



Amazon Product Review Dashboard

Grocery & Pet Supplies

Naomi Tsang | Serena Chan | Thomas Wong

Business Value

Potential Usage

Amazon



- Understand the market trend e.g. what are the hit products
- Increase bargaining power by identify the top/less performed partners and review the trading terms accordingly
- Identify service-related comments for self-improvement
- Make profit by selling the analyzed data to brands for their own review

Other e-commerce platforms



- Identify the brands comment to evaluate whether to cooperate with him/her;
- Identify the market trend e.g. which category most common and inform the brands

Retailer/Wholesalers



- Determine consumer preference for inventory allocation
- Identify market trend for new product opportunities

Manufacturer



- Identify the best/worst rated products to reallocate the inventory
- New product development/discontinue low rated products

Customer



- Classify performing goods to avoid negative experience

Data Collection

Kaggle Dataset & Web Scrapping

Kaggle Dataset: Amazon Product Reviews

- 568K + consumer reviews on different amazon products
- Total Records: 568454
- Total Columns: 10
- Available Fields: Id, ProductId, UserId, ProfileName, HelpfulnessNumerator, HelpfulnessDenominator, Score, Time, Summary, Text
- <https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews>
- Extracted ~1860 Top reviewed products data

Web Scrapping: Amazon Website Product Page

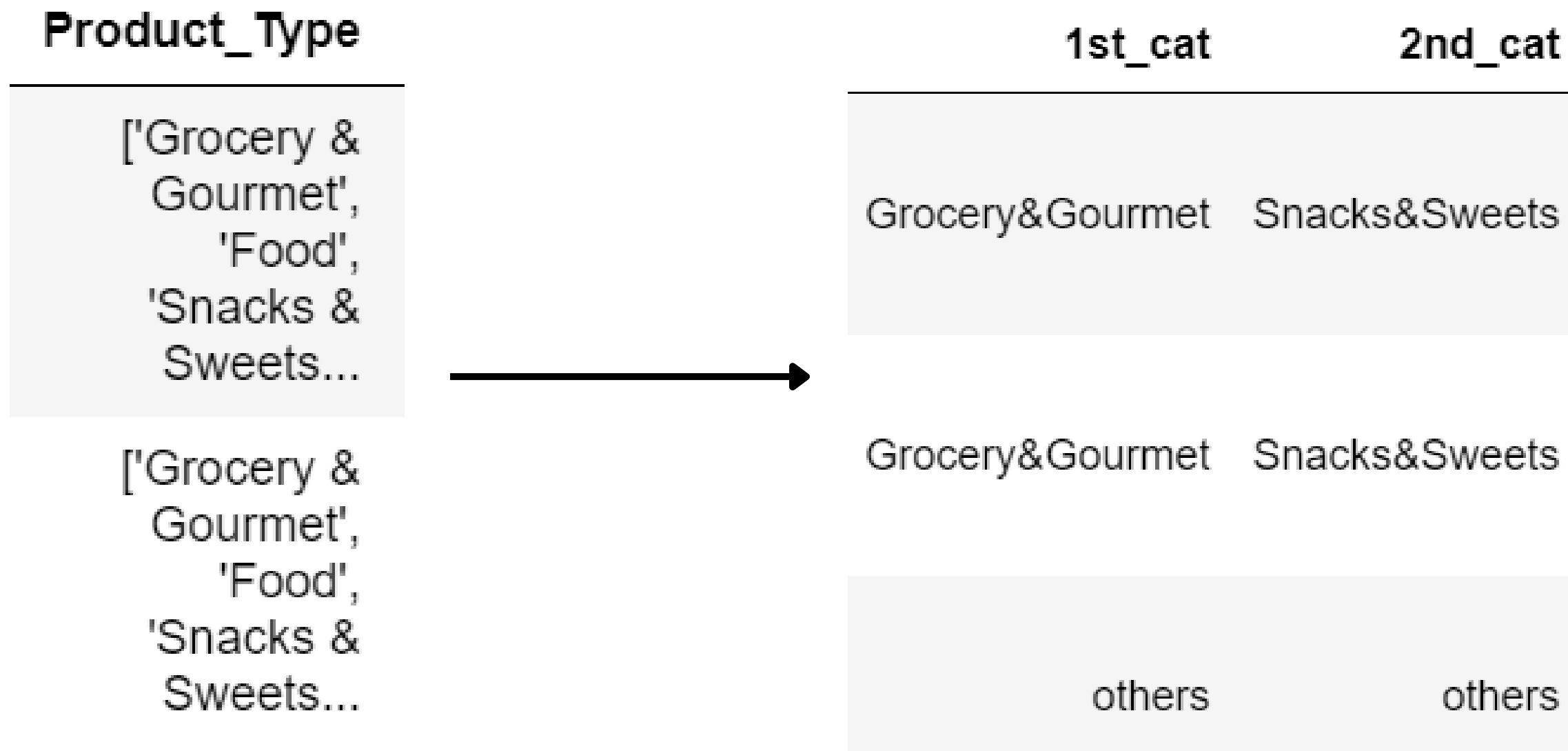
- www.amazon.com/dp/ProductId/
- Values: ProductName, Brand, MainCategory, SubCategory, SubClass, ImageLink, URL
- Libraries: BeautifulSoup, Selenium



Data Preprocessing

Data Cleaning

- Convert the timestamp to readable format
- Categories seperation



Data Preprocessing

Data Cleaning: Word Cloud Preparation

- For displaying word cloud in Tableau
- Only adjectives are extracted from the reviews
- Each adjective are placed in a single column

Text

Great if you
are cutting
back on fat,
they do ...

I have never
met a Kettle
brand chip I
did not...



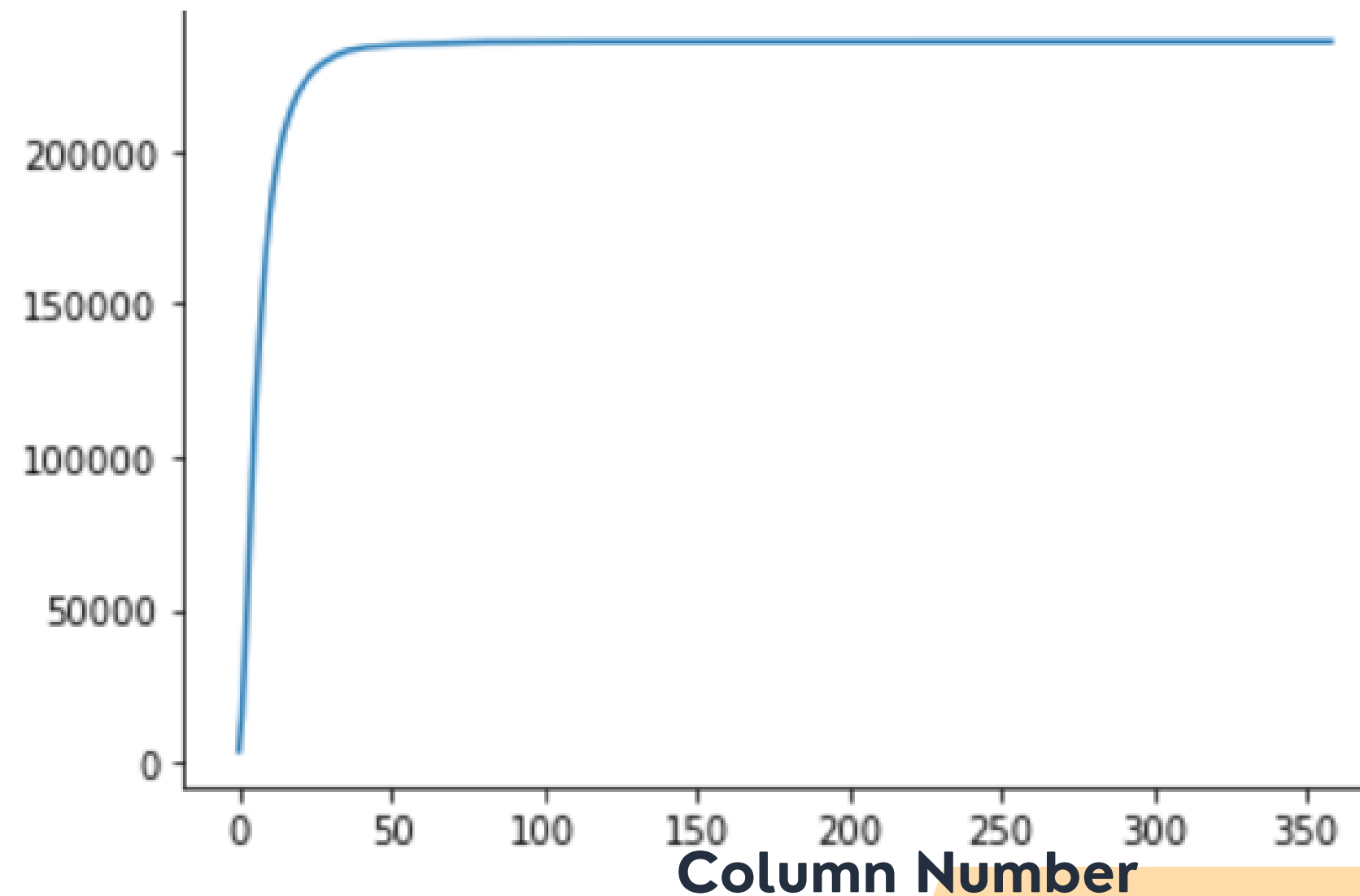
	0	1	2	3	4	5	6	7	8
0	resemble	regular	dry	hard	None	None	None	None	None
1	great	loaded	flavorful	sweet	great-tasting	great	None	None	None
2	bags	hefty	nearby	pleased	jalapeno	slow-building	whole	embarrassed	few

Data Preprocessing

Data Cleaning: Word Cloud Preparation

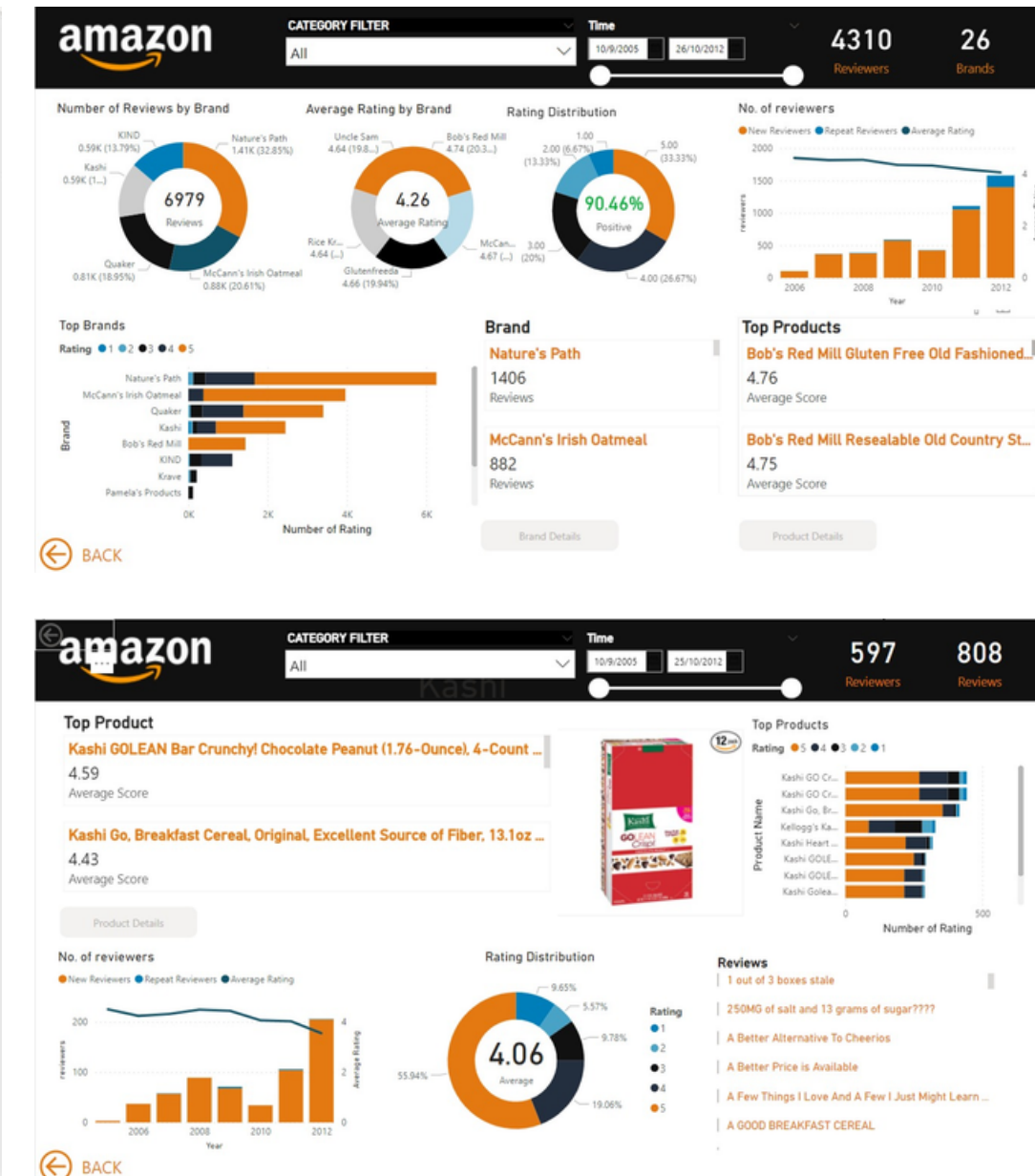
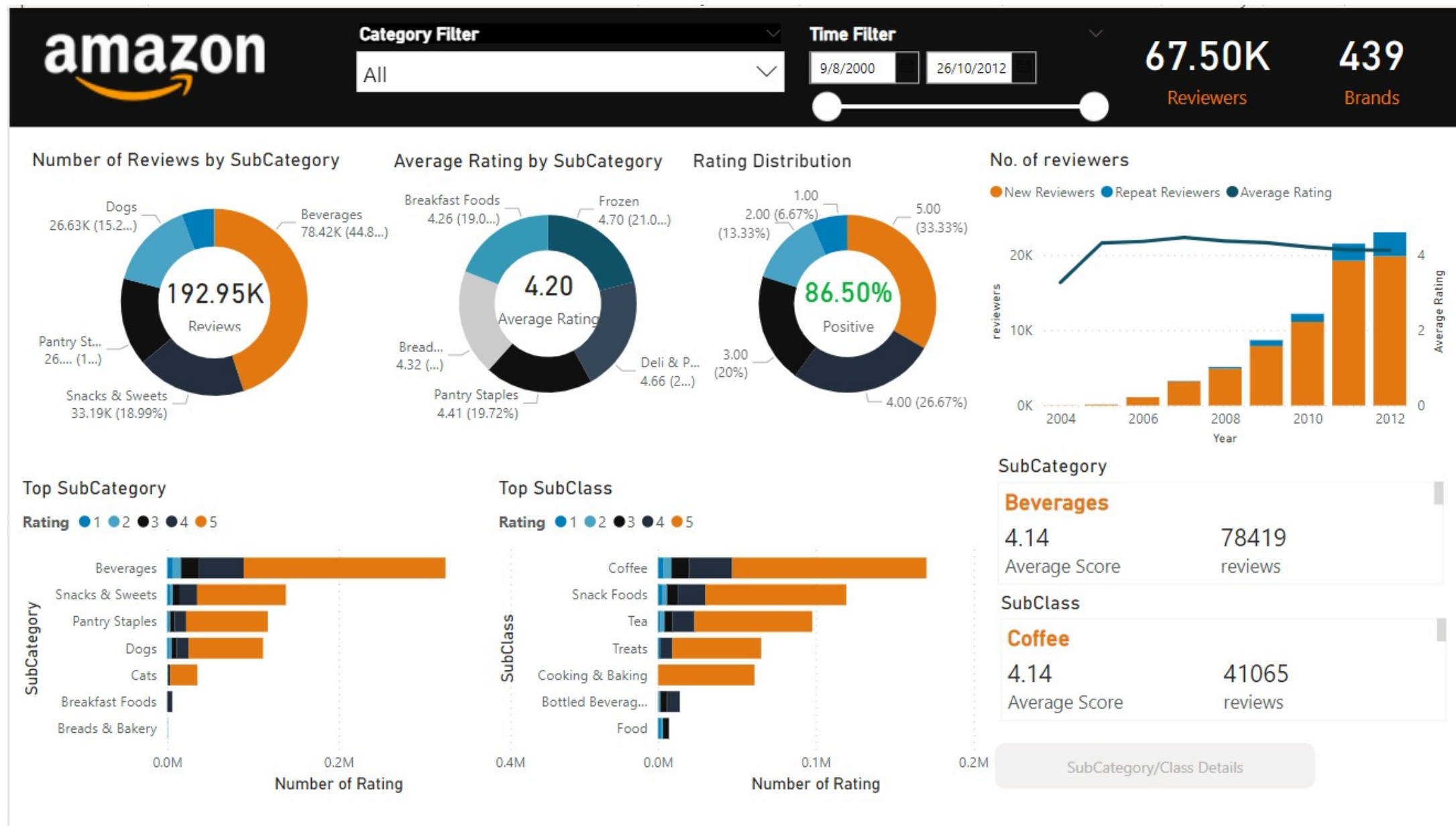
- Total columns: 359 columns
- Only the first 50 columns are kept
- Pivot the columns in table

Number of
empty cells in
each column



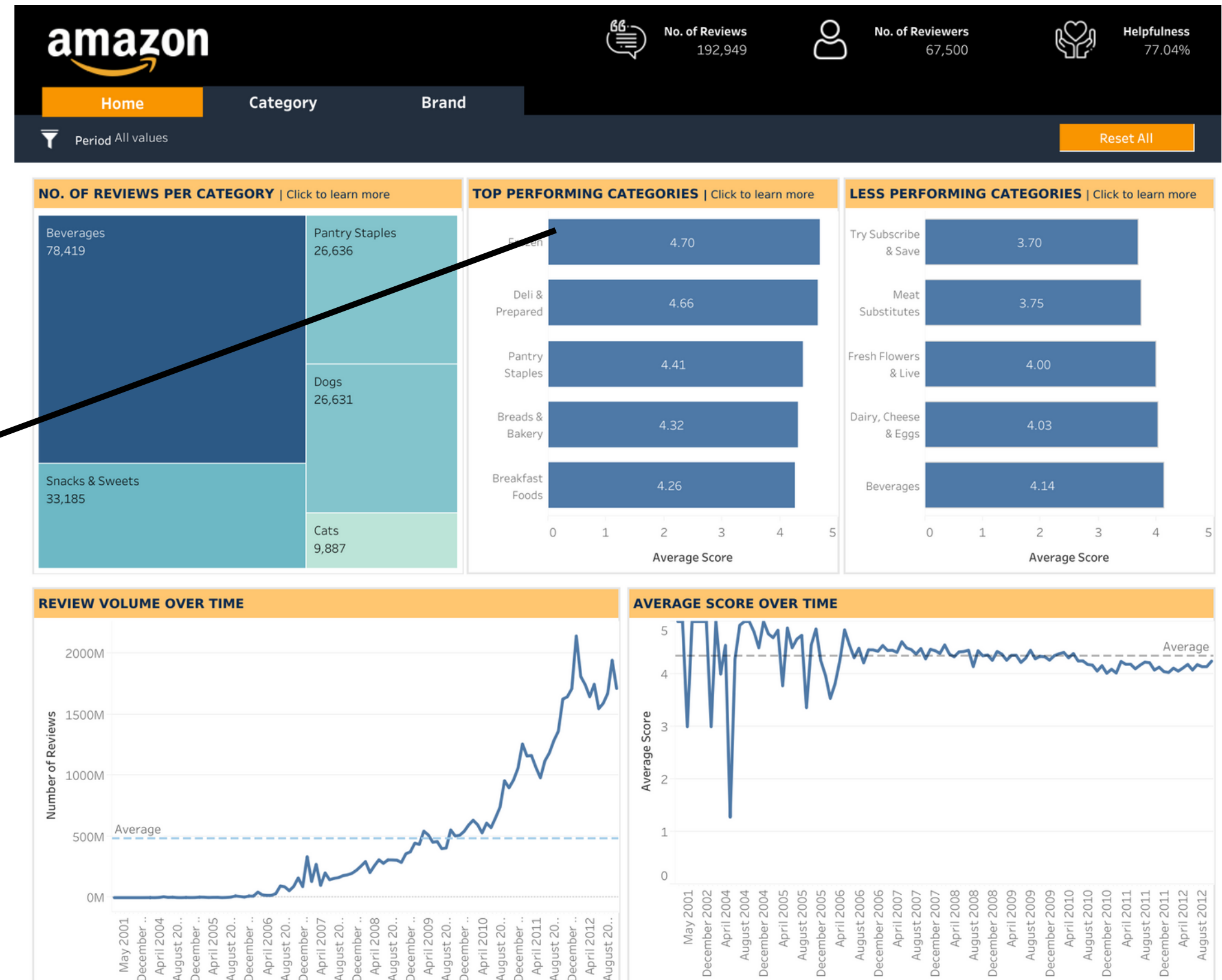
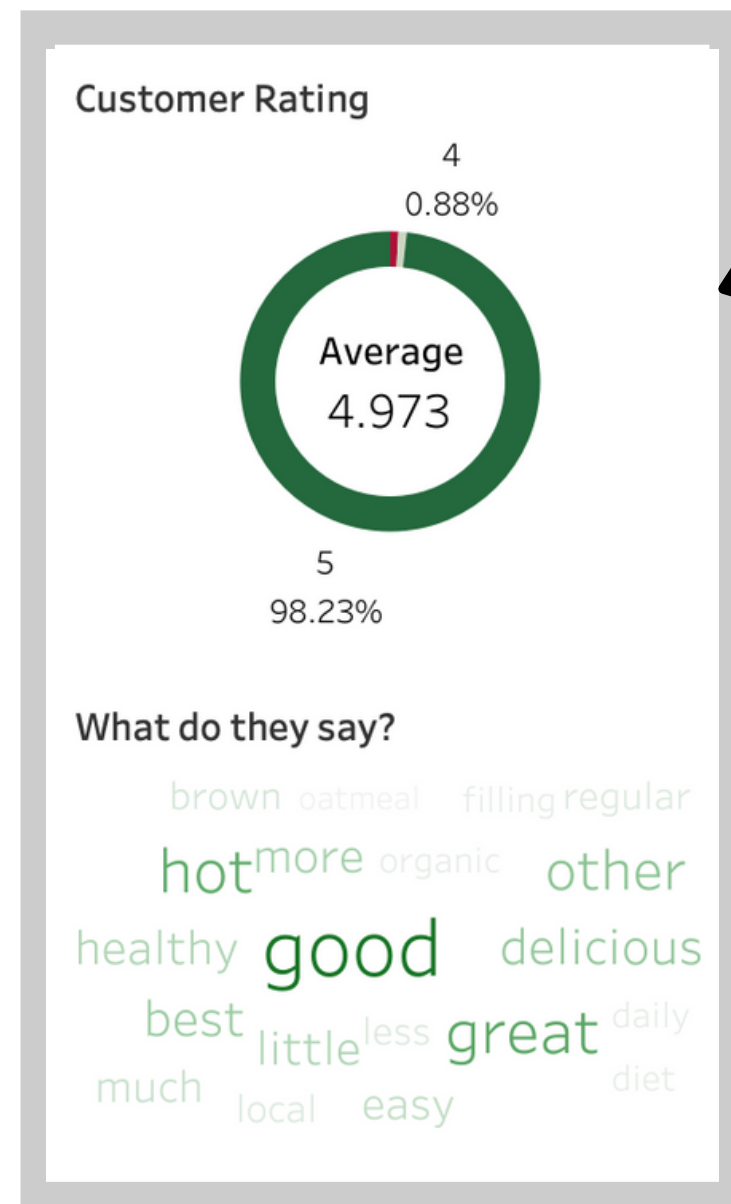
Analysis

Power BI #Focus: Product Rating



Analysis

Tableau #Focus: Product Reviews





Limitations

Future Improvement

Time Limitation

Limited Dimension of data

e.g. Geographical Data per review; latest reviews

Collect the reviews by web scrapping instead of using dataset from Kaggle

Hardware Limitation

NLP running time, software loading time

e.g. Performing NLP on ~240 000 rows of reviews; Tableau takes time to load data

Use a high-end computer; utilise Google Cloud - Cloud Engine

Software Limitation

Collobortion on PowerBi, Tablueau limitations

e.g. PowerBi free version disables the sharing function; max. 15 000 000 rows of data for Tableau free version; Tableau not able to post web-hosted image

Subscribe pro verisions of PowerBi/Tableau

Conclusion

Findings

- Number of reviews are boosted up after 2006
- Reviewers leave comments even if they are satisfied with the products
- The more satisfied the reviewer is, the shorter the comment

Suggestion

- Amazon may split the review box into sections (e.g. products, delivery, customer service, etc,) to reduce analysing time



Thank You!

Q&A