Udemy Course Data Analysis Using Python

The data Contain all course for all Subject.

The datset is available as a csv file. Download from Kaggle

We will analyze the data using the pandas data frame.

Import Library

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

Uploading Csv fle

```
In [2]:   df = pd.read_csv("C:\Users\Syed Arif\Downloads\7. Udemy Courses.csv")
```
```
  Input In [2]
    df = pd.read_csv("C:\Users\Syed Arif\Downloads\7. Udemy Courses.csv")
                                                                        ^
SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in posi
tion 2-3: truncated \UXXXXXXXX escape
```

Remove Unicode error

We are Written Small r before Quatation

```
In [3]:   df = pd.read_csv(r"C:\Users\Syed Arif\Downloads\7. Udemy Courses.csv")
```

In [4]: df

Out[4]:

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

3682 rows × 11 columns

Data Preprocessing

.head()

head is used show to the By default = 5 rows in the dataset

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

Out[5]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
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	Interme L
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Interme L
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Interme L
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Interme L

.tail()

tail is used to show rows by Descending order

In [6]: `df.tail()`

Out[6]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	l
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	Le
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Begir L
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	Le
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	Le
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	Le

.shape

It shoe the total no of rows & Column in the dataset

In [7]: `df.shape`

Out[7]: `(3682, 11)`

.Columns

It show the no of each Column

In [10]: `df.columns`

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

.dtypes

This Attribute show the data type of each column

```
In [11]:  df.dtypes
```

```
Out[11]:  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   object
          subject               object
          dtype: object
```

.unique()

In a column, It show the unique value of specific column.

```
In [12]:  df["num_subscribers"].unique()
```

```
Out[12]:  array([ 3137,  1593,   482, ..., 11892,   924,   723], dtype=int64)
```

nuique()

It will show the total no of unique value from whole data frame

```
In [13]:  df.nunique()
```

```
Out[13]:  course_id             3676
          course_title          3667
          is_paid                  2
          price                   38
          num_subscribers       2198
          num_reviews            511
          num_lectures           229
          level                    4
          content_duration       109
          published_timestamp   3676
          subject                  4
          dtype: int64
```

.describe()

It show the Count, mean , median etc

In [14]: `df.describe()`

Out[14]:

	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

.value_counts

It Shows all the unique values with their count

In [16]: `df["subject"].value_counts()`

Out[16]:
```
Web Development        1200
Business Finance       1199
Musical Instruments     680
Graphic Design          603
Name: subject, dtype: int64
```

Where all different subject for which Udemy is offering couses ?

In [17]: `df.subject.unique()`

Out[17]:
```
array(['Musical Instruments', 'Business Finance', 'Graphic Design',
       'Web Development'], dtype=object)
```

Which Subject has the maximum number of courses.

```
In [20]: df['subject'].value_counts()
```

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

Show all the courses which are Free of Cost

In [25]: `df[df.is_paid == False]`

Out[25]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
41	286070	5 lecciones que todo guitarrista debe tomar	False	Free	4452	263	14	Be
49	696630	7 Ways A Beginner Guitarist Can Sound Better, ...	False	Free	4529	193	7	Be
57	955914	A beginner`s guide to fingerpicking and strumm...	False	Free	3481	29	20	Be
67	270976	A how to guide in HTML	False	Free	7318	205	8	Be
74	1214144	¡Triunfar en La Bolsa de Valores No Requiere d...	False	Free	338	7	6	Be
...	...	...	...	...	...	...	...	
3620	552598	WordPress in under an hour - Create/manage you...	False	Free	6474	86	15	
3638	1068590	Wordpress Website Mastery 2017	False	Free	5181	64	22	Be
3652	752706	Write quicker HTML5 and CSS 3; productivity ha...	False	Free	13016	286	15	
3674	1257976	Your First 10 Guitar Lessons - Learn how to pl...	False	Free	924	4	18	Be
3680	1225194	Zoho Books Gestion Financière d'Entreprise pas...	False	Free	229	0	33	

310 rows × 11 columns

Show all the courses Which are Paid

In [26]:
```python
df[df.is_paid == True]
```

Out[26]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
0	288942	#1 Piano Hand Coordination: Play 10th Ballad i...	True	35	3137	18	68	A
1	1170074	#10 Hand Coordination - Transfer Chord Ballad ...	True	75	1593	1	41	Inter
2	1193886	#12 Hand Coordination: Let your Hands dance wi...	True	75	482	1	47	Inter
3	1116700	#4 Piano Hand Coordination: Fun Piano Runs in ...	True	75	850	3	43	Inter
4	1120410	#5 Piano Hand Coordination: Piano Runs in 2 ...	True	75	940	3	32	Inter
...	...	...	...	...	...	...	...	
3676	498488	Your First Successful Forex Trades - With Case...	True	200	1079	34	16	A
3677	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	A
3678	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	E
3679	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	A
3681	964478	Zombie Apocalypse Photoshop Actions	True	50	12	1	15	A

3372 rows × 11 columns

What are the Top Selling Courses

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

Out[27]:

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

3682 rows × 11 columns

Show all courses of Graphic desingner where the Price is below 100

```
In [45]:  df[(df.subject  == "Graphic Design") & (df.price == "100")]
```

Out[45]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
47	1269190	7 Secrets To Designing In Black and White	True	100	1000	1	12	
169	983444	Adobe Illustrator Vectors For Hand Letterers	True	100	23	2	9	
184	638040	Adobe Photoshop Essentials: Master Adobe Photo...	True	100	1373	5	63	
520	978462	Blender Modelling Series - Volume Two	True	100	1526	3	24	
699	757900	Canva : Book Cover Design	True	100	3377	20	15	
1029	1160004	Design a logo for FREE in Inkscape	True	100	1148	26	42	
1071	1161864	Diseño impreso - Preimpresión para offset en l...	True	100	33	9	29	
1444	490392	Getting Started with Illustrator CC	True	100	92	10	57	
1568	727510	How To Become An Etsy Wholesale Seller	True	100	1013	0	25	
1616	996868	How to Design a Logo in Adobe Illustrator	True	100	972	44	24	
1697	654038	How To Sell Your Art Online With ETSY	True	100	1338	38	35	
1773	1012262	Icon & Logo Symbol Design in Adobe Illustrator	True	100	1043	27	30	
1786	1246168	Illustrator para criação de peças gráficas, De...	True	100	33	3	41	

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	
1797	1266468	Indesign - Para quem quer trabalhar com Design...	True	100	27	7	12	Inte
2137	927186	Learn Blender 3D - The introduction course	True	100	739	12	15	
2481	1027080	Logo Portfolio in 1 Hour in Wordpress	True	100	734	9	36	
2491	771044	Make Money With Photoshop: Logos, Business Car...	True	100	32	4	21	
2576	482924	Mastering Retouching and Restoration (15 proje...	True	100	4927	15	22	
2704	950894	photoshop - CURSO DE PHOTOSHOP DO BÁSICO AO AV...	True	100	1110	183	77	
2706	1243016	Photoshop - Tratamento de pele profissional	True	100	50	5	16	Inte
2713	1147664	Photoshop CC Eğitim Seti (YENİ)	True	100	27	5	202	
3479	575040	Typography From A to Z	True	100	317	15	76	
3594	908664	Word Swag : Create Stunning Images with the Wo...	True	100	2416	19	14	

What are the Courses that are Published in 2015 ?

In [49]:
```python
df["published_timestamp"] = pd.to_datetime(df.published_timestamp)
df["Year"] = df["published_timestamp"].dt.year
df[df.Year == 2015]
```

Out[49]:

	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lectures	lev
10	591880	1 - Concepts of Statistics For Beginners Step ...	True	200	273	4	15	Beginn 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	Intermedia Le
23	550842	16 Guitar Chords to Jam With - (Beginner - Int...	True	20	1224	19	20	Beginn Le
26	486240	2 Easy Steps To Investment And Avoiding Traps	True	20	828	1	20	All Lev
...	...	...	...	...	...	...	...	
75	551170	Your First Course on Piano	True	70	2002	34	33	Beginn Le
76	498488	Your First Successful Forex Trades - With Case...	True	200	1079	34	16	All Lev
77	328960	Your Own Site in 45 Min: The Complete Wordpres...	True	120	1566	29	36	All Lev
78	552700	Your Second Course on Piano: Two Handed Playing	True	70	1018	12	22	Beginn Le
79	631754	Zend Framework 2: Learn the PHP framework ZF2 ...	True	40	723	130	37	All Lev

|4 rows × 12 columns

What are the Maximum Number of Subscriber For each level of Course ?

```
In [52]: df.groupby('level')["num_subscribers"].max()
         df.groupby("level").max()
```

Out[52]:

level	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lecture
All Levels	1277924	６時間でインターバンク市場を攻略！最短距離でトレード基礎力	True	Free	268923	27445	54
Beginner Level	1282064	７日でマスター ビギナー向け Adobe Illustrator どきどきセミナー	True	Free	161029	5924	77
Expert Level	1275790	[Value Investing] Where White People Keep Thei...	True	95	5172	249	15
Intermediate Level	1276182	株式投資で本当のテクニカル分析ができるようになる	True	Free	29167	3326	32

In [56]: `df.groupby('level').max()`

Out[56]:

level	course_id	course_title	is_paid	price	num_subscribers	num_reviews	num_lecture
All Levels	1277924	６時間でインターバンク市場を攻略！最短距離でトレード基礎力	True	Free	268923	27445	54
Beginner Level	1282064	７日でマスター ビギナー向け Adobe Illustrator どきどきセミナー	True	Free	161029	5924	77
Expert Level	1275790	[Value Investing] Where White People Keep Thei...	True	95	5172	249	15
Intermediate Level	1276182	株式投資で本当のテクニカル分析ができるようになる	True	Free	29167	3326	32

In [ ]: