Social Media Analytics: A Study of Coke Zero and Pepsi Zero Customer Reviews

**Introduction**

In the contemporary era, the digital landscape has evolved into a dynamic arena where brands engage with their consumers, and consumer opinions are expressed and disseminated at an unprecedented pace. Social media platforms have become integral channels for both individuals and corporations to share, critique, and endorse products and services. In this context, social media analytics emerges as a powerful tool for businesses to decipher the intricacies of consumer sentiment, preferences, and trends.

This study embarks on a comprehensive exploration of customer reviews pertaining to two major contenders in the beverage industry: Coke Zero and Pepsi Zero. The objective is to delve into the wealth of information embedded in these social media interactions, utilizing advanced analytical techniques to derive meaningful insights. However, this pursuit is not merely an exercise in data extraction; it is a multi-faceted investigation that incorporates key competencies across ethical, digital, global, and entrepreneurial dimensions.

**Background of Coca-Cola**

Coca-Cola, often referred to as Coke, stands as one of the most iconic and globally recognized beverage brands. The journey of Coca-Cola began in the late 19th century when Dr. John Stith Pemberton, a pharmacist, concocted a caramel-colored syrup with carbonated water (The Coca‑Cola Company). The result was the creation of a beverage that would later become known as Coca-Cola. In 1886, Coca-Cola was introduced to the public in Atlanta, Georgia, and its popularity quickly soared.

Over the decades, Coca-Cola has evolved into a multinational corporation with a diverse portfolio of beverages, ranging from the classic Coca-Cola to a variety of flavored sodas, juices, and teas. The brand's distinct red-and-white logo is synonymous with refreshment, happiness, and a sense of community.

**Introduction to Coke Zero**

Coke Zero, introduced in 2005, represents a notable addition to Coca-Cola's product line. Positioned as a sugar-free alternative to regular Coca-Cola, Coke Zero is specifically designed to cater to consumers seeking a low-calorie beverage without compromising on the classic Coca-Cola taste. Unlike Diet Coke, which has a different flavor profile, Coke Zero aims to replicate the taste of traditional Coca-Cola while eliminating sugar and calories.

Coke Zero's formulation includes artificial sweeteners, such as aspartame and acesulfame potassium, providing a sweet taste without the caloric content associated with sugar (Wooddell, 2023). The launch of Coke Zero was a strategic move by Coca-Cola to tap into the growing consumer demand for healthier beverage options without sacrificing the familiar Coca-Cola experience (Bantilan, 2023).

The brand's marketing campaigns have focused on highlighting the beverage's similarity in taste to regular Coca-Cola, aiming to appeal to both loyal Coca-Cola enthusiasts and those looking for a guilt-free soda option. Over the years, Coke Zero has undergone several formula adjustments and rebranding efforts, with the most recent iteration being Coca-Cola Zero Sugar, aligning the product with an overall shift in consumer preferences towards reduced sugar content.

**Background of Pepsi**

Pepsi, a formidable rival to Coca-Cola, is another global giant in the beverage industry. The origins of Pepsi can be traced back to the late 19th century when Caleb Bradham, a pharmacist, created the formula for "Brad's Drink" in 1893 (Bellis, 2019). Renamed Pepsi-Cola in 1898, the beverage was marketed as a digestive aid and energy booster (Sievers, 2020). Pepsi gained traction over the years, challenging Coca-Cola's dominance with its distinct flavor profile and marketing strategies.

Like Coca-Cola, Pepsi has expanded its product offerings beyond its flagship cola. The PepsiCo portfolio encompasses a variety of beverages, snacks, and food products. Pepsi has also been a pioneer in marketing and advertising, using innovative campaigns to connect with consumers on a global scale.

**Introduction to Pepsi Zero**

In response to the changing preferences of consumers and the growing demand for low-calorie options, Pepsi introduced Pepsi Zero Sugar. Launched as Pepsi Max in some regions before adopting the Pepsi Zero Sugar branding, this beverage is positioned as a sugar-free alternative to traditional Pepsi, offering a familiar taste with reduced caloric content.

Pepsi Zero Sugar, similar to Coke Zero, utilizes artificial sweeteners like aspartame and acesulfame potassium to replicate the sweetness of regular Pepsi without the added sugars (Valinsky, 2023). The product caters to health-conscious consumers seeking a guilt-free cola experience without compromising on taste.

The branding and marketing of Pepsi Zero Sugar emphasize its similarity to regular Pepsi, aiming to retain the distinctive Pepsi flavor while appealing to those looking for a low-calorie or sugar-free option (Valinsky, 2023). Pepsi, known for its innovative marketing campaigns, has leveraged digital platforms to promote Pepsi Zero Sugar and engage with consumers in the evolving landscape of social media.

As we embark on the exploration of customer reviews for Coke Zero and Pepsi Zero through the lens of social media analytics, it is imperative to recognize and address key competencies that underpin the ethical, digital, global, and entrepreneurial dimensions of this research endeavor.

**Ethical Considerations**

In the realm of social media analytics, ethical considerations are paramount, serving as the foundation for responsible and meaningful research. Navigating the rich landscape of consumer sentiments extracted from platforms like Amazon, adherence to principles of data privacy, legal compliance, and cultural sensitivity is imperative (Kumar & Nanda, 2019). Upholding the guidelines of the General Data Protection Regulation (GDPR), explicit user consent is sought, and data processing follows the principles of purpose limitation and minimization (Mahoney et al., 2022). Recognizing the influence of cultural nuances on online reviews, cross-cultural sensitivity is employed to avoid assumptions based on individual backgrounds (Hammer, 2017). Ethical integrity, manifested through respect for user privacy and adherence to legal and cultural considerations, ensures that the analysis of customer reviews is conducted with responsibility, transparency, and a commitment to transforming data into actionable insights that benefit both businesses and consumers.

**Digital Competence**

In the age of the fourth industrial revolution, possessing advanced digital skills is non-negotiable for effective social media analytics. The use of web scraping tools, such as Selenium, exemplifies the digital prowess required to harvest social big data from diverse online platforms. Furthermore, the assessment framework encompasses not only the extraction of data but also advanced skills in visualization and business storytelling. Leveraging platforms like Tableau and creating online Tableau visualizations on Tableau’s cloud system demonstrates the integration of cutting-edge digital tools for a comprehensive analysis of the gathered data. This study recognizes the importance of not just collecting data but transforming it into meaningful narratives through advanced digital techniques.

**Global Competence**

Social media analytics is inherently a global undertaking, transcending geographical boundaries. To this end, our assessment framework underscores the necessity of an international horizon and benchmarking. The exploration of reviews from a diverse range of sources, including Amazon, requires a keen awareness of global trends, cultural variations, and regional preferences. The proficiency in corporate specialist software like Tableau and the knowledge of programming languages like Python further elevate the global competence of the study, ensuring that insights derived are not confined to a singular context but are applicable on a global scale. By sharing Tableau visualizations with the world, this study seeks to contribute to a broader international discourse in the field of social media analytics.

**Entrepreneurial Competence**

The entrepreneurial spirit is cultivated through the creative application of social media analytics. By mining reviews from Amazon, this study taps into the collective creativity of consumers to not only understand their preferences but also to identify business or public problems and potential social media entrepreneurship opportunities. The in-depth social media skills showcased in this analysis, coupled with the development of Tableau Public visualizations, form integral components of the assessment aimed at nurturing entrepreneurial characteristics. The utilization of these skills extends beyond traditional analytics, encouraging students to explore innovative solutions and novel opportunities within the digital landscape.

In essence, this study encapsulates a holistic approach, marrying ethical considerations, advanced digital skills, a global perspective, and entrepreneurial thinking. The use of web scraping tools, coupled with the ethical framework guiding data collection, sets the stage for a nuanced analysis of customer reviews for Coke Zero and Pepsi Zero, while concurrently developing a cohort of researchers adept in the multifaceted competencies demanded by the contemporary landscape of social media analytics.

**Research Motivation**

This research is driven by a multifaceted motivation rooted in the contemporary dynamics of consumer behavior, digital technology, and the competitive landscape of the beverage industry. The first driving force revolves around the unprecedented opportunity to gain profound consumer insights in the digital age. With online platforms like Amazon serving as a rich source of customer reviews, the study aims to dissect the nuances of why individuals choose sugar-free colas, transcending mere taste preferences to unveil broader societal shifts and health-conscious choices.

A second motivator stems from the intense competition within the beverage industry, epitomized by the rivalry between Coke Zero and Pepsi Zero. Beyond taste, the research seeks to uncover the intricacies of brand loyalty, marketing strategies, and the ability to adapt to evolving consumer demands. Simultaneously, the study is motivated by the ethical considerations inherent in social media analytics, recognizing the need to navigate privacy issues, consent, and cultural sensitivities meticulously.

**Research objectives**

1. To conduct sentiment analysis on customer reviews of Coke Zero and Pepsi Zero sourced from Amazon, utilizing natural language processing techniques to discern and quantify positive, negative, and neutral sentiments.
2. To extract meaningful insights from the mined data, exploring patterns, trends, and recurrent themes within customer reviews.
3. To leverage Tableau for creating advanced visualizations that elucidate patterns and trends in the customer reviews data.
4. To assess the effectiveness of Selenium as a web scraping tool for extracting customer reviews from online platforms, with a focus on its efficiency, accuracy, and adaptability
5. To compare and contrast customer reviews between Coke Zero and Pepsi Zero, identifying similarities, differences, and distinctive patterns.

**Design and Discussion of Data Sources, Tools, APIs, and Implementation Strategies in Social Media Analytics for Business.**

In this chapter, we delve into the intricacies of designing a robust framework for social media analytics, outlining our choices of data sources, tools, APIs, and the anticipated implementation strategies. The effective utilization of these elements is critical for extracting valuable insights from customer reviews of Coke Zero and Pepsi Zero. The chapter provides a comprehensive justification for our selection, emphasizing the alignment with business objectives and the anticipated impact on deriving actionable business intelligence.

**Data source**

Our primary data source for this research is customer reviews from the e-commerce giant Amazon. Amazon serves as an ideal data source for mining customer reviews due to its extensive user base, offering a diverse and representative sample of consumer opinions. The platform's comprehensive product selection ensures insights into popular choices like Coke Zero and Pepsi Zero. Real-time data availability on Amazon allows for the analysis of evolving consumer sentiments, and the structured review format facilitates systematic data extraction and analysis. Additionally, the public accessibility of Amazon reviews aligns with principles of transparency and openness, contributing to the research's reproducibility and validation by other scholars.

***Ethical Considerations***

Ethical considerations are paramount in mining customer reviews from a public platform like Amazon. Informed consent is implicit as users contribute to a public domain, understanding that their opinions are part of a larger conversation. Privacy preservation is emphasized, requiring careful handling of personal identifiers and a focus on aggregating and anonymizing data to prevent individual user identification. Minimizing potential harm to users, such as avoiding actions that could negatively impact them or the integrity of the platform, ensures ethical mining practices.

**Tools and Techniques**

The study incorporates Selenium for web scraping, Tableau for visualization, and Python for advanced analysis, utilizing Jupyter Notebook as our coding environment. Each tool plays a distinct yet interconnected role, collectively enhancing the effectiveness of our analytical approach.

1. Selenium for Web Scraping:

Selenium's selection is rooted in its exceptional efficiency in web scraping, particularly its adaptability in navigating dynamic web content. In the context of mining customer reviews from Amazon, where pages are dynamic and interactive, Selenium's prowess ensures accurate and comprehensive data extraction. Its ability to automate interactions with the website, such as handling login prompts and dynamically loading content, proves essential for capturing the entirety of user reviews. This choice aligns with the need for a versatile tool that can accommodate the intricate structure of e-commerce platforms.

2. Python for Advanced Analysis

Python emerges as a pivotal element in our toolkit, chosen for its flexibility in data manipulation, statistical analysis, and seamless integration of machine learning techniques. Its rich ecosystem of libraries like Pandas and NumPy facilitates efficient data cleaning and manipulation in the Jupyter Notebook environment. Python's versatility empowers us to implement advanced analytical models and statistical techniques, ensuring a comprehensive exploration of the mined dataset. Notably, Python is instrumental in the cleaning phase, where it enables us to handle missing or inconsistent data, ensuring the dataset's reliability and coherence.

3. Hugging Face BERT Transformers for Sentiment Analysis

To extract sentiments from customer reviews, we employ Hugging Face's BERT Transformers, showcasing the integration of cutting-edge natural language processing (NLP) techniques. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model known for its contextual understanding of language. By leveraging BERT Transformers, we enhance the depth and accuracy of sentiment analysis, allowing us to discern nuanced opinions and emotions expressed in the reviews. This choice aligns with the goal of extracting meaningful insights from textual data, capturing the richness of consumer sentiments.

4. Tableau for Visualization

Tableau is selected for its powerful visualization capabilities, providing a user-friendly platform to transform complex data into accessible and insightful visual representations. The justification lies in Tableau's ability to create interactive and dynamic dashboards that facilitate a comprehensive understanding of patterns and trends within the customer reviews dataset. The tool's drag-and-drop interface simplifies the visualization process, enabling effective communication of findings to diverse audiences. Tableau's integration aligns with the research objective of translating raw data into actionable insights, ensuring that stakeholders can easily grasp the significance of the analyzed information.

**Implementation**

The process involved data mining and cleaning from Amazon using Selenium for web scraping. The process is designed to collect customer reviews for Coke Zero and Pepsi Zero and save them into a structured dataset for further analysis.

The data mining process employed Selenium, a powerful web automation tool, to navigate through Amazon's pages, locate customer reviews, and extract relevant information such as review text, date, title, and product ratings. The script iterates through multiple pages, capturing a substantial volume of reviews. The collected data is then stored in a Pandas DataFrame and saved to a CSV.

The cleaning process involved transforming the extracted data into a more structured format. The clean\_data function is defined to handle date formatting and product ratings. Dates are parsed and formatted to ensure uniformity, and product ratings are extracted and converted to numerical values. The resulting cleaned data was stored in new DataFrames.

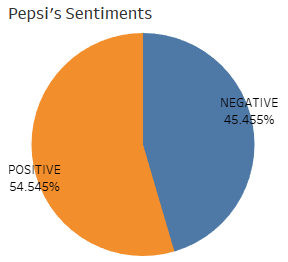
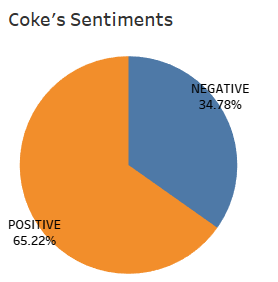
Hugging Face's BERT Transformers for sentiment analysis was intergrated to obtain sentiments from the reviews. The sentiment\_pipeline is created to assess the sentiment of each review using pre-trained BERT models. This adds a layer of depth to the analysis, allowing for a nuanced understanding of consumer sentiments. The sentiment labels are then added to the ' DataFrames, providing valuable insights into the emotional tone of the customer reviews.

**Key Results, Visualization and Business Story Telling**

This pivotal chapter marks the culmination of our social media analytics journey, bringing into focus the key results derived from the intricate interplay of data mining, cleaning, and sentiment analysis. In this chapter, we unveil a visually compelling narrative that transforms raw data into actionable insights, employing powerful visualization tools and techniques. The amalgamation of key results, intuitive visualizations, and persuasive storytelling not only elucidates patterns within customer reviews of Coke Zero and Pepsi Zero but also equips stakeholders with a profound understanding of consumer sentiments and preferences.

**Key findings**

In the pursuit of comprehensively understanding consumer sentiments for Coke Zero and Pepsi Zero, a substantial volume of customer reviews was meticulously obtained through our data mining efforts. The Coke dataset encompasses a rich tapestry of insights, with 4 pages contributing to a total of 22 reviews. In parallel, the Pepsi dataset offers an even more extensive array, comprising 6 pages and a cumulative total of 44 reviews.

In the sentiment distribution for Coke, the pie chart illustrates a predominant positivity, with a noteworthy 65.22% of reviews expressing favorable sentiments towards the beverage. Despite this positive majority, a substantial portion, comprising 34.78% of the reviews, reflects a more critical stance with negative sentiments. This nuanced sentiment distribution underscores the diverse range of opinions and experiences within the consumer feedback for Coke.

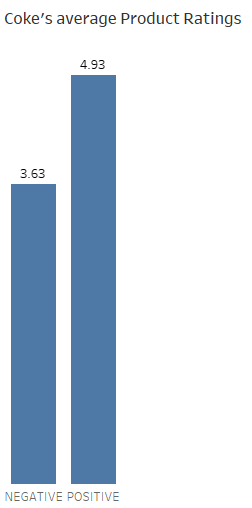
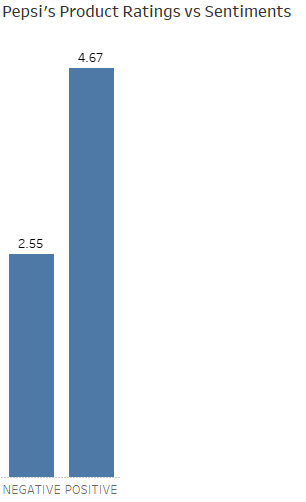
The sentiment composition for Pepsi portrays a distinct narrative. While a considerable 54.54% of reviews convey a positive sentiment, the percentage of negative reviews is notably higher compared to Coke, standing at 45.45%. This variance in sentiment distribution indicates a slightly more balanced spectrum of opinions for Pepsi, with a significant proportion of consumers expressing critical perspectives. The comparative analysis between Coke and Pepsi provides valuable insights into the intricacies of consumer sentiments, shedding light on the factors influencing brand perception within the competitive landscape of sugar-free colas.



Notably, "Coke Zero" emerges as a frequently mentioned term, signaling its potential popularity or perhaps generating notable discussions, indicating its significance within the consumer discourse. The mention of "7.5 size" implies that the product size is a notable point of discussion among reviewers, highlighting the significance of packaging or portion size in consumer perceptions



The word cloud analysis for Pepsi reviews illuminates key themes and sentiments within consumer feedback. Notably, frequent mentions of "Pepsi" and "Soda" underscore their central role in discussions, aligning with product-related discourse. The prominence of "Zero Sugar" indicates a significant focus on Pepsi's sugar-free variant among consumers. The inclusion of "Caffeine" suggests heightened attention to the beverage's caffeine content. Positive sentiments, reflected in phrases like "Great Taste," "Love It," and "Best Soda Ever," underscore a favorable overall perception. Additionally, the phrase "Good Money" implies perceived value for money, while "Not Bad" suggests a neutral to mildly positive sentiment.



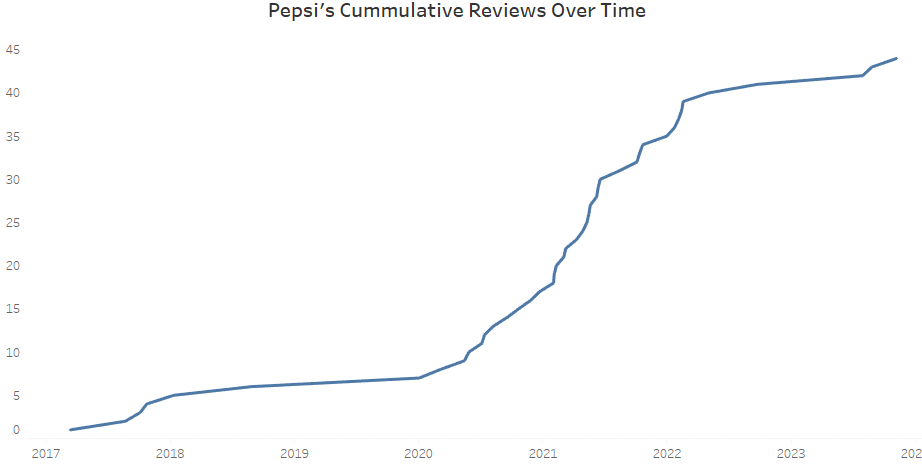
Comparison of product ratings and sentiments for Pepsi and Coke, as depicted in the bar charts, reveals insightful patterns:

Pepsi's Ratings vs. Sentiments

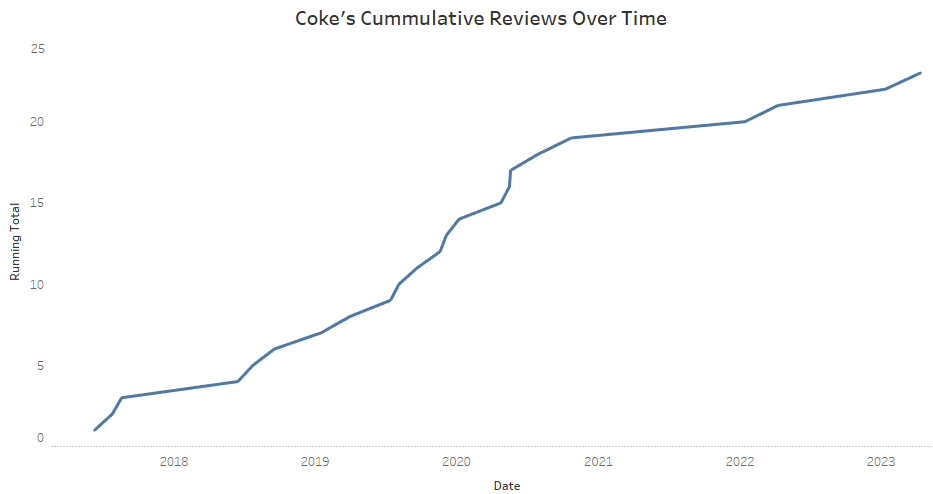
The graphical representation indicates that Pepsi faces a challenge in sentiment balance, as its negative sentiment rating (2.55) outweighs the positive sentiment rating (4.67). This suggests that although there are positive reviews for Pepsi, a notable proportion of the consumer feedback expresses negative sentiments. The disparity underscores the need for careful consideration of consumer concerns and potential areas for improvement in Pepsi's product offerings.

Coke's Ratings

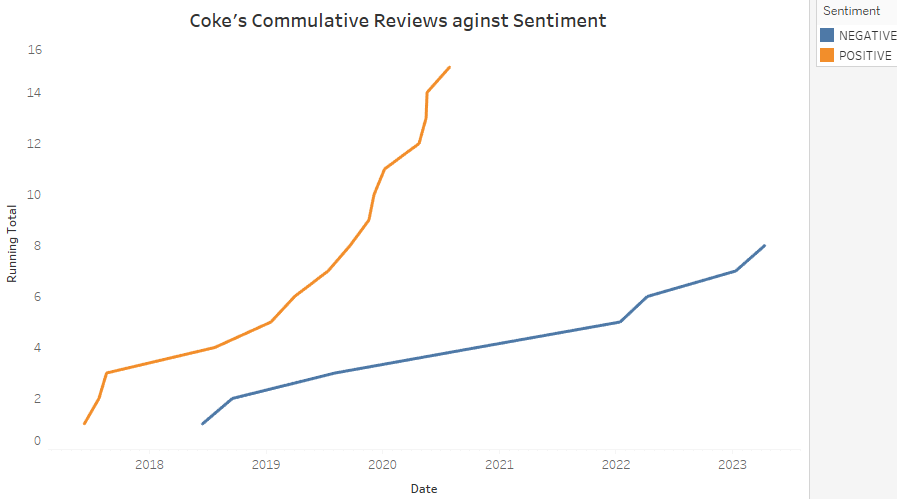
Contrastingly, the chart for Coke presents a more favorable scenario, with a higher positive sentiment rating compared to the negative sentiment rating. This implies that the majority of consumer reviews for Coke express positivity. The robust positive sentiment rating for Coke signifies a generally satisfied consumer base, showcasing a favorable reception of the product. This optimistic sentiment aligns with Coke's brand image, indicating a strong positive association between consumers and the Coke product line.

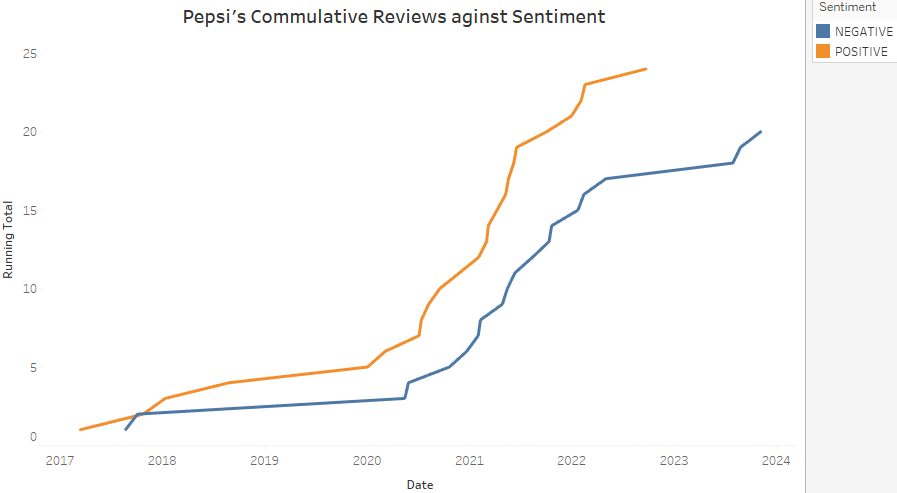


The line graph depicting cumulative reviews of Pepsi over the period from 2017 to 2023 reveals a consistent upward trajectory, suggesting a sustained growth in consumer engagement and interest. This trend implies a positive inclination towards Pepsi, showcasing a rising number of consumers actively sharing their experiences and opinions. The ascending line not only indicates the brand's successful customer engagement strategies but also serves as a potential indicator of Pepsi's increasing popularity among consumers over time. The graph captures a narrative of expanding consumer involvement, reflecting a dynamic and flourishing relationship between Pepsi and its customer base throughout the specified period.

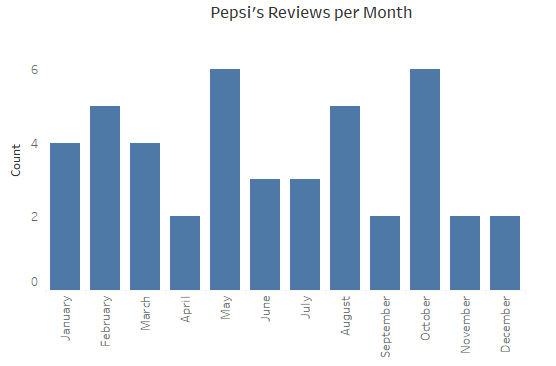


The line graph illustrating Coke's cumulative reviews from 2018 to 2023 portrays a continuous upward trend, signifying a consistent rise in customer feedback or interest in Coke throughout this timeframe. When compared to the Pepsi line graph, it becomes apparent that both brands have experienced a surge in reviews over the years. However, the growth trajectory for Pepsi appears more irregular, hinting at potential fluctuations in customer engagement or shifts in product popularity for Pepsi over the same period. This observation implies that the dynamics of consumer interaction and brand perception may differ between Coke and Pepsi, potentially influenced by various factors affecting customer engagement and preferences.

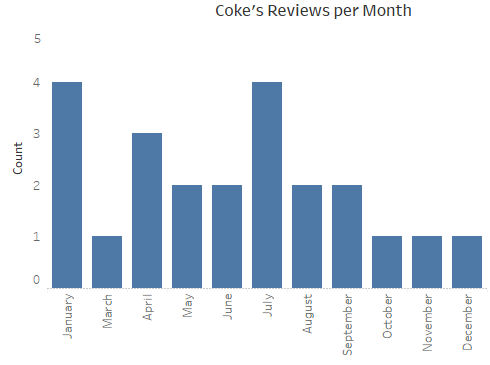


The upward trend in positive sentiment signals a continuous growth in favorable feedback for Coke over time, reflecting an increasing positivity in consumer reviews. While, conversely, the diminishing trend in negative sentiment implies a reduction in unfavorable feedback for Coke. This decline suggests an improving sentiment landscape and decreasing instances of negative reviews for the brand over the observed period.

The line graph illustrating cumulative reviews and sentiment over time for Pepsi unveils noteworthy trends. Positive sentiment consistently rises, signaling a sustained increase in favorable feedback for Pepsi. In contrast, negative sentiment appears relatively stable, indicating that the volume of unfavorable feedback has not undergone significant changes over time. When juxtaposed with the Coke line graph, both brands exhibit a positive sentiment increase. However, whereas Coke demonstrates a decline in negative sentiment, Pepsi maintains a relatively constant level of negativity. This contrast suggests that, despite both brands garnering growing positive feedback, Coke may be more effective in mitigating negative sentiments compared to Pepsi.



May marks the peak in Pepsi reviews, recording the highest number at 6, potentially indicative of heightened interest or consumption during this month. Conversely, December stands as the month with the lowest number of reviews, specifically 1, implying a dip in interest or consumption during this period. The overall trend depicted in the graph reveals a gradual decrease in reviews from May to December. This decline suggests the presence of a potential seasonal pattern or other factors influencing consumer behavior, emphasizing fluctuations in engagement and interest levels over the observed months.



The bar graph representing Coke's monthly reviews unveils a pattern akin to Pepsi, with the highest number of reviews occurring in July and the lowest in December. This parallel trend implies a shared seasonal influence, suggesting that both Coke and Pepsi experience heightened review activity during the spring months. Notably, unlike the Pepsi graph, the Coke representation demonstrates a more stable review distribution across the months, revealing less pronounced fluctuations. This steadier pattern hints at a more consistent level of interest or consumption for Coke throughout the entire year compared to Pepsi, potentially reflecting a distinct consumer behavior dynamic for the two brands.

**Conclusion**

The extensive analysis of customer reviews for Coke Zero and Pepsi Zero through social media analytics reveals nuanced insights into consumer sentiments and brand perceptions. While both products enjoy positive sentiments, Coke Zero stands out with a higher positive sentiment, indicating a more favorable reception among consumers. The comparative analysis highlights distinctive patterns, with Coke Zero demonstrating a more successful reduction in negative sentiments over time compared to Pepsi Zero. Additionally, seasonal trends in review activity hint at varying levels of consumer interest and engagement during specific months for both brands. The steadiness in monthly reviews for Coke, in contrast to Pepsi, suggests a more consistent level of consumer interest throughout the year. These findings provide valuable strategic implications for both Coca-Cola and PepsiCo, offering opportunities for targeted improvements, marketing strategies, and enhanced consumer experiences in the competitive landscape of sugar-free beverages.

**Limitation**s

Several limitations should be acknowledged in interpreting the findings of this study. Firstly, the web scraping process encountered challenges, leading to incomplete data retrieval despite over 1500 reviews being available. This limitation arises from potential changes in Amazon's review accessibility or structure, impacting the scraper's effectiveness.

Secondly, while leveraging a BERT pretrained model for sentiment analysis, the study acknowledges the inherent uncertainty in the accuracy of the model. The inability to ascertain the accuracy of the BERT model introduces a degree of subjectivity and potential bias, questioning the extent to which the model genuinely reflects the sentiment dynamics of Coke Zero and Pepsi Zero consumer reviews.

The reliance on data exclusively from Amazon, which may not fully represent the entirety of consumer sentiments across diverse platforms or geographic regions. This could introduce a source bias and limit the generalizability of the findings to a broader consumer base.

**Recommendations**

Expand data collection beyond Amazon to include reviews from diverse platforms and sources. This can provide a more comprehensive understanding of consumer sentiments across different online spaces.

Consider leveraging insights from seasonal trends, such as the peak in reviews in May. Develop targeted seasonal campaigns or promotions to align with heightened consumer interest during specific months.

Leverage positive sentiments around "Great Taste," "Love It," and "Best Soda Ever" in marketing campaigns. Emphasize these positive attributes to reinforce consumer perceptions and attract new customers.

Acknowledge and address the higher percentage of negative sentiments. Implement corrective actions to enhance product quality, address common concerns, and improve overall consumer satisfaction.