Ramen Ratings

IST 652

Final Project Presentation

Jeffrey Chao

Data Sources

- The structured data used in this assignment was the Ramen Ratings data on Kaggle.com. This dataset is taken from the product review site theramentrater.com and contains the Brand, Variety, Country, Style, Rating, and URL for each of the ramen reviews made on the website.

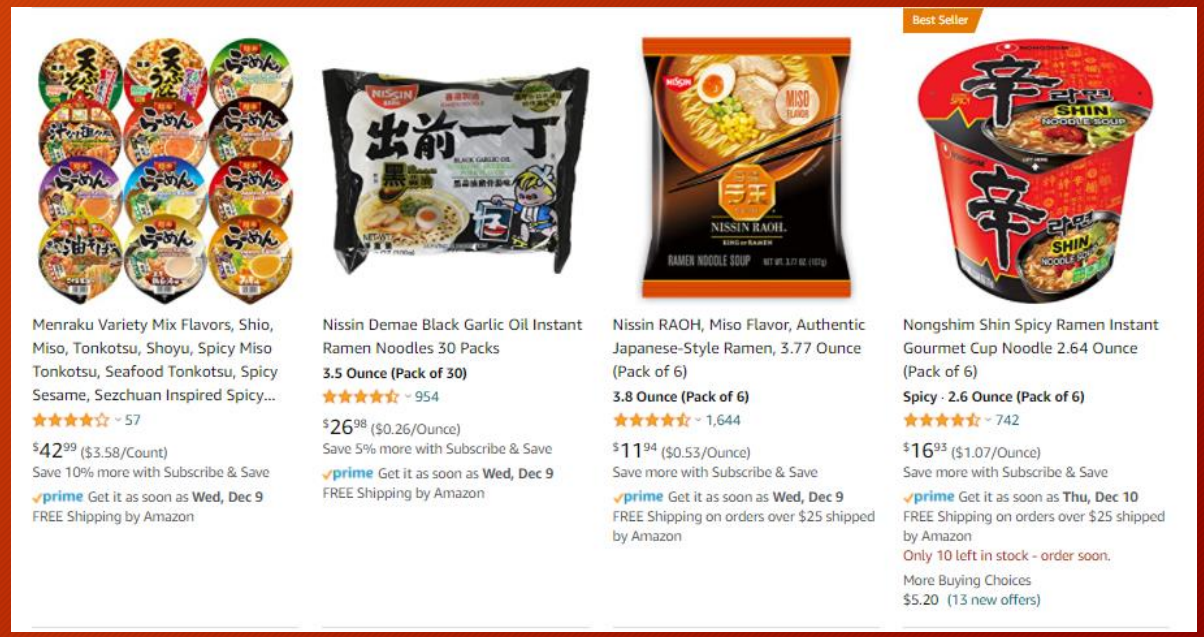
- The unstructured data was scraped from Amazon using the BeautifulSoup and Selenium packages.

```
In [5]: ramendf
```

Out[5]:

	ID	URL	Brand	Variety	Style	Country	Stars
0	3473	https://www.theramenrater.com/2020/04/05/3473-...	MyKuali	White Fish Broth Noodle	Pack	Malaysia	5.00
1	3472	https://www.theramenrater.com/2020/04/05/3472-...	MyKuali	Penang White Curry Noodle (New Recipe)	Pack	Malaysia	5.00
2	3471	https://www.theramenrater.com/2020/04/05/3471-...	Ve Wong	Instant Oriental Noodles Soup Chinese Herb - G...	Pack	Taiwan	3.75
3	3470	https://www.theramenrater.com/2020/04/04/3470-...	Myojo	Ippeichan Rich & Sweet Yakisoba	Tray	Japan	5.00
4	3469	https://www.theramenrater.com/2020/04/03/3469-...	Mykuali	Penang Hokkien Prawn Rice Vermicelli Soup	Bowl	Malaysia	5.00
...	...	...	...	...	...	...	...
3468	5	https://www.theramenrater.com/2009/06/22/vifon...	Vifon	Hu Tiu Nam Vang ["Phnom Penh" style] Asian Sty...	Bowl	Vietnam	3.50
3469	4	https://www.theramenrater.com/2009/06/22/wai-w...	Wai Wai	Oriental Style Instant Noodles	Pack	Thailand	1.00
3470	3	https://www.theramenrater.com/2009/06/22/wai-w...	Wai Wai	Tom Yum Shrimp	Pack	Thailand	2.00
3471	2	https://www.theramenrater.com/2009/06/22/wai-w...	Wai Wai	Tom Yum Chili Flavor	Pack	Thailand	2.00
3472	1	https://www.theramenrater.com/2009/06/22/westb...	Westbrae	Miso Ramen	Pack	United States	0.50

3458 rows × 7 columns



Data Exploration and Cleaning (pt.1)

Structured Data

- Import from .csv
- Drop rows with missing data
- Typos/inconsistencies in the Country column

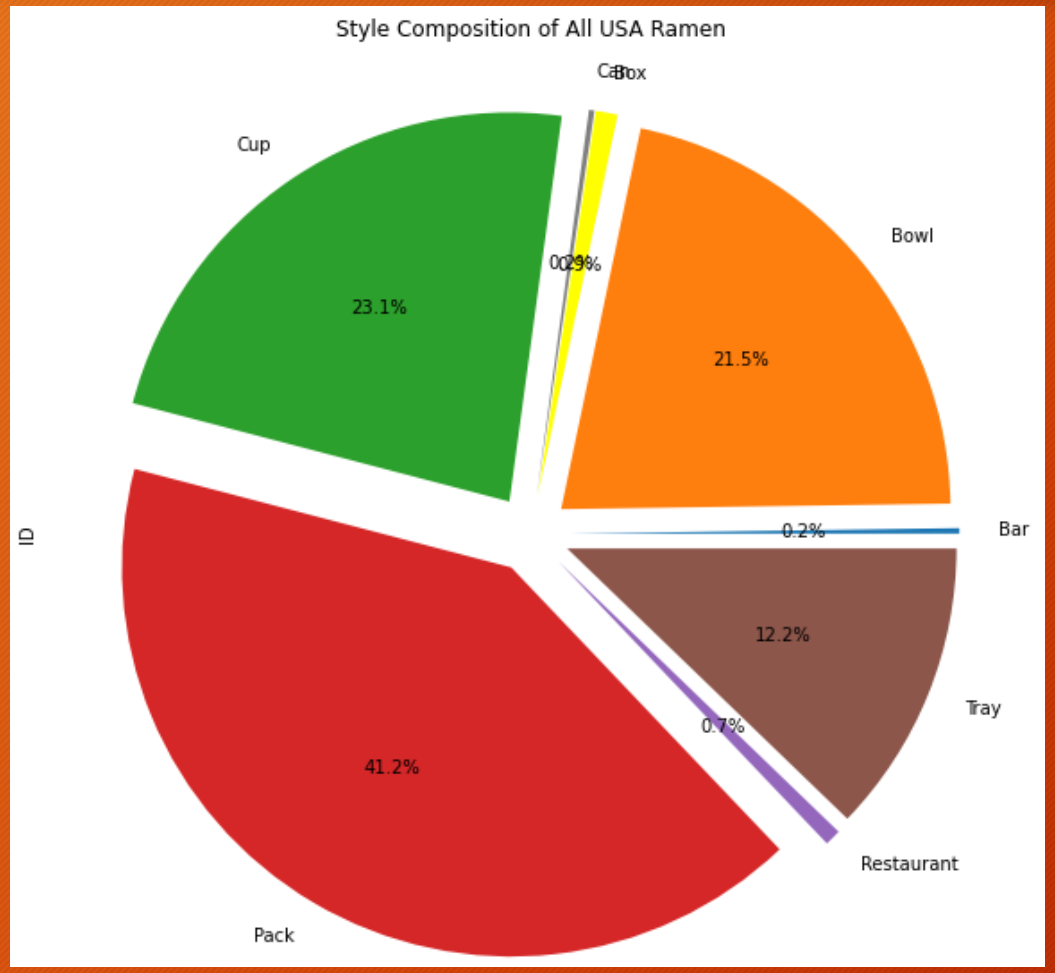
```
ramendf = ramendf.replace({'Country': {'Souh Korea': 'South Korea',
                                       'USA': 'United States',
                                       'UK': 'United Kingdom',
                                       'Phlippines': 'Philippines',
                                       'Sarawak': 'Malaysia'}})
```

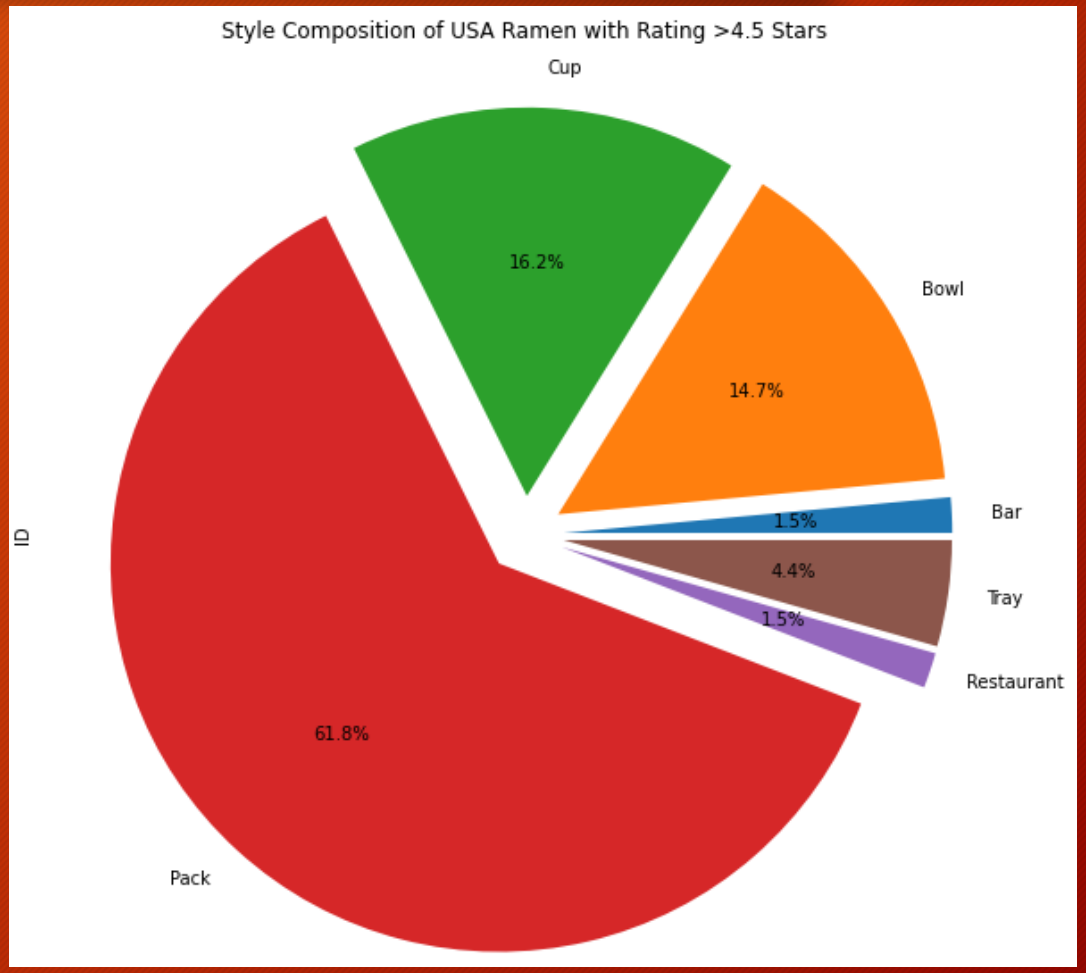
Data Exploration and Cleaning (pt.2)

Unstructured Data

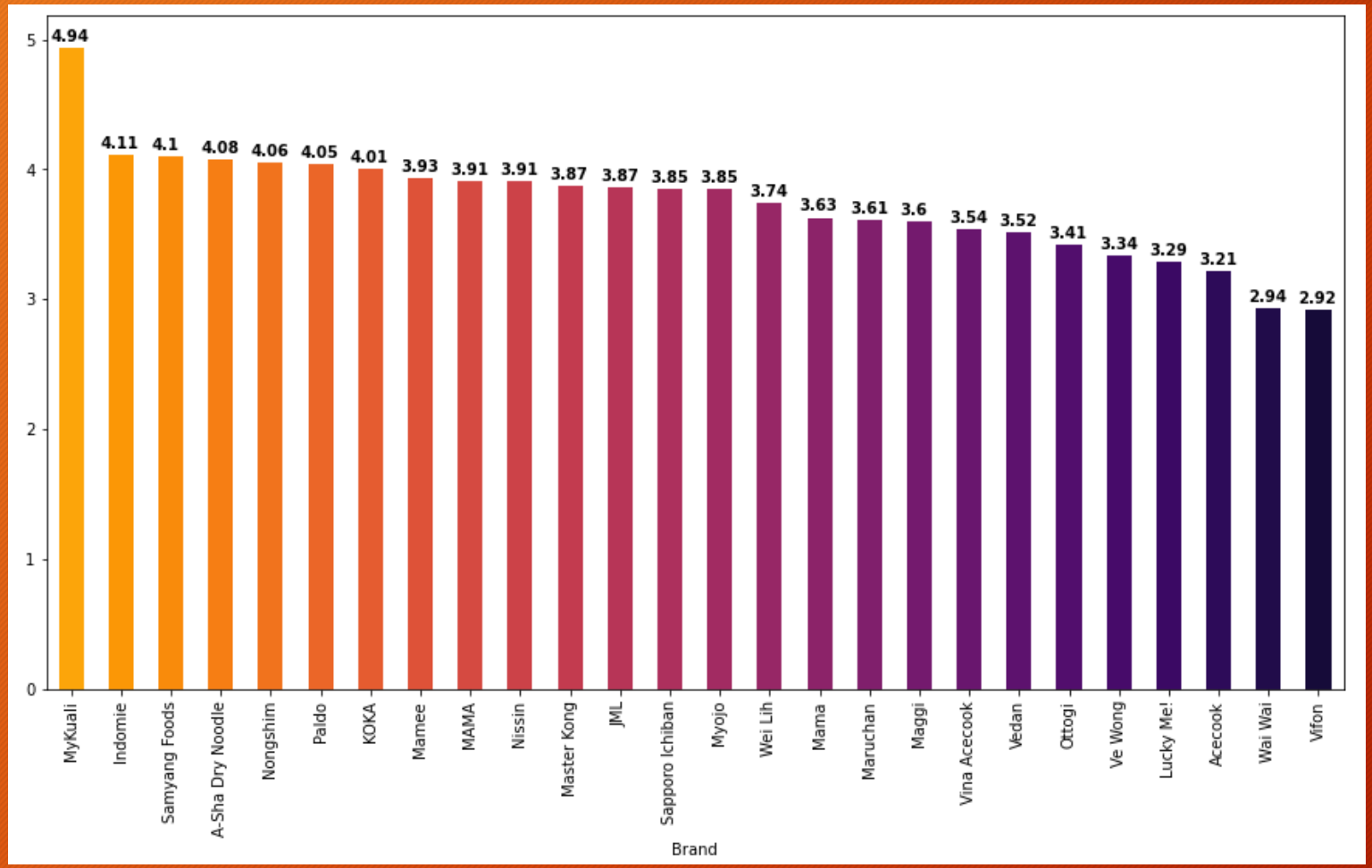
- Generate URL for desired search results
- Load Amazon search page using Selenium
- Pull and parse the HTML through BeautifulSoup to grab relevant information
- Process “product descriptions” so that Brand and Variety are in individual columns
- Manually collect information that was not included in text
- Format information to match the structured data

Question 1:
“Which style of ramen manufactured in the USA is the tastiest?”

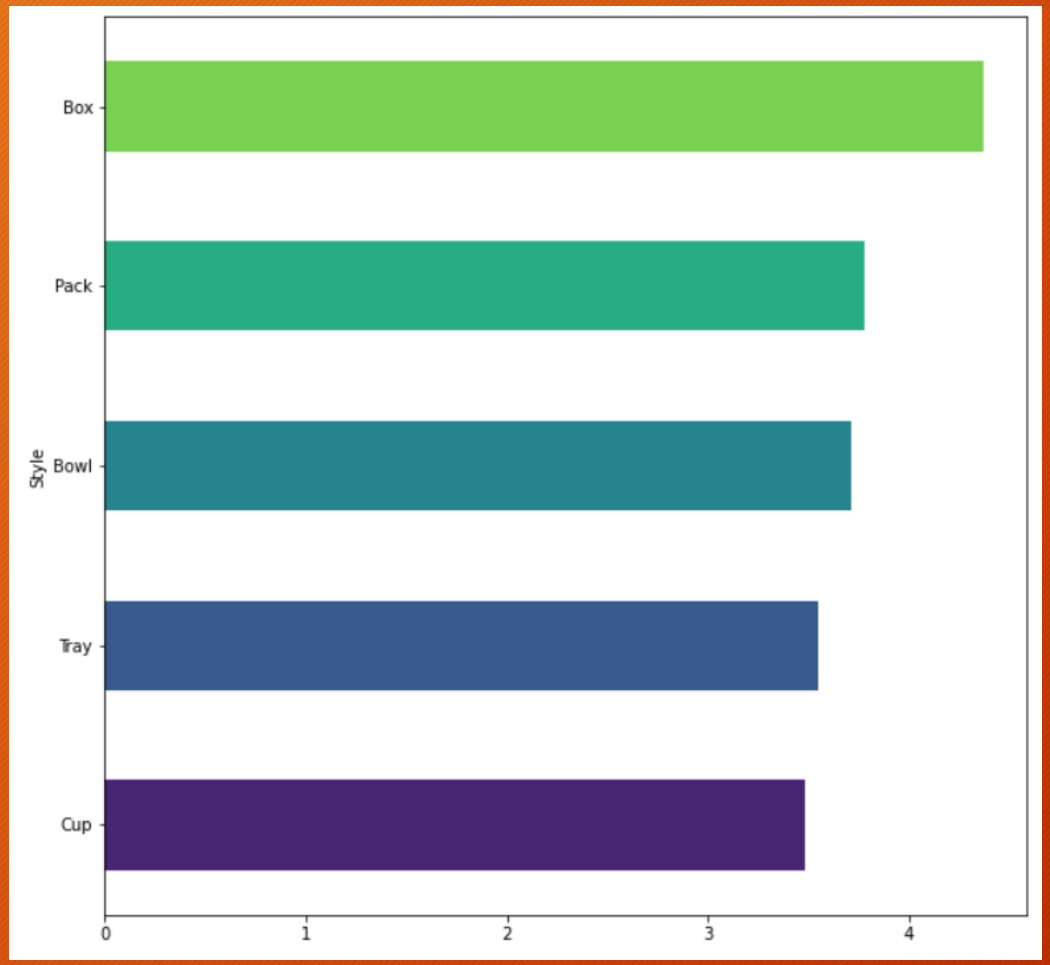




Question 2:
“Which brand produces the tastiest instant ramen?”



Question 3:
“Are there correlations between the form-factor (style) of the ramen and rating?”



Thank you for listening!

Questions?