

## Cleaning and exploring the dataset

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: ramen = pd.read_csv('ramen-ratings.csv')
ramen.head()
```

Out[2]:

	Review #	Brand	Variety	Style	Country	Stars	Top Ten
0	2580	New Touch	T's Restaurant Tantanmen	Cup	Japan	3.75	NaN
1	2579	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan...	Pack	Taiwan	1	NaN
2	2578	Nissin	Cup Noodles Chicken Vegetable	Cup	USA	2.25	NaN
3	2577	Wei Lih	GGE Ramen Snack Tomato Flavor	Pack	Taiwan	2.75	NaN
4	2576	Ching's Secret	Singapore Curry	Pack	India	3.75	NaN

```
In [3]: ramen.shape
```

Out[3]: (2580, 7)

```
In [4]: ramen.isna().sum()
```

```
Out[4]: Review #      0
Brand        0
Variety      0
Style         2
Country       0
Stars         0
Top Ten     2539
dtype: int64
```

```
In [5]: # Top Ten values are formatted as a combined string of year and the ranking so we have to split the string and create new columns and drop the old column
new_rank = ramen["Top Ten"].str.split(" ", n = 1, expand = True)
ramen["Year_rank"] = new_rank[0].fillna(0)
ramen["Top10"] = new_rank[1].fillna(0)
ramen.drop(columns = "Top Ten", inplace = True)
```

```
In [100]: #Check the 2NaNs in Style and fill them with "Pack" after looking them up
ramen[pd.isnull(ramen["Style"])]
ramen["Style"].fillna("Pack", inplace = True)
```

```
In [8]: #renaming Review #
ramen = ramen.rename(columns = {'Review #': 'Review_no'})
```

```
In [9]: #drop the rows that were "Unrated"
ramen = ramen.drop(ramen[ramen.Stars == 'Unrated'].index)
```

```
In [10]: rank = []
for i in ramen['Top10']:
    if i != 0:
        i = i.replace("#", "")
    rank.append(int(i))
ramen['Rank'] = rank
```

```
In [11]: ramen = ramen.drop(columns = ['Top10'])
ramen.head()
```

Out[11]:

	Review_no	Brand	Variety	Style	Country	Stars	Year_rank	Rank
0	2580	New Touch	T's Restaurant Tantanmen	Cup	Japan	3.75	0	0
1	2579	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan...	Pack	Taiwan	1	0	0
2	2578	Nissin	Cup Noodles Chicken Vegetable	Cup	USA	2.25	0	0
3	2577	Wei Lih	GGE Ramen Snack Tomato Flavor	Pack	Taiwan	2.75	0	0
4	2576	Ching's Secret		Singapore Curry	Pack	India	3.75	0

```
In [12]: #Check types
ramen = ramen.astype({"Stars": "float"})
ramen.dtypes
```

```
Out[12]: Review_no      int64
Brand          object
Variety        object
Style          object
Country        object
Stars         float64
Year_rank     object
Rank           int64
dtype: object
```

In [102]: #Number of reviews and average rating by ramen style. Pack is the most popular and seem to have great rating.

```
ramen.groupby('Style').agg({'Review_no':'count','Stars':'mean'}).sort_values(by = "Review_no")
```

Out[102]:

Style	Review_no	Stars
Bar	1	5.000000
Can	1	3.500000
Box	6	4.291667
Tray	108	3.545139
Cup	450	3.498500
Bowl	481	3.670686
Pack	1530	3.700033

In [13]: # 10 Countries that have the most reviews

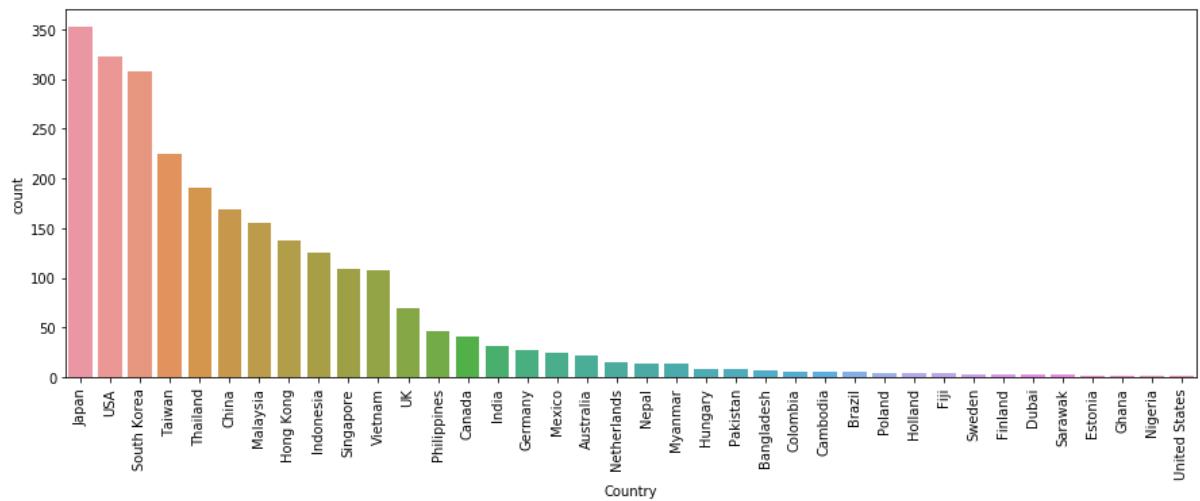
```
review_vol = ramen.groupby(ramen['Country']).agg({'Review_no':'count'}).sort_values('Review_no', ascending = False)
review_vol[:10]
```

Out[13]:

Country	Review_no
Japan	352
USA	323
South Korea	307
Taiwan	224
Thailand	191
China	169
Malaysia	155
Hong Kong	137
Indonesia	126
Singapore	109

```
In [14]: #Plot with seaborn
plt.figure(figsize=(15,5))
chart = sns.countplot(
    data = review_vol,
    x = ramen["Country"],
    order = ramen['Country'].value_counts().index
)
chart.set_xticklabels(chart.get_xticklabels(),rotation=90)
```

```
Out[14]: [Text(0, 0, 'Japan'),
  Text(0, 0, 'USA'),
  Text(0, 0, 'South Korea'),
  Text(0, 0, 'Taiwan'),
  Text(0, 0, 'Thailand'),
  Text(0, 0, 'China'),
  Text(0, 0, 'Malaysia'),
  Text(0, 0, 'Hong Kong'),
  Text(0, 0, 'Indonesia'),
  Text(0, 0, 'Singapore'),
  Text(0, 0, 'Vietnam'),
  Text(0, 0, 'UK'),
  Text(0, 0, 'Philippines'),
  Text(0, 0, 'Canada'),
  Text(0, 0, 'India'),
  Text(0, 0, 'Germany'),
  Text(0, 0, 'Mexico'),
  Text(0, 0, 'Australia'),
  Text(0, 0, 'Netherlands'),
  Text(0, 0, 'Nepal'),
  Text(0, 0, 'Myanmar'),
  Text(0, 0, 'Hungary'),
  Text(0, 0, 'Pakistan'),
  Text(0, 0, 'Bangladesh'),
  Text(0, 0, 'Colombia'),
  Text(0, 0, 'Cambodia'),
  Text(0, 0, 'Brazil'),
  Text(0, 0, 'Poland'),
  Text(0, 0, 'Holland'),
  Text(0, 0, 'Fiji'),
  Text(0, 0, 'Sweden'),
  Text(0, 0, 'Finland'),
  Text(0, 0, 'Dubai'),
  Text(0, 0, 'Sarawak'),
  Text(0, 0, 'Estonia'),
  Text(0, 0, 'Ghana'),
  Text(0, 0, 'Nigeria'),
  Text(0, 0, 'United States')]
```



## Analyzing the labels of different ramen that received 5.0 stars

- Using Natural Language Processing tools to clean, tokenize, lemmatize and vectorize the text in ramen labels
- Visualization by word clouds

```
In [15]: five_star = ramen[ramen["Stars"] == 5].sort_values("Country", ascending = True)
five_star
```

Out[15]:

Review_no	Brand	Variety	Style	Country	Stars	Year_rank	Rank
512	2068	Maggi	Fusian Special Edition Ow... Ow... Spicy Cow M...	Pack	Australia	5.0	0 0
540	2040	Nissin	Cup Noodles Sabor Franhp Com Requeijao	Cup	Brazil	5.0	0 0
721	1859	Mama	Instant Noodles chicken Green Curry Flavour	Pack	Cambodia	5.0	0 0
849	1731	Mama	Instant Noodles Creamy Shrimp Tom Yum Flavour	Pack	Cambodia	5.0	0 0
1031	1549	Sapporo Ichiban	Chow Mein Japanese Style Noodles Yakisoba	Pack	Canada	5.0	0 0
...	...	...	...	...	...	...	...
45	2535	Nissin	Cup Noodles Very Veggie Chicken Flavor Ramen N...	Cup	USA	5.0	0 0
756	1824	Nongshim	Soon Veggie Noodle Soup	Cup	USA	5.0	0 0
716	1864	Nissin	Raoh Ramen Noodle Soup Umami Tonkotsu Flavor	Pack	USA	5.0	0 0
448	2132	Maruchan	Instant Lunch Chipotle Chicken Flavor Ramen No...	Cup	USA	5.0	0 0
226	2354	Vifon	Viet Cuisine Bun Rieu Cua Sour Crab Soup Insta...	Bowl	Vietnam	5.0	0 0

386 rows × 8 columns

```
In [17]: #Combine list of text into 1 big large string
def combine_txt(list_of_txt):
    combined_txt = ' '.join(list_of_txt)
    return combined_txt
```

```
In [18]: df = five_star.groupby("Country").agg(name = ("Variety", lambda x: combine_txt(x)))
df.head()
```

Out[18]:

Country	name
Australia	Fusian Special Edition Owl... Owl... Spicy Cow M...
Brazil	Cup Noodles Sabor Franhp Com Requeijao
Cambodia	Instant Noodles chicken Green Curry Flavour In...
Canada	Chow Mein Japanese Style Noodles Yakisoba Cuis...
China	Emperer Instant Noodles Spicy Chicken Spicy Ki...

```
In [19]: import re
import string

import nltk
nltk.download('stopwords')
stopwords = nltk.corpus.stopwords.words('english')

nltk.download('wordnet')
wn = nltk.WordNetLemmatizer()
```

```
[nltk_data] Downloading package stopwords to /Users/jo/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /Users/jo/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
In [20]: #remove punctuation, numbers, tokenize, lemmatize and apply lowercase
def clean_text(text):
    text = " ".join([word.lower() for word in text if word not in string.punctuation])
    text = re.sub('\w*\d\w*', ' ', text)
    tokens = re.split('\W+', text)
    text = [wn.lemmatize(word) for word in tokens]
    text = " ".join([line for line in text])
    return text
round1 = lambda x: clean_text(x)
```

```
In [27]: df_clean = pd.DataFrame(df.name.apply(round1))
df_clean['country'] = df_clean.index
df_clean = df_clean[['country', "name"]]
#Calculating the length of the labels to determine which ones have more words
df_clean['length'] = [len(i) for i in df_clean.name]
df_clean.sort_values('length', ascending = False)
```

Out[27]:

	Country	country	name	length
Malaysia	Malaysia	penang red tom yum goong noodle taste of malay...		2024
Taiwan	Taiwan	xo sauce noodle rice noodle shallot with brais...		1790
Japan	Japan	cup noodle singapore laksa cup noodle milk chi...		1740
Singapore	Singapore	extra spicy curry instant noodle sens laksa in...		1294
South Korea	South Korea	maesaengytangmyun baked noodle jin jjambpong ...		1114
Hong Kong	Hong Kong	demae ramen spicy seafood flavour instant nood...		995
USA	USA	yokohama tonkotsu shoyu cup noodle very veggie...		962
Thailand	Thailand	shrimp creamy tom yum mi goreng creamy tom yum...		767
Indonesia	Indonesia	mi goreng ala jawa dengan sambal cabe asli ins...		653
China	China	emperor instant noodle spicy chicken spicy kin...		391
Myanmar	Myanmar	seafood tom yum sichet flavour instant noodle ...		147
UK	UK	instant noodle taste sensation creamy shrimp t...		133
Cambodia	Cambodia	instant noodle chicken green curry flavour ins...		87
Canada	Canada	chow mein japanese style noodle yakisoba cui...		75
India	India	atta noodle jhatpat banao befikr khao minute n...		64
Vietnam	Vietnam	viet cuisine bun rieu cua sour crab soup insta...		64
Philippines	Philippines	pancit canton sweet spicy pancit canton toyo m...		50
Australia	Australia	fusian special edition ow ow spicy cow mi goreng		48
Mexico	Mexico	cup noodle gourmet sabor a caldo de camaron		43
Brazil	Brazil	cup noodle sabor franhp com requeijao		37
Nepal	Nepal	preeti instant noodle veg soup base		35
Sarawak	Sarawak	instant kampua dark soy sauce		29
Germany	Germany	soba yakitori chicken		21

```
In [58]: #select the ones that have more ramen labels than others based on the total length of the labels >50 words
df_selected = df_clean[df_clean["length"] >50].sort_values("length", ascending = False)
```

```
In [59]: from sklearn.feature_extraction.text import CountVectorizer
```

In [60]: #Produce document term matrix that includes bigrams

```
cv = CountVectorizer(ngram_range = (1,2))
df_cv = cv.fit_transform(df_selected['name'])
df_dtm = pd.DataFrame(df_cv.toarray(), columns=cv.get_feature_names())
df_dtm.index = df_selected.index
df_dtm
```

Out[60]:

Country	adventure	adventure chicken	aka	aka tonkotsu	ala	ala jawa	aloe	aloe noodle	aloe thin	aloe vera	...	yum mi	n
Malaysia	0	0	0	0	0	0	0	0	0	0	0	...	0
Taiwan	0	0	0	0	0	0	6	3	1	2	...	0	0
Japan	0	0	0	0	0	0	0	0	0	0	0	...	0
Singapore	0	0	0	0	0	0	0	0	0	0	0	...	0
South Korea	0	0	0	0	0	0	0	0	0	0	0	...	0
Hong Kong	0	0	1	1	0	0	0	0	0	0	0	...	0
USA	0	0	0	0	0	0	0	0	0	0	0	...	0
Thailand	0	0	0	0	0	0	0	0	0	0	0	...	1
Indonesia	0	0	0	0	1	1	0	0	0	0	0	...	0
China	0	0	0	0	0	0	0	0	0	0	0	...	0
Myanmar	0	0	0	0	0	0	0	0	0	0	0	...	0
UK	0	0	0	0	0	0	0	0	0	0	0	...	0
Cambodia	0	0	0	0	0	0	0	0	0	0	0	...	0
Canada	1	1	0	0	0	0	0	0	0	0	0	...	0
India	0	0	0	0	0	0	0	0	0	0	0	...	0
Vietnam	0	0	0	0	0	0	0	0	0	0	0	...	0

16 rows × 1652 columns

```
In [61]: desc = df_dtm.transpose()
desc
```

Out[61]:

Country	Malaysia	Taiwan	Japan	Singapore	South Korea	Hong Kong	USA	Thailand	Indonesia	China
adventure	0	0	0	0	0	0	0	0	0	0
adventure chicken	0	0	0	0	0	0	0	0	0	0
aka	0	0	0	0	0	1	0	0	0	0
aka tonkotsu	0	0	0	0	0	1	0	0	0	0
ala	0	0	0	0	0	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
yum sichek	0	0	0	0	0	0	0	0	0	0
zha	0	0	0	0	1	0	0	0	0	0
zha wang	0	0	0	0	1	0	0	0	0	0
zhou	0	1	0	0	0	0	0	0	0	0
zhou meat	0	1	0	0	0	0	0	0	0	0

1652 rows × 16 columns

```
In [62]: #Store top 20 common words in a dictionary where countries are keys and
      lists of words are values
top_dict = {}
for c in desc.columns:
    top = desc[c].sort_values(ascending = False).head(20)
    top_dict[c] = list(zip(top.index,top.values))
top_dict
```

```
Out[62]: {'Malaysia': [('penang', 34),  
                      ('noodle', 28),  
                      ('curry', 25),  
                      ('taste', 15),  
                      ('white', 15),  
                      ('penang white', 14),  
                      ('white curry', 14),  
                      ('malaysia', 11),  
                      ('malaysia penang', 10),  
                      ('curry noodle', 9),  
                      ('tom', 8),  
                      ('yum', 7),  
                      ('red', 7),  
                      ('instant noodle', 7),  
                      ('tom yum', 7),  
                      ('penang red', 7),  
                      ('instant', 7),  
                      ('vermicelli', 6),  
                      ('hokkien', 6),  
                      ('soup', 6)],  
          'Taiwan': [('noodle', 41),  
                      ('sauce', 20),  
                      ('with', 18),  
                      ('flavor', 16),  
                      ('noodle with', 12),  
                      ('oil', 9),  
                      ('sesame', 6),  
                      ('aloe', 6),  
                      ('instant', 6),  
                      ('spicy', 6),  
                      ('sichuan', 5),  
                      ('rice', 5),  
                      ('onion', 5),  
                      ('feast', 4),  
                      ('instant noodle', 4),  
                      ('flavor instant', 4),  
                      ('han feast', 4),  
                      ('han', 4),  
                      ('curry', 4),  
                      ('pork', 4)],  
          'Japan': [('ramen', 21),  
                      ('noodle', 16),  
                      ('cup', 12),  
                      ('yakisoba', 12),  
                      ('cup noodle', 11),  
                      ('flavor', 8),  
                      ('raoh', 8),  
                      ('udon', 7),  
                      ('chikin', 7),  
                      ('tonkotsu', 6),  
                      ('ramen cup', 6),  
                      ('chikin ramen', 6),  
                      ('miso', 5),  
                      ('sauce', 4),  
                      ('shoyu', 4),  
                      ('men', 4),  
                      ('curry', 4),
```

```
('shio', 3),  
('hakata', 3),  
('tonkotsu ramen', 3)],  
'Singapore': [(['noodle', 16),  
('instant', 10),  
('flavour', 9),  
('soup', 9),  
('singapore', 8),  
('spicy', 8),  
('mian', 7),  
('la', 7),  
('la mian', 7),  
('laksa', 7),  
('flavor', 7),  
('instant noodle', 7),  
('curry', 7),  
('creamy', 6),  
('mee', 5),  
('with', 5),  
('crab', 4),  
('creamy soup', 4),  
('crushed', 4),  
('crushed noodle', 4)],  
'South Korea': [(['noodle', 14),  
('ramyun', 6),  
('ramen', 6),  
('rice', 4),  
('seafood', 4),  
('spicy', 4),  
('jinjja', 4),  
('topokki', 3),  
('jjampong', 3),  
('flavor', 3),  
('jjamppong', 3),  
('cheese', 3),  
('bowl', 3),  
('soup', 3),  
('crab', 3),  
('bibim', 3),  
('king', 3),  
('king bowl', 3),  
('with', 2),  
('namja', 2)],  
'Hong Kong': [(['noodle', 23),  
('flavour', 17),  
('instant', 14),  
('instant noodle', 13),  
('flavour instant', 12),  
('demae', 8),  
('ramen', 6),  
('demae ramen', 6),  
('tonkotsu', 5),  
('seafood', 5),  
('seafood flavour', 4),  
('nupasta', 4),  
('tonkotsu flavour', 4),  
('spicy', 4),
```

```
('soup', 3),  
('cup noodle', 3),  
('cup', 3),  
('vermicelli', 3),  
('sauce', 3),  
('sauce flavour', 3)],  
'USA': [('ramen', 17),  
('noodle', 16),  
('flavor', 13),  
('soup', 11),  
('noodle soup', 11),  
('tonkotsu', 6),  
('ramen noodle', 6),  
('flavor ramen', 5),  
('chicken', 5),  
('veggie', 4),  
('spicy', 4),  
('chicken flavor', 3),  
('noodle very', 3),  
('shoyu', 3),  
('beef', 3),  
('very veggie', 3),  
('cup', 3),  
('very', 3),  
('miso', 3),  
('cup noodle', 3)],  
'Thailand': [('noodle', 12),  
('flavour', 12),  
('tom', 11),  
('instant', 11),  
('yum', 9),  
('tom yum', 9),  
('cup', 6),  
('shrimp', 6),  
('instant noodle', 6),  
('creamy', 5),  
('rice', 4),  
('creamy tom', 4),  
('jumbo', 3),  
('noodle tom', 3),  
('yentafo', 3),  
('yentafo tom', 3),  
('instant rice', 3),  
('shrimp creamy', 3),  
('cup noodle', 3),  
('flavour cup', 3)],  
'Indonesia': [('mi', 16),  
('mi goreng', 15),  
('goreng', 15),  
('noodle', 9),  
('fried', 6),  
('jumbo', 4),  
('instant', 4),  
('rasa', 4),  
('fried noodle', 3),  
('local', 3),  
('beef', 3),
```

```
('curly', 3),  
('noodle mi', 3),  
('goreng jumbo', 3),  
('chicken', 3),  
('cup', 3),  
('curly noodle', 3),  
('instant cup', 3),  
('chicken mi', 2),  
('sate', 2)],  
'China': [('spicy', 10),  
('noodle', 9),  
('king', 4),  
('spicy king', 4),  
('beef', 3),  
('spicy chicken', 3),  
('chongqing', 3),  
('chongqing noodle', 3),  
('flavor', 3),  
('noodle spicy', 3),  
('king spicy', 3),  
('chicken', 3),  
('potato', 3),  
('dry noodle', 2),  
('dry', 2),  
('noodle burning', 2),  
('chicken spicy', 2),  
('noodle chongqing', 2),  
('pork', 2),  
('burning dry', 2)],  
'Myanmar': [('flavour', 4),  
('instant', 3),  
('flavour instant', 3),  
('noodle', 3),  
('instant noodle', 3),  
('milk flavour', 2),  
('tom', 2),  
('coconut milk', 2),  
('coconut', 2),  
('tom yum', 2),  
('yum', 2),  
('noodle coconut', 2),  
('milk', 2),  
('sichek flavour', 1),  
('yum flavour', 1),  
('noodle shrimp', 1),  
('sichek', 1),  
('seafood tom', 1),  
('creamy tom', 1),  
('shrimp', 1)],  
'UK': [('street', 2),  
('taste', 1),  
('authentic street', 1),  
('kong street', 1),  
('monk vegetable', 1),  
('monk', 1),  
('kong', 1),  
('shaolin', 1),
```

```
('shaolin monk', 1),
('authentic', 1),
('shrimp tom', 1),
('street food', 1),
('noodle', 1),
('street beef', 1),
('instant noodle', 1),
('taste sensation', 1),
('shrimp', 1),
('satay', 1),
('instant', 1),
('creamy', 1)],
'Cambodia': [('instant', 2),
('noodle', 2),
('instant noodle', 2),
('flavour', 2),
('chicken', 1),
('creamy', 1),
('creamy shrimp', 1),
('shrimp', 1),
('tom', 1),
('curry flavour', 1),
('flavour instant', 1),
('tom yum', 1),
('noodle creamy', 1),
('shrimp tom', 1),
('chicken green', 1),
('green curry', 1),
('yum', 1),
('yum flavour', 1),
('green', 1),
('curry', 1)],
'Canada': [('adventure', 1),
('yakisoba cuisine', 1),
('mein japanese', 1),
('mein', 1),
('noodle yakisoba', 1),
('chicken', 1),
('chicken pho', 1),
('pho soup', 1),
('chow', 1),
('chow mein', 1),
('style noodle', 1),
('style', 1),
('cuisine', 1),
('cuisine adventure', 1),
('japanese style', 1),
('japanese', 1),
('soup', 1),
('yakisoba', 1),
('pho', 1),
('noodle', 1)],
'India': [('noodle', 2),
('khao minute', 1),
('minute noodle', 1),
('minute', 1),
('banao', 1),
```

```
('banao befikr', 1),  
('noodle jhatpat', 1),  
('spicy', 1),  
('atta', 1),  
('atta noodle', 1),  
('befikr khao', 1),  
('khao', 1),  
('noodle masala', 1),  
('masala', 1),  
('masala spicy', 1),  
('befikr', 1),  
('jhatpat banao', 1),  
('jhatpat', 1),  
('ippin', 0),  
('instant rice', 0)],  
'Vietnam': [('soup instant', 1),  
('instant rice', 1),  
('cua', 1),  
('soup', 1),  
('crab', 1),  
('viet cuisine', 1),  
('viet', 1),  
('cuisine', 1),  
('instant', 1),  
('cuisine bun', 1),  
('bun rieu', 1),  
('rice', 1),  
('rieu cua', 1),  
('sour crab', 1),  
('sour', 1),  
('rieu', 1),  
('bun', 1),  
('vermicelli', 1),  
('crab soup', 1),  
('rice vermicelli', 1)]}
```

```
In [63]: #Print top common words in ramen labels by country
for category, top_words in top_dict.items():
    print(category)
    print(', '.join([word for word, count in top_words[0:15]]))
    print('---')
```

**Malaysia**

penang, noodle, curry, taste, white, penang white, white curry, malaysi a, malaysia penang, curry noodle, tom, yum, red, instant noodle, tom yum

---

**Taiwan**

noodle, sauce, with, flavor, noodle with, oil, sesame, aloe, instant, spicy, sichuan, rice, onion, feast, instant noodle

---

**Japan**

ramen, noodle, cup, yakisoba, cup noodle, flavor, raoh, udon, chikin, tonkotsu, ramen cup, chikin ramen, miso, sauce, shoyu

---

**Singapore**

noodle, instant, flavour, soup, singapore, spicy, mian, la, la mian, la ksa, flavor, instant noodle, curry, creamy, mee

---

**South Korea**

noodle, ramyun, ramen, rice, seafood, spicy, jinjja, topokki, jjampong, flavor, jjamppong, cheese, bowl, soup, crab

---

**Hong Kong**

noodle, flavour, instant, instant noodle, flavour instant, demae, ramen, demae ramen, tonkotsu, seafood, seafood flavour, nupasta, tonkotsu flavour, spicy, soup

---

**USA**

ramen, noodle, flavor, soup, noodle soup, tonkotsu, ramen noodle, flavor ramen, chicken, veggie, spicy, chicken flavor, noodle very, shoyu, beef

---

**Thailand**

noodle, flavour, tom, instant, yum, tom yum, cup, shrimp, instant noodle, creamy, rice, creamy tom, jumbo, noodle tom, yentafo

---

**Indonesia**

mi, mi goreng, goreng, noodle, fried, jumbo, instant, rasa, fried noodle, local, beef, curly, noodle mi, goreng jumbo, chicken

---

**China**

spicy, noodle, king, spicy king, beef, spicy chicken, chongqing, chongqing noodle, flavor, noodle spicy, king spicy, chicken, potato, dry noodle, dry

---

**Myanmar**

flavour, instant, flavour instant, noodle, instant noodle, milk flavour, tom, coconut milk, coconut, tom yum, yum, noodle coconut, milk, sichuk flavour, yum flavour

---

**UK**

street, taste, authentic street, kong street, monk vegetable, monk, kong, shaolin, shaolin monk, authentic, shrimp tom, street food, noodle, street beef, instant noodle

---

**Cambodia**

instant, noodle, instant noodle, flavour, chicken, creamy, creamy shrimp, shrimp, tom, curry flavour, flavour instant, tom yum, noodle creamy,

shrimp tom, chicken green

---

Canada

adventure, yakisoba cuisine, mein japanese, mein, noodle yakisoba, chicken, chicken pho, pho soup, chow, chow mein, style noodle, style, cuisine, cuisine adventure, japanese style

---

India

noodle, khao minute, minute noodle, minute, banao, banao befikr, noodle jhatpat, spicy, atta, atta noodle, befikr khao, khao, noodle masala, masala, masala spicy

---

Vietnam

soup instant, instant rice, cua, soup, crab, viet cuisine, viet, cuisine, instant, cuisine bun, bun rieu, rice, rieu cua, sour crab, sour

---

```
In [64]: # Look at the most common top words --> add them to the stop word list
from collections import Counter

# the top 20 words for each category
words = []
for category in desc.columns:
    top = [word for (word, count) in top_dict[category]]
    for t in top:
        words.append(t)

words
```

```
Out[64]: ['penang',
 'noodle',
 'curry',
 'taste',
 'white',
 'penang white',
 'white curry',
 'malaysia',
 'malaysia penang',
 'curry noodle',
 'tom',
 'yum',
 'red',
 'instant noodle',
 'tom yum',
 'penang red',
 'instant',
 'vermicelli',
 'hokkien',
 'soup',
 'noodle',
 'sauce',
 'with',
 'flavor',
 'noodle with',
 'oil',
 'sesame',
 'aloe',
 'instant',
 'spicy',
 'sichuan',
 'rice',
 'onion',
 'feast',
 'instant noodle',
 'flavor instant',
 'han feast',
 'han',
 'curry',
 'pork',
 'ramen',
 'noodle',
 'cup',
 'yakisoba',
 'cup noodle',
 'flavor',
 'raoh',
 'udon',
 'chikin',
 'tonkotsu',
 'ramen cup',
 'chikin ramen',
 'miso',
 'sauce',
 'shoyu',
 'men',
 'curry',
```

```
'shio',
'hakata',
'tonkotsu ramen',
'noodle',
'instant',
'flavour',
'soup',
'singapore',
'spicy',
'mian',
'la',
'la mian',
'laksa',
'flavor',
'instant noodle',
'curry',
'creamy',
'mee',
'with',
'crab',
'creamy soup',
'crushed',
'crushed noodle',
'noodle',
'ramyun',
'ramen',
'rice',
'seafood',
'spicy',
'jinjja',
'topokki',
'jjampong',
'flavor',
'jjamppong',
'cheese',
'bowl',
'soup',
'crab',
'bibim',
'king',
'king bowl',
'with',
'namja',
'noodle',
'flavour',
'instant',
'instant noodle',
'flavour instant',
'demae',
'ramen',
'demae ramen',
'tonkotsu',
'seafood',
'seafood flavour',
'nupasta',
'tonkotsu flavour',
'spicy',
```

```
'soup',
'cup noodle',
'cup',
'vermicelli',
'sauce',
'sauce flavour',
'ramen',
'noodle',
'flavor',
'soup',
'noodle soup',
'tonkotsu',
'ramen noodle',
'flavor ramen',
'chicken',
'veggie',
'spicy',
'chicken flavor',
'noodle very',
'shoyu',
'beef',
'very veggie',
'cup',
'very',
'miso',
'cup noodle',
'noodle',
'flavour',
'tom',
'instant',
'yum',
'tom yum',
'cup',
'shrimp',
'instant noodle',
'creamy',
'rice',
'creamy tom',
'jumbo',
'noodle tom',
'yentafo',
'yentafo tom',
'instant rice',
'shrimp creamy',
'cup noodle',
'flavour cup',
'mi',
'mi goreng',
'goreng',
'noodle',
'fried',
'jumbo',
'instant',
'rasa',
'fried noodle',
'local',
'veef',
```

'curly',  
'noodle mi',  
'goreng jumbo',  
'chicken',  
'cup',  
'curly noodle',  
'instant cup',  
'chicken mi',  
'sate',  
'spicy',  
'noodle',  
'king',  
'spicy king',  
'beef',  
'spicy chicken',  
'chongqing',  
'chongqing noodle',  
'flavor',  
'noodle spicy',  
'king spicy',  
'chicken',  
'potato',  
'dry noodle',  
'dry',  
'noodle burning',  
'chicken spicy',  
'noodle chongqing',  
'pork',  
'burning dry',  
'flavour',  
'instant',  
'flavour instant',  
'noodle',  
'instant noodle',  
'milk flavour',  
'tom',  
'coconut milk',  
'coconut',  
'tom yum',  
'yum',  
'noodle coconut',  
'milk',  
'sichek flavour',  
'yum flavour',  
'noodle shrimp',  
'sichek',  
'seafood tom',  
'creamy tom',  
'shrimp',  
'street',  
'taste',  
'authentic street',  
'kong street',  
'monk vegetable',  
'monk',  
'kong',  
'shaolin',

```
'shaolin monk',
'authentic',
'shrimp tom',
'street food',
'noodle',
'street beef',
'instant noodle',
'taste sensation',
'shrimp',
'satay',
'instant',
'creamy',
'instant',
'noodle',
'instant noodle',
'flavour',
'chicken',
'creamy',
'creamy shrimp',
'shrimp',
'tom',
'curry flavour',
'flavour instant',
'tom yum',
'noodle creamy',
'shrimp tom',
'chicken green',
'green curry',
'yum',
'yum flavour',
'green',
'curry',
'adventure',
'yakisoba cuisine',
'mein japanese',
'mein',
'noodle yakisoba',
'chicken',
'chicken pho',
'pho soup',
'chow',
'chow mein',
'style noodle',
'style',
'cuisine',
'cuisine adventure',
'japanese style',
'japanese',
'soup',
'yakisoba',
'pho',
'noodle',
'noodle',
'khao minute',
'minute noodle',
'minute',
'banao',
```

```
'banao befikr',
'noodle jhatpat',
'spicy',
'atta',
'atta noodle',
'befikr khao',
'khao',
'noodle masala',
'masala',
'masala spicy',
'befikr',
'jhatpat banao',
'jhatpat',
'ippin',
'instant rice',
'soup instant',
'instant rice',
'cua',
'soup',
'crab',
'vet cuisine',
'vet',
'cuisine',
'instant',
'cuisine bun',
'bun rieu',
'rice',
'rieu cua',
'sour crab',
'sour',
'rieu',
'bun',
'vermicelli',
'crab soup',
'rice vermicelli']
```

```
In [65]: Counter(words).most_common()
```

```
Out[65]: [('noodle', 15),  
          ('instant', 10),  
          ('instant noodle', 8),  
          ('soup', 7),  
          ('spicy', 7),  
          ('flavor', 6),  
          ('curry', 5),  
          ('cup', 5),  
          ('flavour', 5),  
          ('chicken', 5),  
          ('tom', 4),  
          ('yum', 4),  
          ('tom yum', 4),  
          ('rice', 4),  
          ('ramen', 4),  
          ('cup noodle', 4),  
          ('creamy', 4),  
          ('shrimp', 4),  
          ('vermicelli', 3),  
          ('sauce', 3),  
          ('with', 3),  
          ('tonkotsu', 3),  
          ('crab', 3),  
          ('flavour instant', 3),  
          ('beef', 3),  
          ('instant rice', 3),  
          ('taste', 2),  
          ('pork', 2),  
          ('yakisoba', 2),  
          ('miso', 2),  
          ('shoyu', 2),  
          ('seafood', 2),  
          ('king', 2),  
          ('creamy tom', 2),  
          ('jumbo', 2),  
          ('yum flavour', 2),  
          ('shrimp tom', 2),  
          ('cuisine', 2),  
          ('penang', 1),  
          ('white', 1),  
          ('penang white', 1),  
          ('white curry', 1),  
          ('malaysia', 1),  
          ('malaysia penang', 1),  
          ('curry noodle', 1),  
          ('red', 1),  
          ('penang red', 1),  
          ('hokkien', 1),  
          ('noodle with', 1),  
          ('oil', 1),  
          ('sesame', 1),  
          ('aloe', 1),  
          ('sichuan', 1),  
          ('onion', 1),  
          ('feast', 1),  
          ('flavor instant', 1),  
          ('han feast', 1),
```

```
('han', 1),  
('raoh', 1),  
('udon', 1),  
('chikin', 1),  
('ramen cup', 1),  
('chikin ramen', 1),  
('men', 1),  
('shio', 1),  
('hakata', 1),  
('tonkotsu ramen', 1),  
('singapore', 1),  
('mian', 1),  
('la', 1),  
('la mian', 1),  
('laksa', 1),  
('mee', 1),  
('creamy soup', 1),  
('crushed', 1),  
('crushed noodle', 1),  
('ramyun', 1),  
('jinjja', 1),  
('topokki', 1),  
('jjampong', 1),  
('jjamppong', 1),  
('cheese', 1),  
('bowl', 1),  
('bibim', 1),  
('king bowl', 1),  
('namja', 1),  
('demae', 1),  
('demae ramen', 1),  
('seafood flavour', 1),  
('nupasta', 1),  
('tonkotsu flavour', 1),  
('sauce flavour', 1),  
('noodle soup', 1),  
('ramen noodle', 1),  
('flavor ramen', 1),  
('veggie', 1),  
('chicken flavor', 1),  
('noodle very', 1),  
('very veggie', 1),  
('very', 1),  
('noodle tom', 1),  
('yentafo', 1),  
('yentafo tom', 1),  
('shrimp creamy', 1),  
('flavour cup', 1),  
('mi', 1),  
('mi goreng', 1),  
('goreng', 1),  
('fried', 1),  
('rasa', 1),  
('fried noodle', 1),  
('local', 1),  
('curly', 1),  
('noodle mi', 1),
```

```
('goreng jumbo', 1),
('curly noodle', 1),
('instant cup', 1),
('chicken mi', 1),
('sate', 1),
('spicy king', 1),
('spicy chicken', 1),
('chongqing', 1),
('chongqing noodle', 1),
('noodle spicy', 1),
('king spicy', 1),
('potato', 1),
('dry noodle', 1),
('dry', 1),
('noodle burning', 1),
('chicken spicy', 1),
('noodle chongqing', 1),
('burning dry', 1),
('milk flavour', 1),
('coconut milk', 1),
('coconut', 1),
('noodle coconut', 1),
('milk', 1),
('sichek flavour', 1),
('noodle shrimp', 1),
('sichek', 1),
('seafood tom', 1),
('street', 1),
('authentic street', 1),
('kong street', 1),
('monk vegetable', 1),
('monk', 1),
('kong', 1),
('shaolin', 1),
('shaolin monk', 1),
('authentic', 1),
('street food', 1),
('street beef', 1),
('taste sensation', 1),
('satay', 1),
('creamy shrimp', 1),
('curry flavour', 1),
('noodle creamy', 1),
('chicken green', 1),
('green curry', 1),
('green', 1),
('adventure', 1),
('yakisoba cuisine', 1),
('mein japanese', 1),
('mein', 1),
('noodle yakisoba', 1),
('chicken pho', 1),
('pho soup', 1),
('chow', 1),
('chow mein', 1),
('style noodle', 1),
('style', 1),
```

```
('cuisine adventure', 1),  
('japanese style', 1),  
('japanese', 1),  
('pho', 1),  
('khao minute', 1),  
('minute noodle', 1),  
('minute', 1),  
('banao', 1),  
('banao befikr', 1),  
('noodle jhatpat', 1),  
('atta', 1),  
('atta noodle', 1),  
('befikr khao', 1),  
('khao', 1),  
('noodle masala', 1),  
('masala', 1),  
('masala spicy', 1),  
('befikr', 1),  
('jhatpat banao', 1),  
('jhatpat', 1),  
('ippin', 1),  
('soup instant', 1),  
('cua', 1),  
('viet cuisine', 1),  
('viet', 1),  
('cuisine bun', 1),  
('bun rieu', 1),  
('rieu cua', 1),  
('sour crab', 1),  
('sour', 1),  
('rieu', 1),  
('bun', 1),  
('crab soup', 1),  
('rice vermicelli', 1)]
```

```
In [84]: #Add more stop words that are very common words and appear more than 5 times
add_stop_words = [word for word, count in Counter(words).most_common() if count > 5]
add_stop_words.extend(['ramen', 'cup', 'flavour', 'taste'])
add_stop_words.extend(list(df_selected.index))
add_stop_words = [word.lower() for word in add_stop_words]
add_stop_words
```

```
Out[84]: ['noodle',
'instant',
'instant noodle',
'soup',
'spicy',
'flavor',
'ramen',
'cup',
'flavour',
'taste',
'malaysia',
'taiwan',
'japan',
'singapore',
'south korea',
'hong kong',
'usa',
'thailand',
'indonesia',
'china',
'myanmar',
'uk',
'cambodia',
'canada',
'india',
'vetnam']
```

```
In [85]: from sklearn.feature_extraction import text
```

```
In [86]: stop_words = text.ENGLISH_STOP_WORDS.union(add_stop_words)

# Recreate document-term matrix without the updated stop words
cv1 = CountVectorizer(stop_words=stop_words)
df_cv1 = cv1.fit_transform(df_selected.name)
df_stop = pd.DataFrame(df_cv1.toarray(), columns=cv1.get_feature_names())
df_stop.index = df_selected.index
df_stop
```

/opt/anaconda3/lib/python3.7/site-packages/sklearn/feature\_extraction/text.py:300: UserWarning: Your stop\_words may be inconsistent with your preprocessing. Tokenizing the stop words generated tokens ['hong', 'kong', 'korea', 'south'] not in stop\_words.  
'stop\_words.' % sorted(inconsistent))

Out[86]:

Country	adventure	aka	ala	aloe	angry	arrabiata	artificial	artificially	asam	asia	...	yakil
Malaysia	0	0	0	0	0	0	0	0	2	0	...	
Taiwan	0	0	0	6	0	0	0	0	0	0	0	...
Japan	0	0	0	0	0	0	1	0	0	0	0	...
Singapore	0	0	0	0	1	0	0	0	0	0	0	...
South Korea	0	0	0	0	0	0	0	0	0	0	0	...
Hong Kong	0	1	0	0	0	0	1	0	0	0	0	...
USA	0	0	0	0	0	1	0	1	0	1	0	...
Thailand	0	0	0	0	0	0	0	0	0	0	0	...
Indonesia	0	0	1	0	0	0	0	0	0	0	0	...
China	0	0	0	0	0	0	0	0	0	0	0	...
Myanmar	0	0	0	0	0	0	0	0	0	0	0	...
UK	0	0	0	0	0	0	0	0	0	0	0	...
Cambodia	0	0	0	0	0	0	0	0	0	0	0	...
Canada	1	0	0	0	0	0	0	0	0	0	0	...
India	0	0	0	0	0	0	0	0	0	0	0	...
Vietnam	0	0	0	0	0	0	0	0	0	0	0	...

16 rows × 436 columns

```
In [87]: from wordcloud import WordCloud
```

```
In [88]: wc = WordCloud(stopwords=stop_words, background_color="white", colormap="Dark2",
max_font_size=150, random_state=42)
```

```
In [91]: import matplotlib.pyplot as plt

plt.rcParams['figure.figsize'] = [70,70]

names = df_selected.country

# Create subplots for each comedian
for index, category in enumerate(desc.columns):
    wc.generate(df_selected.name[category])

    plt.subplot(6,3, index+1)
    plt.imshow(wc, interpolation="bilinear")
    plt.axis("off")
    plt.title(names[index], fontsize = 50)

plt.show()
```



These word clouds help us visualize the main ingredients, the style and cuisine of different 5.0 star ramen labels from different country. The order of these word clouds also reflects the number/quantity of 5.0 star ramen that each country has.