Investigation of the Effectiveness of Starbucks Promotions

Kaitlyn Cardell carde156@umn.edu
Masha Volkova volko028@umn.edu
Nina Cruz cruz0192@umn.edu
Yasangi Withanawasam witha046@umn.edu
University of Minnesota - Twin Cities
Group 2

1. Introduction

This introduction section discusses some background information, the goals of our analysis, and the scope of the project.

1.1 Background

Our project aims to study the effectiveness of Starbucks promotions based on the promotional channel, as well as age, income, gender, and customer data. All of the data was collected over a 714 hour period of time.

1.2 Goals

The goal of this project is to analyze several SQL queries in order to understand how promotions affect customer transactions at Starbucks. Specifically, we aim to understand how different channels (web, email, mobile, social) impact the effectiveness of promotions, how customer frequency impacts the effectiveness of promotions, and how an individual's salary and age impacts their use of Starbucks promotions. We hope that our analysis will be beneficial in learning about how promotions influence consumer behavior, from the viewpoint of promoters. We hope to provide insight that will aid in more effective promotions.

1.3 Scope

The scope of this project is based on the scope of the dataset being analyzed. The data includes 714 hours worth of information about 17,000 customers and their interactions with Starbucks. The interactions ("events") include transaction, offer received, offer viewed, and offer completed.

2. Data Description

This data description section outlines the data source, the necessary data preprocessing, and the ER-Model used to understand the dataset.

2.1 Data Source

Our database includes data from a Kaggle dataset labeled "Starbucks Customer Data." The dataset includes three csv files: portfolio, profile, and transcript.

The portfolio table includes information about each of the ten promotions that were received during the 714 hour data collection period. The information includes a reward, what is received from redeeming a promotional offer; channels, a list of ways that the offer was promoted (web, email, mobile, and/or social); difficulty, the minimum dollar amount that a customer must spend to use the promotion; duration, the number of days the promotion was offered; offer type, the type of offer (bogo, discount, or informational); and offer id, a unique id to identify the promotion. Offers that are informational have a reward equal to 0.

The profile table includes information about each of the 17,000 customers. The information includes an index, an integer 1 through 17,000; gender (M, F, O, or blank); age; customer id, a

unique id for each customer; became_member, the date, formatted as yyyymmdd, that each customer became a member; and income. Age is 118 if the customer did not enter age, salary, and gender. These customers are excluded from analysis of age and income.

The transcript table includes information about events that occurred for each customer. The information includes id, a unique integer 1 through 306533; person, the unique customer id; event, the event that is being recorded (transaction, offer received, offer viewed, or offer completed); value, either the offer id or the value of the completed transaction; and time, the hour at which the event occurred.

2.2 Entity Relationship Model

Figure 2.1 represents the relationships between the three tables; portfolio (promotion), profile (customer), and transcript. Notice that customer id is the primary key for customer, promotion id is the primary key for promotion, and customer id, value, time, and event are all part of the primary key for transcript.

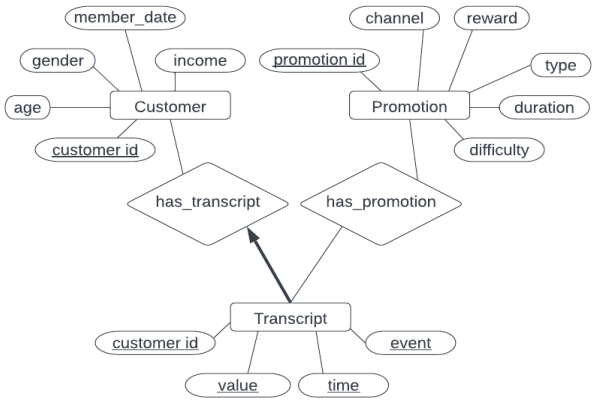

Figure 2.1: Entity Relationship Model

2.3 Data Preprocessing

Data preprocessing (deleting duplicates, and other data clean up) was required before the data was able to be queried.

2.3.1 Deletion of Duplicates

When attempting to create the primary key for the transcript file, the existence of duplicate rows disallowed this operation. We classified a duplicate as a transcript that had the same customer id, event, value, and time. To specify the primary key as {customer id, event, value, time}, we needed to remove these duplicates. We created a new table containing only rows that were duplicates. We then removed these rows from the transcript table where the id (the row id, as opposed to customer id) matched the row id in the transcript table.

2.3.2 Data Clean-Up

Other types of data clean up were also necessary for us to perform our queries. There are several rows in the profile table that contain missing information (gender, age, and income). We assume that member profile information was collected voluntarily, and therefore the missing values correspond to where members chose to not include their information. To ensure that these records did not impact the queries, results, and analysis, we removed the rows with missing data before performing queries that involved gender, age, and income.

3. Method & Queries

The methods and queries section lays out the four queries that were performed on the data.

3.1 Percent of Transactions that Use Promotions

This query was aimed at understanding the relationship between frequency of using promotions and the frequency of completing transactions at Starbucks. We first queried to determine the number of transactions per person and the number of promotions they used. This was done by joining the transaction view with the offer_completed view to create a new view called percentage_completed. We were able to graph the number of transactions against the percentage of promotions used per transaction. Instead of graphing a point for every individual, we grouped by number of transactions and averaged the associated percentages. For example, the promotion percentages for everyone who completed one transaction in the 714 hours were averaged together.

3.2 Income and Use of Promotion

This query was aimed at understanding the relationship between income and the use of promotions in transactions. First, we cleaned the data so that we were only using customer data in which each customer had an income listed. Then, we queried the database to understand the effect of income on the use of promotions. We joined the profile table with the percentage_completed view and grouped the results by income ranges. The resulting table had the income ranges, the average number of transactions, the average percent of transactions that used promotions, and the number of customers that were in the age range. The results of this query were formatted into a graph (can be found in section 4.2) to allow for a visual understanding and a deeper analysis.

3.3 Age and Use of Promotions

Another visualization we wanted from our database was the interaction of age and promotion use. Similar to the comparison of income and promotion use, we first removed data that did not have a true value for the age attribute. In the dataset, if a customer did not fill out their age in their profile, the age was set to 118. The query produced a table that had age ranges, the number of customers that fell into each range, the average transaction amount, and the average promotion usage. The table was a result of merging the profile table and the percentage_complete table on the customer id, then placing each member in an age range, then calculating averages.

3.4 Percent of Promotions Completed

For the final significant query, we aimed to determine how different channels used to disseminate information about a promotion impact the likelihood of a customer to view and subsequently use the promotion. For a company, knowing how many promotions are sent out and how many of them resulted in a sale would be helpful for future marketing. To answer these questions, we first created a view that contained the customerid, promotion ID, and offer ID for each event that was not a transaction (i.e. received offer, viewed offer, and completed offer). We subsequently grouped by promotion ID to create a table containing the count for offer received, offer viewed, and offer completed events for each promotion. We then combined this table with the portfolio of promotions containing information for the channels of dissemination of each promotion. Analysis on this data will primarily compare the percentage of offer viewed over offer received to determine which channels (or combination of channels) are most likely to result in a customer viewing an

offer they receive. We hope to provide insight on how to better distribute promotions to reap a larger gain.

4. Analysis of Queries

The analysis of queries section analyzes the four queries that were performed on the data.

4.1 Percent of Transactions that Use Promotions

Below is a graph (Figure 4.1) that displays the results from our first query, which was written to help us gain an understanding of how the use of promotions change with the frequency of Starbucks visits. Note that the percentages of transactions are computed as averages for each transaction count group. For example, all of the percentages of transactions for the customers who visited Starbucks 3 times during the 714 hour period are averaged together.

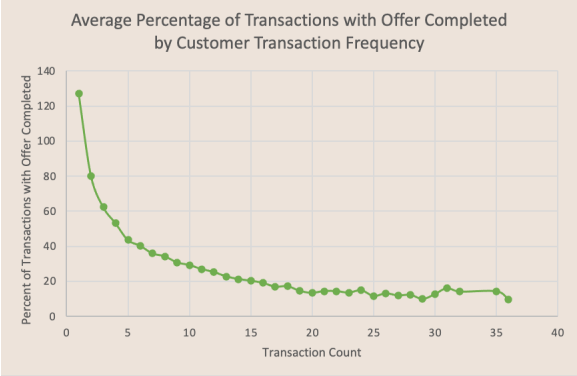

Figure 4.1: Average percent of transactions that use promotions vs. customer transaction frequency.

Figure 4.1 shows that offers are used more often in transactions when customers have a lower transaction frequency. A reason why this trend occurs could be because customers who purchase coffee at Starbucks less frequently are more likely to go if they have a promotion that they can use. This trend could also be because customers that go to Starbucks more frequently aren't able to use promotions as frequently in their transactions due to limited availability of promotion offers.

4.2 Income and Use of Promotion

Below is a graph (Figure 4.2) that displays the results from our second query, which was written to help us gain an understanding of how income affects the use of promotions. Note that the data points are located at the top of the income range for each range. For example, the data point used to express the averages for the $50,000 income group is the average across the $40,000 - $50,000 income range.

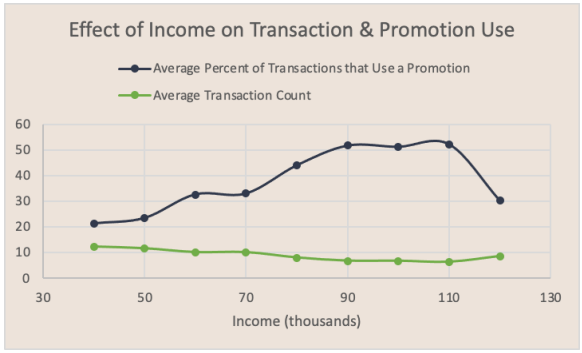

Figure 4.2: Average percent of transactions that use promotions and average transaction count per income grouping.

Based on Figure 4.2, you can see that the average count of transactions is not largely impacted by income. Each income group averaged about 10 transactions in the 714 hour period. On the other hand, it appears that the average number of transactions that used a promotion increases with income until about $110,000 where there is a heavy decrease. This result is both interesting and surprising. We expected to see a decrease in the usage of promotions as income increases due to an increase in financial stability. The heavy decrease in promotion use after $110,000 can be

explained by this, but the initial increase would need more analysis to be understood.

4.3 Age and Use of Promotions

Below is a graph (Figure 4.3) that displays the results from our third query, which was written to help us gain an understanding of how age affects the use of promotions. Note that the data points are located at the top of the age range for each range. For example, the data point used to express the averages for the 40 year old age group is the average across the 30 year old - 40 year old age group.

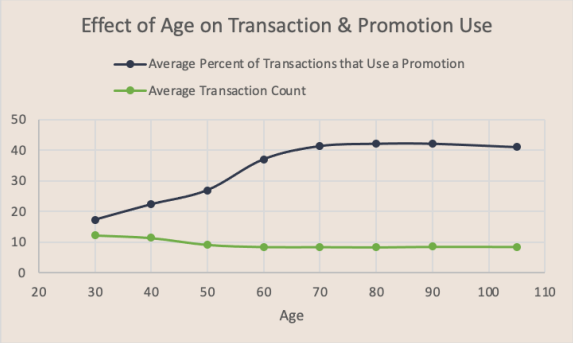

Figure 4.3: Effect of age on average percent of transactions that use promotions and average number of transactions

Figure 4.3 shows that as the age of the customer increases, the average percentage of transactions where promotions are used increases and the average number of transactions decreases slightly. A reason that the average number of transactions decreases as age increases, could be that younger customers tend to go to Starbucks more often than older customers. A reason that the average percent of transactions that use a promotion increases as age increases could be that older customers tend to use coupons and promotions more often than younger customers and because they are motivated by promotions to purchase coffee at Starbucks. Another reason for this trend could be that younger customers are not able to use promotions as often as they buy coffee due to limited availability of offers.

4.4 Percent of Promotions Completed

Figures 4.4.1 and 4.4.2 below show the results from our query comparing the rate at which customers view a received offer by promotion ID. The figures are also colored by which – or how many – channels that each promotion uses to disseminate information.

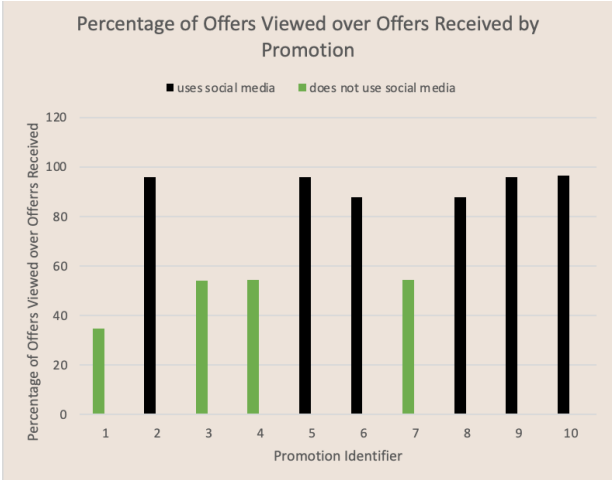

Figure 4.4.1: Percentage of offer viewed occurrences out of offer received occurrences by promotion, colored by whether the promotion uses social media as a channel

Figure 4.4.1 compares promotions which do use social media as a channel and those that do not. The comparison metric is the percentage of offers viewed out of offers received, which we refer to as the view rate. Significantly, the four promotions with the lowest view rate are those that do *not* use social media as a channel (the green bars in Figure 4.4.1). This result implies that social media is necessary to maintain view rate of promotions, which could in turn stimulate use of promotions and Starbucks purchases.

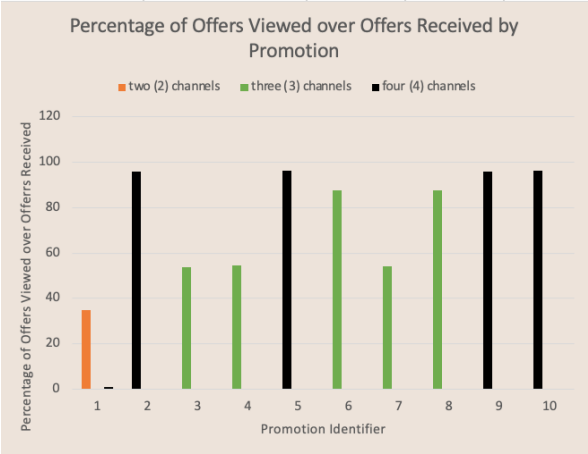

Figure 4.4.2: Percentage of offer viewed occurrences out of offer received occurrences by promotion, colored by number of channels a promotion uses

Figure 4.4.2 is colored by the number of channels that a promotion uses. Notably, the four promotions which use four out of four available channels have the highest view rates. The promotion with the lowest view rate uses only two channels, and the rest of the promotions use three. These results fit with a hypothesis that greater dissemination of promotions (through more channels) will increase the likelihood of a customer to view the promotion.

5. Conclusion

We created a database that contains transaction and profile data for Starbucks customers and promotions. The customer profiles are linked to promotions via transactions that are made with various offers. We had four specific queries that we analyzed for the impact on Starbucks transactions and use of promotions (all four are described in Section 3 and analyzed in Section 4). The first query helped us understand that low frequency customers (we defined this as < 10 transactions per the 714 hour time period, about one month) had a higher usage of promotions than customers that were considered high frequency (we defined this as ≥ 10 transactions per the 714 hour time period, about one month).

With our second query we noted that as customer income increases, the average percentage of transactions that use promotions also increases until you reach the income of $110,000 where there is a significant decrease in promotion usage. This suggests that customers that have an income over 110k have a lower usage of promotions due to financial stability. However, we would need to further investigate to determine why promotion usage increases as income increases before 110k. From our third query we learned that the average number of transactions decrease with age while the percentage of transactions with promotions increase. This suggests that older people are more likely to use promotions. The fourth query gave us an understanding of how the promotions filter from being sent to customers to actually being viewed by customers and how that relates to the methods of communication. We noticed that there was a higher amount of return on promotions that used three or more channels and those that use social media as a channel. This suggests that the more promoted a promotion is, the more people you are likely to draw. It also suggests that mobile communication methods have the largest impact. We also notice that customers in general preferred promotions that were discounts over bogos and informational. Of the top four promotions used, three were discounts. Members also tended to lean towards using promotions that had a better risk to reward ratio. Promotions that had lower difficulty but a moderate reward were more likely to be used. We suggest that Starbucks focuses on promoting to younger ages that are more likely to buy coffee, while maintaining their promotions to older customers and low frequency customers to draw in their business.

6. Acknowledgements

Nina Cruz: Pre-processed data, Managed database and wrote queries, Contributed to data analysis and query descriptions
Masha Volkova: ER Diagram, Query Analysis, Part of Conclusion
Kaitlyn Cardell: Wrote Introduction and most of Data Description, and some Query Analysis.
Yasangi Withanawasam: Data Description, Contributed to analysis of query results, conclusion, Methods

7. References

[1] Muliar, Ihor. “Starbucks Customer Data.” *Kaggle*, 31 Mar. 2021, https://www.kaggle.com/datasets/ihormuliar/starbucks-customer-data?select=transcript.csv.

[2] RAMAKRISHNAN, Raghu, and Johannes Gehrke. *Database Management Systems: (3rd Ed)*. McGraw-Hill, 2003.