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

In [2]:
```python
udemy = pd.read_csv("documents/udemy courses.csv")
```

In [3]:
```python
udemy
```

Out[3]:

	course_id	course_title	url	is_paid	price	num_subscribers	num_reviews	num_lectures	level	content_duration	published_timestamp	subject
0	1070968	Ultimate Investment Banking Course	https://www.udemy.com/ultimate-investment-bank...	True	200	2147	23	51	All Levels	1.5	2017-01-18T20:58:58Z	Business Finance
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	2017-03-09T16:34:20Z	Business Finance
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	2016-12-19T19:26:30Z	Business Finance
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	2017-05-30T20:07:24Z	Business Finance
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	2016-12-13T14:57:18Z	Business Finance
...	...	...	...	...	...	...	...	...	...	...	...	...
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	2016-06-14T17:36:46Z	Web Development
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	2017-03-10T22:24:30Z	Web Development
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	2015-12-30T16:41:42Z	Web Development
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	2016-08-11T19:06:15Z	Web Development
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	2014-09-28T19:51:11Z	Web Development

3678 rows × 12 columns

In [4]:
```python
udemy.groupby('subject')
```

Out[4]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000022F425B6410>
```

In [5]:
```python
udemy.groupby('subject')['num_subscribers'].sum()
```

Out[5]:
```
subject
Business Finance       1868711
Graphic Design         1063148
Musical Instruments     846689
Web Development        7980572
Name: num_subscribers, dtype: int64
```

In [7]:
```python
sub= udemy.groupby('subject')['num_subscribers'].sum().sort_values(ascending = True)
```

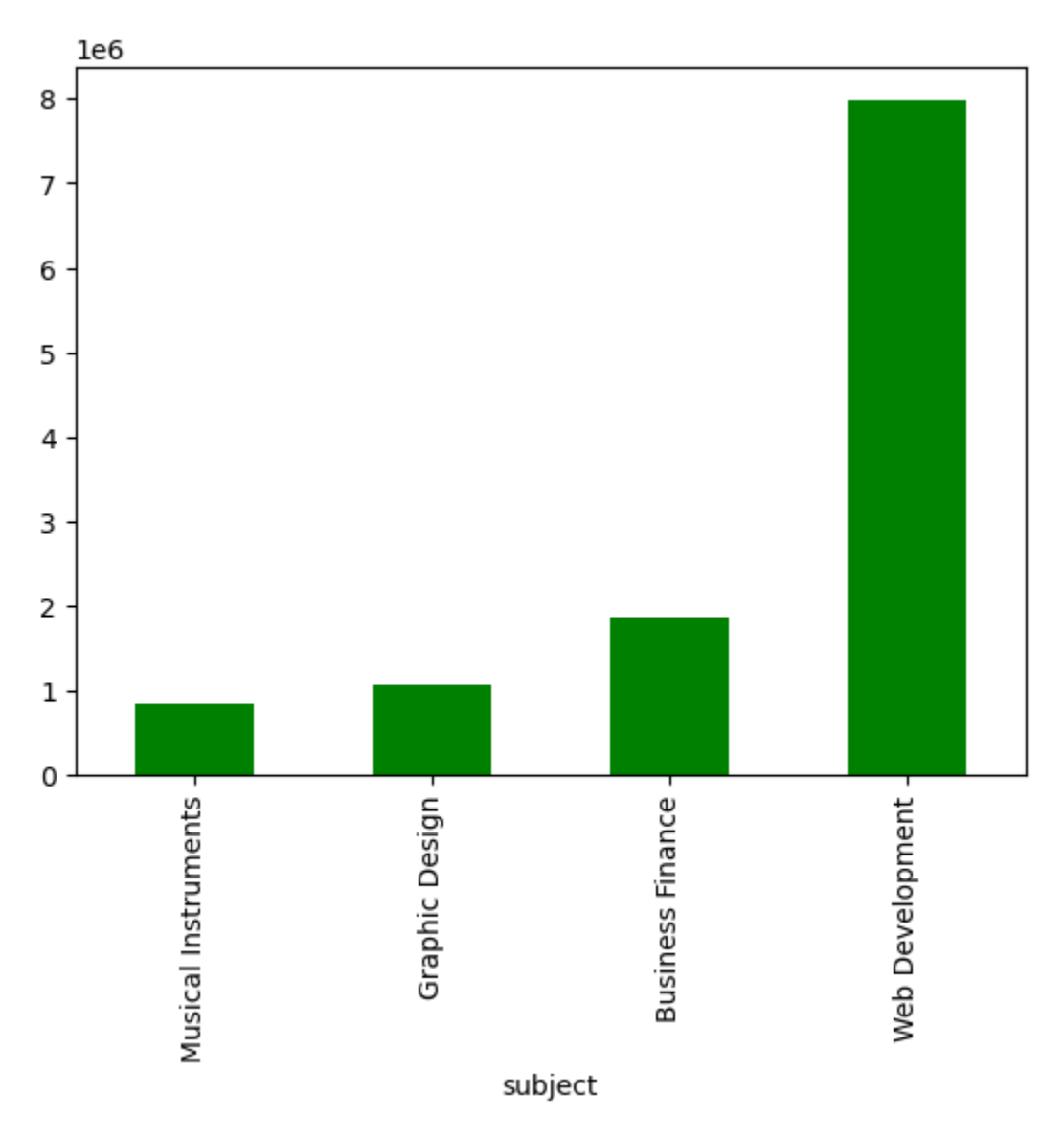
In [8]:
```python
#Total number of students in a particular subject.

sub
```

Out[8]:
```
subject
Musical Instruments     846689
Graphic Design         1063148
Business Finance       1868711
Web Development        7980572
Name: num_subscribers, dtype: int64
```

In [10]:
```python
sub.plot(kind="bar", color = "green")
```

Out[10]:
```
<Axes: xlabel='subject'>
```



In [ ]:
```python
# From the above analysis and graph, it can be deduced that the most popular and well taken course among the students is Web
# developement, while the least poular taken subject among the students is Musical instruments
```

In [83]:
```python
#The number of subscribers in each subjects, in relation to the course been free or paid.
is_paid = udemy.groupby(['subject', 'is_paid'])["num_subscribers" ].sum()
```

In [82]:
```python
is_paid
```

Out[82]:
```
subject              is_paid
Business Finance     False       603228
                     True       1265483
Graphic Design       False       284821
                     True        778327
Musical Instruments  False       304735
                     True        541954
Web Development      False      2382741
                     True       5597831
Name: num_subscribers, dtype: int64
```

In [ ]:
```python
#From the above analysis, it can be seen that more students took the free version of web development courses than other
# subjects free courses, while least amount students took the free courses under graphic design
```

In [12]:
```python
price = udemy.groupby('subject')['price'].sum()
```

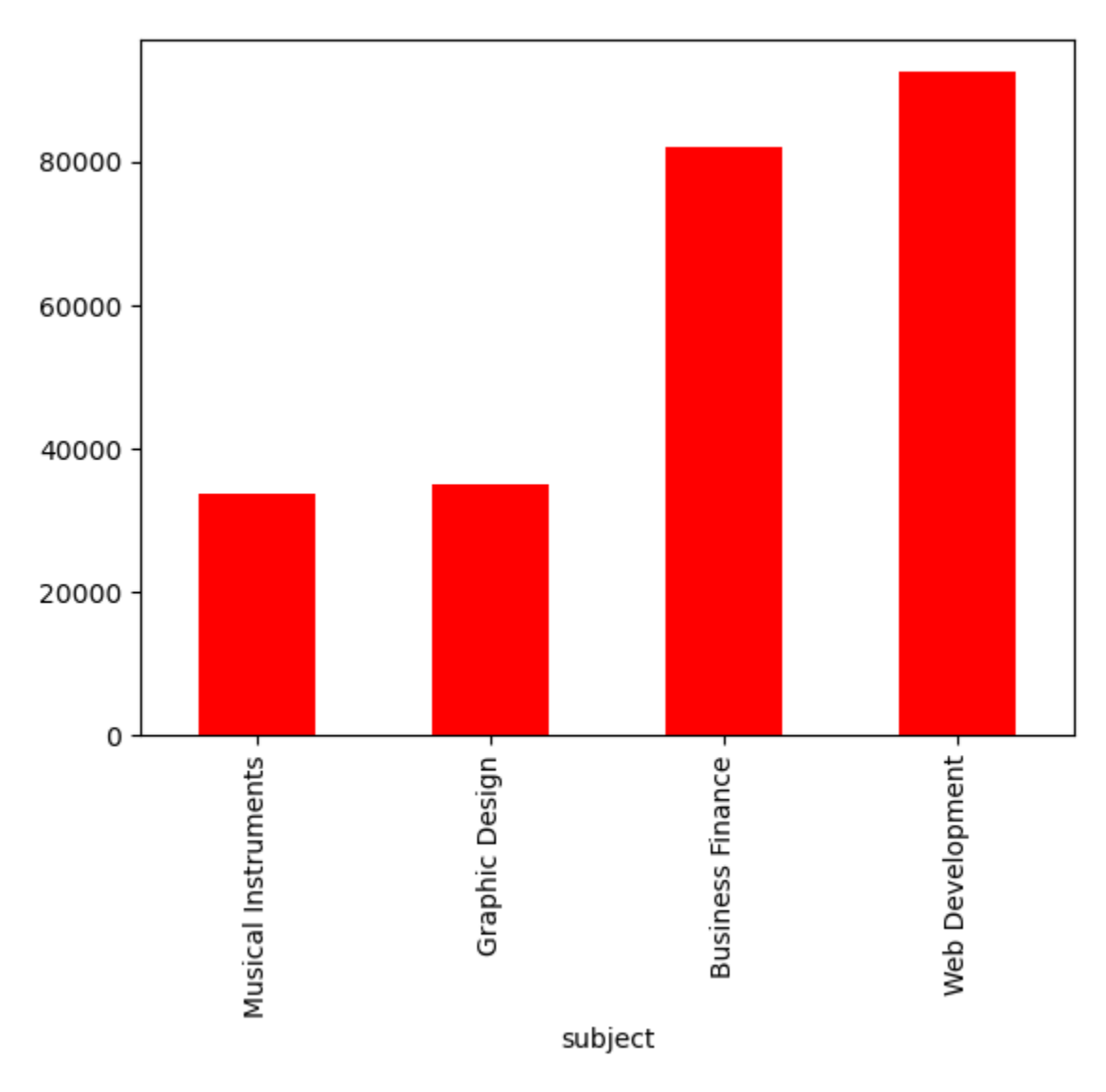
In [14]:
```python
#This is the total price of the various courses in the different subjects.

price.sort_values(ascending=True)
```

Out[14]:
```
subject
Musical Instruments    33700
Graphic Design         34870
Business Finance       81920
Web Development        92440
Name: price, dtype: int64
```

In [20]:
```python
price.sort_values(ascending=True).plot(kind="bar", color = "red")
```

Out[20]:
```
<Axes: xlabel='subject'>
```



In [ ]:
```python
# From the above analysis and graph, it can be deduced that the most priced or costliest courses fall under Web
# developement, while the cheapest courses fall under Musical instruments.
# In other words, more students went for the costliest courses while little amount of students went for the cheapest courses.
```

In [25]:
```python
#Here, I tried to find if there is any correlation between the price of the courses and the review it had.

udemy.groupby('price')['num_reviews'].sum().sort_values(ascending=False)
```

Out[25]:
```
price
0      132038
200    100831
195     52469
190     34075
175     29616
150     29611
20      27107
50      26427
180     16839
100     14778
75      13015
30      11505
95       9962
60       9244
40       7306
120      6874
145      6533
25       6003
110      5183
140      5127
80       4052
35       3836
45       3704
65       2916
115      2268
125      2196
85       1543
70       1454
55       1374
185      1353
90       1302
105      1213
165      1171
160       546
155       526
170       336
135       302
130        86
Name: num_reviews, dtype: int64
```

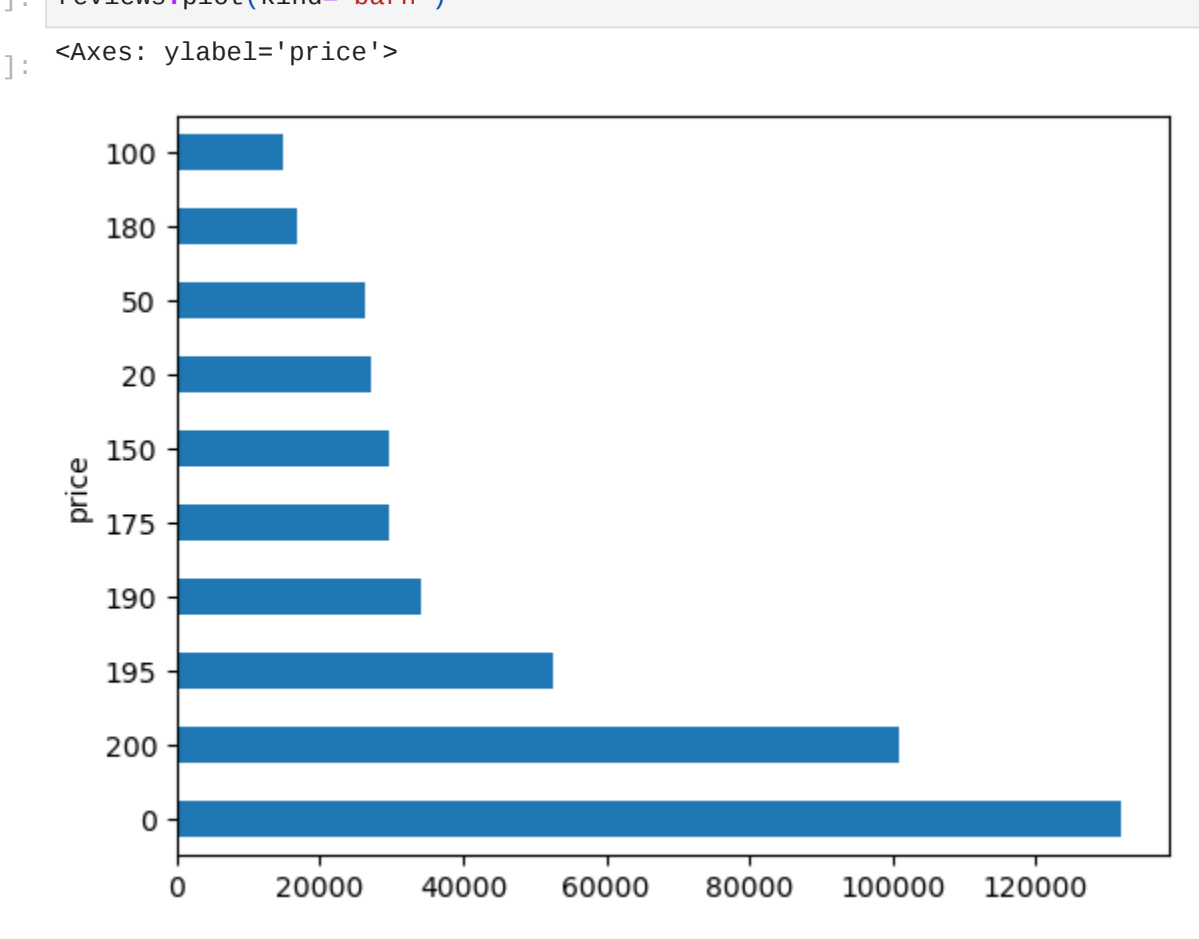
In [27]:
```python
reviews =udemy.groupby('price')['num_reviews'].sum().sort_values(ascending=False).head(10)
```

In [28]:
```python
reviews
```

Out[28]:
```
price
0      132038
200    100831
195     52469
190     34075
175     29616
150     29611
20      27107
50      26427
180     16839
100     14778
Name: num_reviews, dtype: int64
```

In [29]:
```python
reviews.plot(kind="barh")
```

Out[29]:
```
<Axes: ylabel='price'>
```



In [ ]:
```python
# With the above analysis and graph, it can be deduced that the courses that received the most accumulative reviews were the
# free ones that the students paid nothing, the least number of reviews was received by the courses priced at 130$. Inorder to
# be able to plot a graph/chart showing this, the top 10 courses with the highest reviews were taken starting
# with the highest reviewed.
```

In [23]:
```python
#Here, I tried to find if there is any correlation between the price of the courses and the subscribers it had.
udemy.groupby('price')['num_subscribers'].sum()
```

Out[23]:
```
price
0      3575525
20     1320658
25      170861
30      286714
35      116300
40      200217
45       99011
50      717383
55       47561
60      157151
65       61281
70       46791
75      410166
80      100207
85       28806
90       41585
95      347949
100     410491
105      21613
110      37344
115      34497
120     167452
125      48257
130        964
135      16909
140      45169
145     124118
150     376202
155      12356
160       8904
165      13377
170      20350
175     167675
180      90366
185      16886
190     142439
195     939254
200    1336331
Name: num_subscribers, dtype: int64
```

In [30]:
```python
udemy.groupby('price')['num_subscribers'].sum().sort_values(ascending=False)
```

Out[30]:
```
price
0      3575525
200    1336331
20     1320658
195     939254
50      717383
100     410491
75      410166
150     376202
95      347949
30      286714
40      200217
25      170861
175     167675
120     167452
60      157151
190     142439
145     124118
35      116300
80      100207
45       99011
180      90366
65       61281
125      48257
55       47561
70       46791
140      45169
90       41585
110      37344
115      34497
85       28806
105      21613
170      20350
135      16909
185      16886
165      13377
155      12356
160       8904
130        964
Name: num_subscribers, dtype: int64
```

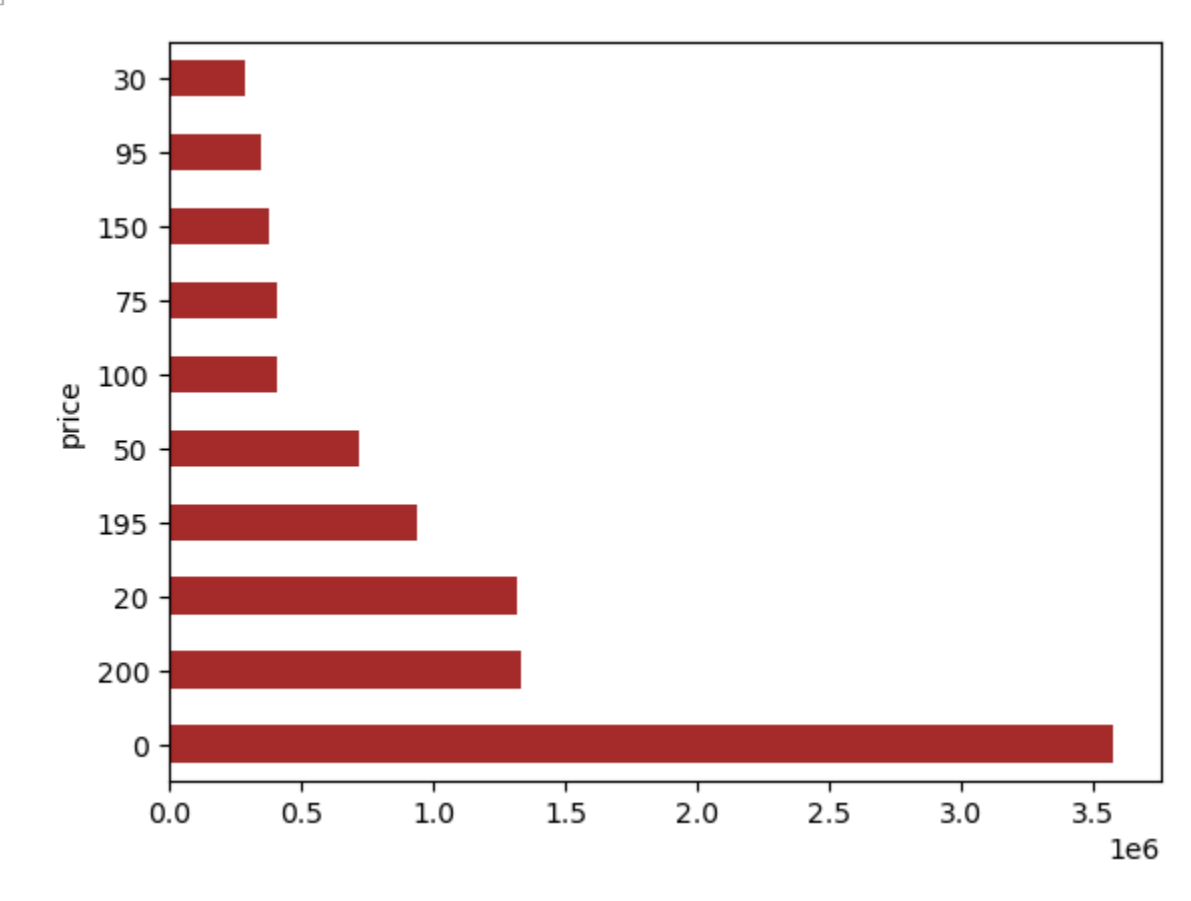
In [32]:
```python
sub =udemy.groupby('price')['num_subscribers'].sum().sort_values(ascending=False).head(10)
```

In [33]:
```python
sub
```

Out[33]:
```
price
0      3575525
200    1336331
20     1320658
195     939254
50      717383
100     410491
75      410166
150     376202
95      347949
30      286714
Name: num_subscribers, dtype: int64
```

In [35]:
```python
sub.plot(kind="barh", color = "brown")
```

Out[35]:
```
<Axes: ylabel='price'>
```



In [ ]:
```python
# With the above analysis and graph, it can be deduced that the courses that had the most subscribers were the free courses
# that the students paid nothing, and the courses priced at 130$ had the least number of subscribers. Inorder to
# be able to plot a graph/chart showing this, the top 10 courses with the highest subscribers were taken starting
# with the highest reviewed.
```

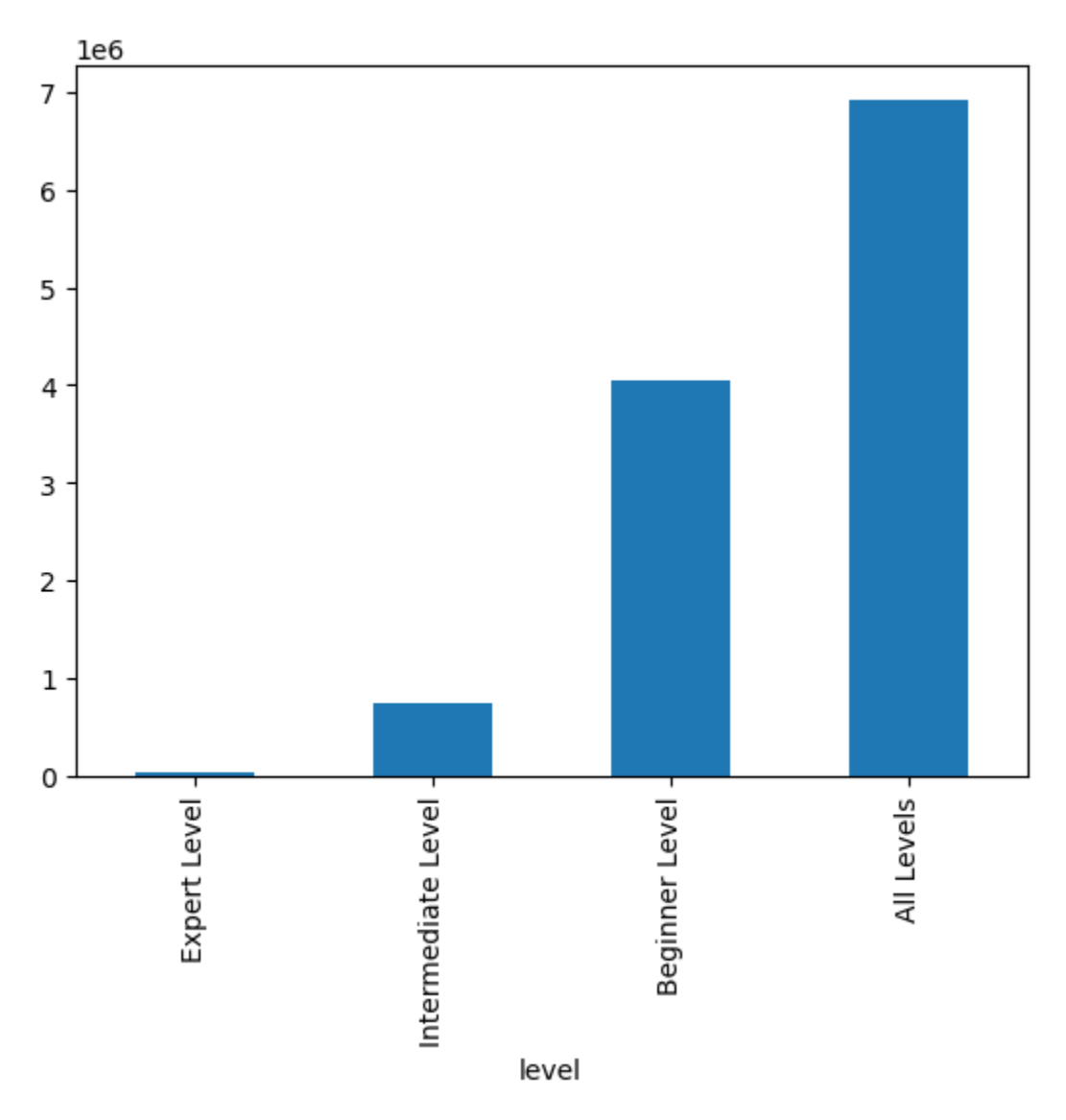
In [38]:
```python
level = udemy.groupby('level')['num_subscribers'].sum()
```

In [39]:
```python
level
```

Out[39]:
```
level
All Levels            6915076
Beginner Level        4051843
Expert Level            50196
Intermediate Level     742005
Name: num_subscribers, dtype: int64
```

In [44]:
```python
level.sort_values(ascending = True).plot(kind="bar")
```

Out[44]:
```
<Axes: xlabel='level'>
```



In [ ]:
```python
# With the above analysis and graph, it is clear that the expert level courses had the least number of subscribers, and all
# levels had the highest number of subscribers, with the beginner level being right next to it.
```

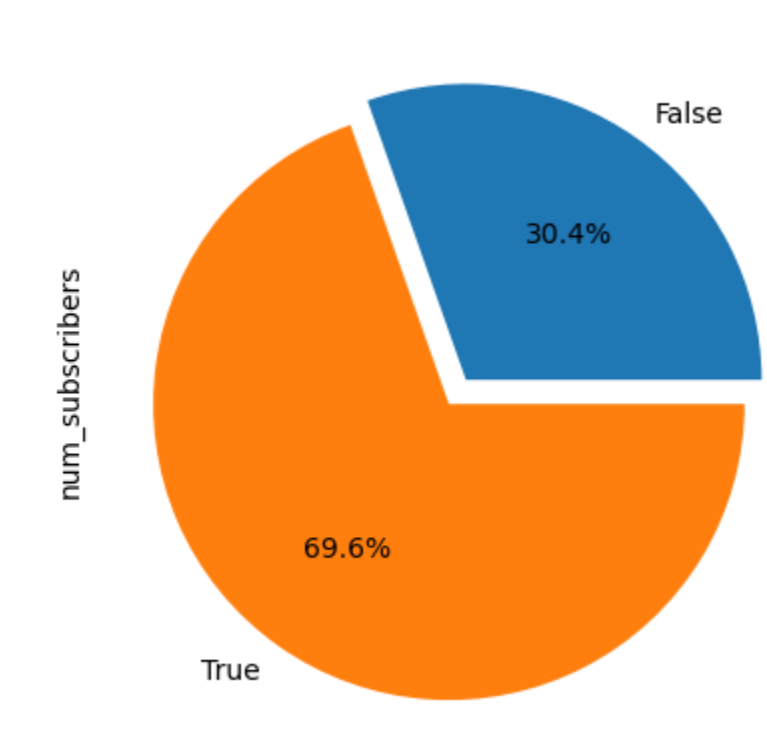
In [46]:
```python
paid = udemy.groupby('is_paid')['num_subscribers'].sum()
```

In [47]:
```python
paid
```

Out[47]:
```
is_paid
False    3575525
True     8183595
Name: num_subscribers, dtype: int64
```

In [50]:
```python
paid.plot(kind="pie", autopct="%1.1f%%", explode =( 0,0.1,))
```

Out[50]:
```
<Axes: ylabel='num_subscribers'>
```



In [ ]:
```python
# Looking at the pie chart above, yes there are free courses and we had more that double the amount of students in the
# free classes, in the paid classes.
```

In [62]:
```python
udemy.groupby('num_lectures')['num_reviews'].sum().sort_values(ascending=False)
```

Out[62]:
```
num_lectures
342    27445
329    24324
304    22473
45     19315
85     18238
       ...
321        1
121        0
156        0
127        0
0          0
Name: num_reviews, Length: 229, dtype: int64
```

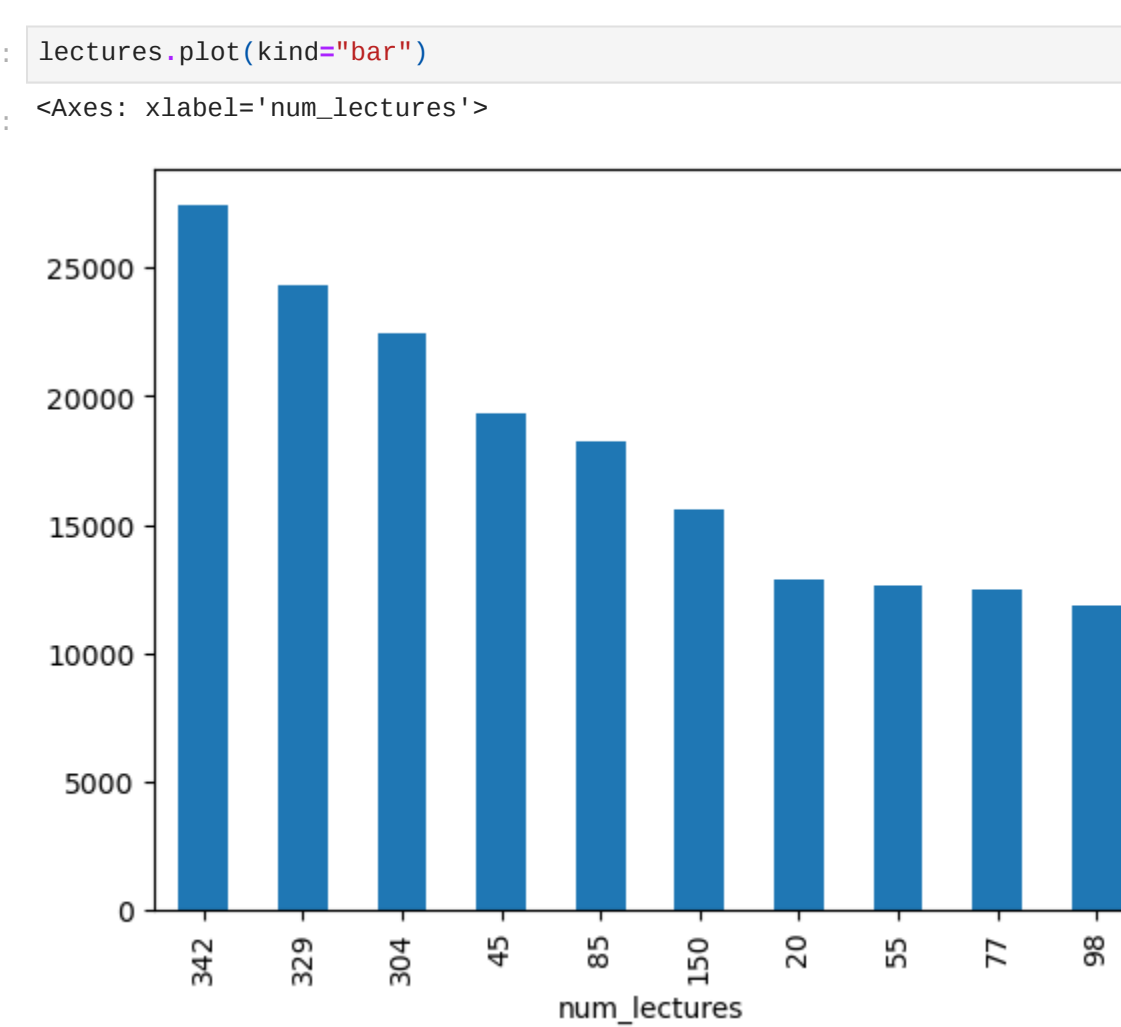
In [56]:
```python
lectures = udemy.groupby('num_lectures')['num_reviews'].sum().sort_values(ascending=False).head(10)
```

In [57]:
```python
lectures
```

Out[57]:
```
num_lectures
342    27445
329    24324
304    22473
45     19315
85     18238
150    15582
20     12893
55     12605
77     12493
98     11822
Name: num_reviews, dtype: int64
```

In [58]:
```python
lectures.plot(kind="bar")
```

Out[58]:
```
<Axes: xlabel='num_lectures'>
```



In [ ]:
```python
# With the above analysis, the courses wth the highest number of lectures had the most reviews, and the courses with the
# lowest number of lectures had no reviews alongside courses with 121, 156 & 127 number of lectures.
```

In [66]:
```python
#Number of free and paid courses in all the subjects.
udemy.groupby('subject')['is_paid'].value_counts()
```

Out[66]:
```
subject              is_paid
Business Finance     True       1099
                     False        96
Graphic Design       True        568
                     False        35
Musical Instruments  True        634
                     False        46
Web Development      True       1067
                     False       133
Name: count, dtype: int64
```

In [74]:

In [ ]:

In [ ]:

In [ ]: