Table of Contents

Introduction

This report analyses the Cozyorganic dataset to present actionable insights and recommendations to help Cozyorganic improve its business strategy. It also aims to illustrate a number of key concepts and processes in text mining.

The report starts with an explanation of documents, corpus and tokens/terms. It then examines the customer satisfaction by state, which includes looking at which state has more positive feedback, as well as the overall positive to negative feedback ratio. This is followed by a term frequency representation of the first line in the dataset. Finally, three key questions regarding sales performance, supplier performance, and product preference by region are addressed; and recommendations are provided accordingly.

Documents, Corpus and Tokens/Terms

A document is a single unit of text, generally, from a single source. For example, in regard to the Cozyorganic dataset, the product feedback for January can be considered a document.
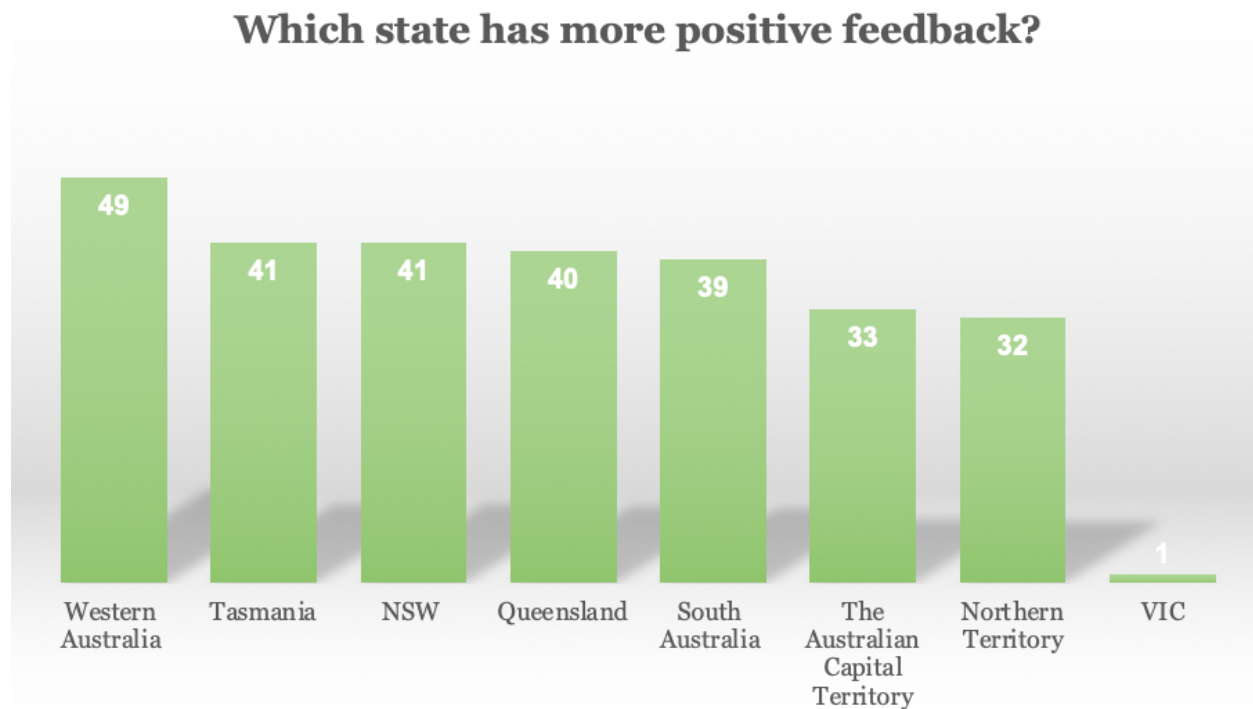
Corpus is a collection of documents. For example, the product feedback for all 12 months in the Cozyorganic dataset is a corpus.

Tokens/terms are single words or meaningful combinations of words. For example, some tokens in the Cozyorganic dataset are "store", "supplier", and "agree".

Customer Satisfaction

Product feedback is categorised into: 'agree', 'strongly agree', 'disagree', 'strongly disagree' and 'extremely poor', among which 'agree' and 'strongly agree' indicates positive feedback while the others indicate negative feedback. This report compares the amount of positive feedback received by each state as well as examines the percentage of positive and negative feedback in general to gain an overview of customer satisfaction.

Which state has more positive feedback

Which state has more positive feedback?

| Western Australia	Tasmania	NSW	Queensland	South Australia	The Australian Capital Territory	Northern Territory	VIC
49	41	41	40	39	33	32	1

Figure 1. The amount of positive feedback received by each state

As shown in Figure 1, Western Australia ranks first in providing customer satisfaction with 49 positive feedback. Tasmania and NSW both have 41 positive feedback, followed closely by Queensland, South Australia, The Australian Capital Territory and Northern Territory. Noticeably, Victoria only has 1 positive feedback. However, this is not likely to accurately reflect its performance in terms of customer satisfaction because there is only 1 data point in total for Victoria in the Cozyorganic dataset, which means the sample size is not big enough to draw a conclusion.

Praise-to-Criticism Ratio

When we compute the percentage of positive and negative feedback for all of the states (as shown in Figure 2), it is noticed that positive feedback only accounts for about 40% or two-fifths of the total feedback. Thus, the ratio of positive to negative feedback is currently 2 to 3. Zenger and Folkman (2013) confer that the ideal praise to criticism ratio is 5 to 1. Therefore, Cozyorganic should take actions to improve customer satisfaction to achieve 5 positive feedback for every negative one.
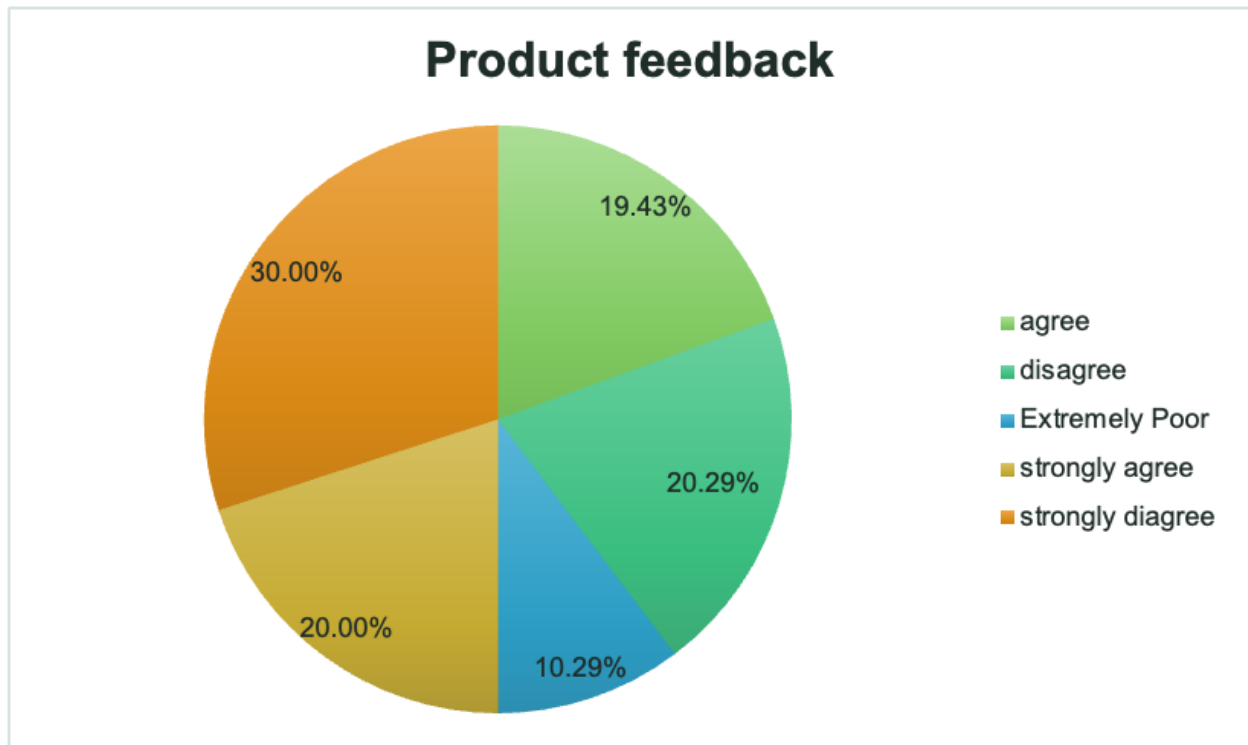
Figure 2. Product feedback

Term Frequency Representation

This report uses the bag of words approach to create a table of term frequency representation. It starts with extracting the terms from the first line in the Cozyorganic dataset, followed by case normalisation, stopword removal and stemming (using the PorterStemmer algorithm). The following table demonstrates the normalisation process, including transforming all terms into lowercase and stemming them.

Terms		Stemming (PorterStemmer)
Item_id	item id	item id
Item	item	item
Store	store	store
supplier	supplier	supplier
supplier location	supplier location	supplier loc
quantity	quantity	quantiti
Item sold	item sold	item sold

purchase price	purchase price	purchase pric
total cost	total cost	total cost
sold price	sold price	sold pric
Discounts	discounts	discount
total sales	total sales	total sal
Date	date	date
Month number	month number	month numb
Month name	month name	month nam
Year	year	year
product feedback	product feedback	product feedback

Table 1. Text pre-processing of the first line in the Cozyorganic dataset

There are no stopwords so no terms have to be discarded. The stemming was done using the PorterStemmer algorithm. Figure 5 shows the Python source code that was used to conduct stemming.

```
[ ]  #Import Python Natural Language Tool Kit Package
     import nltk
     nltk.download('stopwords')

     [nltk_data] Downloading package stopwords to /root/nltk_data...
     [nltk_data]   Package stopwords is already up-to-date!
     True

[ ]  from nltk.stem import PorterStemmer
     porter = PorterStemmer()
     words = ['item id', 'item', 'store', 'supplier', 'supplier location', 'quantity', 'item sold', 'purchase price', 'total cost', 'sold price', 'discounts', 't
     for word in words:
         print(word,"--->",porter.stem(word))

     item id ---> item id
     item ---> item
     store ---> store
     supplier ---> supplier
     supplier location ---> supplier loc
     quantity ---> quantiti
     item sold ---> item sold
     purchase price ---> purchase pric
     total cost ---> total cost
     sold price ---> sold pric
     discounts ---> discount
     total sales ---> total sal
     date ---> date
     month number ---> month numb
     month name ---> month nam
     year ---> year
     product feedback ---> product feedback
```

**Figure 3. PorterStemmer algorithm**

It should be noted that the PorterStemmer algorithm usually generates stems that are not actual English words. However, its advantage is simplicity and speed (DataCamp, n.d.), which is the reason why it was used to do the stemming in this case. After the terms have been normalised, they can be used to create a term frequency table, as shown in Table 2 below.

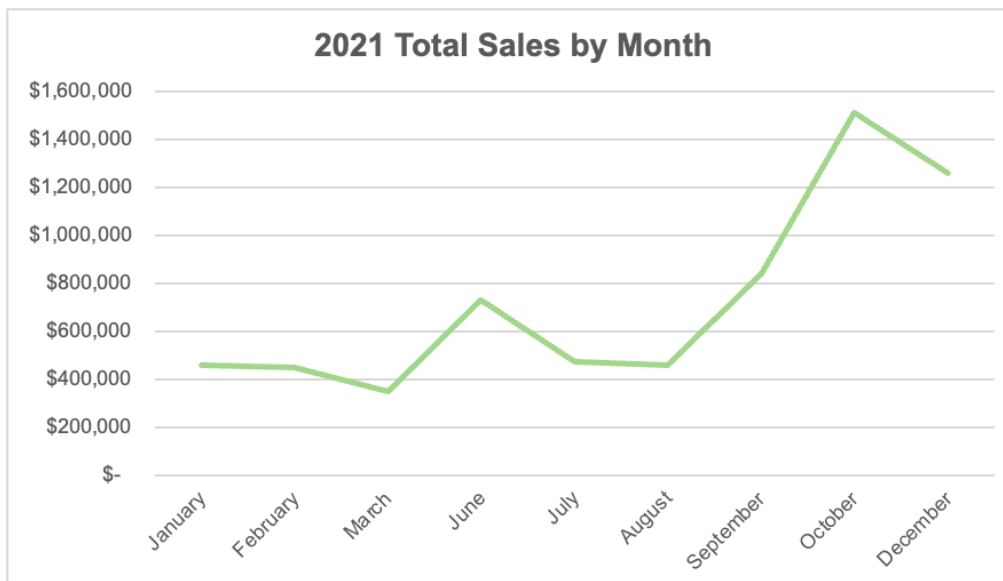| Terms	Frequency
item	3
supplier	2
sold	2
pric	2
total	2
month	2
id	1
store	1
loc	1
quantiti	1
purchase	1
cost	1
discount	1
sal	1
date	1
numb	1
nam	1
year	1
product	1
feedback	1

Table 2. Term Frequency

Other Key Insights and Recommendations

How is Cozyorganic's monthly sales performance?
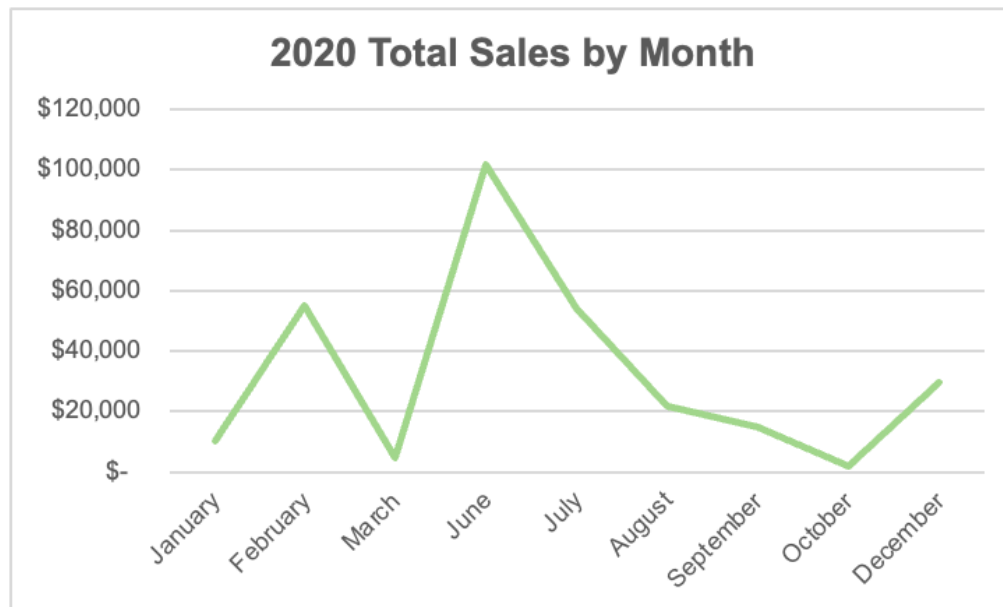
**Figure 4. Cozyorganic's 2020 Sales Month on Month**

Figure 5. Cozyorganic's 2021 Sales Month on Month

The data for 2021 sales has been filtered to exclude April, May and November so that it is comparable to the data for 2020 sales. It can be seen in Figure 7 and Figure 8 that Cozyorganic's sales fluctuate considerably during each year. However, the variation patterns for 2020 and 2021 are not similar. In 2020, Cozyoragnic saw the deepest fall in its sales from June to October - a decrease of around 98%. Meanwhile, in 2021, it only saw a slight decline in sales from June to August, followed by a significant climb from August to October - an increase of approximately 230%. It can also be noticed that while sales increased from October to December in 2020, they declined from October to December in 2021. Therefore, this fluctuation in sales is not likely to be seasonal.

If it was a seasonable decline because product demands change between seasons, then no large-scale counter-actions are needed other than closely monitoring the situation. However, in the case of Cozyogranic, it is not a problem of seasonality. Thus, there must be other reasons that are causing a decline in sales in certain periods. It is not apparent from this dataset why Cozyorganic's sales fluctuate. It is, therefore, recommended that Cozyorganic look into other data sources to see why sales decreased during a specific period and develop strategies to combat this. For example, in 2020, there was a spiked drop in sales between February and March, which indicates that there could have been some unexpected event that caused sales to decline. It can be an unexpected shortage of the most favoured products due to supply chain issues. The suitable action, then, is to simply fix the supply chain interruption. However, the gradual 4-month decline in sales from June to October 2020 might show that customers were not finding value in Cozyorganic's products, which would require a different course of action.

What are Cozyorganic's main suppliers and how do they perform in terms of customer satisfaction?
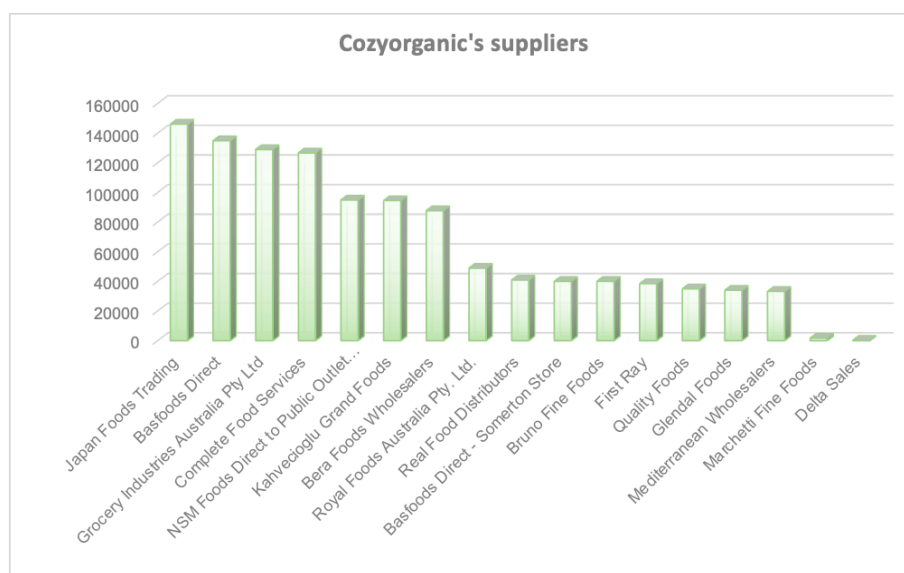
**Figure 6. Cozyorganic's suppliers in terms of quantity supplied**
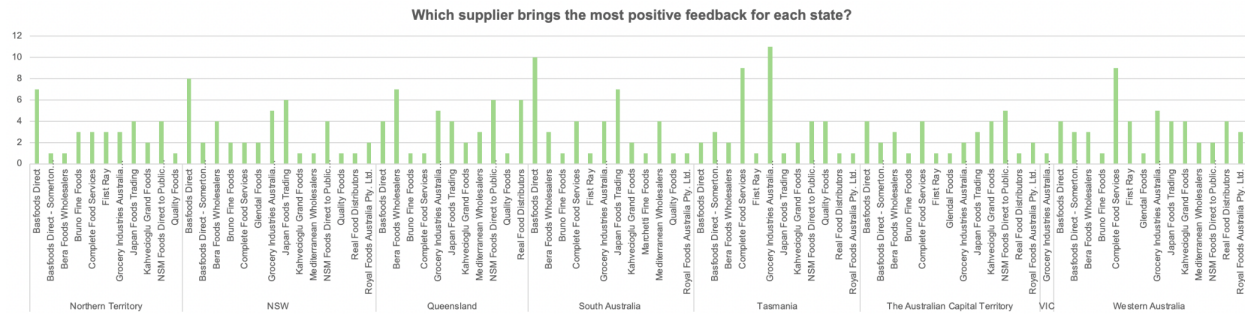
Figure 7. Suppliers' customer satisfaction performance

Figure 6 demonstrates the quantity supplied by each Cozyorganic's supplier while Figure 7 shows the supplier that brings in the most positive feedback for each state. It can be seen that Cozyorganic imports the bulk of its products from Japan Foods Trading, Basfoods Direct, Grocery Industries Australia Pty Ltd and Complete Food Services, each supplying more than 120,000 items.

Nevertheless, despite being Cozyorganic's largest supplier, Japan Foods Trading does not win the highest customer satisfaction in any of the surveyed states. Meanwhile, Basfoods Direct brings in the most positive feedback for one-third of the states, namely NSW, South Australia and Northern Territory. Grocery Industries Australia and Complete Food Services won the most positive feedback for Tasmania and Western Australia, respectively.

Thus, it is recommended that Cozyorganic import more products from Basfoods Direct, Grocery Industries Australia and Complete Food Services as products provided by these suppliers are receiving great feedback from customers. It should consider replacing Japan Foods Trading with Basfoods Direct as the main supplier. A strategic partnership with Basfoods Direct is recommended. This will allow Cozyorganic to buy in bulk from Basfoods Direct at a discounted price, thus increasing its profitability. However, this does not mean working with one supplier only. It is still a good strategy for Cozyorganic to diversify suppliers to reduce risks of product shortage, for example, in the case where a supplier encounters an unexpected supply chain issue.

Which state has a high demand for which product(s)?
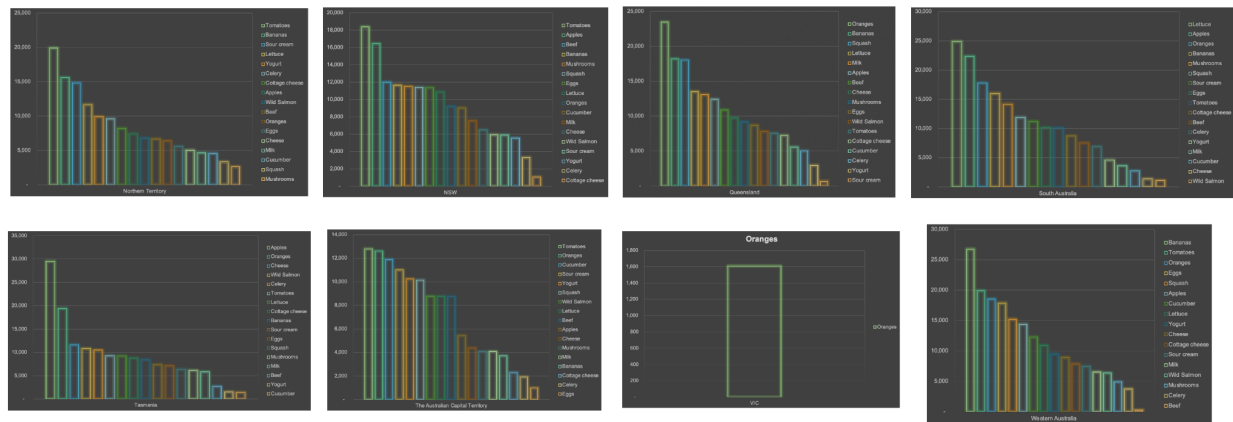
Figure 8. Best seller items by state

Figure 8 gives an overview of product preference by state. Tomatoes are in high demand in Northern Territory, NSW and The Australian Capital Territory. Apples and bananas are especially popular in Tasmania and Western Australia, respectively, while lettuce gains great traction in South Australia. Meanwhile, customers in Queensland and Victoria have a strong preference for oranges.

Thus, Cozyorganic should tailor its product offerings and marketing messages to customers in each state based on their product preferences. An example is a special promotion program for tomatoes in Northern Territory, NSW and The Australian Capital Territory where customers buying tomatoes before 10 am every morning will receive a 10% discount. This will help to further stimulate demand for this product.

Conclusion

In terms of customer satisfaction, Western Australia achieved the most positive feedback while Basfoods Direct stood out as the supplier that was doing best in this aspect. Besides, it was found that the current ratio of positive to negative feedback is 2 to 3, which should be changed to 5 to 1.

Cozyorganic's sales greatly fluctuation by month and the data patterns showed that this was not a seasonal problem. Thus, Cozyorganic should investigate the root cause of the decline in sales in specific periods to take necessary actions.

In addition, Japan Foods Trading, Basfoods Direct, Grocery Industries Australia Pty Ltd and Complete Food Services are currently Cozyorganic's main suppliers. It is recommended that Cozyorganic lform a strategic partnership with Basfoods Direct while also keeping a sufficiently

diversified portfolio of suppliers to ensure products are always in stock to cater to customers' demands. Besides, Cozyorganic should consider each state's different preferences for its products in order to adopt a suitable differentiated marketing approach.

References

DataCamp. (n.d.). *Stemming and Lemmatization in Python*. https://www.datacamp.com/tutorial/stemming-lemmatization-python.

Zenger, J. & Folkman, J. (2013, March 15). The Ideal Praise-to-Criticism Ratio. *Harvard Business Review*. https://hbr.org/2013/03/the-ideal-praise-to-criticism.