

GoToWebinar | Event Details x Home Page - Select or create a notebook x EDA Using Python - Jupyter Notebook x

localhost:8888/notebooks/EDA%20Using%20Python.ipynb

jupyter EDA Using Python Last Checkpoint: a day ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

in [1]: `import pandas as pd
import numpy as np
import seaborn as sns`

In [3]: `data = pd.read_csv("C:/Users/Waseem/Desktop/datasets/student.csv")`

In [5]: `#1 understanding the data`

In [6]: `data.head()`

Out[6]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

In []:

```
In [7]: data.tail()
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

GoToWebinar | Event Details x Home Page - Select or create a x EDA Using Python - Jupyter Note x +

localhost:8888/notebooks/EDA%20Using%20Python.ipynb

jupyter EDA Using Python Last Checkpoint: a day ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [10]: data.describe()

Out[10]:

	math score	reading score	writing score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

In [9]: data.shape

Out[9]: (1000, 8)

GoToWebinar | Event Details x Home Page - Select or create a notebook x EDA Using Python - Jupyter Notebook x +

localhost:8888/notebooks/EDA%20Using%20Python.ipynb

jupyter EDA Using Python Last Checkpoint: a day ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

max 100.00000 100.000000 100.000000

```
In [9]: data.shape
```

```
Out[9]: (1000, 8)
```

```
In [12]: data.columns
```

```
Out[12]: Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',  
              'test preparation course', 'math score', 'reading score',  
              'writing score'],  
              dtype='object')
```

```
In [ ]:
```

GoToWebinar | Event Details x Home Page - Select or create a notebook x EDA Using Python - Jupyter Notebook x +

localhost:8888/notebooks/EDA%20Using%20Python.ipynb

jupyter EDA Using Python Last Checkpoint: a day ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [16]: `data.nunique()`

Out[16]:

gender	2
race/ethnicity	5
parental level of education	6
lunch	2
test preparation course	2
math score	81
reading score	72
writing score	77
dtype:	int64

In [18]: `data['race/ethnicity'].unique()`

Out[18]: `array(['group B', 'group C', 'group A', 'group D', 'group E'], dtype=object)`

In []:

GoToWebinar | Event Details x Home Page - Select or create a x EDA Using Python - Jupyter Note x +

localhost:8888/notebooks/EDA%20Using%20Python.ipynb

jupyter EDA Using Python Last Checkpoint: a day ago (unsaved changes)

Python 3

Logout

File Edit View Insert Cell Kernel Widgets Help

Run

Code

In [20]: `#cleaning the data`

In [21]: `data.isnull().sum()`

Out[21]:

gender	0
race/ethnicity	0
parental level of education	0
lunch	0
test preparation course	0
math score	0
reading score	0
writing score	0
dtype: int64	

In []: |

```
reading score
writing score
dtype: int64
```

In [27]: `student = data.drop(['race/ethnicity', 'parental level of education'], axis=1)`

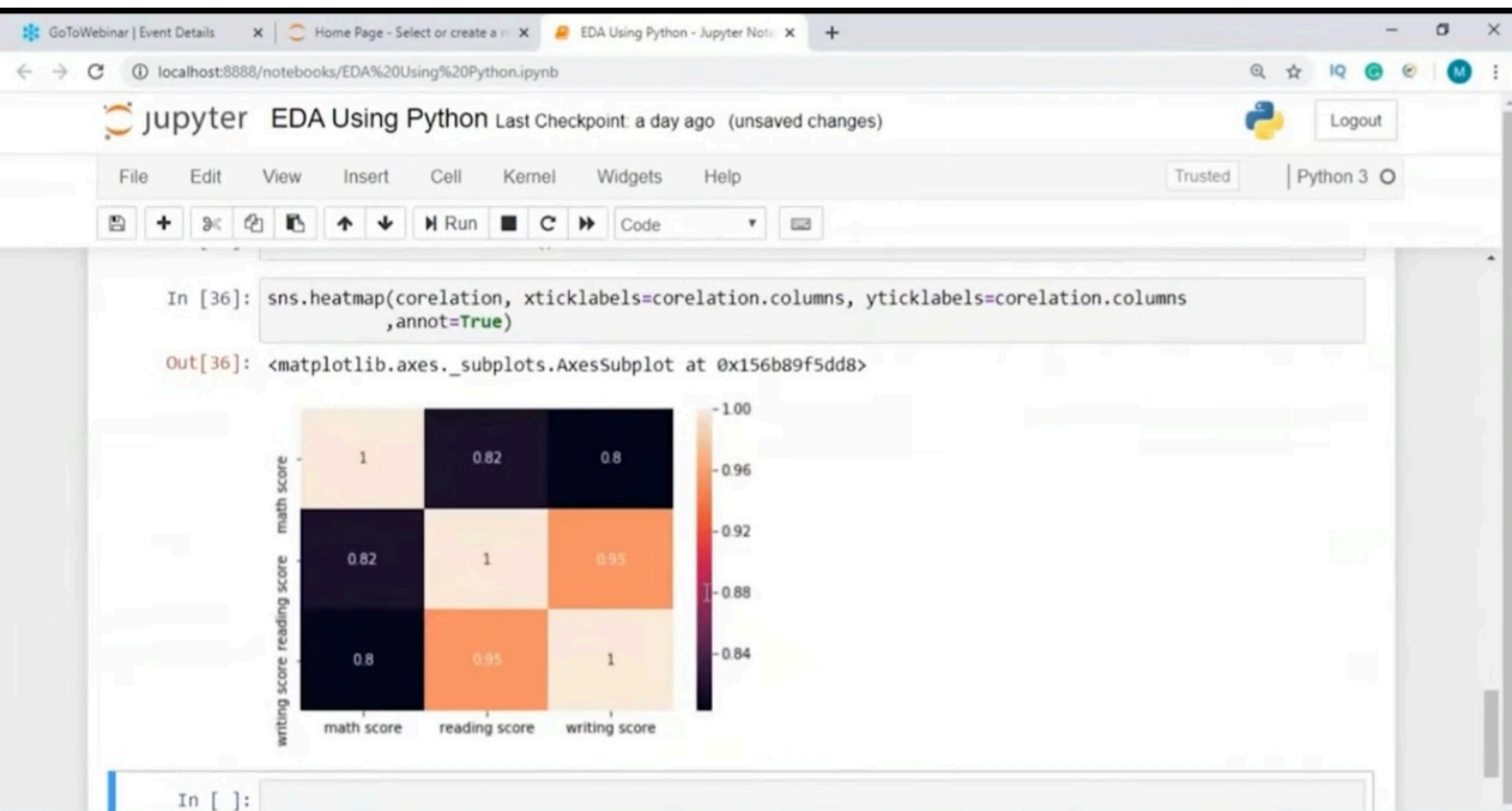
In [28]: `student.head()`

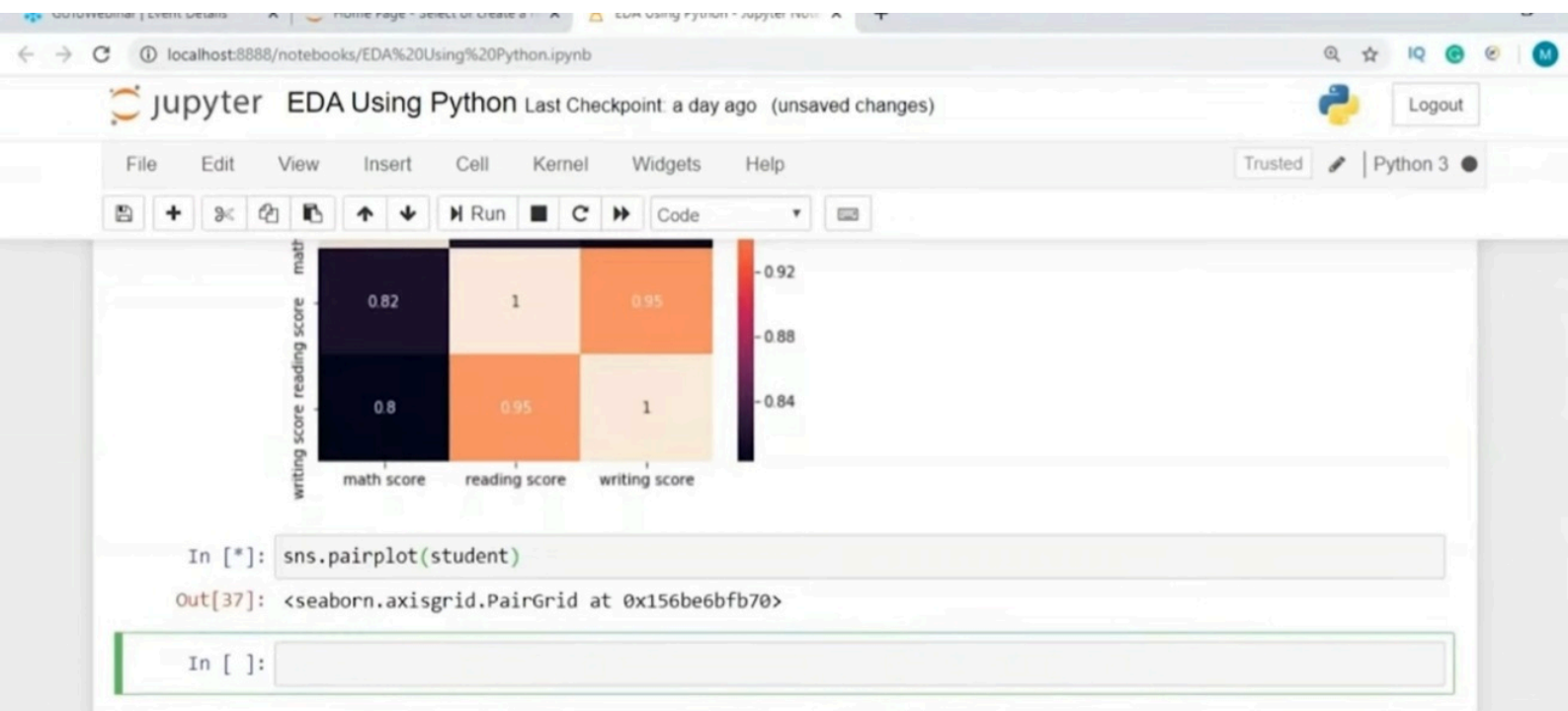
Out[28]:

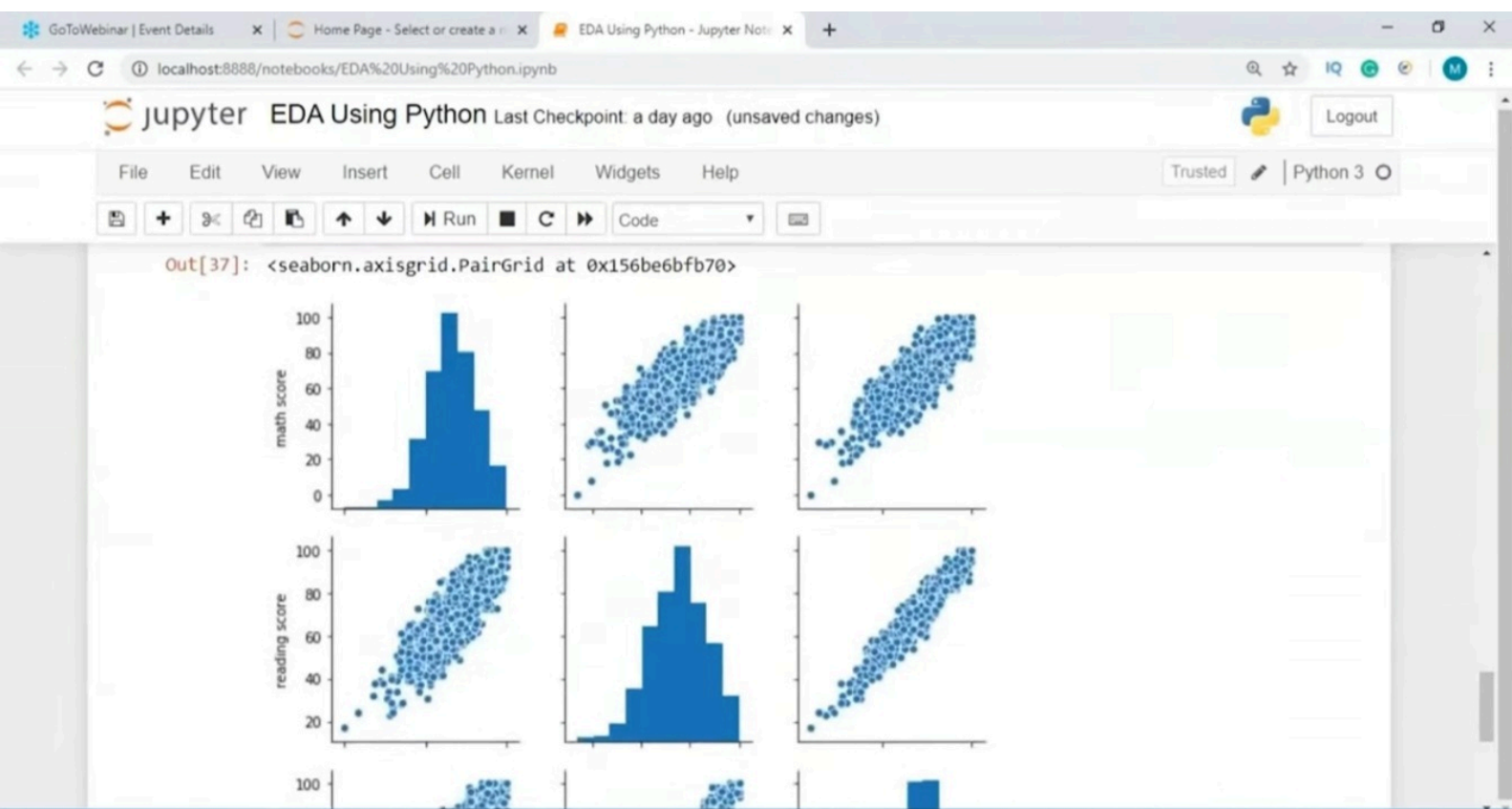
	gender	lunch	test preparation course	math score	reading score	writing score
0	female	standard	none	72	72	74
1	female	standard	completed	69	90	88
2	female	standard	none	90	95	93
3	male	free/reduced	none	47	57	44
4	male	standard	none	76	78	75

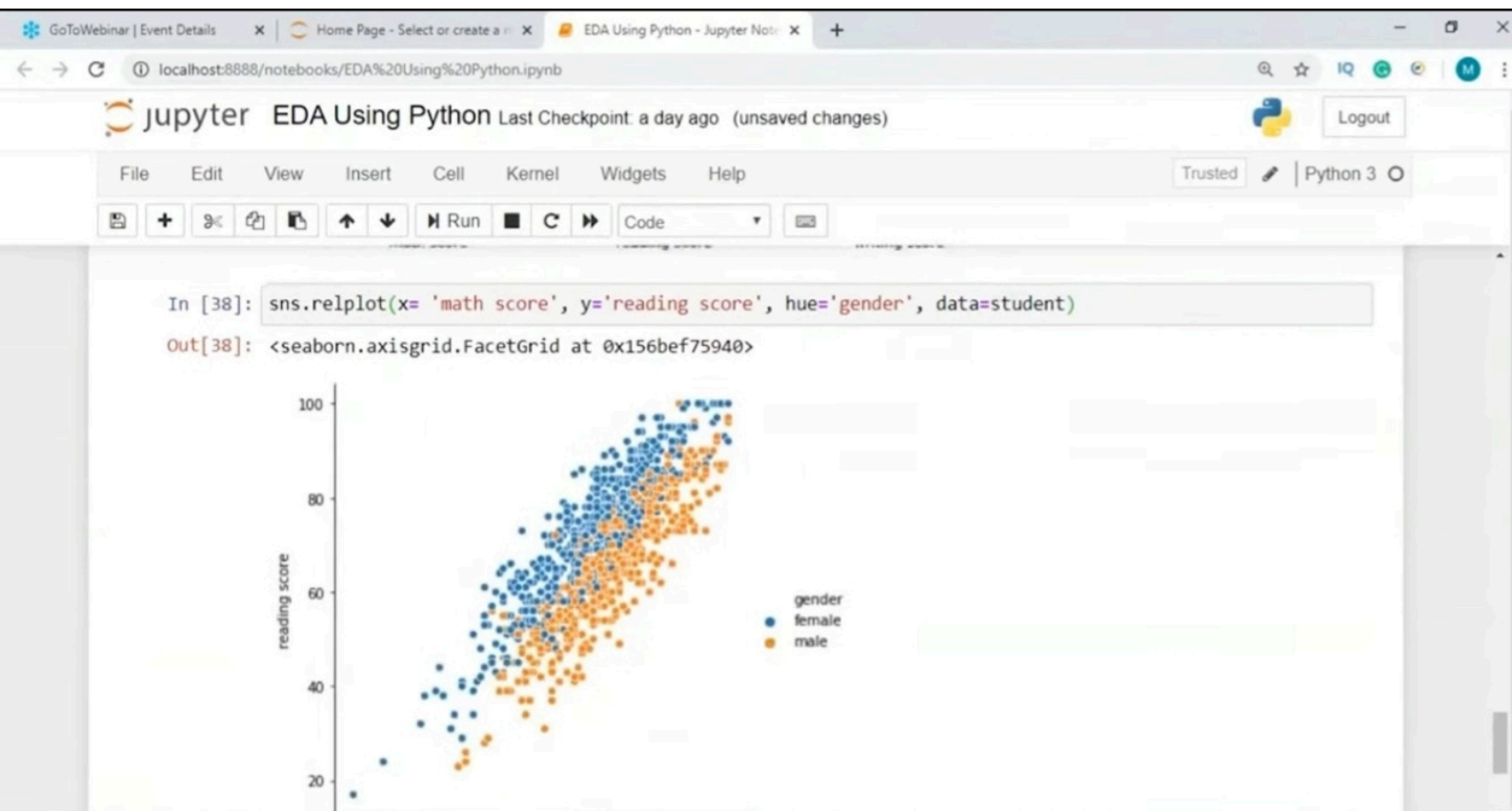
In []:

```
In [32]: correlation = student.corr()
```







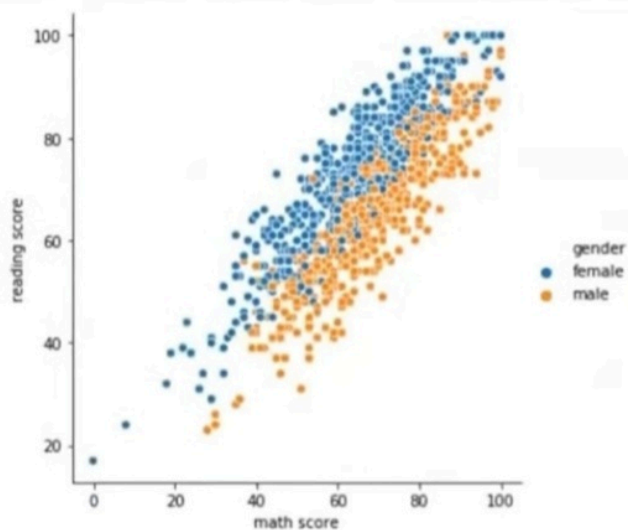
jupyter EDA Using Python Last Checkpoint: a day ago (unsaved changes)

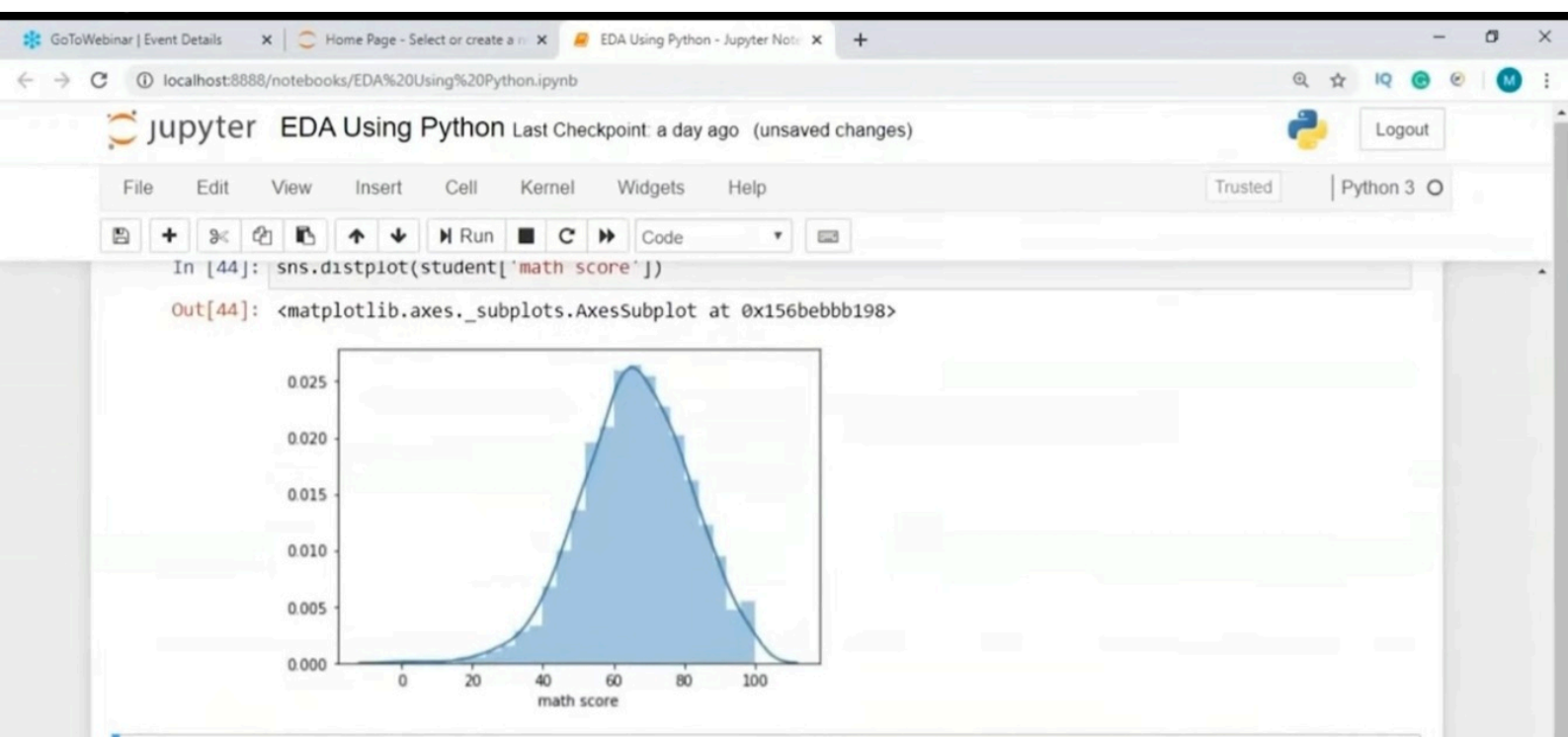
Logout

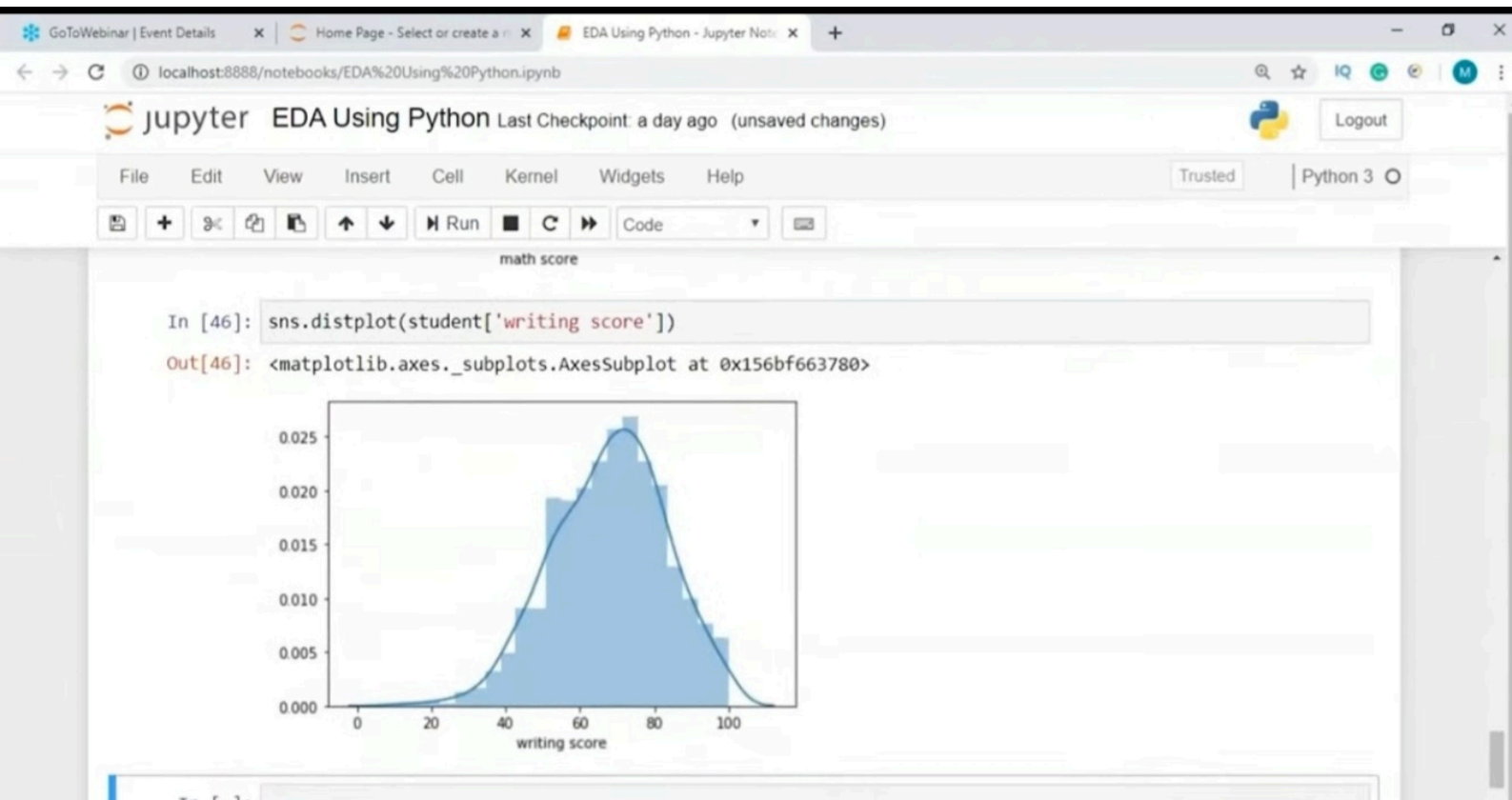
File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Run Code







```
In [48]: sns.distplot(student['writing score'], bins=5)
```

```
Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x156c0750d30>
```

