UdemyAnalysis

In [2]:
```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [5]:
```python
data=pd.read_csv("udemy_course_data.csv")
```

In [6]:
```python
data
```

`Out[6]:`

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	nu
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	
2	1006314	Financial Modeling for Business Analysts and C...	https://www.udemy.com/financial-modeling-for-b...	True	45	2174	74	
3	1210588	Beginner to Pro - Financial Analysis in Excel ...	https://www.udemy.com/complete-excel-finance-c...	True	95	2451	11	
4	1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-pro...	True	200	1276	45	
...	...	...	...	...	...	...	...	
3678	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040	14	
3679	1088178	How To Design A WordPress Website With No Codi...	https://www.udemy.com/how-to-make-a-wordpress-...	True	25	306	3	
3680	635248	Learn and Build using Polymer	https://www.udemy.com/learn-and-build-using-po...	True	40	513	169	
3681	905096	CSS Animations: Create Amazing Effects on Your...	https://www.udemy.com/css-animations-create-am...	True	50	300	31	
3682	297602	Using MODX CMS to Build Websites: A Beginner's...	https://www.udemy.com/using-modx-cms-to-build-...	True	45	901	36	

3683 rows × 18 columns

`In [7]:` `data.dtypes`

```
Out[7]:  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        object
         published_timestamp     object
         subject                 object
         profit                   int64
         published_date          object
         published_time          object
         year                     int64
         month                    int64
         day                      int64
         dtype: object
```

1. Display Top 10 Rows of The Dataset

```
In [8]:  data.head(10)
```

Out[8]:

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_l
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	
1	1113822	Complete GST Course & Certification - Grow You...	https://www.udemy.com/goods-and-services-tax/	True	75	2792	923	
2	1006314	Financial Modeling for Business Analysts and C...	https://www.udemy.com/financial-modeling-for-b...	True	45	2174	74	
3	1210588	Beginner to Pro - Financial Analysis in Excel ...	https://www.udemy.com/complete-excel-finance-c...	True	95	2451	11	
4	1011058	How To Maximize Your Profits Trading Options	https://www.udemy.com/how-to-maximize-your-pro...	True	200	1276	45	
5	192870	Trading Penny Stocks: A Guide for All Levels I...	https://www.udemy.com/trading-penny-stocks-a-g...	True	150	9221	138	
6	739964	Investing And Trading For Beginners: Mastering...	https://www.udemy.com/investing-and-trading-fo...	True	65	1540	178	
7	403100	Trading Stock Chart Patterns For Immediate, Ex...	https://www.udemy.com/trading-chart-patterns-f...	True	95	2917	148	
8	476268	Options Trading 3 : Advanced Stock Profit and ...	https://www.udemy.com/day-trading-stock-option...	True	195	5172	34	
9	1167710	The Only Investment Strategy You Need For Your...	https://www.udemy.com/the-only-investment-stra...	True	200	827	14	

2. Check Last 5 Rows of The Dataset

In [9]:
```python
data.tail(5)
```

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

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_l
3678	775618	Learn jQuery from Scratch - Master of JavaScri...	https://www.udemy.com/easy-jquery-for-beginner...	True	100	1040	14	
3679	1088178	How To Design A WordPress Website With No Codi...	https://www.udemy.com/how-to-make-a-wordpress-...	True	25	306	3	
3680	635248	Learn and Build using Polymer	https://www.udemy.com/learn-and-build-using-po...	True	40	513	169	
3681	905096	CSS Animations: Create Amazing Effects on Your...	https://www.udemy.com/css-animations-create-am...	True	50	300	31	
3682	297602	Using MODX CMS to Build Websites: A Beginner's...	https://www.udemy.com/using-modx-cms-to-build-...	True	45	901	36	

3. Find Shape of Our Dataset (Number of Rows And Number of Columns)

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

Out[10]:
```
(3683, 18)
```

4.Getting Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement

In [11]:
```
data.info
```

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

```
Out[11]: <bound method DataFrame.info of       course_id                                              co
         urse_title  \
         0        1070968                  Ultimate Investment Banking Course
         1        1113822  Complete GST Course & Certification - Grow You...
         2        1006314  Financial Modeling for Business Analysts and C...
         3        1210588  Beginner to Pro - Financial Analysis in Excel ...
         4        1011058          How To Maximize Your Profits Trading Options
         ...          ...                                                  ...
         3678      775618  Learn jQuery from Scratch - Master of JavaScri...
         3679     1088178  How To Design A WordPress Website With No Codi...
         3680      635248                         Learn and Build using Polymer
         3681      905096  CSS Animations: Create Amazing Effects on Your...
         3682      297602  Using MODX CMS to Build Websites: A Beginner's...

                                                          url  is_paid  price  \
         0     https://www.udemy.com/ultimate-investment-bank...     True    200
         1            https://www.udemy.com/goods-and-services-tax/     True     75
         2     https://www.udemy.com/financial-modeling-for-b...     True     45
         3     https://www.udemy.com/complete-excel-finance-c...     True     95
         4     https://www.udemy.com/how-to-maximize-your-pro...     True    200
         ...                                                 ...      ...    ...
         3678  https://www.udemy.com/easy-jquery-for-beginner...     True    100
         3679  https://www.udemy.com/how-to-make-a-wordpress-...     True     25
         3680  https://www.udemy.com/learn-and-build-using-po...     True     40
         3681  https://www.udemy.com/css-animations-create-am...     True     50
         3682  https://www.udemy.com/using-modx-cms-to-build-...     True     45

               num_subscribers  num_reviews  num_lectures              level  \
         0                 2147           23            51          All Levels
         1                 2792          923           274          All Levels
         2                 2174           74            51   Intermediate Level
         3                 2451           11            36          All Levels
         4                 1276           45            26   Intermediate Level
         ...                ...          ...           ...                 ...
         3678              1040           14            21          All Levels
         3679               306            3            42       Beginner Level
         3680               513          169            48          All Levels
         3681               300           31            38          All Levels
         3682               901           36            20       Beginner Level

               content_duration    published_timestamp            subject  profit  \
         0            1.5 hours  2017-01-18T20:58:58Z  Business Finance  429400
         1             39 hours  2017-03-09T16:34:20Z  Business Finance  209400
         2            2.5 hours  2016-12-19T19:26:30Z  Business Finance   97830
         3              3 hours  2017-05-30T20:07:24Z  Business Finance  232845
         4              2 hours  2016-12-13T14:57:18Z  Business Finance  255200
         ...                ...                   ...               ...     ...
         3678           2 hours  2016-06-14T17:36:46Z   Web Development  104000
         3679         3.5 hours  2017-03-10T22:24:30Z   Web Development    7650
         3680         3.5 hours  2015-12-30T16:41:42Z   Web Development   20520
         3681           3 hours  2016-08-11T19:06:15Z   Web Development   15000
         3682           2 hours  2014-09-28T19:51:11Z   Web Development   40545

               published_date published_time  year  month  day
         0         18-01-2017       20:58:58Z  2017      1   18
         1         09-03-2017       16:34:20Z  2017      3    9
         2         19-12-2016       19:26:30Z  2016     12   19
         3         30-05-2017       20:07:24Z  2017      5   30
         4         13-12-2016       14:57:18Z  2016     12   13
         ...              ...             ...   ...    ...  ...
         3678      14-06-2016       17:36:46Z  2016      6   14
         3679      10-03-2017       22:24:30Z  2017      3   10
         3680      30-12-2015       16:41:42Z  2015     12   30
         3681      11-08-2016       19:06:15Z  2016      8   11
```
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
3682     28-09-2014     19:51:11Z  2014     9   28

[3683 rows x 18 columns]>
```

5. Check Null Values In The Dataset

In [12]:
```python
data.isnull().sum()
```

Out[12]:
```
course_id            0
course_title         0
url                  0
is_paid              0
price                0
num_subscribers      0
num_reviews          0
num_lectures         0
level                0
content_duration     0
published_timestamp  0
subject              0
profit               0
published_date       0
published_time       1
year                 0
month                0
day                  0
dtype: int64
```

6. Check For Duplicate Data and Drop Them

In [13]:
```python
df=data.duplicated().any()
print("are there any duplicate value in data?",df)
```

```
are there any duplicate value in data? True
```

In [33]:
```python
data=data.drop_duplicates()
```

In [34]:
```python
df=data.duplicated().any()
print("are there any duplicate value in data?",df)
```

```
are there any duplicate value in data? False
```

7. Find Out Number of Courses Per Subjects

In [35]:
```python
data.head(1)
```

Out[35]:

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_le
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	

7. Find Out Number of Courses Per Subjects

In [36]:
```python
data["subject"].value_counts()
```

```
subject
Web Development      1199
Business Finance     1195
Musical Instruments   681
Graphic Design        602
Name: count, dtype: int64
```

In [50]:
```python
sns.countplot(x="subject",data=data)
plt.xlabel("subject",fontsize=13)
plt.ylabel("Number of courses per subject",fontsize=13)
plt.xticks(rotation=65)
plt.show()
```
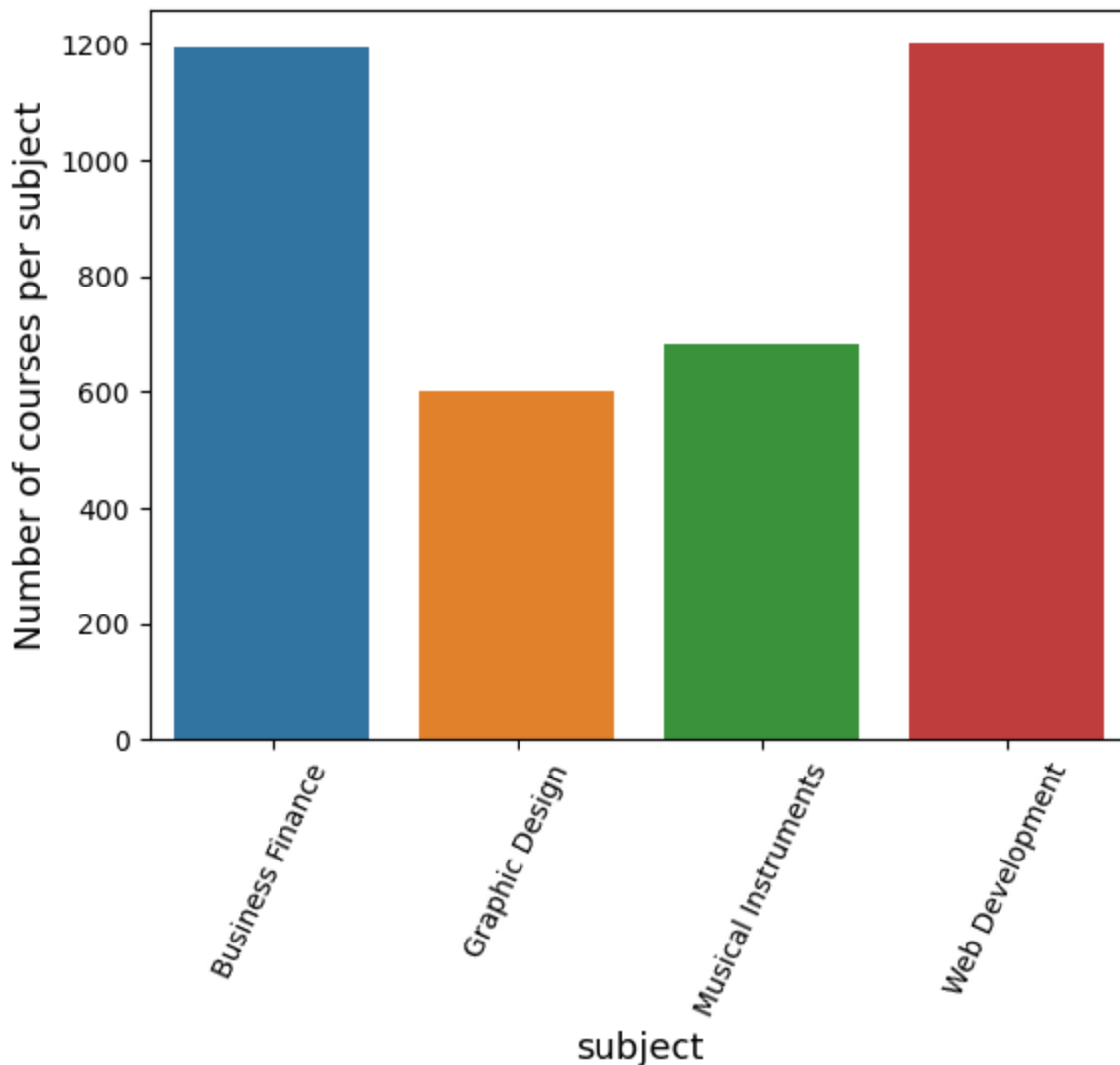

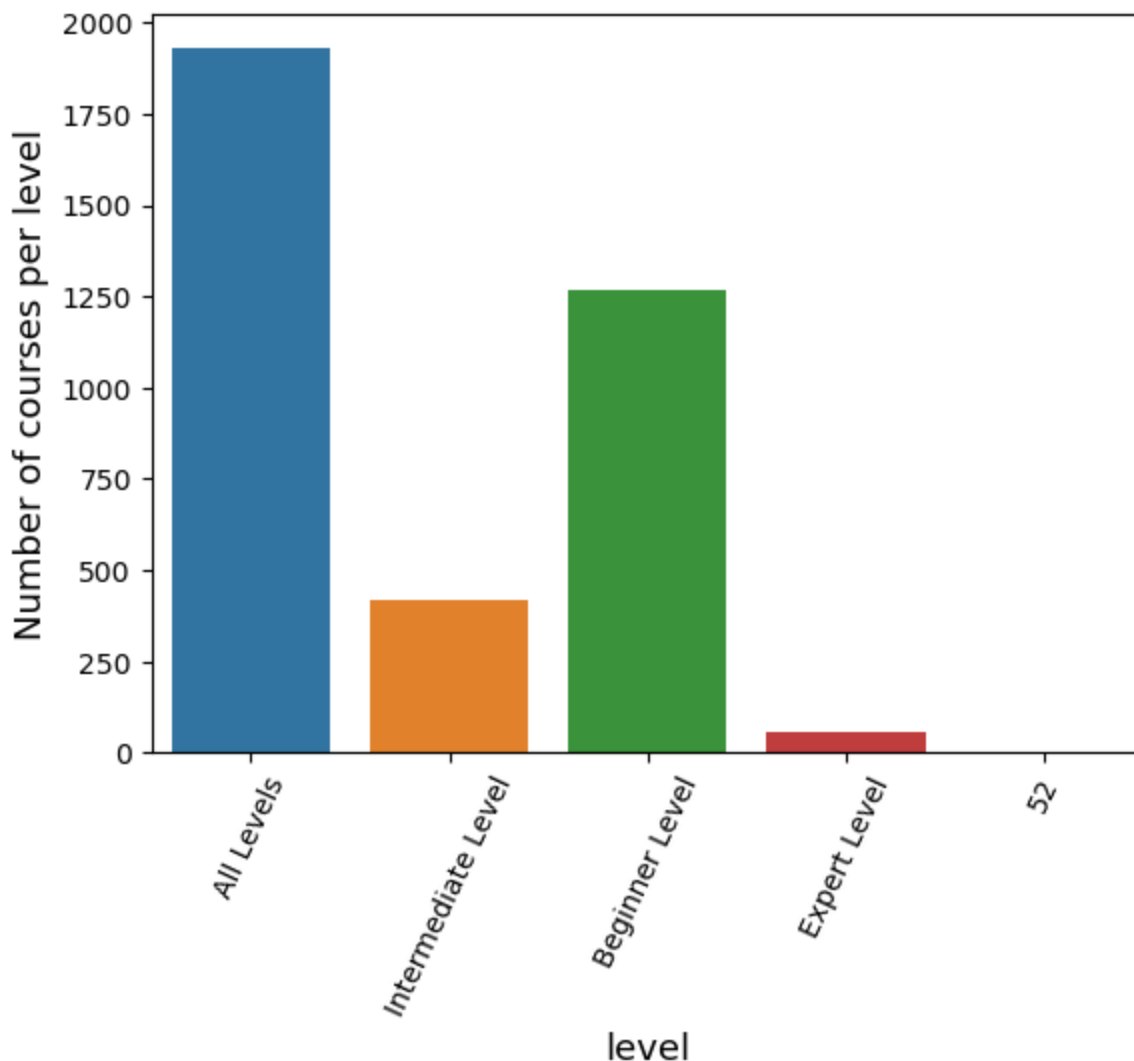## 8. For Which Levels, Udemy Courses Providing The Courses

In [52]:
```python
data["level"].value_counts()
```

Out[52]:
```
level
All Levels          1928
Beginner Level      1269
Intermediate Level   421
Expert Level          58
52                     1
Name: count, dtype: int64
```

In [53]:
```python
sns.countplot(x="level",data=data)
plt.xlabel("level",fontsize=13)
```
er level",fontsize=13)

```
plt.xticks(rotation=65)
plt.show()
```


9. Display The Count of Paid and Free Courses

In [56]: `data.columns`

Out[56]:
```
Index(['course_id', 'course_title', 'url', 'is_paid', 'price',
       'num_subscribers', 'num_reviews', 'num_lectures', 'level',
       'content_duration', 'published_timestamp', 'subject', 'profit',
       'published_date', 'published_time', 'year', 'month', 'day'],
      dtype='object')
```

In [58]:
```
sns.countplot(x="is_paid",data=data)
plt.xlabel("is_paid",fontsize=13)
plt.ylabel("Number of free and paid courses",fontsize=13)
plt.xticks(rotation=65)
plt.show()
```
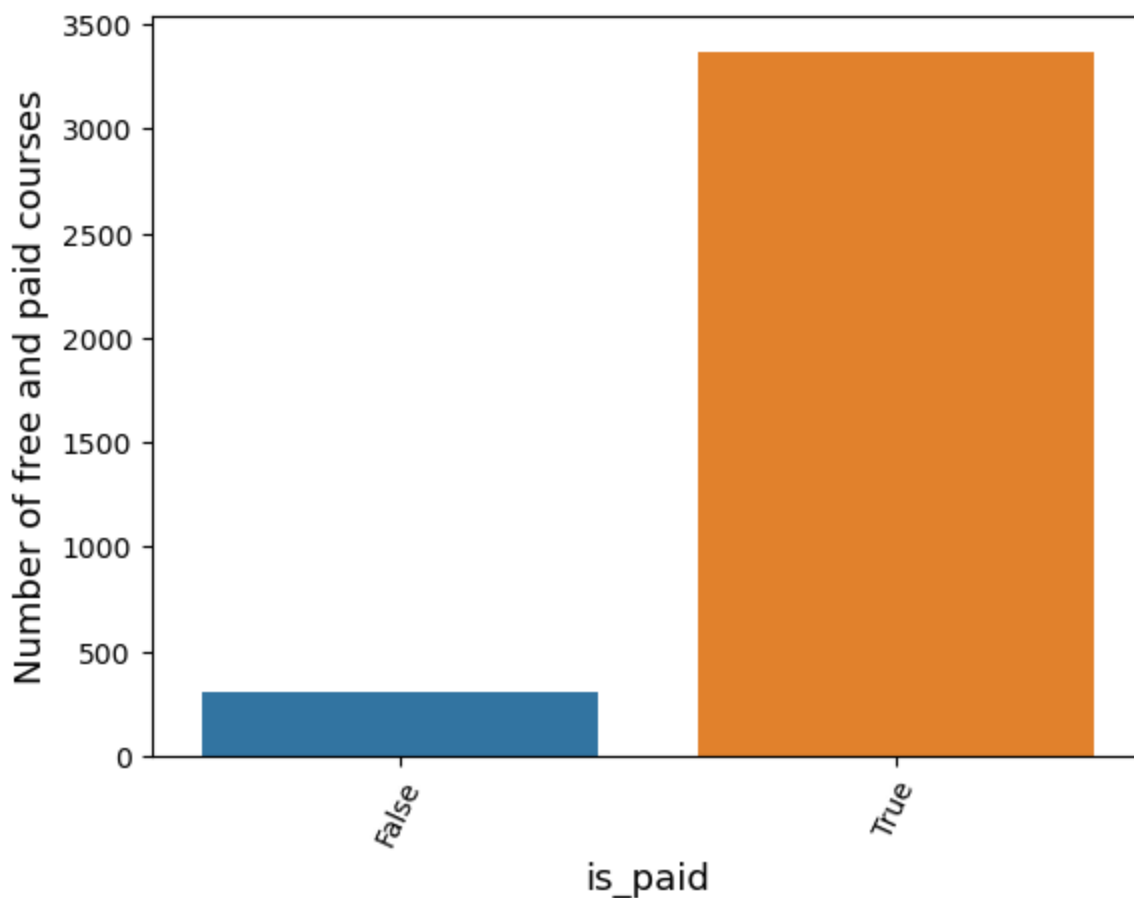
10. Which Courses Have A Higher Number of Subscribers Free or Paid?
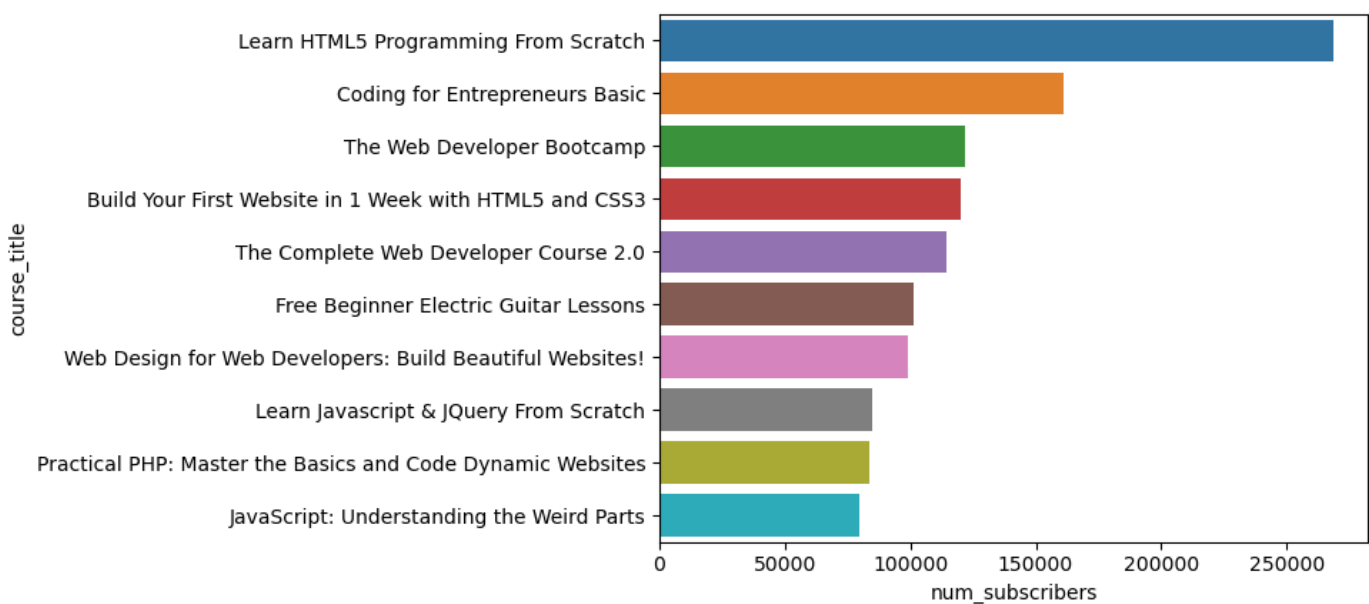
```
In [68]: data.columns
```

```
Out[68]: Index(['course_id', 'course_title', 'url', 'is_paid', 'price',
               'num_subscribers', 'num_reviews', 'num_lectures', 'level',
               'content_duration', 'published_timestamp', 'subject', 'profit',
               'published_date', 'published_time', 'year', 'month', 'day'],
             dtype='object')
```

```
In [69]: sns.barplot(x="is_paid",y="num_subscribers",data=data)
```

```
Out[69]: <Axes: xlabel='is_paid', ylabel='num_subscribers'>
```
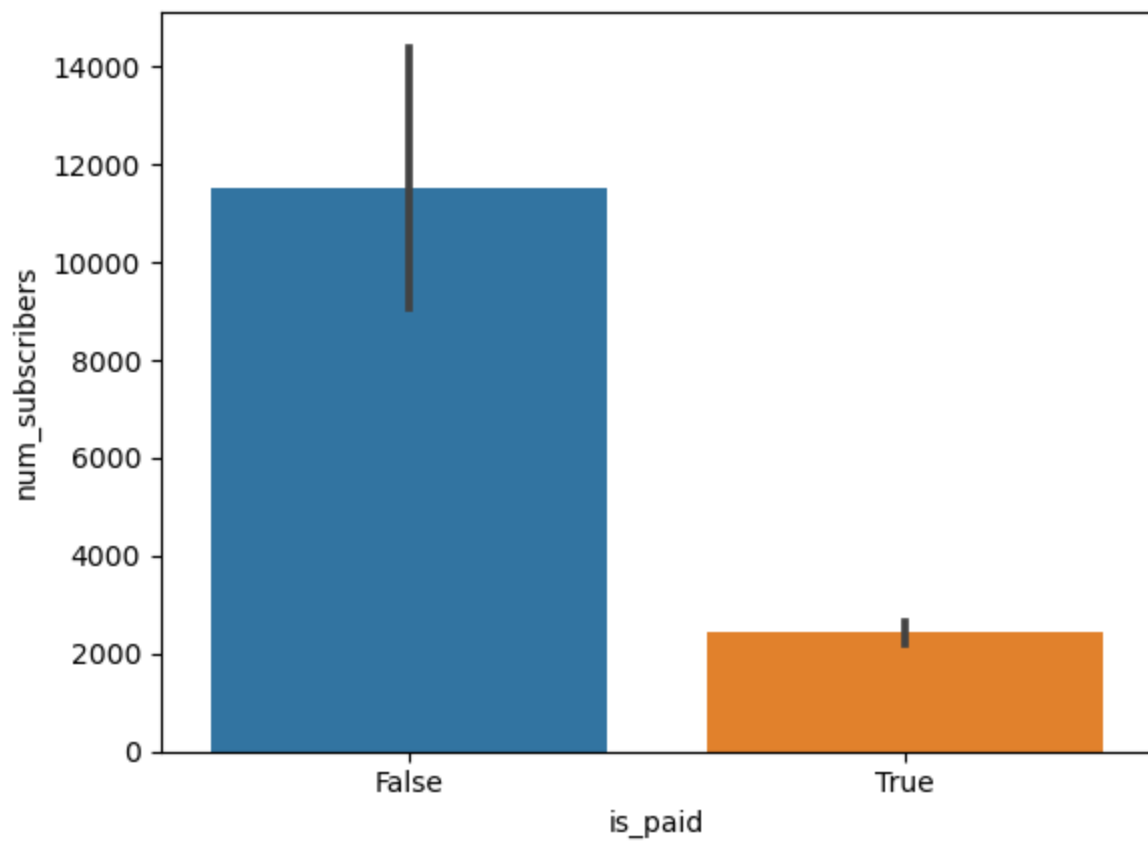
11. Which Level Has The Highest Number of Subscribers?

```
In [85]:  sns.barplot(x="level",y="num_subscribers",data=data)
          plt.xticks(rotation=60)
          plt.show()
```

```
<Figure size 800x400 with 0 Axes>
```
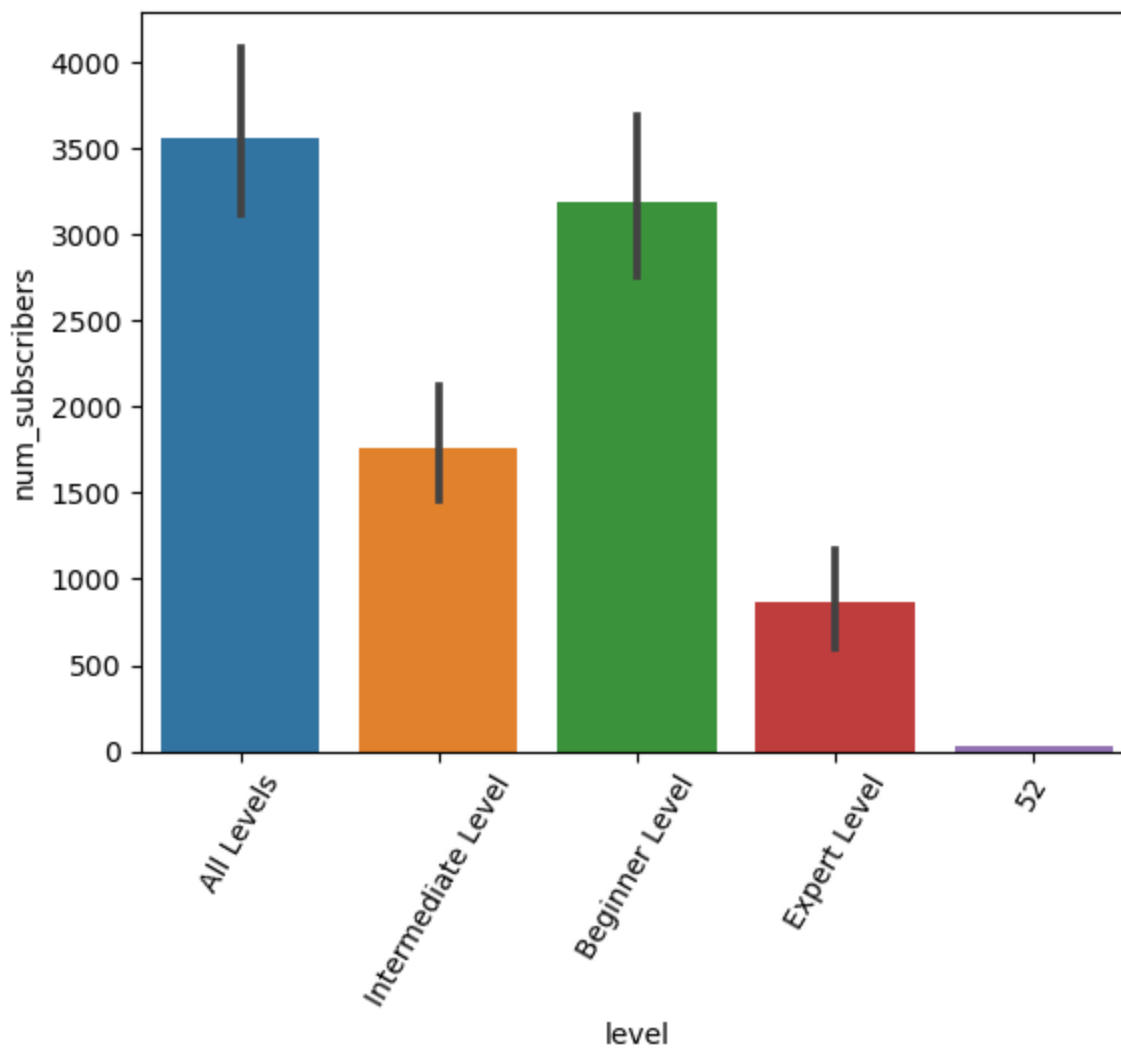
13. Display 10 Most Popular Courses As Per Number of Subscribers

```
In [77]: data.columns
```

```
Out[77]: Index(['course_id', 'course_title', 'url', 'is_paid', 'price',
               'num_subscribers', 'num_reviews', 'num_lectures', 'level',
               'content_duration', 'published_timestamp', 'subject', 'profit',
               'published_date', 'published_time', 'year', 'month', 'day'],
              dtype='object')
```

```
In [14]: top=data.sort_values(by="num_subscribers",ascending=False).head(10)
```

```
In [82]: sns.barplot(x="num_subscribers",y="course_title",data=top)
```
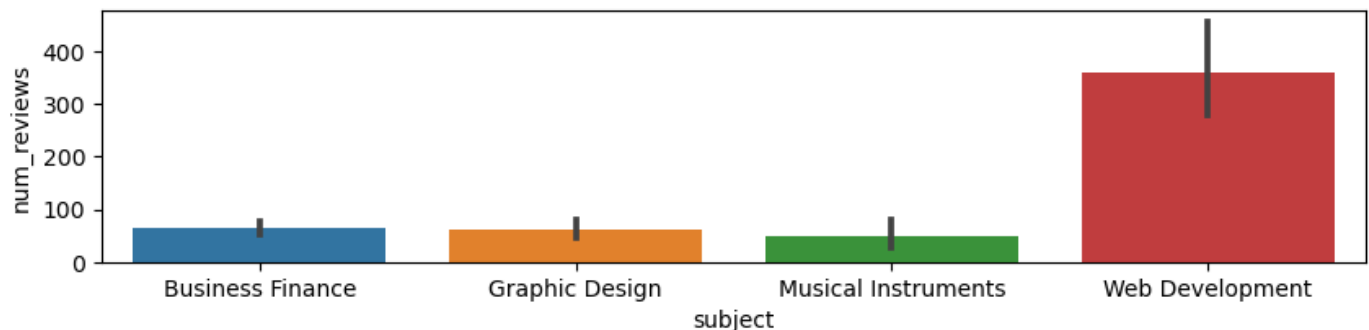
```
Out[82]: <Axes: xlabel='num_subscribers', ylabel='course_title'>
```

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

14. Find The Course Which Is Having The Highest Number of Reviews.

In [15]:
```python
plt.figure(figsize=(10,2))
sns.barplot(x="subject",y="num_reviews",data=data)
```

Out[15]:
```
<Axes: xlabel='subject', ylabel='num_reviews'>
```
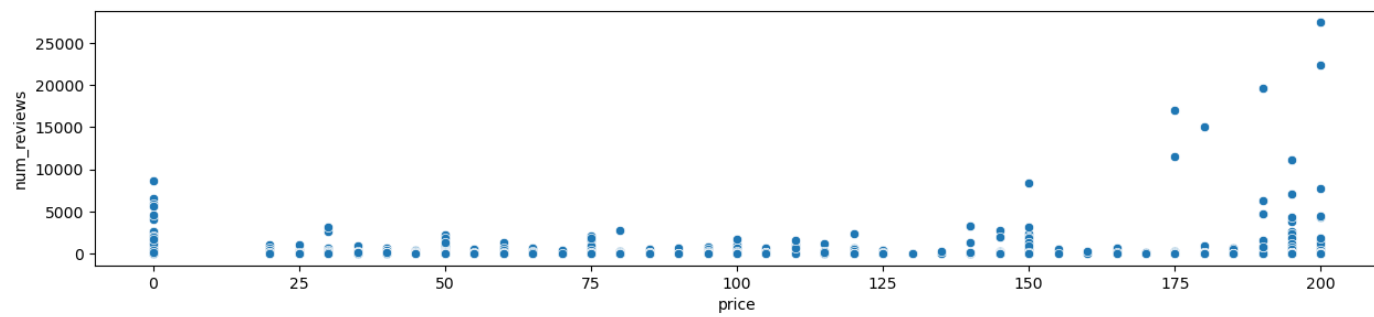

15. Does Price Affect the Number of Reviews?

In [16]:
```python
data.columns
```

Out[16]:
```
Index(['course_id', 'course_title', 'url', 'is_paid', 'price',
       'num_subscribers', 'num_reviews', 'num_lectures', 'level',
       'content_duration', 'published_timestamp', 'subject', 'profit',
       'published_date', 'published_time', 'year', 'month', 'day'],
      dtype='object')
```

In [18]:
```python
plt.figure(figsize=(15,3))

sns.scatterplot(x="price",y="num_reviews",data=data)
```

Out[18]:
```
<Axes: xlabel='price', ylabel='num_reviews'>
```

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

In [ ]: