

udemy dataset

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
data = pd.read_csv("udemy_courses.csv")
data
```

Out[1]:

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	51	All Levels	1.5
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	274	All Levels	39.0
2	1006314	Financial Modeling for Business Analysts and C...	https://www.udemy.com/financial-modeling-for-b...	True	45	2174	74	51	Intermediate Level	2.5
3	1210588	Beginner to Pro - Financial Analysis in Excel ...	https://www.udemy.com/complete-excel-finance-c...	True	95	2451	11	36	All Levels	3.0
4	1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-pro...	True	200	1276	45	26	Intermediate Level	2.0
...
3673	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040	14	21	All Levels	2.0
3674	1088178	How To Design A WordPress Website With No Codi...	https://www.udemy.com/how-to-make-a-wordpress-...	True	25	306	3	42	Beginner Level	3.5
3675	635248	Learn and Build using Polymer	https://www.udemy.com/learn-and-build-using-po...	True	40	513	169	48	All Levels	3.5
3676	905096	CSS Animations: Create Amazing Effects on Your...	https://www.udemy.com/css-animations-create-am...	True	50	300	31	38	All Levels	3.0
3677	297602	Using MODX CMS to Build Websites: A Beginner's...	https://www.udemy.com/using-modx-cms-to-build-...	True	45	901	36	20	Beginner Level	2.0

3678 rows × 12 columns

```
In [2]: # published_timestamp couldnt read data inside it correctly
data.dtypes
```

```
Out[2]: course_id          int64
course_title       object
url               object
is_paid           bool
price            int64
num_subscribers   int64
num_reviews       int64
num_lectures      int64
level            object
content_duration  float64
published_timestamp object
subject          object
dtype: object
```

```
In [3]: # published_timestamp read data inside it correctly
data = pd.read_csv("udemy_courses.csv", parse_dates = ['published_timestamp'])
data.dtypes
```

```
Out[3]: course_id                int64
course_title              object
url                      object
is_paid                  bool
price                   int64
num_subscribers          int64
num_reviews              int64
num_lectures             int64
level                   object
content_duration         float64
published_timestamp      datetime64[ns, UTC]
subject                 object
dtype: object
```

```
In [4]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3678 entries, 0 to 3677
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   course_id             3678 non-null  int64  
 1   course_title          3678 non-null  object  
 2   url                   3678 non-null  object  
 3   is_paid               3678 non-null  bool    
 4   price                 3678 non-null  int64  
 5   num_subscribers       3678 non-null  int64  
 6   num_reviews           3678 non-null  int64  
 7   num_lectures          3678 non-null  int64  
 8   level                 3678 non-null  object  
 9   content_duration      3678 non-null  float64 
10   published_timestamp   3678 non-null  datetime64[ns, UTC]
11   subject               3678 non-null  object  
dtypes: bool(1), datetime64[ns, UTC](1), float64(1), int64(5), object(4)
memory usage: 319.8+ KB
```

check for duplicate data and drop them

```
In [5]: duplicated = data.duplicated().any()
duplicated
```

```
Out[5]: True
```

```
In [6]: data = data.drop_duplicates()
data
```

Out[6]:

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	51	All Levels	1.5
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	274	All Levels	39.0
2	1006314	Financial Modeling for Business Analysts and C...	https://www.udemy.com/financial-modeling-for-b...	True	45	2174	74	51	Intermediate Level	2.5
3	1210588	Beginner to Pro - Financial Analysis in Excel ...	https://www.udemy.com/complete-excel-finance-c...	True	95	2451	11	36	All Levels	3.0
4	1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-pro...	True	200	1276	45	26	Intermediate Level	2.0
...
3673	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040	14	21	All Levels	2.0
3674	1088178	How To Design A WordPress Website With No Codi...	https://www.udemy.com/how-to-make-a-wordpress-...	True	25	306	3	42	Beginner Level	3.5
3675	635248	Learn and Build using Polymer	https://www.udemy.com/learn-and-build-using-po...	True	40	513	169	48	All Levels	3.5
3676	905096	CSS Animations: Create Amazing Effects on Your...	https://www.udemy.com/css-animations-create-am...	True	50	300	31	38	All Levels	3.0
3677	297602	Using MODX CMS to Build Websites: A Beginner's...	https://www.udemy.com/using-modx-cms-to-build-...	True	45	901	36	20	Beginner Level	2.0

3672 rows × 12 columns

```
In [7]: duplicated = data.duplicated().any()
duplicated
```

Out[7]: False

Which levels do Udemy courses provide courses for?

```
In [8]: data.columns
```

```
Out[8]: Index(['course_id', 'course_title', 'url', 'is_paid', 'price',
              'num_subscribers', 'num_reviews', 'num_lectures', 'level',
              'content_duration', 'published_timestamp', 'subject'],
              dtype='object')
```

```
In [9]: data['level'].value_counts()
```

```
Out[9]: level
All Levels      1925
Beginner Level  1268
Intermediate Level  421
Expert Level     58
Name: count, dtype: int64
```

how many paid and free courses in udemy

```
In [10]: data.columns
```

```
Out[10]: Index(['course_id', 'course_title', 'url', 'is_paid', 'price',  
             'num_subscribers', 'num_reviews', 'num_lectures', 'level',  
             'content_duration', 'published_timestamp', 'subject'],  
            dtype='object')
```

```
In [11]: data['is_paid'].value_counts()
```

```
Out[11]: is_paid  
True      3362  
False     310  
Name: count, dtype: int64
```

which course has more lectures (free or paid)?

```
In [13]: data.groupby(['is_paid']).max()
```

```
Out[13]:
```

	course_id	course_title	url	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp
is_paid										
False	1268616	财务分析与估值建模入门	https://www.udemy.com/zprzqgfl/	0	268923	8629	145	Intermediate Level	21.0	1
True	1282064	7日でマスター ビジネス向け Adobe Illustrator どきどきセミナー	https://www.udemy.com/zombie-apocalypse-photos...	200	121584	27445	779	Intermediate Level	78.5	2

```
In [14]: data.groupby(['is_paid']).max()['num_lectures']
```

```
Out[14]: is_paid  
False    145  
True     779  
Name: num_lectures, dtype: int64
```

او

```
In [25]: data['is_paid'] = data['is_paid'].replace({True: 'Paid', False: 'Free'})  
data.groupby(['is_paid']).max()['num_lectures']
```

```
Out[25]: is_paid  
Free     145  
Paid     779  
Name: num_lectures, dtype: int64
```

which courses have a higher number of subscribers free or paid ?

```
In [15]: result = data[['course_title', 'num_subscribers']].sort_values('num_subscribers', ascending=False)  
result
```

```
Out[15]:
```

	course_title	num_subscribers
2827	Learn HTML5 Programming From Scratch	268923
3032	Coding for Entrepreneurs Basic	161029
3230	The Web Developer Bootcamp	121584
2783	Build Your First Website in 1 Week with HTML5 ...	120291
3232	The Complete Web Developer Course 2.0	114512
...
1234	MicroStation - Células	0
886	Stop Creditors from Harassing you and Avoid Ba...	0
887	Learn to trade using technical analysis	0
888	How to Invest Using ETFs from Former Wall Stre...	0
777	Kickstarter success in 5 easy steps	0

3672 rows × 2 columns

```
In [16]: data[data['num_subscribers'].max() == data['num_subscribers']][['course_title', 'num_subscribers']]
```

```
Out[16]:
```

	course_title	num_subscribers
2827	Learn HTML5 Programming From Scratch	268923

which level has the highest number of subscribers

```
In [17]: result = data[['level', 'num_subscribers']].sort_values('num_subscribers', ascending=False)  
result
```

```
Out[17]:
```

	level	num_subscribers
2827	All Levels	268923
3032	Beginner Level	161029
3230	All Levels	121584
2783	Beginner Level	120291
3232	All Levels	114512
...
1234	Beginner Level	0
886	Beginner Level	0
887	Intermediate Level	0
888	All Levels	0
777	All Levels	0

3672 rows × 2 columns

find the courses which having the highest number of reviews

```
In [18]: result = data[['course_title', 'num_reviews']].sort_values('num_reviews', ascending=False)  
result
```

```
Out[18]:
```

	course_title	num_reviews
3230	The Web Developer Bootcamp	27445
3232	The Complete Web Developer Course 2.0	22412
3204	Angular 4 (formerly Angular 2) - The Complete ...	19649
3247	JavaScript: Understanding the Weird Parts	16976
3254	Modern React with Redux	15117
...
1416	Design Tickets and Vouchers in Photoshop and I...	0
514	How to invest in Russian equity market	0
515	Handle difficult questions the bank manager wi...	0
516	Credit Control Procedures	0
2157	Learn Lion (Saroo) Soundtrack On The Piano By Ear	0

3672 rows × 2 columns

```
In [19]: data[data['num_reviews'].max() == data['num_reviews']][['course_title', 'num_reviews']]
```

```
Out[19]:
```

	course_title	num_reviews
3230	The Web Developer Bootcamp	27445

find total number of courses related to python

```
In [20]: data[data['course_title'].str.contains('python',case=False)]
```

```
Out[20]:
```

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_dur
14	1196544	Python Algo Trading: Sentiment Trading with News	https://www.udemy.com/hedge-fund-strategy-trad...	True	200	294	19	42	All Levels	
30	1170894	Python Algo Stock Trading: Automate Your Trading!	https://www.udemy.com/algorithmic-stock-tradin...	True	95	1165	21	41	Beginner Level	
41	1035472	Python for Finance: Investment Fundamentals & ...	https://www.udemy.com/python-for-finance-inves...	True	195	3811	278	103	All Levels	
147	1070886	Python Algo Trading: FX Trading with	https://www.udemy.com/python-algo-trading-fx-t...	True	200	453	42	33	Intermediate Level	

```
In [21]: len(data[data['course_title'].str.contains('python',case=False)])
```

```
Out[21]: 29
```

display 10 most popular python courses per number of subscribers

```
In [22]: data[data['course_title'].str.contains('python',case=False)]
```

536	529828	Trading & Investing	https://www.udemy.com/python-for-trading-inves...	True	95	638	25	36	All Levels	
762	1088656	Quantitative Trading Analysis with Python	https://www.udemy.com/quantitative-trading-ana...	True	50	256	17	49	All Levels	
863	902888	Investment Portfolio Analysis with Python	https://www.udemy.com/investment-portfolio-ana...	True	50	209	13	37	All Levels	
1682	546848	Learn to code in Python and learn Adobe Photos...	https://www.udemy.com/learn-to-code-in-python-...	True	50	1132	1	29	All Levels	
2497	16646	Web Programming with Python	https://www.udemy.com/web-programming-with-pyt...	True	50	35267	217	53	All Levels	
----	-----	Learn Python and Django	https://www.udemy.com/learn-	---	---	-----	---	---	---	

```
In [23]: data[data['course_title'].str.contains('python',case=False)].sort_values(by="num_subscribers",ascending=False)[["course_title","num_subscribers"]].head(10)
```

```
Out[23]:
```

	course_title	num_subscribers
2497	Web Programming with Python	35267
2570	Coding for Entrepreneurs: Learn Python, Django...	23412
2528	Learn Python and Django: Payment Processing	17714
3200	Python and Django Full Stack Web Developer Boo...	11832
3507	Fun and creative web engineering with Python a...	10917
2553	The Complete Ethical Hacking Course 2.0: Pytho...	7827
3138	Complete Python Web Course: Build 8 Python Web...	7489
3197	Try Django 1.9 Build a Blog and Learn Python...	7407
2681	Python for Beginners: Python Programming Langu...	6153
3203	REST APIs with Flask and Python	5151

```
In [ ]:
```

```
In [ ]:
```