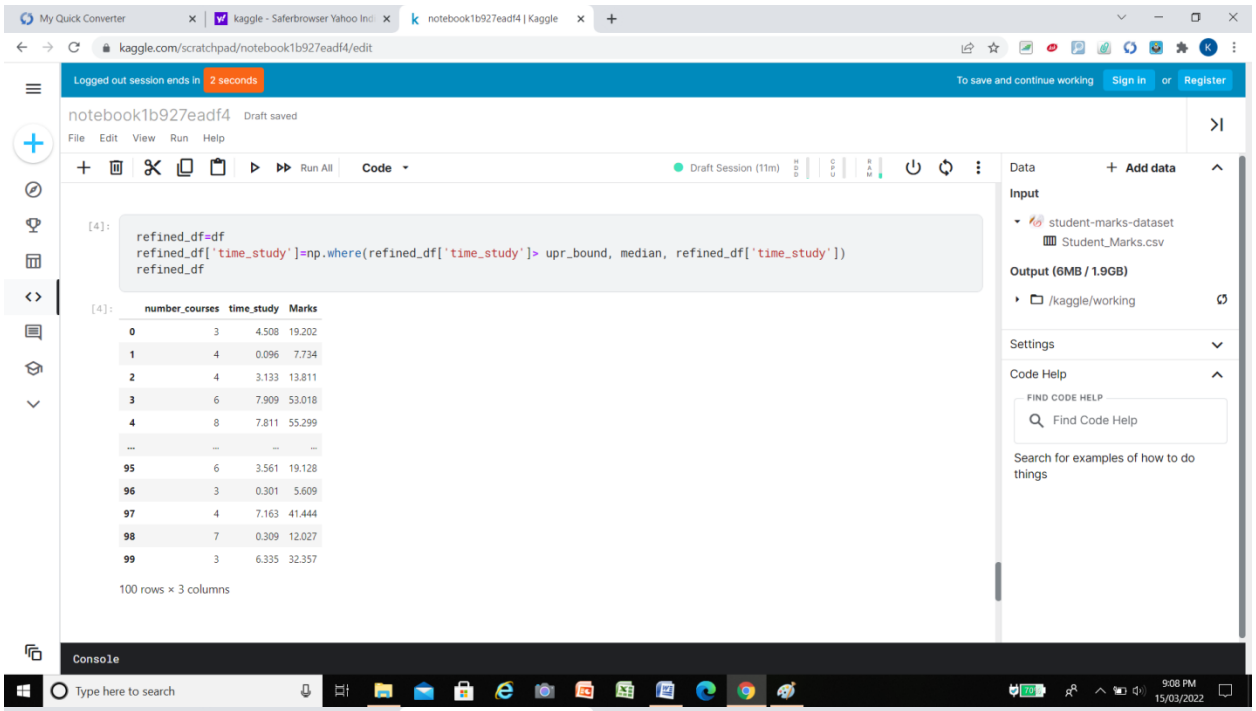
Find Code Help

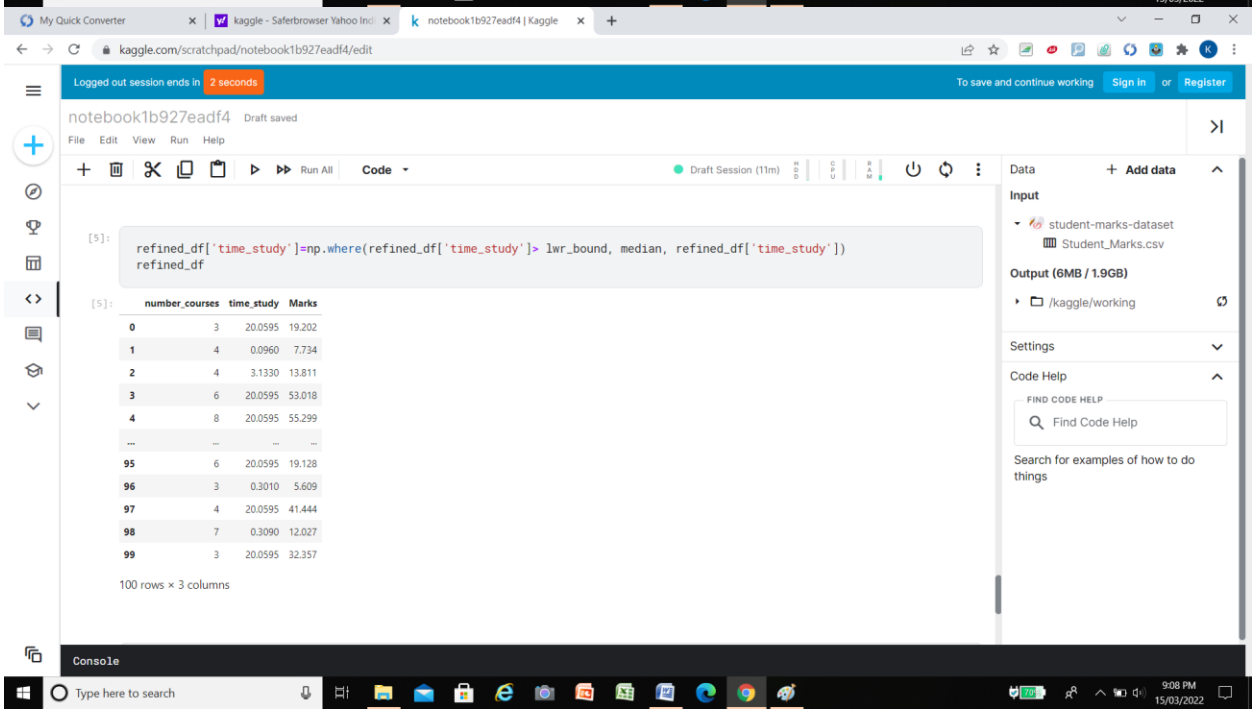
Search for examples of how to do things

Console

Type here to search

9:08 PM 15/03/2022


My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (11m)
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[4]:
refined_df=df
refined_df['time_study']=np.where(refined_df['time_study']> upr_bound, median, refined_df['time_study'])
refined_df
[4]:
number_courses time_study Marks
0 3 4.508 19.202
1 4 0.096 7.734
2 4 3.133 13.811
3 6 7.909 53.018
4 8 7.811 55.299
... ... ... ...
95 6 3.561 19.128
96 3 0.301 5.609
97 4 7.163 41.444
98 7 0.309 12.027
99 3 6.335 32.357
100 rows × 3 columns
Console
Type here to search
9:08 PM
15/03/2022


My Quick Converter
kaggle - Saferbrowser Yahoo Indi
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All Code Draft Session (11m)
Data Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[5]:
refined_df['time_study']=np.where(refined_df['time_study']> lwr_bound, median, refined_df['time_study'])
refined_df
[5]:
number_courses time_study Marks
0 3 20.0595 19.202
1 4 0.0960 7.734
2 4 3.1330 13.811
3 6 20.0595 53.018
4 8 20.0595 55.299
... ... ... ...
95 6 20.0595 19.128
96 3 0.3010 5.609
97 4 20.0595 41.444
98 7 0.3090 12.027
99 3 20.0595 32.357
100 rows × 3 columns
Console
Type here to search
9:08 PM
15/03/2022

New Tab | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

notebook1b927eadf4 Draft saved

File Edit View Run Help

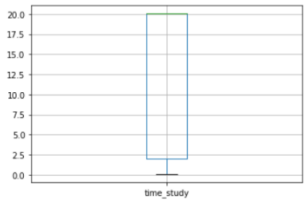
Run All | Code | Draft Session (11m)

| | | | |
|---|---|---|---|
| 99 | 3 | 20.0595 | 32.357 |

100 rows × 3 columns

[6]:
```
col=['time_study']
refined_df.boxplot(col)
```

[6]: <AxesSubplot:>




[8]:

Console

Data | + Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Type here to search — 9:08 PM 15/03/2022

New Tab | kaggle - Saferbrowser Yahoo Ind | notebook1b927eadf4 | Kaggle

kaggle.com/scratchpad/notebook1b927eadf4/edit

Logged out session ends in 2 seconds — To save and continue working Sign in or Register

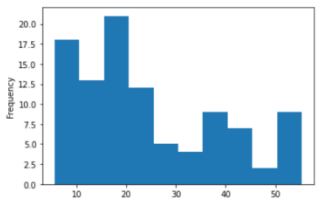
notebook1b927eadf4 Draft saved

File Edit View Run Help

Run All | Code | Draft Session (11m)

[8]:
```
import matplotlib.pyplot as plt
new_df=df
new_df['Marks'].plot(kind='hist')
```

[8]: <AxesSubplot:ylabel='Frequency'>




[10]:
```
df['log_math']=np.log10(df['Marks'])
df['log_math'].plot(kind='hist')
```

Console

Data | + Add data

Input

student-marks-dataset

Student_Marks.csv

Output (6MB / 1.9GB)

/kaggle/working

Settings

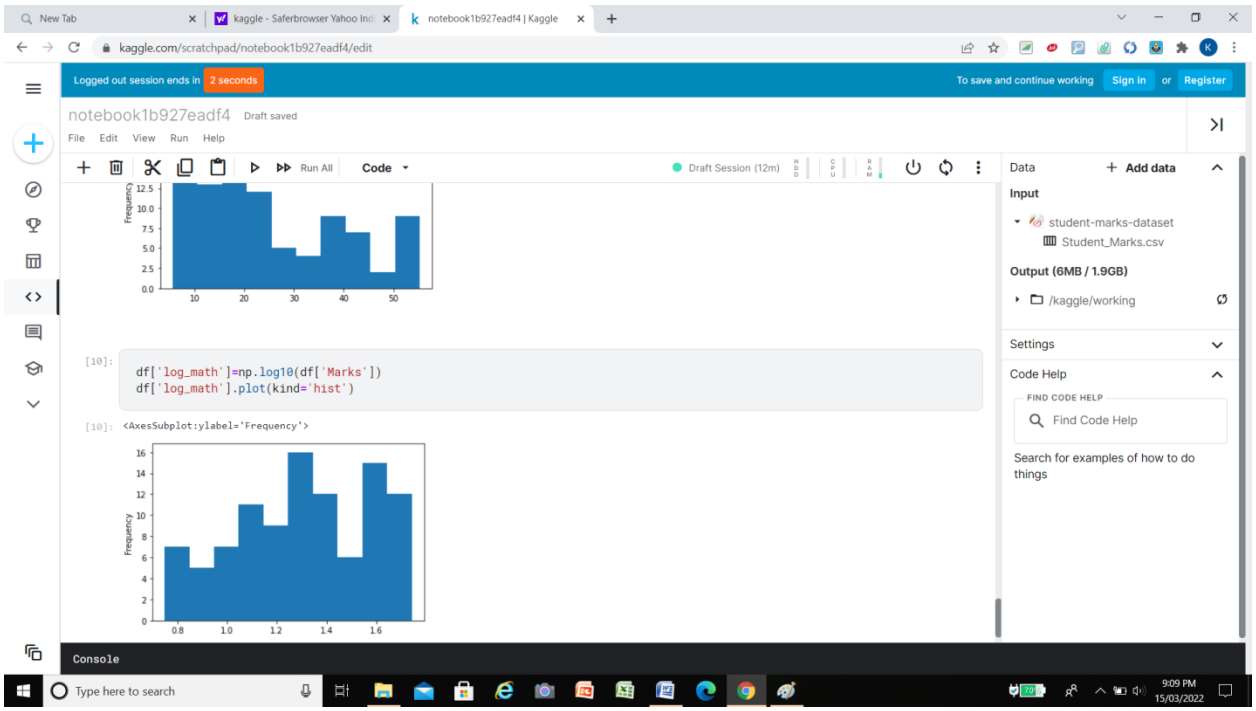
Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

Type here to search — 9:09 PM 15/03/2022


New Tab
kaggle - Saferbrowser Yahoo Ind
notebook1b927eadf4 | Kaggle
kaggle.com/scratchpad/notebook1b927eadf4/edit
Logged out session ends in 2 seconds
To save and continue working Sign in or Register
notebook1b927eadf4 Draft saved
File Edit View Run Help
Run All
Code
Draft Session (12m)
Frequency
12.5
10.0
7.5
5.0
2.5
0.0
10
20
30
40
50
[10]:
df['log_math']=np.log10(df['Marks'])
df['log_math'].plot(kind='hist')
[10]: <AxesSubplot:ylabel='Frequency'>
Frequency
16
14
12
10
8
6
4
2
0
0.8
1.0
1.2
1.4
1.6
Console
Data
Add data
Input
student-marks-dataset
Student_Marks.csv
Output (6MB / 1.9GB)
/kaggle/working
Settings
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
9:09 PM
15/03/2022