```python
import pandas as pd
df = pd.read_csv(r'C:\Users\PC-chetan\Downloads\7. Udemy Courses.csv')
df
```

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Le
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermed Le
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermed Le
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermed Le
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermed Le
...	...	...	...	...	...	...	...	
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Le
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Begir Le
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Le
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Le
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Le

3682 rows × 11 columns

```python
df.head(3)
```

Out[15]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Levels
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermediate Level
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermediate Level

In [20]:
```python
df.groupby('subject').course_title.count()
```

Out[20]:
```
subject
Business Finance       1199
Graphic Design          603
Musical Instruments     680
Web Development        1200
Name: course_title, dtype: int64
```

In [22]:
```python
df[df['is_paid'] == False].count()
```

Out[22]:
```
course_id            310
course_title         310
is_paid              310
price                310
num_subscribers      310
num_reviews          310
num_lectures         310
level                310
content_duration     310
published_timestamp  310
subject              310
dtype: int64
```

In [25]:
```python
df.subject.unique()
```

Out[25]:
```
array(['Musical Instruments', 'Business Finance', 'Graphic Design',
       'Web Development'], dtype=object)
```

In [26]:
```python
df.subject.value_counts()
```

Out[26]:
```
Web Development        1200
Business Finance       1199
Musical Instruments     680
Graphic Design          603
Name: subject, dtype: int64
```

In [27]:
```python
df.sort_values('num_subscribers', ascending=False)
```

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	l
2230	41295	Learn HTML5 Programming From Scratch	False	Free	268923	8629	45	All Le
776	59014	Coding for Entrepreneurs Basic	False	Free	161029	279	27	Begi L
3385	625204	The Web Developer Bootcamp	True	200	121584	27445	342	All Le
640	173548	Build Your First Website in 1 Week with HTML5 ...	False	Free	120291	5924	30	Begi L
3316	764164	The Complete Web Developer Course 2.0	True	200	114512	22412	304	All Le
...	...	...	...	...	...	...	...	
2290	1223240	Learn Pirates of the Caribbean by Ear on the P...	True	20	0	0	6	All Le
1267	1258666	Financial Statement Auditing Cycles	True	50	0	0	9	Intermed L
2070	1215926	Kickstarter success in 5 easy steps	True	20	0	0	12	All Le
1931	1247992	Introduction to Project Management for Finance...	True	50	0	0	9	Begi L
3165	1087466	Stop Creditors from Harassing you and Avoid Ba...	True	20	0	0	7	Begi L

3682 rows × 11 columns

In [28]:
```python
df.sort_values('num_subscribers', ascending=True)
```

Out[28]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	l
649	1233314	Building a Balanced Scorecard	True	50	0	0	11	Intermed L
3259	1232282	The Cash Flow Statement - An Introduction	True	50	0	0	10	Begi L
3261	1275872	The Complete Adobe Spark Course	True	150	0	0	29	All Le

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	l
910	726314	Create Beautiful Image Maps for Your Website	True	20	0	0	8	Intermed L
3328	715476	The Cuckoo Strategy to get European Funding	True	160	0	0	6	All Le
...	...	...	...	...	...	...	...	
3316	764164	The Complete Web Developer Course 2.0	True	200	114512	22412	304	All Le
640	173548	Build Your First Website in 1 Week with HTML5 ...	False	Free	120291	5924	30	Begi L
3385	625204	The Web Developer Bootcamp	True	200	121584	27445	342	All Le
776	59014	Coding for Entrepreneurs Basic	False	Free	161029	279	27	Begi L
2230	41295	Learn HTML5 Programming From Scratch	False	Free	268923	8629	45	All Le

3682 rows × 11 columns

In [37]:
```python
df[(df.subject == 'Graphic Design') & (df.price < '100')]
```

Out[37]:

course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level	conten

In [39]:
```python
df.course_title.unique()
```

Out[39]:
```
array(['#1 Piano Hand Coordination: Play 10th Ballad in Eb Key songs',
       '#10 Hand Coordination - Transfer Chord Ballad 9 - C & Bb Key',
       '#12 Hand Coordination: Let your Hands dance with Ballad 9', ...,
       'Zend Framework 2: Learn the PHP framework ZF2 from scratch',
       "Zoho Books Gestion Financière d'Entreprise pas à pas",
       'Zombie Apocalypse Photoshop Actions'], dtype=object)
```

In [41]:
```python
df.published_timestamp = pd.to_datetime(df.published_timestamp)
```

In [42]:
```python
df['year'] = df.published_timestamp.dt.year
```

In [44]:
```python
df.head(5)
```

Out[44]:

course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	level
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Levels
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermediate Level
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermediate Level
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermediate Level
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermediate Level

In [49]:
```python
df[df['year'] == 2015]
```

Out[49]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le...
10	591880	1 - Concepts of Statistics For Beginners Step ...	True	200	273	4	15	Begin... Le...
16	302450	10 Numbers Every Business Owner Should Know	True	20	13	1	9	All Lev...
19	384928	101 Blues riffs - learn how the harmonica supe...	True	200	1350	65	55	Intermedi... Le...
23	550842	16 Guitar Chords to Jam With - (Beginner - Int...	True	20	1224	19	20	Begin... Le...
26	486240	2 Easy Steps To Investment And Avoiding Traps	True	20	828	1	20	All Lev...
...	...	...	...	...	...	...	...	
3675	551170	Your First Course on Piano	True	70	2002	34	33	Begin... Le...

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
3676	498488	Your First Successful Forex Trades - With Case...	True	200	1079	34	16	All Lev
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Lev
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Begin Le
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Lev

1014 rows × 12 columns

In [51]:
```python
df[df.course_title.str.contains('Python')]
```

Out[51]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
216	599504	Advanced Scalable Python Web Development Using...	True	120	1299	56	71	Interme
777	47963	Coding for Entrepreneurs: Learn Python, Django...	True	195	23412	799	251	All L
814	631128	Complete Python Web Course: Build 8 Python Web...	True	110	7489	941	173	All L
842	186096	Core: A Web App Reference Guide for Django, Py...	True	195	2497	98	154	All L
1400	394832	Fun and creative web engineering with Python a...	False	Free	10917	319	25	All L
1932	1038538	Introduction to QGIS Python Programming	True	85	197	26	28	Beg
1969	902888	Investment Portfolio Analysis with Python	True	50	209	13	37	All L

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
2294	391546	Learn Python and Django: Payment Processing	True	70	17714	198	23	All L
2295	1063722	Learn Python Django - A Hands-On Course	True	50	1339	21	18	Beg
2341	546848	Learn to code in Python and learn Adobe Photos...	True	50	1132	1	29	All L
2906	368340	Professional Python Web Development Using Flask	True	120	3420	489	102	Beg
2908	1035940	Professional RESTful API Design using Python F...	True	120	578	25	36	Interme
2929	270808	Projects in Django and Python	True	60	1764	53	28	All L
2952	1170894	Python Algo Stock Trading: Automate Your Trading!	True	95	1165	21	41	Beg
2953	1070886	Python Algo Trading: FX Trading with Oanda	True	200	453	42	33	Interme
2954	1196544	Python Algo Trading: Sentiment Trading with News	True	200	294	19	42	All L
2955	822444	Python and Django Full Stack Web Developer Boo...	True	200	11832	1883	191	All L
2956	477702	Python for Beginners: Python Programming Langu...	True	150	6153	125	84	Beg
2957	1035472	Python for Finance: Investment Fundamentals & ...	True	195	3811	278	103	All L
2958	529828	Python for Trading & Investing	True	95	638	25	36	All L
2959	70640	Python Programming: Create an Digital Marketpl...	True	195	4198	145	161	All L

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
2960	523312	Python Web Programming	True	100	1020	46	60	Beg
2962	1088656	Quantitative Trading Analysis with Python	True	50	256	17	49	All L
3013	970600	REST APIs with Flask and Python	True	110	5151	737	115	Interme
3154	815482	Stock Technical Analysis with Python	True	50	409	35	46	All L
3272	938560	The Complete Ethical Hacking Course 2.0: Pytho...	True	195	7827	268	66	All L
3460	76052	Try Django 1.9	Build a Blog and Learn Python...	True	50	7407	172	147	All L
3573	16646	Web Programming with Python	True	50	35267	217	53	All L
3574	574082	Web Scraping with Python, Ruby & import. io	True	75	973	50	46	All L

In [53]:
```python
df.groupby('level')['num_subscribers'].max()
```

Out[53]:
```
level
All Levels            268923
Beginner Level        161029
Expert Level            5172
Intermediate Level     29167
Name: num_subscribers, dtype: int64
```

In [61]:
```python
df.groupby('subject').is_paid.value_counts()
```

Out[61]:
```
subject              is_paid
Business Finance     True       1103
                     False        96
Graphic Design       True        568
                     False        35
Musical Instruments  True        634
                     False        46
Web Development       True       1067
                     False       133
Name: is_paid, dtype: int64
```

In [63]:
```python
df.groupby('is_paid').num_subscribers.sum()
```

Out[63]:
```
is_paid
False    3575525
True     8185631
Name: num_subscribers, dtype: int64
```

```
In [64]:    df.head(30)
```

Out[64]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	lev
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Leve
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermedia Lev
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermedia Lev
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermedia Lev
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermedia Lev
5	1122832	#6 Piano Hand Coordination: Play Open 10 Balla...	True	65	2015	3	21	Intermedia Lev
6	1123184	#7 Piano Hand Coordination: Play Piano Runs in...	True	65	2015	2	25	Intermedia Lev
7	768028	* An Integrated Approach to the Fundamentals o...	True	200	134	14	123	All Leve
8	206592	:::Beginners Guide To Creating Amazing Images ...	True	25	72	5	17	Beginn Lev
9	738910	[Value Investing] Where White People Keep Thei...	True	200	2715	70	20	Expert Lev
10	591880	1 - Concepts of Statistics For Beginners Step ...	True	200	273	4	15	Beginn Lev
11	64605	1 Hour CSS	True	100	18324	333	16	Beginn Lev
12	64422	1 Hour HTML	True	200	16212	514	20	Beginn Lev

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	lev
13	63612	1 Hour JavaScript	True	200	22999	602	18	Beginn Lev
14	72262	1 hour jQuery	True	100	5456	133	25	Beginn Lev
15	145220	1. Principles of Simple Interest	True	20	1074	6	13	All Leve
16	302450	10 Numbers Every Business Owner Should Know	True	20	13	1	9	All Leve
17	157336	10. Bonds and Bond Pricing	True	20	308	4	10	All Leve
18	897162	101 Awesome Rockabilly Guitar Licks	True	40	407	64	33	Intermedia Lev
19	384928	101 Blues riffs - learn how the harmonica supe...	True	200	1350	65	55	Intermedia Lev
20	1165268	15 Motion Graphic Elements in After Effect (S...	True	30	8	0	20	Intermedia Lev
21	315716	15 Mandamientos para ganar dinero en Bolsa	True	20	45	2	49	All Leve
22	759972	150 Rock Guitar Licks You MUST Know	True	45	252	14	21	All Leve
23	550842	16 Guitar Chords to Jam With - (Beginner - Int...	True	20	1224	19	20	Beginn Lev
24	975916	17 Complete JavaScript projects explained st...	True	185	327	26	106	Beginn Lev
25	638966	188% Profit in 1Year - Forex ALGO Robot Tradin...	True	120	357	71	20	All Leve
26	486240	2 Easy Steps To Investment And Avoiding Traps	True	20	828	1	20	All Leve
27	146454	2. Principles of Simple Interest - advanced pr...	True	20	10	1	12	Intermedia Lev

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	lev
28	54246	3 documentos clave para administrar tu empresa...	True	45	40	6	18	Intermedia Lev
29	1089610	3 Little Pigs: A Multiple Timeframe Forex Trad...	True	20	18	0	6	All Leve

In [65]:
```python
df.content_duration.unique()
```

Out[65]:
```
array(['1.5 hours', '1 hour', '37 mins', '44 mins', '5.5 hours',
       '4 hours', '31 mins', '2.5 hours', '6.5 hours', '3 hours',
       '8.5 hours', '4.5 hours', '9.5 hours', '2 hours', '3.5 hours',
       '9 hours', '5 hours', '36 mins', '16 hours', '32 mins', '14 hours',
       '35 mins', '6 hours', '41 mins', '7.5 hours', '14.5 hours',
       '24.5 hours', '33 mins', '29 hours', '13.5 hours', '42 mins',
       '33 hours', '12.5 hours', '8 hours', '7 hours', '34 mins',
       '11.5 hours', '62 hours', '12 hours', '39 mins', '19 mins',
       '15.5 hours', '10.5 hours', '21 hours', '68.5 hours', '19.5 hours',
       '20.5 hours', '22 hours', '10 hours', '38 mins', '43 mins',
       '40 mins', '24 hours', '13 hours', '31.5 hours', '44.5 hours',
       '17.5 hours', '48.5 hours', '60 hours', '11 hours', '27.5 hours',
       '18.5 hours', '25.5 hours', '22.5 hours', '16.5 hours', '30 mins',
       '57 hours', '45 hours', '20 hours', '39 hours', '34 hours',
       '25 hours', '26 hours', '29.5 hours', '15 hours', '37.5 hours',
       '71.5 hours', '47 hours', '18 hours', '38 hours', '28.5 hours',
       '8 mins', '32.5 hours', '31 hours', '17 hours', '27 mins',
       '29 mins', '46.5 hours', '43.5 hours', '66.5 hours', '26.5 hours',
       '0', '30 hours', '30.5 hours', '19 hours', '23 hours',
       '23.5 hours', '24 questions', '218 questions', '18 questions',
       '82 questions', '78.5 hours', '21.5 hours', '51 hours',
       '76.5 hours', '42 hours', '43 hours', '70 hours', '28 mins'],
      dtype=object)
```

In [69]:
```python
df.describe()
```

Out[69]:
	course_id	num_subscribers	num_reviews	num_lectures	year
count	3.682000e+03	3682.000000	3682.000000	3682.000000	3682.000000
mean	6.766121e+05	3194.230310	156.093156	40.065182	2015.432917
std	3.436355e+05	9499.378361	934.957204	50.373299	1.185800
min	8.324000e+03	0.000000	0.000000	0.000000	2011.000000
25%	4.078430e+05	110.250000	4.000000	15.000000	2015.000000
50%	6.885580e+05	911.500000	18.000000	25.000000	2016.000000
75%	9.617515e+05	2540.250000	67.000000	45.000000	2016.000000
max	1.282064e+06	268923.000000	27445.000000	779.000000	2017.000000

In [74]:
```python
df1 = df.replace({"content_duration" :'[A-Za-z]'},'', regex =True)
```

In [75]:
```python
df1
```

	ice	num_subscribers	num_reviews	num_lectures	le

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Le
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermed Le
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermed Le
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermed Le
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermed Le
...	...	...	...	...	...	...	...	
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Le
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Begir Le
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Le
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Le
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Le

3682 rows × 12 columns

In [84]: `df.dtypes`

Out[84]:
```
course_id        int64
course_title    object
is_paid           bool
price           object
```

```
num_subscribers                    int64
num_reviews                        int64
num_lectures                       int64
level                             object
content_duration                  object
published_timestamp    datetime64[ns, UTC]
subject                           object
year                               int64
dtype: object
```

In [82]:
```python
df1.convert_dtypes()
```

Out[82]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	All Le
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Intermed Le
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Intermed Le
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Intermed Le
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Intermed Le
...	...	...	...	...	...	...	...	
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Le
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Begin Le
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Le
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	All Le

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	le
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	All Le

3682 rows × 12 columns

In [85]:
```python
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3682 entries, 0 to 3681
Data columns (total 12 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   datetime64[ns, UTC]
 10  subject            3682 non-null   object
 11  year               3682 non-null   int64
dtypes: bool(1), datetime64[ns, UTC](1), int64(5), object(5)
memory usage: 320.1+ KB
```

In [87]:
```python
df1['content_duration'] = df1['content_duration'].astype(float)
```

In [89]:
```python
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3682 entries, 0 to 3681
Data columns (total 12 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   float64
 9   published_timestamp  3682 non-null   datetime64[ns, UTC]
 10  subject            3682 non-null   object
 11  year               3682 non-null   int64
dtypes: bool(1), datetime64[ns, UTC](1), float64(1), int64(5), object(4)
memory usage: 320.1+ KB
```

In [91]:
```python
df1.content_duration.mean()
```

Out[91]:
```
7.3738457360130365
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js