Capstone Project

Customer Churn for ClientCo

We hereby certify that this report and the accompanying presentation is our own original work inits entirety, unless where indicated and referenced.

By Emilio Lopez, Eugenio Autrique, Guillermo Brun, Kangjie Yu, Mislav Ilijas

IE Master in Big Data | July 10th, 2024

Table of Contents

1. Introduction

ClientCo is an international construction and renovation products distributor. Its core business consists of buying products from suppliers and reselling them through a network of point distributions dedicated mostly to B2B Channels.

ClientCo CEO(58) has been leading the company for 3 years and is likely to remain at his position for the coming years. The company's chairman is at the beginning of her second 4 years term and is known to easily endorse the CEO´s strategy. The company is publicly held, with no clear shareholder majority.

ClientCo’s financial statements reveal notable trends. The balance sheet shows a steady rise in goodwill and inventories, suggesting a possible acquisition. This theory is supported by the significant growth in goodwill from 2006 to 2008, alongside smaller peaks in inventories, total assets, and total liabilities during the same period. Additionally, long-term debt also peaked in 2008, likely due to both the financial crisis and the acquisition.

The cash flow statement analysis shows that ClientCo did not pay taxes or interest in cash before 2006. Cash paid for taxes peaked in 2012, while cash paid for interest peaked in 2007. Surprisingly, dividends paid peaked in 2005, even though the highest income was recorded in 2011 and 2012. Typically, dividends are highest when income is highest, suggesting that the lack of significant dividends in 2011 and 2012 may indicate potential business expansion or reinvestment.

Regular increases in long-term debt could hinder ClientCo's growth. These increases may stem from the company's focus on acquisitions. While expansion is crucial, ClientCo should prioritise consolidation and strengthening its data science and data management departments. This approach will facilitate quicker long-term expansion through data-driven decision-making.

ClientCo faces several challenges, including a very weak online presence, an increasing debt portfolio, and significant environmental threats. The SWOT analysis reveals that while ClientCo has a strong presence in various markets and a wide customer base supported by demand prediction models and targeted ads, its limited liquidity and financial constraints are major weaknesses.

Opportunities lie in the early phases of data-driven development roadmaps that could facilitate adoption processes for companies. Addressing the weak online presence and optimising liquidity through demand prediction are critical areas of focus to mitigate financial risks and enhance overall market performance.

The main objective of this project is to provide some data-driven insights in order to help ClientCo take action. We want to help them tackle their problems in the most efficient and objective way.

2. Data Sources

Financial Dataset
The financial dataset contains the income statement, balance sheet, and cash flow statement from 2000 to 2019.

Transactional Dataset
The transactional Dataset contains 2 years of transaction data. Within this dataset, we can find the following columns:

- *date_order.* Date on which the order happened.
- *date_invoice:* date of the invoice
- *product_id:* unique product ID key
- *client_id:* unique client ID key
- *sales_net:* net amount purchased in currency
- *quantity:* quantity of items purchase

This dataset is the dataset on which we based most of our Exploratory Data Analysis, shown in *Section 4*.

Client Dataset Creation
Based on the Transactional Data mentioned above, we performed some feature engineering steps that resulted in the following Client Dataset:

- *Recency:* Last *date_order* from the customer.
- *Frequency:* Frequency of the customer
- *Monetary:* the total amount of sales_net of the customers
- *Average days between purchases:* Average time between the customer's purchase dates.
- *Variance of purchase intervals:* Variation in time intervals between consecutive purchases.
- *Most common order channel:* The sales channel used most frequently by the customer.
- *Number of returns:* Total instances of the customer returning products.
- *Most purchased product:* The product most frequently bought by the customer.
- *Number of unique products purchased:* Count of different products purchased by the customer.
- *Most common branch:* Store location where the customer shops most often.
- *Total quantity purchased:* Sum of all items purchased by the customer.
- *First purchase date:* Date of the customer's initial purchase.
- *CLV:* Customer lifecycle value.

The dataset would be used for analyzing the purchase behavior of each client and supporting marketing strategies.These marketing strategies will be based on customer clustering.

3. Methodology

The primary objective of this project is to explain customer churn observed in ClientCo and provide insights and recommendations about improving customer retention and maximizing sales. Churn, or the loss of customers, poses a significant threat to revenue and growth. For ClientCo, understanding and mitigating churn is crucial for maintaining a stable and profitable business. The specific goals of this project are to identify the underlying causes of customer churn, develop predictive models to anticipate churn and implement effective retention strategies to reduce churn rates.

To address the problem of churn, we adopted a structured approach that involves several key steps:

1. **Exploratory Data Analysis (EDA)**
    a. Conducted a thorough EDA to understand the distribution and trends within the transaction data. This included analysing sales over time, identifying seasonal patterns, and detecting anomalies.
    b. Utilised statistical methods and visualisations to uncover insights and guide subsequent analyses.
2. **Feature Engineering**
    a. Developed new features from the raw transaction data to enhance the predictive power of our models. This included creating variables such as recency, frequency, and monetary value (RFM), as well as calculating the Customer Lifetime Value (CLV) for each client.
    b. Engineered additional features to capture customer behavior and engagement metrics, such as purchase frequency and order channel preferences.
3. **Customer Segmentation**
    a. Implemented RFM analysis to segment customers based on their purchasing behaviour.
    b. Segmented customers by sales channels (e.g., phone, online, in-store) to understand the churn dynamics within each channel.
4. **RFM and CLV Calculation**
    a. Calculated the RFM scores for each customer to quantify their recency, frequency, and monetary value. This segmentation helped in tailoring specific retention strategies for different customer groups.
    b. Estimated the average Customer Lifetime Value (CLV) to understand the long-term revenue potential of retaining customers.
5. **Churn Prediction Model Development**
    a. Developed a churn prediction model using XGBoostClassifier. The model was trained on historical transaction data and engineered features.
    b. Validated the model using accuracy metrics to ensure its reliability in predicting churn.
6. **Demand Prediction Model Development**
    a. Created a demand prediction model (XGBoostRegressor) to forecast future sales and inventory needs. This model utilised historical sales data and identified patterns to predict demand accurately.
    b. Validated the demand prediction model using the Mean Absolute Percentage Error (MAPE) to measure its forecasting accuracy.

7. **Model Validation and Refinement**
    a. Continuously monitored the performance of both the churn and demand prediction models. Regular validation and refinement were conducted to improve accuracy and adapt to changing customer behaviours and market conditions.
8. **Business Impact Analysis**
    a. Quantified the business impact of applying the predictive models and identified patterns. This involved estimating potential revenue losses due to churn and the financial benefits of retaining customers.
    b. Calculated the impact of improved inventory management and demand forecasting on operational efficiency and customer satisfaction.
9. **Actionable Insights and Recommendations**
    a. Provided actionable insights based on the findings from the predictive models and pattern analysis. Recommendations included targeted retention strategies, personalized marketing campaigns, and improvements in customer service.
    b. Suggested specific actions for each customer segment and sales channel to maximise retention and revenue.

4. Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) phase was critical in understanding the dataset, identifying potential issues, and uncovering key insights that would guide the subsequent analyses and model development. Here are the detailed steps and findings from the EDA conducted for ClientCo:

Note: All the graphs and images showed in the content were created by the authors of this documents

Review of Missing Values and Adjustment of Data Types

The initial step in the EDA involved a thorough review of the dataset to identify any missing values and ensure that the data types were appropriate for analysis.

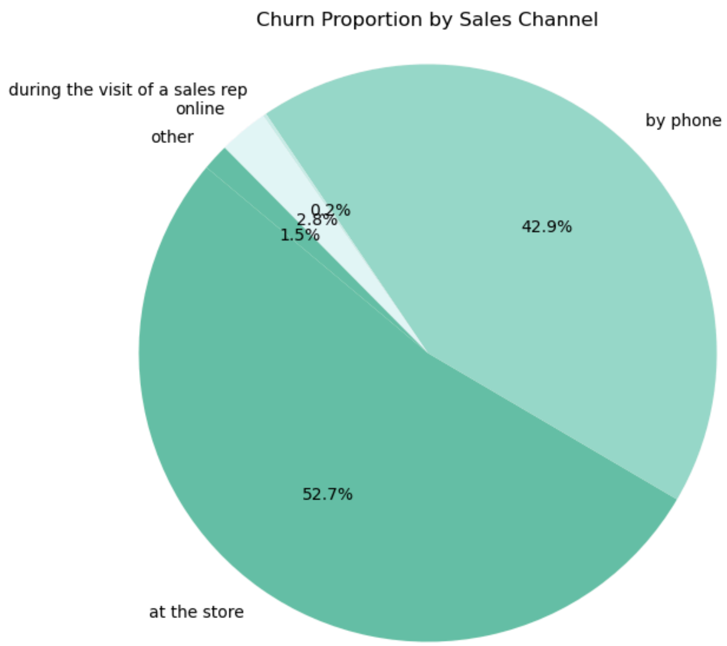
Missing Values: We assessed each column and determined that there were not missing values in the transactions dataset

Data Types: We adjusted the data types of key columns, such as converting date columns (date_order and date_invoice) to datetime format and ensuring that identifiers (product_id, client_id, branch_id) were treated as strings. This ensured accurate calculations and efficient processing during the analysis.

Data Analysis for Problem Detection

To detect potential problems and set the stage for further analysis, we conducted several key analyses:

- **Churn Rate in the Last 6 Months**: We calculated the churn rate by identifying customers who had not made a purchase in the last six months. This involved defining a threshold for churn and applying it to the transaction data. We found that 30% of ClientCo's customers had not made a purchase in the past six months, indicating a significant churn issue.
- **Proportion of Churn Rate by Order Channel**: We analyzed the churn rate across different sales channels (phone, online, in-store, and others). This revealed that 53% of churned customers were from in-store sales, 43% from phone sales, and the remainder from other channels. Understanding the distribution of churn across channels helped identify areas requiring targeted retention efforts.

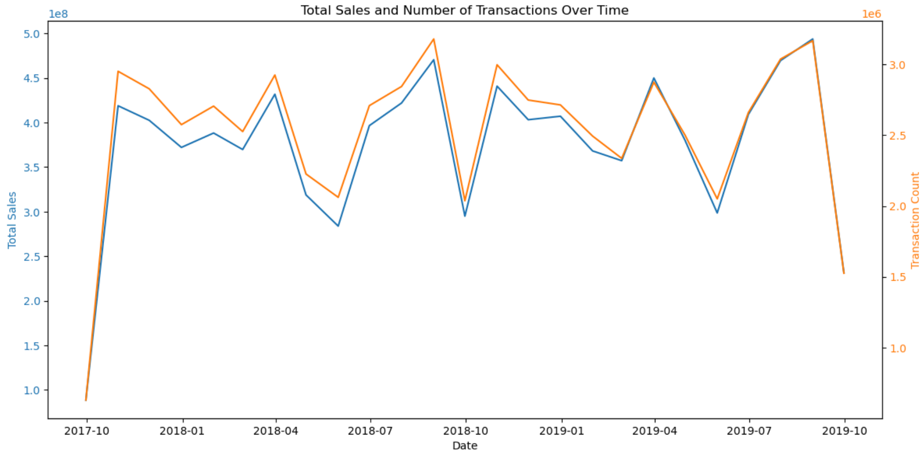


- **Number of Customers that Churned in the Last 6 Months**: We quantified the exact number of customers who had churned, which amounted to 50,772 customers. This provided a clear picture of the scale of the churn issue.
- **Average CLV at a Company Level**: We calculated the average Customer Lifetime Value (CLV) to understand the revenue potential of retaining customers. The average CLV was found to be $109,826.08, highlighting the significant financial impact of each churned customer.
- **Financial Impact**: By combining the number of churned customers with the average CLV, we estimated the financial impact of churn. The potential loss due to churn was calculated to be approximately $5.5 billion, underscoring the urgency to address the churn issue.

Analysis of Total Sales and Transactions Over Time

We conducted a temporal analysis of total sales and the number of transactions to identify trends, seasonality, and anomalies. This involved plotting sales and transaction volumes over time to visually inspect patterns. Key findings included:

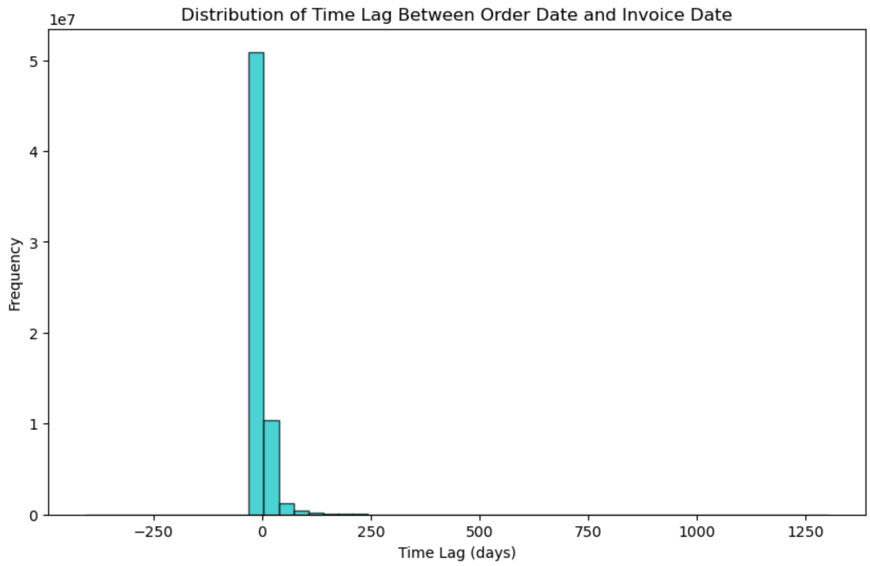
- **Anomalies**: Significant drops in sales during June 2018 and 2019 were identified, warranting further investigation to understand the underlying causes.



Analysis of the Time Lag Between Order and Invoice Date

We analyzed the time lag between the order date and invoice date to assess the efficiency of the invoicing process.

- **Findings**: We observed lots of negative values in the time lag (date invoice - date order) and that insight took us to use date order for further analysis and predictive modelling



Analysis of Returns

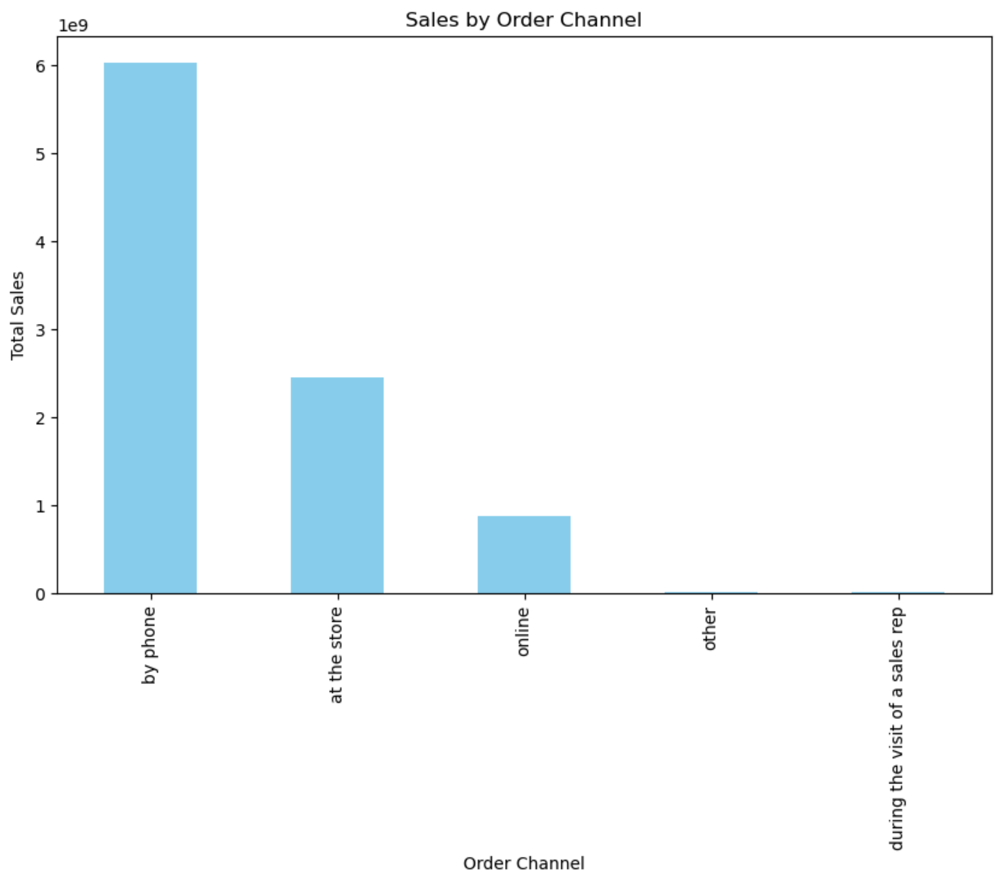
We examined the return transactions to identify the most frequently returned products and the reasons for returns.

- **Findings**: Certain products had a higher return rate, indicating potential quality or suitability issues. Understanding these patterns helps in addressing product-related concerns and improving customer satisfaction.
- Top 5 of most returned products (Product IDs)
    1. 539670
    2. 2681878
    3. 1492103
    4. 463156
    5. 932375

Analysis of Channels Driving Sales

We analyzed the effectiveness of different sales channels in driving revenue. This involved comparing sales volumes and customer acquisition rates across channels.

- **Findings**: In-store and phone sales were the dominant channels, but online sales showed significant potential for growth. Tailoring strategies to enhance online sales could diversify revenue streams and reduce dependence on traditional channels.



Analysis of the Branches

We conducted a comparative analysis of sales and transactions performance across different branches to identify high-performing and underperforming locations.

- **Findings**: Certain branches consistently outperformed others, suggesting the need for best practice sharing and performance optimization across all locations. Underperforming branches were flagged for further investigation and targeted support.
- We were able to detect the IDs of the following relevant branches
    - Branch with highest sales: 3318 ($122620211.81118721)
    - Branch with lowest sales: 524 ($-2.842170943040401e-14)
    - Branch with highest transaction count: 3318 (643183 transactions)
    - Branch with lowest transaction count: 2925 (1 transactions)

5. Model Developments

In this part, we developed 2 models, one for churn prediction and the other for demand forecasting. In each of these models, we conducted feature engineering, model selection and development and model evaluation.

Churn Prediction

To decrease the churn rate of the customers and increase customer lifetime value, we developed a churn model based on the window dataset that we generated for customer behaviour analysis.

a. **Feature Engineering (churn defining)**

To develop a churn mode, we need to define churn to serve as the target variable and we explored various methods to define churn:

- ***Confidence Interval***

In this approach, we consider whether a customer's purchase behavior (i.e., the interval between purchases) follows a normal distribution. We then assess if the recency (time since the last purchase) falls within the 95% confidence interval. If it lies in the right tail, we infer that the event's probability is below 2.5%, indicating the customer has likely churned. However, we found that most customers' purchase behaviors do not follow a normal distribution, rendering this method ineffective.

- ***K-means***

By employing K-means clustering, we aimed to classify customers and identify clusters with common churn characteristics. We identified a cluster with high recency but low frequency and lifespan, suggesting these customers likely made a single purchase and then churned. While this approach successfully identified these customers, it failed to define churn for high-CLV (Customer Lifetime Value) customers.

- ***K-means + Confidence Interval***

  Given the limitations, we analyzed whether high-value customers' purchase behaviors followed a normal distribution to define churn more accurately. However, these behaviors also did not conform to a normal distribution, making this method infeasible.

- ***K-means + Average Purchase Interval***

  To define churn for high-value customer clusters, we adopted a new method: if a customer's recency is three times greater than their average purchase interval, we classify them as churned.

- ***Windows***

  In this approach, we defined a two-month window based on the company's sales drop-off period. We analysed customer purchase behaviour in these two-month intervals. If a customer made a purchase within a window, it was marked as 1; otherwise, it was 0. This resulted in a new dataset containing 13 windows (0-12), each recording whether a customer made a purchase. We used purchase behaviour within these windows to define churn.

  In conclusion, for this model training, we used a window dataset that includes data on whether each customer made a purchase within each window (1 if purchased, 0 if not).

b. Model Selection & Development

We utilised the XGBoostClassifier for training. The training started with the first 3 windows to train and predict the fourth window, then used the first 4 windows to train and predict the fifth window, and so on.

In each training iteration, we fine-tuned the model to maintain optimal performance. We recorded the F1-score, AUC-ROC, and accuracy at each step to assess the impact of new data on the model.

c. Model evaluation

After fine-tuning, the model achieved an F1-score of 0.84, an AUC-ROC of 0.89, and an accuracy of 0.82. The F1-score improved from 0.78 to 0.84 but dropped to 0.75 in the final window, primarily due to the purchase ratio decreasing from 0.56 to 0.35 during that period. The AUC-ROC improved from 0.85 to 0.89.

Iteration	Window	F1_score	ROC_AUC	Accuracy
0	3	0.79	0.85	0.78
1	4	0.79	0.85	0.79
2	5	0.80	0.85	0.78
3	6	0.80	0.86	0.79

4	7	0.80	0.86	0.79
5	8	0.79	0.87	0.80
6	9	0.80	0.88	0.80
7	10	0.80	0.88	0.81
8	11	0.84	0.89	0.81
9	12	0.75	0.89	0.83

d. Other Approaches

For this model, we used a dataset derived from a 30-day transaction dataset, containing client data such as RFM, CLV, variance of purchase intervals, and average days between purchases, to predict whether a customer would churn after 30 days.

Here, we defined churn using the K-means + Average Purchase Interval method. By comparing churn labels generated from the 30-day dataset with those from the previous 30-day dataset, and incorporating other data points from 30 days prior (e.g., CLV, RFM), we could create a model to predict customer churn status.

However, this model did not account for customer behaviour evolution, such as customers buying more earlier and less later. Despite achieving a high F1-score of 0.86, the model's interpretability and significance were limited, so it was not adopted.

Demand Prediction

The main objective of the Demand Prediction model is to support Supply Chain Operations. With this model, we will be able to help ClientCo prevent some of ClientCo supply chain main problems: i) Stockouts, ii) Reducing Delivery.

To develop this model, our starting point has been our ***Transactional Dataset.***

a. Feature engineering

To start with, we created one additional column for each date feature of the *date_order* field. We extracted year, month, week, day of the week, and a combined date week field. Moreover, we also optimised some of the other fields, storing them as numpy arrays. This helped us to improve Space Efficiency and Processing Speed.

The second challenge we encountered was how to deal with all of the information contained in the *product_id, client_id, branch_id* and *order_channel* fields. These fields have a huge amount of unique values, which hinders simple encoding methods such as *One-Hot Encoding* or *Label Encoding.* So we decided to use *Target Encoding* with the 4 fields mentioned above. Target encoding replaces categorical features with values linked to the target variable, leveraging its information to potentially improve model performance.

For this task, our target variable was ***quantity*** and the encoded variables were ***product_id, client_id, branch_id*** **and** ***order_channel.***

b. Model Selection & Development

After having prepared our data, we selected a ***XGBoostRegressor*** model from python's library ***xgboost.***

The explanatory variables of this model were: i) *product_id* (target encoded), ii) *client_id* (target encoded), iii) *branch_id* (target encoded), iv) *order_channel* (target encoded) and v) *date_order_month* (month of the date order).

The target variable of the model was ***quantity***.

Due to the size of the dataset provided, we attempted to run randomized grid search cross validation to obtain relevant improvement, but due to lack of computational power, we couldn't successfully run it.

c. Model Evaluation

We used MAPE to evaluate the model.

Model Evaluation	
MAE	49.826
MAPE	1.265

MAE and MAPE both measure average prediction error, but MAPE expresses it as a percentage, making it easier to compare models for different data scales.

Customer Segmentation

For customer segmentation, we explored two different approaches. The first approach involved classification based on the churn model's predictions combined with client dataset data. The second approach used k-means clustering on the client dataset, which resulted in higher entropy. Therefore, we ultimately chose the first clustering method for analysis and recommendations.

a. Churn Prediction-Based Clustering

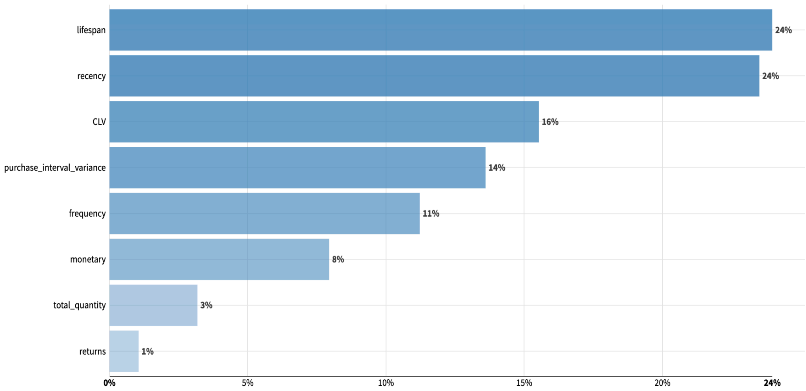
We created two clusters for analysis:

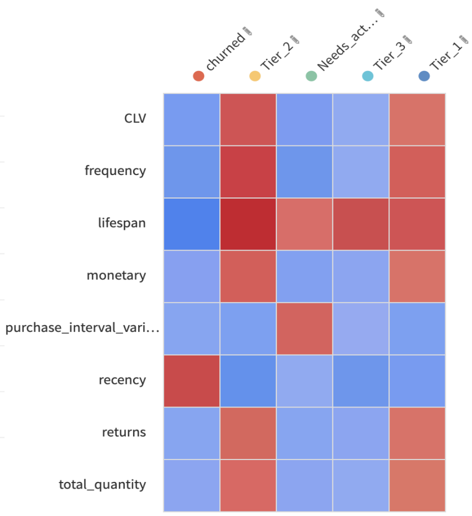
- **Churn Prediction + Most Common Order Channel**:
  i. We combined churn predictions with the most common order channels to form 10 clusters (5 order channels and 2 predictions).
  ii. This clustering allows us to tailor promotions to churned customers based on their preferred purchase channels.
- **Churn Prediction + Customer Lifetime Value (CLV)**:
  iii. We combined churn predictions with customer lifetime value to form 10 clusters (5 CLV tiers and 2 predictions).
  iv. This clustering helps us target and engage valuable customers more effectively.

b. K-Means Clustering

In this approach, we used Dataiku to perform k-means clustering on the client dataset. We determined that 5 to 6 clusters were optimal using the elbow method. Additionally, we applied PCA to reduce the impact of multicollinearity. The resulting clusters were categorised as follows:

- Three clusters of high-value customers
- One cluster of customers who churned after only one or two purchases
- One cluster of[1] customers with irregular purchasing patterns





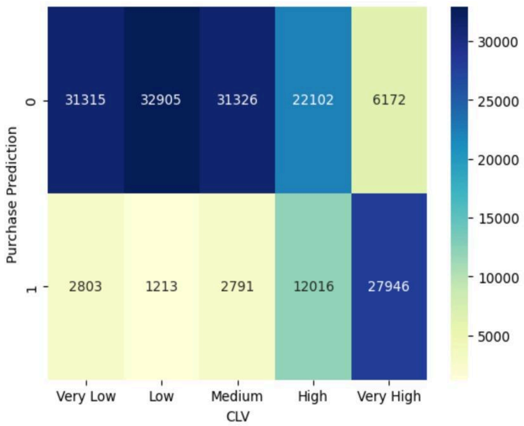
[1] Feature importance of clustering and heatmap of each cluster

6. Insights and Recommendations

Insights Obtained

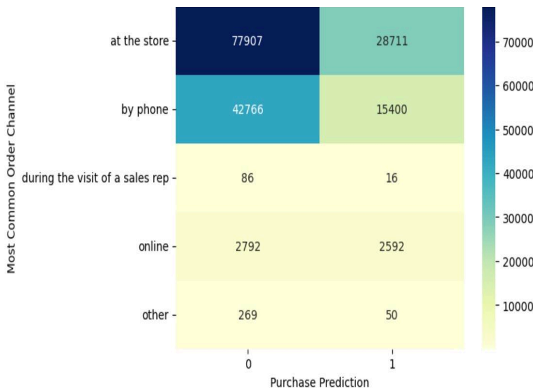
To start our search for insights we first decided to analyse all the minutes provided by ClientCo. We found three relevant points that help us understand in a better way the problem and what insights to persuade. The Data Science Team highlighted the need for more detailed data to refine customer segmentation, based on this we realized our customer analysis. From Sales Management, we learned that personalized interactions, such as direct phone calls and customized recommendations, resonate much better with customers than broad campaigns. Lastly, the Supply Chain Management team emphasized the importance of accurate forecasting for effective inventory control, for this we decided to realized our demand prediction model.

Our analysis has provided several critical insights that are essential for understanding customer behaviour and improving ClientCo's operations. During our EDA, we noticed key patterns and trends in how customers make their purchases. Firstly, we found that sales tend to peak during certain seasons, showing a strong seasonal trend in customer buying habits. Specifically, there were noticeable drops in sales during June 2018 and June 2019. These trends can help ClientCo better plan its inventory and marketing efforts to make the most of high-demand periods. Additionally, we encountered lots of negative values in the time lag between the date of invoice and the date of order. This anomaly led us to decide to use the order date for further analysis to ensure accuracy.



Then we realized a heatmap where customers with 'Very High' CLV show consistent engagement, suggesting a strong connection with the brand. However, there's a noticeable gap in the 'High' CLV segment where potential purchases aren't being realized. This gap represents a great opportunity for ClientCo to improve its engagement strategies and boost conversions. By designing marketing efforts that meet the unique needs and preferences of these segments, we can significantly improve customer retention and satisfaction.

Our second heatmap highlights a strong preference for traditional sales channels like in-store and phone transactions across all Customer Lifetime Value (CLV) categories. However, there's a notable trend of increasing online engagement among higher CLV customers.. By improving ClientCo's online platforms, we can not only better serve our high-value customers but also attract a broader audience that prioritizes convenience and speed in their shopping experience[2].



[2] Heatmap CLV nad heatmap with most common order chanel

Our churn prediction model has pinpointed critical factors that drive customer churn, specifically the recency of a customer's last purchase and their total spending. Customers who have been inactive recently or spend less are at a higher risk of discontinuing their relationship with ClientCo. Early identification of these customers allows us to deploy targeted retention strategies like personalized offers and engagement campaigns, effectively reducing churn rates and enhancing loyalty.

Similarly, our demand prediction model delivers precise forecasts that are vital for inventory management. By accurately predicting high-demand products, we ensure optimal stock levels, prevent shortages, and minimize surplus inventory costs. These insights are crucial for developing strategies that not only reduce churn and boost customer lifetime value but also streamline inventory management, driving increased sales and profitability for ClientCo.

Potential Solutions

To tackle the challenges of customer churn and enhance business performance at ClientCo, we've tailored our strategy to focus on practical, impactful solutions. Drawing on insights from our predictive models and team feedback, we've crafted two main approaches, each carefully considered for its feasibility, data requirements, and potential business impact.

a. Chosen Solution: Enhanced Customer Engagement

Our selected strategy leverages deep insights from our models to improve the way we interact with customers through personalized marketing and loyalty programs. This approach is designed to strengthen relationships with our customers, significantly reduce churn rates, and enhance customer lifetime value.

The sales management team's feedback was instrumental in shaping this strategy. They noted that personalized interactions, such as direct phone calls and tailored recommendations, have a much more positive impact compared to broader campaigns. Based on this, we plan to identify at-risk customers—those who have not made recent purchases or tend to spend less—and engage them with customized marketing campaigns.

We are also introducing a tiered loyalty program that encourages repeat business and fosters long-term loyalty by offering exclusive rewards and benefits tailored to customer behaviors and preferences, as revealed by our data analysis. This program will include perks like early access to new products and special discounts.

b. Non-Selected Alternative: Advanced Predictive Analytics and Real-time Optimization

While we recognized the potential benefits of implementing cutting-edge predictive analytics and real-time optimization, this solution was not chosen due to its complexity and extensive data requirements. This approach would have involved sophisticated machine learning models and real-time data integration to dynamically manage inventory and customer interactions. Despite its promise for optimizing operations, the substantial investment in resources and the need for significant technological upgrades made it less feasible in the short term. Instead, we chose to leverage existing capabilities and insights to make a more immediate impact.

Our decision to implement the Enhanced Customer Engagement solution is based on its alignment with ClientCo's current capabilities and immediate goals. It offers a practical and impactful way to reduce churn and enhance customer value, using the data and resources we already have. This strategy sets the stage for future advancements in analytics and customer engagement, building a strong foundation for sustained growth.

Business Impact

Business impact of the Chosen Solution:

- Reduced Churn Rates
  - By proactively addressing the needs and preferences of customers at risk of churn, we expect to see a significant reduction in churn rates.
- Increased Customer Lifetime Value
  - Our personalized engagement and loyalty initiatives will deepen relationships with customers, thereby increasing their overall value to ClientCo.
- Boosted Sales and Revenue
  - Targeted marketing and loyalty incentives are designed to increase purchase frequency and create upselling opportunities, directly enhancing sales.
- Enhanced Brand Loyalty and Market Positioning
  - Implementing personalized engagement strategies not only meets immediate sales targets but also strengthens ClientCo's brand loyalty and competitive edge in the market. This strategic positioning helps in expanding our market share and solidifying our presence in the industry.

Financial Impact Calculation of Churn Prediction Model

To quantify the financial impact of implementing the churn prediction model, we conducted a detailed analysis using the provided customer base and model performance metrics. Below are the steps and calculations performed to estimate the potential revenue saved by reducing churn through targeted interventions based on our model's predictions.

Customer Base and Purchase Behavior

- **Total Customer Base**: 170,589
- **Customers Who Did Not Buy in the Last Two Months**: 46.05% of the customer base
- **Number of Customers Who Did Not Buy**:
  - Number of customers who did not buy = 170,589×0.4605 = 78,565

Model Performance Metrics

- **F1-Score of the Model**: 0.75
- **Average Customer Lifetime Value (CLV)**: $109,826.08

Calculation of True Positives (TP)

Using the F1-score, we can estimate the number of customers correctly predicted as likely to churn (True Positives):

True Positives (TP) = F1-Score × Number of customers who did not buy

TP= 0.75 × 78,565 = 58,924

Financial Impact Estimation

The total revenue saved by correctly identifying and retaining these customers can be calculated as:

Total Revenue Saved = TP × Average CLV

Total Revenue Saved = 58,924 × 109,826.08 = 6,471,425,500.32

Development

In the development phase of our project aimed at reducing customer churn and enhancing engagement at ClientCo. We have organized all of our code and project files, which are readily accessible in a dedicated GitHub repository.

- Link to GitHub repository: https://github.com/lohemilio/BCG-Capstone-Project

For the use-case implementation, our approach is tailored to address specific use-cases identified from team insights and data analysis, focusing on re-engaging at-risk customers those less active or spending less.

Quality Standards and Practices:
- Our commitment to high coding standards meant that all scripts are well-commented and adhere to best practices for readability and maintenance.
- We conducted rigorous testing to validate our predictive models, ensuring reliability and effectiveness. Specifically, our churn prediction model achieved an F1-score of 0.84, demonstrating a strong balance between precision and recall. The model's AUC-ROC of 0.89 indicates a high degree of separability, meaning it can effectively distinguish between customers who will churn and those who will not. An overall accuracy of 82% across testing phases reassures us of the model's robustness in real-world scenarios.
- Additionally, the demand prediction model showed a Mean Absolute Percentage Error (MAPE) of 1.265%, reflecting high forecast accuracy and minimal deviation from actual values, which is crucial for effective inventory management.

Our development efforts were thoughtfully designed to align with ClientCo's strategic goals, utilizing the right technology tools and analytical methods. By taking care of the quality standards and ensuring data integrity, we've established a solid foundation to support future expansions and enhancements.

Recommendations

Based on the main challenges that ClientCo is facing, we propose the following recommendations to help the company overcome these issues. Our suggestions will not only include how to leverage the models we created but also provide theoretical insights and future maintenance advice for sustainable development.

1. **Enhancing Online Channel Presence**
   Given ClientCo's relatively stronger performance with direct phone marketing, we can target customers who are likely to make purchases and primarily use the phone as their order channel. By offering incentives such as discounts, we can encourage these customers to transition from phone to online purchases.

   It’s also important to develop the e-commerce platform focusing on user-friendly interfaces, seamless checkout processes, and personalized recommendations.

2. **Addressing Customer Churn**
   To mitigate customer churn, we recommend reactivating customers with high Customer Lifetime Value (CLV) who are not expected to make a purchase in the next two months. This can be achieved through personalised outreach such as emails, phone calls, or even in-person visits to bring these high-value customers back into our sales funnel.

3. **Demand Forecasting**

   Utilise the demand prediction model to accurately forecast inventory needs, reducing stockouts and costs related to excess inventory. Ensure optimal stock levels by aligning inventory management with predicted demand patterns.

4. **Branch Performance Improvement**

   Share best practices from high-performing branches with underperforming ones. Provide targeted support and resources to improve their sales and transaction performance.

5. **Monitoring for data drift**

   As the models are used on real data it is important to monitor for data drift. The sample used to train the model might not be fully representative of real world customer populations, hence the models should be retrained often with new data. If retraining is too expensive, data drift detection techniques like Kolmogorov-Smirnov test for continuous features and Chi-squared test for categorical features can be used to determine the optimal retraining period.