# Coffee Makers Customer Review Analysis

Kelly Wang, Jacqueline Hsu

## Research Questions and Results Summary

In this project, we will compare three coffee makers in the market: Nespresso Vertuo Plus Coffee and Espresso Maker, Keurig K-Cafe Special Edition Single Serve K-Cup Pod Coffee, Latte and Cappuccino Maker, and Mr. Coffee Espresso and Cappuccino Maker.

Our research questions and summary are as the following:

1. Is there a correlation between the number of helpful votes and rating? Do people find negative reviews or positive reviews more helpful?
    - There is no meaningful correlation between the number of helpful votes and rating. However, the extremes (1 star and 5 stars) tend to have more helpful votes. People generally find positive reviews more helpful.
2. When do customers submit the most reviews? Do people submit more reviews during the holiday season? How does customers' sentiment (positive review rate) change over time for each product?
    - Customers tend to submit more reviews towards the beginning and end of the year. There are more reviews submitted during and after the holiday season, Black Friday and Christmas to be specific. There isn't an identifiable trend between customers' sentiment and time.
3. For each product, what do most customers think about their purchase? Which coffee maker stands out as the most efficient or user-friendly? (For the customer reviews of each coffee maker, what are their most appeared positive and negative keywords?)
    - For all three coffee makers, people who chose to submit a rating either love or hate the coffee maker product they purchased. Based on what customers have commented, Nespresso's coffee maker stands out as the most efficient whereas Keurig and Mr. Coffee's coffee makers stand out as more user-friendly.

## Motivation

Our motivation for this research is to understand how three different coffee makers compare to one another. By analyzing product reviews using the several aspects of each product, we could gain a holistic understanding of which product stands out and is preferred by customers. The main implication of our research conclusion is to help users recognize which coffee maker best suits their needs, hence would be the ideal purchase. Both team members are regular coffee drinkers, so the results would be insightful for us too.

Aside from this, we are also interested in learning a new Python tool that can be applied to future studies and works. Hence, we chose web scraping because of its various uses such as monitoring customer behavior, data aggregation, enriching machine learning models, etc.

## Dataset

Our data will be scraped from Amazon.com, one of the largest E-commerce sites. In particular, we are looking at three coffee makers (with frother) sold on Amazon: Nespresso Vertuo Plus, Keurig K-Cafe, and Mr. Coffee Cafe Barista. Below are the URLs we will be using.

Nespresso's Vertuo Plus:
https://www.amazon.com/Nespresso-VertuoPlus-Espresso-DeLonghi-Aeroccino/product-reviews/B01MTZ419O/ref=cm_cr_arp_d_paging_btm_next_1?ie=UTF8&reviewerType=all_reviews&pageNumber=1

Keurig K-Cafe:

https://www.amazon.com/Keurig-Single-Serve-K-Cup-Special/product-reviews/B07J5FV7WS/ref=cm_cr_arp_d_paging_btm_next_1?ie=UTF8&reviewerType=all_reviews&pageNumber=1

Mr. Coffee Cafe Barista:

https://www.amazon.com/Mr-Coffee-Espresso-Cappuccino-Barista/product-reviews/B007K9OIMU/ref=cm_cr_arp_d_paging_btm_next_1?ie=UTF8&reviewerType=all_reviews&pageNumber=1

From each of these sources, we will be using their Reviews section, which consists of review date, product detail, star ratings, comments, etc.

The `write_csv.py` file writes three CSV files: `nespresso_reviews.csv`, `keurig_reviews.csv`, and `mr_coffee_reviews.csv`. The three datasets for the three products are tabular data in CSV format. Each row in the dataset represents a customer review submission. The columns are described as the following:

- review_date: The date of customer review submission
- rating: Customer rating on the scale of 1-5
- helpful_vote: Number of people who found the review helpful
- review_title: The title of the customer review, which summarizes review_body
- review_body: The main content of the customer review

## Method

Before we did any data analysis to answer the research questions, we first scraped customer review data from Amazon with the following steps:

1. Find URL of products
2. Web scrape data with the 'Beautiful Soup' library for each product (URL)

3. Extract the first page of customer reviews (sorted by top reviews) and any data that is relevant to answer the questions (review_date, rating, helpful_vote, review_title, review_body)
4. To extract data from multiple pages of customer reviews, break the URL into parts (texts and page numbers) and iterate through the first 10 pages/URLs.
5. Clean data to ensure each dataset is usable and reliable for further analysis
6. Write extracted data into a CSV file

Then, to perform data analysis and answer our research questions, we follow these steps:

1. Is there a correlation between the number of helpful votes and rating? Do people find negative reviews or positive reviews more helpful?
    a. Create a scatter plot that shows the relationship between the number of helpful votes and rating.
    b. For each product, classify reviews into "positive" and "negative" with "rating > 3" as positive, "rating < 3" as negative, and "rating == 3" as neutral and make a new dataframe with an added 'sentiment' column
    c. Create a bar chart that shows the number of helpful votes each sentiment category has
2. When do customers submit the most reviews? Do people submit more reviews during the holiday season? How does customers' sentiment (positive review rate) change over time for each product?
    a. For each review, extract the month portion of the 'review_date' and add it into a new 'review_month' column in the dataframe
    b. Create a bar chart that displays the number of customer reviews submitted in each calendar month.
    c. For each product, classify reviews into "positive" and "negative" with "rating > 3" as positive, "rating < 3" as negative, and "rating == 3" as neutral and make a new dataframe with an added 'sentiment' column
    d. Count the total number of reviews by month and the total number of reviews with a positive sentiment by month
    e. Compute each month's positive review rate by dividing the positive sentiment reviews count by the total number of reviews
    f. Visualize positive review rate over each month of the year using a point plot
3. Which coffee maker stands out as the most efficient or user-friendly? (For the customer reviews of each coffee maker, what are their most appeared positive and negative keywords?)

a. Create a customer rating histogram that shows how most customers rate the products for each coffee maker to see the rating distribution, use plotly for interactive data visualization

b. Filter out any stop words like articles, prepositions, pronouns, or conjunctions that do not add much meaningful information to the text

c. Count the number of times each word appeared and find the first 30 most frequently appeared words in 'review_title'

d. Create a bar chart that displays the most frequently appeared words in the product review comments

e. For each product, classify reviews into "positive" and "negative" with "rating > 3" as positive, "rating < 3" as negative, and "rating == 3" as neutral and make a new dataframe with an added 'sentiment' column

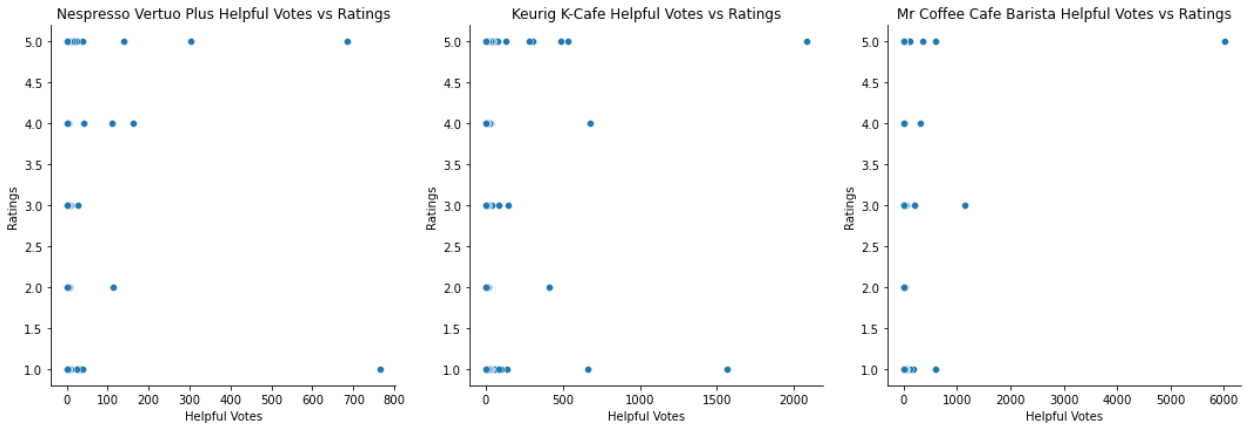
f. For each coffee maker, find the first 30 most frequently appeared words separately for positive and negative reviews and create a bar chart for each category

g. Based on the bar chart, look for the most frequently appeared keywords that are relevant to efficiency (e.g. "convenient", "time-saving", etc.) or user-friendly ("easy-to-use", "easy-to-clean", etc.) and compare customer opinions between each product based on the bar charts

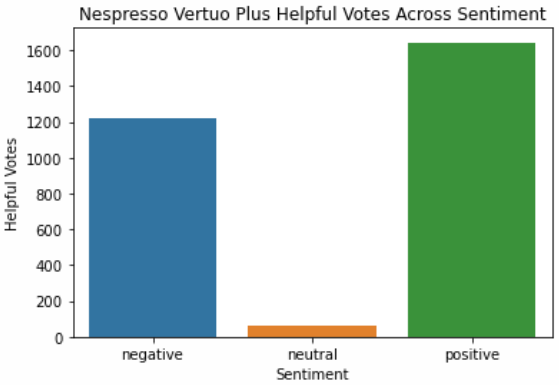
## Results

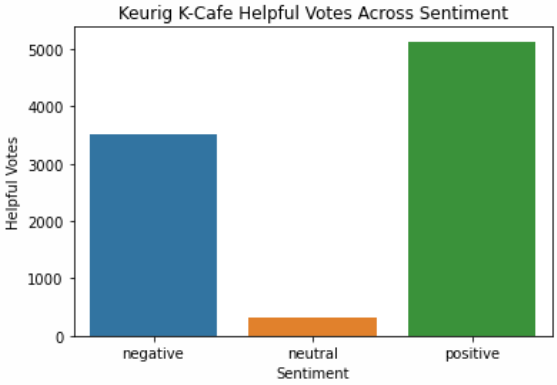
### Research Question 1

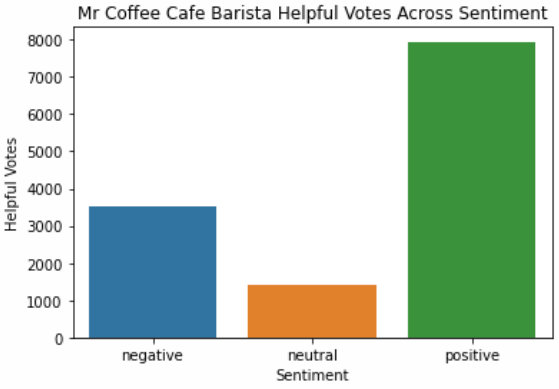
Is there a correlation between the number of helpful votes and rating? Do people find negative reviews or positive reviews more helpful?



To answer this research question, we first made a scatterplot for each of the products that graphed the star ratings against how many helpful votes they received. At first, looking at the graph, there isn't an obvious relationship between the two. For all three products, there seem to be many helpful votes associated with all five possible star ratings. However, reviews with the highest helpful votes tend to be those that rate the product either a one-star or a five-star.



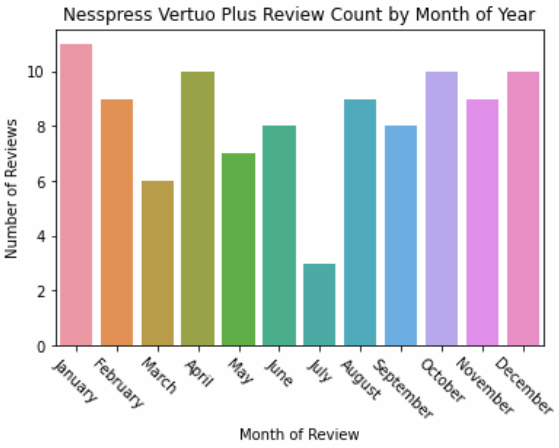

Next, we looked at whether customers find positive or negative reviews more helpful. This was done by categorizing the reviews by sentiment then distributing them using histograms. From the histograms, it can be interpreted that the positive reviews (with a 4 or 5-star rating) were found to be more helpful than reviews that were neutral or negative. This may mean that positive reviews contribute more to whether a customer wants to purchase the coffee maker. As a result, brands may rely more on these reviews to boost their purchases, which is why some offer free samples in exchange for positive feedback. Another commonality between these histograms is that neutral reviews (with a 3-star rating) are not favored. This is because extremes - in this case, negatives and positives - tend to have a long-lasting impression in people's minds.

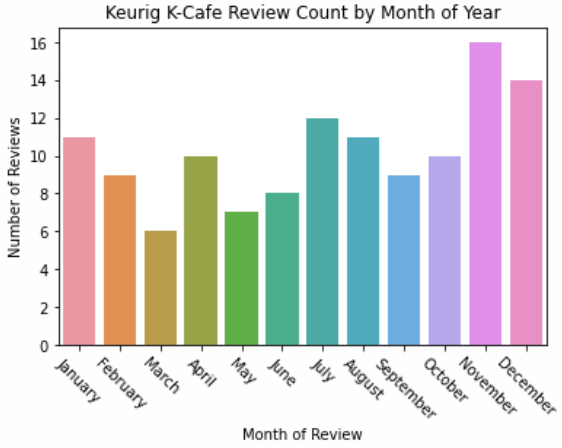
**Research Question 2**

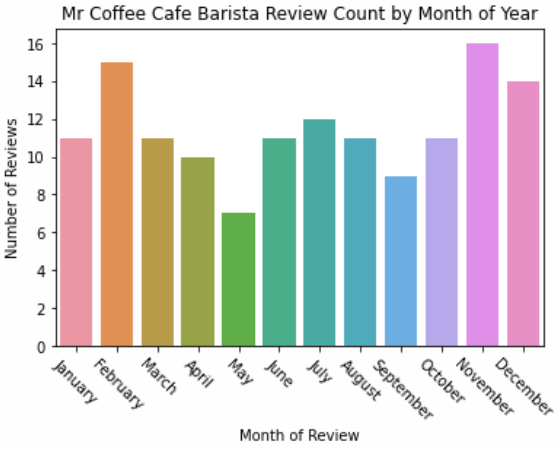
When do customers submit the most reviews? Do people submit more reviews during the holiday season? How does customers' sentiment (positive review rate) change over time for each product?

To answer this research question, we first made three histograms, one for each product. These histograms graphed each month's review count. Generally, all three plots displayed more review towards the beginning and end of the year, with a slight dent in the middle. However, the trends cannot be explicitly described with any distribution (normal, skew, or uniform) because there is not an obvious relationship.

For Nespresso, the number of reviews peaked in January, whereas for both Keurig and Mr. Coffee, the review count peaked in November. Both months can be related to some holiday season. November is the month of Black Friday when the biggest sale of the year occurs. This is when customers tend to "clear their shopping carts" and spend an excessive amount on non-necessities, such as coffee makers. Similarly, January is the month right after Christmas. Christmas is all about gift shopping for friends and family. By the time they receive the present and use it, it might be January already, which makes the month they are reviewing make sense.

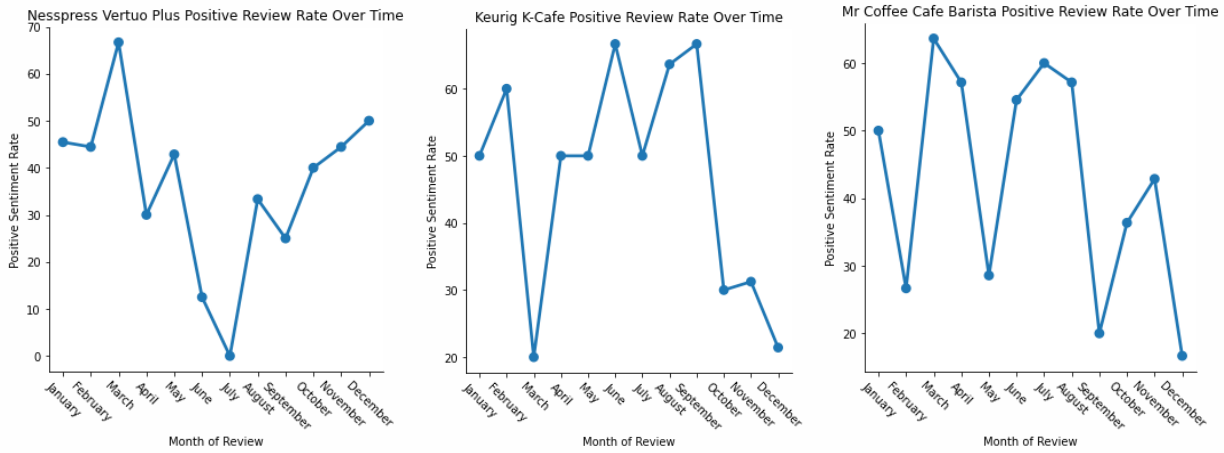


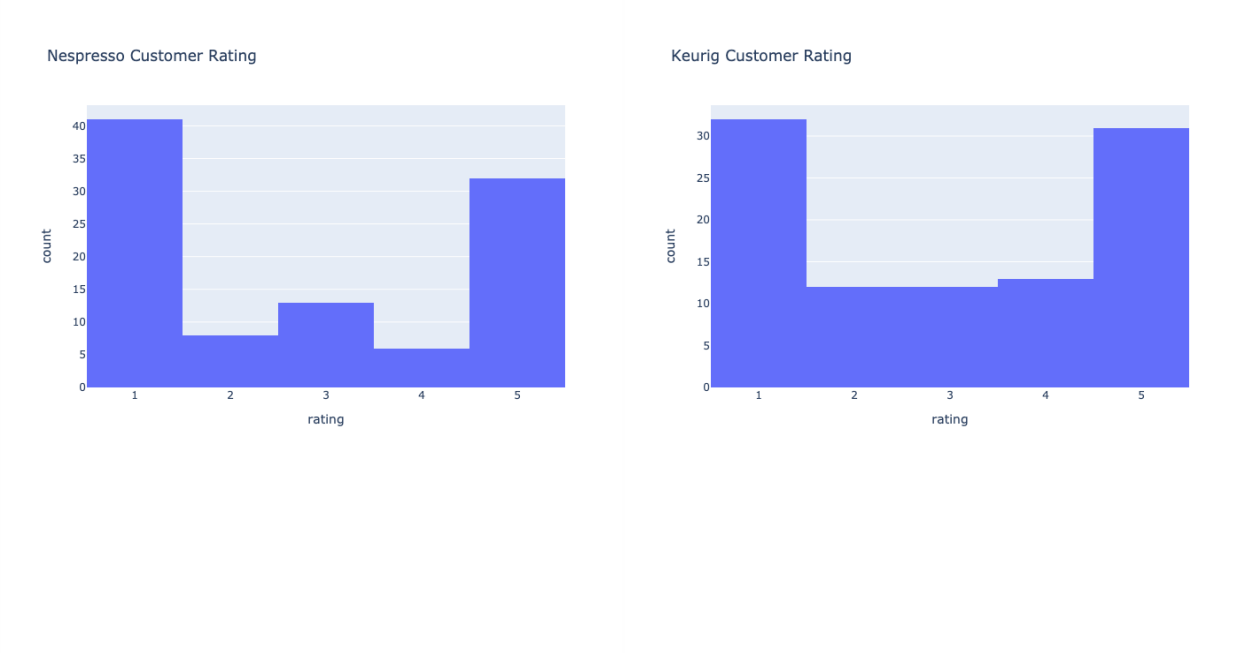

Next, we took a look at how the positive review rate changed over time. This was done by calculating the percentage of reviews that were of 4 or 5 stars. Again, there is no consistency across the

three products. For Nespresso and Mr. Coffee, positive reviews peaked in March, whereas for Keurig, March is when they had their lowest positive rating count. It is also noteworthy that Nespresso has a 0 positive sentiment rate in July even though they do have reviews for that month. There is not an identifiable trend between customers' sentiment and time. The positive sentiment rate fluctuates throughout the year without an obvious pattern.



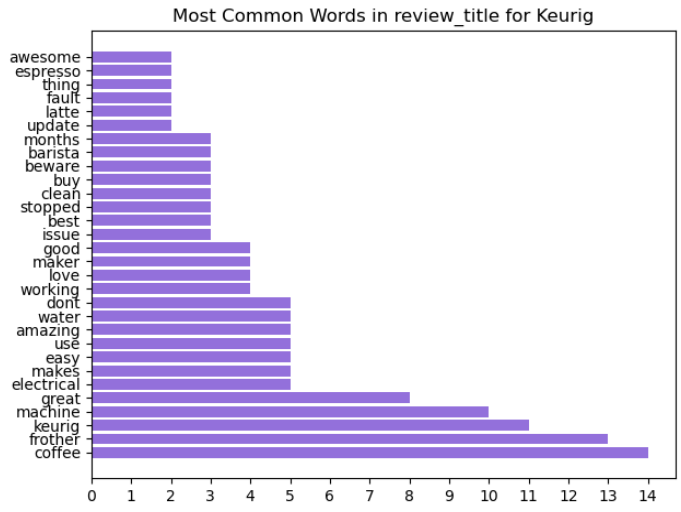

**Research Question 3**

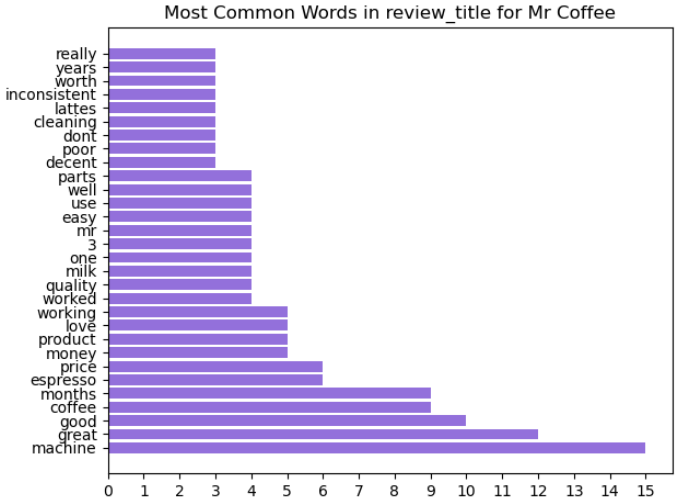
For each product, what do most customers think about their purchase? Which coffee maker stands out as the most efficient or user-friendly? (For the customer reviews of each coffee maker, what are their most appeared positive and negative keywords?)




To answer this research question, we first made a customer rating distribution to see how most customers rate the coffee maker product they bought. By looking at the graphs, we can see that the counts for Nespresso range from about 5 to 40. Comparing the counts of different ratings, most customers rated the Nespresso coffee maker with only 1 star, followed by 5 stars. For Keurig, the counts range from about 10 to 30 and most customers rated the Keurig coffee maker with only 1 star or 5 stars. Similarly, the counts for Mr. Coffee range from about 10 to 30, and the majority of the customers rated the Mr. Coffee coffee maker with 5 stars, followed by 1 star, 2 stars.

From these three histograms, we can see that they can all be described as a U-shaped distribution. This might tell us that people who chose to submit a rating either love or hate the coffee maker product they purchased. This could also be explained by the Beginning-Ending list bias, which addresses that people tend to choose items at the beginning or end of a list given options.

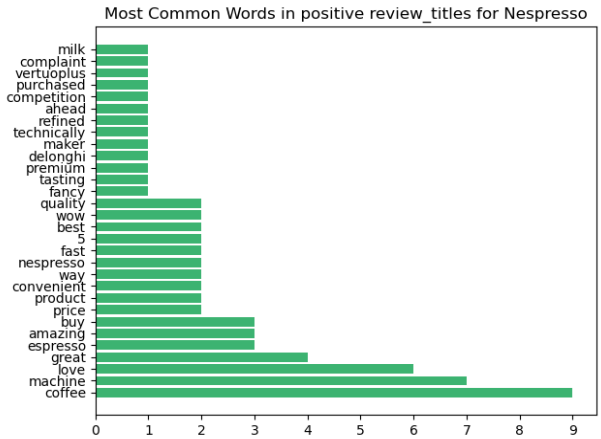



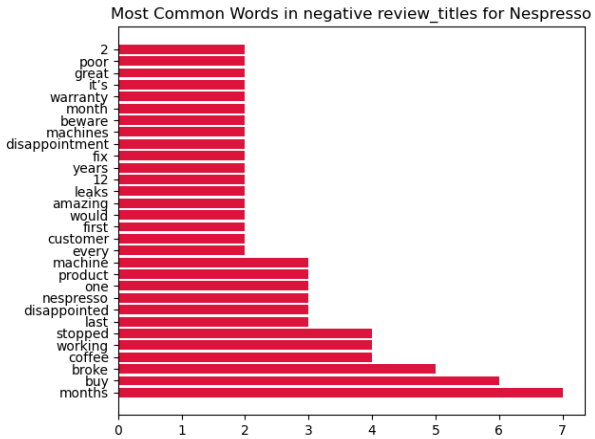



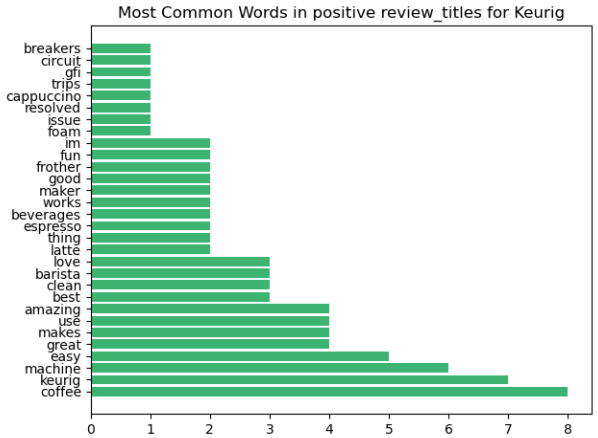

Next, we charted the most frequently appeared words in ‘review_title’ for each product after we counted the number of times each word appeared filtered to the first 30 most common words. Looking at the bar charts above, some of the most meaningful popular words can be summarized as the following table:

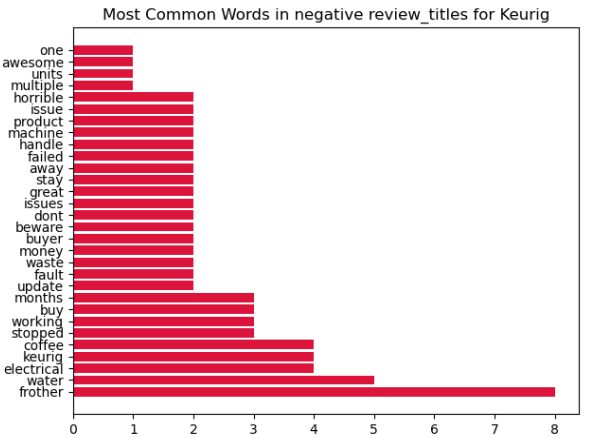
| Coffee Maker Coffee Brand | Meaningful Popular Words |
|---|---|
| Nespresso | “great”, “love”, “broke”, “amazing”, “good”, “stopped”, “disappointed”, “leaks”, “best”, “convenient”, “fast”, “leaking” |
| Keurig | “great”, “easy”, “amazing”, “love”, “good”, “best”, “stopped”, “beware”, “awesome” |
| Mr. Coffee | “great”, “good”, “love”, “easy”, “well”, “poor”, “inconsistent” |

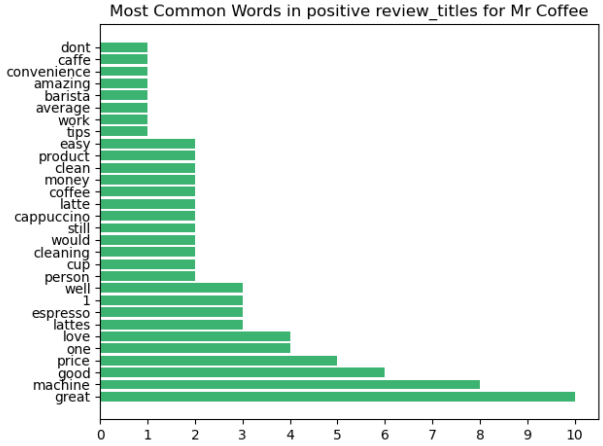
We can see that the most frequently appeared meaningful words in all three products are associated with both positive and negative sentiments. This means that most reviews in the datasets express both positive and negative feelings, which also indicates that people who chose to submit a rating either love or hate their coffee maker.

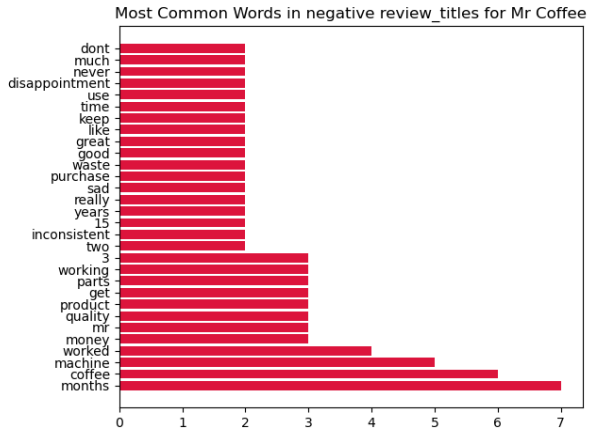


To investigate more into how customers describe their purchased coffee maker products, we also separated positive and negative reviews by classifying them based on how many stars a customer rated the product. Then, we did a similar analysis and created a bar chart for each sentiment and brand. The tables below show the most meaningful positive and negative popular words that appeared in the bar charts:

| Coffee Maker Coffee Brand | Meaningful Popular Words (Positive) |
|---|---|
| Nespresso | “love”, “great”, “amazing”, “price”, **“convenient”, “fast”**, “quality” |
| Keurig | **“easy”**, “great”, “amazing”, “best”, “love”, “good” |
| Mr. Coffee | “great”, “good”, “price”, “love”, “well”, “cleaning”, “money”, **“clean”**, **“easy”** |

| Coffee Maker Coffee Brand | Meaningful Popular Words (Negative) |
|---|---|
| Nespresso | “broke”, “stopped”, “working”, “disappointed”, “leaks”, “disappointment”, “beware”, “poor” |
| Keurig | “stopped”, “working”, “waste”, “money”, “beware”, “issues”, “stay”, “away”, “failed”, “horrible” |
| Mr. Coffee | “money”, “quality”, “inconsistent”, “sad”, “waste”, “disappointment” |

These tables were created manually by looking at the bar charts we plotted previously since it was not possible to make this in Python. This is because even though we were able to filter out stop words like articles or prepositions, other words like “coffee”, “thing”, and “months” were still present in our bar charts, which do not provide any meaningful information.

Based on the summary of the results, we noticed that Nespresso contains keywords like “convenient” and “fast”, which are relevant to “efficient”. Keurig and Mr. Coffee both contain keywords like “easy” and “clean”, which are relevant to “user-friendly”. Therefore, we can argue that the Nespresso coffee maker stands out as the most efficient whereas Keurig and Mr. Coffee coffee makers stand out as more user-friendly.

## Impact and Limitations

The main implication of our research conclusion is to help users recognize which coffee maker best suits their needs, hence would be the ideal purchase. If their primary source of consideration is

existing users' experience with the product, or whether the product is efficient and user-friendly, our findings would be sufficient to help them grasp a holistic perspective.

Another potential implication of our result is gift shopping. When someone wants to give a coffee maker but does not have extensive knowledge about the product, they could refer to our customer-centralized analysis to narrow down their choices and potentially make a final decision.

The group of people who will benefit most from our findings are coffee drinkers, especially those who are shifting to making their own coffee at home as they are most likely to be purchasing a coffee maker. We are also helping the three brands we chose to investigate because we are indirectly asking our audience to choose from this selection.

An excluded population of our analysis is non-coffee drinkers. Since they do not consume coffee, they would not find our results useful. Other coffee maker brands aside from Nespresso, Keurig, and Mr. Coffee may be harmed by our analysis. If people solely use our analysis to determine which coffee maker to purchase, we are inherently harming the brands we did not choose to investigate because they are not in the scope of consideration of the customers.

As aforementioned in our analysis results section, we noticed that a lot of the reviews displayed beginning-ending list bias, which is when customers tend to choose the most extreme options - either a 1 star or 5-star rating. This impacts how truthful the reviews may be because there seem to be minimal "in-betweens".

Below are some limitations of our analysis and how others should or should not use our conclusions:

1. We only web-scraped the first ten pages of reviews for each product, which comes down to 100 reviews each. With only a small dataset out of thousands of reviews, we may not have holistically grasped the entire picture. Therefore, our conclusions should not be used to make decisions that rely solely on what customers have written in their reviews.
2. We only used reviews from purchases in the United States. Therefore, audiences of other countries should not base their purchase decisions on our conclusions because customers of their countries may not have had the same experiences with the product.
3. We did not extract all the data available about reviews. For example, we did not include the purpose of reviewing as part of our analysis. The purpose of leaving a review includes an exchange of free products. This means that customers may be intentionally leaving positive feedback because they want samples, rather than being genuine about their experiences. Therefore, our conclusions should not be used to determine how truthful customer reviews are.

## Challenge Goals

1. **Messy Data**: We planned to write code to scrape data from Amazon customer reviews websites, which means that a lot of pre-processing is needed for the data to be usable.

2. **New Library**: With our goal of extracting information from websites for coffee maker product customer reviews, we planned to learn new Python libraries 'BeautifulSoup' and 'requests' to help with our web scraping process. As we started on our coding portion of the project, we also learned other new libraries that we found useful to our analysis:
    - Plotly: for creating interactive visualizations
    - Kaleido: for exporting static images for plotly figures
    - Nltk: for stop words

## Work Plan Evaluation

1. Find product review URLs for coffee makers on Amazon (Estimated Work Time: 10 min, Actual: 10 min)
    - Finding coffee maker products on Amazon was quick so our proposed work time estimate was accurate.
2. Inspect the websites, use Beautiful Soup to scrape data by extracting customer reviews, perform data cleaning, and transform data into a dataframe or other formats that are easy to work with (Estimated Work Time: 15 hrs, Actual: 17 hrs)
    - Beautiful Soup and web scraping were definitely fun to learn, but it was also notoriously difficult to master. Taking this into account, we dedicated a lot of time to learning and using web scraping tools.
    - As we started scraping the customer review data, we also ran into a lot of different roadblocks so writing `write_csv.py` took longer than expected, costing a few more hours.
3. Work through each research question by performing data manipulation and creating visualizations (Estimated Work Time: 10 hrs, Actual: 23 hrs)
    - Our estimated work time for this task was far from reality because we did not allocate time to test our code so the coding portion for our analysis took significantly longer than expected.
    - We also changed some of our methods as we worked through the analysis so it also ended up taking more time to learn other new libraries that were not mentioned in our project proposal.
4. Present the insights gained from analysis and answer the research questions in a report (Estimated Work Time: 3 hrs, Actual: 5 hrs)
    - We did not account for potential bias that may exist in our data analysis so our work time took longer in reality.

We expected task 2 (web scraping) to be the toughest and take the longest to complete. To best support each other, we communicated often and shared the resources we have found. If there were any issues, we also addressed them promptly. We planned to work together on task 1 and learn web scraping in the first couple of weeks, then divide the rest of the tasks once we successfully scrape the data we need. Our primary development environments will be in Jupyter notebook, Ed workspaces, and local developments like Sublime and command line.

Initially, we wanted to study the reviews of three smart speakers sold on BestBuy.com. We have decided to move away from this idea because of the trouble we faced while testing web scraping code. The reviews for BestBuy products are nested inside Javascript, which requires us to use Selenium instead of Beautiful Soup to conduct the scraping. Due to this, we are now looking at products sold on Amazon.com, which can be successfully scrapped using the Beautiful Soup library.

## Testing

Doing any reliable testing for our code was quite challenging for us. First of all, our datasets were collected from web scraping customer review sections of three coffee maker products on Amazon. The `write_csv.py` file writes the datasets specifically for the websites that we chose and does not necessarily work for any other customer review websites.

But to check if our generated dataset is correct and usable, we randomly picked a customer review on the website and tested if it is one of the rows in our dataset. We also made sure that we included reviews with images or videos because they have different HTML structures. Since no errors were raised, we could argue that our web scraper worked and our datasets are reliable.

Secondly, we don't have any other plots that can be compared to our results so there is no reference for what a plot should look like. In this case, we tested other functions that do not create plots with small example dataframes and `assert_equals`. For example, we tested if our code correctly added a new 'sentiment' column to the data frames by using the `assert_equals` function and a small example dataframe. We also tested if the month could be correctly extracted from the review dates and added into a new column. For functions that create plots, we checked our data to make sure that the results made sense and found no apparent issues.

## Collaboration

Some questions were answered by course staff about testing methods and some aspects of matplotlib. Other resources we consulted include Beautiful Soup and Matplotlib documentation, web scraping tutorials, and Stack Overflow.