**Assignments: ALY6080 90325 Integrated Experiential Learn SEC 03 Summer 2023 CPS [BOS-1-HY]**

**Module 10 Assignment — Individual Project Proposal Draft**

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by**

Zihan Ma

ma.zihan1@northeastern.edu

**Instructor**

 Prof. Valerie A. Atherley

**Date**

08/20/2023

**Advanced Analytics Approach to Customer Lifetime Value Optimization for Locally Inspired**

1. **Statement of Purpose**

Locally Inspired is an innovative, community-focused retail store that thrives on providing high-quality, handcrafted products representing the talents of local Wisconsin artisans. While the retail store has been successful, there is a need to take a strategic step towards enhancing the customers' experience, improving customer retention, and increasing overall profitability.

The underlying business question is, "How can Locally Inspired predict and maximize Customer Lifetime Value (CLTV)?" Understanding and enhancing CLTV is a proven approach to boost a company's long-term financial performance. By proposing this project, we aim to develop a predictive model and an interactive dashboard to better understand the buying behavior of our customers and subsequently increase their CLTV. This proposal will help Locally Inspired not only to retain loyal customers but also to increase the monetary value each customer brings.

1. **Scope of the Project**

The proposed project focuses on the development and implementation of an analytical tool that can accurately predict the CLTV of Locally Inspired's customers. The deliverables for this project include:

1. An analytical model that accurately predicts the CLTV based on customer purchase history.
2. A dynamic, interactive dashboard that allows real-time visualization of critical customer metrics (Recency, Frequency, and Monetary Value).
3. A comprehensive report providing strategic recommendations to enhance customer engagement and increase CLTV based on the insights derived from the model and the dashboard.
4. **Background Research and Literature**

Understanding and enhancing CLTV is a proven approach to boost a company's long-term financial performance. This proposal will help Locally Inspired not only to retain loyal customers but also to increase the monetary value each customer brings. For instance, a prominent online retail store implemented a CLTV predictive model, resulting in a significant increase in its customer retention rate and overall profitability. This past success offers promising prospects for Locally Inspired as well.

Further research, such as that by Peinkofer, S. T., Esper, T. L., Smith, R. J., & Williams, B. D. (2023) shed light on the role of price promotions in enhancing customer retention in the face of product stockouts, thereby informing inventory management strategies and promotional activities (Peinkofer, Esper, Smith, & Williams, 2023).

Another critical aspect of online retail is customer segmentation. Two studies by Turkmen (2022) delve into the exploration and comparison of several customer segmentation methods, employing machine learning techniques on online retail data, which can be beneficial for Locally Inspired's data processing (Turkmen, 2022).

1. **Design and Data Collection Methods**

Our proposed CLTV project for Locally Inspired involves a comprehensive plan for identifying key customer metrics, calculating CLTV, and presenting this information in a format that is both accessible and easy to interpret. In order to ensure the accuracy and effectiveness of our project, we will follow a methodical approach based on the following steps:

1. **Identify necessary columns:** From the dataset provided, we'll isolate the columns that will be pivotal in our calculation of CLTV. The necessary columns from the Locally\_Inspired\_Order\_1\_1\_23\_5\_31\_23 dataset include 'Email', 'Total', and 'Created at'. The 'Email' field will serve to identify unique customers, 'Total' will inform us of the amount spent by the customer, and 'Created at' will allow us to determine the customer's tenure. Since there isn't a specific column indicating customer churn, we'll derive this by considering customers with no purchases within a recently defined period as churned.
2. **Calculate individual customer metrics:** Our focus will be to establish the Recency, Frequency, and Monetary Value for each customer. The recency metric will be determined by how recent the customer's last purchase was. The frequency will refer to the number of purchases made by the customer, thus indicating their purchase habits. Monetary Value will represent the average amount the customer spends. These three metrics will form the backbone of our CLTV calculations.
3. **Calculate CLTV:** The calculation of Customer Lifetime Value hinges on the metrics derived above. We'll calculate the CLTV using the following formula: CLTV = (Customer\_Value / Churn) \* Profit\_margin. Here, Customer Value is determined by the product of Average\_Order\_Value and Purchase\_Frequency. We'll assume a profit margin as part of our calculations since it's not provided in the dataset. It is important to note that while this approach provides a straightforward way to calculate CLTV, it is dependent on the specific nature of the business, the products it offers, and the purchase cycle of its customers. Therefore, we might need to refine our calculations to better suit Locally Inspired's unique business context.
4. **Visualization and Presentation:** To ensure that our findings are easily understandable, we'll resort to visual representation techniques. We'll use graphical methods like histograms, pie charts, and bar graphs, which are intuitive and simple to interpret. These visuals will be created using the 'ggplot2' and 'plotly' packages in RStudio. The goal is to create a visual narrative that highlights key takeaways about Locally Inspired's customer value.

Ultimately, the purpose of this analysis is to uncover who the most valuable customers are, how they can be retained, and what strategies can be employed to increase the value of less profitable customers. The data collection and design methods we plan to use are geared towards this goal, promising a fruitful result for Locally Inspired.

1. **Implementation Methodology and Strategies**

The implementation of this project will be informed by the methodologies discussed in two key research studies: "An Analysis of Mechanism for Customers' Purchase Amount and Number of Visits in Department Store" by Hiroki Yamada and Tadahiko Sato, and a chapter focusing on customer lifetime value (CLV), customer retention, and churn.

The first research piece suggests the use of a hierarchical Bayes regression model to analyze customers' purchasing amounts and a hierarchical Bayes Poisson regression model to estimate the number of visits. These models aim to reveal the hidden heterogeneity within customer behaviors and thus offer more precise predictions. In our case, we will apply these models to analyze and predict the Recency, Frequency, and Monetary Value of Locally Inspired's customers, which are key components for CLTV calculation.

The second study illuminates how to calculate CLV effectively in different market situations and emphasizes the critical role of customer retention and churn in the calculation of CLV. Based on the market nature of Locally Inspired, which is a non-contractual market, we will focus on churn probability, customer migration, and lifespan for CLTV calculation. Given the potential heterogeneity in our customer base, we will lean towards probabilistic models rather than deterministic ones. This decision aligns with the recommendation from the research, as deterministic models may overlook individual variations in customer behaviors.

Data analysis and model development will be performed in RStudio, leveraging its powerful and versatile libraries for statistical analysis and modeling. To address the issue of unobserved customer defection, survival analysis methods will be adopted, which are particularly effective in estimating customer churn rates and lifespan in non-contractual settings.

The interactive dashboard will be built using the 'Shiny' package in R. It will provide real-time visualization of key customer metrics and CLTV predictions. The dashboard will also allow for data manipulation and scenario analysis to facilitate more informed decision-making.

Upon completion of the model development and dashboard creation, we will conduct a validation phase where we test the model against unseen data to assess its prediction accuracy. Refinements will be made as necessary to improve the model's performance.

A screenshot of a computer

Description automatically generated

https://onlinelibrary.wiley.com/doi/full/10.1111/jbl.12095?casa\_token=T8CayCCFk\_sAAAAA%3AAh\_L1tle1QhG-183N\_Pr9S2jAn\_9t3SvoyBUPMkNtv4oUUl0vYerMtQzV9ztAf6rjLRnZ0d1mgdl9g8

As a result of the project, a model like above will be built that can cluster and predict how the promotion will affect different groups of customers’ buying will.

1. **Conclusion**

Predicting and optimizing CLTV is a strategic initiative that can significantly influence Locally Inspired's growth trajectory. Through the development and implementation of a predictive model and a real-time dashboard, the retail store can gain valuable insights into its customers' buying behavior. This knowledge can further drive effective strategies that enhance customer retention and boost the monetary value of each customer.

By focusing on Recency, Frequency, and Monetary Value of customer transactions, Locally Inspired can significantly improve its understanding of its customers, which is the cornerstone of any successful retail business. With the proposed tools in place, Locally Inspired can expect a stronger customer base, increased customer loyalty, and consequently, improved profitability.

**Annotated Bibliography**

Peinkofer, S. T., Esper, T. L., Smith, R. J., & Williams, B. D. (2023). Assessing the Impact of Price Promotions on Consumer Response to Online Stockouts. Journal of Business Logistics, 46(4), 317-333. <https://onlinelibrary.wiley.com/doi/full/10.1111/jbl.12095?saml_referrer>

This empirical study focuses on how consumers react to stockouts, particularly of price-promoted products, in the e-commerce landscape. The authors deploy the expectation-disconfirmation theory (EDT) to scrutinize how price promotions impact consumers' expectations of product availability and their subsequent response to online stockouts. They discovered that in low involvement shopping scenarios, consumers demonstrate less dissatisfaction and a lower likelihood of switching to a competitor's online platform when they encounter a stockout of a discounted item. The research introduces a potential limitation of EDT in high involvement scenarios, suggesting the need for additional theoretical refinement in such contexts.

Peinkofer et al.'s research is noteworthy for its exploration of an under-researched area in e-commerce, providing valuable insights into consumer behavior and inventory management in the online retail context. Their experimental approach and robust theoretical framework elucidate significant implications for both academic study and practical applications in supply chain management. The authors convincingly argue that price promotions, despite potentially leading to stockouts, could also play a vital role in customer retention in the online retail sphere. However, they also recognize the study's limitations and suggest future research directions, including examining the impacts of different promotional activities and product types on consumer responses to stockouts, and the long-term effects of repeated stockouts.

Turkmen, B. (2022). Customer Segmentation With Machine Learning for Online Retail Industry. The European Journal of Social & Behavioural Sciences, Volume 31(Issue 2), 111-136. <https://doi.org/10.15405/ejsbs.316>

In this seminal work, Turkmen (2022) delves into the exploration and comparison of several customer segmentation methods, employing machine learning techniques on online retail data. The research underscores the critical role of customer segmentation in comprehending purchasing behavior and its consequential effects on pricing and demand forecasting in business. A variety of unsupervised machine learning clustering models, including K-means clustering, hierarchical clustering, Density-based Spatial Clustering of Applications with Noise (DBSCAN), and the conventional model based on recency, frequency, and monetary (RFM) values are scrutinized.

The author provides a comprehensive literature review on the evolution and applications of artificial intelligence, clustering models, and customer segmentation problems across industries. The work further discusses the adoption of artificial intelligence as a tool for learning in information systems, forecasting, prediction, and optimization across various industries, and future research directions. Finally, it provides a financial impact analysis of AI, asserting that revenues from the AI market worldwide could surpass $3,060 billion by 2024.

Turkmen, B. (2022). Customer Segmentation With Machine Learning for Online Retail Industry. The European Journal of Social & Behavioural Sciences, Volume 31(Issue 2), 111-136. <https://doi.org/10.15405/ejsbs.316>

This research paper explores the application of various clustering algorithms such as Mean-shift, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Agglomerative Clustering, and K-Means in conjunction with Recency, Frequency, and Monetary value (RFM) analysis on online retail transactions. The purpose of this multidimensional analysis is to identify distinct customer groups and understand their purchasing behaviors. The researchers found that these clustering algorithms, when combined with RFM analysis, can reveal valuable insights into customer segmentation based on their RFM scores, thus informing business strategies for customer retention and profit maximization.

The findings of this research are particularly relevant to our current project that involves applying K-means clustering to an online retail store. The paper demonstrates how these algorithms can identify high-value customer segments and potentially drive strategic decisions. Future research suggested by the authors includes the application of other clustering algorithms and the use of classification algorithms to predict the purchasing behaviors of new customers.

**References:**

Castéran, H., Meyer-Waarden, L., Reinartz, W. (2022). Modeling Customer Lifetime Value, Retention, and Churn. In: Homburg, C., Klarmann, M., Vomberg, A. (eds) Handbook of Market Research. Springer, Cham. <https://doi.org/10.1007/978-3-319-57413-4_21>

Yamada, H., &amp; Sato, T. (2021, January 31). An analysis of mechanism for customers' purchase amount and number of visits in Department Store. Journal of the Operations Research Society of Japan. <https://www.jstage.jst.go.jp/article/jorsj/64/1/64_12/_article/-char/ja/>

Peinkofer, S. T., Esper, T. L., Smith, R. J., & Williams, B. D. (2023). Assessing the Impact of Price Promotions on Consumer Response to Online Stockouts. Journal of Business Logistics, 46(4), 317-333.

Turkmen, B. (2022). Customer Segmentation With Machine Learning for Online Retail Industry. The European Journal of Social & Behavioural Sciences, Volume 31(Issue 2), 111-136. <https://doi.org/10.15405/ejsbs.316>

Y. Parikh and E. Abdelfattah, "Clustering Algorithms and RFM Analysis Performed on Retail Transactions," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2020, pp. 0506-0511, doi: 10.1109/UEMCON51285.2020.9298123.