Restaurants ratings

September 21, 2020

This data is all about restaurants ratings, style of food served, brand name and the country involved in the process. Popular style of ramen in the country … More this data is based on visualization, how the squarify plot works ? use of groupby operations .

[1]:

```
import pandas as pd
import zipfile
data = pd.read_csv('ramen-ratings.csv')
```

[2]:

```
data.head(10)
```

[2]:

```
   Review #           Brand  \
0      2580       New Touch   
1      2579        Just Way   
2      2578          Nissin   
3      2577         Wei Lih   
4      2576  Ching's Secret   
5      2575   Samyang Foods   
6      2574         Acecook   
7      2573     Ikeda Shoku   
8      2572      Ripe'n'Dry   
9      2571            KOKA   

                                            Variety Style      Country Stars  \
0                        T's Restaurant Tantanmen    Cup        Japan  3.75   
1  Noodles Spicy Hot Sesame Spicy Hot Sesame Guan...  Pack       Taiwan     1   
2                     Cup Noodles Chicken Vegetable   Cup          USA  2.25   
3                     GGE Ramen Snack Tomato Flavor  Pack       Taiwan  2.75   
4                                   Singapore Curry  Pack        India  3.75   
5                            Kimchi song Song Ramen  Pack  South Korea  4.75   
6               Spice Deli Tantan Men With Cilantro   Cup        Japan     4   
7                             Nabeyaki Kitsune Udon  Tray        Japan  3.75   
8                          Hokkaido Soy Sauce Ramen  Pack        Japan  0.25   
9             The Original Spicy Stir-Fried Noodles  Pack    Singapore   2.5   

  Top Ten
0     NaN
1     NaN
```

```
2     NaN
3     NaN
4     NaN
5     NaN
6     NaN
7     NaN
8     NaN
9     NaN
```

[3]:
```
data.describe(include='all')
```

[3]:
```
           Review #   Brand  Variety Style Country Stars Top Ten
count   2580.000000    2580     2580  2578    2580  2580      41
unique          NaN     355     2413     7      38    51      38
top             NaN  Nissin  Chicken  Pack   Japan     4      \n
freq            NaN     381        7  1531     352   384       4
mean    1290.500000     NaN      NaN   NaN     NaN   NaN     NaN
std      744.926171     NaN      NaN   NaN     NaN   NaN     NaN
min        1.000000     NaN      NaN   NaN     NaN   NaN     NaN
25%      645.750000     NaN      NaN   NaN     NaN   NaN     NaN
50%     1290.500000     NaN      NaN   NaN     NaN   NaN     NaN
75%     1935.250000     NaN      NaN   NaN     NaN   NaN     NaN
max     2580.000000     NaN      NaN   NaN     NaN   NaN     NaN
```

[4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2580 entries, 0 to 2579
Data columns (total 7 columns):
Review #    2580 non-null int64
Brand       2580 non-null object
Variety     2580 non-null object
Style       2578 non-null object
Country     2580 non-null object
Stars       2580 non-null object
Top Ten     41 non-null object
dtypes: int64(1), object(6)
memory usage: 141.2+ KB
```

[5]:
```
data['Stars'].value_counts()
```

[5]:
```
4          384
5          369
3.75       350
3.5        326
3          173
3.25       170
```

```
4.25       143
4.5        132
2.75        85
2           68
2.5         67
4.75        64
1.5         37
1.75        27
1           26
0           26
2.25        21
0.5         14
0.25        11
5.0         10
1.25        10
3.50         9
5.00         7
4.00         6
4.3          4
4.0          3
4.50         3
3.8          3
Unrated      3
1.1          2
4.125        2
2.9          2
2.8          2
2.3          2
3.0          2
3.1          2
3.65         1
3.4          1
0.1          1
2.85         1
2.125        1
3.3          1
2.1          1
0.75         1
3.7          1
1.8          1
3.125        1
3.6          1
0.9          1
3.2          1
3.00         1
Name: Stars, dtype: int64
```

If I check the data, stars needs to be converted into integers !!! Rest all are ok and perfect !!! Apart from that unrated value needs to be changed , since there are only three values so they can be put to zero , will not make much difference, or can be given a rating of average 2.5

```
[6]: data['Stars'] = data['Stars'].str.replace('Unrated', '0').astype(float)
```

```
[7]: data['Stars'].value_counts()
```

```
[7]: 4.000    393
     5.000    386
     3.750    350
     3.500    335
     3.000    176
     3.250    170
     4.250    143
     4.500    135
     2.750     85
     2.000     68
     2.500     67
     4.750     64
     1.500     37
     0.000     29
     1.750     27
     1.000     26
     2.250     21
     0.500     14
     0.250     11
     1.250     10
     4.300      4
     3.800      3
     2.900      2
     2.800      2
     3.100      2
     2.300      2
     1.100      2
     4.125      2
     3.650      1
     3.600      1
     3.700      1
     3.400      1
     3.125      1
     2.850      1
     0.100      1
     1.800      1
     3.200      1
     2.100      1
     3.300      1
```

```
2.125      1
0.750      1
0.900      1
Name: Stars, dtype: int64
```

[8]:

```
import squarify
data_rating = data[(data['Stars']>=3)]
data_rating['Stars'].value_counts()
data_rating['Country'].value_counts()
```

[8]:

```
Japan            321
South Korea      273
USA              256
Taiwan           183
Thailand         150
Malaysia         145
China            137
Indonesia        119
Hong Kong        119
Singapore        102
Vietnam           79
UK                45
Philippines       36
Germany           27
Mexico            25
India             24
Australia         15
Canada            15
Nepal             13
Myanmar           12
Netherlands        8
Hungary            8
Bangladesh         7
Pakistan           7
Brazil             5
Colombia           5
Cambodia           5
Fiji               4
Poland             4
Holland            4
Dubai              3
Finland            3
Sarawak            3
Sweden             3
Ghana              2
Estonia            2
United States      1
```

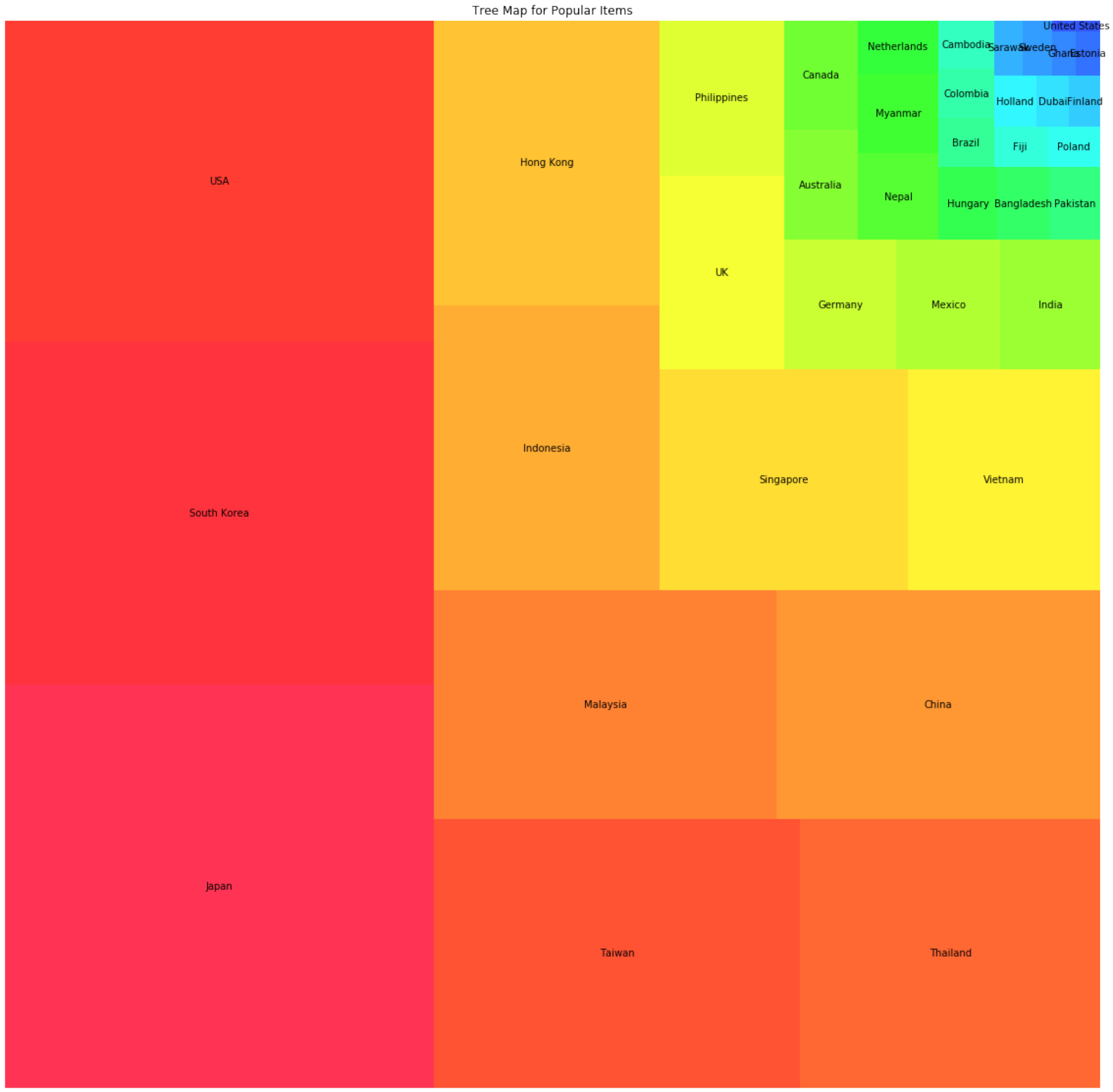
```
Name: Country, dtype: int64
```

If we check the above data, Japan, USA and South korea have the maximum ratings !!! Let's plot them using squarify plots !!! This tool cluster the data in squares , well organised and size according to the detail and information !!

[22]:

```python
import numpy as np
import matplotlib.pyplot as plt
y = data_rating['Country'].value_counts().head(50).to_frame()
y.index

plt.rcParams['figure.figsize'] = (20, 20)
color = plt.cm.gist_rainbow(np.linspace(0, 1, 50))
squarify.plot(sizes = y.values, label = y.index, alpha=.8, color = color)
plt.title('Tree Map for Popular Items')
plt.axis('off')
plt.show()
```



[10]:
```
data.isnull().sum()
```

[10]:
```
Review #       0
Brand          0
Variety        0
Style          2
Country        0
Stars          0
Top Ten     2539
dtype: int64
```

Can delete the Top Ten values !!! as it is more then 90 % null

[11]:
```
data.shape
```

```
[11]: (2580, 7)
```

```
[12]: data.describe()
```

```
[12]:          Review #        Stars
      count  2580.000000  2580.000000
      mean   1290.500000     3.650426
      std     744.926171     1.022358
      min       1.000000     0.000000
      25%     645.750000     3.250000
      50%    1290.500000     3.750000
      75%    1935.250000     4.250000
      max    2580.000000     5.000000
```

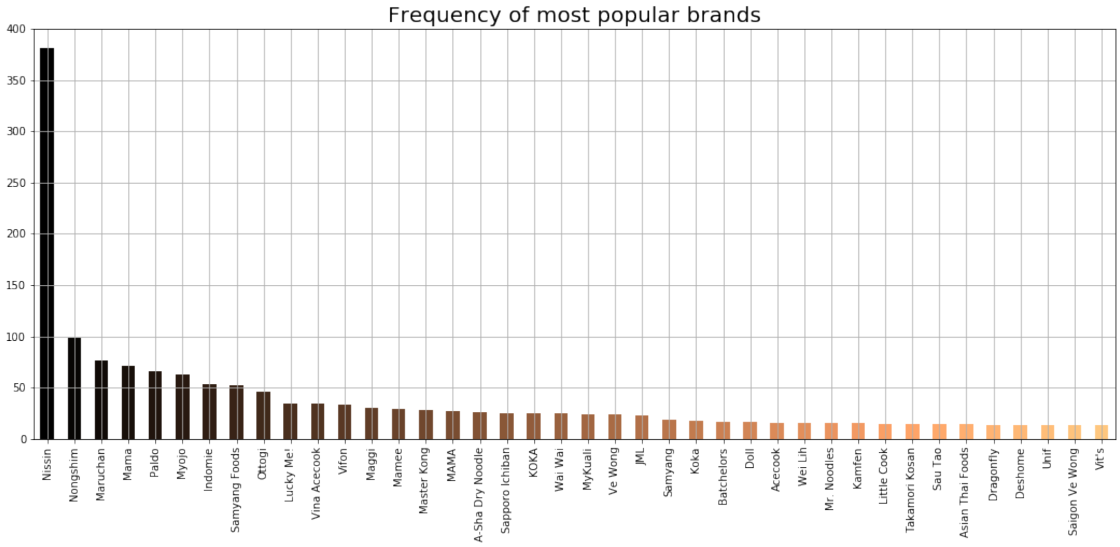
```
[13]: del data['Top Ten']
```

Let's check the brands, value counts in every specific item

```
[14]: data['Brand'].value_counts()
```

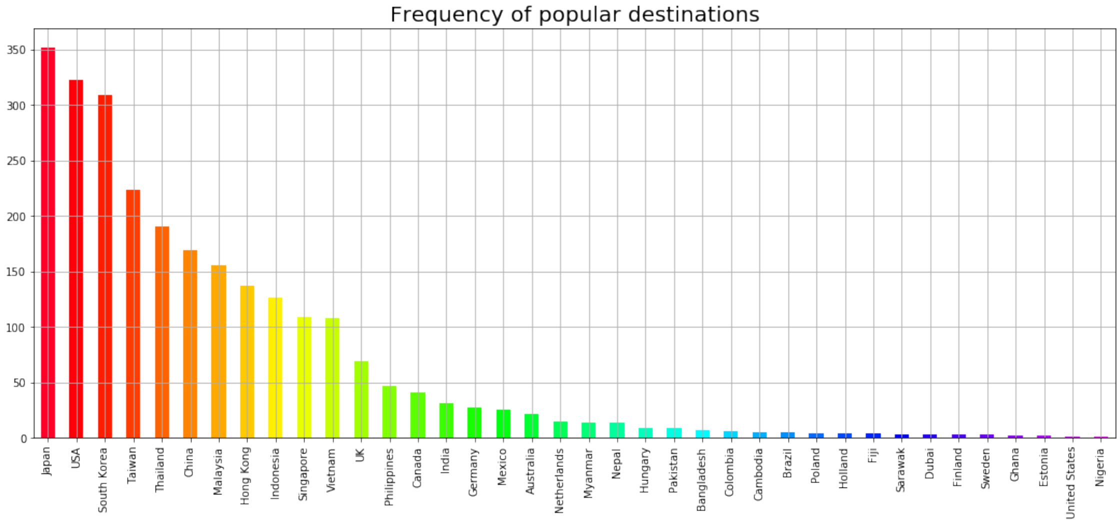
```
[14]: Nissin            381
      Nongshim           98
      Maruchan           76
      Mama               71
      Paldo              66
                       ...
      Qin Zong            1
      Fuji Mengyo         1
      Peyang              1
      President Rice      1
      Q                   1
      Name: Brand, Length: 355, dtype: int64
```

```
[15]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      plt.rcParams['figure.figsize'] = (18, 7)
      color = plt.cm.copper(np.linspace(0, 1, 40))
      data['Brand'].value_counts().head(40).plot.bar(color = color)
      plt.title('Frequency of most popular brands', fontsize = 20)
      plt.xticks(rotation = 90 )
      plt.grid()
      plt.show()
```



[16]:

```
import pandas as pd
import numpy as np
plt.rcParams['figure.figsize'] = (18, 7)
color = plt.cm.gist_rainbow(np.linspace(0, 1, 40))
data['Country'].value_counts().head(40).plot.bar(color = color)
plt.title('Frequency of popular destinations', fontsize = 20)
plt.xticks(rotation = 90 )
plt.grid()
plt.show()
```



Nissin is the most popular brand and the country is Japan !! Let's check which country tops the data in ratings above 4 stars

[17]:
```
data['Stars'].value_counts()
data[(data['Stars']>4)]
```

[17]:
```
      Review #           Brand                           Variety Style  \
5         2575   Samyang Foods            Kimchi song Song Ramen  Pack
10        2570     Tao Kae Noi       Creamy tom Yum Kung Flavour  Pack
11        2569        Yamachan          Yokohama Tonkotsu Shoyu  Pack
12        2568        Nongshim  Mr. Bibim Stir-Fried Kimchi Flavor  Pack
13        2567          Nissin        Deka Buto Kimchi Pork Flavor  Bowl
...        ...             ...                               ...   ...
2535        45         Indomie                   Mi Goreng Sate  Pack
2536        44         Indomie                  Special Chicken  Pack
2552        28          Nissin                     Chikin Ramen  Pack
2557        23          Nissin          Top Ramen Creamy Chicken  Pack
2567        13  Sapporo Ichiban                        Chow Mein  Pack

          Country  Stars
5     South Korea   4.75
10       Thailand   5.00
11            USA   5.00
12    South Korea   4.25
13          Japan   4.50
...           ...    ...
2535    Indonesia   5.00
2536    Indonesia   4.25
2552        Japan   5.00
2557          USA   4.50
2567        Japan   5.00

[734 rows x 6 columns]
```

[18]:
```
grouped = data.groupby('Country')
grouped.size().sort_values(ascending=False)
```

[18]:
```
Country
Japan            352
USA              323
South Korea      309
Taiwan           224
Thailand         191
China            169
Malaysia         156
Hong Kong        137
Indonesia        126
Singapore        109
Vietnam          108
UK                69
```

```
Philippines       47
Canada            41
India             31
Germany           27
Mexico            25
Australia         22
Netherlands       15
Myanmar           14
Nepal             14
Hungary            9
Pakistan           9
Bangladesh         7
Colombia           6
Brazil             5
Cambodia           5
Poland             4
Fiji               4
Holland            4
Dubai              3
Finland            3
Sweden             3
Sarawak            3
Estonia            2
Ghana              2
United States      1
Nigeria            1
dtype: int64
```

[19]:

```python
# Count number of style in each country
ramen_style = data.groupby(['Country','Style']).agg({'Variety':'count'})
ramen_style = ramen_style.reset_index()
ramen_style.head()
```

[19]:

```
      Country Style  Variety
0   Australia   Cup       17
1   Australia  Pack        5
2  Bangladesh  Pack        7
3      Brazil   Cup        2
4      Brazil  Pack        3
```

Here we are grouping country and style together based on the count of variety !!!

Let's create bar chart of countires which have more than 30 products in review.

[20]:

```python
stars = data.groupby(['Country','Brand']).agg({'Stars': ['mean', 'median'],␣
 ↪'Review #': 'count'})
stars = stars.reset_index()
stars.columns = ['Country','Brand','Mean Stars', 'Median Stars', 'Review#']
```

```
stars = stars.sort_values('Median Stars', ascending = False)

# Create new column for label
stars['Country Brand'] = stars['Brand'] + ' (' + stars['Country'] + ')'
stars.head()
```

```
[20]:       Country     Brand  Mean Stars  Median Stars  Review#       Country Brand
      162      Japan  Torishi    5.000000           5.0        1     Torishi (Japan)
      123      Japan   Kimura    5.000000           5.0        1      Kimura (Japan)
      171   Malaysia   CarJEN    4.928571           5.0        7   CarJEN (Malaysia)
      141      Japan   Peyang    5.000000           5.0        1      Peyang (Japan)
      174   Malaysia    Daddy    5.000000           5.0        1    Daddy (Malaysia)
```

```
[21]: stars = data.groupby(['Country','Brand']).agg({'Stars': ['mean', 'max'], 'Review␣
        ↪#': 'size'})
      stars = stars.reset_index()
      stars.columns = ['Country','Brand','Mean Stars', 'Max Stars', 'Review#']
      stars = stars.sort_values('Max Stars', ascending = False)

      # Create new column for label
      stars['Country Brand'] = stars['Brand'] + ' (' + stars['Country'] + ')'
      stars.sort_values('Max Stars', ascending = False)
```

```
[21]:       Country           Brand  Mean Stars  Max Stars  Review#  \
      111      Japan         Daikoku    3.875000        5.0        6
      198   Malaysia           Vit's    4.211538        5.0       13
      377        USA           Myojo    3.625000        5.0        6
      331   Thailand          Nissin    3.808824        5.0       17
      329   Thailand            Mama    3.545690        5.0       58
      ..         ...             ...         ...        ...      ...
      417    Vietnam   Uni-President    0.000000        0.0        1
      300     Taiwan           Tiger    0.000000        0.0        1
      384        USA          Roland    0.000000        0.0        2
      380        USA   One Dish Asia    0.000000        0.0        1
      359        USA  Dr. McDougall's    0.000000        0.0        1

                     Country Brand
      111          Daikoku (Japan)
      198         Vit's (Malaysia)
      377              Myojo (USA)
      331        Nissin (Thailand)
      329          Mama (Thailand)
      ..                       ...
      417  Uni-President (Vietnam)
      300           Tiger (Taiwan)
      384             Roland (USA)
      380      One Dish Asia (USA)
```

```
359    Dr. McDougall's (USA)

[423 rows x 6 columns]
```

In the last two columns, we have grouped data by count and size of reviews of country and brands. Daikoku in Japan has the most reviews … followed by vits in malaysia !!!

```
[ ]:
```