

In [2]:

```
import pandas as pd
import numpy as np
import pylab as py
import matplotlib as plt
%matplotlib inline
df = pd.read_csv('/Users/Tanakorn/Desktop/Bamm.csv')
```

In [3]:

```
df.head(10)
```

Out[3]:

	user_id	submission_count	problem_solved	contribution	country	follower_count	last_online_time_seconds	max_rating	rating
0	user_3311	47	40	0	NaN	4	1504111645	348.337	330.8
1	user_3028	63	52	0	India	17	1498998165	405.677	339.4
2	user_2268	226	203	-8	Egypt	24	1505566052	307.339	284.4
3	user_480	611	490	1	Ukraine	94	1505257499	525.803	471.3
4	user_650	504	479	12	Russia	4	1496613433	548.739	486.3
5	user_3177	3	1	0	NaN	0	1484244659	331.135	331.7
6	user_2032	216	199	0	India	28	1505534677	387.328	387.3
7	user_3306	2	1	0	NaN	0	1505568051	304.186	281.2
8	user_1483	91	79	0	NaN	0	1502437836	318.807	258.8
9	user_667	91	71	0	NaN	4	1494953565	383.028	362.3

In [4]:

```
df.describe()
```

Out[4]:

	submission_count	problem_solved	contribution	follower_count	last_online_time_seconds	max_rating	rating	regist
count	3571.000000	3571.000000	3571.000000	3571.000000	3.571000e+03	3571.000000	3571.000000	3.5710
mean	299.481098	267.894427	4.102492	46.690563	1.502680e+09	390.374392	350.165578	1.4340
std	366.102887	344.139688	16.552256	211.494638	5.114850e+06	92.428788	106.592503	4.7500
min	1.000000	0.000000	-64.000000	0.000000	1.484237e+09	303.899000	0.000000	1.2640
25%	66.500000	53.000000	0.000000	4.000000	1.502691e+09	317.661000	279.243000	1.4160
50%	169.000000	146.000000	0.000000	13.000000	1.505054e+09	355.791000	329.702000	1.4490
75%	390.000000	349.000000	0.000000	40.000000	1.505551e+09	444.954000	413.417500	1.4700
max	4570.000000	4476.000000	171.000000	10575.000000	1.505595e+09	983.085000	911.124000	1.4840

In [5]:

```
df['rating'].value_counts()
```

Out[5]:

```
304.186 53
314.507 11
320.814 11
295.298 11
272.649 10
289.851 10
322.248 10
249.140 10
341.170 10
284.690 9
347.477 9
292.144 9
386.755 9
319.954 9
302.179 9
```

```
457.282 9
305.046 9
315.940 9
256.594 9
279.817 9
298.165 8
317.374 8
381.307 8
263.475 8
394.209 8
285.550 8
277.810 8
321.101 8
297.592 8
288.704 8
```

```
..
522.936 1
415.711 1
664.851 1
434.060 1
419.725 1
448.394 1
434.346 1
185.206 1
675.745 1
489.679 1
211.869 1
192.087 1
497.706 1
771.502 1
339.163 1
72.821 1
634.174 1
622.993 1
561.927 1
685.206 1
490.252 1
507.454 1
351.778 1
572.534 1
526.089 1
508.601 1
563.647 1
600.631 1
491.112 1
350.344 1
```

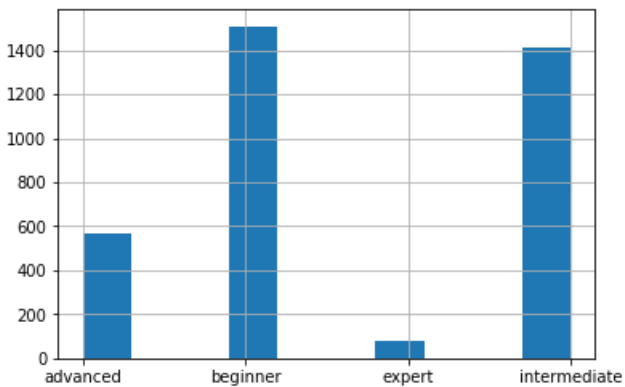
Name: rating, Length: 1307, dtype: int64

In [6]:

```
#Histogram, This is show the number of rank in each of them.
df['rank'].hist(bins=10)
```

Out[6]:

<matplotlib.axes._subplots.AxesSubplot at 0x106c8d470>

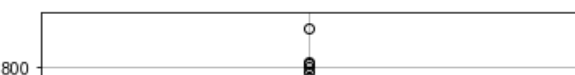


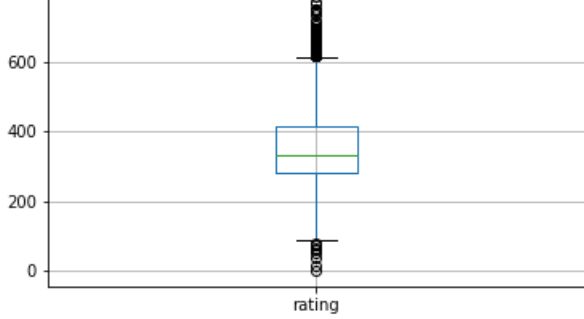
In [7]:

```
#boxplot of the rating.
df.boxplot(column='rating')
```

Out[7]:

<matplotlib.axes._subplots.AxesSubplot at 0x1067da4a8>



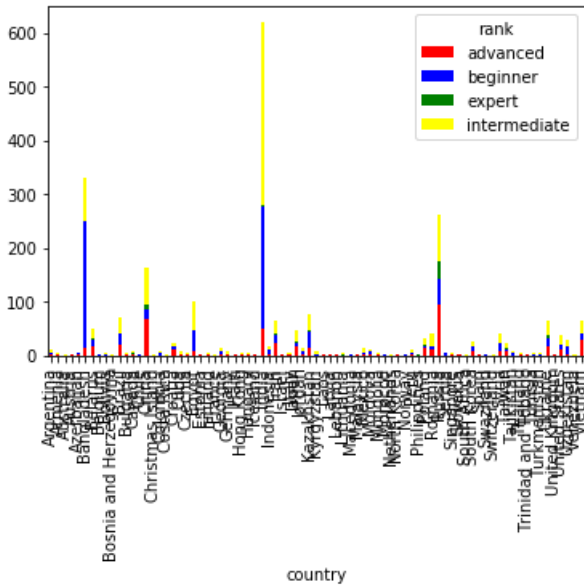


In [8]:

```
#show the rank in each country
temp1 = pd.crosstab(df['country'], df['rank'])
temp1.plot(kind='bar', stacked=True, color=['red','blue','green','yellow'], grid=False)
```

Out[8]:

<matplotlib.axes._subplots.AxesSubplot at 0x113949978>

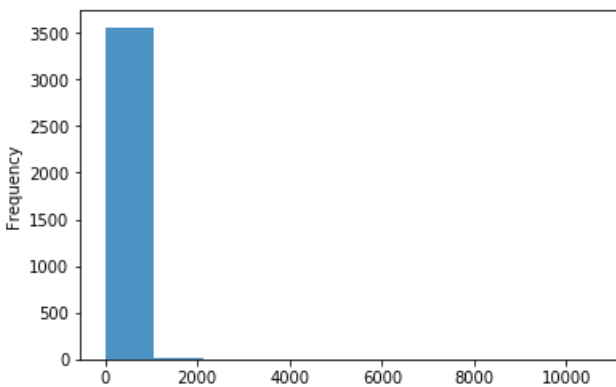


In [9]:

```
df['follower_count'].plot.hist(alpha=0.8)
```

Out[9]:

<matplotlib.axes._subplots.AxesSubplot at 0x113a59ba8>

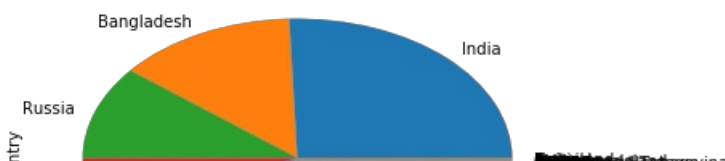


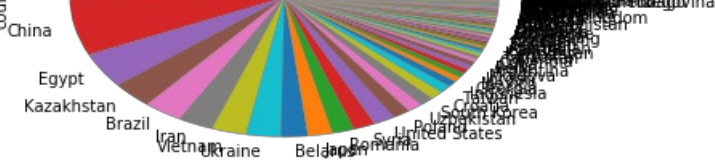
In [10]:

```
#pie chart show the amount of country
df.country.value_counts().plot(kind='pie')
```

Out[10]:

<matplotlib.axes._subplots.AxesSubplot at 0x113b430f0>



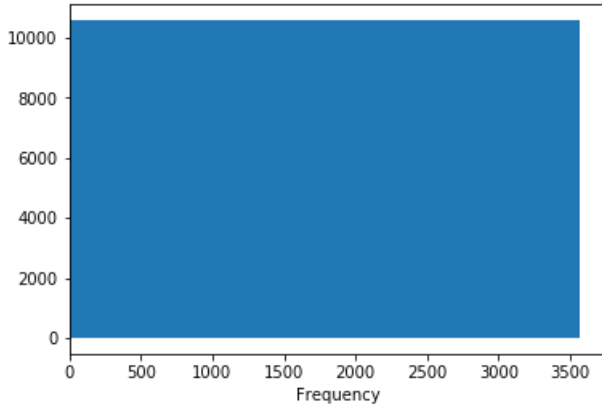


In [11]:

```
df['follower_count'].plot.hist(orientation='horizontal', cumulative=True)
```

Out[11]:

<matplotlib.axes._subplots.AxesSubplot at 0x113e37668>

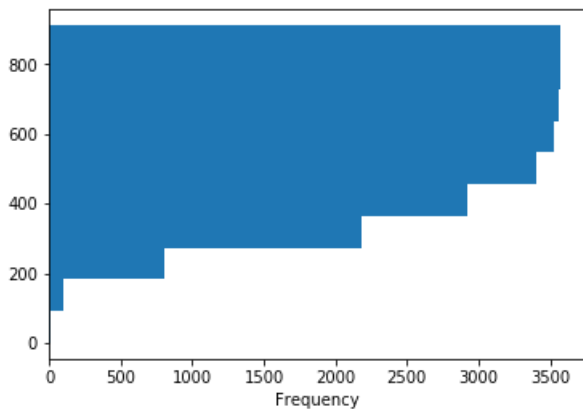


In [12]:

```
#the horizontal histogram show the frequency of rating
df['rating'].plot.hist(orientation='horizontal', cumulative=True)
```

Out[12]:

<matplotlib.axes._subplots.AxesSubplot at 0x113d04cf8>

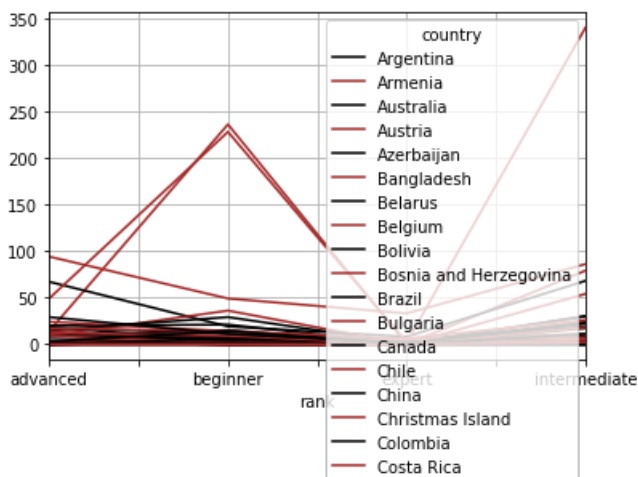


In [13]:

```
#The line graph
temp1 = pd.crosstab(df['rank'], df['country'],)
temp1.plot(kind='line', stacked=False, color=['black', 'brown'], grid=True)
```

Out[13]:

<matplotlib.axes._subplots.AxesSubplot at 0x1140a3198>



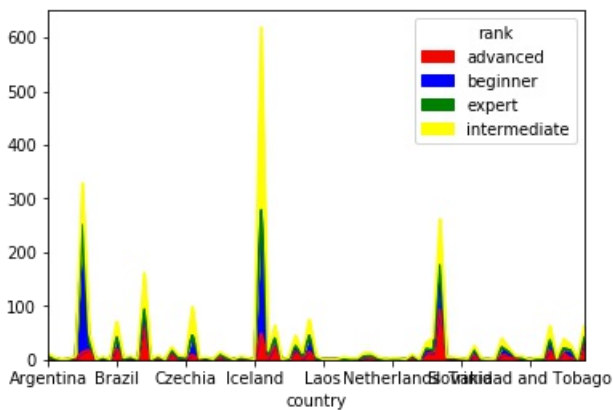
- Croatia
- Cuba
- Czechia
- Egypt
- Estonia
- Finland
- France
- Georgia
- Germany
- Haiti
- Hong Kong
- Hungary
- Iceland
- India
- Indonesia
- Iran
- Israel
- Italy
- Japan
- Jordan
- Kazakhstan
- Kyrgyzstan
- Laos
- Latvia
- Lebanon
- Lithuania
- Macedonia
- Malaysia
- Mexico
- Moldova
- Mongolia
- Morocco
- Netherlands
- North Korea
- Norway
- Peru
- Philippines
- Poland
- Romania
- Russia
- Serbia
- Singapore
- Slovakia
- South Africa
- South Korea
- Spain
- Swaziland
- Switzerland
- Syria
- Taiwan
- Tajikistan
- Thailand
- Trinidad and Tobago
- Tunisia
- Turkmenistan
- Ukraine
- United Kingdom
- United States
- Uzbekistan
- Venezuela
- Vietnam

In [14]:

```
temp1 = pd.crosstab(df['country'], df['rank'])
temp1.plot(kind='area', stacked=True, color=['red', 'blue', 'green', 'yellow'], grid=False)
```

Out[14]:

<matplotlib.axes._subplots.AxesSubplot at 0x1140b6eb8>



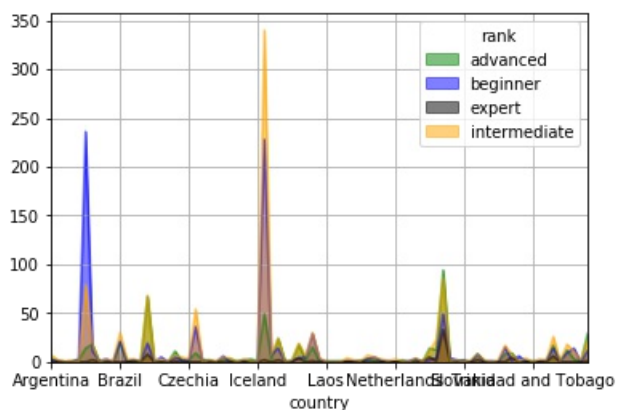
In [15]:

```
temp1 = pd.crosstab(df['country'], df['rank'])
```

```
temp1 = pd.crosstab(df['country'], df['rank'])
temp1.plot(kind='area', stacked=False, color=['green', 'blue', 'black', 'orange'], grid=True)
```

Out[15]:

<matplotlib.axes._subplots.AxesSubplot at 0x1145620f0>

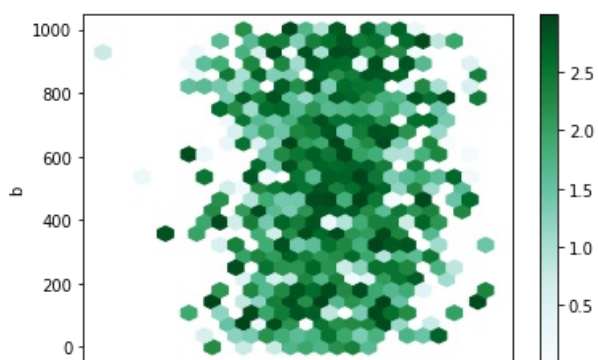


In [16]:

```
#This is the Hexagonal Bin Plot
df = pd.DataFrame(np.random.randn(1000, 2), columns=['a', 'b'])
df['b'] = df['b'] + np.arange(1000)
df['z'] = np.random.uniform(0, 3, 1000)
df.plot.hexbin(x='a', y='b', C='z', reduce_C_function=np.max, gridsize=25)
```

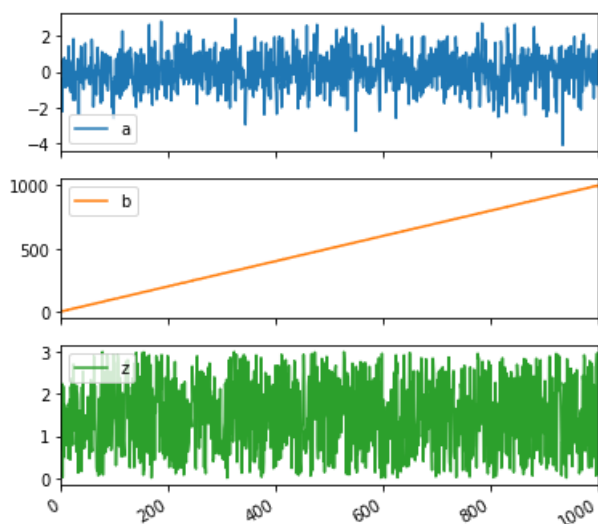
Out[16]:

<matplotlib.axes._subplots.AxesSubplot at 0x114654a20>



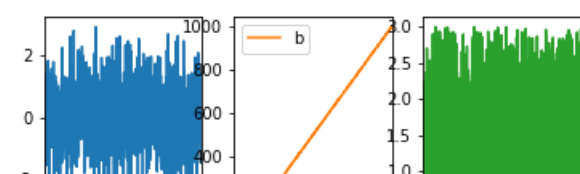
In [17]:

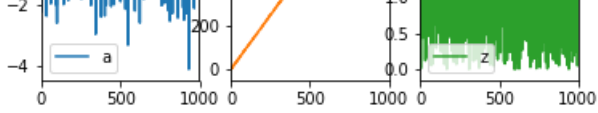
```
df.plot(subplots=True, figsize=(6, 6));
```



In [18]:

```
df.plot(subplots=True, layout=(2, 3), figsize=(6, 6), sharex=False);
```





In [1]:

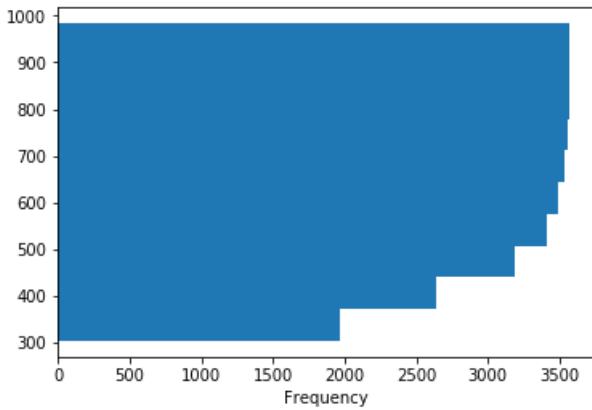
```
import pandas as pd
import numpy as np
import pylab as py
import matplotlib as plt
%matplotlib inline
df = pd.read_csv('/Users/Tanakorn/Desktop/Bamm.csv')
```

In [2]:

```
df['max_rating'].plot.hist(orientation='horizontal', cumulative=True)
```

Out[2]:

<matplotlib.axes._subplots.AxesSubplot at 0x107f10470>

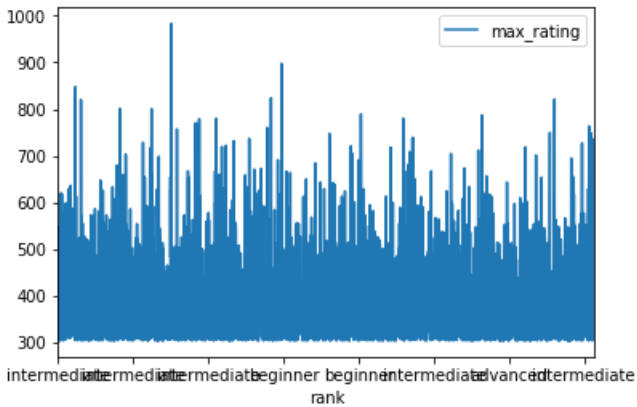


In [3]:

```
df.plot(x='rank', y='max_rating')
```

Out[3]:

<matplotlib.axes._subplots.AxesSubplot at 0x10bfa2780>

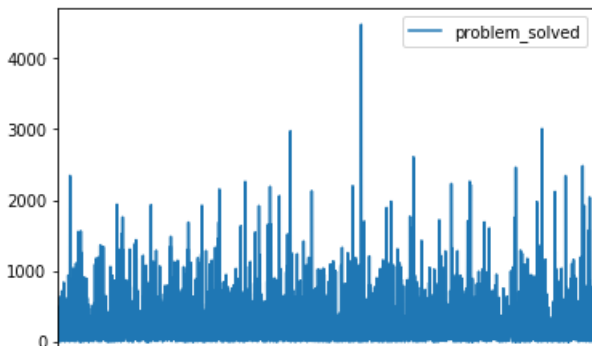


In [4]:

```
df.plot(x='country', y='problem_solved')
#show the relation between country and problem solved
```

Out[4]:

<matplotlib.axes._subplots.AxesSubplot at 0x10bfab4a8>

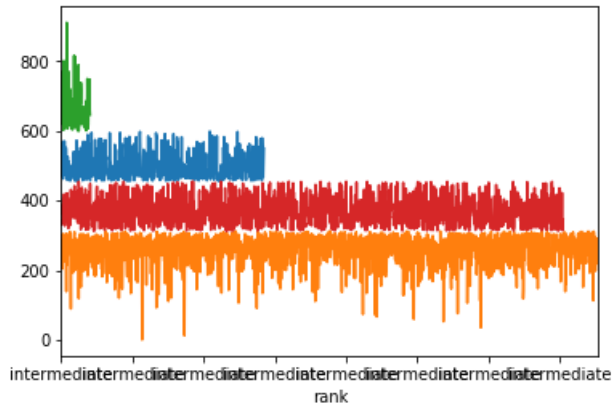



```
In [6]: #This is the combination of the graph above, I convert those into one graph.
```

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots()
df.groupby('rank').plot(x='rank', y='rating', ax=ax, legend=False)
```

Out[6]:

```
rank
advanced    AxesSubplot(0.125,0.125;0.775x0.755)
beginner    AxesSubplot(0.125,0.125;0.775x0.755)
expert      AxesSubplot(0.125,0.125;0.775x0.755)
intermediate AxesSubplot(0.125,0.125;0.775x0.755)
dtype: object
```

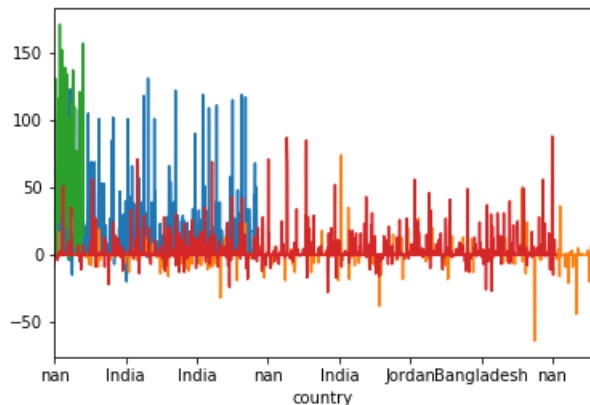


In [7]:

```
fig, ax = plt.subplots()
df.groupby('rank').plot(x='country', y='contribution', ax=ax, legend=False)
```

Out[7]:

```
rank
advanced    AxesSubplot(0.125,0.125;0.775x0.755)
beginner    AxesSubplot(0.125,0.125;0.775x0.755)
expert      AxesSubplot(0.125,0.125;0.775x0.755)
intermediate AxesSubplot(0.125,0.125;0.775x0.755)
dtype: object
```



In [24]:

```
# Mean of the solved problems for each country
df.groupby("country")["problem_solved"].mean()
```

Out[24]:

```
country
Argentina    253.583333
Armenia      526.250000
Australia     97.000000
Austria      599.500000
Azerbaijan   332.833333
Bangladesh   240.799392
Belarus      627.795918
Belgium       8.000000
Bolivia      269.000000
Bosnia and Herzegovina 290.000000
Brazil       339.239437
Bulgaria     504.500000
Canada       177.714286
Chile        136.500000
China        264.364198
```

```

Christmas Island      232.000000
Colombia              230.625000
Costa Rica            164.000000
Croatia              229.125000
Cuba                 450.000000
Czechia              202.166667
Egypt                622.474747
Estonia              259.333333
Finland              356.400000
France               559.000000
Georgia              475.500000
Germany              306.285714
Haiti                23.000000
Hong Kong            476.333333
Hungary              283.500000
...
Morocco              225.000000
Netherlands          159.666667
North Korea          389.666667
Norway               34.000000
Peru                 532.100000
Philippines          240.000000
Poland               328.500000
Romania              281.750000
Russia               439.687023
Serbia               134.000000
Singapore            187.200000
Slovakia             360.333333
South Africa         98.500000
South Korea          187.230769
Spain                447.333333
Swaziland            11.000000
Switzerland          374.000000
Syria                606.275000
Taiwan               742.318182
Tajikistan           336.285714
Thailand             234.333333
Trinidad and Tobago  115.000000
Tunisia              275.500000
Turkmenistan         409.000000
Ukraine              549.079365
United Kingdom       417.333333
United States        342.358974
Uzbekistan           519.142857
Venezuela            777.000000
Vietnam              360.809524

```

Name: problem_solved, Length: 79, dtype: float64

In [23]:

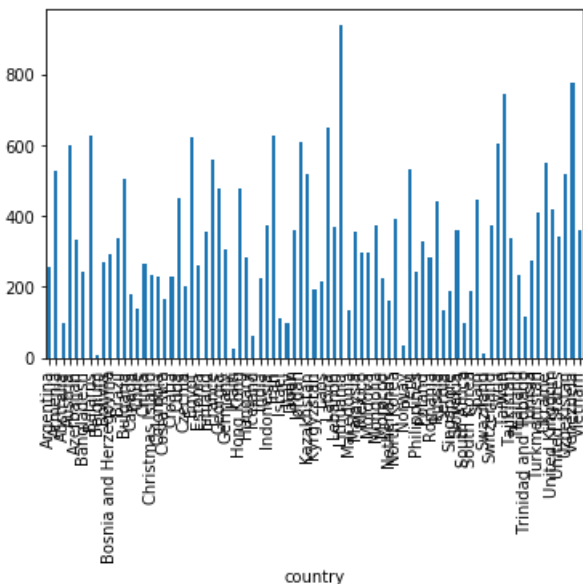
```

#Mean of the solved problemes in graph
df.groupby("country")["problem_solved"].mean().plot(kind='bar')

```

Out[23]:

<matplotlib.axes._subplots.AxesSubplot at 0x1157a2198>



In [26]:

```

df.plot(x='country', y='rating')

```

Out[26]:

<matplotlib.axes._subplots.AxesSubplot at 0x115d472b0>

