**Improving Customer Retention Through the Loyalty App Program of The Canadian Brewhouse**

Northern Alberta Institute of Technology

DATA 3960 – Data Capstone Project

Esha Basharat, Tsz Fung Cheung, Michael J Harris, Kurtis Li

April 23, 2024

**Executive Summary:**

This comprehensive report delves into an in-depth analysis of customer engagement and loyalty program effectiveness within The Canadian Brewhouse. Through data analysis and strategic insights, this report aims to provide actionable recommendations to optimize customer satisfaction, loyalty, and business success. With the Check in dataset, we conducted customer segmentation to create groups that would be more susceptible to targeted marketing strategies. Additionally, we performed clustering analysis on the guest dataset and identified high-spending customer segments also primed for targeted marketing efforts. With the referral and redemption datasets we developed insights that allows for tailored strategies to incentivize referrals and encourage redemption among all customers. Overall, the recommendations include developing the "VIP Rewards" program, focusing on digital engagement via Facebook campaigns, customizing the menu to entice craft beer enthusiasts, and enhancing event experiences during NHL game showcases. In conclusion, this report provides a roadmap for The Canadian Brewhouse to optimize its loyalty program, enhance customer engagement, and drive business growth.

**Business Case:**

The Canadian Brewhouse is a successful sports bar chain with 45 locations across Canada, and in recent years has begun looking more towards data to increase their business and customer base. With a thorough analysis of the available data and by uncovering any trends there is the possibility for important insights and data-driven decisions to be made. How is customer retention affected by the loyalty app program and the rewards and opportunities it offers. If there is a greater understanding of the segments of customers that participate in the app and what their behavior are, then this information can be applied to marketing campaigns that would have much more focused strategy and would likely lead to greater success. As a food service business, they are relying on finding ways to bring customers into their establishments and getting them to engage in their promotions and activities, but the answer to the best strategy for that may present itself in the data.

**Research Questions:**

The main goal for our project was to use the data provided by the Brewhouse from their loyalty program to find a way to increase customer retention for the Canadian Brewhouse. First this would be done by understanding how those relationships present themselves presently based on the data, and then how that can be leveraged. We looked to accomplish this in a couple ways.

First was to look at how customer segmentation could be used on the users through these datasets to separate them based on their behaviors. Common groupings could then be part of a more focused approach. Next was to understand how ratings and feedback affect customer retention and if there is a trend in how customers are reacting to their experience. Is there an opportunity to improve the overall dining or loyalty app experience. We also aimed to understand how external factors can maximize customer retention, such as the implementation of craft beer sales. Finally, how to increase redemption and referral engagement with the loyalty app users, and how the loyalty reward programs can be adjusted to provide a better experience and attract more customers.

**Dataset Description:**

The Canadian Brewhouse provided us with many different datasets related to their loyalty app run by the program PUNCHH, such as receipt data, coupons, and visits to locations. We chose to focus on five datasets in total: check in, guest, user feedback, referral, and redemption.

Check in contained data on each individual visit by any user to a Brewhouse location. Guest contained the record of every user of the app and their points and number of check ins. Referral contained information on the guest that has been referred to the loyalty app, and the user that referred them. Redemption contained information on each time a user redeemed a reward and what the reward was. And finally, user feedback contained each customer review and their rating of their visit. Together these datasets would provide us with a good understanding of customer behaviors and trends, and we could use various types of analysis to potentially understand what causes customer retention, and what recommendations we could provide to the Canadian Brewhouse. We also used a few outside datasets about online Yelp reviews and craft beer.

**LRFM Analysis – Check in:**

One of the ways we can look to understand the potential ways to maintain or improve customer retention for patrons of the Canadian Brewhouse is by examining the data surrounding the details of their visits. The check-in dataset contains the information on each individual check-in for each customer, which when combined together can provide a larger picture of several important factors about the customer base. In this case we can examine the number of times a customer checked in, how recently they checked in, the time from the first and most recent check in, and most importantly how much money they spent. The type of analysis is known as LRFM Analysis and can be used on a variety of data types but is best suited for the kind of data available in this dataset.

LRFM analysis is a form of customer segmentation and is typically used to predict how a customer may act in the future by examining what they have done in the past. It can also be used by marketers to create specific customer groups that would be susceptible to specific kinds of campaign or promotions. (Delval, 2024) For the purposes of this project we used it for customer segmentation but with the intention that if this model was implemented on active data including more information on the customers themselves there was more potential for substantial results.

**Methodology:**

For each of the aspects of LRFM the first step was to determine the actual value of each measure, then they were assigned a score based on how they compared to each other. For each measure we used a scale of 1 to 5, with 1 indicating the worst values and 5 indicating the best values. In each case the scores were determined by using qcut with q=5, so that they were equally divided based on their own values. This score would be the main way to analyze the different measures and the way that the customer segmentations could be made.

Recency was calculated by comparing each check in to the date of the ‘end’ of the dataset, to determine how recently they had visited. The end date was determined to be the day after the final date available in the dataset, which was September 10, 2022, so every date was compared to September 11, 2022, in terms of recency. Frequency was calculated by determining the total number of check ins for each customer, tracked by User ID. This was done by first grouping by User ID and taking a count of the rows for each user. This number was applied as the index for that grouping, and when merged back into the main dataset each User ID would have the same frequency value for each of their check ins. Monetary was calculated by taking a sum of all the receipt amount data to determine much money a customer had spent across all their visits. Length was calculated by how many days it has been from the first instance of a check in, to the most recent check in.

After creating the LRFM scores we can now assign them labels of customer levels based on their combined value. The scores were summed and then assigned into 4 new categories based on their value. Customers were assigned a label and sorted into the following groups: Poor (4-5), Low (6-8), Medium (9-15), High (16-20). Each label was an indication of their quality as a customer to the Brewhouse since higher scores meant they visited more often and spent more money. An additional category was created called ‘New’ and was assigned to the customers who had only one value in Recency but had a Frequency of less than 32, to indicate those beginning their loyalty profile in the last month of data.

**Findings:**

The results of this analysis provided some interesting results regarding the range of customers and their behaviors. The highest group was Low with 54,747 users (48.7%), next was Medium with 29,831 users (26.5%), then Poor with 20,256 users (18.0%), then High with 5803 users (5.2%), and finally New with 1732 users (1.5%) (Figure 1). In terms of locations, the Windemere location had the highest average of LRFM Score (Sum) with 11.05, with St. Albert South at 10.49, and Sherwood Park at 9.67. The locations with the lowest scores were Camrose with 6.44, Edmonton Downtown with 6.64, and Calgary Mahogany with 7.36 (Figure 4). The overall average score for all users was 8.61, which would be considered Medium. High level customers had 89.7 check ins and 76284.8 loyalty points, Medium customers had 18.1 check ins and 14651.0 points, Low customers had 3.5 check ins, and 2608.1 points, and Poor customers had 2.0 check ins and 1116.7 points.

To further understand the relationship between two of the more important factors, recency and monetary, we also conducted correlational analysis. They were found to have a correlation coefficient of 0.87 which was by far the highest value (Figure 3). Next, we aimed to learn to what degree frequency of visits affected the amount of money spent by performing linear regression analysis. They were found to have a R2 value of 0.786, and after creating a regression model it was clear that with each visit there was an increase of $47.37. Considering that the average frequency value was 8.52 that would equate to approximately $424.625 spent by the average user.

With the results of the LRFM analysis, there seems to be a trend towards the customer loyalty level of Low being the most prominent, users with a LRFM score of 6-8. From looking at the distribution of the counts of each score, there is always a decrease in score counts as the value of that score increases (Figure 2). The probable reason is that the average customer isn’t going to be exceptional in terms of these factors, as is likely the case with most consumer products and services, the upper percentage of users are much fewer in number, especially when it comes to spending money. It is much easier for a customer to be in the lowest score group and therefore it would be more likely. These results indicate that it would take more of a push to move customers into the higher score groups and higher customer levels.

With this created dataset of LRFM scores and segmentations for each customer there are multiple ways that the data can be utilized further, and a few stand out as more important to focus on. There are customers with higher monetary scores but lower frequency scores, which means the customers who spend more money but do not visit as often. The data indicates that if those customers can be bumped up by one frequency score bracket, they could theoretically increase their spending by 33% on average, or $297.48. The other group is those with high frequency but low recency, which means the customers who used to visit often but have not been back in a while. If they were also bumped up by one score bracket, they would increase their number of visits by 35% on average, or 42.8 visits.

Both groups show a strong potential for high-quality visits if they can be brought back on track and continue to engage with the loyalty program.

**Recommendations:**

To increase customer engagement in the loyalty app and improve customer retention, an extra push would need to be given to those customer segments that were described as high value. The main way to accomplish this would be with targeted marketing efforts to take advantage of the traits of the specific segments. The best strategy would likely be differentiated marketing, which involves marketing in subsets of the population (Lewison & Hawes, 2007), as the groups of customers are proven to be diverse. This could be done through the implementation of special rewards or promotional offers or raising awareness of the special events happening at their local location. This is also complemented well with our analysis of the rewards program where the most popular rewards that provided the most customer engagement could be used to attract these customers. However, there is also the possibility for orchestrated marketing, where the marketing efforts are concentrated on a cross section of customers from multiple groups (Lewison & Hawes, 2007), such all those from a certain location but different LRFM groups. This would take advantage of location specific aspects of the customer base that may be useful such as population or geographic location.

By extension of that, since the customers will have a preferred location, there is another strategy that could be used to understand how each location is performing in terms of the customers that are visiting there, and this would be through the use of surveys. For example, the customers who frequent the Windemere location could be surveyed to understand why they tend to visit there and what keeps them coming back. Likewise, the same could be done for the Edmonton Downtown location to understand why they are only attracting lower-level customers, since further analysis of the Camrose location indicates it has since been closed. Geographic location could be a factor here considering Windemere’s location on the edge of the city in a more affluent neighborhood, along with Edmonton Downton being frequented by university students who may not spend as much, but this survey technique would be one of the best ways to understand some of the nuances.

Overall, the main recommendation would be to combine the LRFM analysis we’ve conducted here with the personal information available from the real-time data, as we feel that this could be a good opportunity to increase sales and customer retention by focusing on these groups of customers that show the most potential.

**User Feedback:**

**Mothodology:**

To streamline the dataset, we initially replaced all null values in the feedback column with 'No Comment' to determine how many people did not provide any comment. Next, unnecessary columns such as First & Last Name, Feedback Reply, Date of Reply, and Requires Response were removed to enhance dataset cleanliness. We then extracted the Year section from the Date of Feedback column to enable year-wise rating analysis. Lastly, sentiment analysis was conducted using the feedback section, focusing on ratings and individual customer feedback for each location. We selected this dataset specifically for its rating and feedback sections. By utilizing these two, we can analyze the results of each location based on ratings and examine the feedback from individual customers.

**Findings:**

In Figure 8, Camrose scores the highest at 4.87, while Banff, despite its tourist appeal, scores the lowest at 4.50. This is significant as a poor restaurant experience in a tourist destination like Banff could greatly affect overall trip satisfaction. In Figure 9, Lewis Estates receives the most feedback with over 9000 entries, whereas Banff has one of the lowest counts at 125 entries. This scarcity of feedback in places like Banff makes pinpointing issues challenging. Looking specifically at Banff in Figure 10, we notice a concerning drop in average rating from a perfect 5 in 2022 to 4.5 in 2023, signaling a 10% decrease. This decline should prompt thorough investigation and prompt action from The Canadian Brewhouse. Conversely, Camrose remains stable over the years with consistently high scores. Figure 11 reveals that about 94% of entries have 'No Comment,' highlighting a lack of detailed feedback crucial for improvement. Finally, Figure 12 shows sentiment analysis results, indicating predominantly positive feedback at HQ Simphony Lab but an overall negative sentiment across all entries, suggesting room for improvement in customer satisfaction.

**Recommendations:**

The Canadian Brewhouse should identify areas for improvement and address the issues to enhance customer satisfaction. For example, regarding the average rating by location and year. Banff could learn from Camrose to understand how they maintain a consistently high average score year after year. In addition, to address the “No Comment” issue, the Canadian Brewhouse could offer customers a chance to win a gift card in exchange for leaving feedback. Furthermore, to discourage customers from leaving less valuable feedback such as 'Good' or 'Bad' without explanation, it's important for employees to follow up on feedback. Currently, there is no data in the feedback reply column of this dataset. Following up with customers demonstrates that the restaurant values their input and encourages them to provide more detailed and thoughtful feedback in the future as well as enhancing retention. Lastly, regarding the average sentiment score, recommendations would be to improve customer service, product quality, communication strategies, and other aspects of the business to enhance overall satisfaction and sentiment in the future.

**Yelp Review**

**Methodology:**

To mitigate bias from solely relying on internal data sources from The Canadian Brewhouse, we've incorporated an external data source from Yelp. In the 'Yelp Review' dataset, we gathered reviews spanning from 2019 to the most recent ones for four different locations, namely: 6093 Currents Dr NW, 840 St. Albert Trail Suite 105, 9535 Ellerslie Rd SW, and 10338 109 St NW.

**Finding:**

Figure 13 reveals the average rating. While the lowest, 2.7, is found at St. Albert. This discrepancy indicates that customers may be more critical compared to the ratings from internal data sources, where the lowest is 4.5 at Banff. This underscores the importance of gathering feedback from diverse sources to gain a comprehensive understanding of various perspectives.

**Recommendation:**

Adopting this holistic approach to feedback collection enables the restaurant to foster stronger relationships with customers, drive continuous improvement, and enhance overall satisfaction and loyalty.

**User\_feedback with total spend**

**Methodology:**

In this dataset, we merged the 'Total Spend' data from the 'guest\_NAIT' dataset with the 'user\_feedback\_NAIT' dataset using the 'User ID'. This allowed us to create LRFM analysis, a method through which businesses can segment their customers based on their purchasing behavior. By analyzing this data, businesses gain a deeper understanding of their customer base and can tailor marketing strategies to specific segments. For the LRFM analysis, we utilized the following variables: 'Date of Feedback', 'User ID',  and 'Total Spend'.

**Findings:**

In Figure 14, the monetary category boasts the highest score at 3.46, signaling increased customer spending. Conversely, the frequency category records the lowest score at 1.14, indicating fewer visits per customer. A higher monetary score suggests a willingness to spend more per visit, while a lower frequency score implies infrequent visits. Moving to Figure 15, the LRFM segmentation uncovers valuable insights into the customer base composition. The 'Lost' category dominates, comprising 57.69% of customers, while 'Potential Loyalists' represent only 2.31%. This imbalance towards 'Lost' customers indicates a significant portion of the customer base hasn't made recent purchases, necessitating re-engagement efforts. The high percentage of 'Lost' customers underscores the risk of churn, highlighting the need for targeted strategies to win them back. Conversely, the presence of 'Potential Loyalists' presents an opportunity for fostering long-term loyalty, albeit at a lower engagement level.

**Recommendations:**

We recommend doing more promotions during hockey game nights and providing discounts for future visits. This strategy can incentivize customers to visit more frequently and increase their spending, ultimately enhancing customer retention and revenue. Other than that, targeted retention strategies such as craft beer tasting events can be employed to nurture these customers and encourage their progression towards higher engagement levels. Overall, these findings underscore the significance of both customer retention and reactivation efforts in improving overall customer retention and revenue growth.

**Craft Beers**

**Methodology:**

It's clear that Canadian consumers have a strong interest in craft beer, evident in venues like farmers' markets and liquor stores. Dining establishments offering diverse craft beer selections stand to attract customers seeking unique beverage experiences. To explore this trend further, we obtained two craft beer datasets from Kaggle and conducted a linear regression analysis. This enabled us to predict, forecast, and understand the relationships between variables related to craft beer consumption and sales. As part of the data transformation process, we merged the 'breweries', 'Product\_range', and 'Transactions' datasets. Additionally, we segmented the date into year and quarter segments to analyze profitability trends more effectively.

**Findings:**

The first finding (Figure 16) reveals that there are 45 different Canadian brewing companies worldwide. The second finding (Figure 17) underscores the seasonal trend in craft beer sales, with the highest profit observed during the summer in Canada. This suggests that consumers are more inclined to enjoy craft beer during the warmer seasons. The scatter plot analysis (Figure 18) reveals a strong positive relationship between sales amount and profit, indicating that as sales amount increases, so does profit. Applying a linear regression model to the data, we derive the formula Profit = 31.075 \* Amount + 595.485. This equation underscores the direct impact of sales amount on profitability, affirming that higher sales volumes lead to increased profits.

**Recommendations:**

It's recommended that The Canadian Brewhouse capitalizes on this seasonal trend by implementing more craft beer promotions during the summer. By offering enticing promotions and events centered around craft beer during the warmer months, The Canadian Brewhouse can potentially increase business revenue and attract more customers seeking refreshing beverage options. Given this evidence of the positive relationship between amount and profit, it's evident that promoting craft beers during the summer months, when sales are typically higher, can significantly boost business revenue for The Canadian Brewhouse.

**Literature Review**

Like "LRFMV: An efficient customer segmentation model for Superstores" (*LRFMV: An efficient customer segmentation model for superstores* 2022), we utilized the LRFM model to segment customers based on their purchase behavior, focusing on variables like recency, frequency, and monetary value. While our project built upon this existing research, we focused specifically on a dataset incorporating user feedback and total spend information. Using variables such as Date of Feedback, User ID, and Total Spend, we forecasted customer behavior and identified distinct customer segments. Similar to previous work by Guney et al., we categorized customers into groups like "Lost", "About To Sleep", "Recent Customers", "Loyal Customers", and "Potential Loyalists". Our project offered a unique perspective by applying LRFM analysis to a different dataset and adapting the segmentation approach to our specific context, aiming to enhance customer understanding and marketing strategy.

Our project focused on text feedback analysis for a thorough grasp of customer sentiment. Using the VADER model, we assigned scores to categorize feedback as "Positive," "Negative," or "Neutral," like an SVM classifier. By considering linguistic and contextual factors like word usage, morphology, and semantics, we achieved accurate sentiment analysis. Additionally, we compared average scores, providing insights for decision-making and improvements. This approach not only deepens understanding of customer sentiment but also offered actionable recommendations for enhancing satisfaction and business performance (*(PDF) literature review on sentiment analysis in social media: Open challenges toward applications* 2020).

**Redemption and Referral:**

The analysis begins with data preparation, where two main datasets are utilized: Redemption containing information on customer redemptions and Referral with details on customer referrals. These datasets provide crucial insights into customer interactions with the loyalty program. The Redemption dataset provides detailed information on customer redemptions. Each entry represents a transaction where a customer exchanges loyalty points for a reward. This dataset includes valuable details such as the type of item redeemed, loyalty points used, and redemption location. These elements collectively offer a comprehensive view of customer preferences and engagement levels. The referral data set focuses on customer referrals within the loyalty program. It records instances where customers refer others to join the program. Details like the referrer's ID, the referred user's ID, and the referral timestamp shed light on the social dynamics within the customer base. By cleaning, merging, and structuring these datasets, we gain a holistic view of customer engagement. Combining redemption and referral records allows us to explore correlations and patterns beyond individual transactions. This approach provides a deeper understanding of how customers interact with and benefit from the loyalty program.

**Statistical Analysis: Chi-square Tests**

A pivotal aspect of the analysis involves the use of Chi-square tests to evaluate the relationship between customer referrals and redemptions. The Chi-square test assesses whether there is a significant association between these two categorical variables. By comparing observed and expected frequencies of referrals and redemptions, we determine if the observed pattern is statistically significant. The Chi-square test for the relationship between customer referrals and redemptions was a key analytical step in this study (Figure 5). The Chi-square test is a statistical method used to determine if there is a significant association between two categorical variables, in this case, referrals and redemptions. This test was conducted to explore whether the act of referring a friend to the Canadian Brewhouse loyalty program had any impact on the likelihood of a customer making a redemption. To perform the Chi-square test, the data from the redemption and referral datasets were merged based on the common column 'User ID'. This allowed us to combine information about customer redemptions and referrals into a single dataset for analysis. The resulting dataset contained information about each customer's redemption status (redeemed or not) and referral status (referred to a friend or not). The Chi-square test was then applied to a contingency table created from this combined dataset. The contingency table was constructed with rows representing the redemption status (redeemed or not) and columns representing the referral status (referred or not referred). Each cell in the table represented the count of customers falling into a specific combination of these two categories. The null hypothesis for the Chi-square test was that there is no significant association between referrals and redemptions. The alternative hypothesis was that there is a significant association between these variables. The Chi-square test statistic was calculated based on the observed frequencies in the contingency table. After conducting the Chi-square test, the results indicated a p-value of less than 0.05, suggesting that there is a significant association between customer referrals and redemptions. This means that the act of referring a friend to the loyalty program is associated with a higher likelihood of a customer making a redemption. This finding has important implications for the Canadian Brewhouse, as it suggests that their referral program is effective in driving customer engagement and redemptions.

**Results of Chi-square Analysis**

The Chi-square test statistic of 247.44, along with the p-value of less than 0.001, indicates a strong association between these two categorical variables. In practical terms, this means that customers who engage in referrals, by referring friends to the Canadian Brewhouse loyalty program, are significantly more likely to participate in redemptions (Figure 6). This finding is crucial for the Canadian Brewhouse, as it highlights the potential effectiveness of their referral program in driving customer engagement and participation in the loyalty program. Customers who refer friends to the loyalty program likely have a vested interest in the program's success, whether it be due to the benefits they receive from successful referrals or their overall satisfaction with the program. This positive association suggests that referrals act as a catalyst for increased engagement, potentially leading to more frequent visits and purchases at the Canadian Brewhouse. The data suggests that focusing on encouraging and incentivizing referrals could yield substantial benefits by not only increasing the customer base through referrals but also boosting redemptions and overall program engagement. Furthermore, when customers personally recommend the loyalty program to their friends, it carries a certain level of trust and credibility that traditional advertising may not achieve. As a result, the Canadian Brewhouse may want to consider expanding and enhancing their referral program, offering compelling incentives for both referrers and new customers to drive this positive cycle of engagement. In conclusion, the Chi-square analysis unequivocally demonstrates the significant relationship between customer referrals and redemptions in the Canadian Brewhouse loyalty program. This insight not only validates the effectiveness of the referral program but also provides actionable recommendations for the Canadian Brewhouse to leverage this relationship to drive increased engagement, customer acquisition, and ultimately, loyalty program success.

**Combining User IDs for Comprehensive Analysis**

To ensure a comprehensive analysis, Python functions were meticulously crafted to merge user IDs from both the redemption and referral datasets within the Canadian Brewhouse loyalty program. This strategic approach enabled the creation of an exhaustive list encompassing all customers who either engaged in redemptions or referrals, thus affording us a holistic perspective on customer engagement within the program. The process of combining user IDs involved several steps to ensure accuracy and completeness. First, we carefully extracted user IDs from the Redemption and Referral datasets, recognizing the pivotal role user IDs play in linking customer activities across both datasets. Next, Python functions were designed to merge these user IDs, ensuring that no duplicates or missing entries were overlooked.

This meticulous merging process was essential for constructing a unified dataset that captured the entirety of customer interactions with the loyalty program. By consolidating user IDs from redemptions and referrals, we were able to establish a comprehensive list of customers who actively participated in the program. This not only facilitated streamlined data analysis but also provided a foundational basis for identifying patterns and trends in customer behavior.

In summary, the process of combining user IDs from both redemption and referral datasets through Python functions was instrumental in creating a comprehensive view of customer engagement. By leveraging this unified dataset, we were able to uncover insights that inform strategic decisions to enhance customer participation and loyalty within the Canadian Brewhouse loyalty program.

**Customer Segmentation: Redeemed vs. Referred**

Customer segmentation based on their activities within the Canadian Brewhouse loyalty program unveiled compelling insights into engagement patterns. Out of the total 242,833 unique customers analyzed, most of 240,303 customers actively engaged in redemptions. This indicates a strong inclination among customers to utilize their accumulated points or benefits within the program, showcasing the appeal and effectiveness of the loyalty rewards system. However, the analysis revealed that 18,517 customers participated in referrals, showing a notable interest in advocating for the program to friends and acquaintances. These "referring customers" play a crucial role in expanding the program's reach and acquiring new customers through word-of-mouth promotion. Their engagement highlights the potential of referral programs as a powerful marketing tool, leveraging existing customers as brand advocates. Furthermore, the segmentation identified 2,530 customers who did not redeem their points or benefits. This subset of customers presents an opportunity for targeted engagement strategies to encourage them to utilize their rewards, potentially through personalized offers or promotions tailored to their preferences. Notably, a significant portion of 224,316 customers did not engage in referrals. This understanding is essential for developing tailored strategies to enhance customer satisfaction, loyalty, and program effectiveness. By leveraging these insights, the Canadian Brewhouse can optimize its approach to customer engagement and retention, ultimately fostering a stronger and more loyal customer base.

**Top Redeemable Items Analysis**

Understanding the popularity of redeemable items is crucial for optimizing the loyalty program's offerings. To achieve this, we grouped data by location and redeemable item, identifying the top 10 most redeemed items for locations such as Calgary Mahogany, Lethbridge, and Lewis Estates. This analysis provides actionable insights into which items resonate most with customers at specific locations. The analysis revealed a consistent preference for specific redeemable items across all three locations, suggesting their universal popularity among customers. These items, such as the Golden Garlic Fingers and the Tri-Stacked Nachos, are highly sought after and present an opportunity for The Canadian Brewhouse to leverage their popularity for targeted promotions and marketing campaigns (Figure 7).

**Recommendations for The Canadian Brewhouse**

Based on the analysis conducted, several recommendations can be made to enhance customer engagement and loyalty program effectiveness:

Enhance Referral Program: The Canadian Brewhouse should implement targeted campaigns to incentivize customers who have redeemed items but have not yet referred anyone to participate in the referral program. Incentives such as exclusive redeemable items or discounts for successful referrals can drive engagement.

Promote Top Redeemable Items: Capitalize on the popularity of the top redeemable items by creating enticing promotions and combo deals. This strategic approach can attract more customers and boost redemption rates, ultimately leading to increased foot traffic and revenue.

Diversify Redeemable Items: While the top items are customer favorites, introducing new and unique redeemable items can create excitement and encourage repeat visits. Regularly updating the redeemable items menu can keep customers engaged and eager to participate in the program.

Location-Specific Promotions: Tailor promotions and offerings based on location-specific preferences. What resonates with customers in Calgary Mahogany may differ from those in Lethbridge or Lewis Estates. By personalizing promotions, The Canadian Brewhouse can strengthen its connection with local communities and drive customer loyalty.

**Conclusion**

In conclusion, the analysis presented in this report provides valuable insights into customer engagement and loyalty program effectiveness for The Canadian Brewhouse. The use of statistical methods such as Chi-square tests, combined with Python programming for data manipulation, allowed for a comprehensive exploration of customer behavior. The identified trends and recommendations offer a roadmap for optimizing the loyalty program, ultimately leading to increased customer satisfaction and loyalty.

**Guest Clustering:**

Automating customer segmentation from the guest dataset is essential for optimizing marketing strategies and enhancing customer engagement. By leveraging attributes such as demographics, behavior, loyalty points, device type, gender, signup channel, and other relevant factors, businesses can gain deeper insights into customer preferences and tailor their marketing efforts accordingly. Automating the segmentation process allows for more efficient and timely identification of distinct customer segments, enabling targeted messaging, personalized offers, and enhanced customer experiences. This proactive approach not only improves the effectiveness of marketing campaigns but also fosters stronger customer relationships, driving long-term loyalty and profitability for the business.

**Methodology**

Data Cleaning:

As a crucial initial step, the dataset underwent meticulous cleaning procedures. Categorical data within the dataframe was transformed into numerical data using the OneHotEncoder technique. Additionally, the 'Birthday' data was standardized into datetime format to accurately calculate guest ages. By subtracting the birth year from the current year, accurate guest ages were computed. An upper and lower bound for guest ages was established to rectify inconsistencies, transforming out-of-range ages to the mean age. Furthermore, non-essential time data was removed to optimize resource utilization.

Clustering Algorithm Application:

Following dataset preparation, the standardized data underwent the K-means clustering algorithm. Prior to clustering, the data was standardized using the StandardScaler to ensure appropriate weight allocation. The K-means algorithm effectively partitioned guests into distinct clusters based on similarities, facilitating a comprehensive segmentation analysis.

Feature Selection:

To facilitate clustering analysis, categorical variables such as device type, gender, and signup channel were subjected to one-hot encoding. This preprocessing step ensured effective utilization of categorical variables within the clustering algorithm.

Evaluation:

The efficacy of the clustering results was assessed based on the within-cluster sum of squares (WCSS). This metric served as a robust indicator of cluster quality, enabling thorough evaluation and refinement of the clustering outcomes.

**Findings**

The analysis revealed three predominant customer segments - Clusters 1, 0, and 7 - characterized by their high spending behavior and accumulation of substantial loyalty points. These segments represent highly engaged guests likely to respond positively to targeted marketing efforts.

An in-depth examination of the loyalty points earned across different signup channels unveiled MobileFacebook as the most lucrative channel, displaying significantly higher activity compared to other channels. This underscores the importance of targeted digital engagement strategies on platforms like Facebook, particularly for customers signing up through mobile devices.

**Implementation Plan:**

VIP Rewards Program Development:

A comprehensive marketing strategy outlining the benefits and rewards of the "VIP Rewards" program will be developed. Personalized email campaigns targeting top-spending segments will be designed and implemented to communicate the value proposition of the program and encourage guest participation. Continuous monitoring and optimization of engagement metrics will guide adjustments to the marketing strategies to enhance customer engagement and retention.

Digital Engagement Initiatives:

Collaboration with the marketing team will facilitate the creation of targeted Facebook campaigns tailored to customers signing up through mobile devices. Engaging content such as exclusive deals, contests, and behind-the-scenes footage will be developed to encourage customer interaction on the platform. Utilization of social media analytics tools will enable thorough tracking and analysis of engagement metrics to refine digital engagement strategies over time.

Menu Customization:

 As we continue to showcase NHL games at Canadian Brewhouse, we'll further enhance the experience by highlighting our in-house craft beers to entice craft beer enthusiasts. Leveraging insights from our collaboration with Tze Fung, we'll refine our craft beer offerings to better cater to this audience.

Event Planning:

Building on our existing NHL game showcases, we'll introduce special deals on game-day snacks and drinks to complement the excitement. To deepen engagement, we'll organize additional activities such as hockey trivia contests and jersey giveaways. Moreover, we'll implement a data collection initiative to learn about customers' favorite NHL teams, enabling us to personalize invitations for future game nights at Canadian Brewhouse. This holistic approach will not only enrich the NHL viewing experience but also foster a sense of community and camaraderie among patrons.

**Conclusion**

The guest clustering analysis has furnished invaluable insights into customer segmentation and behavior, providing Canadian Brewhouse with the foundation to develop tailored marketing campaigns and digital engagement strategies. Leveraging initiatives such as the "VIP Rewards" program and capitalizing on digital platforms like Facebook will be instrumental in enhancing customer satisfaction, loyalty, and overall business success. Continuous monitoring and optimization of these strategies will be pivotal in sustaining customer engagement and driving long-term growth and profitability.

**Recommendations for Data system**

Standardize Data Collection Procedures: Standardizing data collection procedures is essential, especially when dealing with inconsistent values in the 'Location name' column. By establishing clear guidelines for data collection, including the format and structure of location data, you can ensure consistency across different datasets. This will facilitate the connection between datasets and improve the accuracy of location-based analysis.

Ensure Data Quality: Given the challenges with inconsistent values and failed visualizations, prioritizing data quality is crucial. Implement measures to verify the accuracy, completeness, and consistency of the collected data. Conduct thorough data validation and cleansing processes to address any errors or discrepancies before proceeding with analysis. This will help mitigate issues related to varied dataset names and ensure reliable results.

Use Multiple Data Collection Methods: Employing multiple data collection methods can provide a more comprehensive understanding of customer behavior and preferences. Consider combining quantitative methods such as transaction records and website analytics with qualitative methods such as surveys or interviews. This diversified approach will help capture a wide range of insights and enhance the robustness of your analysis, particularly when dealing with complex issues like customer segmentation and engagement.

**Figures:**

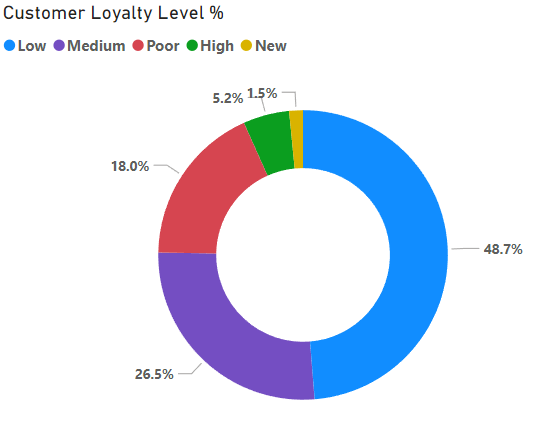


Figure 1. Customer Loyalty Levels from LRFM Analysis

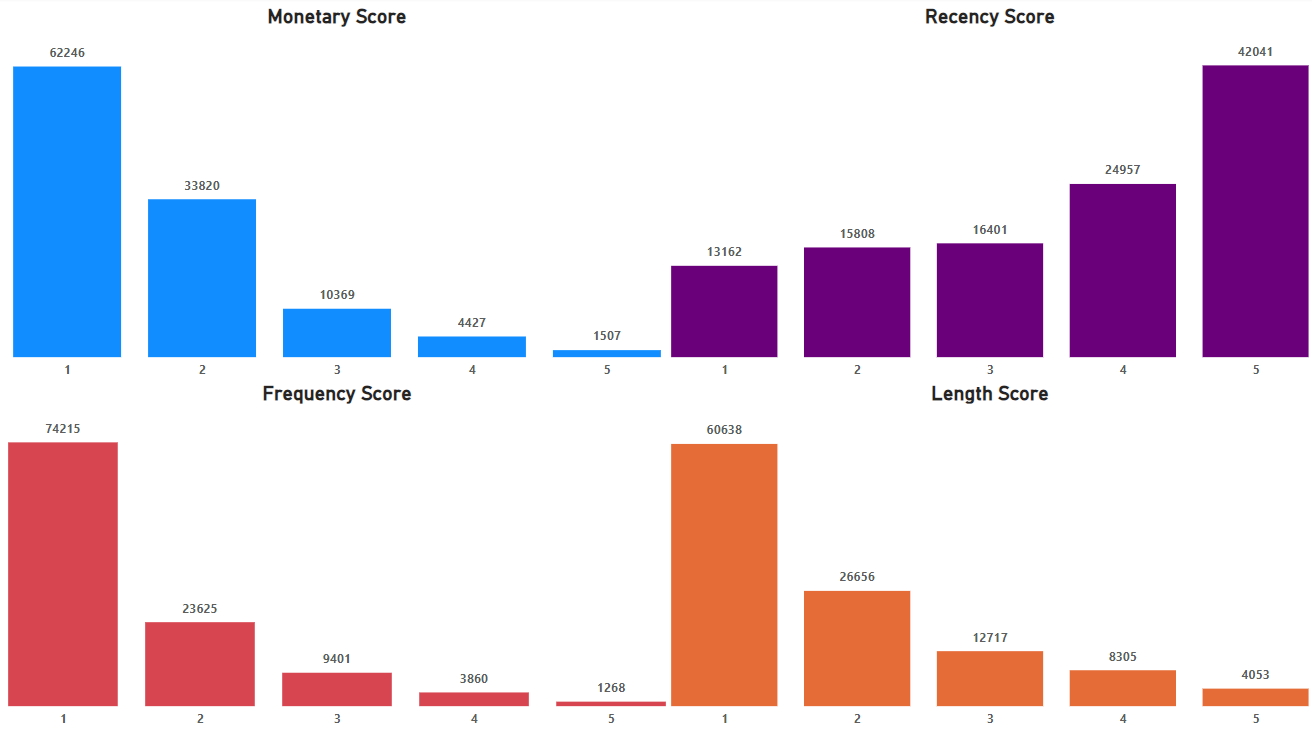


Figure 2. LRFM Score Counts

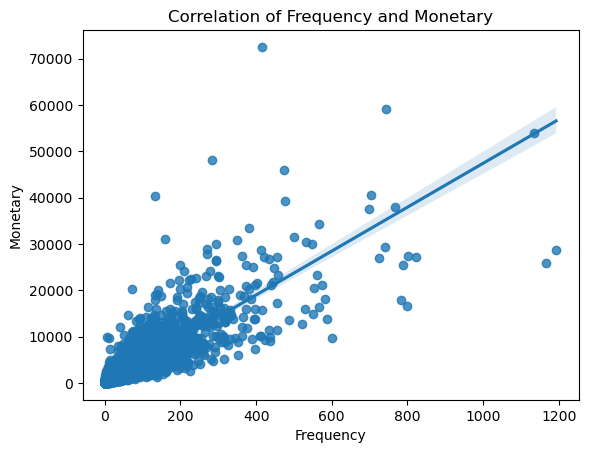


Figure 3. Correlation between Monetary and Frequency Values

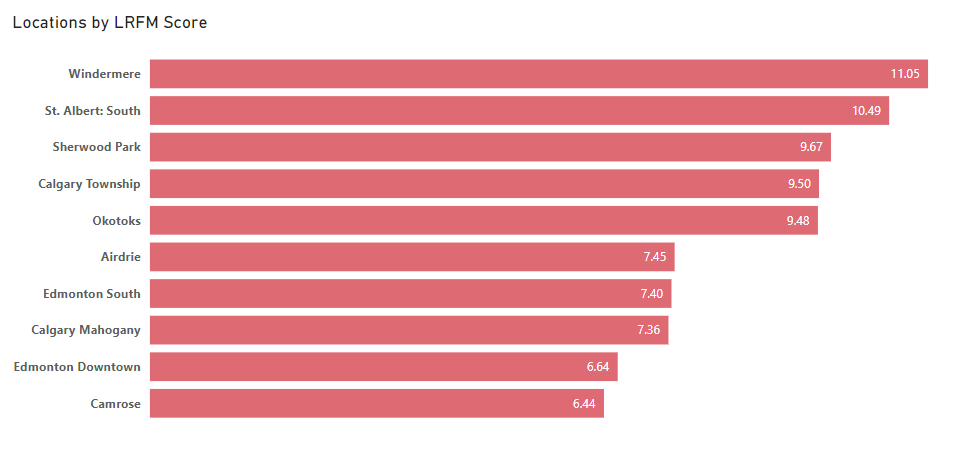


Figure 4. Top 5 and Bottom 5 Locations based on average LRFM Score.

A graph of blue rectangular bars with white text

Description automatically generated

Figure 5: Redemption Location and Redeem counts

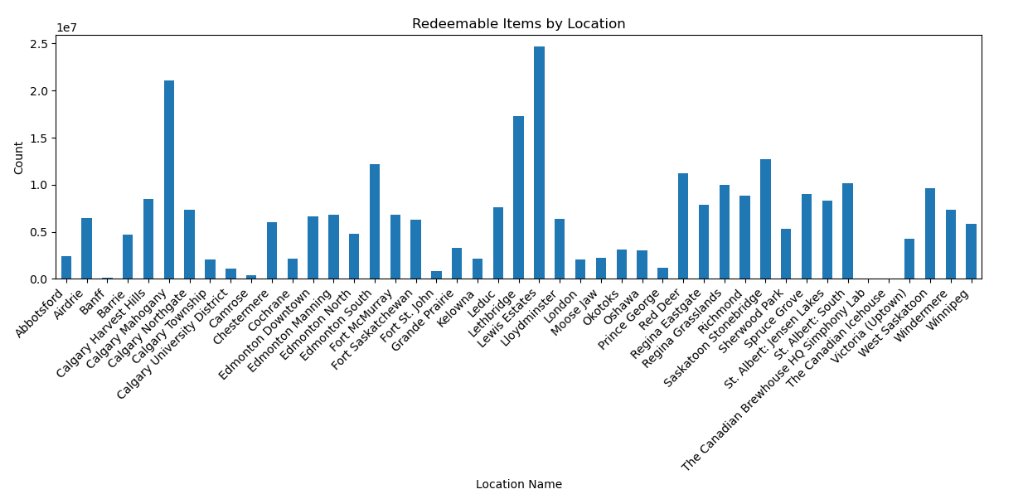
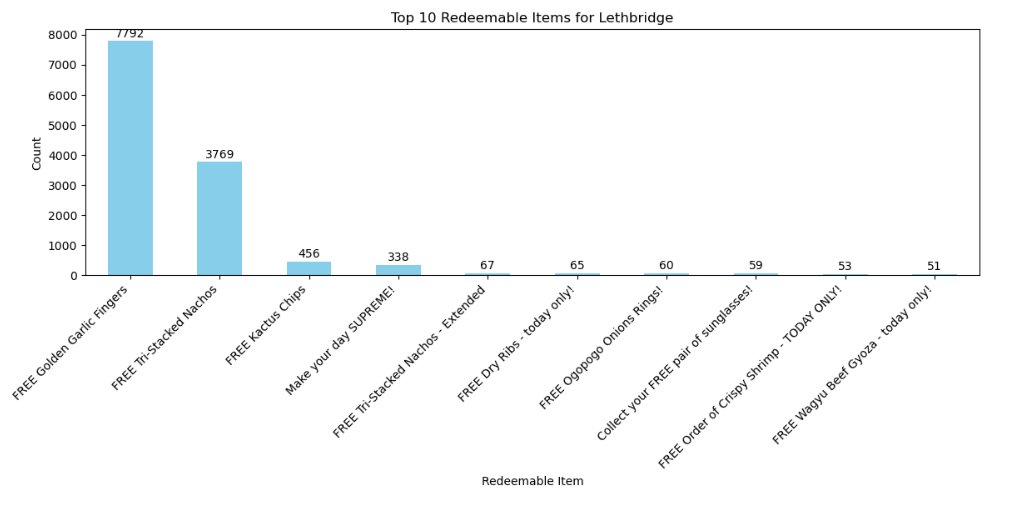
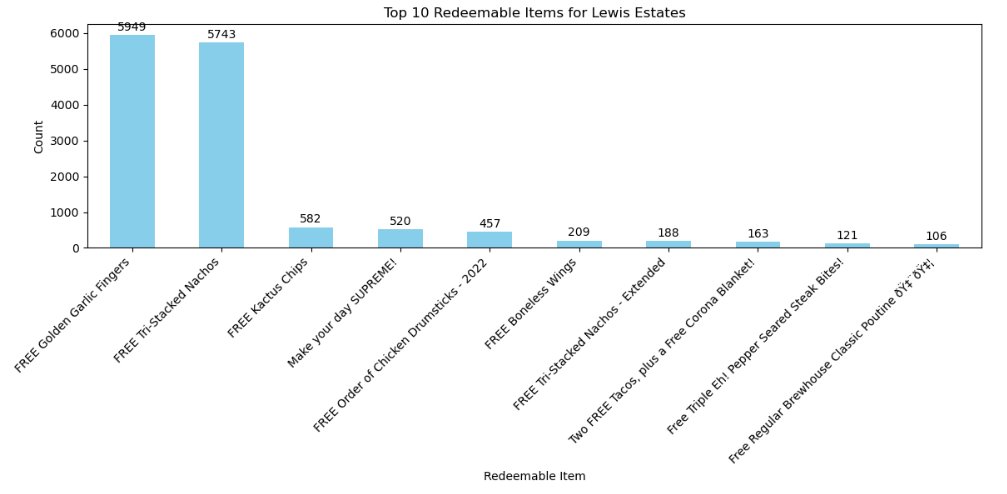


Figure 6: Referral locations and Redeem counts





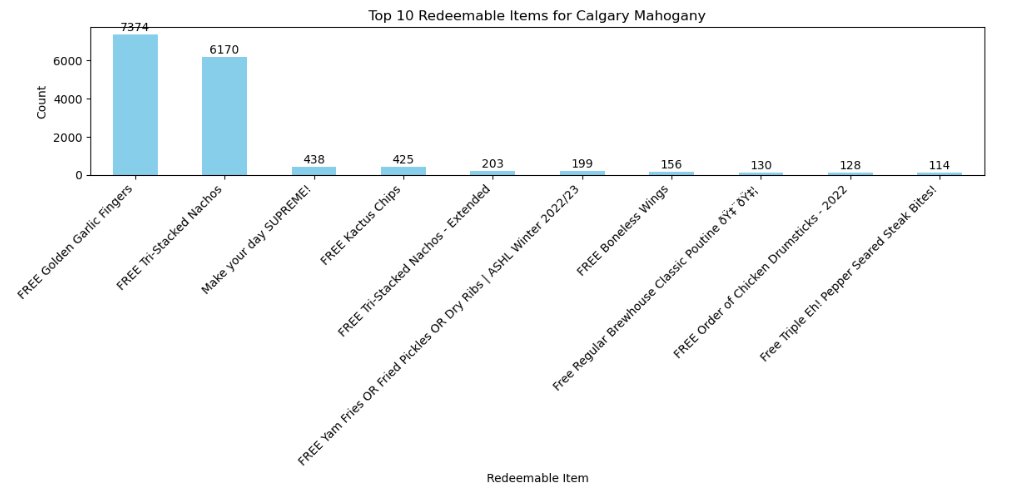


Figure 7: Top 3 Locations and Redeem Items

A line of numbers and letters

Description automatically generated with medium confidence

Figure 8: Average Rating of each location

A blue graph with white text

Description automatically generated

Figure 9: Count of Feedback by location

A colorful graph with text

Description automatically generated with medium confidence

Figure 10: Average Rating by location and year

A blue circle with a number of words

Description automatically generated

Figure 11: Pie chart of Feedback

A blue rectangles with black text

Description automatically generated

Figure 12: Sentiment score by location

A group of blue rectangular bars

Description automatically generated

Figure 13: Average Rating by location (Yelp Review)

A graph of different colored bars

Description automatically generated with medium confidence

Figure 14: LRFM scores for each category

A pie chart with different colored sections

Description automatically generated

Figure 15: Pie chart of LRFM segmentatio

A blue rectangle with white border

Description automatically generated

Figure 16: Canadian brewing companies

A line graph with a blue line

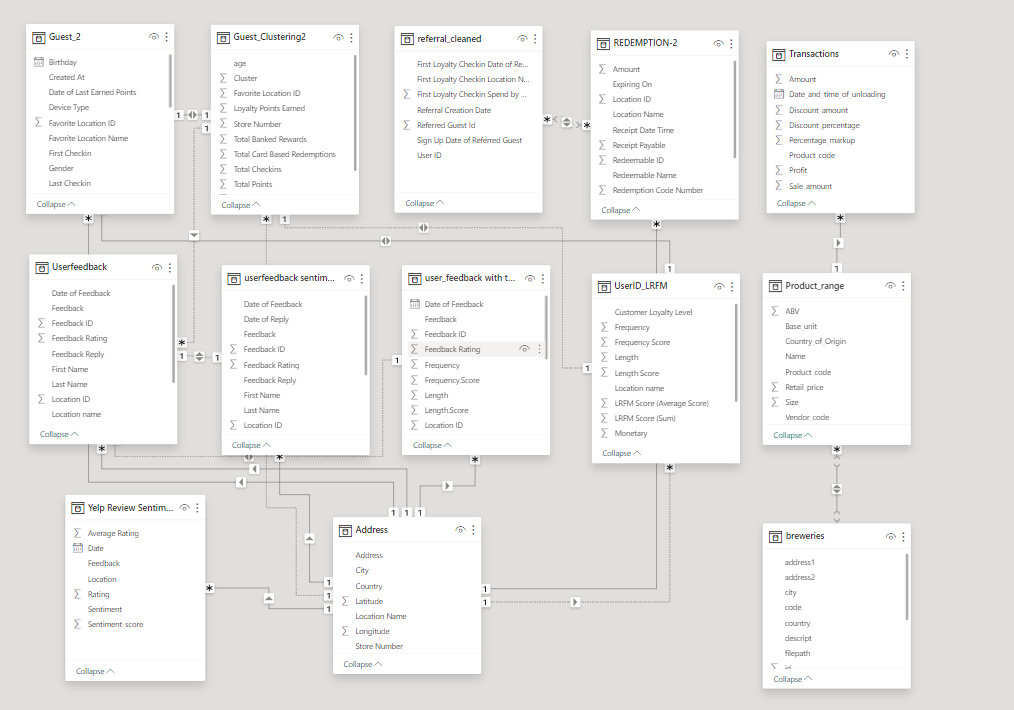
Description automatically generated

Figure 17: Profit vs Month

A graph with blue dots

Description automatically generated

Figure 18: Amount vs Profit



Appendix 1.

**References:**

Lewison, D. & Hawes, J. (2007). “Student Tarket Marketing Strategies for Universities” *Journal*  *of College Admission* (Summer), 14-19.

Delval, F. (2024, February 21). *What is RFM analysis?*. ActionIQ. https://www.actioniq.com/blog/what-is-rfm-analysis/

Craft Beer Bar Sales: <https://www.kaggle.com/datasets/podsyp/sales-in-craft-beer-bar/data>

Craft Beers: <https://www.kaggle.com/datasets/joebeachcapital/craft-beers/data?select=styles.csv>

Mahfuza, R., Islam, N., Toyeb, M., Emon, M. A. F., Chowdhury, S. A., & Alam, M. G. R. (2022, December 20). *LRFMV: An efficient customer segmentation model for superstores*. PloS one. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9767363/

Devi, G. D., & Somasundaram, K. (2020, January). *(PDF) literature review on sentiment analysis in social media: Open challenges toward applications*. ResearchGate. https://www.researchgate.net/publication/341913399\_Literature\_Review\_on\_Sentiment\_Analysis\_in\_Social\_Media\_Open\_Challenges\_toward\_Applications