In [14]:

import pandas as pd
df = nd road csy(r)

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

d.

14]:		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 L
	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 r	ows × 11 c	columns						

```
course_id course_title is_paid price num_subscribers num_reviews num_lectures
Out[15]:
                                                                                                       level
                           #1 Piano
                              Hand
          0
                                                                              18
                                                                                            68
               288942 Coordination:
                                       True
                                               35
                                                              3137
                                                                                                   All Levels
                           Play 10th
                           Ballad i...
                          #10 Hand
                        Coordination
                                                                                                Intermediate
          1
              1170074
                           - Transfer
                                       True
                                               75
                                                              1593
                                                                               1
                                                                                            41
                                                                                                       Level
                        Chord Ballad
                          #12 Hand
                       Coordination:
                                                                                                Intermediate
          2
                                                                               1
              1193886
                                               75
                                                               482
                            Let your
                                       True
                                                                                                       Level
                        Hands dance
                               wi...
In [20]:
           df.groupby('subject').course_title.count()
          subject
Out[20]:
          Business Finance
                                   1199
          Graphic Design
                                    603
          Musical Instruments
                                    680
          Web Development
                                   1200
          Name: course title, dtype: int64
In [22]:
          df[df['is paid'] == False].count()
                                   310
          course_id
Out[22]:
          course title
                                   310
                                   310
          is paid
          price
                                   310
          num subscribers
                                   310
          num reviews
                                   310
          num lectures
                                   310
          level
                                   310
          content_duration
                                   310
          published timestamp
                                   310
                                   310
          subject
          dtype: int64
In [25]:
           df.subject.unique()
          array(['Musical Instruments', 'Business Finance', 'Graphic Design',
Out[25]:
                  'Web Development'], dtype=object)
In [26]:
          df.subject.value counts()
                                   1200
          Web Development
Out[26]:
          Business Finance
                                   1199
          Musical Instruments
                                    680
          Graphic Design
                                    603
          Name: subject, dtype: int64
In [27]:
           df.sort values('num subscribers', ascending=False)
                                                  rice num_subscribers num_reviews num_lectures
```

df.head(3)

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

		course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	1
:	2230	41295	Learn HTML5 Programming From Scratch	False	Free	268923	8629	45	All L€
	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 L€
	640	173548	Build Your First Website in 1 Week with HTML5	False	Free	120291	5924	30	Begi L
3	3316	764164	The Complete Web Developer Course 2.0	True	200	114512	22412	304	All Le
2	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
7	2070	1215926	Kickstarter success in 5 easy steps	True	20	0	0	12	All L€
:	1931	1247992	Introduction to Project Management for Finance	True	50	0	0	9	Begi L
3	3165	1087466	Stop Creditors from Harassing you and Avoid Ba	True	20	0	0	7	Begi l

3682 rows × 11 columns

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

Course

course_title is_paid price num_subscribers num_reviews num_lectures course_id Out[28]: Building a Intermed 11 649 Balanced True 50 0 0 1233314 Scorecard The Cash Flow

Begi Statement -3259 1232282 True 50 0 0 10 An Introduction The Complete Adobe Spark 3261 1275872 True 150 0 0 29 All L€

### State			_				-	_	_	
3328 715476 Strategy to get European Funding True 160 0 0 0 6 All Le		910	726314	Beautiful Image Maps for Your	True	20	0	0	8	Intermed L
3316 764164 The Complete Web Developer Course 2.0 114512 22412 304 All Let Course 2.0 114512 22412 304 All Let Course 2.0 173548 First Website in 1 Week with HTML5 False Free 120291 5924 30 Beging 1 1745		3328	715476	Strategy to get European	True	160	0	0	6	All L€
3316 764164										
640		3316	764164	Web Developer	True	200	114512	22412	304	All L€
3385 625204 Developer Bootcamp True 200 121584 27445 342 All Le Coding for Form Coding for Basic 230 41295 Programming Form Scratch 3682 rows × 11 columns In [37]: 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]: df.course_title.unique() array(['#1 Piano Hand Coordination: Play 10th Ballad in Eb Key songs', '#10 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]: df.published_timestamp = pd.to_datetime(df.published_timestamp) In [44]: df.head(5)		640	173548	First Website in 1 Week	False	Free	120291	5924	30	Begi l
776 59014 Entrepreneurs Basic False Free 161029 279 27 Beging Basic Part Basic Free Basic Programming False Free 268923 8629 45 All Lk 3682 rows × 11 columns In [37]: df[(df.subject == 'Graphic Design') & (df.price < '100')] Out[37]: course_id course_title is_paid price num_subscribers num_reviews num_lectures level contenum[39]: df.course_title.unique() 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',, 'Zoho Books Gestion Financière d'Entreprise pas à pas'', 'Zoho Books Gestion Financière d'Entreprise pas à pas'', 'Zombie Apocalypse Photoshop Actions'], dtype=object) In [41]: df.published_timestamp = pd.to_datetime(df.published_timestamp) In [42]: df.head(5)		3385	625204	Developer	True	200	121584	27445	342	All Le
2230 41295 Programming Found From Scratch Found From Scratch Found Scratch From Scr		776	59014	Entrepreneurs	False	Free	161029	279	27	Begi L
<pre>In [37]: df[(df.subject == 'Graphic Design') & (df.price < '100')] Out[37]:</pre>		2230	41295	Programming	False	Free	268923	8629	45	All Le
<pre>In [37]: df[(df.subject == 'Graphic Design') & (df.price < '100')] Out[37]:</pre>		3682 ro	NS × 11 C	rolumne						
Out[37]: course_id course_title is_paid price num_subscribers num_reviews num_lectures level conten In [39]: df.course_title.unique() Out[39]: array(['#1 Piano Hand Coordination: Play 10th Ballad in Eb Key songs',		3002 10	//3 // II C	olannis						
<pre>In [39]: df.course_title.unique() Out[39]: array(['#1 Piano Hand Coordination: Play 10th Ballad in Eb Key songs',</pre>	In [37]:	df[(df	.subject	== 'Graphic	Design') & (df	.price < '100')	1		
Out[39]: array(['#1 Piano Hand Coordination: Play 10th Ballad in Eb Key songs',	Out[37]:	cours	e_id cour	rse_title is_pa	id price	num_s	ubscribers num_	reviews num_	lectures level	conten
<pre>'#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]: df.published_timestamp = pd.to_datetime(df.published_timestamp) In [42]: df['year'] = df.published_timestamp.dt.year In [44]: df.head(5)</pre>	In [39]:	df.cou	rse_titl	e.unique()						
<pre>In [42]: df['year'] = df.published_timestamp.dt.year</pre> In [44]: df.head(5)	Out[39]:	array(['#10 Han '#12 Han 'Zend Fr "Zoho Bo	d Coordination Coordination Coordination Coordination Communication Francisco Francisco Coordination Coordination Francisco Coordination Coordinatio	on - Tra on: Let earn the inanciè	nsfer C your Ha PHP fr re d'En	hord Ballad 9 - nds dance with amework ZF2 fro treprise pas à	C & Bb Key', Ballad 9', om scratch', pas",		
In [44]: df.head(5)	In [41]:	df.pub	lished_t	imestamp = p	od.to_da	tetime(df.published_ti	mestamp)		
di.nead(3)	In [42]:	df['ye	ar'] = d	f.published_t	:imestam	p.dt.yea	ar			
Out[44]: course_id course_title is_paid price num_subscribers num_reviews num_lectures level	In [44]:	df.hea	d(5)							
	Out[44]:	cour	se_id co	urse_title is_p	aid pric	e num_	subscribers nur	n_reviews nur	n_lectures	level

course_id course_title is_paid price num_subscribers num_reviews num_lectures

I

10 Seginners Step Step		С	ourse_id	course_tit	le is_	paid	price	nu	m_subscribers	num_reviev	vs ni	um_lectures		level
1		0	288942	Har Coordinatio Play 10	nd n: th	True	35		3137	1	18	68		All Levels
2		1	1170074	Coordination - Transf	on er ad	True	75		1593		1	41	Inte	
Numbers Stephen Step		2	1193886	Coordinatio Let yo Hands dan	n: ur ce	True	75		482		1	47	Inte	
Manual M		3	1116700	Har Coordinatio Fun Piar	nd n: no	True	75		850		3	43	Inte	
		4	1120410	Har Coordinatio Piano Runs	nd n: in	True	75		940		3	32	Inte	
10 591880 For Beginners Step True 200 273 4 15 Beginners Step	In [49]:	df[df['year	'] == 2015	5]									
10 591880 of Statistics For Beginners Step True 200 273 4 15 Begin Les 16 302450 Numbers Every Business Owner Should Know True 20 13 1 9 All Les 19 384928 101 Blues riffs - learn how the harmonica supe True 200 1350 65 55 Intermedi Les 23 550842 16 Guitar Chords to Jam With - (Beginner - Int True 20 1224 19 20 Begin Les 26 486240 True 20 828 1 20 All Les 3675 551170 Your First Course on Plano True 70 2002 34 33 Begin Les	Out[49]:		course	_id course	_title	is_pai	id pri	ce	num_subscribe	ers num_re	views	num_lectur	es	le
16 302450 Every Business Owner Should Know True 20 13 1 9 All Lev Owner Should Know 19 384928 101 Blues riffs - learn harmonica supe True 200 1350 65 55 Intermedia Le 23 550842 16 Guitar Chords to Jam With - (Beginner - Int) True 20 1224 19 20 Begin Le 26 486240 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		10) 5918	of Sta 380 Begi	For nners	Tru	ue 2	00	2	273	4		15	
19 384928 how the harmonica supe True 200 1350 65 55 Intermediate 23 550842 loginar - Int True loginar - Int 20 loginar - Int 1224 loginar - Int 19 loginar - Int 20 loginar - Int 26 486240 loginary logina		10	5 3024	150 Bus	Every siness Owner	Trı	ıe	20		13	1		9	All Lev
23 550842 Chords to Jam With - (Beginner - Int) True 20 1224 19 20 Begin Le 26 486240 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		19	9 3849	riffs - 928 ho harm	learn w the ionica	Tru	ue 2	00	13	350	65		55	
26 486240 Investment And Avoiding Traps True 20 828 1 20 All Lev		23	3 5508	Cho 342 Jam	rds to With - nner -	Tru	ıe	20	12	224	19		20	
Your First 3675 551170 Course on True 70 2002 34 33 Begin Le		20	5 4862	240 Inves	To tment oiding	Tru	ıe	20	8	328	1		20	All Lev
3675 551170 Course on True 70 2002 34 33 Lε														
		367	5 5511		se on	Tru	ıe	70	20	002	34		33	

	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 L€
3679	631754	Zend Framework 2: Learn the PHP framework ZF2	True	40	723	130	37	All Lev

1014 rows × 12 columns

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

	_	_			•				
Out[51]:		course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
	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
nJax]/jax/	output/Comm	onHTML/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]:	df.gr	oupby('le	vel')['num_su	bscriber	s'].ma	ax()			
Out[53]:	Expert Interm	er Level Level ediate Le	268923 161029 5172 vel 29167 ribers, dtype						
In [61]:	df.gr	oupby(' <mark>sul</mark>	oject').is_pa	id.value	_count	cs()			
Out[61]:	subjec Busine	t ss Financ	is_paid e True False	1103 96					
	Graphi	c Design	True False	568 35	3				
	Musica	l Instrum		63 ⁴	1				
		velopment		1067 133	7				
In [63]:			_paid').num_s	uhscribe	are elin	n ()			
Out[63]:	is_pai False True	d 357552 818563	5			/			

Name: num subscribers dtype: int6/ Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

df.head(30)In [64]: course_title is_paid price num_subscribers num_reviews num_lectures course id Out[64]: lev #1 Piano Hand 0 288942 Coordination: True 35 3137 18 68 All Leve Play 10th Ballad i... #10 Hand Coordination -Intermedia 1170074 1 41 1 True 75 1593 Transfer Chord Lev Ballad ... #12 Hand Coordination: Intermedia 2 1193886 1 47 Let your 75 482 True Le۱ Hands dance wi... #4 Piano Hand Intermedia 3 1116700 Coordination: True 75 850 3 43 Lev Fun Piano Runs in ... #5 Piano Hand Intermedia 4 1120410 Coordination: True 75 940 3 Le۱ Piano Runs in 2 ... #6 Piano Hand Intermedia 5 3 21 1122832 Coordination: True 2015 65 Lev Play Open 10 Balla... #7 Piano Hand Intermedia 6 2 25 1123184 Coordination: True 65 2015 Le۱ Play Piano Runs in... * An Integrated Approach to 7 768028 True 200 134 14 123 All Leve the **Fundamentals** 0... :::Beginners Guide To Beginn 8 206592 25 72 5 17 Creating True Le۱ Amazing Images ... [Value Investing] 9 70 738910 Where White True 200 2715 20 Expert Lev People Keep Thei... 1 - Concepts Beginn of Statistics 10 591880 True 200 273 4 15 For Beginners Le۱ Step ... Beginn 11 64605 1 Hour CSS True 100 18324 333 16 Lev Beginn 12 64422 1 Hour HTML True 200 16212 514 20 Le۱

	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_title is_paid price num_subscribers num_reviews num_lectures
                                     course_id
                                                                                                                                                                                                                                                                              lev
                                                                3 documentos
                                                                                                                                                                                                                                                             Intermedia
                                                                        clave para
                          28
                                             54246
                                                                                                           True
                                                                                                                               45
                                                                                                                                                                            40
                                                                                                                                                                                                                                                    18
                                                                 administrar tu
                                                                                                                                                                                                                                                                              Le۱
                                                                        empresa...
                                                                3 Little Pigs: A
                                                                              Multiple
                                        1089610
                                                                                                                                                                                                                  0
                                                                                                                                                                                                                                                       6
                          29
                                                                                                          True
                                                                                                                               20
                                                                                                                                                                            18
                                                                                                                                                                                                                                                                     All Leve
                                                                        Timeframe
                                                                     Forex Trad...
In [65]:
                            df.content duration.unique()
                         Out[65]:
                                             '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', '30 mins', '46 5 hours', '47 hours', '48 finally 146 5 hours', '48 finally 146 5 hours', '49 finally 146 fina
                                              '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]:
                            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 4.000000 15.000000 2015.000000 110.250000 **50**% 6.885580e+05 911.500000 18.000000 25.000000 2016.000000 **75**% 9.617515e+05 2540.250000 67.000000 45.000000 2016.000000 268923.000000 27445.000000 max 1.282064e+06 779.000000 2017.000000

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

In [75]: df1

		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 r	ows × 12 c	columns						
:	df.dt	ypes							
]:	course course is_pa:	e_title			int6 objec boo	t il			

In [84]

Out[84]

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

num_subscribers	int64
num_reviews	int64
num_lectures	int64
level	object
content_duration	object
<pre>published_timestamp</pre>	<pre>datetime64[ns, UTC]</pre>
subject	object
year	int64
dtype: object	

In [82]: df1.convert_dtypes()

) p = 5 ()	convert_ut	
l€	num_lectures	num_reviews	num_subscribers	price	is_paid	course_title	course_id	
All Le	68	18	3137	35	True	#1 Piano Hand Coordination: Play 10th Ballad i	288942	0
Intermed Le	41	1	1593	75	True	#10 Hand Coordination - Transfer Chord Ballad 	1170074	1
Intermed Le	47	1	482	75	True	#12 Hand Coordination: Let your Hands dance wi	1193886	2
Intermed Lo	43	3	850	75	True	#4 Piano Hand Coordination: Fun Piano Runs in	1116700	3
Intermed L	32	3	940	75	True	#5 Piano Hand Coordination: Piano Runs in 2	1120410	4
All Le	36	29	1566	120	True	Your Own Site in 45 Min: The Complete Wordpres	328960	577
Begir L	22	12	1018	70	True	Your Second Course on Piano: Two Handed Playing	552700	678
All Le	37	130	723	40	True	Zend Framework 2: Learn the PHP framework ZF2	631754	679
All Le	33	0	229	Free	False	Zoho Books Gestion Financière d'Entreprise pas	1225194	680

```
Zombie
                            Apocalypse
          3681
                  964478
                                          True
                                                  50
                                                                   12
                                                                                  1
                                                                                               15
                                                                                                      All Le
                             Photoshop
                                Actions
         3682 \text{ rows} \times 12 \text{ columns}
In [85]:
          dfl.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 3682 entries, 0 to 3681
          Data columns (total 12 columns):
          #
               Column
                                     Non-Null Count
                                                      Dtype
          - - -
               -----
                                                       ----
           0
                                                       int64
               course id
                                     3682 non-null
           1
               course_title
                                     3682 non-null
                                                       object
           2
                                     3682 non-null
               is paid
                                                       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
                                     3682 non-null
               published timestamp
                                                       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]:
          df1['content duration'] = df1['content duration'].astype(float)
In [89]:
          df1.info()
          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 3682 entries, 0 to 3681
          Data columns (total 12 columns):
          #
               Column
                                     Non-Null Count
                                                      Dtype
          - - -
               _ _ _ _ _ _
                                                       _ _ _ _ _
           0
                                     3682 non-null
                                                       int64
               course id
           1
               course title
                                     3682 non-null
                                                       object
           2
                                     3682 non-null
               is paid
                                                       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
                                     3682 non-null
                                                       float64
               content duration
           9
               published timestamp
                                     3682 non-null
                                                       datetime64[ns, UTC]
           10
                                     3682 non-null
               subject
                                                       object
                                     3682 non-null
               year
                                                       int64
          dtypes: bool(1), datetime64[ns, UTC](1), float64(1), int64(5), object(4)
          memory usage: 320.1+ KB
In [91]:
          dfl.content duration.mean()
          7.3738457360130365
Out[91]:
```

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

course_id course_title is_paid price num_subscribers num_reviews num_lectures

lε