

Udemy Course Data Visualization

Import libraries

```
In [1]: import pandas as pd
import numpy as np

import seaborn as sb
import matplotlib.pyplot as plt
```

Read dataset

```
In [4]: d=pd.read_csv("C:/Users/Ajay/Downloads/7. Udemy Courses.csv")
```

```
In [5]: d
```

Out[5]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Levels	1.5 hours	2014-09-18T05:07:05Z	Musical Instruments
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermediate Level	1 hour	2017-04-12T19:06:34Z	Musical Instruments
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermediate Level	1.5 hours	2017-04-26T18:34:57Z	Musical Instruments
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermediate Level	1 hour	2017-02-21T23:48:18Z	Musical Instruments
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermediate Level	37 mins	2017-02-21T23:44:49Z	Musical Instruments
...
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Levels	4 hours	2015-04-20T22:15:17Z	Web Development
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Beginner Level	5 hours	2015-10-26T20:04:21Z	Musical Instruments
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Levels	6.5 hours	2015-11-11T18:55:45Z	Web Development
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Levels	2 hours	2017-05-26T16:45:55Z	Business Finance
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Levels	1.5 hours	2016-09-26T22:19:48Z	Graphic Design

3682 rows × 11 columns

```
In [5]: # showing first five rows with head()
d.head()
```

Out[5]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Levels	1.5 hours	2014-09-18T05:07:05Z	Musical Instruments
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermediate Level	1 hour	2017-04-12T19:06:34Z	Musical Instruments
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermediate Level	1.5 hours	2017-04-26T18:34:57Z	Musical Instruments
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermediate Level	1 hour	2017-02-21T23:48:18Z	Musical Instruments
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermediate Level	37 mins	2017-02-21T23:44:49Z	Musical Instruments

```
In [6]: d.head(5)
```

Out[6]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Levels	1.5 hours	2014-09-18T05:07:05Z	Musical Instruments
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermediate Level	1 hour	2017-04-12T19:06:34Z	Musical Instruments
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermediate Level	1.5 hours	2017-04-26T18:34:57Z	Musical Instruments
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermediate Level	1 hour	2017-02-21T23:48:18Z	Musical Instruments
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermediate Level	37 mins	2017-02-21T23:44:49Z	Musical Instruments

```
In [7]: # Last 10 rows
d.tail()
```

Out[7]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Levels	4 hours	2015-04-20T22:15:17Z	Web Development
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Beginner Level	5 hours	2015-10-26T20:04:21Z	Musical Instruments
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Levels	6.5 hours	2015-11-11T18:55:45Z	Web Development
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Levels	2 hours	2017-05-26T16:45:55Z	Business Finance
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Levels	1.5 hours	2016-09-26T22:19:48Z	Graphic Design

```
In [8]: d.tail(5)
```

Out[8]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Levels	4 hours	2015-04-20T22:15:17Z	Web Development
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Beginner Level	5 hours	2015-10-26T20:04:21Z	Musical Instruments
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Levels	6.5 hours	2015-11-11T18:55:45Z	Web Development
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Levels	2 hours	2017-05-26T16:45:55Z	Business Finance
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Levels	1.5 hours	2016-09-26T22:19:48Z	Graphic Design

```
In [6]: ## shape of the data
d.shape
```

Out[6]: (3682, 11)

```
In [7]: # view dataframe summary
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3682 entries, 0 to 3681
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   course_id             3682 non-null   int64
 1   course_title          3682 non-null   object
 2   is_paid               3682 non-null   bool
 3   price                 3682 non-null   object
 4   num_subscribers       3682 non-null   int64
 5   num_reviews           3682 non-null   int64
 6   num_lectures          3682 non-null   int64
 7   level                 3682 non-null   object
 8   content_duration      3682 non-null   object
 9   published_timestamp   3682 non-null   object
10   subject               3682 non-null   object
dtypes: bool(1), int64(4), object(6)
memory usage: 291.4+ KB
```

```
In [9]: # to view the maximum (num_subscribers)
d['num_subscribers'].max()
```

```
Out[9]: 268923
```

```
In [10]: #to show the unique
d['subject'].unique()
```

```
Out[10]: array(['Musical Instruments', 'Business Finance', 'Graphic Design',
               'Web Development'], dtype=object)
```

```
In [11]: # check for missing values
print("\nMissing values: ",d.isnull().sum().values.sum())
```

```
Missing values:  0
```

```
In [12]: d.describe()
```

```
Out[12]:
```

	course_id	num_subscribers	num_reviews	num_lectures
count	3.682000e+03	3682.000000	3682.000000	3682.000000
mean	6.766121e+05	3194.230310	156.093156	40.065182
std	3.436355e+05	9499.378361	934.957204	50.373299
min	8.324000e+03	0.000000	0.000000	0.000000
25%	4.078430e+05	110.250000	4.000000	15.000000
50%	6.885580e+05	911.500000	18.000000	25.000000
75%	9.617515e+05	2540.250000	67.000000	45.000000
max	1.282064e+06	268923.000000	27445.000000	779.000000

```
In [13]: d.isnull().any()
```

```
Out[13]: course_id      False
course_title    False
is_paid         False
price           False
num_subscribers False
num_reviews     False
num_lectures    False
level           False
content_duration False
published_timestamp False
subject         False
dtype: bool
```

```
In [15]: d['subject'].value_counts()
```

```
Out[15]: Web Development      1200
Business Finance             1199
Musical Instruments          680
Graphic Design               603
Name: subject, dtype: int64
```

```
In [16]: d.iloc[d['num_lectures'].idxmax()]
```

```
Out[16]: course_id      79154
course_title    Back to School Web Development and Programming...
is_paid         True
price           200
num_subscribers 18170
num_reviews     117
num_lectures    779
level           Beginner Level
content_duration 44.5 hours
published_timestamp 2013-08-19T14:53:00Z
subject         Web Development
Name: 394, dtype: object
```

```
In [17]: d.iloc[d['num_reviews'].idxmax()]
```

```
Out[17]: course_id      625204
course_title    The Web Developer Bootcamp
is_paid         True
price           200
num_subscribers 121584
num_reviews     27445
num_lectures    342
level           All Levels
content_duration 43 hours
published_timestamp 2015-11-02T21:13:27Z
subject         Web Development
Name: 3385, dtype: object
```

```
In [21]: print(d.groupby('content_duration')['published_timestamp'].value_counts())
```

```
content_duration  published_timestamp
0                2015-12-17T05:38:38Z    1
1 hour           2016-12-15T14:56:17Z    2
                2012-06-18T17:06:55Z    1
                2012-08-31T22:06:29Z    1
                2013-02-16T16:08:05Z    1
                ..
9.5 hours        2017-01-04T23:51:26Z    1
                2017-02-01T23:42:53Z    1
                2017-02-08T00:40:08Z    1
                2017-03-23T05:54:27Z    1
                2017-06-02T18:58:51Z    1
Name: published_timestamp, Length: 3676, dtype: int64
```

```
In [22]: d.groupby('level')['subject'].value_counts()
```

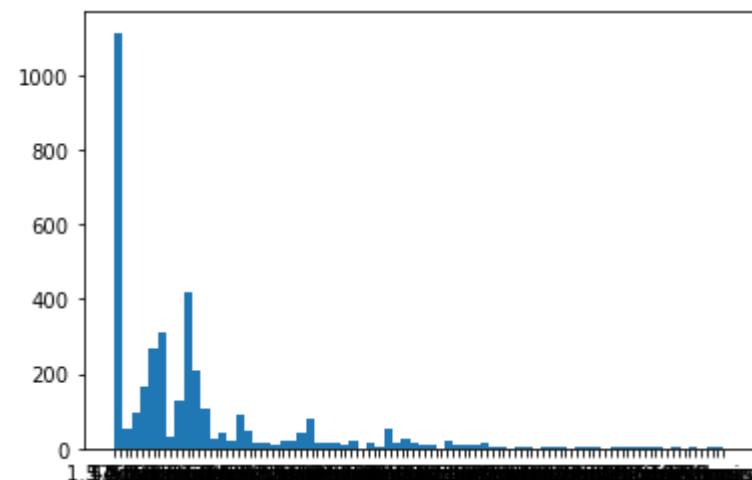
```
Out[22]: level      subject
All Levels      Business Finance    699
               Web Development     659
               Graphic Design      298
               Musical Instruments  276
Beginner Level  Web Development     391
               Business Finance     341
               Musical Instruments  296
               Graphic Design      243
Expert Level    Business Finance     31
               Web Development      15
               Musical Instruments   7
               Graphic Design        5
Intermediate Level Web Development  135
               Business Finance     128
               Musical Instruments  101
               Graphic Design        57
Name: subject, dtype: int64
```

```
In [6]: d['level'].value_counts()
```

```
Out[6]: All Levels      1932
Beginner Level    1271
Intermediate Level  421
Expert Level       58
Name: level, dtype: int64
```

```
In [7]: plt.hist(d['content_duration'],bins=70)
```

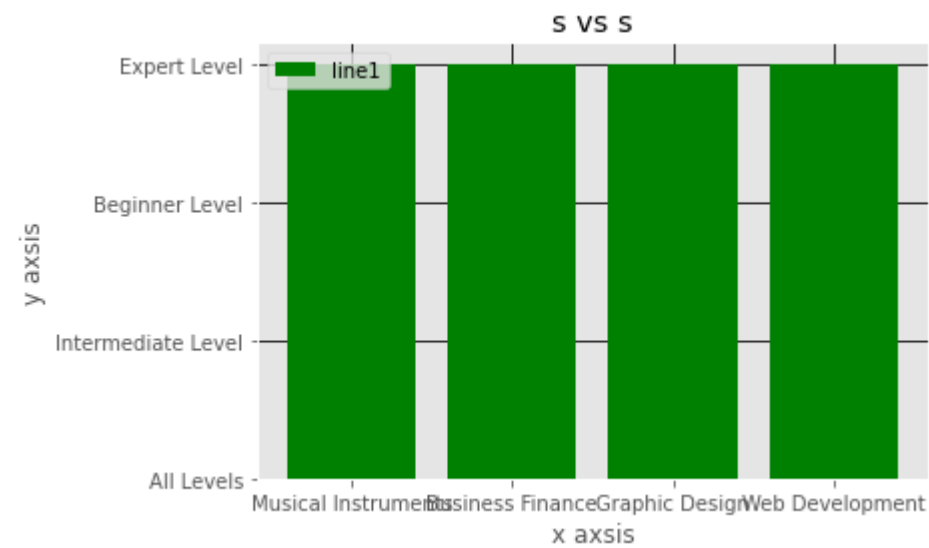
```
Out[7]: (array([1.114e+03, 5.500e+01, 9.400e+01, 1.630e+02, 2.690e+02, 3.090e+02,
 2.900e+01, 1.300e+02, 4.190e+02, 2.100e+02, 1.080e+02, 2.800e+01,
 4.000e+01, 1.800e+01, 8.900e+01, 4.800e+01, 1.600e+01, 1.600e+01,
 1.000e+01, 2.000e+01, 2.300e+01, 4.100e+01, 8.200e+01, 1.700e+01,
 1.700e+01, 1.500e+01, 9.000e+00, 2.300e+01, 1.000e+00, 1.600e+01,
 2.000e+00, 5.100e+01, 1.600e+01, 2.600e+01, 1.600e+01, 8.000e+00,
 1.000e+01, 1.000e+00, 1.900e+01, 8.000e+00, 7.000e+00, 7.000e+00,
 1.600e+01, 2.000e+00, 6.000e+00, 1.000e+00, 5.000e+00, 6.000e+00,
 1.000e+00, 2.000e+00, 4.000e+00, 2.000e+00, 1.000e+00, 4.000e+00,
 3.000e+00, 2.000e+00, 1.000e+00, 2.000e+00, 2.000e+00, 2.000e+00,
 5.000e+00, 2.000e+00, 2.000e+00, 1.000e+00, 2.000e+00, 1.000e+00,
 2.000e+00, 1.000e+00, 2.000e+00, 2.000e+00]),
array([ 0.          ,  1.54285714,  3.08571429,  4.62857143,
 6.17142857,  7.71428571,  9.25714286, 10.8          ,
12.34285714, 13.88571429, 15.42857143, 16.97142857,
18.51428571, 20.05714286, 21.6          , 23.14285714,
24.68571429, 26.22857143, 27.77142857, 29.31428571,
30.85714286, 32.4          , 33.94285714, 35.48571429,
37.02857143, 38.57142857, 40.11428571, 41.65714286,
43.2          , 44.74285714, 46.28571429, 47.82857143,
49.37142857, 50.91428571, 52.45714286, 54.          ,
55.54285714, 57.08571429, 58.62857143, 60.17142857,
61.71428571, 63.25714286, 64.8          , 66.34285714,
67.88571429, 69.42857143, 70.97142857, 72.51428571,
74.05714286, 75.6          , 77.14285714, 78.68571429,
80.22857143, 81.77142857, 83.31428571, 84.85714286,
86.4          , 87.94285714, 89.48571429, 91.02857143,
92.57142857, 94.11428571, 95.65714286, 97.2          ,
98.74285714, 100.28571429, 101.82857143, 103.37142857,
104.91428571, 106.45714286, 108.          ]),
<BarContainer object of 70 artists>)
```



```
In [6]: from matplotlib import pyplot as plt
from matplotlib import style
```

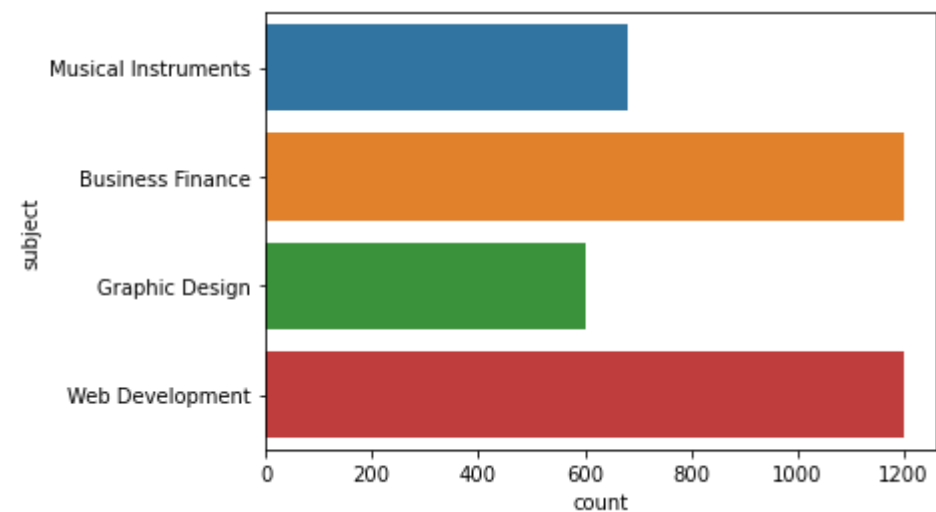
```
In [12]: style.use('ggplot')
x=d["subject"]
y=d["level"]
plt.bar(x,y,label='line1',color='g',linewidth=4)

plt.title("s vs s")
plt.ylabel('y axis')
plt.xlabel('x axis')
plt.legend()
plt.grid(True,color="k")
plt.show()
```

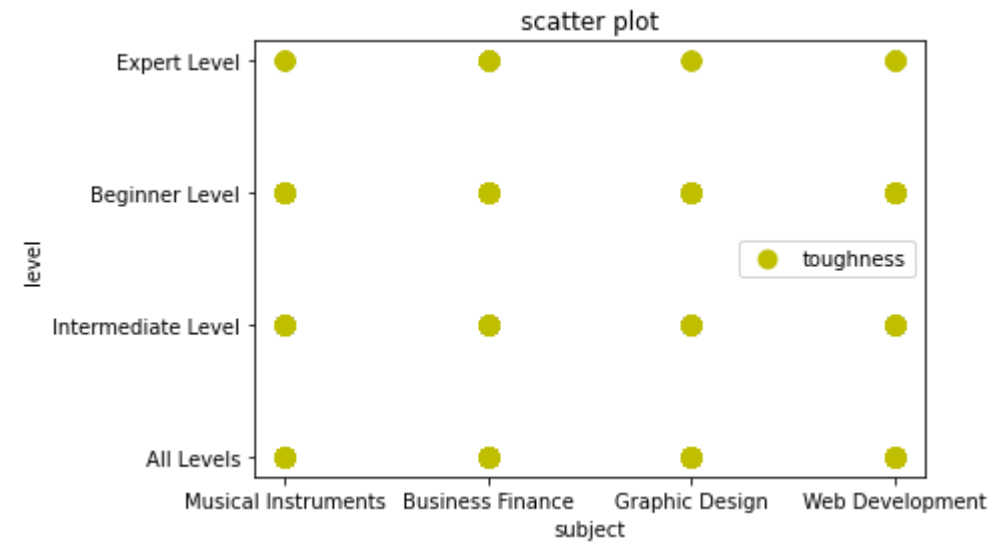


```
In [5]: sb.countplot(y='subject',data=d)
```

```
Out[5]: <AxesSubplot:xlabel='count', ylabel='subject'>
```



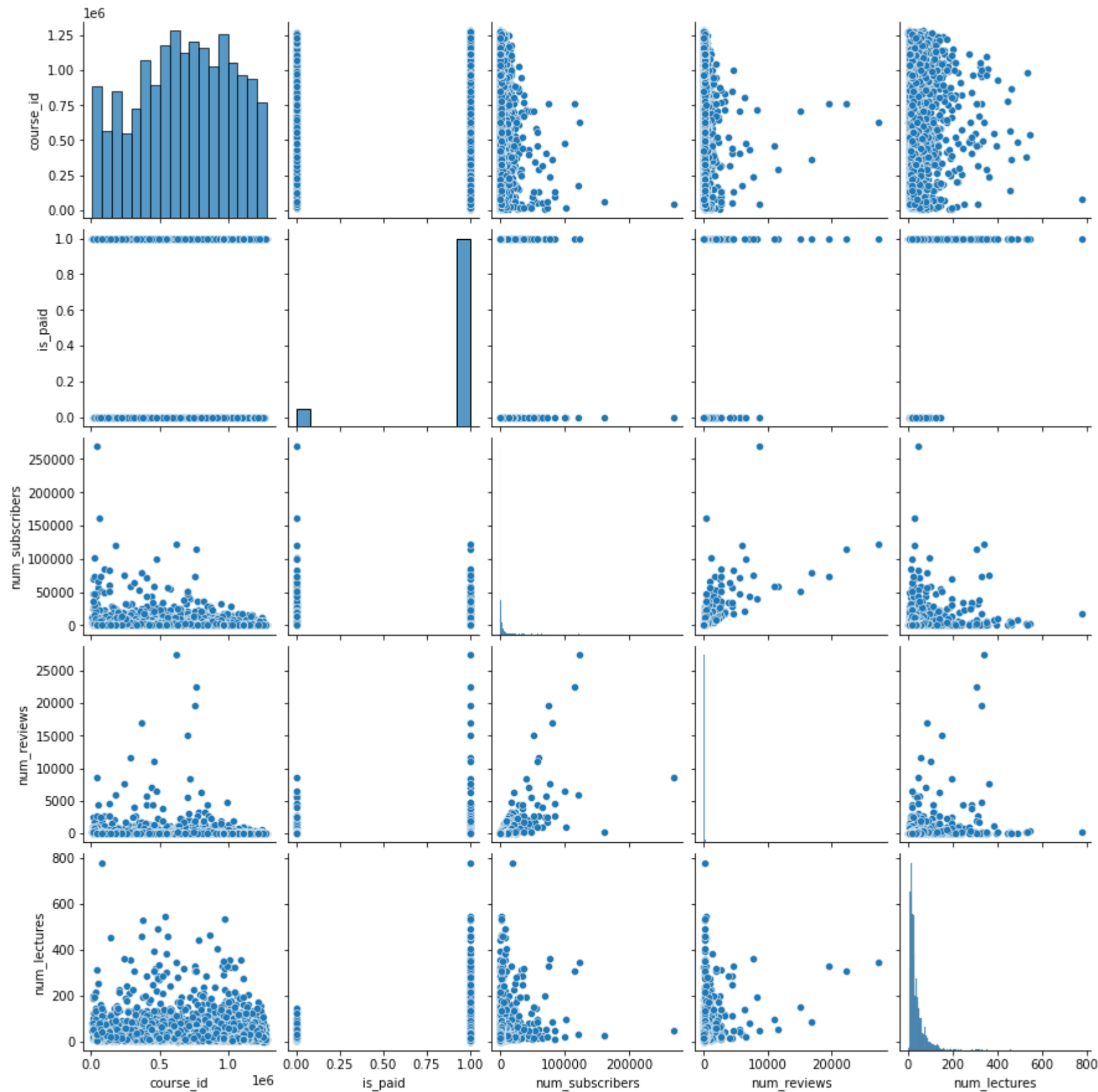
```
In [8]: x=d['subject']  
y=d['level']  
plt.scatter(x,y,label="toughness",color="y",linewidth=4)  
plt.title("scatter plot")  
plt.ylabel('level')  
plt.xlabel('subject')  
plt.legend()  
plt.show()
```




```
In [9]: sb.pairplot(d)
```

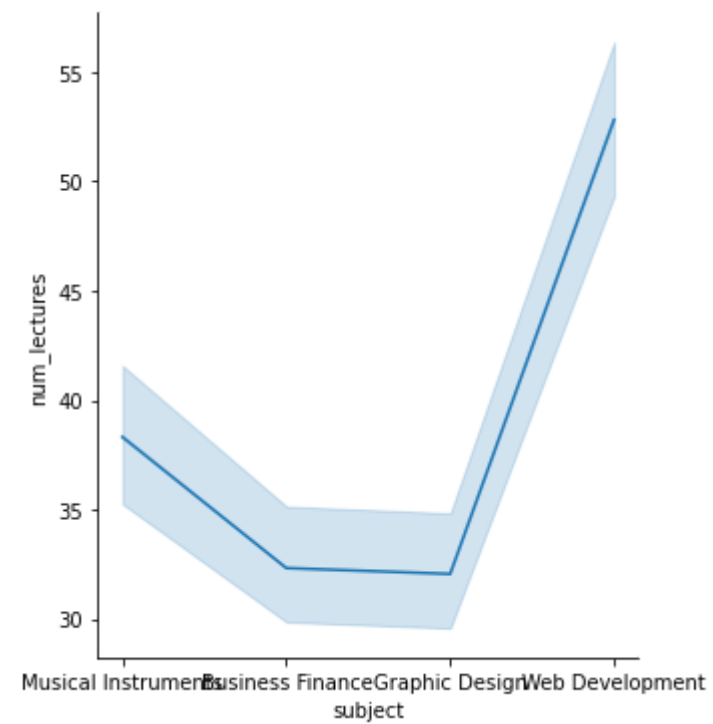
```
<_array_function__ internals>:5: RuntimeWarning: Converting input from bool to <class 'numpy.uint8'> for compatibility.  
<_array_function__ internals>:5: RuntimeWarning: Converting input from bool to <class 'numpy.uint8'> for compatibility.
```

```
Out[9]: <seaborn.axisgrid.PairGrid at 0x13af62e9e20>
```



```
In [11]: sb.relplot(x="subject",y="num_lectures",kind="line",data=d)
```

```
Out[11]: <seaborn.axisgrid.FacetGrid at 0x13af6335970>
```



```
In [ ]:
```

