DATA 3960 – Group 1 Project Proposal – Canadian Brewhouse Dataset

Group 1 Roles:

As a Group:

* Determining hypotheses
* Analyzing results of tests and predictions
* Writing report and preparing presentation

Esha Basharat: Creating dashboards/visualizations

Tsz Fung Cheung: Data cleaning/exploration

Michael Harris: Modeling

Kurtis Li: Data cleaning/exploration

Statement of Problem:

In our pursuit of enhancing the success of the Canadian Brewhouse, our team has undertaken a project centered around a crucial objective: improving the existing customer retention rate. Much like others in the hospitality sector, the Canadian Brewhouse confronts the immediate necessity to create and strengthen a dedicated customer following amid a competitive market. Acknowledging the pivotal role of customer loyalty in ensuring lasting success, our primary aim is to explore the diverse factors impacting customer retention and, thereafter, formulate strategic measures to enhance its resilience.

At the core of our project lies a fundamental question: How can we enhance the customer retention rate for the Canadian Brewhouse? To tackle this overarching question, we have outlined a comprehensive approach focused on comprehending, analyzing, and innovating within the sphere of customer retention. Key inquiries guiding our exploration encompass: How can we elevate customer retention? What factors exert sway over customer retention? And, what inventive strategies can be implemented to strengthen this vital metric?

Our group discussions have produced a range of potential strategies, ranging from introducing more captivating rewards and customizable offers aligned with individual buying habits to incentivizing referrals and implementing surprise perks to enhance the overall customer experience. Additionally, we have delved into the prospect of analyzing customer buying habits, using historical purchase data to customize points offers and tailor rewards to individual preferences. The objective is clear: to not only attract customers but, more importantly, to keep them engaged and loyal through a nuanced and personalized loyalty program.

Our project aims to achieve two primary goals. Firstly, we aim to comprehend the existing impact of the Canadian Brewhouse’s loyalty app program on customer retention. Secondly, we aspire to discern potential pathways for improvement. To achieve these goals, we will scrutinize the datasets at our disposal, identifying patterns in customer behaviors and trends. We seek to answer specific questions, such as whether the rewards offered through the app prompt increased customer visits and spending, if special offers such as Oilers Game Days promotions and reduced ticket prices lead to a surge in customer visits, and whether personalized offers and the referral program contribute significantly to heightened customer engagement.

In essence, our project is a strategic exploration into the intricacies of customer behavior, with the ultimate aim of providing the Canadian Brewhouse with actionable insights. By aligning our goals with the broader objectives of the Canadian Brewhouse, we envision crafting a proposal that not only outlines our commitment to addressing the customer retention challenge but also positions us as catalysts for positive change within the organization. Through this proposal, we hope to embark on a journey that not only analyzes past performance but, more crucially, charts a course for innovative and effective customer retention strategies in the future.

Description of the Dataset:

The Canadian Brewhouse internal dataset is a comprehensive collection of information spanning from December 2018 to December 2023, shedding light on various facets of the establishment's operations. Two primary datasets, Guest\_NAIT and User\_Feedback\_NAIT, form the core of this repository.

Guest\_NAIT Dataset:

This dataset encompasses crucial user data, tracking customer interactions and preferences within the Canadian Brewhouse. It includes details such as first and last check-ins, total check-ins, loyalty points, and the channel through which users sign up. Additionally, it holds valuable information for customer segmentation and targeted marketing, offering insights into customer behavior over the five-year period.

User\_Feedback\_NAIT Dataset:

Focused on customer sentiments, this dataset captures direct feedback from patrons. It includes information about the store, promotions, and other aspects, providing a qualitative understanding of customer experiences. By analyzing this dataset, the establishment gains insights into customer satisfaction, areas of improvement, and the correlation between feedback and customer retention.

Pros of Using the Dataset:

Holistic Customer Insight: The dataset allows for a 360-degree view of customer behavior, combining quantitative metrics (from Guest\_NAIT) with qualitative insights (from User\_Feedback\_NAIT), providing a holistic understanding.

Longitudinal Analysis: The five-year span allows for longitudinal analysis, enabling the identification of trends, seasonality, and changes in customer behavior over time. This historical context is invaluable for strategic decision-making.

Real-Time Monitoring: The incorporation of real-time metrics, such as check-ins, facilitates the creation of dynamic dashboards for ongoing monitoring of customer behavior. This enables timely responses to emerging trends and issues.

Customer Segmentation: The dataset provides the opportunity for robust customer segmentation based on various metrics like loyalty points, allowing for targeted marketing strategies tailored to specific customer groups.

Cons of Using the Dataset:

Data Quality Challenges: Over a five-year period, data quality issues may arise, including missing or inconsistent entries. This could impact the accuracy of analyses and decision-making.

Privacy Concerns: Handling user data necessitates a robust privacy framework to ensure compliance with regulations. Sensitive customer information requires careful handling and protection.

Complexity in Feedback Analysis: Analyzing qualitative feedback can be intricate due to the subjective nature of responses. The interpretation of sentiments and identifying actionable insights may pose challenges.

Resource Intensive: Creating real-time dashboards and implementing sophisticated analyses requires substantial computational resources and skilled personnel, potentially posing resource challenges.

The Goal of Our Analysis:

Ultimately the goal of our analysis is to better understand how customer retention is currently affected by the Canadian Brewhouse’s loyalty app program, and what are some potential ways that it can be increased. We plan to do this by looking into the datasets to determine patterns in the data that can help us understand customer behaviors and trends, and then use these patterns to create predictions for how customer retention can be affected by altering certain aspects of the app program.

A few important factors that we intend to investigate are which aspects of the data influence when people visit the Brewhouse and how often they visit, and how successful the current rewards and campaigns are in their effort to drive customer retention. More specifically some goals of our analysis in terms of questions that we hope to test and answer are:

* Do the rewards offered through the app increase customer visits compared to a customer that is offered less rewards, and do they spend more money on a visit.
* Would special offers such as Oilers Game Days promotions and reduced tickets prices cause an increase in customers’ visits.
* Would personalized offers or rewards increase customer visits and spending.
* Does the referral program cause an increase in visits for the referring customer

These are the goals of our analysis at this initial stage of the project, however if through our analysis we find that there is a strong relationship between customer visits and certain other aspects of the loyalty program, we plan to make that a focus going forward.

Technical Approach:

While we have several datasets provided to us by the Canadian Brewhouse, some of them are quite large and contain plenty of missing values. The first step of our technical approach will be to perform some basic exploratory data analysis by cleaning and organizing the datasets so they can be used. Of particular importance will be the null data that can be a result of errors in the data collection, or values that are not pertinent to our uses such as personal info. For cases of entire columns that are missing it will be best to remove them to make the dataset more manageable. If there are missing values within a column then they should be dealt with based on what they represent.

Missing categorical data may mean that the user did not fill out that information and that may not be as important to our needs. However missing numerical data could be a zero or an actual missing data point, and depending on the specific data type that data may need to be ignored or changed.

We will then use this data to create one or more predictive models that can predict how often a customer will return to the Brewhouse based on factors such as the rewards they received or the amount of money they spend. In terms of the types of models, we will use a regression model, but others may become good candidates as we progress in our analysis. Throughout this semester we will be learning various other modeling types and techniques, and we intend to apply those to this project. The types of models used will depend on the specific factors that we find most informative about customer retention and how those specific models will produce good results for our project.

Any cleaning or manipulation of the data, and the creation of predictive models, will be done by using the python coding language, more specifically through Visual Studio Code. Visualizations of the data and the results of our models will be created using Microsoft Power BI.

As we progress through this project, we may encounter technical problems with the data or the modeling aspects of our analysis. There may be issues with the data itself or in the way that the model is coded and structured. Python can be a confusing language at times and there may be a small error that has created an issue with our analysis. In each case we plan to work through any issues by first attempting to understand the cause of the issue, if the issue is the quality or cleanliness of the data then we need to go back and organize it better. If the issue is in the coding, then consulting other resources or class notes may be necessary to understand the syntax or logical error we have encountered. Implementing various error checks within the code would help to make this process much quicker.

Data Products our Project will Produce:

For the data products our project will produce, the first one is descriptive statistics. We can use mean, median, mode, and standard deviation for key customer retention metrics. The mean can give us an overall view of customer retention performance, while a higher standard deviation indicates greater variability, helping us understand the stability of our retention metrics. In addidtion, we can create distribution plots to visualize the spread of customer behavior data. For example, histograms can be used to visualize the frequency distribution of customer retention rates, helping identify the shape of the distribution. For example, we can use the 'guest\_NAIT' dataset to create a histogram for loyalty points to determine whether the member is active or inactive.

Secondly, we can use a dashboard to provide real-time updates on customer behavior metrics and alerts for significant changes in customer retention patterns. For example, by utilizing the 'checkin' dataset, we can design a dashboard to pinpoint the date with the highest customer visits. This dashboard can also help identify total active customers, churn rate, and the influx of new customers. Additionally, we can leverage the 'guest\_NAIT' dataset within the same dashboard to ascertain the average spending per customer.

Finally, we can use customer feedback to integrate survey data and understand customer sentiments. We can visualize customer satisfaction scores across different touchpoints to identify areas of improvement and create visualizations of survey responses and their correlation with retention. For example, we can use the 'user\_feedback\_NAIT' dataset to determine areas of improvement, identifying them based on the number of scores and the feedback.

Project Management Plan: (~12 weeks)

1. Data Cleaning – 2 weeks
2. Data Exploration – 1 week
3. Modeling – 2 weeks
4. Testing Hypothesis – 2 weeks
5. Analyzing Results – 1 week
6. Creating Dashboard/Visualization - 1 week
7. Writing Report – 2 weeks
8. Creating Final Presentation – 1 week

In an effort to ensure our project is moving smoothly we have created the above schedule roughly outlining how long we expect to spend on each stage. If a certain stage takes a longer or shorter amount of time, we will adjust and spend time where it is most necessary. Our group plans to maintain good communication with each other throughout the process, with a group meeting at least once a week. We also have a bi-weekly meeting with our professor Dr. Asif Nashiry to discuss our progress and ask any questions.

Important Deadline Dates:

Mid-Point Presentation: March 5

Final Presentation: April 23