Lab1: Warm Up – Familiarity with Data types and Visualization

```
Step 1: Download the dataset files that belong to the following data formats from internet. The files
may belong to any dataset available online.
Step 2: Read these files inside the python code. Some of the file formats cannot be read using default
python packages. In this case, explore the python packages suitable for reading the files.
Step 3: Print the properties of the data files such as size, shape, dimensions, etc.
```

In [2]:
```
import csv
import pandas as pd
```

In [3]:
```
cars=pd.read_csv("cars.csv",sep=";")
```

In [4]:
```
cars.head()
```

Out[4]:

	Car	MPG	Cylinders	Displacement	Horsepower	Weight	Acceleration	Model	Origin
0	STRING	DOUBLE	INT	DOUBLE	DOUBLE	DOUBLE	DOUBLE	INT	CAT
1	Chevrolet Chevelle Malibu	18.0	8	307.0	130.0	3504.	12.0	70	US
2	Buick Skylark 320	15.0	8	350.0	165.0	3693.	11.5	70	US
3	Plymouth Satellite	18.0	8	318.0	150.0	3436.	11.0	70	US
4	AMC Rebel SST	16.0	8	304.0	150.0	3433.	12.0	70	US

In [5]:
```
cars.shape
```

Out[5]:
```
(407, 9)
```

In [6]:
```
cars.size
```

Out[6]:
```
3663
```

In [7]:
```
cars.ndim
```

Out[7]: 2

In [8]:
```
cars.shape[0]
```

Out[8]: 407

In [9]:
```
cars.shape[1]
```

Out[9]: 9

In [10]:
```
type(cars)
```

Out[10]: pandas.core.frame.DataFrame

In [13]:
```
import json
import pandas as pd
```

In [14]:
```
sample=pd.read_json('employee.json')
sample.head()
```

Out[14]:

	Employees
0	{'userId': 'krish', 'jobTitle': 'Developer', '...
1	{'userId': 'devid', 'jobTitle': 'Developer', '...
2	{'userId': 'tin', 'jobTitle': 'Program Directo...

In [15]:
```
sample.shape
```

Out[15]: (3, 1)

In [16]:
```
sample.size
```

Out[16]: 3

In [17]:
```
sample.ndim
```

Out[17]: 2

In [18]:
```
import csv
import pandas as pd
```

In [19]:
```
data=pd.read_csv("forestfires.csv")
```

In [20]:
```
data.head()
```

Out[20]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

In [47]:
```
data.tail()
```

Out[47]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

In [21]:
```
data.shape
```

Out[21]: (517, 13)

In [22]:
```
data.size
```

Out[22]: 6721

In [23]:
```
data.ndim
```

Out[23]: 2

In [24]:
```
data.shape[0]
```

Out[24]: 517

In [25]:
```
data.shape[1]
```

Out[25]: 13

In [26]:
```
type(data)
```

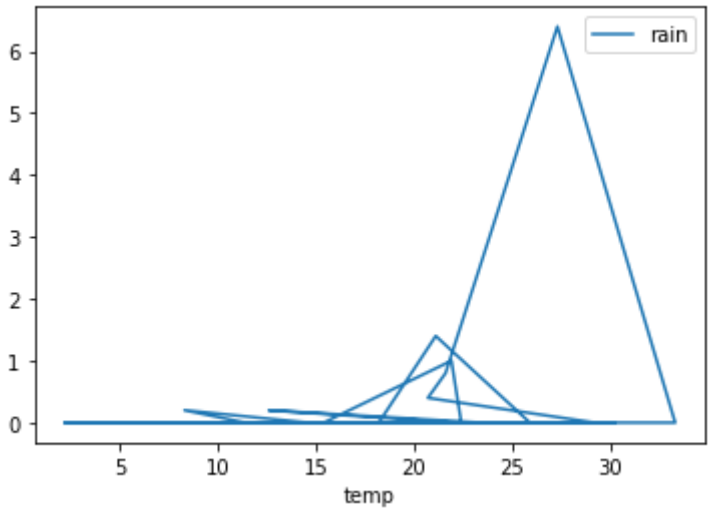
Out[26]: pandas.core.frame.DataFrame

Step 4: Visualize each of these data files using graphs, diagrams, etc.

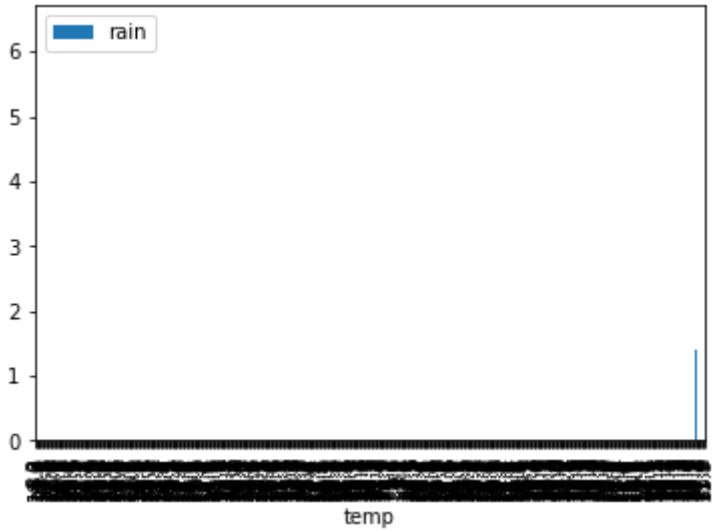
Table data visualization: line graph, bar graph, histogram chart, pie chart, scatter plot

In [27]:
```
import matplotlib.pyplot as plt
```

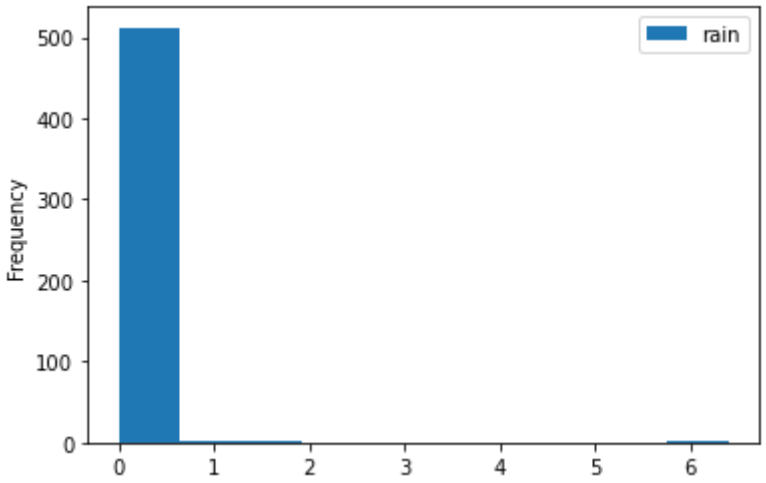
In [28]:
```
data.plot(kind="line",x="temp",y="rain")
plt.show()
```



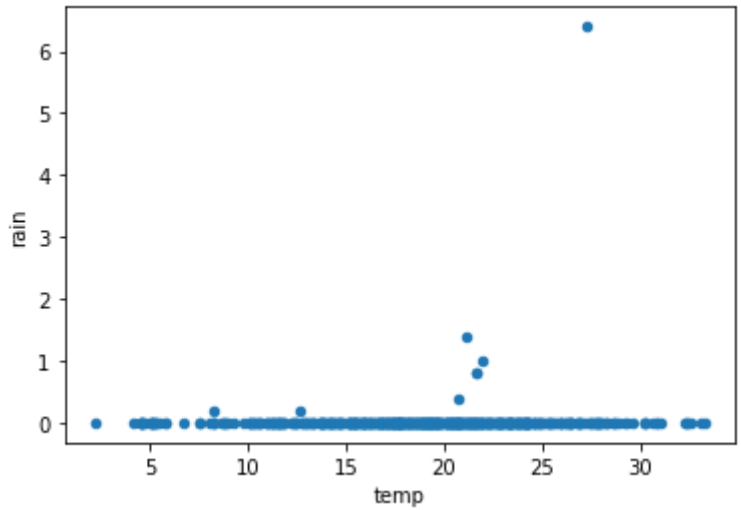
In [29]:
```
data.plot(kind="bar",x="temp",y="rain")
plt.show()
```



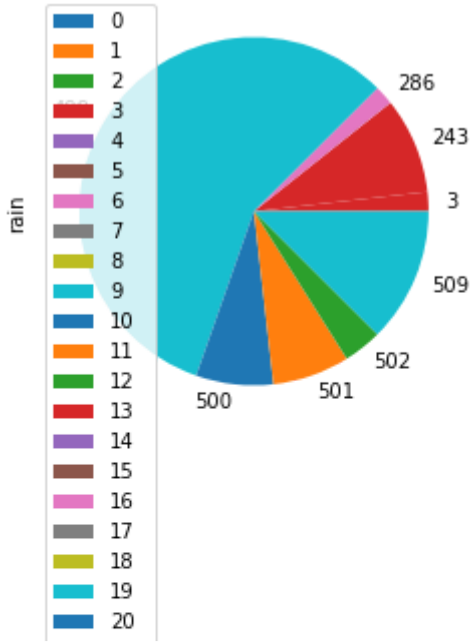
In [30]:
```
data.plot(kind="hist",x="temp",y="rain")
plt.show()
```



In [31]:
```
data.plot(kind="scatter",x="temp",y="rain")
plt.show()
```



In [38]:
```
data.plot(kind="pie",x="temp",y="rain")
plt.show()
```



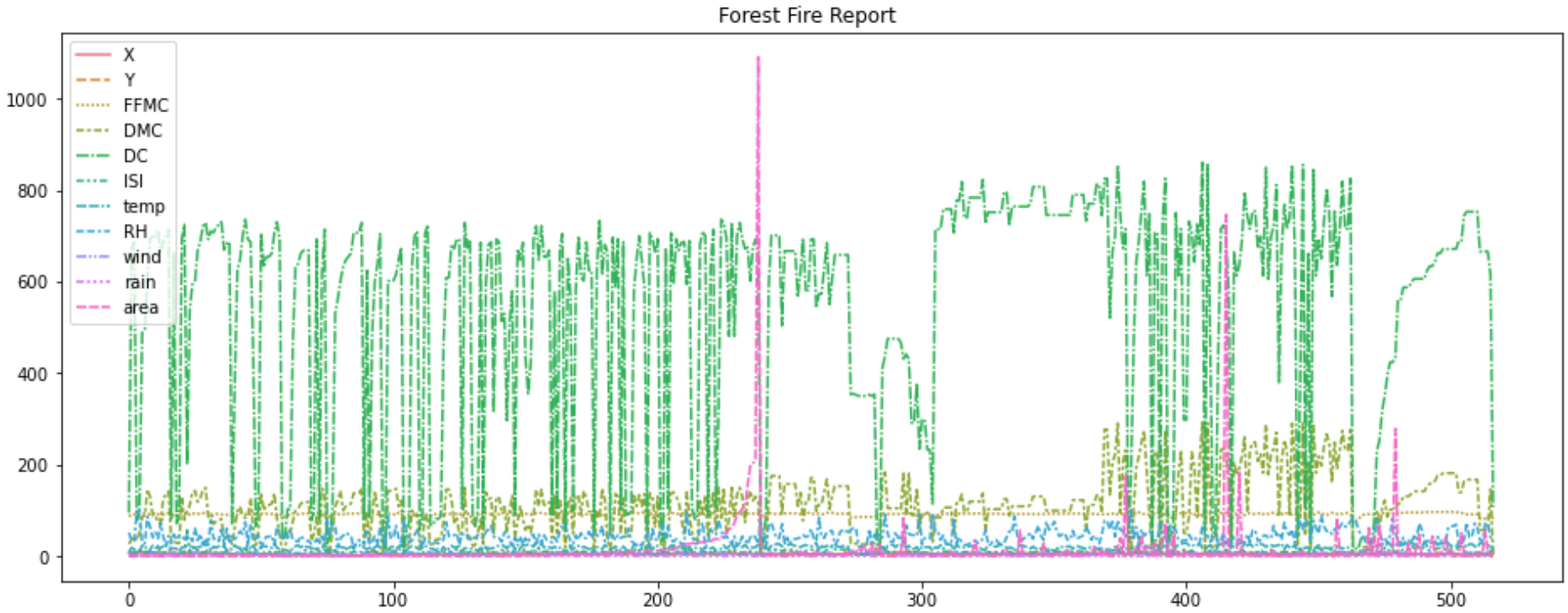
In [43]:
```
!pip3 install seaborn
```

```
Requirement already satisfied: seaborn in c:\programdata\anaconda3\lib\site-packages (0.11.0)
Requirement already satisfied: matplotlib>=2.2 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (3.3.2)
Requirement already satisfied: numpy>=1.15 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (1.19.2)
Requirement already satisfied: pandas>=0.23 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (1.1.3)
Requirement already satisfied: scipy>=1.0 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (1.5.2)
Requirement already satisfied: certifi>=2020.06.20 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.2-
>seaborn) (2020.6.20)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.2->s
eaborn) (1.3.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.2
->seaborn) (2.8.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in c:\programdata\anaconda3\lib\site-packages
(from matplotlib>=2.2->seaborn) (2.4.7)
Requirement already satisfied: pillow>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.2->seabo
rn) (8.0.1)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib>=2.2->seabor
n) (0.10.0)
Requirement already satisfied: pytz>=2017.2 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.23->seaborn)
(2020.1)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.1->matpl
otlib>=2.2->seaborn) (1.15.0)
```

In [44]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

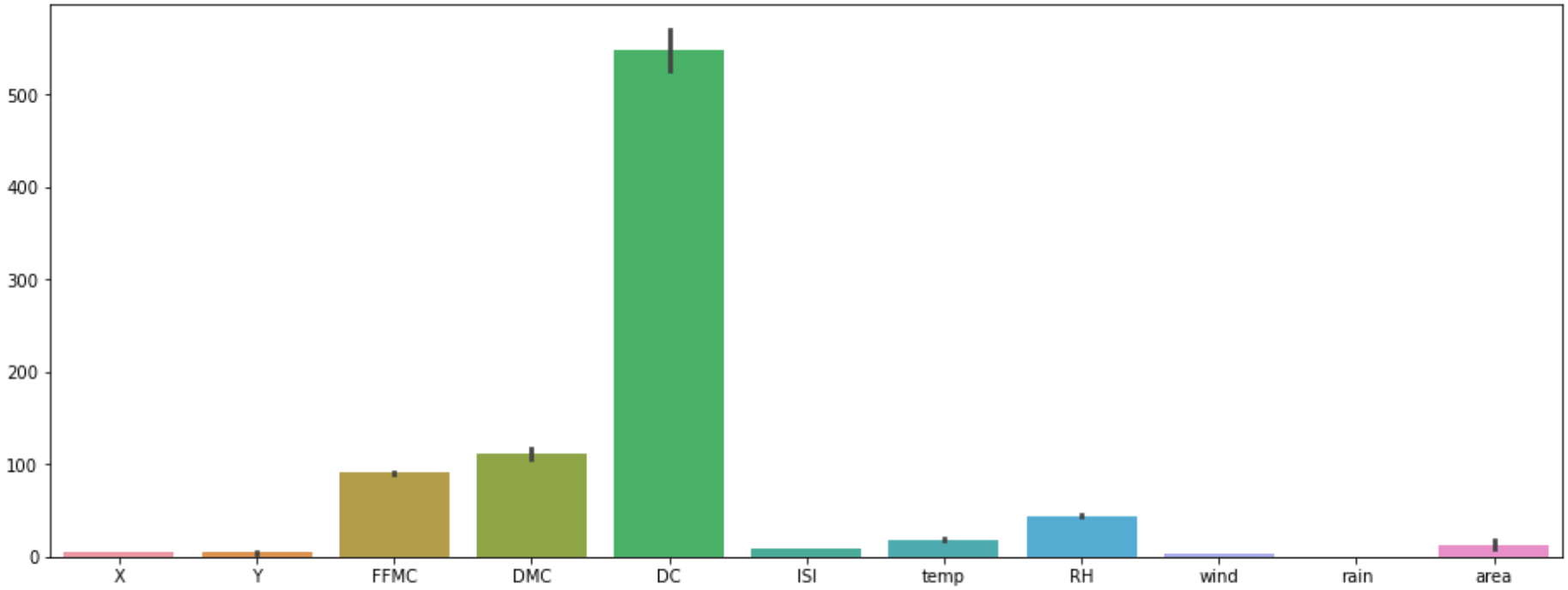
In [49]:
```
plt.figure(figsize=(16,6))
plt.title("Forest Fire Report ")
sns.lineplot(data=data)
```

Out[49]:
```
<AxesSubplot:title={'center':'Forest Fire Report '}>
```



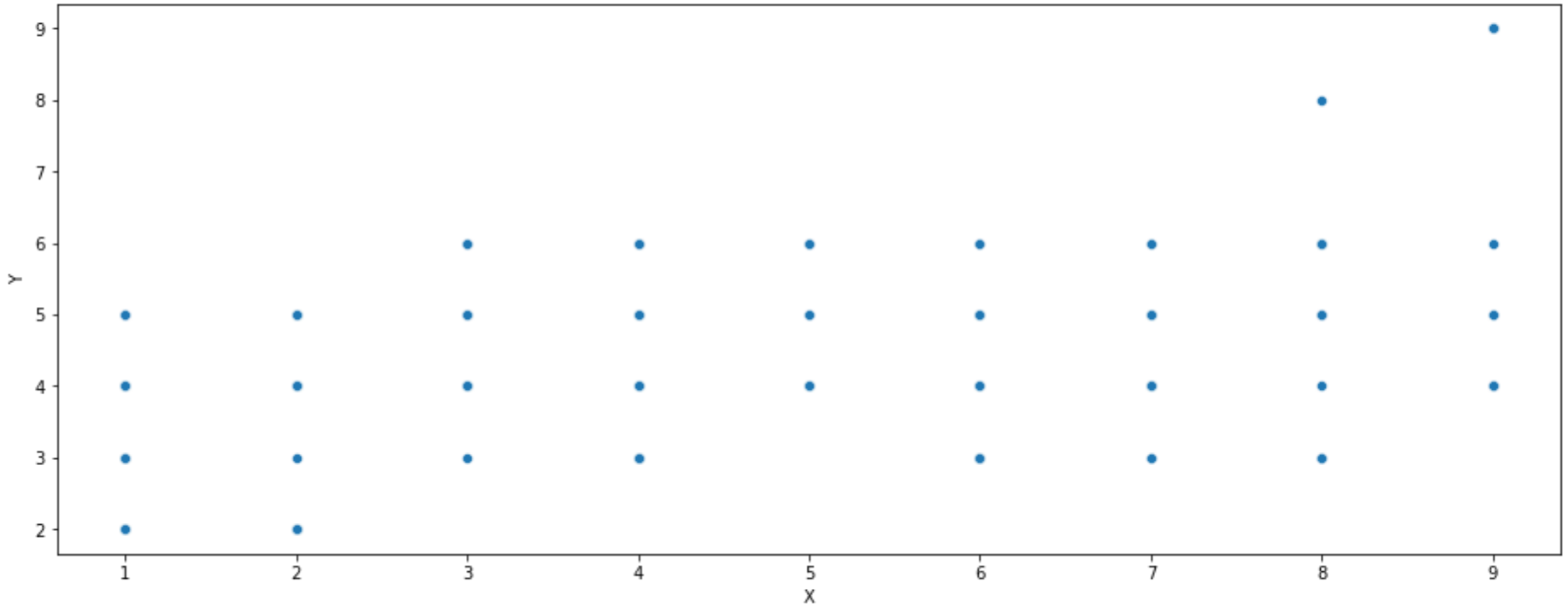
In [50]:
```
plt.figure(figsize=(16,6))
sns.barplot(data=data)
```

Out[50]: <AxesSubplot:>



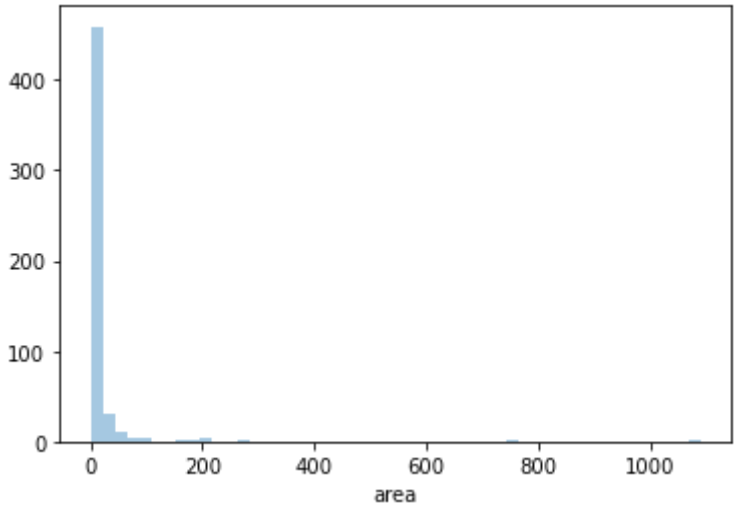
In [51]:
```
plt.figure(figsize=(16,6))
sns.scatterplot(x=data['X'] , y=data['Y'])
```

Out[51]: <AxesSubplot:xlabel='X', ylabel='Y'>

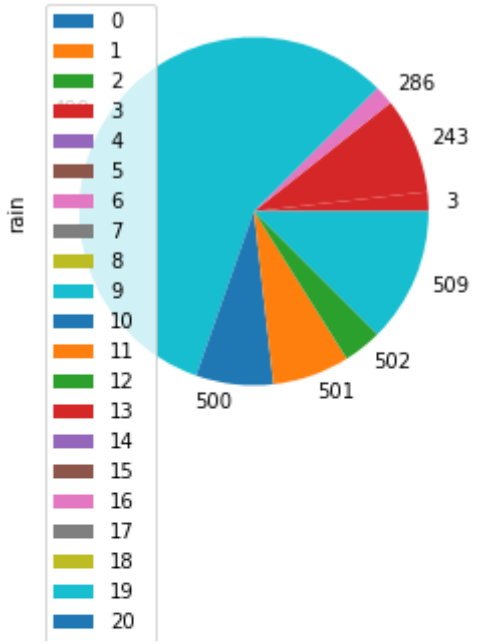


In [54]:
```
sns.distplot(a=data['area'],kde=False)
```

Out[54]: <AxesSubplot:xlabel='area'>



In [72]:
```
data.plot(kind="pie",x="area",y="rain")
plt.show()
```



In [73]:
```
!pip install ijson
```

```
Requirement already satisfied: ijson in c:\programdata\anaconda3\lib\site-packages (3.1.4)
```

In [74]:
```
import ijson
from pandas.io.json import json_normalize
```

In [75]:
```
iri = pd.read_json("iris.json")
iri.head()
```

Out[75]:

	sepalLength	sepalWidth	petalLength	petalWidth	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [77]:
```
iri.tail()
```

Out[77]:

	sepalLength	sepalWidth	petalLength	petalWidth	species
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

In [78]:
```
iri.shape
```

Out[78]:
```
(150, 5)
```

In [79]:
```
iri.size
```

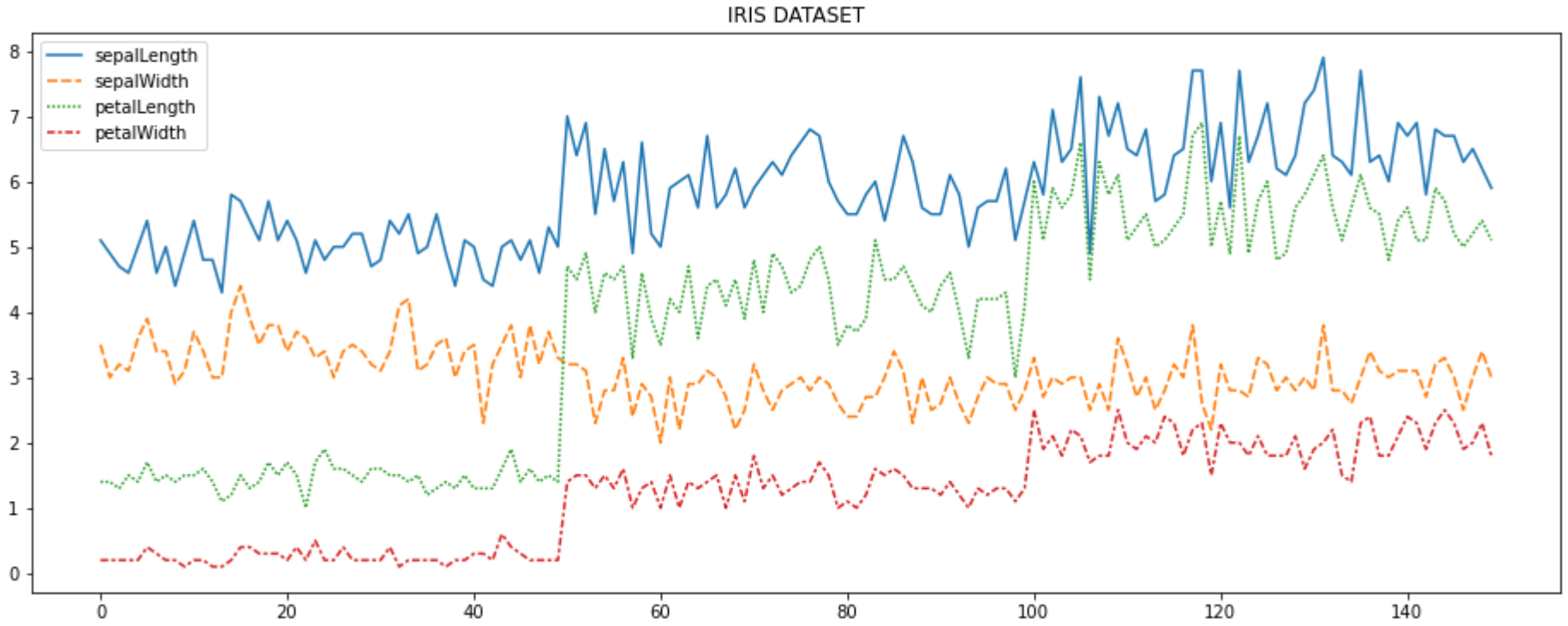
Out[79]:
```
750
```

In [80]:
```
iri.ndim
```

Out[80]:
```
2
```

In [81]:
```
plt.figure(figsize=(16,6))
plt.title("IRIS DATASET")
sns.lineplot(data=iri)
```

Out[81]:
```
<AxesSubplot:title={'center':'IRIS DATASET'}>
```



In [82]:
```
iri.set_index('petalWidth')
```

Out[82]:

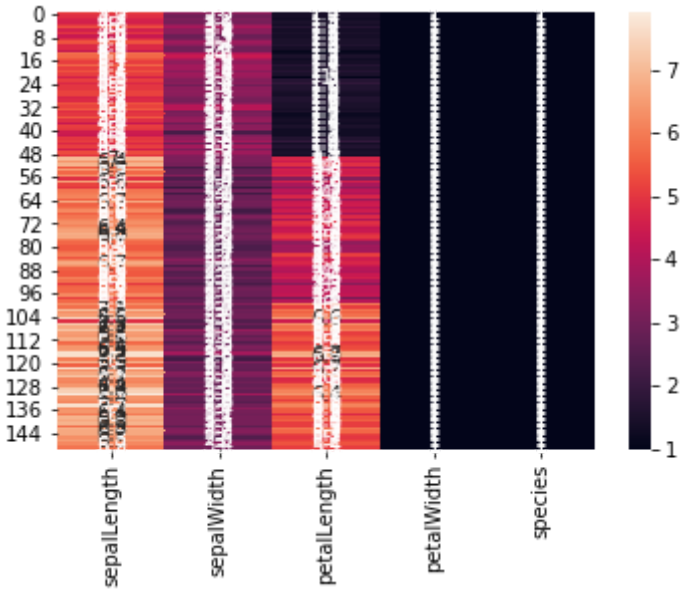
petalWidth	sepalLength	sepalWidth	petalLength	species
0.2	5.1	3.5	1.4	setosa
0.2	4.9	3.0	1.4	setosa
0.2	4.7	3.2	1.3	setosa
0.2	4.6	3.1	1.5	setosa
0.2	5.0	3.6	1.4	setosa
...	...	...	...	...
2.3	6.7	3.0	5.2	virginica
1.9	6.3	2.5	5.0	virginica
2.0	6.5	3.0	5.2	virginica
2.3	6.2	3.4	5.4	virginica
1.8	5.9	3.0	5.1	virginica

150 rows × 4 columns

In [87]:
```
iri['species']=1
```

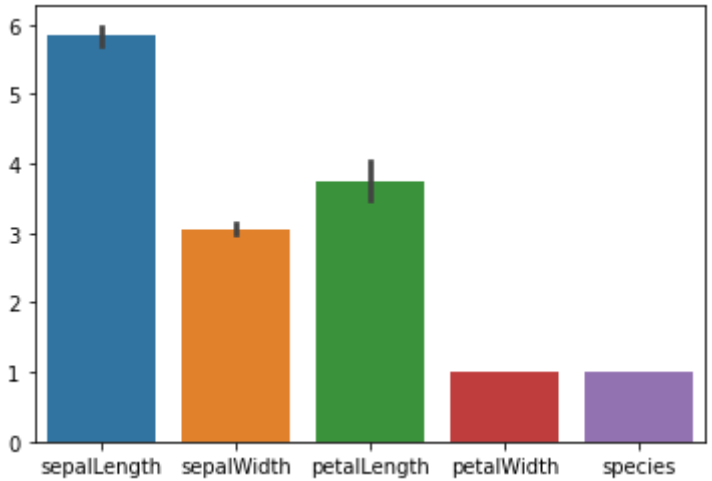
In [88]:
```
sns.heatmap(data=iri,annot=True)
```

Out[88]: <AxesSubplot:>



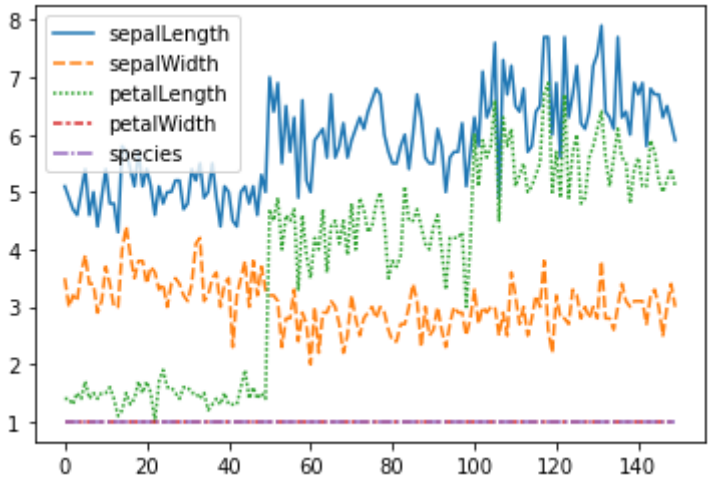
In [89]:
```
sns.barplot(data=iri)
```

Out[89]: <AxesSubplot:>



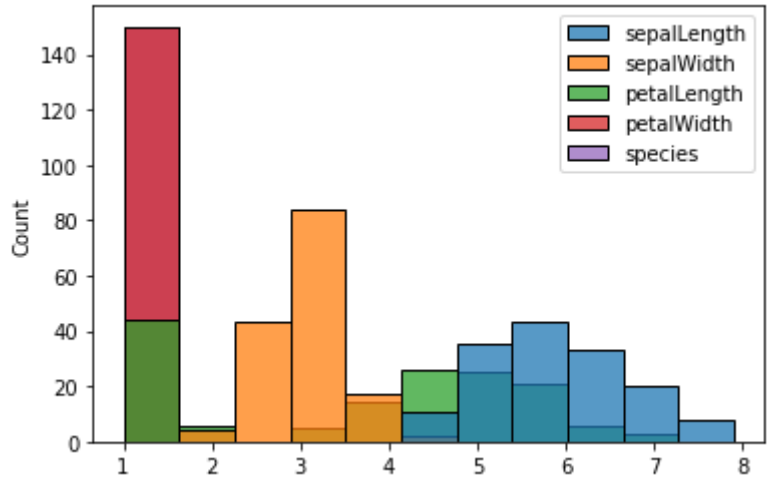
In [91]:
```
sns.lineplot(data=iri)
```

Out[91]: <AxesSubplot:>



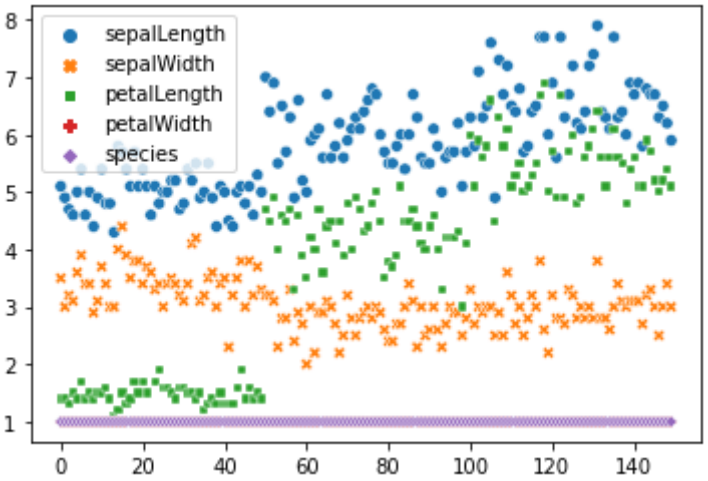
In [92]:
```
sns.histplot(data=iri)
```

Out[92]: <AxesSubplot:ylabel='Count'>

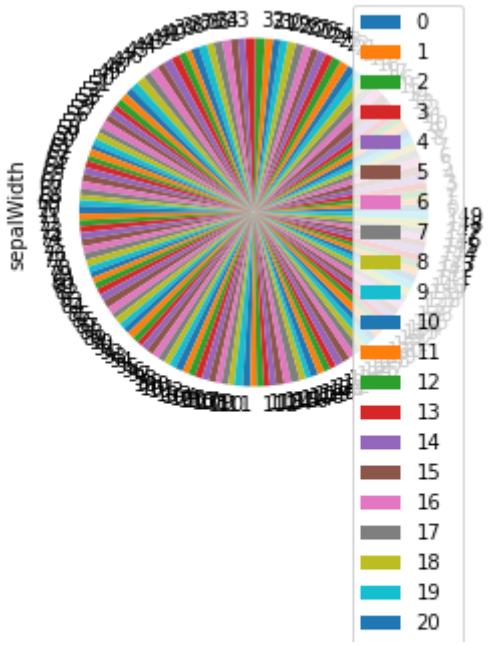


In [93]:
```
sns.scatterplot(data=iri)
```

Out[93]: <AxesSubplot:>



In [97]:
```
iri.plot(kind="pie",x="sepalLength",y="sepalWidth")
plt.show()
```



In [120]:
```
import pandas as pd
dat= pd.read_csv('iris.tsv', sep='\t')
```

In [121]:
```
dat
```

Out[121]:

	id	sepal_length	sepal_width	petal_length	petal_width	species
0	1	5.1	3.5	1.4	0.2	setosa
1	2	4.9	3.0	1.4	0.2	setosa
2	3	4.7	3.2	1.3	0.2	setosa
3	4	4.6	3.1	1.5	0.2	setosa
4	5	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	virginica
146	147	6.3	2.5	5.0	1.9	virginica
147	148	6.5	3.0	5.2	2.0	virginica
148	149	6.2	3.4	5.4	2.3	virginica
149	150	5.9	3.0	5.1	1.8	virginica

150 rows × 6 columns

In [126]:
```
dat.set_index('id')
```

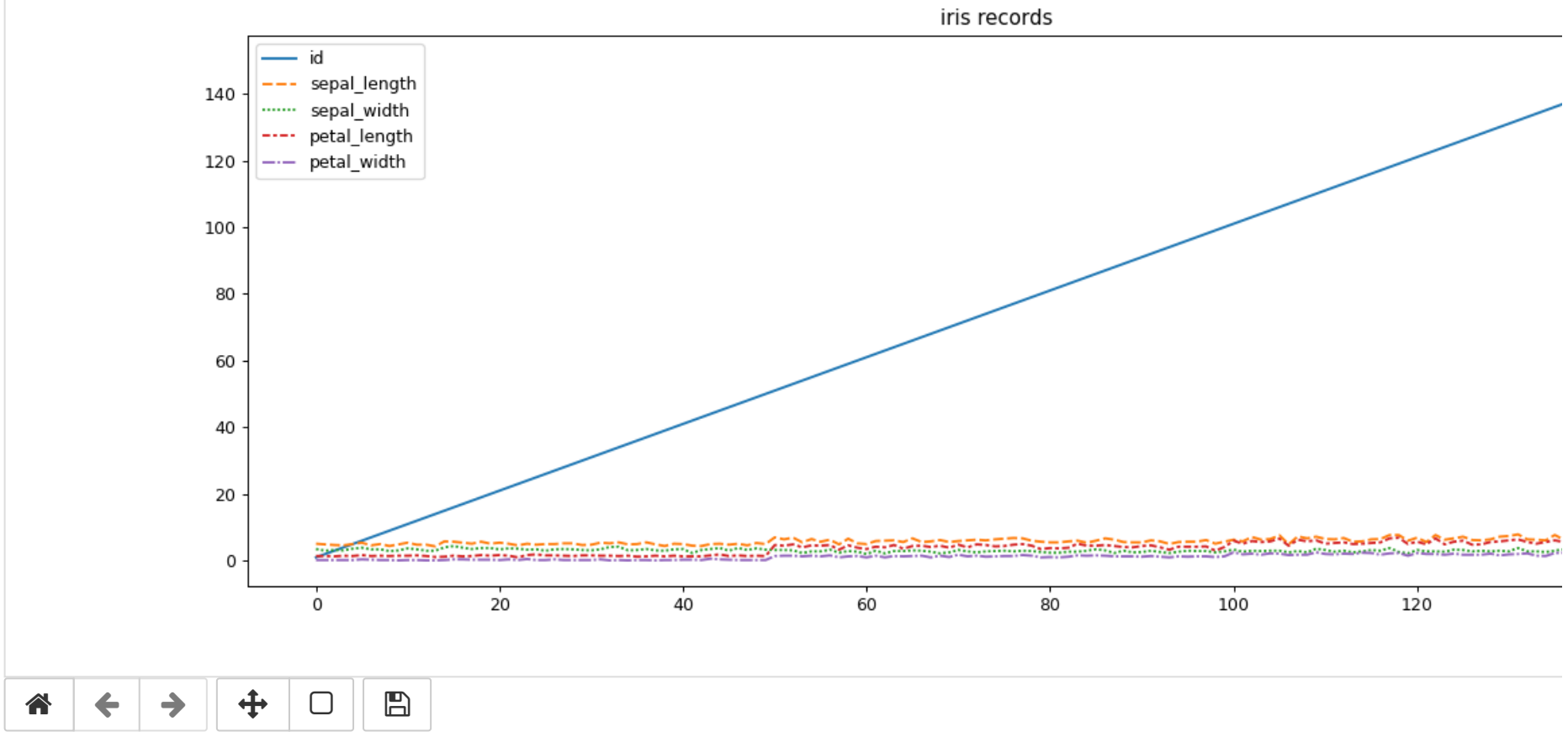
Out[126]:

id	sepal_length	sepal_width	petal_length	petal_width	species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [127]:
```
plt.figure(figsize=(16,6))
plt.title("iris records ")
sns.lineplot(data=dat)
```

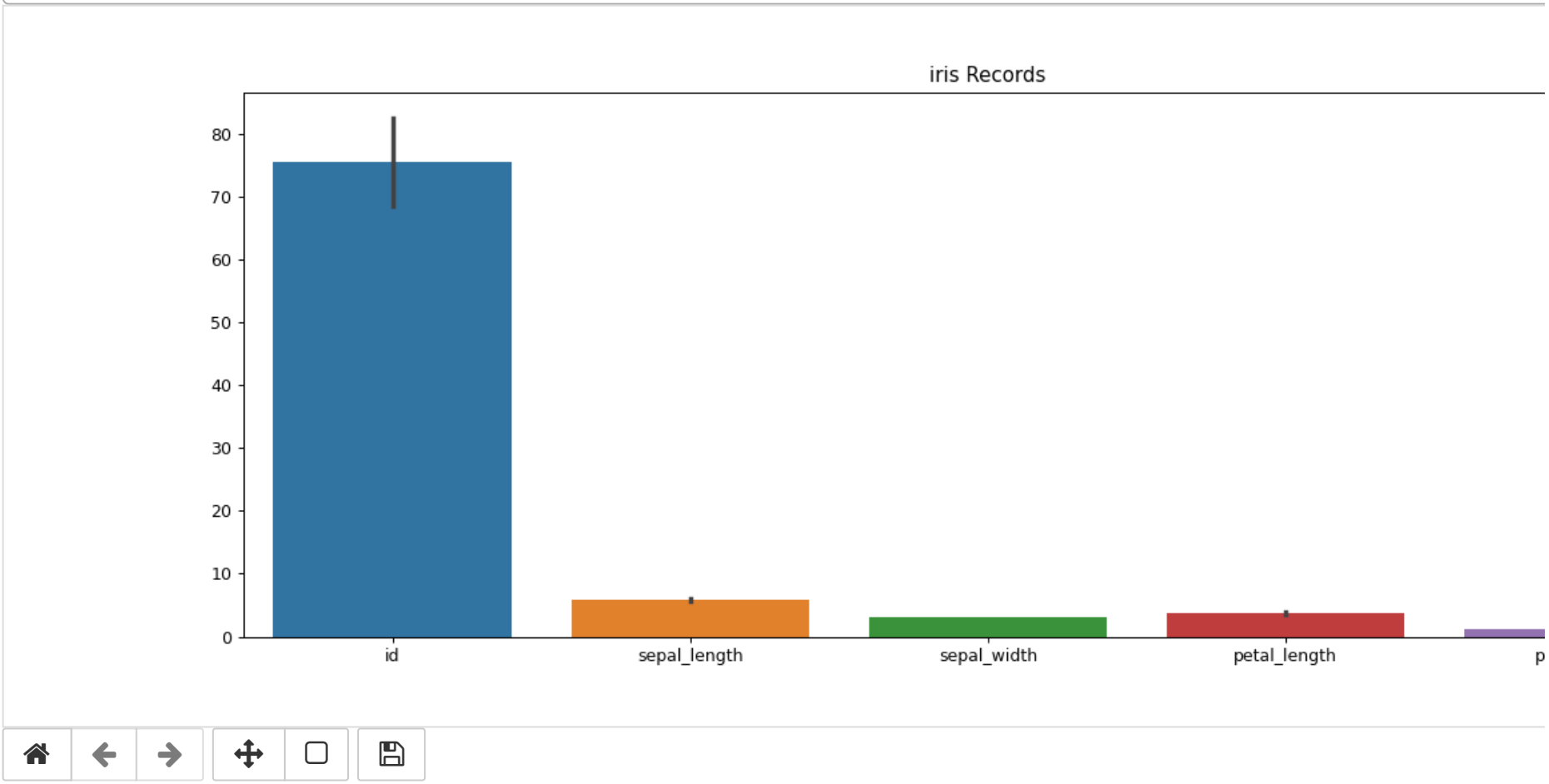
Figure 5



Out[127]: <AxesSubplot:title={'center':'iris records '}>

In [128]:
```
plt.figure(figsize=(16,6))
plt.title("iris Records")
sns.barplot(data=dat)
```

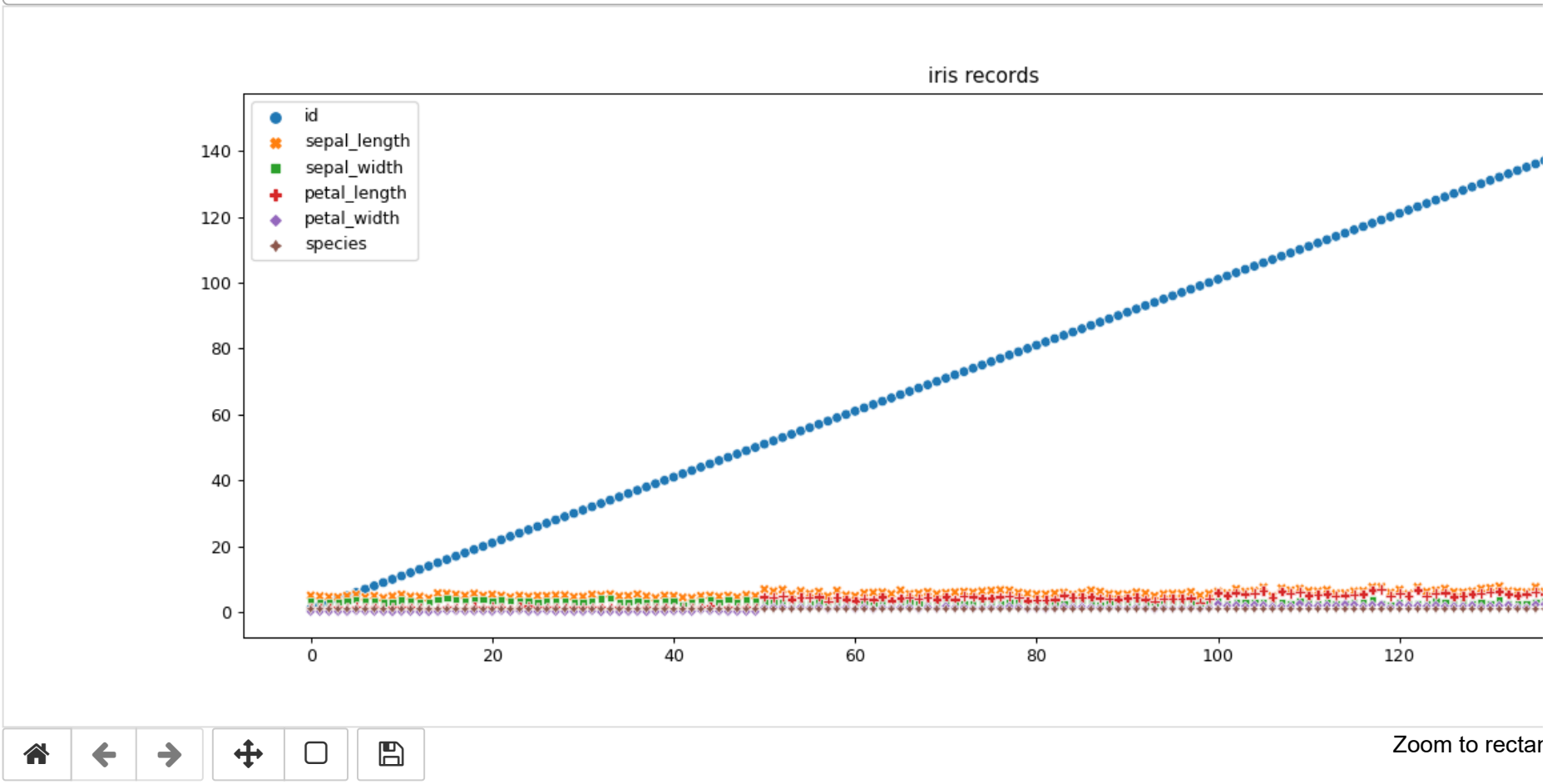
Figure 6



Out[128]: <AxesSubplot:title={'center':'iris Records'}>

In [139]:
```
plt.figure(figsize=(16,6))
plt.title("iris records ")
sns.scatterplot(data=dat)
```

Figure 7



Out[139]:
```
<AxesSubplot:title={'center':'iris records '}>
```

In [140]:
```
from scipy.io import arff
```

In [141]:
```
daf = arff.loadarff('EPL.arff')
df = pd.DataFrame(daf[0])
```

In [142]:
```
df
```

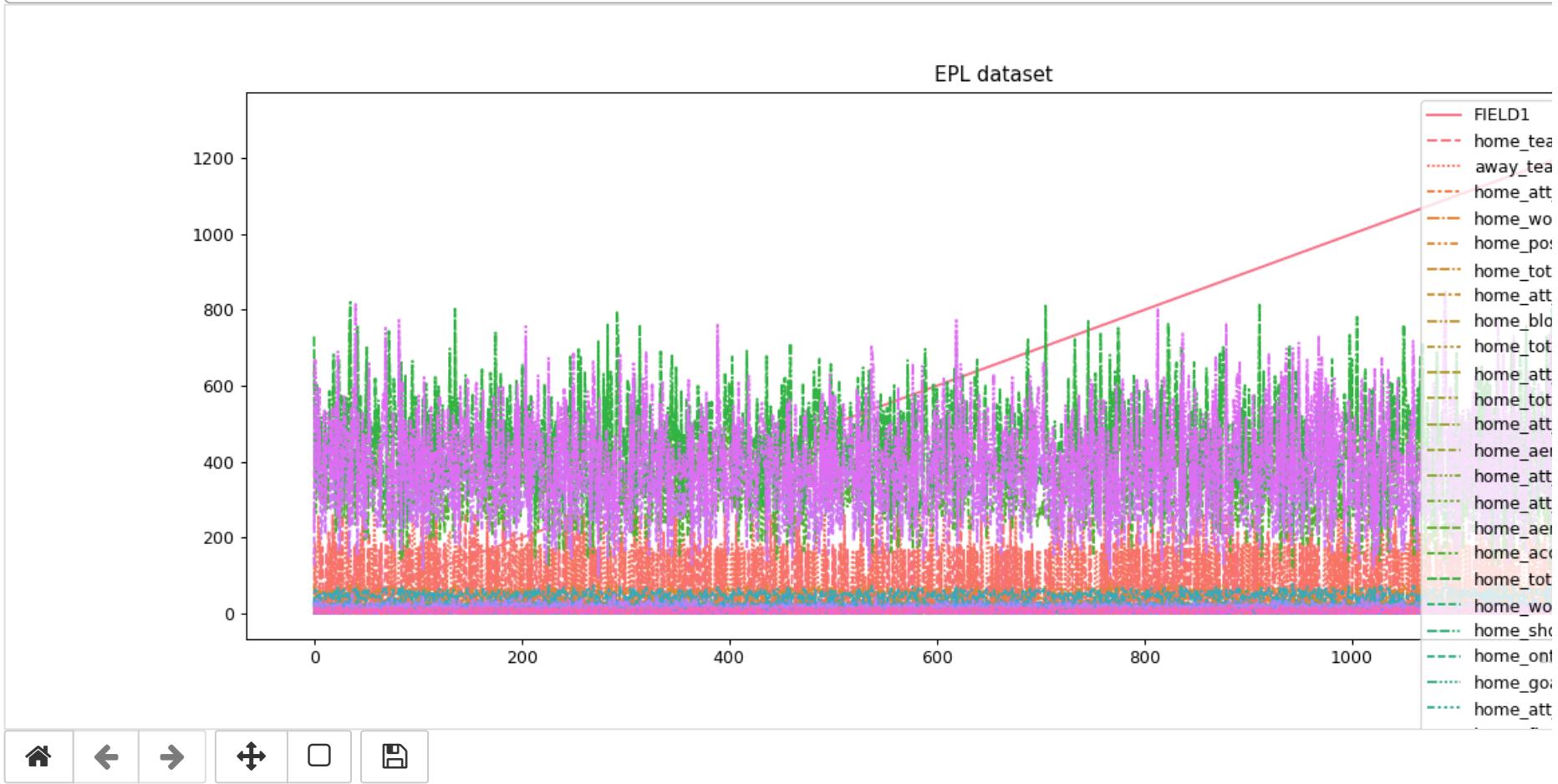
Out[142]:

	FIELD1	season	home_team_id	away_team_id	home_team_name	away_team_name	date_string	half_time_score	full_time_score	full_time_r
0	0.0	b'14-15'	13.0	162.0	b'Arsenal'	b'Crystal Palace'	b'16/08/2014 17:30:00'	b'1 : 1'	b'2 : 1'	
1	1.0	b'14-15'	184.0	15.0	b'Burnley'	b'Chelsea'	b'18/08/2014 20:00:00'	b'1 : 3'	b'1 : 3'	
2	2.0	b'14-15'	14.0	31.0	b'Leicester'	b'Everton'	b'16/08/2014 15:00:00'	b'1 : 2'	b'2 : 2'	
3	3.0	b'14-15'	26.0	18.0	b'Liverpool'	b'Southampton'	b'17/08/2014 13:30:00'	b'1 : 0'	b'2 : 1'	
4	4.0	b'14-15'	32.0	259.0	b'Manchester United'	b'Swansea'	b'16/08/2014 12:45:00'	b'0 : 1'	b'1 : 2'	
...	...	...	...	...	...	...	...	...	...	
1305	1305.0	b'17-18'	30.0	183.0	b'Tottenham'	b'Bournemouth'	b'14/10/2017 15:00:00'	b'0 : 0'	b'1 : 0'	
1306	1306.0	b'17-18'	26.0	175.0	b'Liverpool'	b'West Bromwich Albion'	b'13/12/2017 20:00:00'	b'0 : 0'	b'0 : 0'	
1307	1307.0	b'17-18'	27.0	13.0	b'Watford'	b'Arsenal'	b'14/10/2017 17:30:00'	b'0 : 1'	b'2 : 1'	
1308	1308.0	b'17-18'	30.0	211.0	b'Tottenham'	b'Brighton'	b'13/12/2017 20:00:00'	b'1 : 0'	b'2 : 0'	
1309	1309.0	b'17-18'	15.0	27.0	b'Chelsea'	b'Watford'	b'21/10/2017 12:30:00'	b'1 : 1'	b'4 : 2'	

1310 rows × 58 columns

In [143]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.lineplot(data=df)
```

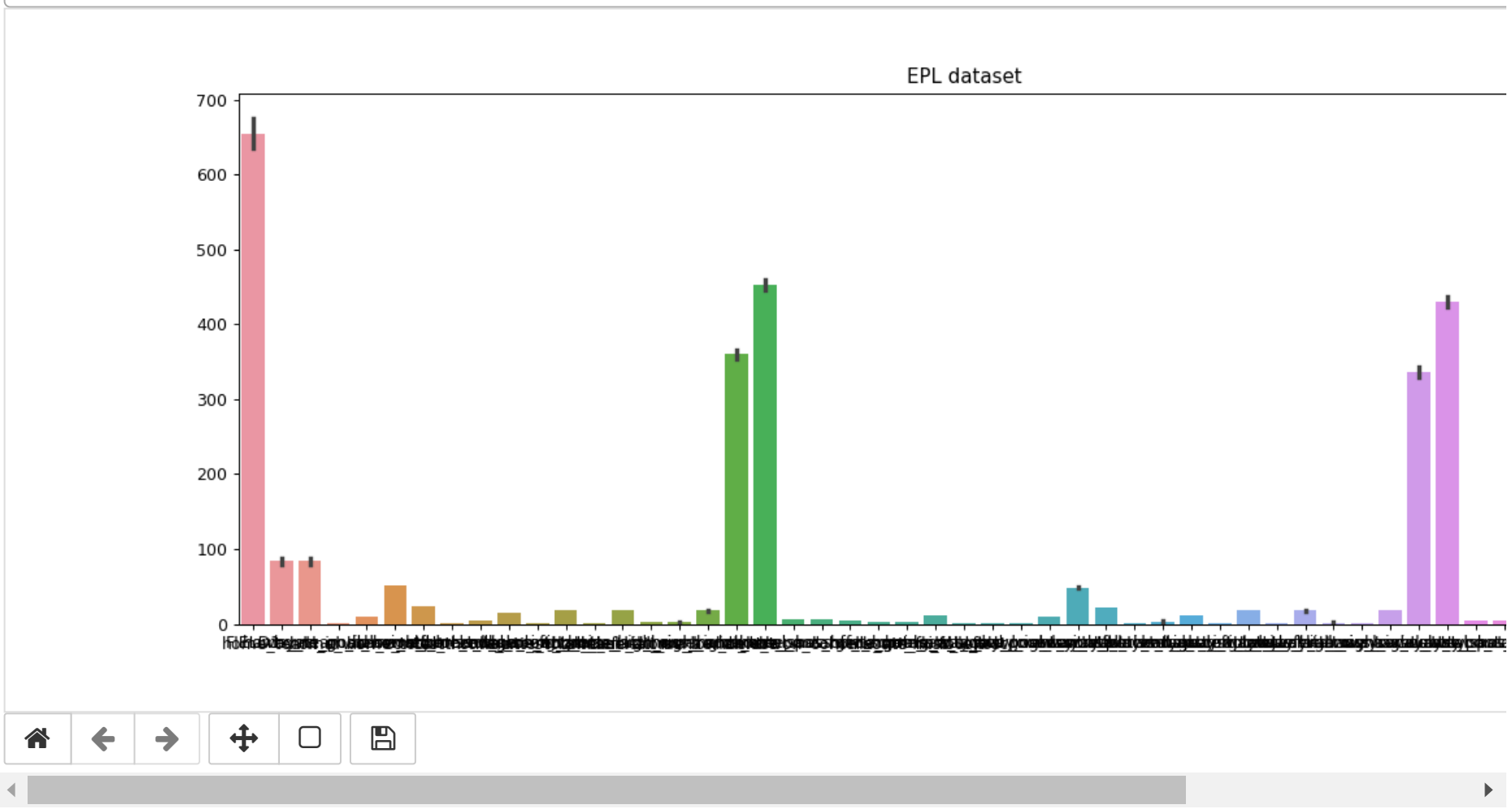
Figure 8



Out[143]:
```
<AxesSubplot:title={'center':'EPL dataset'}>
```

In [144]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.barplot(data=df)
```

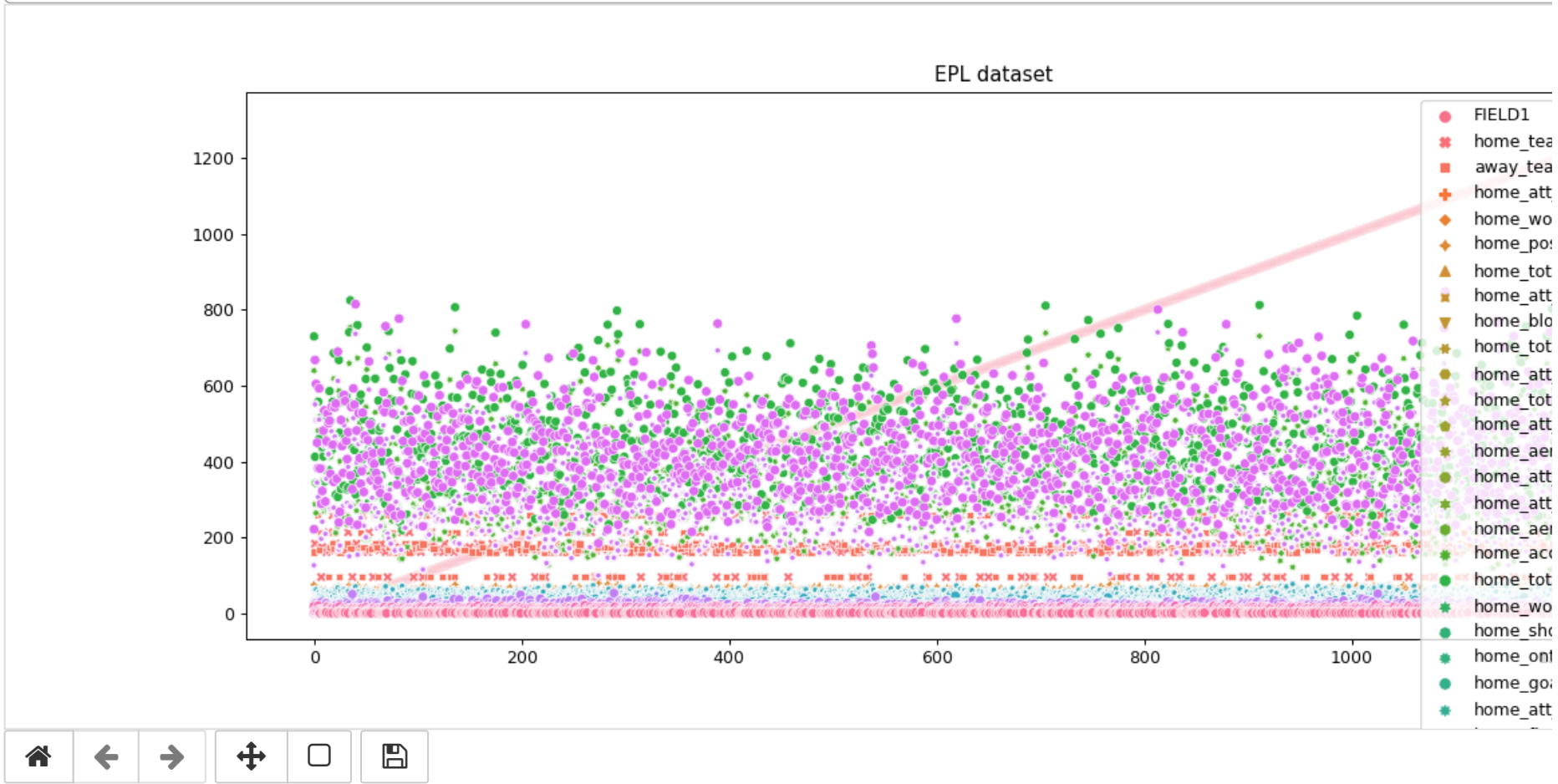
Figure 9



Out[144]:
```
<AxesSubplot:title={'center':'EPL dataset'}>
```

In [145]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.scatterplot(data=df)
```

Figure 10



Out[145]:
```
<AxesSubplot:title={'center':'EPL dataset'}>
```

In [146]:
```
!pip install xlrd
```

```
Requirement already satisfied: xlrd in c:\programdata\anaconda3\lib\site-packages (1.2.0)
```

In [147]:
```
!pip install openpyxl
```

```
Requirement already satisfied: openpyxl in c:\programdata\anaconda3\lib\site-packages (3.0.5)
Requirement already satisfied: et-xmlfile in c:\programdata\anaconda3\lib\site-packages (from openpyxl) (1.0.1)
Requirement already satisfied: jdcal in c:\programdata\anaconda3\lib\site-packages (from openpyxl) (1.4.1)
```

In [148]:
```
import xlrd
import openpyxl
```

In [150]:
```
df = pd.read_excel ('meu.xlsx')
dfxx = pd.DataFrame(df)
```

In [151]:
```
dfxx
```

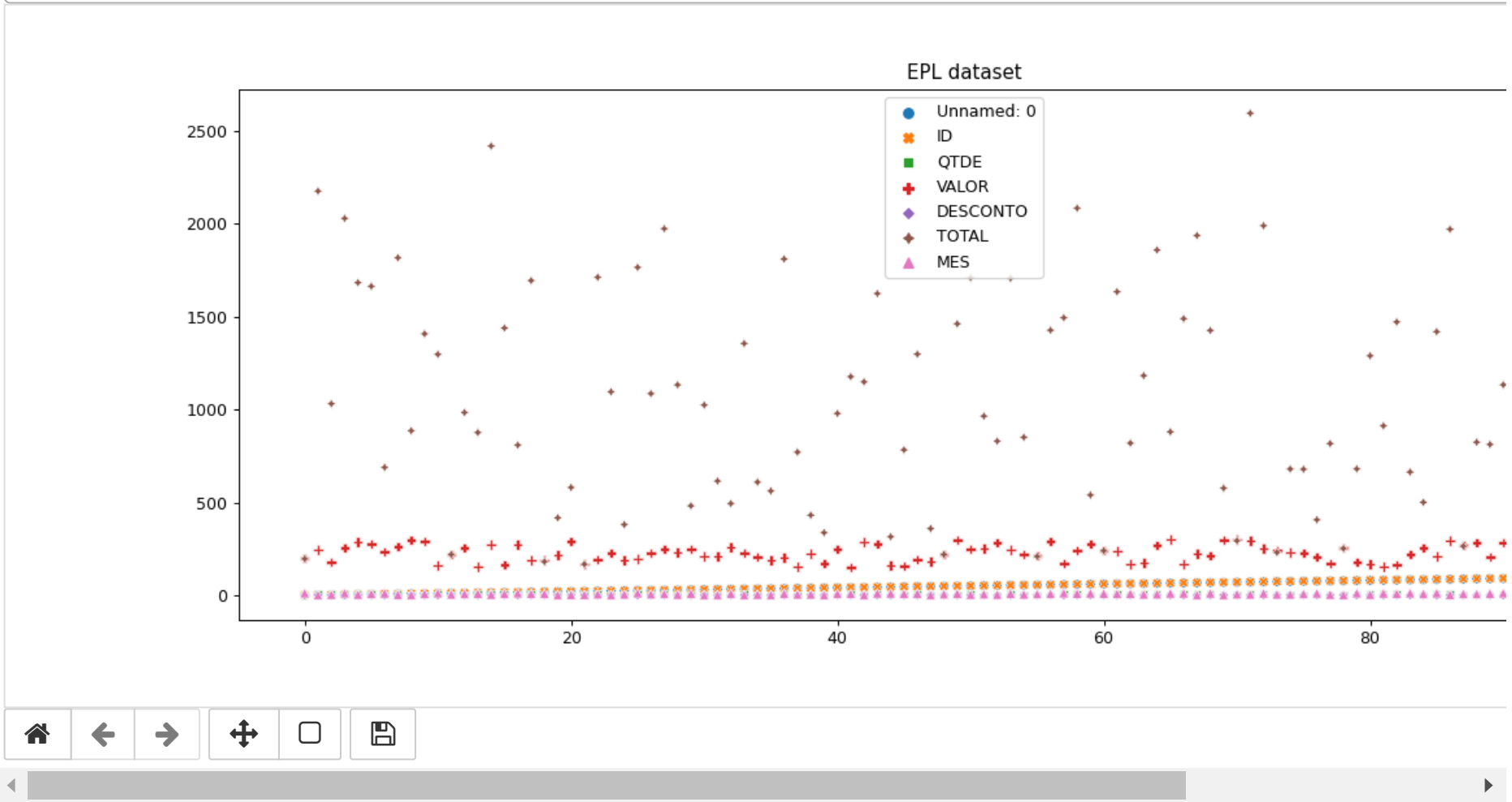
Out[151]:

	Unnamed: 0	ID	PRODUTO	CATEGORIA PRODUTO	QTDE	VALOR	DESCONTO	TOTAL	LOJA	VENDEDOR	CLIENTE	SEXO	ESTADO CIVIL	DATA	PERIC
0	0	1	I	BA	1	198	0.01	196.02	GGFF	DDEEE	WNMJUZ LDGWSXCV	-	d	2019-12-26	201
1	1	2	J	AC	9	244	0.01	2174.04	GGFG	EEEDE	WXFRMV YIDLUDFU	m	d	2019-01-28	201
2	2	3	D	BB	6	177	0.03	1030.14	GFGF	DEEEE	VOTGLV PGKAMASM	f	-	2019-04-08	201
3	3	4	O	AC	8	256	0.01	2027.52	GGFG	DEDED	HYTRVB LNLUCDLI	m	-	2019-12-29	201
4	4	5	F	CA	6	286	0.02	1681.68	GGGF	EEDDD	YTIIIO GNEOHMQT	f	c	2019-04-04	201
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
95	95	96	E	AC	3	296	0.04	852.48	GFFF	DEEDD	BJPWAO SDOQEACC	m	d	2019-05-05	201
96	96	97	H	BB	2	269	0.01	532.62	FGGG	EDDEE	QFFEVB POHRVWLA	f	-	2019-03-19	201
97	97	98	P	AA	5	186	0.03	902.10	FGGG	DEDEE	TDYDKL WAWNHEEA	f	d	2019-02-08	201
98	98	99	I	CB	1	164	0.01	162.36	GFGG	EEDDD	VMWTHS XMLLCBTE	a	s	2019-10-23	201
99	99	100	C	BA	5	157	0.04	753.60	FFGF	EEDDD	EVADZF YPCZFMAM	m	-	2019-09-13	201

100 rows × 16 columns

In [152]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.scatterplot(data=dfxx)
```

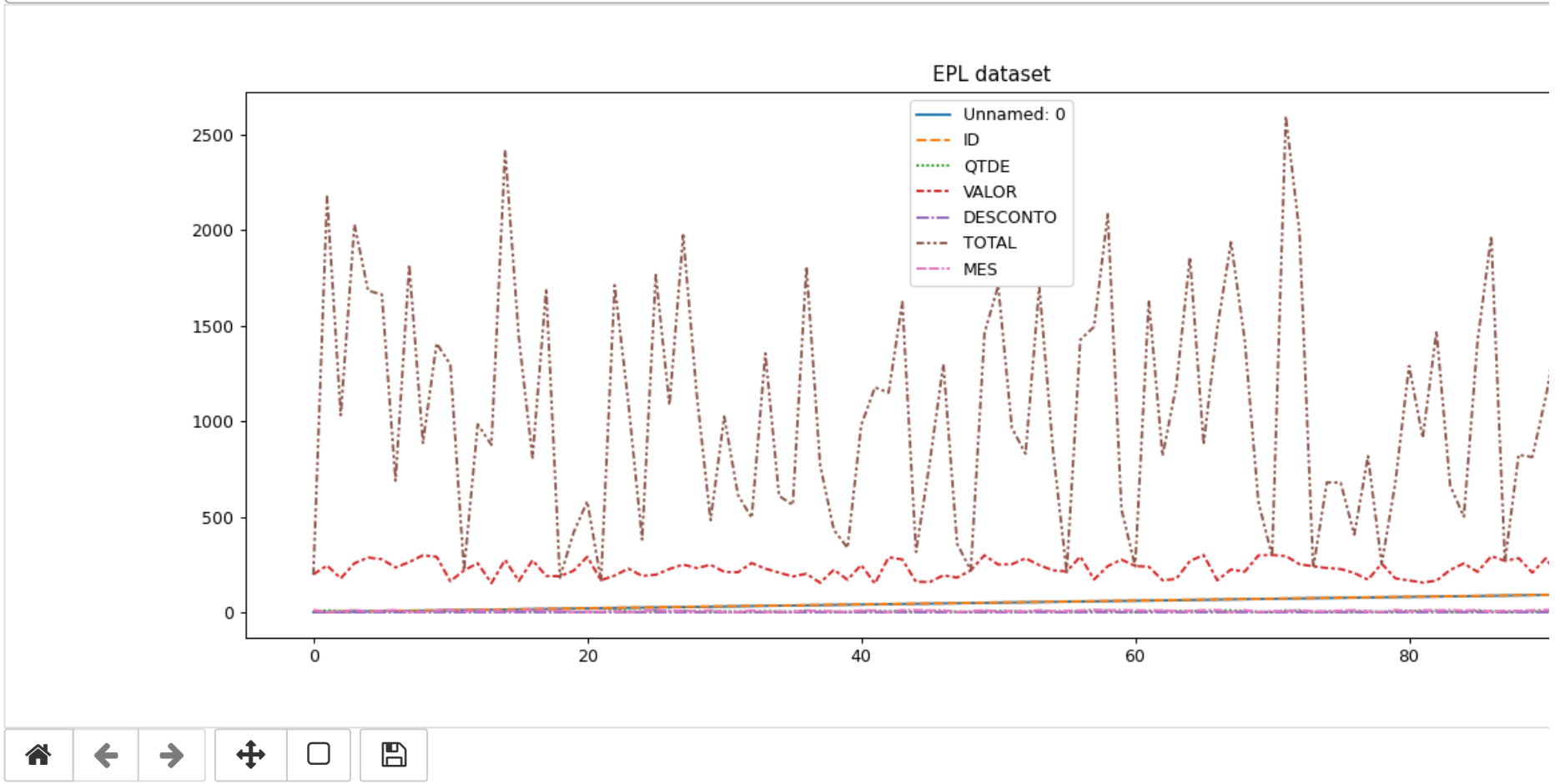
Figure 11



Out[152]:
```
<AxesSubplot:title={'center':'EPL dataset'}>
```

In [153]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.lineplot(data=dfxx)
```

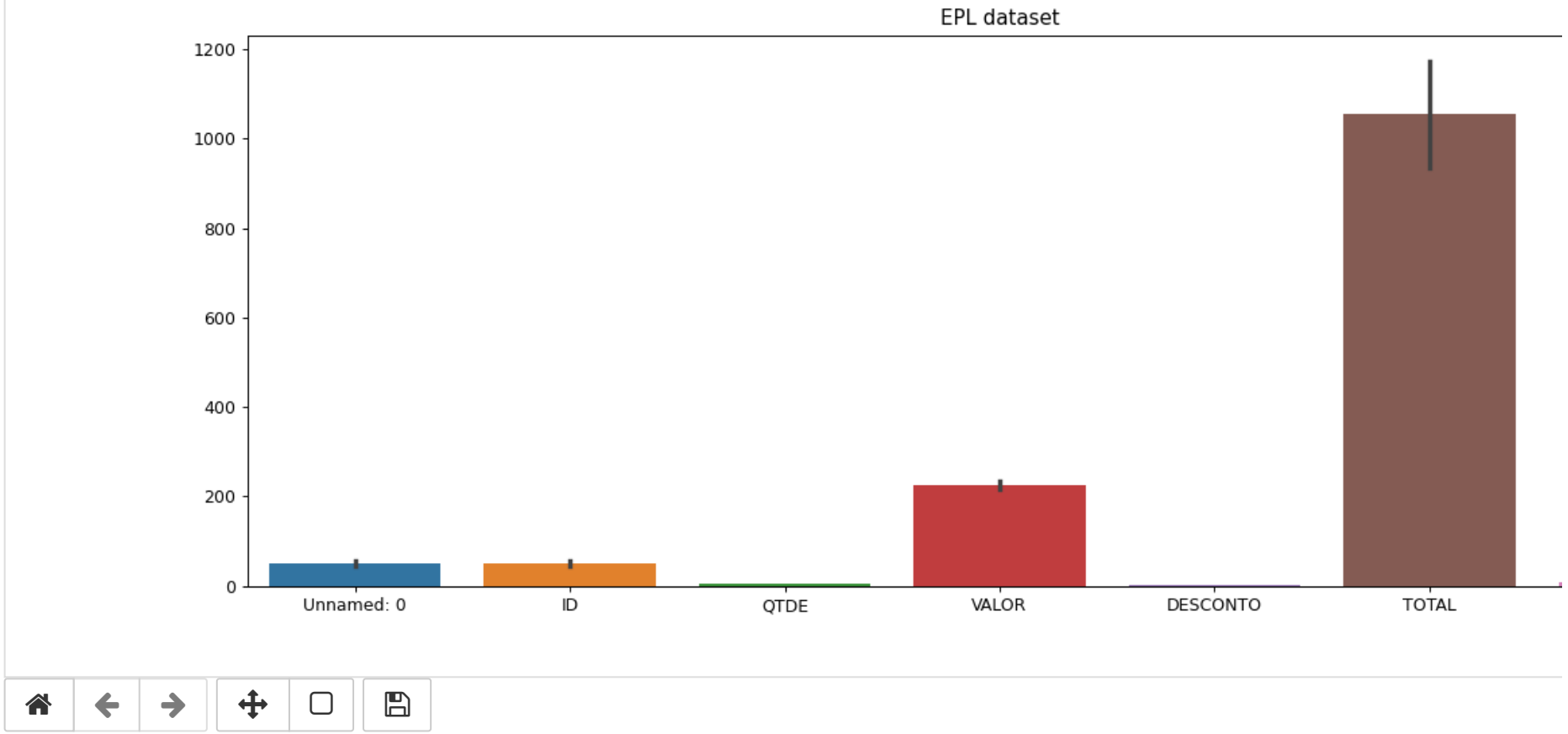
Figure 12



Out[153]: <AxesSubplot:title={'center':'EPL dataset'}>

In [154]:
```
plt.figure(figsize=(16,6))
plt.title("EPL dataset")
sns.barplot(data=dfxx)
```

Figure 13



Out[154]:
```
<AxesSubplot:title={'center':'EPL dataset'}>
```

In [155]:
```
df = pd.read_excel ('US Super.xls')
dfss = pd.DataFrame(df)
```

In [156]:
```
dfss
```

Out[156]:

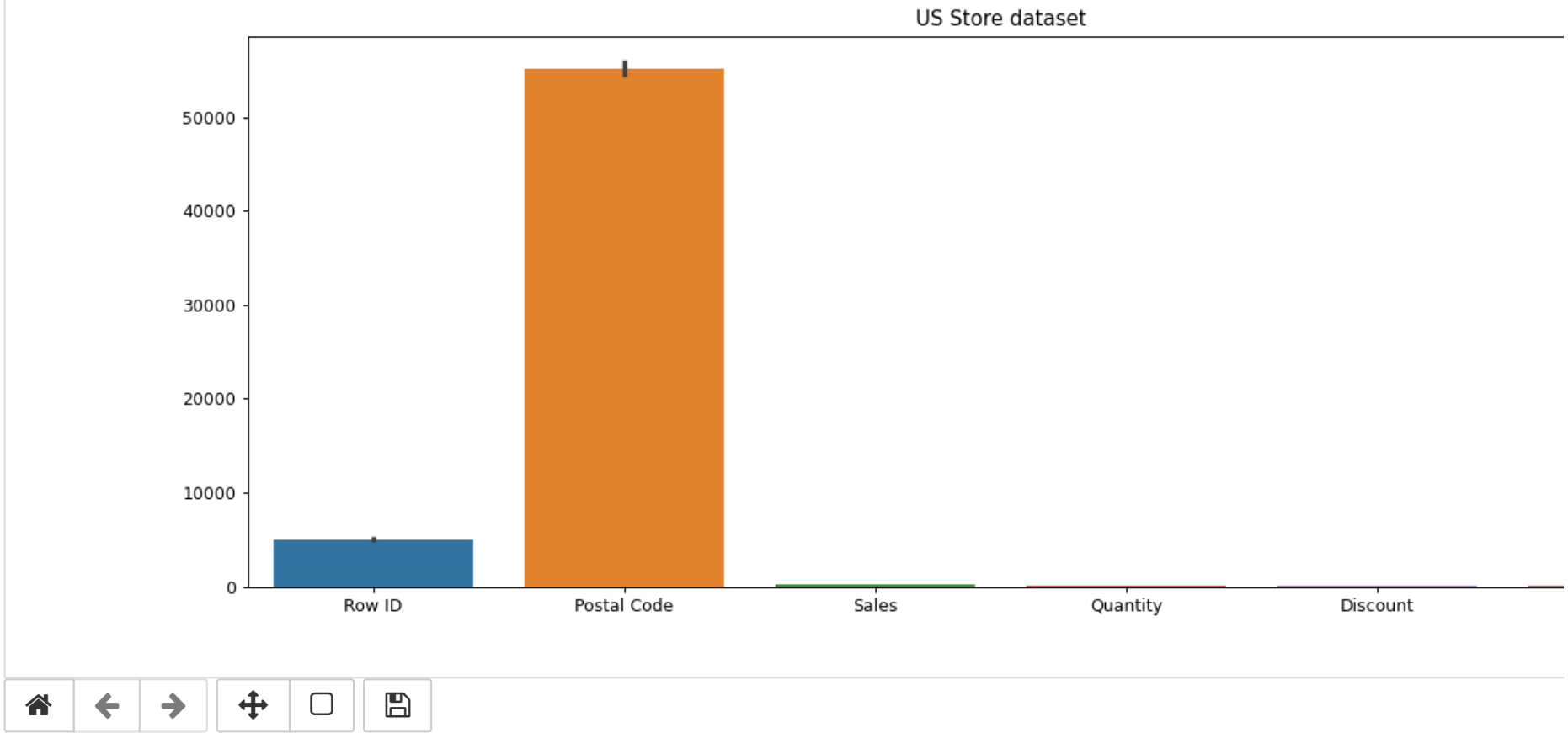
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	
0	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-BO-10001798	Furniture	B
1	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-CH-10000454	Furniture	
2	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-LA-10000240	Office Supplies	
3	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-TA-10000577	Furniture	
4	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-ST-10000760	Office Supplies	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
9989	9990	CA-2014-110422	2014-01-21	2014-01-23	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	33180	South	FUR-FU-10001889	Furniture	Fu
9990	9991	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	FUR-FU-10000747	Furniture	Fu
9991	9992	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	TEC-PH-10003645	Technology	

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	
9992	9993	CA-2017-121258	2017-02-26	2017-03-03	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	OFF-PA-10004041	Office Supplies	
9993	9994	CA-2017-119914	2017-05-04	2017-05-09	Second Class	CC-12220	Chris Cortes	Consumer	United States	Westminster	...	92683	West	OFF-AP-10002684	Office Supplies	A

9994 rows × 21 columns

In [157]:
```
plt.figure(figsize=(16,6))
plt.title("US Store dataset")
sns.barplot(data=dfss)
```

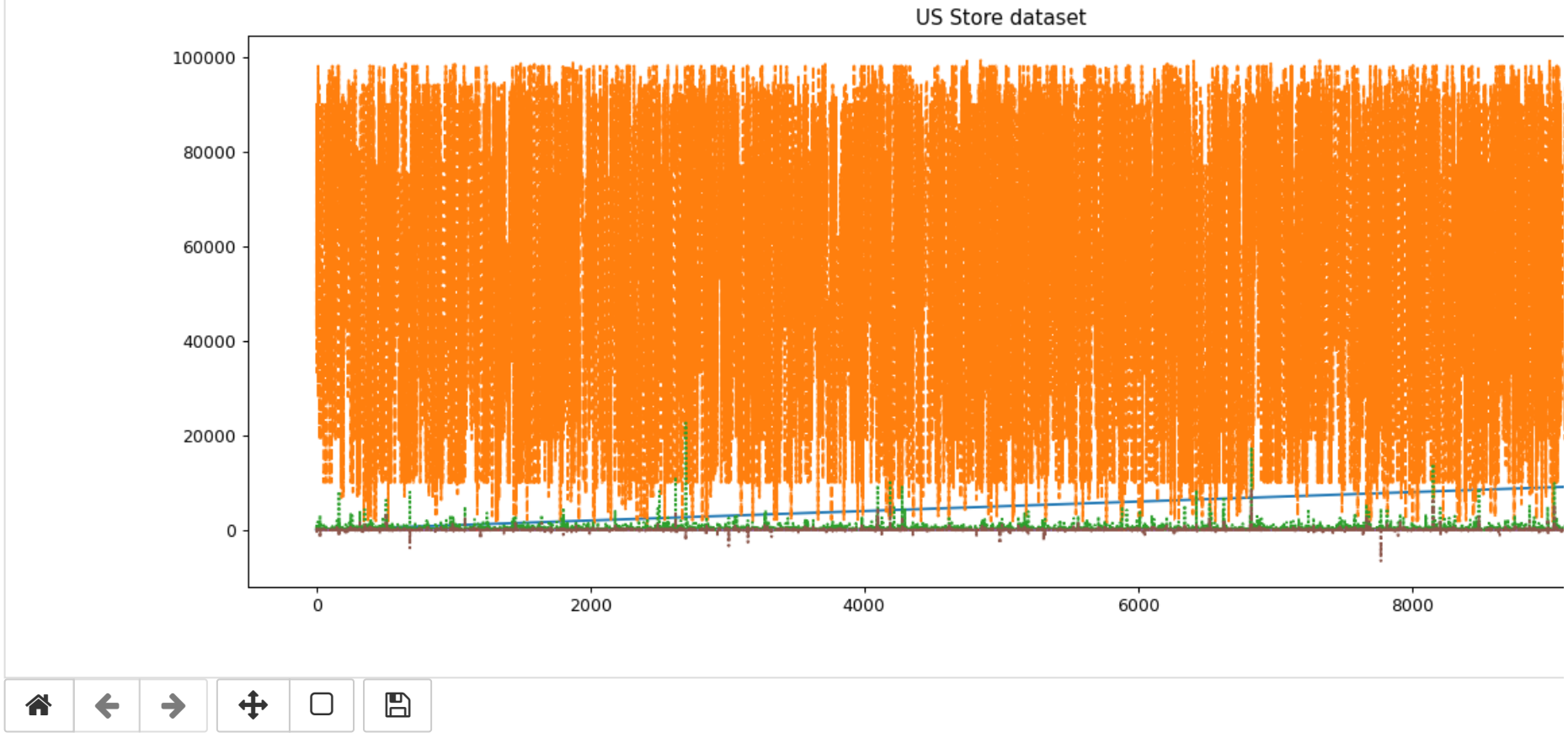
Figure 14



Out[157]: <AxesSubplot:title={'center':'US Store dataset'}>

In [158]:
```
plt.figure(figsize=(16,6))
plt.title("US Store dataset")
sns.lineplot(data=dfss)
```

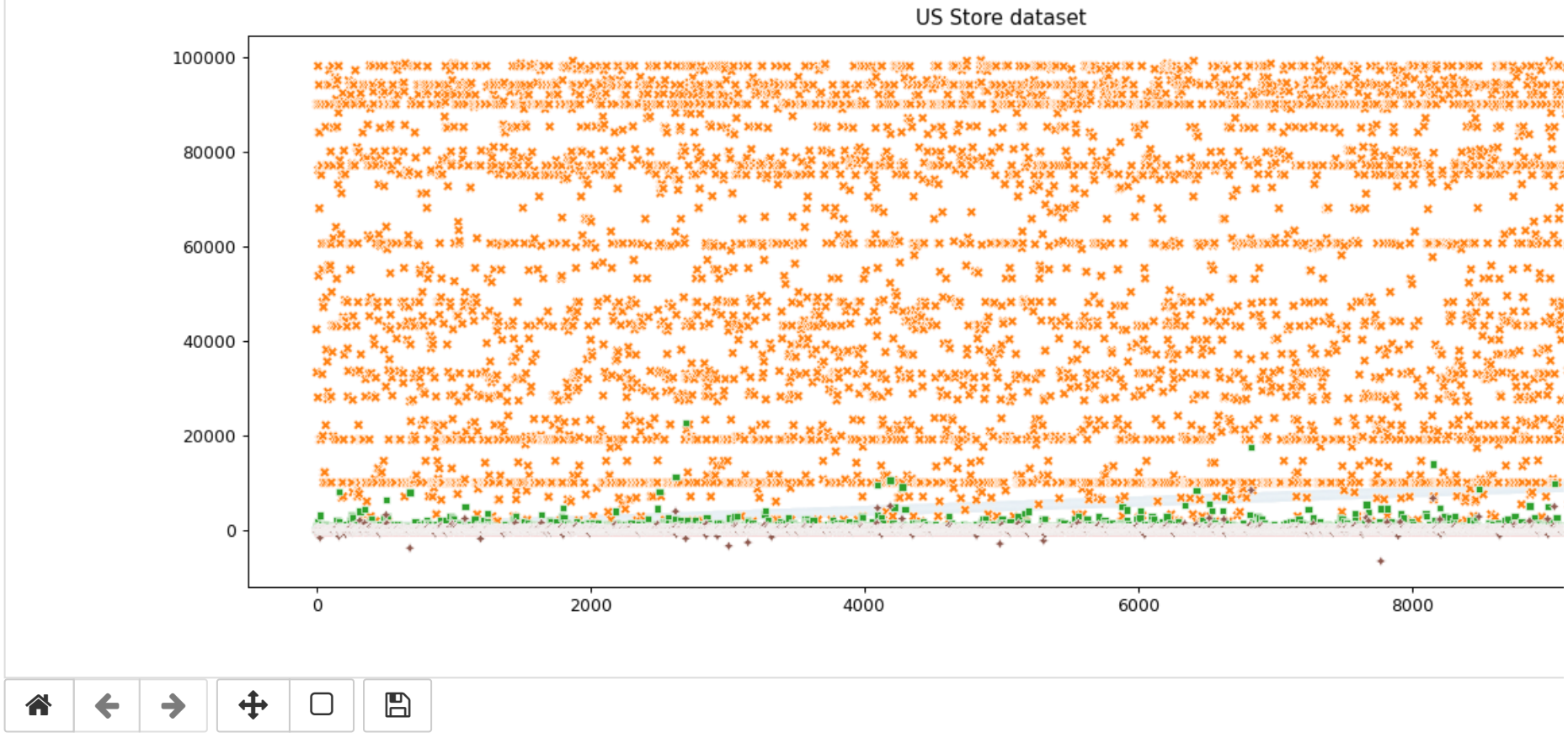
Figure 15



Out[158]: <AxesSubplot:title={'center':'US Store dataset'}>

In [159]:
```
plt.figure(figsize=(16,6))
plt.title("US Store dataset")
sns.scatterplot(data=dfss)
```

Figure 16



Out[159]: <AxesSubplot:title={'center':'US Store dataset'}>

In [160]:
```
!pip install lxml
```

```
Requirement already satisfied: lxml in c:\programdata\anaconda3\lib\site-packages (4.6.1)
```

In [161]:
```
from lxml import objectify
import pandas as pd
```

In [162]:
```
xml_data = objectify.parse('API.xml')  # Parse XML data
root = xml_data.getroot()  # Root element

data = []
cols = []

for i in range(len(root.getchildren())):
    child = root.getchildren()[i]
    data.append([subchild.text for subchild in child.getchildren()])
    cols.append(child.tag)

dfl = pd.DataFrame(data).T
dfl.columns = cols
print(dfl)
```

```
       data
0      None
1      None
2      None
3      None
4      None
...     ...
15835  None
15836  None
15837  None
15838  None
15839  None

[15840 rows x 1 columns]
```

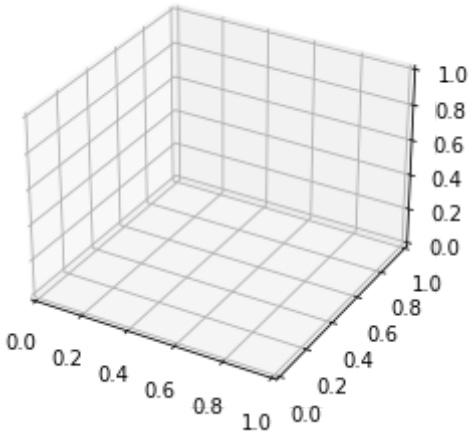
Image visualization: image plot, 3d plot

In [99]:
```
%matplotlib inline
import imageio
import matplotlib.pyplot as plt
import matplotlib.cbook

pic=imageio.imread('bhc.jpg')
plt.figure(figsize=(6,6))
plt.imshow(pic)
plt.axis('off')
```

Out[99]: (-0.5, 473.5, 646.5, -0.5)

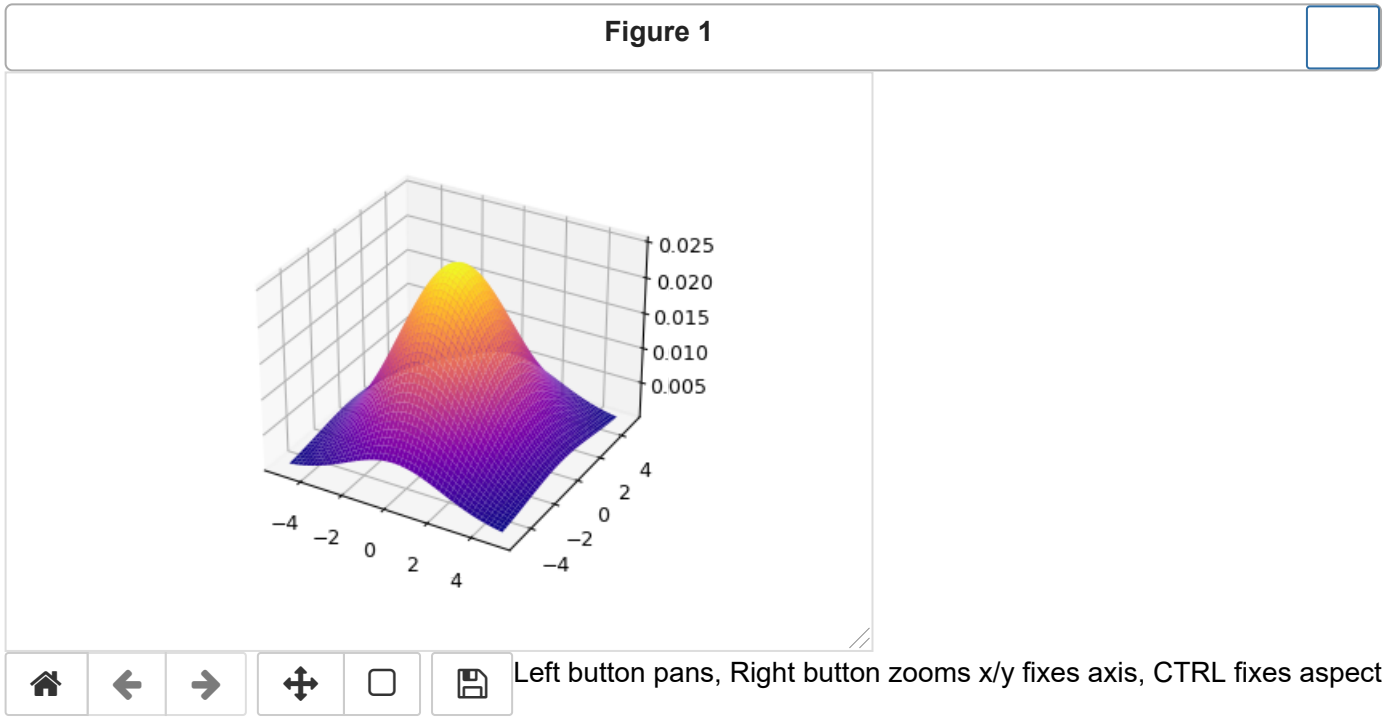
In [39]:
```
import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import Axes3D
```

In [40]:
```
fig = plt.figure(figsize=(4,4))

ax = fig.add_subplot(111, projection='3d')
```



In [100]:

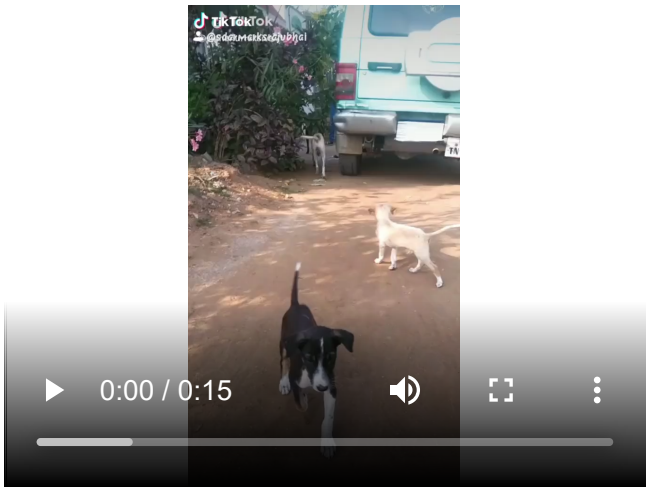
```
%matplotlib notebook
import matplotlib.pyplot as plt

from mpl_toolkits.mplot3d import Axes3D

import numpy as np

from scipy.stats import multivariate_normal

X = np.linspace(-5,5,50)

Y = np.linspace(-5,5,50)

X, Y = np.meshgrid(X,Y)

X_mean = 0; Y_mean = 0

X_var = 5; Y_var = 8

pos = np.empty(X.shape+(2,))

pos[:,:,0]=X

pos[:,:,1]=Y

rv = multivariate_normal([X_mean, Y_mean],[[X_var, 0], [0, Y_var]])

fig = plt.figure()

ax = fig.add_subplot(111, projection='3d')

ax.plot_surface(X, Y, rv.pdf(pos), cmap="plasma")

plt.show()
```



Video visualization: video player

In [109]:
```
from IPython.display import HTML

HTML("""
    <video width="320" height="240" controls>
        <source src="dog.mp4" type="video/mp4">
    </video>
""")
```

Out[109]:



Audio visualization: audio player, spectrogram

In [114]:
```
!pip install librosa
```

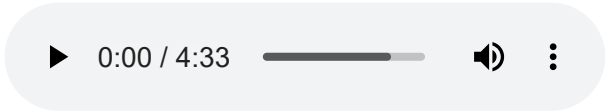
```
Requirement already satisfied: librosa in c:\programdata\anaconda3\lib\site-packages (0.8.0)
Requirement already satisfied: numba>=0.43.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (0.51.2)
Requirement already satisfied: soundfile>=0.9.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (0.10.
3.post1)
Requirement already satisfied: scipy>=1.0.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (1.5.2)
Requirement already satisfied: decorator>=3.0.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (4.4.2)
Requirement already satisfied: audioread>=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (2.1.9)
Requirement already satisfied: numpy>=1.15.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (1.19.2)
Requirement already satisfied: joblib>=0.14 in c:\programdata\anaconda3\lib\site-packages (from librosa) (0.17.0)
Requirement already satisfied: scikit-learn!=0.19.0,>=0.14.0 in c:\programdata\anaconda3\lib\site-packages (from lib
rosa) (0.23.2)
Requirement already satisfied: resampy>=0.2.2 in c:\programdata\anaconda3\lib\site-packages (from librosa) (0.2.2)
Requirement already satisfied: pooch>=1.0 in c:\programdata\anaconda3\lib\site-packages (from librosa) (1.3.0)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in c:\programdata\anaconda3\lib\site-packages (from numba
>=0.43.0->librosa) (0.34.0)
Requirement already satisfied: setuptools in c:\programdata\anaconda3\lib\site-packages (from numba>=0.43.0->libros
a) (50.3.1.post20201107)
Requirement already satisfied: cffi>=1.0 in c:\programdata\anaconda3\lib\site-packages (from soundfile>=0.9.0->libro
sa) (1.14.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from scikit-lear
n!=0.19.0,>=0.14.0->librosa) (2.1.0)
Requirement already satisfied: six>=1.3 in c:\programdata\anaconda3\lib\site-packages (from resampy>=0.2.2->librosa)
(1.15.0)
Requirement already satisfied: packaging in c:\programdata\anaconda3\lib\site-packages (from pooch>=1.0->librosa) (2
0.4)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from pooch>=1.0->librosa) (2.
24.0)
Requirement already satisfied: appdirs in c:\programdata\anaconda3\lib\site-packages (from pooch>=1.0->librosa) (1.
4.4)
Requirement already satisfied: pycparser in c:\programdata\anaconda3\lib\site-packages (from cffi>=1.0->soundfile>=
0.9.0->librosa) (2.20)
Requirement already satisfied: pyparsing>=2.0.2 in c:\programdata\anaconda3\lib\site-packages (from packaging->pooch
>=1.0->librosa) (2.4.7)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages
(from requests->pooch>=1.0->librosa) (1.25.11)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from requests->pooch
>=1.0->librosa) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests->pooc
h>=1.0->librosa) (2020.6.20)
```

```
Requirement already satisfied: idna<3,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->pooch>=1.0
->librosa) (2.10)
```

In [115]:
```
import IPython.display as ipd
import librosa
import librosa.display
import matplotlib.pyplot as plt
```

In [116]:
```
ipd.Audio('Ennavale_Adi_Ennavale_-_Kaadhalan_-_HD.mp3')
```

Out[116]:



Text visualization: Word cloud, bubble cloud (some more in [http://vallandingham.me/textvis talk/](http://vallandingham.me/textvis%02talk/))

In [113]:
```
!pip install wordcloud
```

```
Collecting wordcloud
  Downloading wordcloud-1.8.1-cp38-cp38-win_amd64.whl (155 kB)
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (3.3.2)
Requirement already satisfied: numpy>=1.6.1 in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (1.19.2)
Requirement already satisfied: pillow in c:\programdata\anaconda3\lib\site-packages (from wordcloud) (8.0.1)
Requirement already satisfied: certifi>=2020.06.20 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->word
cloud) (2020.6.20)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcloud)
(0.10.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in c:\programdata\anaconda3\lib\site-packages
(from matplotlib->wordcloud) (2.4.7)
Requirement already satisfied: python-dateutil>=2.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wor
dcloud) (2.8.1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib->wordcl
oud) (1.3.0)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-packages (from cycler>=0.10->matplotlib->wordcl
oud) (1.15.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.8.1
```

In [124]:

```
# Python program to generate WordCloud

# importing all necessery modules
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
import pandas as pd

# Reads 'Youtube04-Eminem.csv' file
df = pd.read_csv(r"Youtube04-Eminem.csv", encoding ="latin-1")

comment_words = ''
stopwords = set(STOPWORDS)

# iterate through the csv file
for val in df.CONTENT:
    val = str(val)
    tokens = val.split()
# Converts each token into lowercase
for i in range(len(tokens)):
    tokens[i] = tokens[i].lower()
    comment_words += " ".join(tokens)+" "

wordcloud = WordCloud(width = 800, height = 800,
background_color ='white',
stopwords = stopwords,
min_font_size = 10).generate(comment_words)

# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```

Figure 3

x=402. y=584. [255, 255, 255]