NespressoMetropolisCustomerReviewAnalysis

Kunal Jeshang - Coffee Specialist *(Project Timeline: January 2023 to May 2023)*

Table of Contents

Introduction

This project was created to assist the Nespresso Metrotown Boutique understand its service quality in terms of customer experience and optimal execution of business operations. As management of the Boutique predominantly refers to Google Reviews to periodically assess service quality, the data utilized for this project were Google Reviews received between 2019 to 2022. Thus, reviews received in 2023 were not considered for analysis. In turn, this project is a cumulative annual analysis of our service quality until 2022 year-end.

Nespresso Metrotown management periodically refers to Google reviews to understand customer perception. Although, the Google reviews online shows very basic informaton such as aggregate star rating along with star rating and written review per reviewer. This is not as informative unless one were to read each and every review in great detail. Thus, this project aimed to better understand customer perception of service quality for the Nespresso Metrotown branch, albeit from an analytical perspective. Using machine learning & natural language processing, sentiment analysis was performed on the pre-processed written Google reviews.

Working on this project has been a fruitful experience. I am thankful for this to be a part of my Employee Performance Development program under Nespresso & Randstad. This experience greatly improved my technical ability, but also challenged myself to consider what management would potentially be thinking about. I have written this project report in the form of a research study. Although, being a Coffee Specialist, I may have a frame of mind that may differ from that of members of staff that at a supervisory and/or management level. This could lead to analysis and interpretation of results that does not emulate the level of depth a member of staff of management level would have. Therefore, please consider consider this project with an open mind and with some scrutiny.

Below is a list of the five stages the project followed through.

1. Data Scraping and Extraction
2. Data Cleaning
3. Exploratory Data Analysis
4. Sentiment Analysis - Exploration
5. Sentiment Analysis - Predictive Modeling

This project was completed using the following devices, tools, and technologies.

- Devices & Platforms
  - ASUS VivoBook - Windows 11 (8GB RAM); used for data scraping/extraction
  - Macbook Air 2014 - Mac OS X BigSur (4GB RAM); used for all stages *after* data scraping/extraction
- Scripting
  - Programming Language: Python
    - Official Version 3.10.1 using Asus VivoBook; used for data scraping/extraction
    - Anaconda Navigator version (3.9.11) using Macbook Air 2014; used for all stages *after* data scraping/extraction
  - Important Packages: Selenium (with ChromeDriver.exe), Datetime/Time/Dateutil, Pandas, Numpy, Re, Matplotlib, Seaborn, WordCloud, TQDM, Itertools, NLTK, String, TextBlob, NRCLex, Scikit Learn
- Development Environment
  - Jupyter Notebook
  - Software
    - Microsoft Visual Studio Code; used for data scraping/extraction
    - Anaconda Navigator; used for all stages *after* data scraping/extraction

Data Scraping and Extraction

This is the first stage of the project but it is also the hardest. Despite customer Google Reviews being freely accessible online via Google Maps, I am unable to retrieve the data in a traditional sense (i.e., download a CSV/JSON file). Being simply a Coffee Specialist at the Boutique, I would not have access to save this data locally on an endpoint because I am neither in a management-level position nor part of the Nespresso Canada Social Media team. Therefore, the only option was to programmatically extract the data using the Selenium package and the ChromeDriver executable file.

In other words, I coded a scraper bot to perform the following.

1. The scraper bot is switched on and it will open the Google Chrome web browser.
2. Go to Google Maps webpage of Nespresso Metrotown branch via a pre-specified URL link.
3. Retrieve ***Overall Rating*** and ***Total Number of Reviews*** values via a pre-specified xpath, and save the values in a pandas dataframe.
4. Retrieve the count in the ***Number of Reviews*** per Star Rating via a pre-specified xpath, and save the counts in a pandas dataframe.
5. Access the full reviews list within the Google Maps webpage of Nespresso Metrotown branch via a pre-specified URL link.
6. Sort the reviews by recency by toggling the *Sort* button and then selecting the *Newest* option via pre-specified xpaths.
7. As not all reviews can be seen within the browser window, the scraper bot would continuously scroll to the bottom of the page/review list to load all reviews until the very first that was received in 2019. This is performed with the pre-specified xpath of the review list scroll-bar.
8. After all reviews from oldest to most recent are loaded by the webpage, using a partial xpath refering to the reviewer name, the scraper bot will retrieve the reivewer names and save them as elements in a list to be used in the forthcoming steps.

9. As some reviews recieved for Nespresso Metrotown branch can be quite long, the full written review is not shown to conserve browser screen space. There are *More* buttons under each review to show the full written review in case it is too long. Thus, the scraper bot must click all of the *More* buttons in the review list with the button HTML Tag name to unveil the full written reviews for retrieval.
10. Using previously retrieved reviewer name list and a pre-specified parent xpath, the scraper bot will loop through the entire review list from most recent to oldest, and retrieve the values of ***Reviewer Name***, ***Total Reviews Given*** by the reviewer, ***Time of Review***, ***Review*** (i.e., the actual review received), and ***Stars Given***. Then the aforementioned values are saved in a pandas dataframe.
11. All pandas dataframes are altogether saved in an Excel workbook, as well as individual CSV files, and then placed into a "data" folder.
12. The scraper bot is then shutdown.

Please refer to the Scrape Jupyter Notebook to view the codebase of the data scraping & extraction.

Data Cleaning

In this stage of the project, the raw Google Reviews data is imported and is cleaned for any data scraping errors & inconsistencies. There is transformation of the dataset to include new columns that are more meaningful for the later stages of the project. It is imperative to take a peek of the data by checking the first few rows and summary of column names, non-NULL count, and data types, prior to performing data cleaning & transformation. After the appropriate data cleaning & transformation steps are completed, the cleaned data is then loaded to a CSV file to be used in the next stage of the project. That being said, below is a list of the most important actions taken in this stage of the project leading up to the aforementioned data loading step.

- Retreive ***Retriever Title*** from ***Total Reviews Given*** column values, and create a new column from the retrieved values.
- Clean the the ***Total Reviews Given*** column values such that the word "reviews" is removed, and only the numeric count of reviews remains.
- Parse the column values for ***Total Reviews Given*** to integer, remove any instance of punctuation, and print rows whereby the column values could not be parsable to integer. Make manual adjustments by row when necessary.
- Clean the ***Time of Review*** column values such that there is no extra whitespace.
- Due to a web scraping error, for some rows the value for ***Total Reviews Given*** is moved to ***Time of Review*** column, and the value for ***Time of Review*** is moved to ***Stars Given***. Find the rows where this error has occurred and perform the appropriate shuffling so that the appropriate values are in the correct columns.
- Retrieve the ***Year of Review*** from the *Webscraping Datetime* value and ***Time of Review*** column value, and create a new column from the retrieved values.
- Clean the ***Review*** column such such that each row does not contain "response from owner" or NULL values.
- Clean the ***Stars Given*** column by removing the word "stars" and any extra whitespace. Then convert the column data type to numeric.
- Change the order of columns to be more logical, and change column names when necessary.

Below is a tabular breakdown of the cleaned Google Review data in terms of columns and data types.

Column	Data Type

Column	Data Type
Reviewer Name	object
Reviewer Title	object
Total Reviews Given	int64
Review	object
Stars Given	int64
Year of Review	int64
Time of Review	object
Webscraping Datetime	object

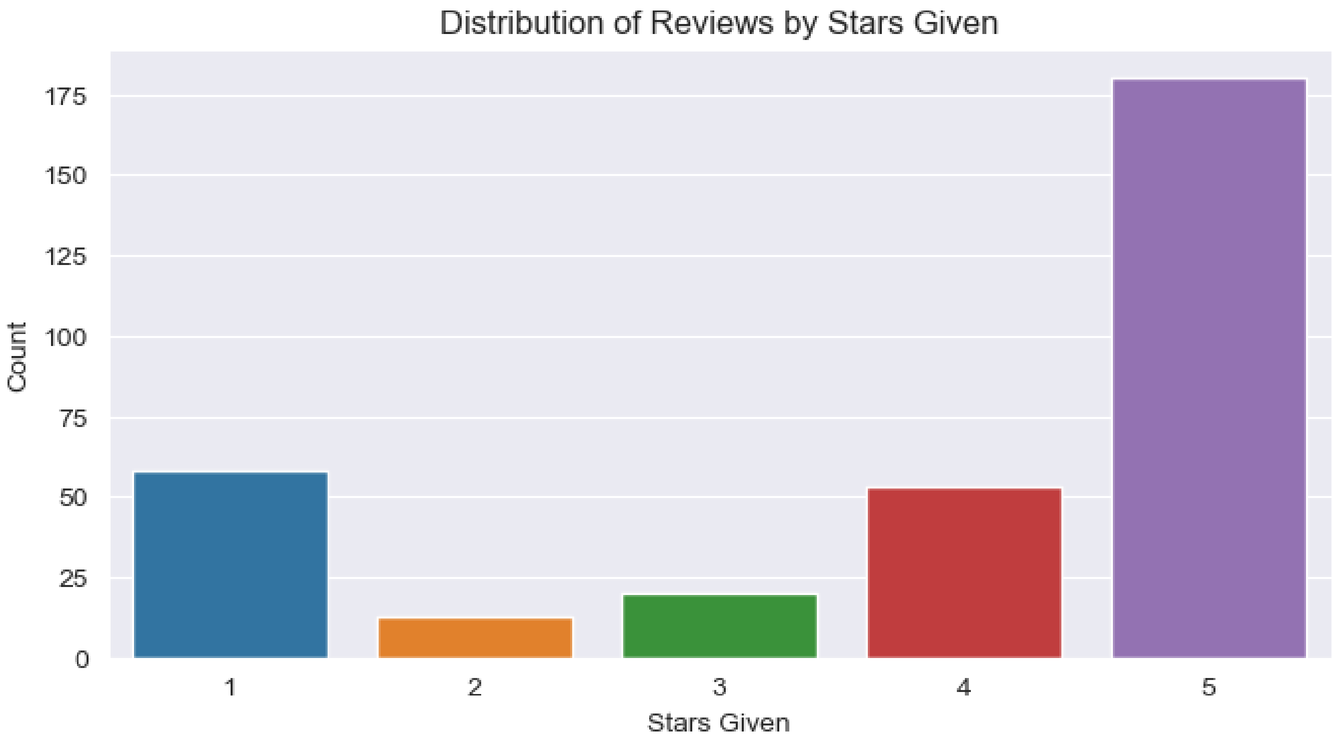
Please refer to the Data Cleaning Jupyter Notebook to view the codebase of the data cleaning.

Exploratory Data Analysis

This stage is somewhat self explanatory, but it is important in order to get "a feel" of the now cleaned Google Reviews dataset by performing Exploratory Data Analysis (EDA). Prior to this, it is important to import the data and filter it such that the ***Year of Review*** is between 2019 to 2022. In other words, the data should be filtered to reflect 2019 to 2022 as the current year, 2023, is still in progress.

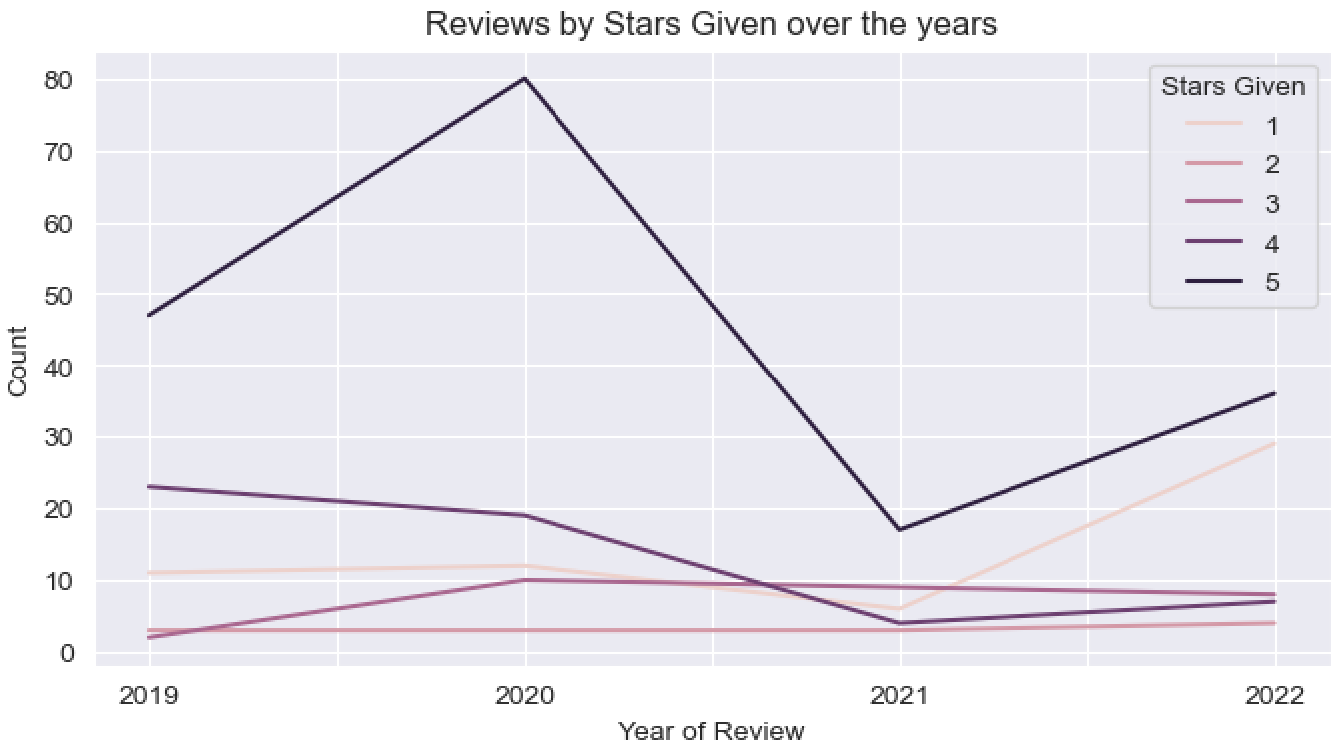
To see a list of the top reviewers overall as well as top reviewers based on specific condition, please refer to the latter half of the EDA Jupyter Notebook. This is to protect both the Google Accounts and identities of the customers that provided reviews to not be parsed by any internet search engines.

Fig 1. Distribution by Stars



Overall, our service quality is good but not consistent. There is a higher distribution of 5-star reviews received throughout the years. The Boutique has also received a relatively lower proportionate amount of reviews with 1-star and 2-star rating.

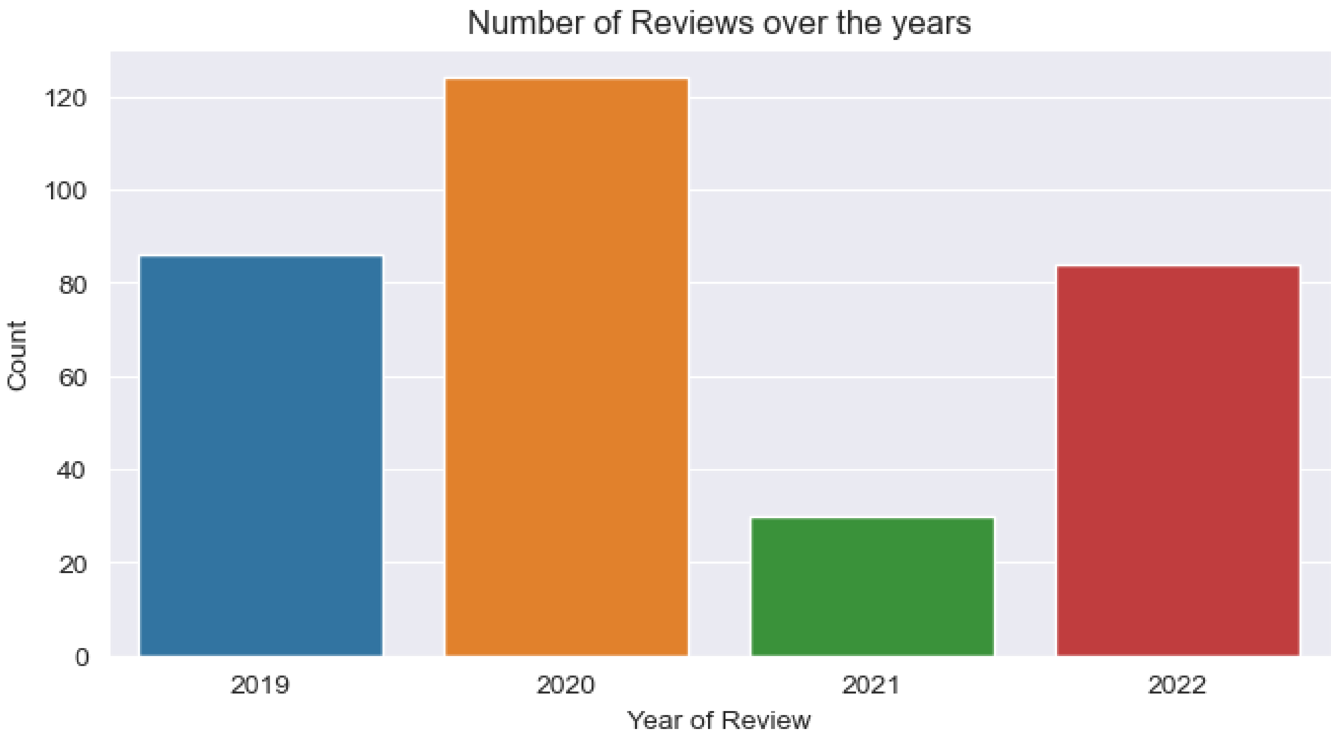
Fig 2. Reviews by Stars Given over the years



There has been a significant dip in the number of reviews received in 2021, and then there was a sharp increase of the number of reviews in 2022. This is due to to the Covid-19 pandemic where there was limited in-person services provided at Nespresso. Therefore, customers would not be able to experience our in-person customer service to provide Google Reviews. Below are some general observations:

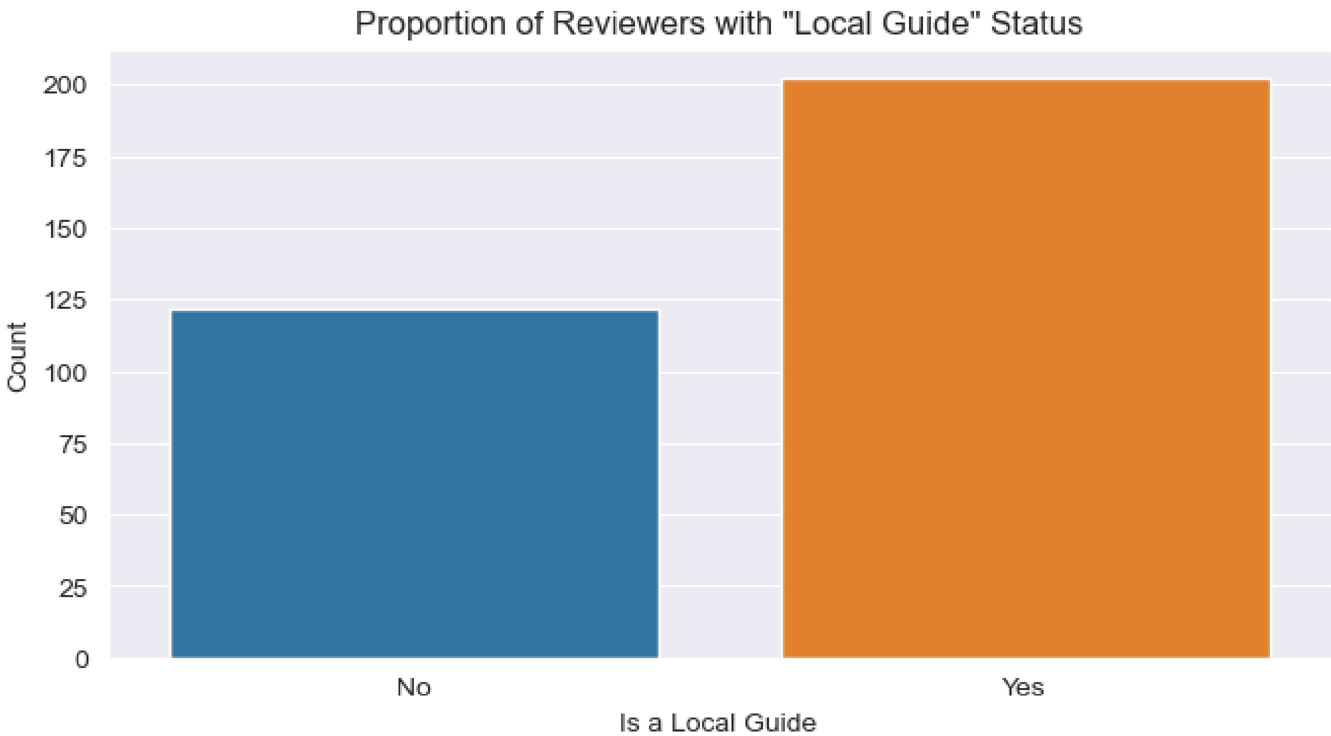
- Significant increase in 1-star reviews in 2022 comapared to 2021.
- Stagnant 2-star reviews throughout the years.
- Historically, there has been a decrease in 4-star reviews but there was a slight uptick in 2022.
- There has been a significant increase and then a sharp dip in 5-star reviews post-2020, and then a sharp increase post-2021.

Fig 3. Number of Reviews over the years



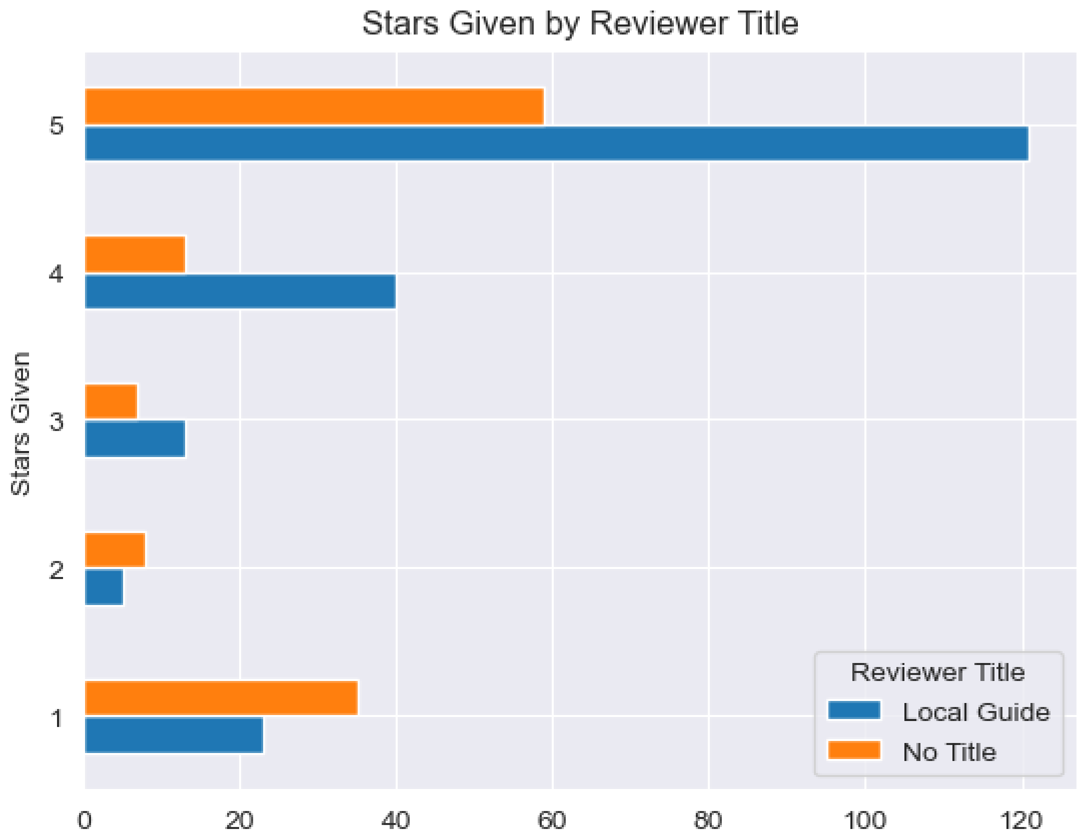
Interestingly, the boutique received maximum reviews in 2020, yet minimal reviews in 2021 despite the Covid pandemic still at large. Post-pandemic levels of reviews are equivalent to pre-pandemic levels of reviews.

Fig 4. Proportion of Reviewers with "Local Guide" Status



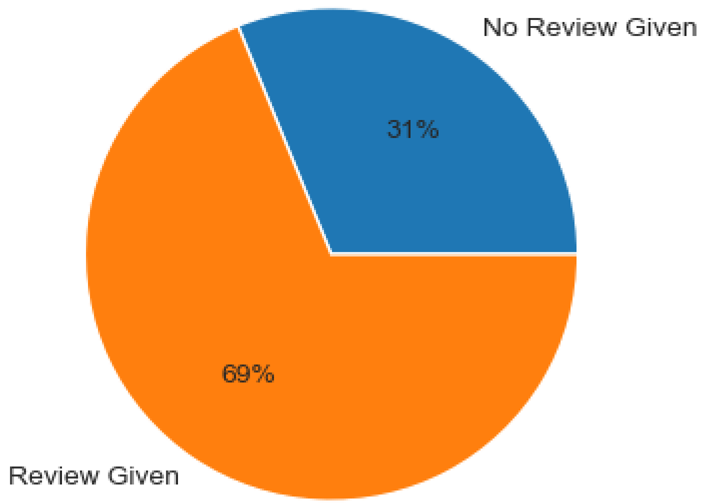
There is a higher proportion of reviews received from customers that are of "Local Guide" status, yet there are a lot of reviews received from customers that are not of "Local Guide" status that suggests a higher amount of subjective reviews.

Fig 5. Stars Given by Reviewer Title



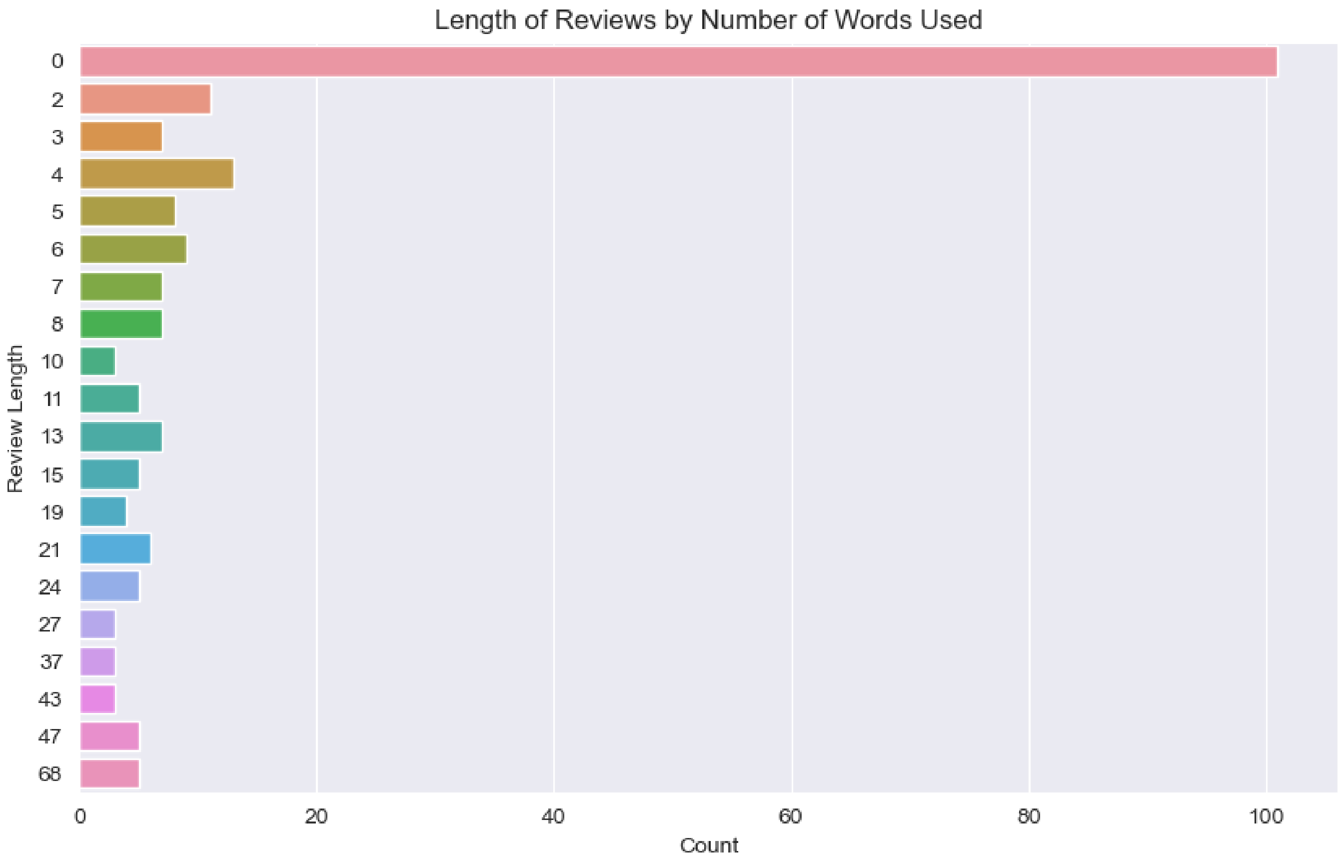
The interesting observation in this chart is that almost 50% more reviewers with "Local Guide" status provided 5-star rating compared to reviewers without "Local Guide" status. Furthermore, reviewers with "Local Guide" status also provided more 4-star and 3-star rating compared to reviewers without "Local Guide" status. This infers that Nespresso Metrotown's service quality is excellent, and it is truly acknowledged as well as legitimate. A confirmation of this notion could be that there are more 2-star and 1-star rating provided by reviewers without "Local Guide" status. That being said, there is still a sizeable proportion of 1-star reviews form reviewers with "Local Guide" status. Thus, certain negative perceptions about Nespresso Metrotown's service quality may be legitimate.

Fig 6. Percentage Proportion of Review Given vs No Review Given



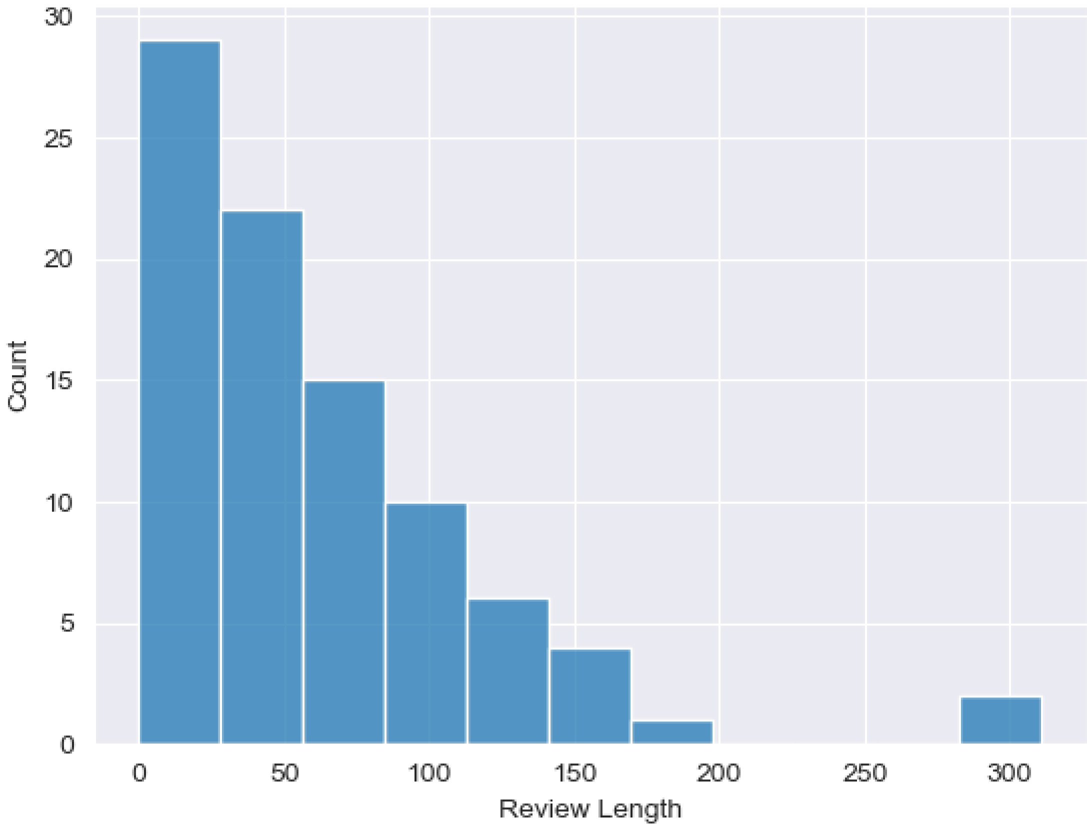
The pie chart above indicates that only nearly 70% of the reviewers wrote a review in addition to providing a star rating. Thus, little more than 30% of the reviewers only provided a star rating, but did not write a review. Regardless whether the written reviews are positive, negative, or in-between, the service provided to the customers is memorable enough for them to provide a written review.

Fig 7. Length of Reviews by Number of Words Used - Bar Chart



Despite nearly 70% of reviewers providing a written review, when comparing the word count of written reviews, there is a variation in terms of the number of words used per reviewer based on the bar chart. The number of reviews without a written component outweighs the number of reviews with a written component that is more than or equal to two words.

Fig 8. Length of Reviews by Number of Words Used - Histogram



In the histogram, it can be seen that the majority of the number of words used for written reviews are less than 150. There is also an anomaly whereby there has been a handful of reviewers that wrote around 300 words for their review. This is indicative of an anomaly occurrence whereby the reviewer may have experienced a very positive or negative experience in terms of service quality at Nespresso Metrotown branch. Despite the anomaly, the histogram indicates a right-skewed distribution with the majority of the reviews having a word count less than 150.

Sentiment Analysis Exploration

This stage of the project, after importing the necessary packages, the cleaned Google Reviews is imported and filtered to only include reviews received from 2019 to 2022. Prior to performing any natural language processing (NLP), the Google Reviews data must be pre-processed; specifically, the ***Review*** column values for each row. NLP pre-processing is necessary to perform appropriate sentiment analysis. Below are the important steps involved in the NLP pre-processing:

1. Perform word tokenization such that each word in the ***Review*** column value is separated by a comma in a list. Now, each word are referred to as a token.
2. Perform lemmatization such that each token that is a word in its extended form is reduced to its base form (i.e., Caring --> Care).
3. Remove any tokens that are punctuation or english stop words.
4. Perform Part-of-Speech tagging on each token to only include adjectives, verbs, nouns, and adverbs.
5. For each token, reduce any occurrence of additional whitespace.
6. Combine all tokens, in the comma-separated list, together into a unified sentence.
7. Repeat steps 1 through 5 per row for each ***Review*** column value.
8. Create a new column called ***Review Cleaned*** and save the pre-processed column values to it.

Post-completion of NLP pre-processing, sentiment analysis is performed. There are essentially four methods.

1. **Word Cloud:** A simplistic method that considers the most highly used words overall, and displays them in the form of a word cloud. Interpretation of the overall sentiment of Nespresso Metrotown is determined by the reader.
2. **VADER Sentiment Scoring:** Able to provide a score for positivity, negativity, and neutrality, as well as an overall compound score to the reviews. VADER refers to Valence Aware Dictionary and sentiment. This method incorporates a Bag-of-Words approach, which considers simply the frequency of the words used.
3. **Textblob Sentiment Scoring:** This method functions similarly to VADER method, but determines a numerical score for the subjectivity and polarity of the written review. Therefore, the level of objectivity & validity (or lackthereof) can be understood using this method.
4. **Emotion Classification using Lexicon based method:** This method is able to provide numerical scores to a collection ten emotions of varying levels of positivity & negativity based on the reviews.

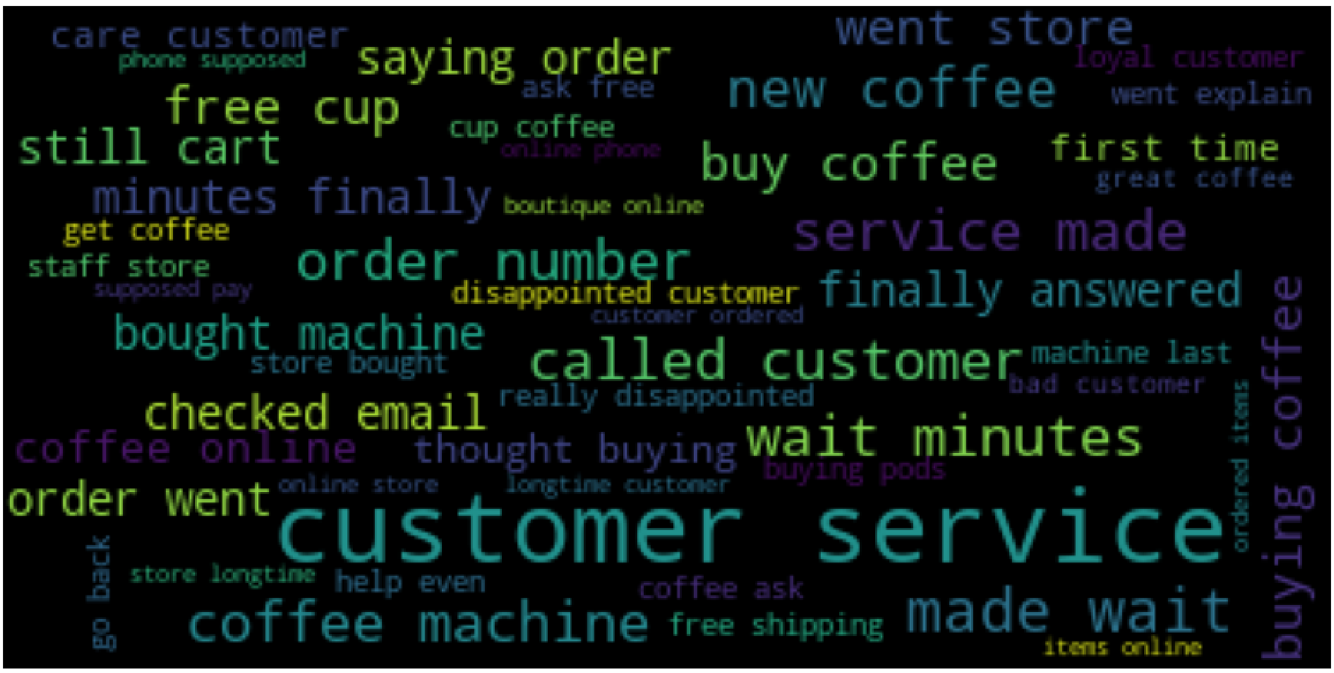
After performing a sentiment analysis exploration using the above four methods, at least three new datasets are created using the cleaned reviews dataset to reflect the sentiment analysis results for each method. Then the newly created datasets are exported as CSV files to be used in the next stage of the project.

Please refer to the Sentiment Analysis Exploration Jupyter Notebook to view the codebase of the sentiment analysis exploration.

In the forthcoming sub-sections, the visualizations derived from each sentiment analysis method are shown, along with brief intepretations of the visualizations and analysis.

Word Cloud

Fig 9. Word Cloud

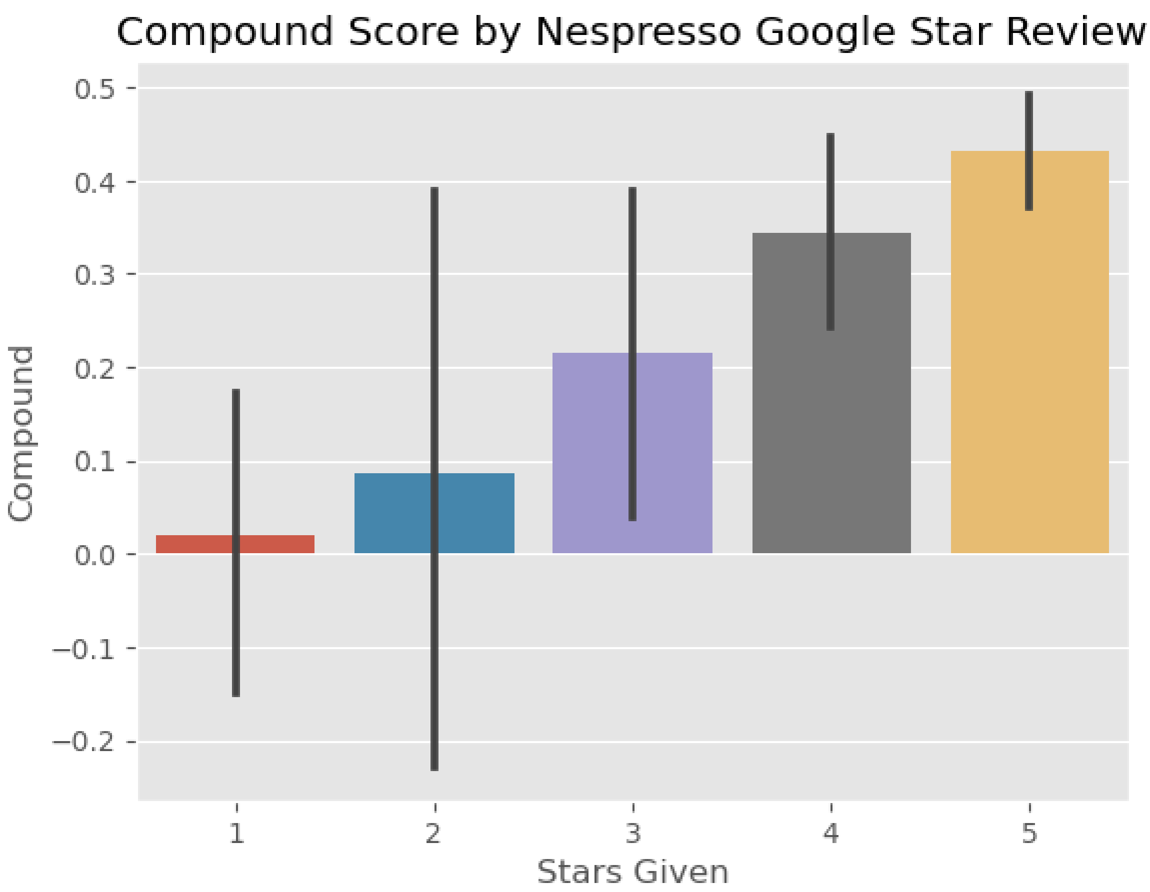


The word cloud is not quite conclusive. Larger the size of the words means higher the occurrence of the word. Most of the words of a larger size are neutral in nature. The remaining words that are of a smaller size are either neutral as well as of negative sentiment. Thus, deciphering the overall sentiment of Nespresso Metrotown's service quality is not very conclusive using the word cloud.

VADER Sentiment Scoring

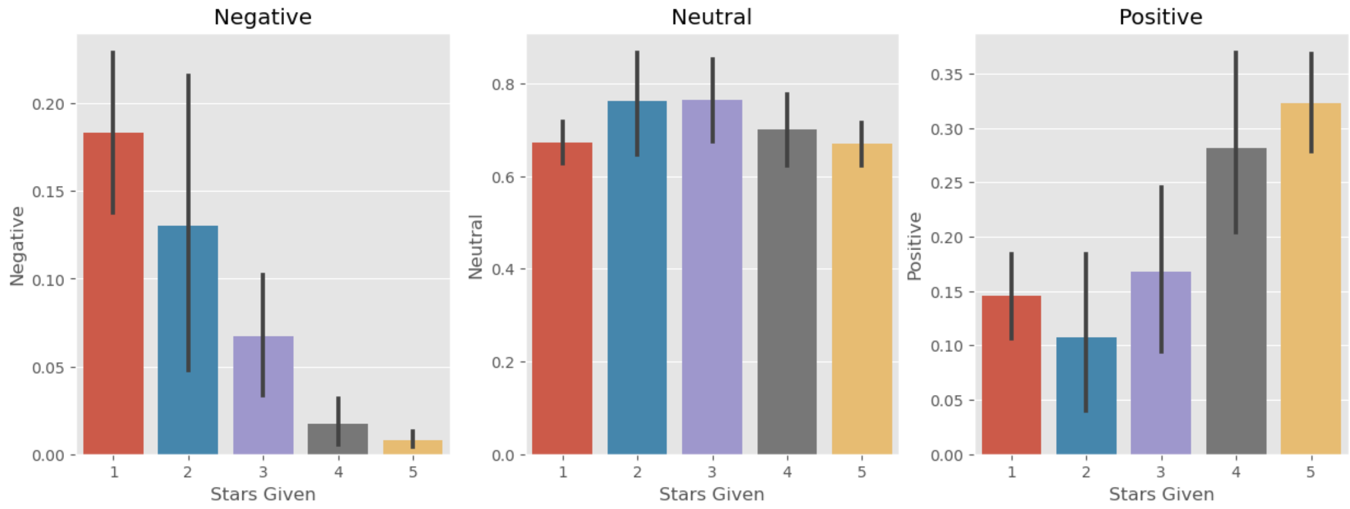
In VADER sentiment scoreing, the compound score range is from -1 to +1. Closer the compound score is to +1, more positive the written review. Closer the compound score is to -1, more negative the written review. Also, close the compound score is to 0, more neutral the written review. This method is also able to provide a score for positivity, negativity, and neutrality.

Fig 10. Compound Score by Stars Given



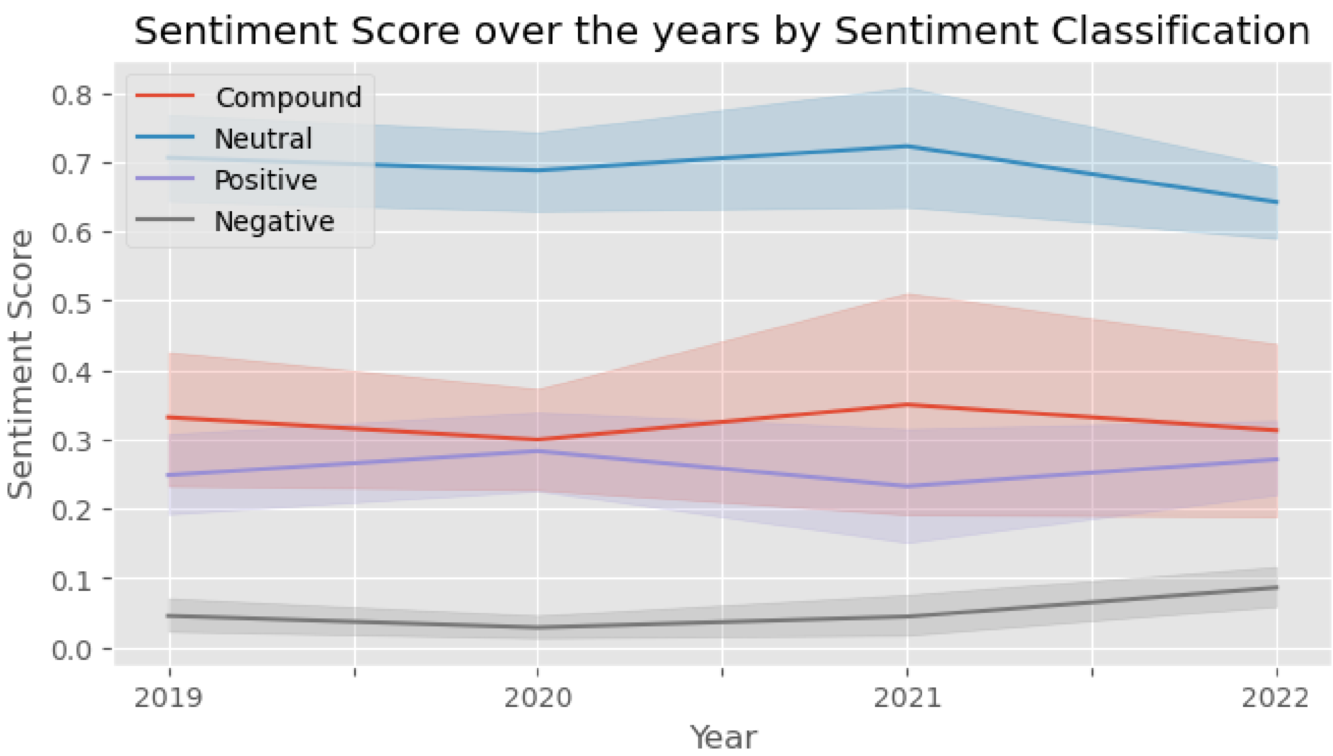
The bar chart indicates a left-skewed distribution regarding the compound score. The chart indicates a higher variation of negative to positive reviews with 1-star & 2-star rating. The latter star ratings indicate lower variation of slightly positive to very positive reviews. The compound score is higher for the higher star ratings, which is testament to a consistent service standard. Further testament to the consistent service standard is that none of the star ratings have a compound score of less than 0.

Fig 11. Compound Score by Sentiment Classification



Sentiment	Analysis
Negative	The chart is representative of a right-skewed distribution. Therefore, lower the star rating, the higher the negative score. This makes sense as the likelihood of a high negative score for a 3 to 5 star rating is unlikely. This is indicative of a consistent service standard.
Neutral	The chart is representative of a normal distribution. This indicates that there are written reviews with a variable star ratings but with a neutral review. This could be due to a lot of reviewers providing star ratings without a written review or a written review that is contains words that are inconclusive of sentiment.
Positive	The chart does not have a conclusive distribution. It is confusing how the positive score for 1-star rating exceeds that of 2-star rating. However, the positive score whilst moving to the latter star ratings shows a sharp increase. This is indicative of a somewhat consistent service standard, despite a contrast when it comes to the positive score for the 1-star rating.

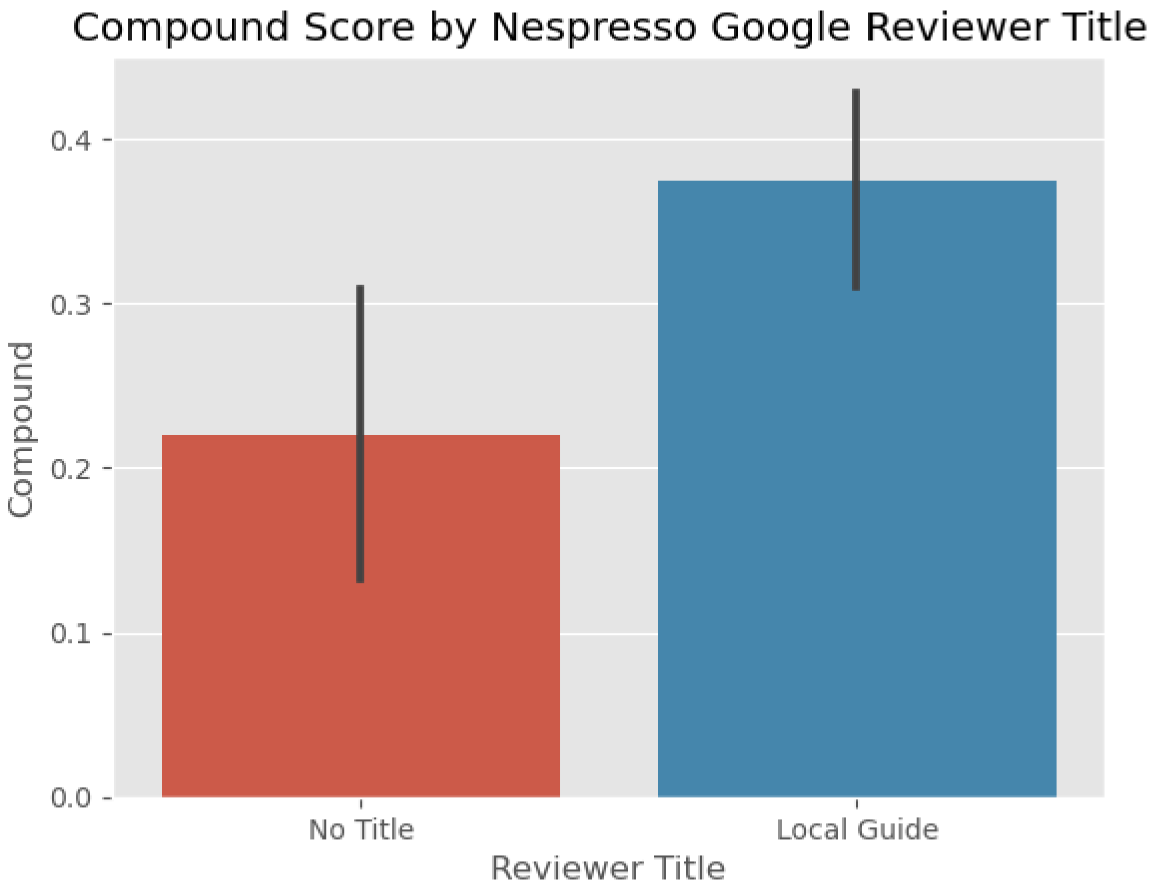
Fig 12. Sentiment Score over the years by Sentiment Classification



The compound score has fluctuated over the years but has remained less than 0.5, which infers that overall the reviews have not been very positive but moderately positive. Although, the moderate level of positive reviews have been consistent over the years. There was an slight decrease in 2020 with a slight uptick in 2021. As of 2022, there is a decrease in compound score that is slightly higher than that of 2020.

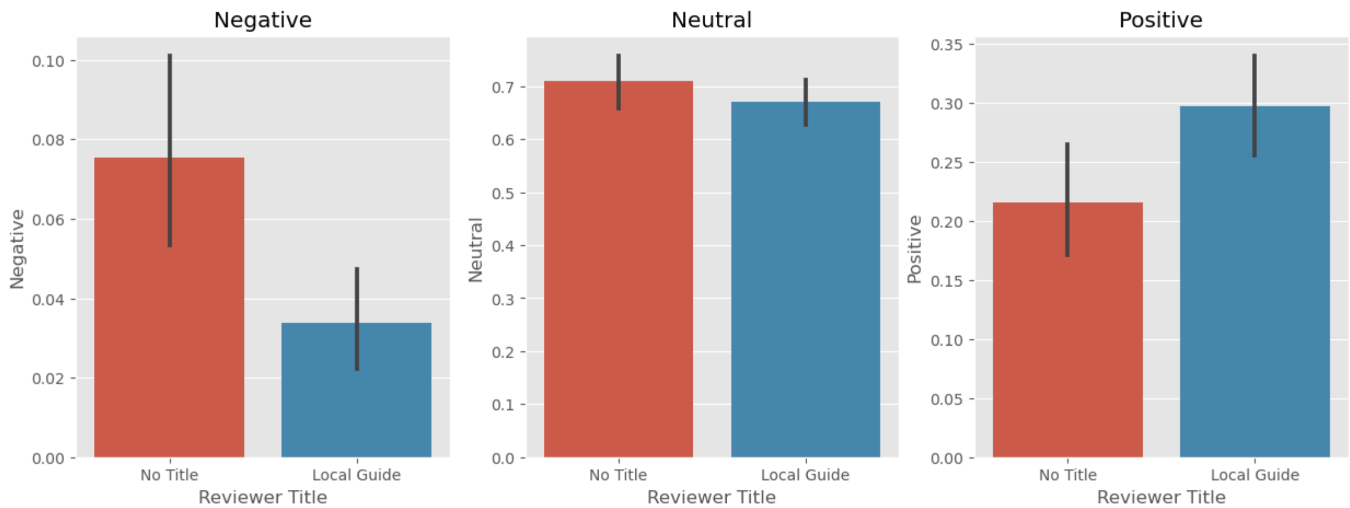
Sentiment	Analysis
Negative	There has been a slight uptick in negative score since 2020, which infers that there has been occurrences of negative customer service experiences.
Neutral	The neutral score has been very high over the years but has experienced a decrease in 2022. This could infer that the reviews received are due to impactful customer service experience irregardless if it is positive or negative.
Positive	The positive score has fluctuated over the years with a slight uptick in 2022, which almost matches 2020.

Fig 13. Compound Score by Reviewer Title



The bar chart suggests that reviewers with "Local Guide" status on Google Reviews overall provide more positive reviews compared to reviewers without "Local Guide" status. This is because compound score for 'Local Guide' is higher than that of 'No Title'. Thus, it can be inferred that the reviews written by those with "Local Guide" status are more constructive and/or reasonable comparatively to the reviewers without "Local Guide" status.

Fig 14. Reviewer Title Sentiment Score by Sentiment Classification



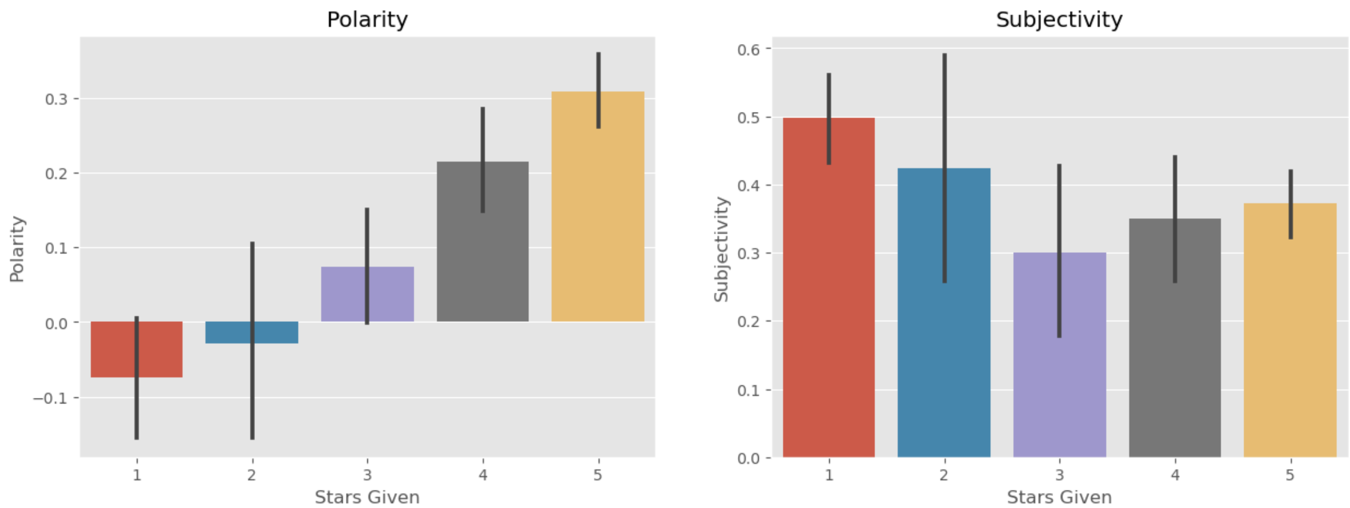
Sentiment Analysis

Sentiment	Analysis
Negative	The negative score for reviewers without "Local Guide" status far exceeds the negative score for those with "Local Guide" status. This can suggest that those without "Local Guide" status have a higher likelihood of writing harsher reviews.
Neutral	Interestingly, the neutral score for reviewers with "No Title" only slightly exceeds the neutral score for reviewers with "Local Guide" status. This can suggest that overall, there are a lot of written reviews are inconclusive in terms of sentiment along with a sizeable portion of reviewers not providing a written review.
Positive	The positive score for reviewers with "Local Guide" status exceeds the positive score for those without "Local Guide" status. This could suggest that reviewers with "Local Guide" status are more reasonable with their expectations.

Textblob Sentiment Scoring

In textblob sentiment scoring, there are two key components; polarity and subjectivity. Polarity functions similarly to the compound score in VADER sentiment scoring; it ranges from -1 to +1. Closer the polarity score is to +1, the more positive the written review is. Conversely, the closer the polarity score is to -1, the more negative the written review is. Subjectivity ranges from 0 to 1. The closer the subjectivity score is to 1, the more subjective the written review is. Conversely, the closer subjectivity score is to 0, the more objective the written review is.

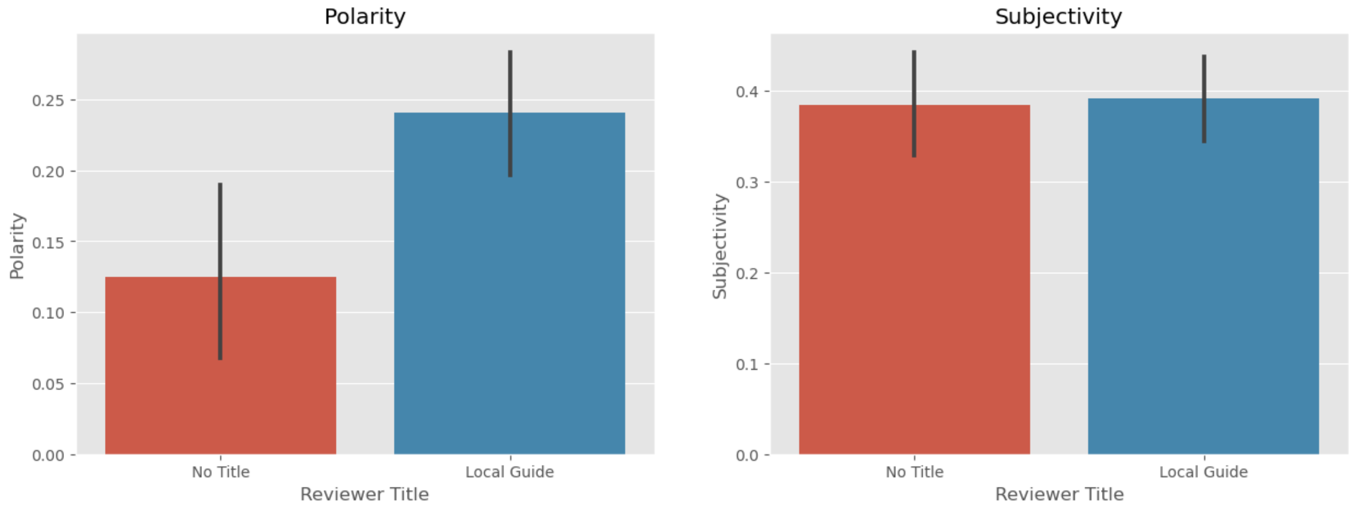
Fig 15. Polarity & Subjectivity by Star Rating



The polarity bar chart shows an understandable trend. The polarity score for 1-star and 2-star rating is below 0. In turn, all latter star ratings have a polarity score greater than 0. This suggests that reviewers that provide a 1-star and 2-star rating had a negative experience at Nespresso Metrotown based on the written reviews. Furthermore, reviewers that provided a 3-star to 5-star rating had a positive experience at Nespresso Metrotown based on the written reviews.

The subjectivty bar chart indicates that the reviews where there is a 1-star and 2-star rating are more objective compared to latter star ratings. This is because the subjectivity scores for 1-star and 2-star rating are highest. Ironically, the reviews that had an accompanying 3-star rating has the lowest subjectivity score, meaning that the respective reviews are most objective compared to the reviews of that of other star ratings.

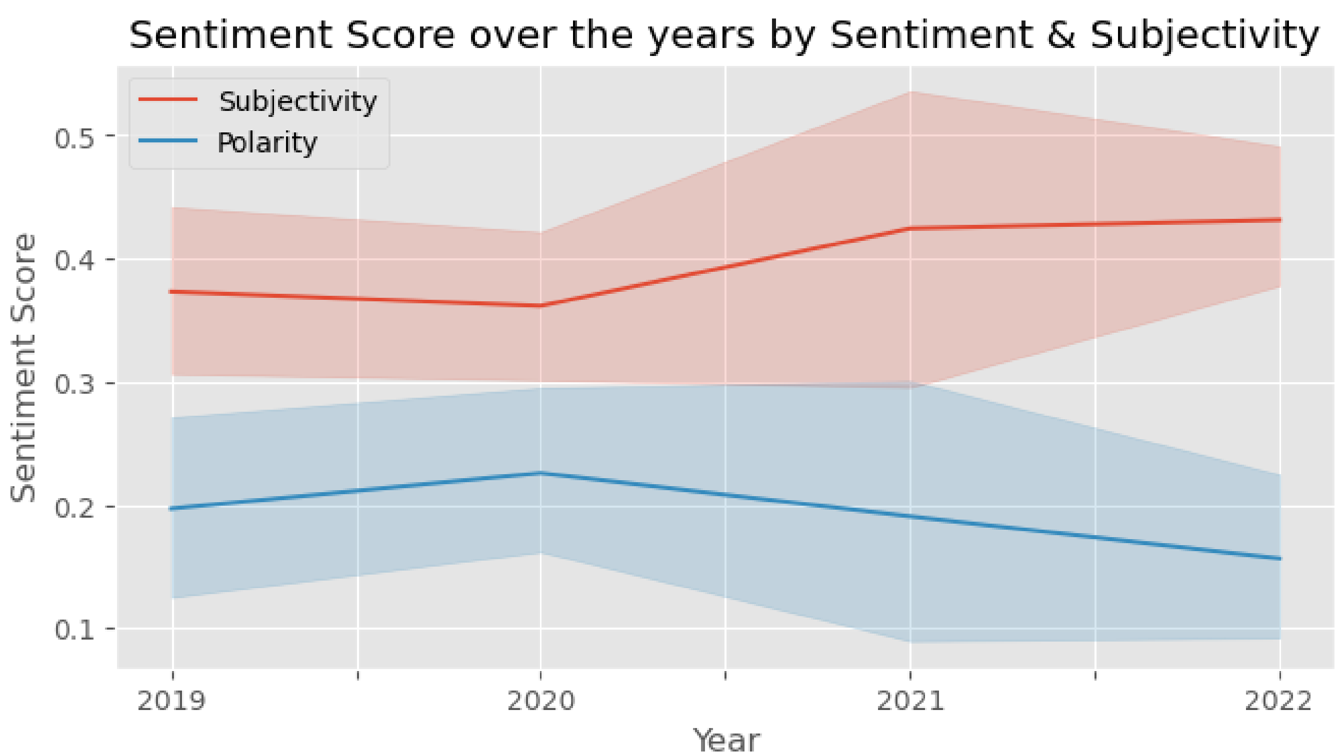
Fig 16. Polarity & Subjectivity by Reviewer Title



In the polarity bar chart, it is clear that reviewers with "Local Guide" status have a higher polarity score than that without "Local Guide" status. Thus, there has been a higher likelihood that reviewers with "Local Guide" status had a more positive experience at Nespressp Metrotown.

The subjectivty bar chart suggests that the subjectivity score for 'No Title' and 'Local Guide' reviewer titles is almost the same. This could infer that the level of subjectivity in the written reviews for both reviewers with and without "Local Guide" status is comparable.

Fig 17. Polarity & Subjectivity over the years

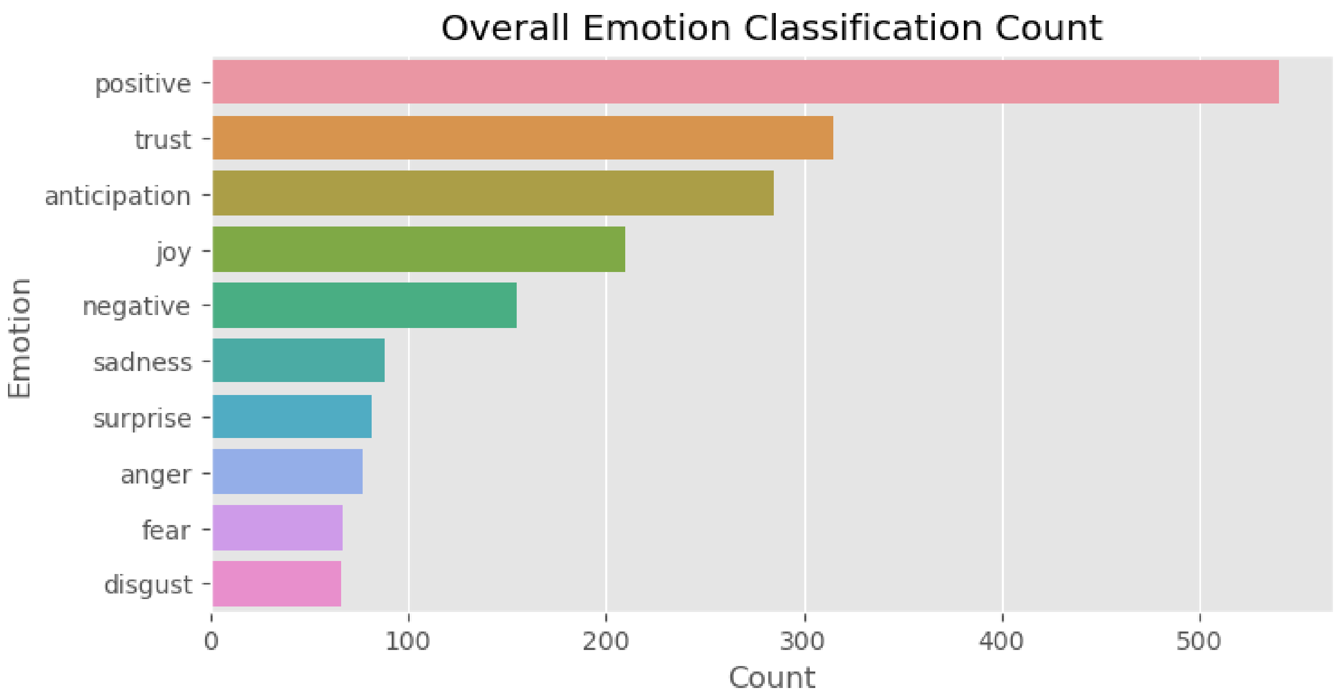


Interestingly there has been an uptick in 2020 in regards to polarity score, but steady dip whilst moving toward 2022. Thus, overall positive sentiment of Nespresso Metrotwon's service quality has decreased. On the other hand, subjectivity score increased by a lot moving towards 2022. This could suggest that Nespresso Metrotown's service quality has been leaving an impression on the reviewers.

Emotion Classification using Lexicon based method

In the Lexicon based method of emotion classification, 10 emotions can be determined. These 10 emotions include: positive, trust, anticipation, joy, negative, sadness, surprise, anger, fear, and disgust. These emotions are determined based on the words used in a written review, although it is prevelant that a given review can reflect more than one emotional sentiment. To maintain simplicity, the Emotion Count and Emotion Frequency will be considered in this part of the sentiment analysis exploration.

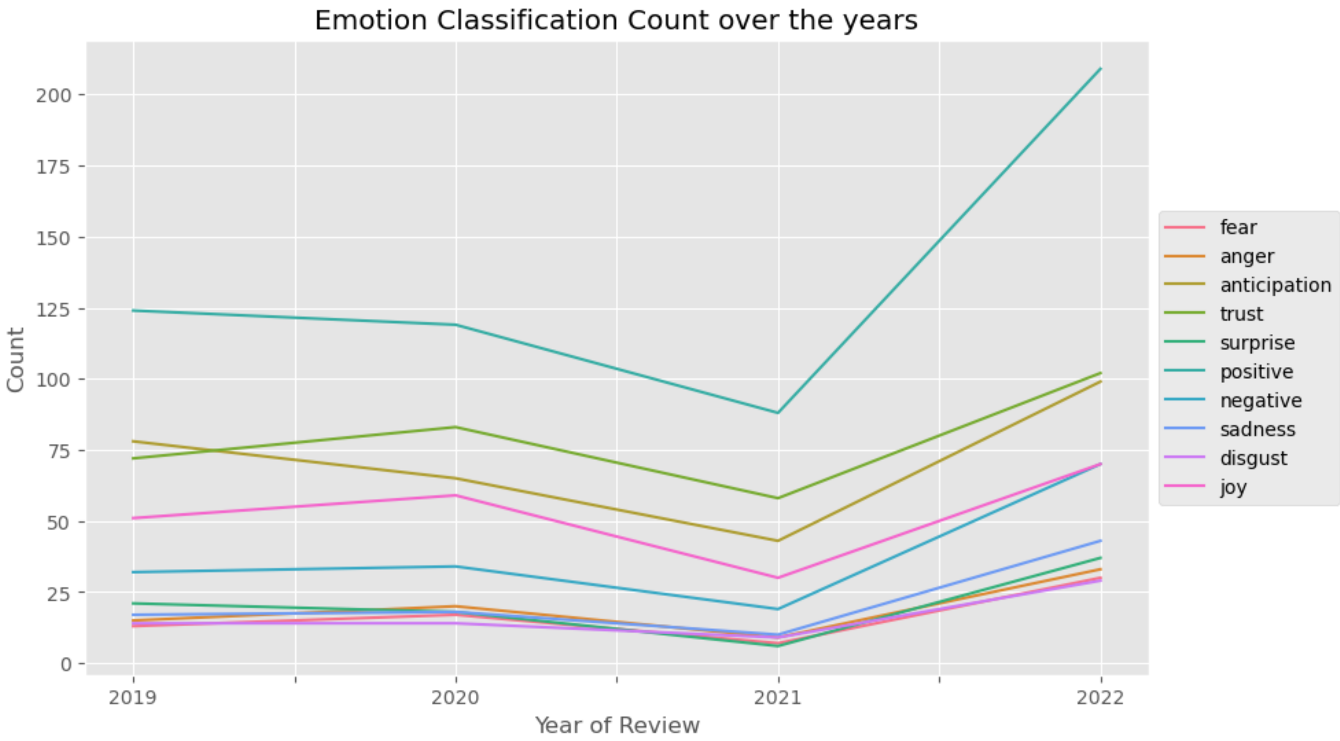
Fig 18. Overall Emotion Classification Count



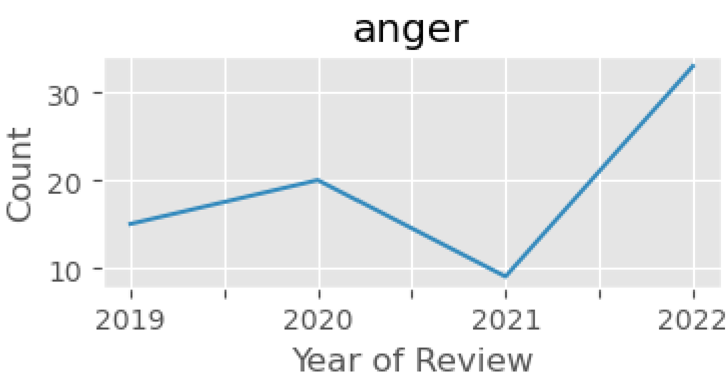
The bar chart indicates that customers that provided written reviews had more of a positive sentiment in regard to their overall experience at Nespresso Metrotown. This is also the case in regard to the other emotions related to positivity such as trust, anticipation, and joy. This is supported by the fact that the negative emotion along with all other related emotions have a lower emotion count than the aforementioned positive emotions.

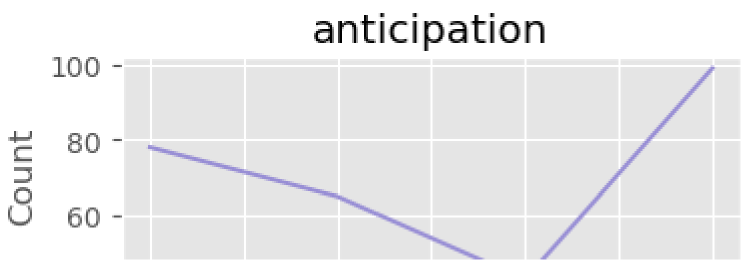
Fig 19. Emotion Classification Count over the years

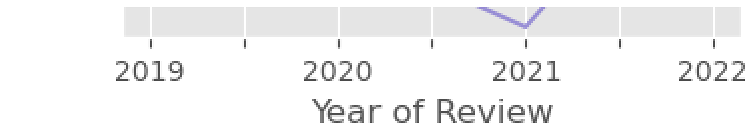
As there are 10 emotions, the Emotion Count over the years has been interpreted in two visualizations. The first chart considers all emotions experienced over the years in a single line chart. The second chart shows the count of each emotions over the years in each individual chart.

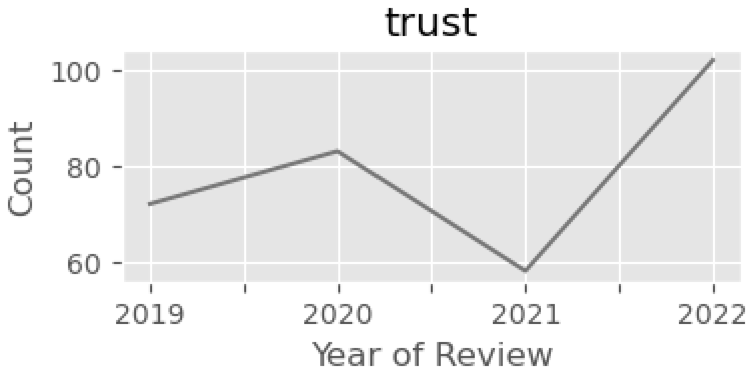


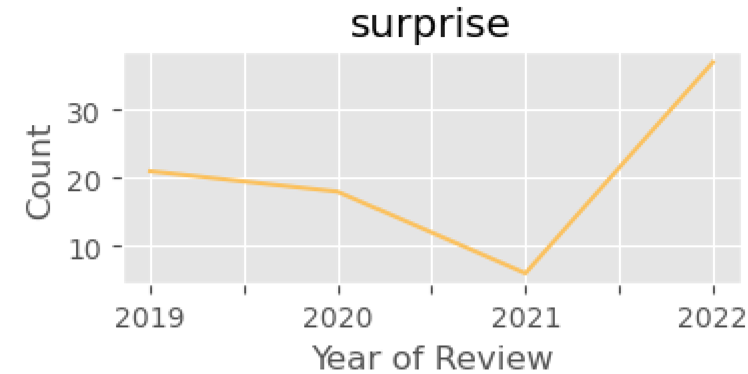
According to the chart above, customers have generally been experiencing the positive, surprise, anticipation, joy, trust, and negative, compared to others. It is a good sign, but there has been an uptick in the negative emotion along with other related emotions. It can be seen that the emotion count for 2021 is lower across all emotions due to limited service hours due to the Covid-19 pandemic.



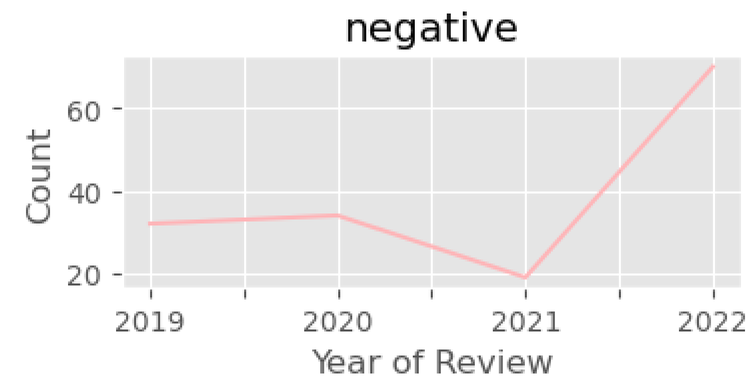


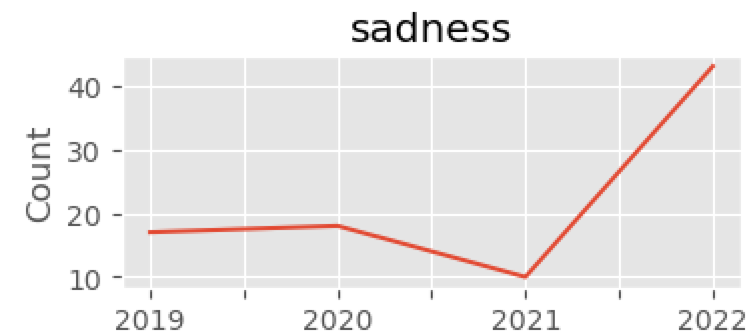

2019
2020
2021
2022
Year of Review

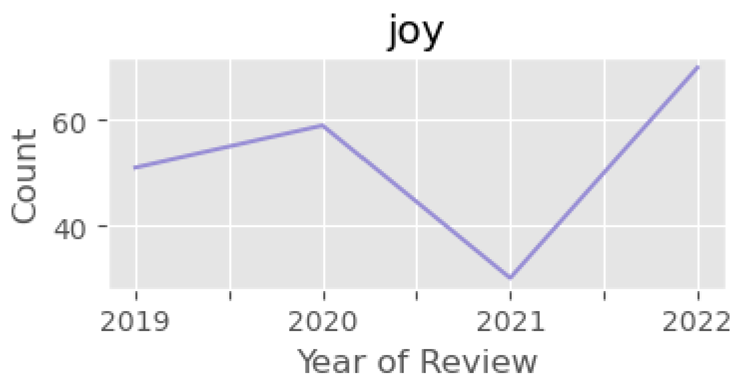

trust
Count
100
80
60
2019
2020
2021
2022
Year of Review


surprise
Count
30
20
10
2019
2020
2021
2022
Year of Review


positive
Count
200
150
100
2019
2020
2021
2022
Year of Review


negative
Count
60
40
20
2019
2020
2021
2022
Year of Review

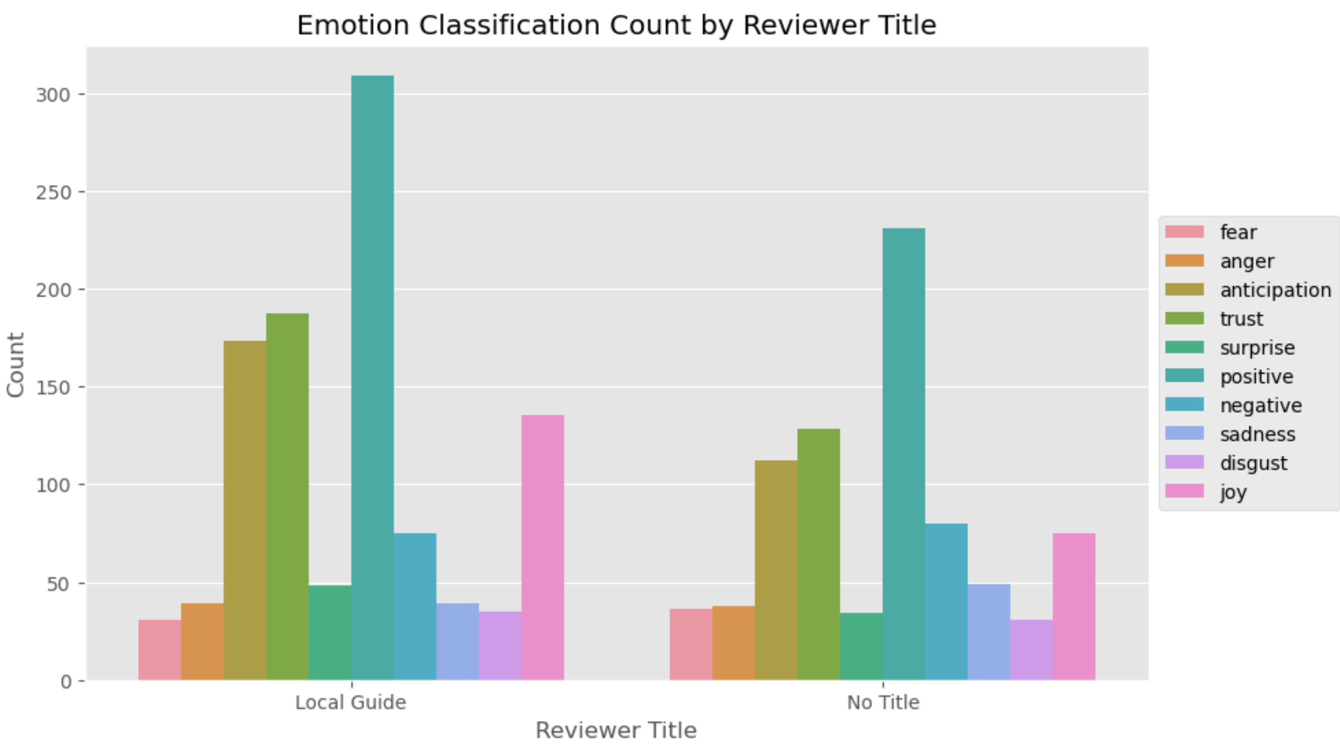

sadness
Count
40
30
20
10
2019
2020
2021
2022



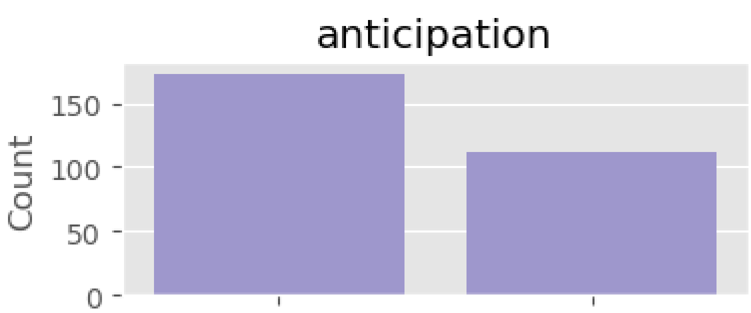
Emotion	Analysis
Fear	Fluctuation of fear in earlier years, and then a sharp increase in 2022.
Anger	Fluctuation of anger ine earlier years, but then a sharp increase in 2022.
Anticipation	A somewhat sharp decline of anticipation leading up to 2021, and then a sharp increase in 2022.
Trust	Fluctuation of trust in earlier years, and then a sharp increase in 2022.
Surprise	A somewhat steady decline of surprise leading up to 2021, and then a sharp increase in 2022.
Positive	A steady decline of positive emotion leading up to 2021, and then a sharp increase in 2022.
Negative	Slight fluctuation of negative emotion leading up to 2021, and then a sharp increase in 2022.
Sadness	Slight fluctuation of sadness leading up to 2021, and then a sharp increase in 2022.
Disgust	Slight fluctuation of disgust leading up to 2021, and then a sharp increase in 2022.
Joy	Fluctuation of joy in earlier years, and then a sharp increase in 2022.

Fig 20. Emotion Classification Count by Reviewer Title

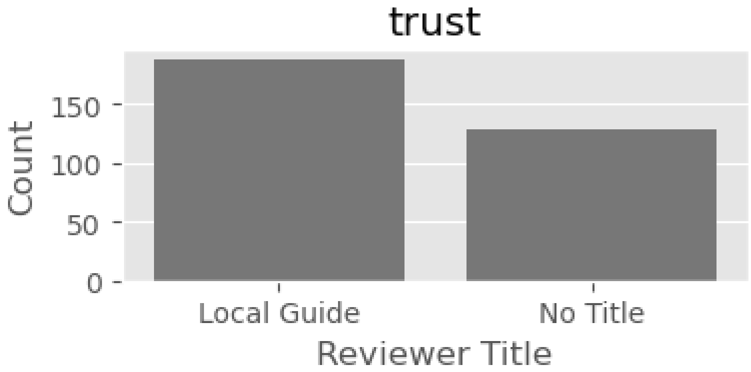
As there are 10 emotions, the Emotion Count by reviewer title has been interpreted in two visualizations. The first chart considers all emotions experienced by reviewer title in a single line chart. The second chart shows the count of each emotions by reviewer title in each individual chart.

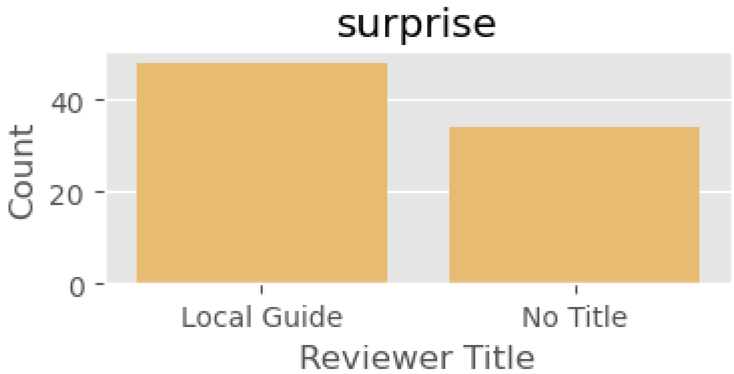


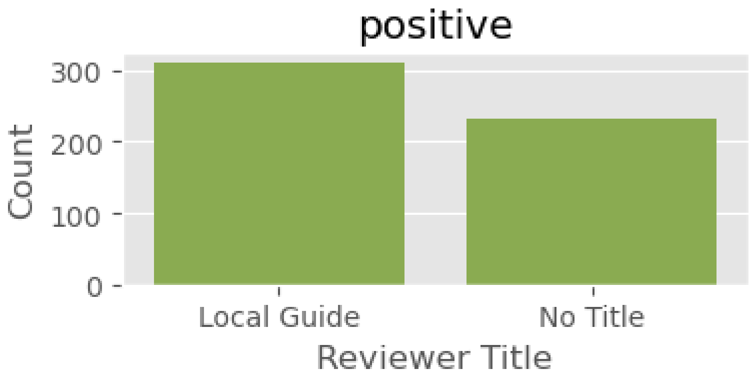
Based on the bar chart, it can be seen that reviewers with "Local Guide" status have a much higher emotion count for the positive emotion along with other related emotions such as anticipation, trust, and joy. All other emotions have comparable emotion counts across reviewer title.

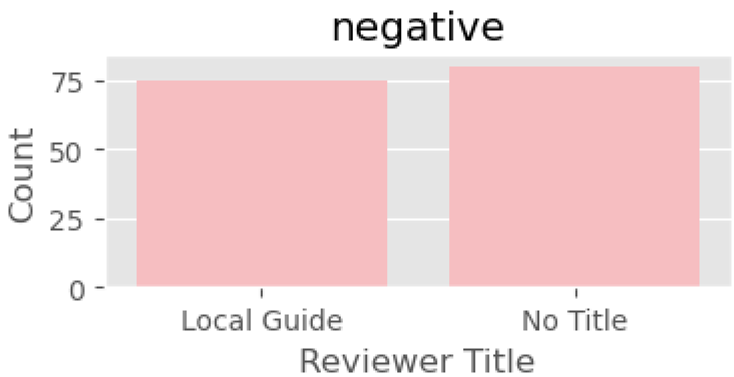


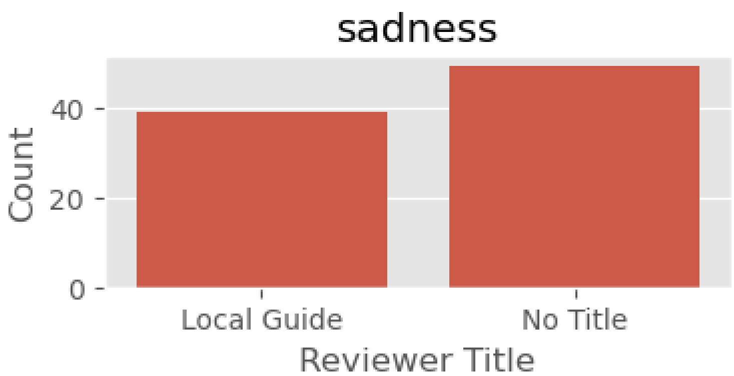

Local Guide
No Title
Reviewer Title

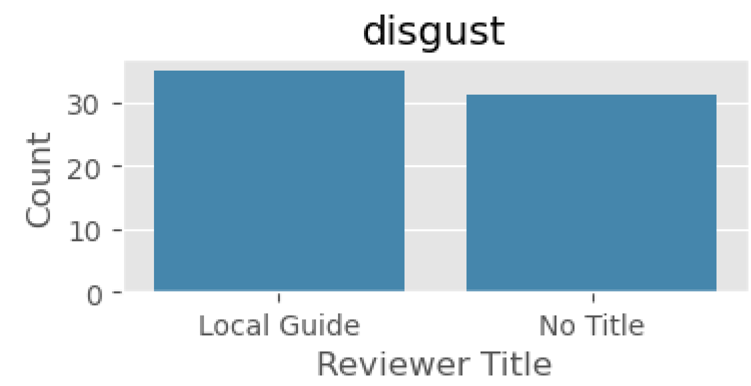

trust
Count
0
50
100
150
Local Guide
No Title
Reviewer Title

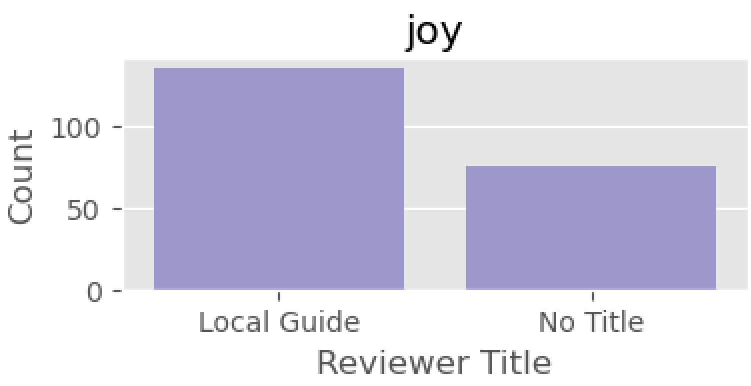

surprise
Count
0
20
40
Local Guide
No Title
Reviewer Title


positive
Count
0
100
200
300
Local Guide
No Title
Reviewer Title


negative
Count
0
25
50
75
Local Guide
No Title
Reviewer Title


sadness
Count
0
20
40
Local Guide
No Title
Reviewer Title

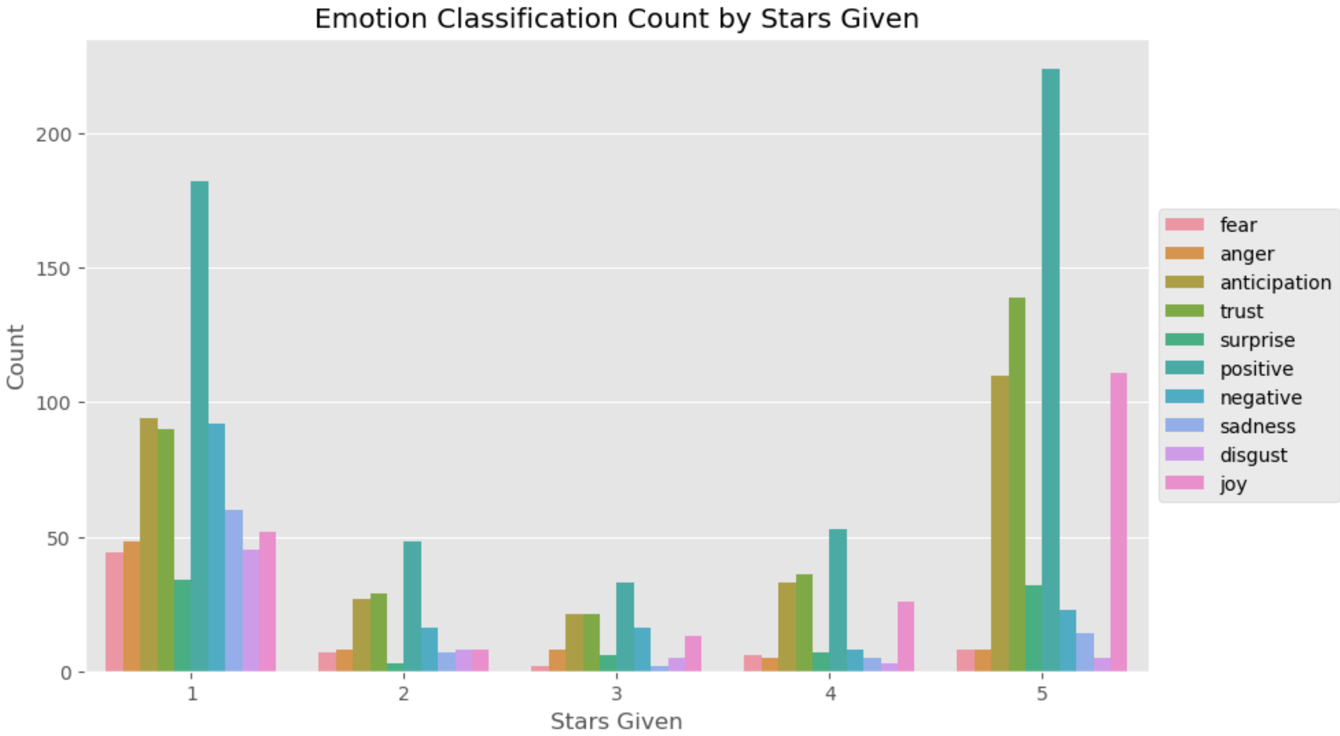




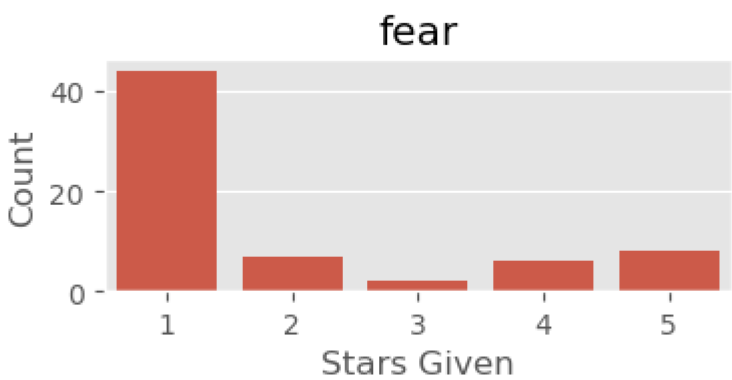
Emotion	Analysis
Fear	Reviewers with no title have experienced emotions of fear slightly more than reviewers with "Local Guide" status.
Anger	Reviewers with "Local Guide" status have experienced emotions of anger at a near identical level to that of reviewers with no title.
Anticipation	Reviewers with "Local Guide" status have experienced emotions of anticipation more than reviewers with no title.
Trust	Reviewers with "Local Guide" status have experienced emotions of trust more than reviewers with no title.
Surprise	Reviewers with "Local Guide" status have experienced emotions of surprise more than reviewers with no title.
Positive	Reviewers with "Local Guide" status have experienced positive emotions more than reviewers with no title.
Negative	Reviewers with "Local Guide" status have experienced slightly less negative emotions than reviewers with no title.
Sadness	Reviewers with "Local Guide" status have experienced less emotions of sadness than reviewers with no title.
Disgust	Reviewers with "Local Guide" status have experienced slightly more emotions of disgust than reviewers with no title.
Joy	Reviewers with "Local Guide" status have experienced much more emotions of joy than reviewers with no title.

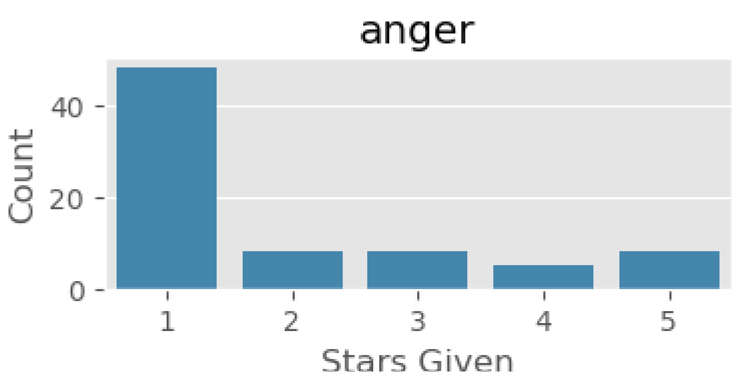
Fig 21. Emotion Classification Count by Stars Given

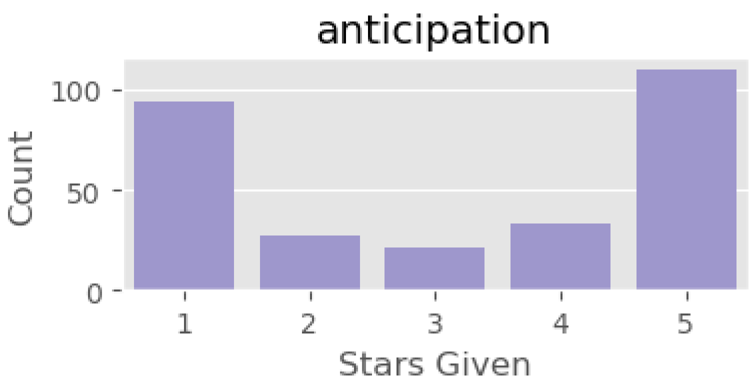
As there are 10 emotions, the Emotion Count by reviewer title has been interpreted in two visualizations. The first chart considers all emotions experienced by stars given in a single line chart. The second chart shows the count of each emotions by stars given in each individual chart.

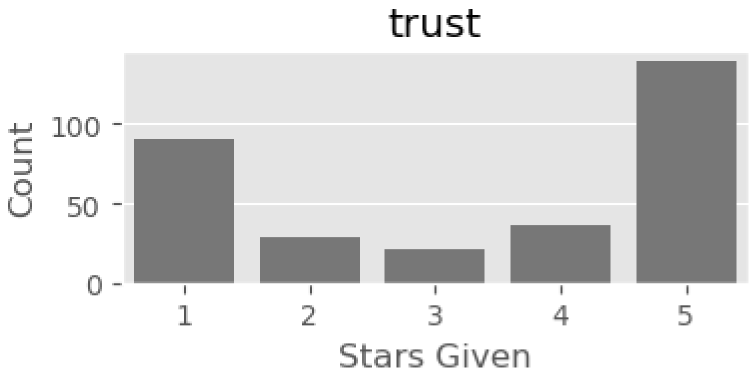


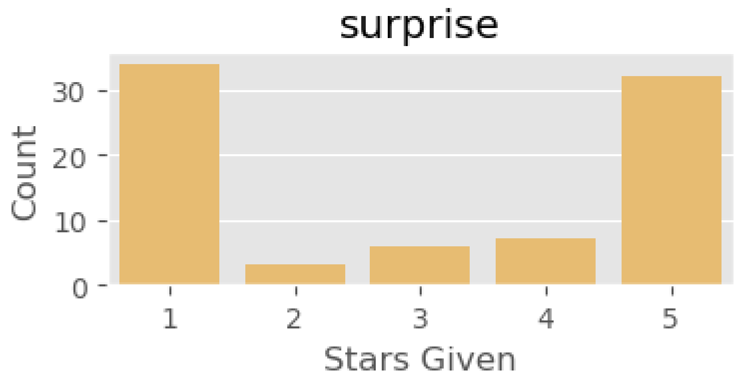
In the bar chart, it can be seen that most reviewers that gave both 1-star and 5-star rating experienced a high influx of various emotions. Although, the positive emotion and some related emotions have a relatively high count for 1-star rating compared to that of 5-star rating. In a way, this does not make sense unless the Lexicon based method interpreted a word to reflect one emotion but the written review in its entirely, as well as in terms of other words, reflects a different emotion or set of emotions.

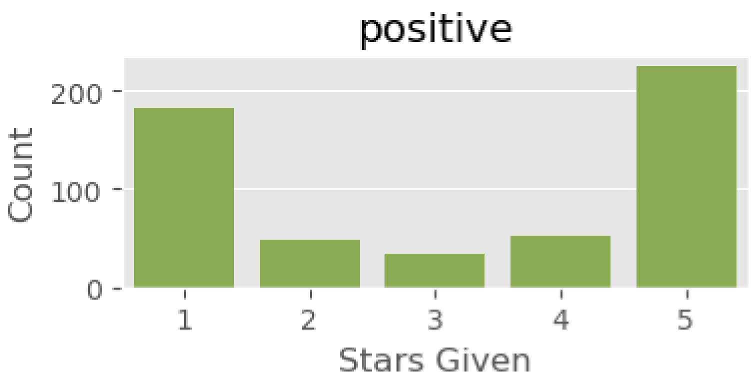


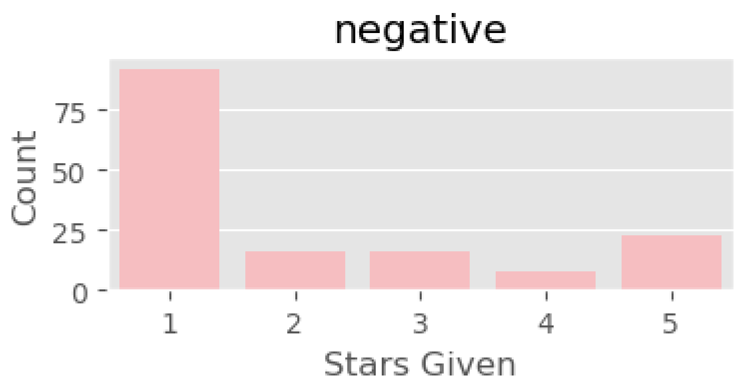



anticipation
Count
100
50
0
1
2
3
4
5
Stars Given

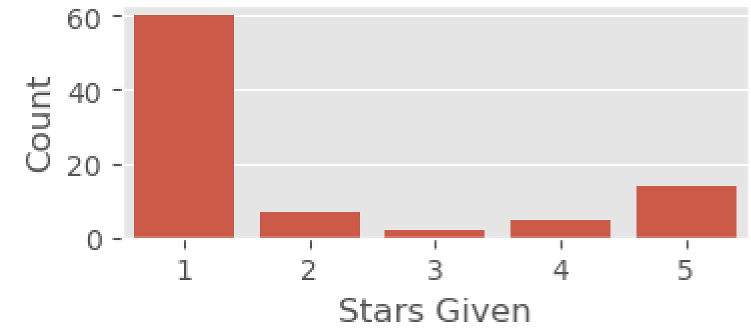

trust
Count
100
50
0
1
2
3
4
5
Stars Given

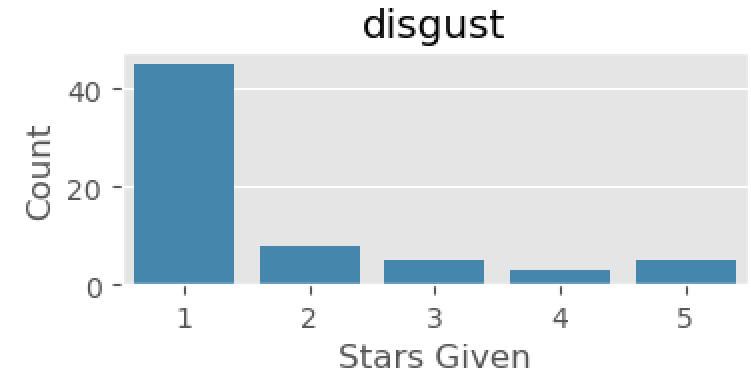

surprise
Count
30
20
10
0
1
2
3
4
5
Stars Given

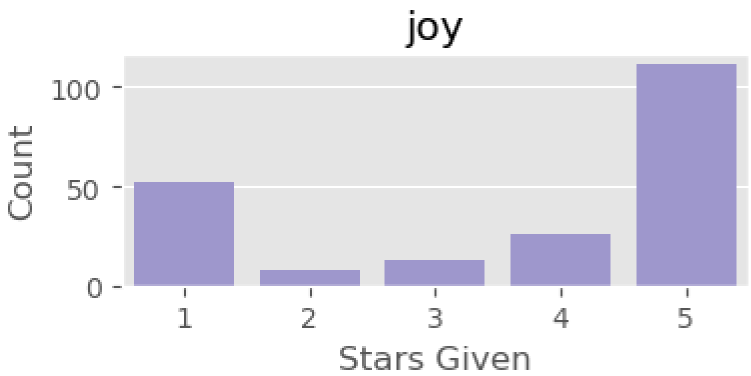

positive
Count
200
100
0
1
2
3
4
5
Stars Given


negative
Count
75
50
25
0
1
2
3
4
5
Stars Given

sadness



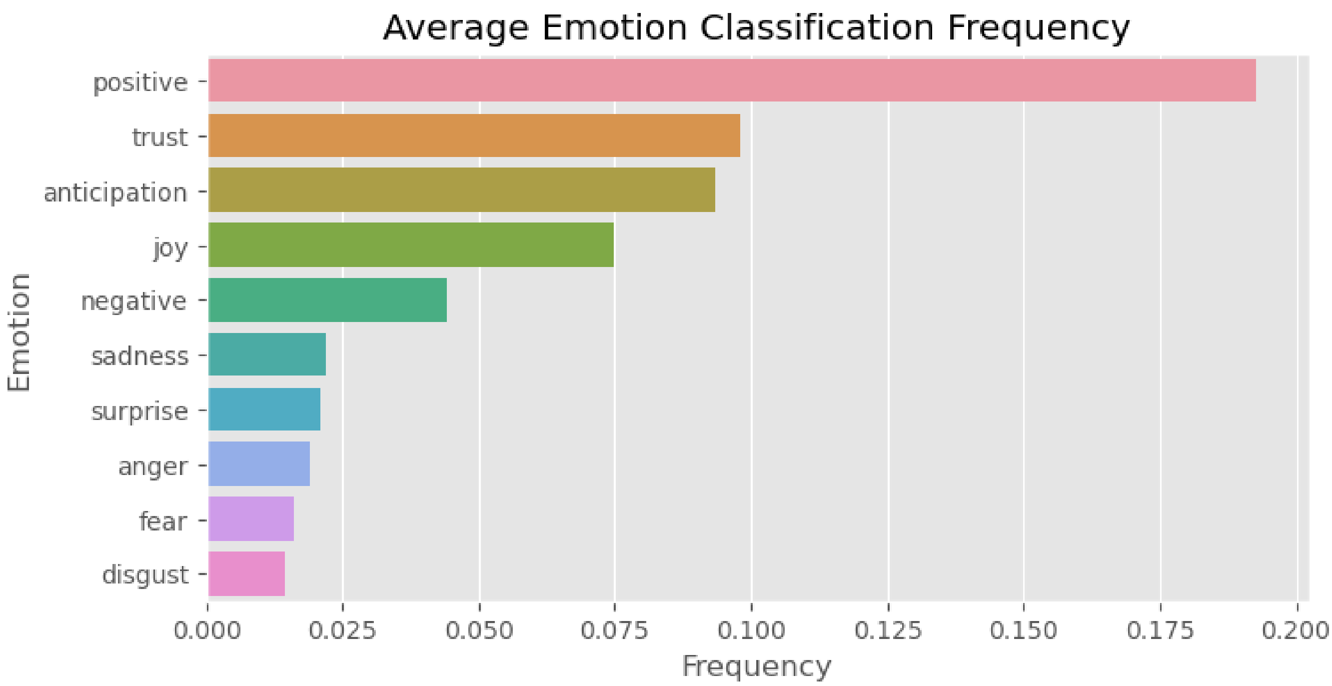




Emotion	**Analysis**
Fear	Reviewers that gave 1-star rating experienced the most emotions of fear, compared to those that gave other star ratings.
Anger	Reviewers that gave 1-star rating experienced the most emotions of anger, compared to those that gave other star ratings.
Anticipation	Reviewers that gave 1-star and 5-star rating experienced the most emotions of anticipation, compared to those that gave other star ratings.
Trust	Majority of reviewers that gave 5-star rating experienced the most emotions of trust, compared to those that gave other star ratings. Also, reviewers that gave 1-star ratings experienced relatively high emotions of trust as well.
Surprise	Reviewers that gave 1-star and 5-star rating experienced the most emotions of surprise, compared to those that gave other star ratings.
Positive	Reviewers that gave 1-star and 5-star rating experienced the most positive emotions, compared to those that gave other star ratings.

Emotion	Analysis
Negative	Reviewers that gave 1-star rating experienced the most negative emotions, compared to those that gave other star ratings.
Sadness	Reviewers that gave 1-star rating experienced the most emotions of sadness, compared to those that gave other star ratings.
Disgust	Reviewers that gave 1-star rating experienced the most emotions of disgust, compared to those that gave other star ratings.
Joy	Reviewers that gave 5-star rating experienced the most emotions of joy, compared to those that gave other star ratings.

Fig 22. Average Emotion Classification Frequency



This bar chart emulates the bar chart in figure 18, although the emotions are measured in terms of a classification frequency that is normalized. This helps to better determine the proportion of each emotion experienced throughout all written reviews. To reiterate, a higher proportion of the reviews reflect the positive emotion, along with related emotions such as trust, anticipation, and joy. Although, a smaller proportion of reviews reflect the emotion of surprise. A sizeable proportion of reviews reflect the negative emotion, although a lower proportion of reviews reflect the related emotions such as sadness, anger, fear, and disgust.

Sentiment Analysis Predictive Modeling

The intent of this part of the project is to essentially perform a classification experiment to determine which sentiment analysis method can be used as a basis to predict polarity of written reviews. In addition, utilize the textblob dataset to determine whether a written review is objective or subjective. In a real world context, this part of the project attempts to address what predictive models could be used by Nespresso Canada to instantenously retrieve a predicted sentiment from a written Google Review.

Vectorization methods are necessary to convert the words in the ***Review Cleaned*** column into numerical values to be utilized in a pipeline prior to applying a classification model. Two types of vectorization methods were used:

1. *Term Frequency Inverse Frequency (TF-IDF)*: This method takes into consideration both the frequency and importance of a words in a written review.
2. *Bag of Words*: This method takes into consideration only the frequency of words in a written review.

After the words in the ***Review Cleaned*** column are vectorized, the features would be prepared to use in the classification model. Below are the four types of classification models that were used:

1. *Logisitic Regression*: The idea is to find a relationship between features and probability of particular outcome.
2. *Multinomial Naive Bayes*: Considers that the presence of a given feature is independent & unrelated to that of all other feautures.
3. *Decision Tree*: Can be used to categorize based on prior set of questions that were answered, and follows a tree-like structure with roots, nodes, and branches.
4. *Support Vector Machine*: Considers that the presence of a given feature may have a relationship to a certain degree to all other feautures.

Please refer to the Sentiment Analysis Predictive Modeling Jupyter Notebook to view the codebase of the sentiment analysis predictive modeling.

Experiment - Positive, Negative, Neutral

Below are the steps taken in this section of the project.

1. Import necessary Python packages to be used for sentiment analysis predictive modeling.
2. Construct a function that delivers an accuracy score using a pipeline based on the prespecified dataset reflecting a sentiment analysis method, vectorization method, features (i.e., cleaned reviews column), target (i.e., prediction result), proportion of training & test set, random state, and the actual classification model that is used to make a prediction.
3. Using similar parameters to the previous step, construct a function that delivers the accuracy score along with the a classification report and a confusion matrix.
4. Import the datasets that reflect each of the sentiment analysis methods as Pandas dataframes. Take a peek at the data, and show an informative summary of each dataset.
5. Perform manual classification for each dataset to determine the actual target for polarity.

    - Manual classification for polarity using vaders dataset.

Determinant	Classification
Compound Score == 0	Neutral
Compound Score > 0	Positive
Compound Score < 0	Negative

    - Manual calssification for polarity using textblob dataset.

Determinant	Classification

Determinant	Classification
Polarity Score == 0	Neutral
Polarity Score > 0	Positive
Polarity Score < 0	Negative

- Manual classification for polarity using lexicon (emotion count) dataset.

  Since this sentiment analysis method involves ten emotions, the emotions are binned into categories that align with the polarty classification levels of positive, negative, and neutral.

Emotions	Emotion Bin
Trust, Surprise, Positive, Joy	Positive
Fear, Anger, Anticipation, Negative, Sadness, Disgust	Negative

  The emotion counts are accumulated for each emotion bin to determine a positive score and negative score. If the positive score and negative score is 0 or equal, then the classification is neutral.

Determinant	Classification
Positive Score == Negative Score	Neutral
Positive Score > Negative Score	Positive
Negative Score < Positive Score	Negative

6. Conduct an experiment to predict the polarity classification and determine the accuracy score with each combination of sentiment analysis dataset, vectorization method, and classification model. Output the results of the experiment in a Pandas dataframe, and then reorder results in descrnding order of accuracy score.
7. Using the combination of sentiment analysis dataset, vectorization method, and classification model that yielded the highest accuracy score, output the accuracy score (again) along with the classification report and confusion matrix.

In the forthcoming sub-sections, the results from the experiment are shown and interpreted.

Fig 23. Polarity Experiment Accuracy Scores

Below are the prediction accuracy scores for each combination of sentiment analysis method, vectorization method, and classification model ordered by most accurate (i.e., descending order of accuracy score). The results in the following table helped to determine which combination of sentiment analysis method (dataset), vectorization method is the most accurate to be used for further prediction.

Sentiment Analysis Method	Vectorization Method	Classification Model	Pipe Score
Emotion Lexicon	Bag-of-Words	Decision Tree	0.915888
Textblob	Bag-of-Words	Decision Tree	0.915888

Sentiment Analysis Method	Vectorization Method	Classification Model	Pipe Score
Textblob	TF-IDF	Decision Tree	0.906542
Emotion Lexicon	TF-IDF	Decision Tree	0.906542
Emotion Lexicon	Bag-of-Words	Multinomial Naive Bayes	0.906542
Textblob	Bag-of-Words	Logistic Regression	0.906542
Textblob	Bag-of-Words	Multinomial Naive Bayes	0.906542
Emotion Lexicon	Bag-of-Words	Logistic Regression	0.906542
Vaders	Bag-of-Words	Multinomial Naive Bayes	0.887850
Emotion Lexicon	TF-IDF	Multinomial Naive Bayes	0.869159
Textblob	TF-IDF	Multinomial Naive Bayes	0.869159
Vaders	Bag-of-Words	Support Vector Machine	0.859813
Textblob	TF-IDF	Logistic Regression	0.859813
Emotion Lexicon	TF-IDF	Logistic Regression	0.859813
Textblob	Bag-of-Words	Support Vector Machine	0.850467
Vaders	TF-IDF	Logistic Regression	0.850467
Vaders	TF-IDF	Support Vector Machine	0.850467
Emotion Lexicon	Bag-of-Words	Support Vector Machine	0.850467
Textblob	TF-IDF	Support Vector Machine	0.841121
Vaders	Bag-of-Words	Logistic Regression	0.841121
Emotion Lexicon	TF-IDF	Support Vector Machine	0.841121
Vaders	TF-IDF	Decision Tree	0.831776
Vaders	TF-IDF	Multinomial Naive Bayes	0.831776
Vaders	Bag-of-Words	Decision Tree	0.803738

Fig 24. Polarity Experiment Result 1

The following combination of sentiment analysis dataset, vectorization method, and classification model yielded one of the highest prediction accuracy scores. Although, when performing the predictive modeling again and outputing the classification results, the accuracy score was less compared to performing the initial experiment. This is strange despite maintaining the same training-to-test set ratio and random state. This could infer that the emotion lexicon sentiment analysis method may be lacking in terms of valid classification, or the manual classification utilized to determine an actual target polarity for the written review was not a valid approach.

Sentiment Analysis Method	Vectorization Method	Classification Model

Sentiment Analysis Method	Vectorization Method	Classification Model
Emotion Lexicon	Bag-of-Words	Decision Tree

> Accuracy Score = **0.6915887850467289**

The accuracy score above indicates that the sentiment analysis method, vectorization method, and classification model does not yield a highly accurate prediction for polarity. Typically, predictive model with an accuracy score greater than 0.80 would typically reflect a high degree of accuracy.

Classification Report:

	Precision	Recall	F1-Score	Support
Negative	0.50	0.47	0.49	19
Neutral	0.74	0.83	0.78	41
Positive	0.72	0.66	0.69	47
Accuracy			0.69	107
Macro Average	0.65	0.65	0.65	107
Weighted Average	0.69	0.69	0.69	107

- Precision
  - Out of all of the written reviews that the model predicted to be negative, only 50% were actually negative.
  - Out of all of the written reviews that the model predicted to be neutral, only 74% were actually neutral.
  - Out of all of the written reviews that the model predicted to be positive, only 72% were actually postive.
- Recall
  - Out of all of the written reviews that actually were negative, the model only predicted this polarity level correctly for 46% of the written reviews.
  - Out of all of the written reviews that actually were neutral, the model only predicted this polarity level correctly for 83% of the written reviews.
  - Out of all of the written reviews that actually were positive, the model only predicted this polarity level correctly for 66% of the written reviews.
- F1-Score
  - As the F1-Score is only 49% for the negative polarity level, the model is not very accurate when correctly predicting written reviews that are actually negative.
  - As the F1-Score is 78% for the neutral polarity level, the model is somewhat accurate when correctly predicting written reviews that are actually neutral.
  - As the F1-Score is 69% for the positive polarity level, the model is somewhat accurate when correctly predicting written reviews that are actually positive.
- Support
  - Out of 107 written reviews in the test set, 19 of the written reviews were determined to be negative as per the model.

- Out of 107 written reviews in the test set, 41 of the written reviews were determined to be neutral as per the model.
- Out of 107 written reviews in the test set, 47 of the written reviews were determined to be positive as per the model.

Confusion Matrix:

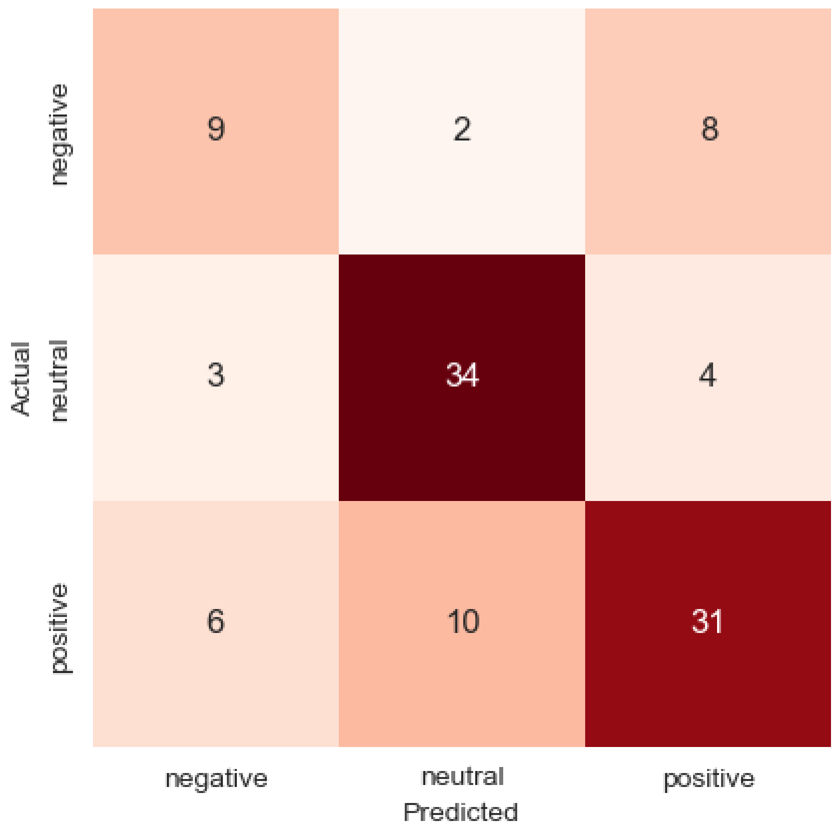


Fig 25. Polarity Experiment Result 2

The following combination of sentiment analysis dataset, vectorization method, and classification model yielded one of the highest prediction accuracy scores. Unlike the previous polarity experiment result, the classification prediction was performed again and yieled exactly the same accuracy score as in Figure 24. In short, the results makes sense. In turn, this would be the predictive model that could potentially be used for further business use case.

Sentiment Analysis Method	Vectorization Method	Classification Model
Textblob	Bag-of-Words	Decision Tree

> Accuracy Score = **0.9158878504672897**

The accuracy score above indicates that the sentiment analysis method, vectorization method, and classification model yields a highly accurate prediction for polarity. The accuracy score is above 0.80, which is an indicator for validity.

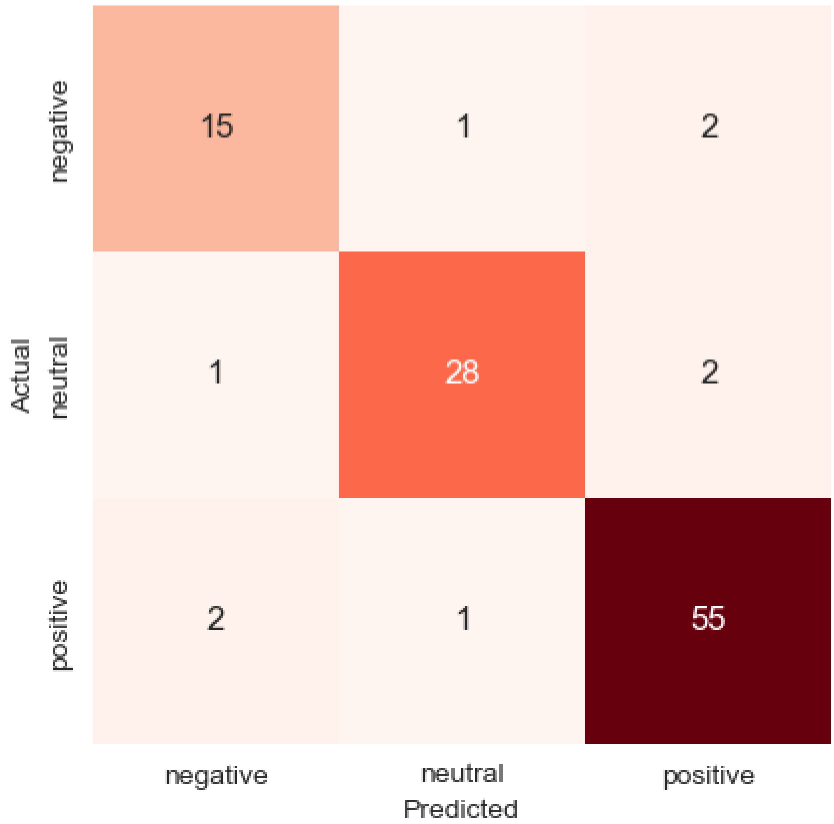
Classification Report:

	Precision	Recall	F1-Score	Support

	Precision	Recall	F1-Score	Support
Negative	0.83	0.83	0.83	18
Neutral	0.93	0.90	0.92	31
Positive	0.93	0.95	0.94	58
Accuracy			0.92	107
Macro Average	0.90	0.89	0.90	107
Weighted Average	0.92	0.92	0.92	107

- Precision
  - Out of all of the written reviews that were predicted to be negative, only 83% of the written reviews are actually negative.
  - Out of all of the written reviews that were predicted to be neutral, only 93% of the written reviews are actually neutral.
  - Out of all of the written reviews that were predicted to be positive, only 93% of the written reviews are actually positive.
- Recall
  - Out of all of the written reviews that actually were negative, the model only predicted this polarity level correctly for 83% of the written reviews.
  - Out of all of the written reviews that actually were neutral, the model only predicted this polarity level correctly for 92% of the written reviews.
  - Out of all of the written reviews that actually were positive, the model only predicted this polarity level correctly for 94% of the written reviews.
- F1-Score
  - As the F1-Score is only 83% for the negative polarity level, the model is quite accurate when correctly predicting written reviews that are actually negative.
  - As the F1-Score is only 92% for the negative polarity level, the model is very accurate when correctly predicting written reviews that are actually neutral.
  - As the F1-Score is only 94% for the negative polarity level, the model is not very accurate when correctly predicting written reviews that are actually positive.
- Support
  - Out of 107 written reviews in the test set, 18 of the written reviews were determined to be negative as per the model.
  - Out of 107 written reviews in the test set, 31 of the written reviews were determined to be neutral as per the model.
  - Out of 107 written reviews in the test set, 58 of the written reviews were determined to be postive as per the model.

Confusion Matrix:



Experiment - Constructive & Subjective Review

Below are the steps taken in this section of the project.

1. Import necessary Python packages to be used for sentiment analysis predictive modeling.
2. Construct a function that delivers an accuracy score using a pipeline based on the prespecified dataset reflecting a sentiment analysis method, vectorization method, features (i.e., cleaned reviews column), target (i.e., prediction result), proportion of training & test set, random state, and the actual classification model that is used to make a prediction.
3. Using similar parameters to the previous step, construct a function that delivers the accuracy score along with the a classification report and a confusion matrix.
4. Import the datasets that reflect each of the sentiment analysis methods as Pandas dataframes. Take a peek at the data, and show an informative summary of each dataset.
5. Perform manual classification using the textblob dataset to determine the actual target for subjectivity.

Determinant	**Classification**
Polarity Score == 0	Inconclusive
Polarity Score != 0 **&** Subjectivity Score < 0.5	Constructive
Polarity Score != 0 **&** Subjectivity Score > 0.5	Subjective

6. Conduct an experiment to predict the subjectivity classification and determine the accuracy score with each combination of vectorization method and classification model. Output the results of the experiment in a Pandas dataframe, and then reorder results in descending order of accuracy score.
7. Using the combination of the vectorization method and classification model that yielded the highest accuracy score, output the accuracy score (again) along with the classification report and confusion

matrix.

In the forthcoming sub-sections, the results from the experiment are shown and interpreted.

Fig 26. Subjectivity Experiment Accuracy Scores

Below are the prediction accuracy scores for each combination of vectorization method and classification model ordered by most accurate (i.e., descending order of accuracy score). The results in the following table helped to determine which combination of vectorization method and classification method is the most accurate to be used for further prediction

Vectorization Method	Classification Model	Pipe Score
TF-IDF	Multinomial Naive Bayes	0.803738
TF-IDF	Logistic Regression	0.757009
Bag-of-Words	Multinomial Naive Bayes	0.747664
Bag-of-Words	Decision Tree	0.738318
Bag-of-Words	Logistic Regression	0.710280
TF-IDF	Support Vector Machine	0.682243
TF-IDF	Decision Tree	0.644860
Bag-of-Words	Support Vector Machine	0.626168

Fig 27. Subjectivity Experiment Result

The following combination of the textblob dataset, vectorization method, and classification model yielded the highest prediction accuracy score. Based on the results of the experiment, the following predictive model could be reasonable for further business use. Although, whilst reading the classification report closely, the results do not make sense in terms of precision, recall, support, and F1-Score.

Vectorization Method	Classification Model
TF-IDF	Multinomial Naive Bayes

> Accuracy Score = **0.8037383177570093**

The accuracy score is around 0.80, which is indicative of a reasonable level of accuracy.

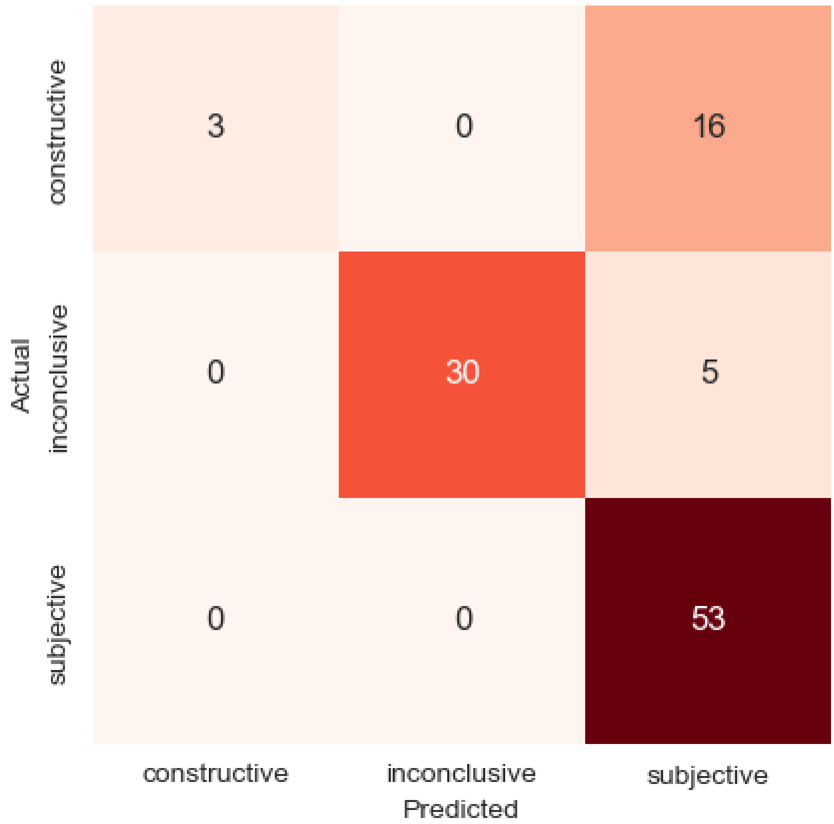
Classification Report:

	Precision	Recall	F1-Score	Support
Constructive	1.00	0.16	0.27	19
Inconclusive	1.00	0.86	0.92	35
Subjective	0.72	1.00	0.83	53

	Precision	Recall	F1-Score	Support
Accuracy			0.80	107
Macro Average	0.91	0.67	0.68	107
Weighted Average	0.86	0.80	0.76	107

- Precision
    - Out of all of the written reviews that the model predicted to be constructive, 100% of the written reviews are actually constructive.
    - Out of all of the written reviews that the model predicted to be inconclusive, 100% of the written reviews are actually inconclusive.
    - Out of all of the written reviews that the model predicted to be subjective, only 72% of the written reviews are actually subjective.
- Recall
    - Out of all the written reviews that were constructive, the model only predicted this subjectivity level correctly for 16% of the written reviews.
    - Out of all of the written reviews that were inconclusive, the model only predicted this subjectivity level correctly for 86% of the reviews.
    - Out of all of the written reviews that were subjective, the model only predicted this subjectivity level correctly for 100% of the reviews.
- F1-Score
    - As the F1-Score is only 27% for the constructive subjectivity level, the model is not accurate at all when predicting written reviews that are actually constructive.
    - As the F1-Score is only 92% for the inconclusive subjectivity level, the model is not accurate at all when predicting written reviews that are actually inconclusive.
    - As the F1-Score is only 83% for the subjective subjectivity level, the model is not accurate at all when predicting written reviews that are actually subjective.
- Support
    - Out of 107 written reviews in the test set, 19 of the written reviews were determined to be constructive as per the model.
    - Out of 107 written reviews in the test set, 35 of the written reviews were determined to be inconclusive as per the model.
    - Out of 107 written reviews in the test set, 53 of the written reviews were determined to be subjective as per the model.

Confusion Matrix:



Conclusion

From the analysis conducted in the exploratory data analysis and sentiment analysis exploration sections, the Nespresso Metrotown branch has been performing quite well. The overall star rating is 3.9 when considering written reviews from 2019 to 2022. The analysis conducted in the aforementioned sections support this, however there has been a fluctuation and incosistency in terms of maintaining a consistent level of service quality. That being said, most of the customer experiences have been positive.

The main takeaway from all of the analysis performed in the previous sections of this project is that Nespresso Metrotown must maintain a level of consistency in its service standard and quality of customer service. This is especially imperative over the long term as product and service lineup change & evolve, and so too does the team of Coffee Specialists, Team Leaders, and Managers. Below are some thoughts and potential recommendations.

- Defining a unique philosophy for customer service that aligns with Nestle Nespresso's mission statement would be advantageous. As there are several competing Nespresso locations in the lower mainland, having a unique philosphy for customer service could transform Nespresso Metrotown's service standard and differentiate itself from other competing locations. This way, Nespresso Metrotown can become the preferred-definitive location for all coffee lovers in the lower mainland. Therefore, despite each member of staff having a different style of selling, the service quality will be the same across the board. Personally, I am a strong believer in having a *customer-first* mentality, and customer satisfaction is my primary motivator. Regardless of whether a customer purchases a small order, a big order, or is simply seeking information, I want the customer to leave the store satisfied with the intent to visit the store again in the near future. I try my best to engage the customer when they first walk-in either by physically approaching them or waving to them. This shows initiative as customers appreciate that we operate at a higher standard than other retail stores. Large sales are

meaningless without long-term customer retention through optimal customer service, creating memorable moments with positivity & enthusiasm, and showing them that **we care**.
    - Fun note: As we share a similar business-to-customer store operations model to Apple, I like to tell them "that I consider Nespresso to be the **Apple** of coffee" to entice them about our service standard. As a joke, I also tell them the only thing we lack is a Genius bar but we can accomodate them with almost everything else!
- Re-introducing a formal training program Coffee Specialists would be meaningful. Building product knowledge and facilitating customer transactions effieciently takes places organically over time, however understanding an organization's history, mission statement, and core competencies must be solidified early on. This way, a Coffee Specialist can truly embody the values of the organization. This shows through the customer interaction via the Coffee Specialist's interpersonal skills and enthusiasm. To strengthen the foundation of sales techniques and customer service, role playing situations could be beneficial to explore, enhance, and diversify the craft of the Coffee Specialist.
- A core aspect to provide excellent customer service is for there to be ***symbiosis between both Team Leaders and Coffee Specialists***. In my perspective, Team Leaders are to be seen as role models and mentors who have ascended beyond the bounds of a Coffee Specialist to a heightened role where their presence has a significant impact on business operations. Coffee Specialists can be viewed as the first point-of-contact for customers and can be considered to be the entities that perform the core business operations. In challenging moments, Coffee Specialists look to Team Leaders for guidance and wisdom, as well as assistance to properly attend customers in the event Coffee Specialists do not have the ability/clearance to fulfill the need of the customer. Two important aspects of the relationship between a Team Leader and Coffee Specialist are *trust* and *approachability*. Despite it being common practice for new hires to shadow an experienced Coffee Specialist, it would be key that new hires also shadow each Team Leader as well, albeit for a limited time. This fosters trust and can help new hires integrate well into the team as well as understand the hierarchy. Furthermore, Team Leaders once upon a time used to be a Coffee Specialist, and so the new hires are receiving direct guidance from those that are considered the best. In short, strong & supportive cross-hierarhical relationship between Team Leaders and Coffee Specialists leads to the culmination of meaningful and efficient customer experiences.

The recommendations above originated not only from my experience working on this project, but also through my time working as a Coffee Specialist as well. That being said, there is much more I have yet to learn. My attempt at applying machine learning processes & models to assess customer sentiment regarding Nespresso Metrotown's service quality was interesting. That being said, not all models are entirely accurate; although they help us get close to the truth. Thus, as humans we must use our own judgement to interpret our findings and make clinical decisions.