Modelling Churn
& Maximizing Customer Life Time Value

MBD Capstone - July 2024

FOR DISCUSSION ONLY

FOR DISCUSSION ONLY

Disclaimer

The following pages contain material provided to the Boston Consulting Group X Team (“BCG”) of the consulting project by the IE School of Science and Technology Team (“Team”) in connection with the project involving churn modelling for their clients' B2B transactions dataset.

The accompanying material was compiled or prepared on a confidential basis solely for the use of BCG in their consulting project and not with a view toward public disclosure under any securities laws or otherwise. The information contained in the accompanying material was obtained from BCG. Any estimates and projections for BCG contained herein have been prepared by the Team. Such estimates and projections, involve numerous and significant subjective determinations, and there is no assurance that such estimates and projections will be realized. The Team does not take responsibility for such estimates and projections, or the basis on which they were prepared. No representation or warranty, express or implied, is made as to the accuracy or completeness of such information and nothing contained herein is, or shall be relied upon as, a representation, whether as to the past, the present or the future. In preparing the accompanying material, the Team assumed and relied, without independent verification, upon the accuracy and completeness of all financial and other information and upon the assurances of BCG that they are not aware of any relevant information that has been omitted or that remains undisclosed to the Team.

The accompanying material was not prepared for use by readers not as familiar with the clients and, accordingly, neither the Team nor their respective legal or analytical or financial advisors take any responsibility for the accompanying material if used by persons other than BCG. The accompanying material is necessarily based upon information available and circumstances existing and disclosed, to the Team as of the date of the accompanying material. The Team does not have any obligation to update or otherwise revise the accompanying material. Nothing contained herein shall be construed as legal, regulatory, tax, accounting or strategic advice.

Agenda for Today

FOR DISCUSSION ONLY

Executive Summary & Team Recommendations

Dynamic Churn Modelling & SQL DBs

Our Churn Modelling is unique for every client and **can trigger actions based on its output probability such as sending promotions or batch discounts.**

To enhance digitization and reduce branch dependency and redundancy, **migrate to a centralized SQL-based DB**. A preliminary forecast **signals an increase of 39% in EBIT**

Increase Revenue by enabling Cross-Selling

The development of a **proprietary website** is an opportunity for integrating all newly gathered data. Consider **Partnering with third-party vendors**.

Our Analysis demonstrates that Sales could increase by **2.4% were they to implement cross-selling**.

Prioritize Retention with Monetary Risk

With a weighted RFM and calculated CLV for each of the segments and Churn Probability, we can calculate the monetary risk associated with leaving.

Top customers with a high probability of churning, are expected to have a mean CLV of 33.9M USD. ClientCo. can now prioritize.

Augment Stock Forecasting Capabilities

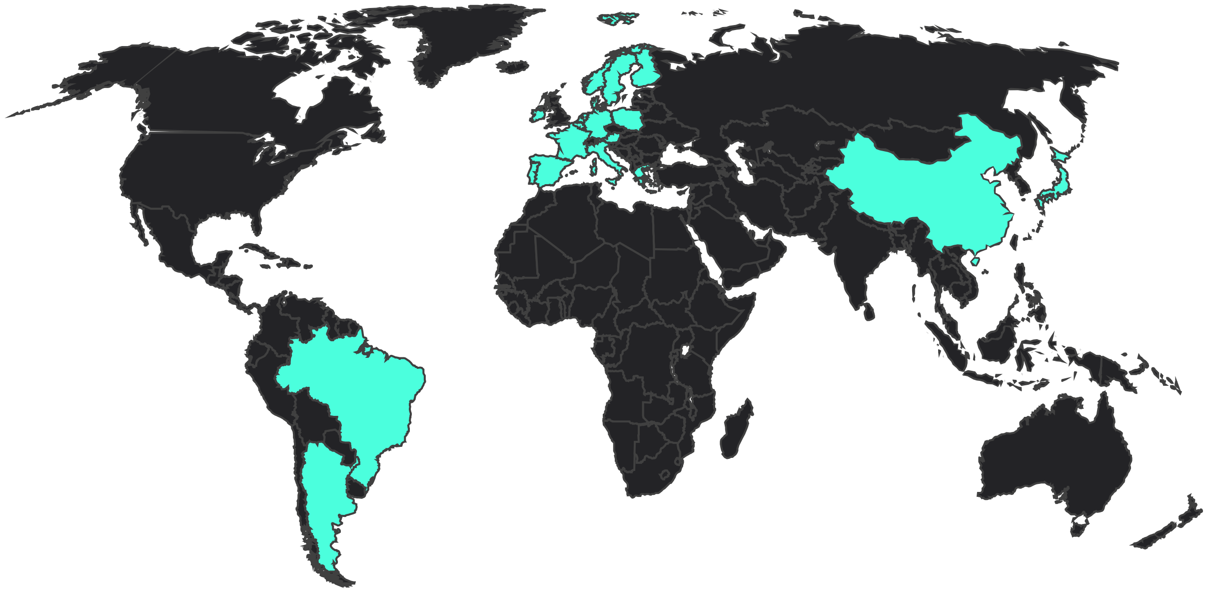
Time Series Decomposition indicates SARIMA models can be used to estimate products in a weekly manner with a high degree of accuracy.

The Team suggests to reduce inventory and **implement this forecast for the top 20K products** (3.9% of total products) **that make 80% of the revenue.**

FOR DISCUSSION ONLY

ClientCo is an international construction and renovation products B2B retailer & distributor…

Geographical Presence



…with over 31,300 employees worldwide

Key Figures (in 2019)

~15.5 B$
in Revenues

~675 M$
in EBIT

450 K
Active Customers

> 1000
Unique branches

Present in more than **15** countries

Source: BCG Report, ClientCo. Documents

… and chose our Team to help them to tackle churn, increase retention and CLTV

FOR DISCUSSION ONLY

Conducted interviews signal a lack of digital implementations across three key areas of the organization: Supply Chain, Sales and Marketing Departments.

		Who?	Role	Described Pain Points	Strengths
Other	Upper Mngmt.	Fernando Suarez	CDO	Initial Lack of Direction, Failed Pilots, Adoption Challenge, Deployment & Program	Adoption by Sales Team, Learning from failure
ClientCo	Upper Mngmt.	Charlotte Davelle	CDO	**Digital Resistance** from Exec. Committee, **Lack of Digital Implement**. by CEO, **Quick Wins** value	Extensive Tenure, CEO discussions, Short-term Feasibility
ClientCo	Upper Mngmt.	Laurent Stacks	Head of Strategy	**Digital Resistance** from Exec. Committee, **Poor Marketing Data Quality**, Digital lack of Experience	Infrastructure Investment, Strong Operational Efficiency, Strategic Digitalization Mindset
ClientCo	Upper Mngmt.	Daniel Alfredo	Head of Sales	Struggling to **measure churn**, **High Client Visit Cost, Online Sales** Efforts are underdeveloped	Strong Customer Relationships, Motivation and Effective Communication
ClientCo	Upper Mngmt.	Laure McKenzie	Commercial Director	**Mngmt Routines** schedule, Resistance by **Traditional Mkting team**, Scaling and **Measurement**	Weekly touchpoints, Churn Prevention Focus and Digitalization of Sales tools
ClientCo	Upper Mngmt.	Tom Hammer	Supply Chain Director	**Manual** Inventory Process, Out of **Stock** and **Delivery Issues**, Need for **digitalization**	Improved Inventory Accuracy, Barcode Technology, Optimized Package Process
ClientCo	Middle Mngmt.	Henri Dubois	Sales Manager	**Note-taking** instead of CRM, **Qualitative client's** interactions standardization, **Data skepticism**	Strong Customer Relationships, Client Visits and Phone calls, Interest in Cross-Selling Tools
ClientCo	Middle Mngmt.	Richard Martin	Data Manager	System **Convergence** Difficulties, **unstructured** data programs, **data similarity** in diff. branches	Improved IT Reliability, Experience in Data Projects, Commitment for Data Structure

Source: ClientCo Meeting Minutes

FOR DISCUSSION ONLY

Combining the concerns of ClientCo with the insights collected from their own employees we managed to have a more holistic view of the AS-IS situation...

Identified Pain Points

AS – IS

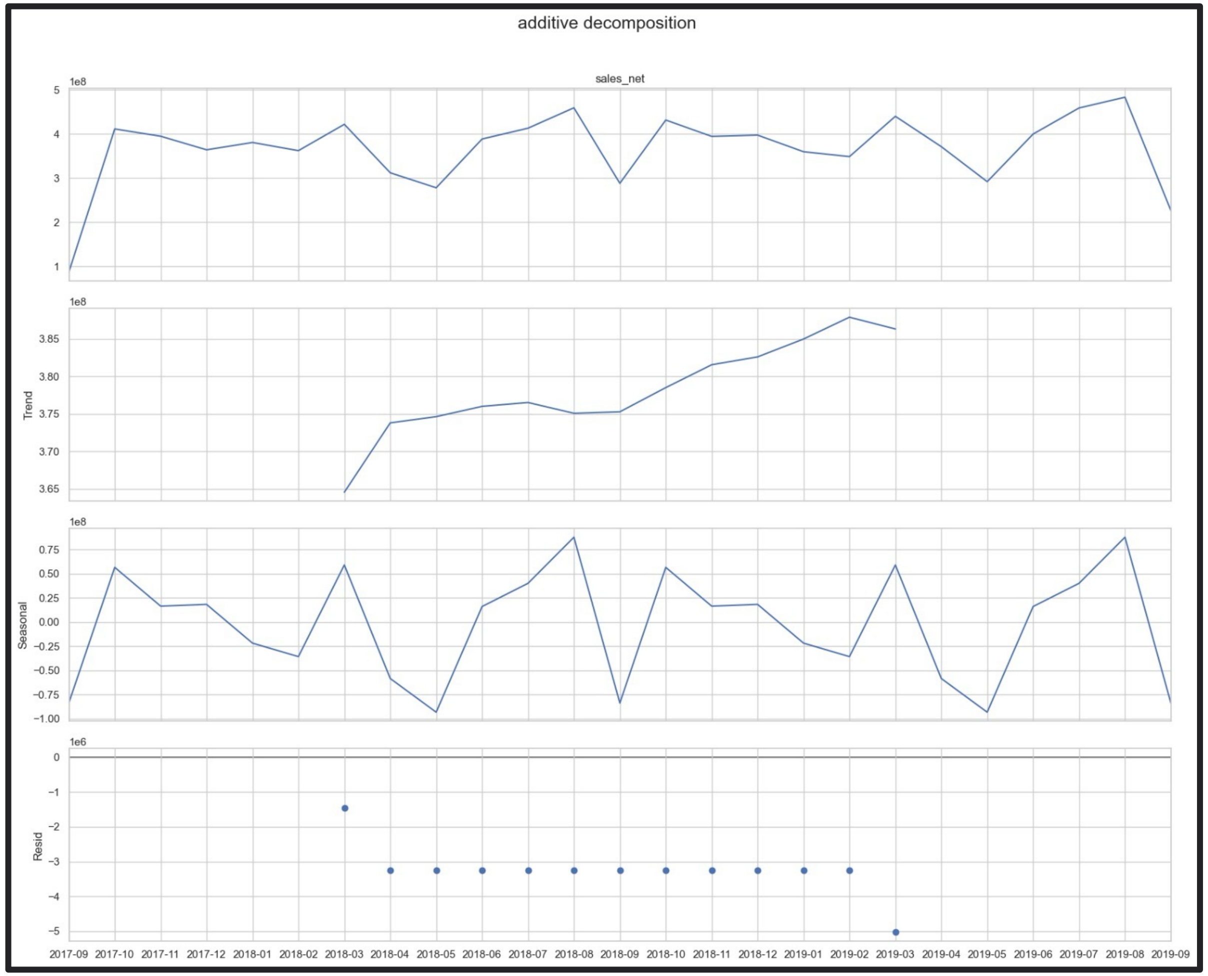
- Very weak online presence. Unacceptable levels of digitization
- Performance in decline in Western Europe and Southeast Asia
- Increasing Debt portfolio and financial costs
- Poor quality of information for the marketing team. No qualitative data.
- Little money to finance big initiatives
- Supply Chain disorders

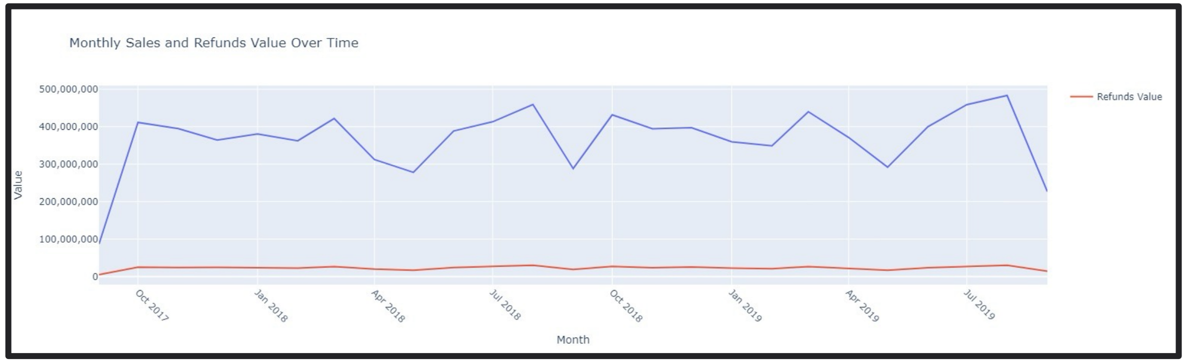
Proposed Solutions

TO - BE

- Improve the digitization of the company
- Creating a standardized inventory management system across branches (SQL)
- Build a churn predictive model to increase retention and drive revenue
- Gather qualitative data on customer to generate tailored promotions and campaigns to increase online presence

Sales and Refunds have seasonality patterns with May and September having the lowest values in the time series. This suggest that forecasting stock demand is feasible.





Seasonal Patterns repeat when decomposing the original time series. Refunds trends are highly correlated with sales trends as well.

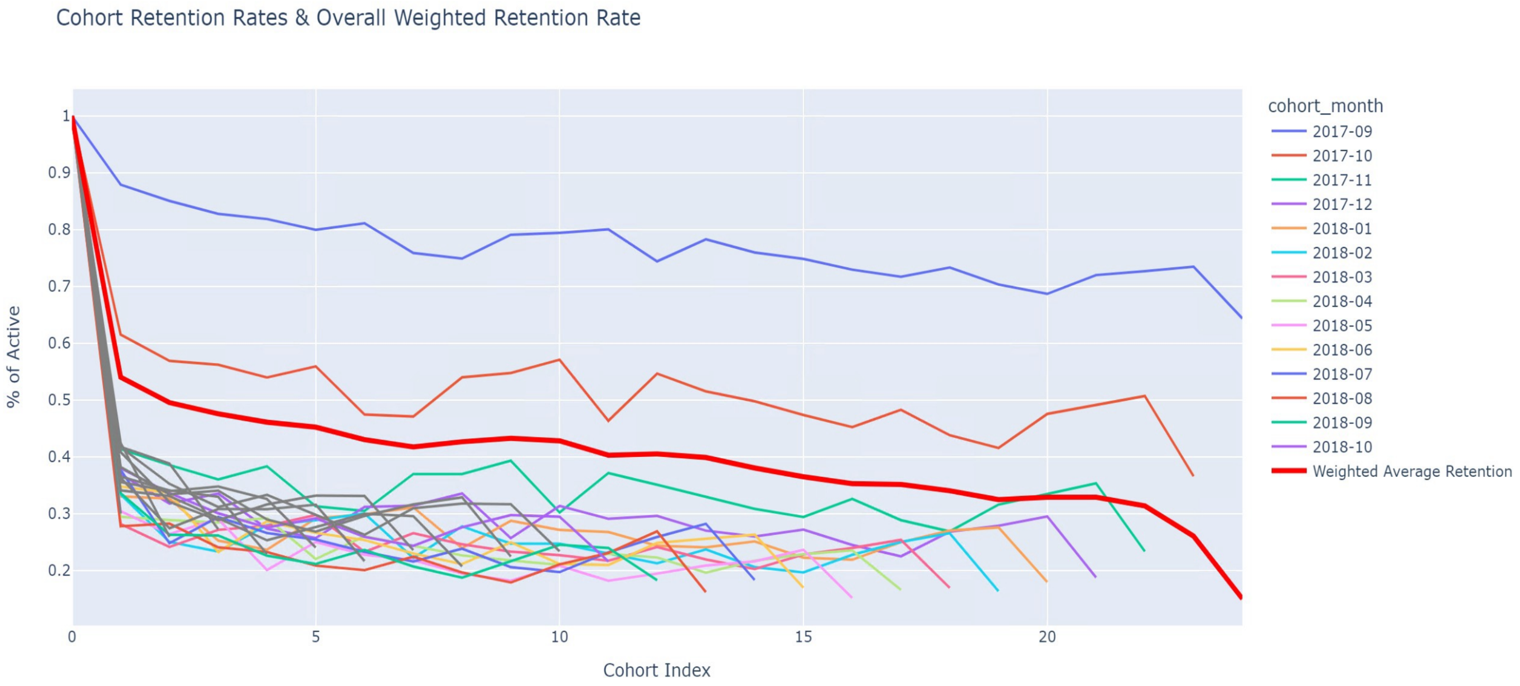
This analysis helps us tackle demand forecasting for a given day. Furthermore, it gives us an intuition on which a very simple model, such as SARIMAX may suffice for the given task in the short-term.

Future lines of work could entail demand forecasting for the top selling products due to shortages in stock explained in the Meeting Minutes.

Source: Team Analysis

FOR DISCUSSION ONLY

A gathering of the customers in cohorts based on first purchases and subsequent ones reveal some additional information on the retention of ClientCo



Source: Team Analysis

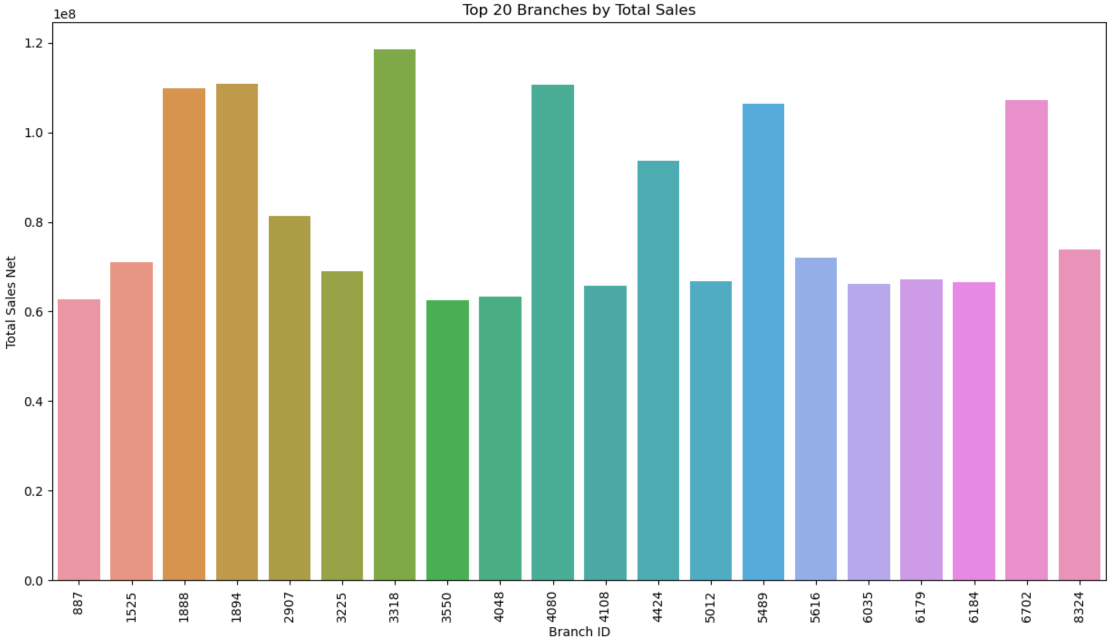
1. Retention of the clients per month from the months of their first purchase. The cohort index shows the number of months passed from the cohort month.

2. The months of the years 2017 and 2019 seem better performing than 2018 in terms of retention in the first few months following the cohort, this could be a sign of the bad performance of 2018 that as stated by Tom Hammer, Supply Chain Director @ ClientCo, OOS and delivery time could have caused losses accountable for up to 50M€ in 2018

3. This line graphs considers the number of customers per cohort and computes a mean retention rate. It shows that ClientCo's concerns about churn are a Tier 1 problem to tackle

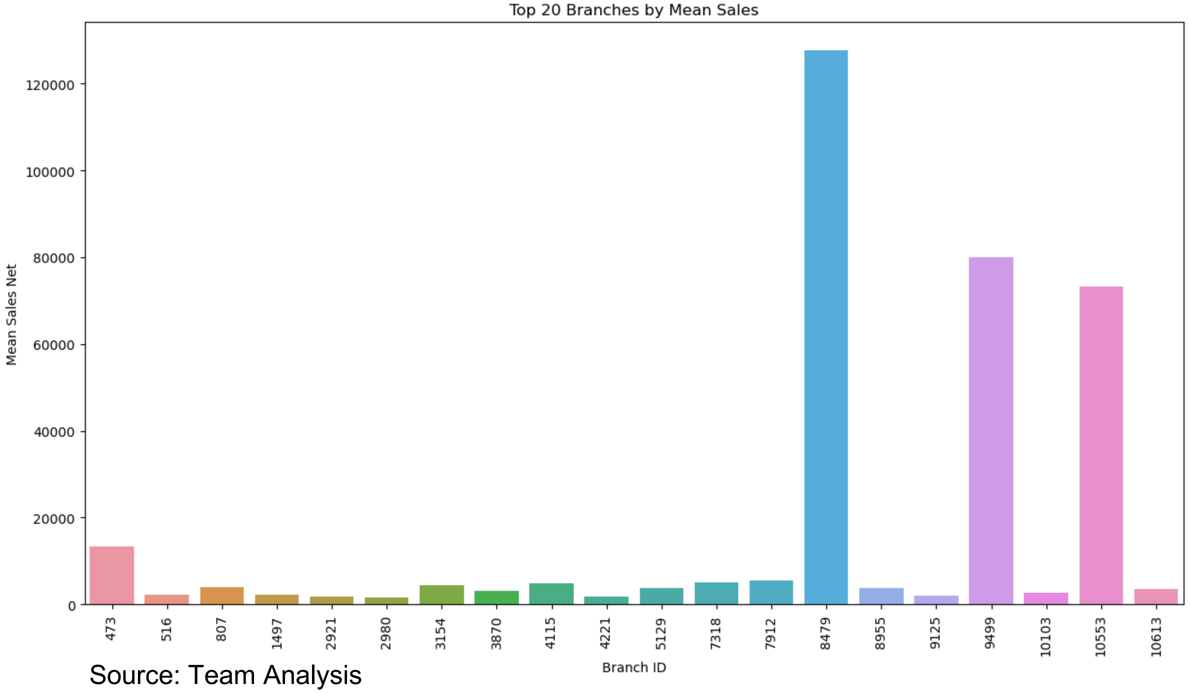
FOR DISCUSSION ONLY

Top Performing Branches show disparities in total and mean sales, which suggests specialized marketing strategies on volume and value transactions…



Source: Team Analysis

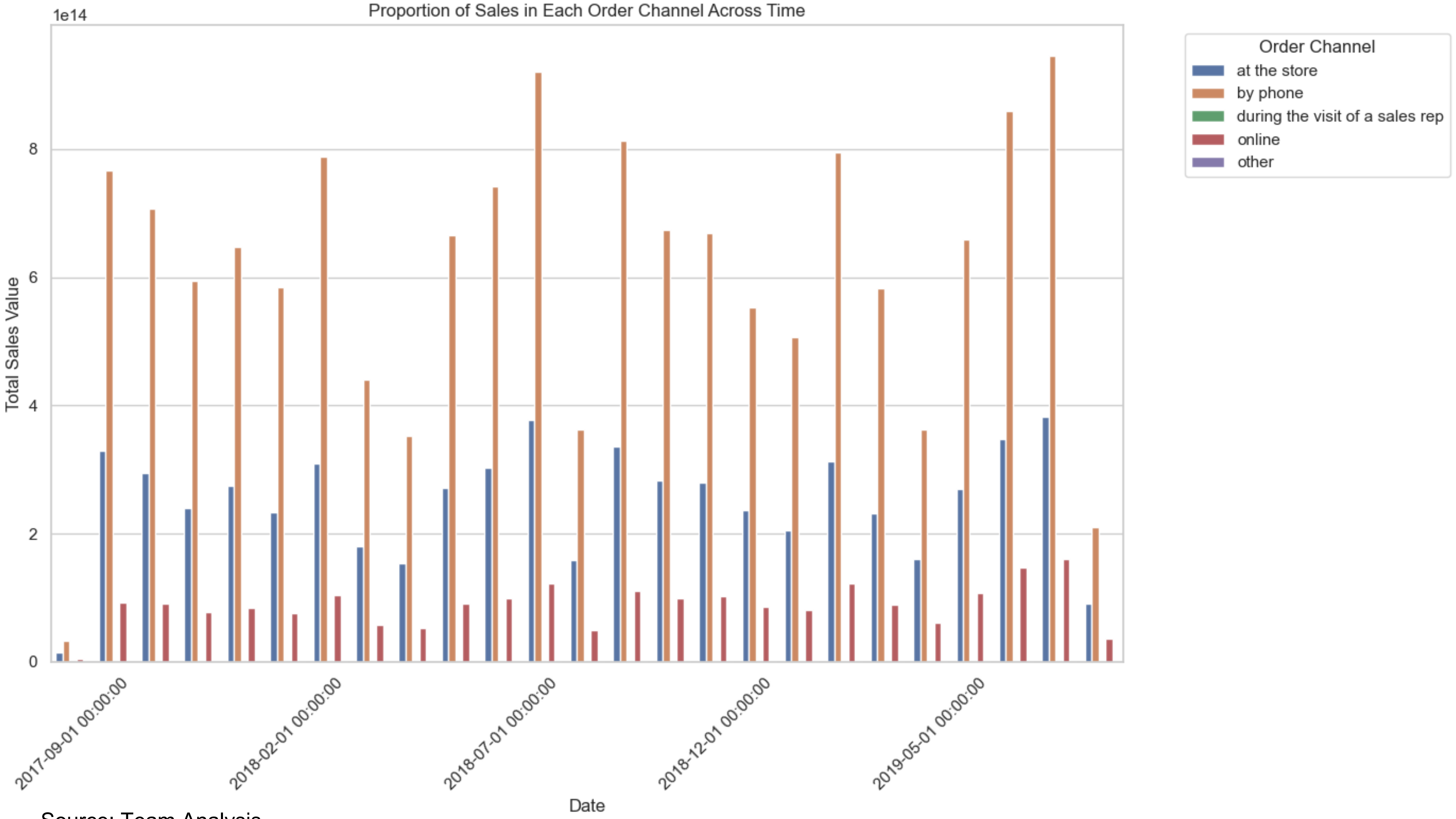
Within the top 20 performing branches by sales, we see a well-balanced and almost uniform distribution, with similar sales across them.



Source: Team Analysis

However, this does not translate when we look at mean sales, where there are clear winners in the top 20 plot. This suggest that these branches (8479, 9499 and 10553) sell high value or specialized items only, when compared to others.

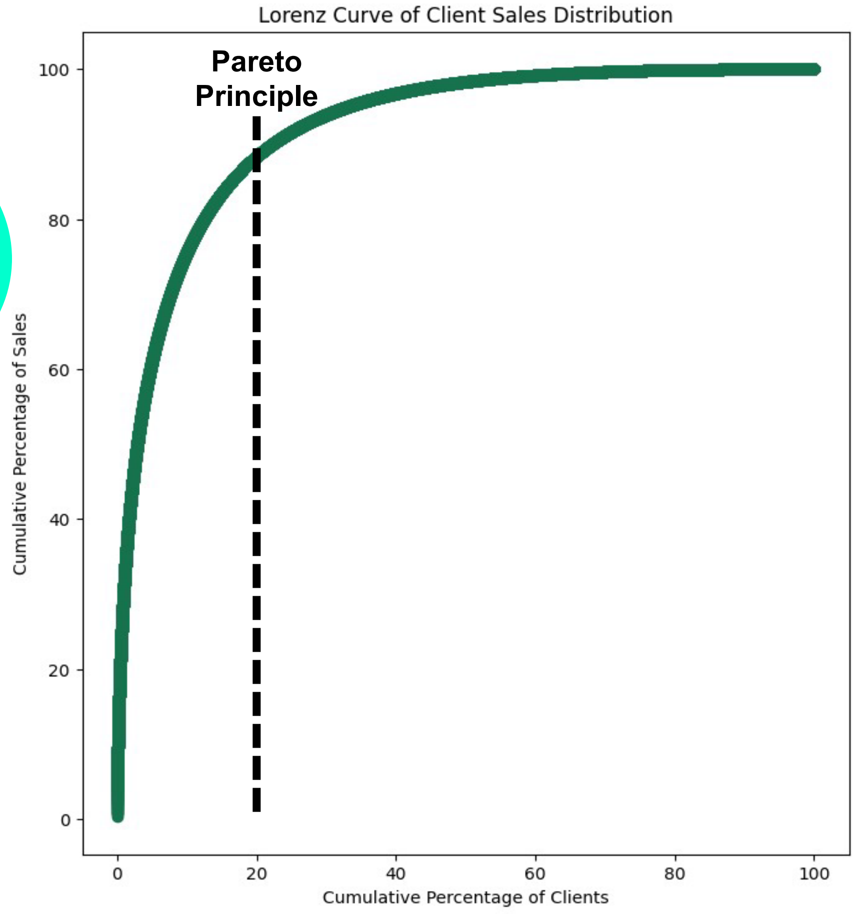
… and the team has also confirmed poor performance through the online channel. Most Sales are conducted through phone or at the store



Source: Team Analysis

FOR DISCUSSION ONLY

Approximately 10% of the client base generates almost 80% of our revenue. We have taken this into account when calculating churn, by calculating more monetary fields to make segmentation stronger…



Source: Team Analysis

Looking at Lorenz Curve, we observe that 10% percent of the clients generate 80% of the revenue or sale invoices. These are valuable customers and the ones to retain and decrease churn.

Furthermore, 20% percent of the clients generate 90% of the revenue, meaning that on the next bracket, the remaining 80% of clients only account for an additional 10% of the revenue. These are the customers to incentivize to purchase more by activating cross-selling initiatives and bundle discounts.

On the other hand we should consider prioritizing the top customers and immediately recognize if they are about to churn so that that sales team can intervene in time by offering discounts or tailored services.

In a nutshell:
For each 100 euros ClientCo earns, 80€ come from 10% of the customers, 10€ from another 10% of the customers and 10€ come from the remaining 80% of the customers.

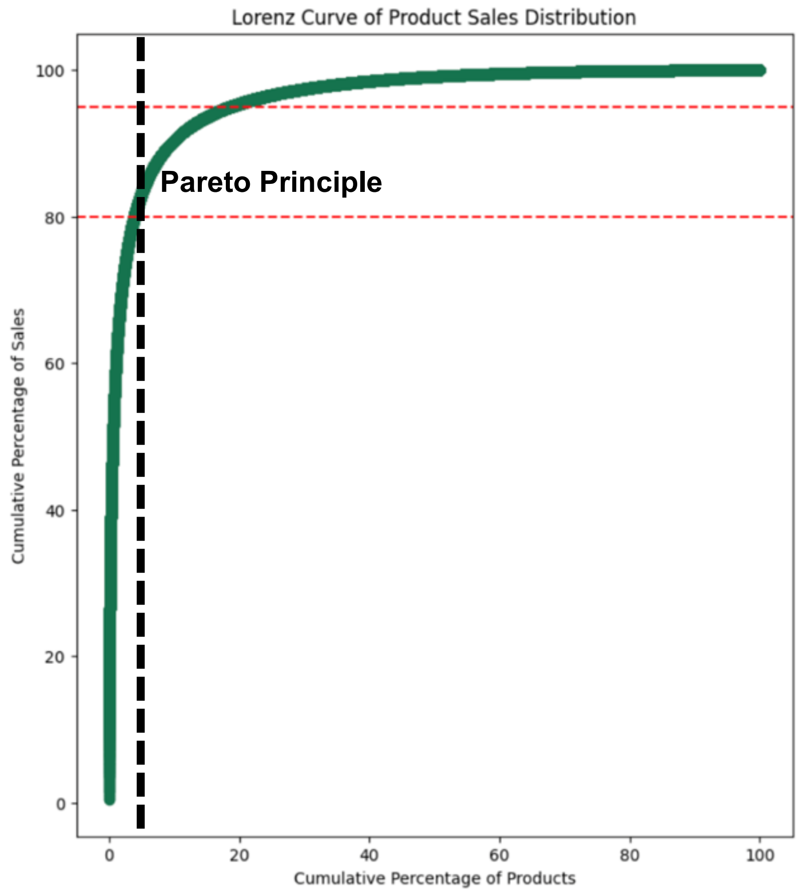
FOR DISCUSSION ONLY

Product Analysis helps ClientCo to bundle products, enabling retention discounts for the best segment of clients and cross-selling to achieve higher revenue streams

1. 3,89% of the products account for 80% of the total revenue.

2. The insight provided by the Lorenz curve also justifies the necessity of improving data management and digitization at ClientCo. With accurate and readily available data, the company can continuously monitor and adjust its focus on the high-performing products, ensuring sustained growth and competitive advantage.

3. The strategic focus helps mitigate risks associated with overstocking or understocking less important products, optimizing the overall supply chain and inventory management processes.



Source: Team Analysis

The starting point in Churn Modelling was defining Churn, exploring Data Distributions and customer behavior.

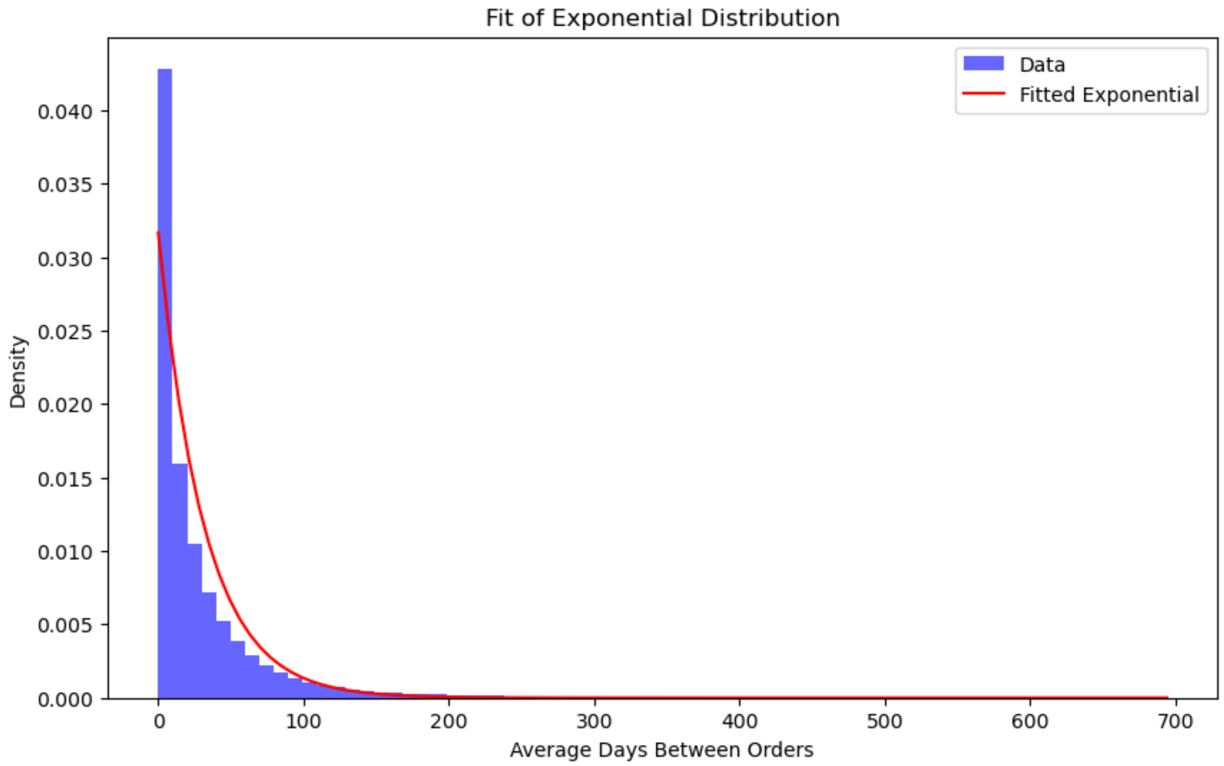
1. At first, we considered using time windows to define churn (ex. recency > 90 days) but we believed this approach was lacking explainability, different clients had different purchasing patterns.

2. We then looked at clients' days between orders to see if their recency significantly deviated from their 'usual' behavior but in order to establish any rule based on the deviations we had to identify a distribution. Even after transformations, normality was not a possibility. Exponential / Poisson distributions when its lambda value is very low were also tried. Despite the latter seemed a good fit, the statistical tests proved otherwise.

3. Most clients only wait 1.8 days on average to make purchases. If not a distribution:

   I. How can we label those clients that order with less frequency but in bigger batches?

   II. How can we model occasional purchasers differently? without directly saying "Churn" or dropping them?



Source: Team Analysis

FOR DISCUSSION ONLY

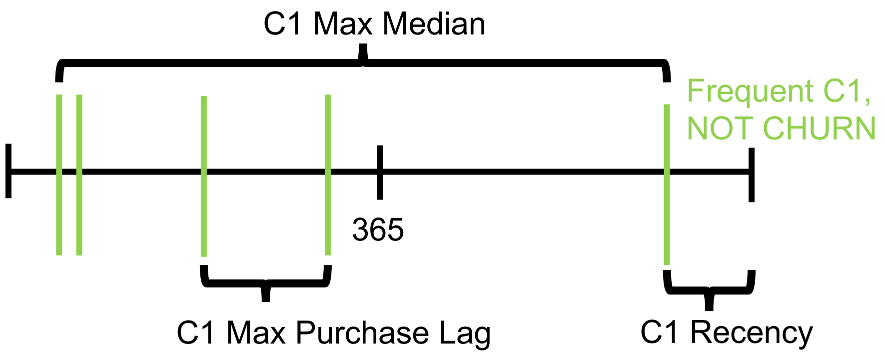
The first step to develop a churn model without labels is to make reasonable assumptions. We defined churn based on Pattern Analysis: Recency, Max. Purchase Lag and Median days btw. orders…

Criteria 1 – Frequent Purchasers – Num tickets > 1

Recency > (Max Days Between orders + Median Purchase Days)

But Why?

- Median is robust, mean average days and standard deviations between purchases are NOT.
- Maximum Days Between Orders is a dynamic proxy and customized for every client.
- Customer that buy each day, each 7 days and each 30 days will be targeted with the proxy.

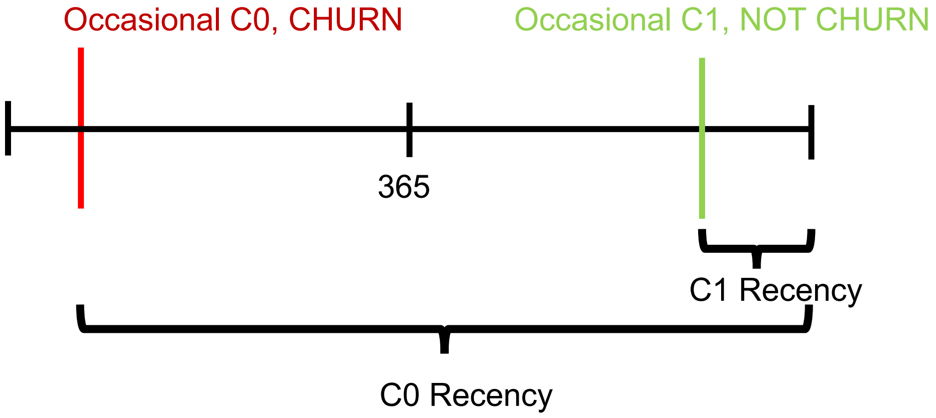


Criteria 2 – One-time Purchasers – Num tickets = 1

Recency > 365 Days

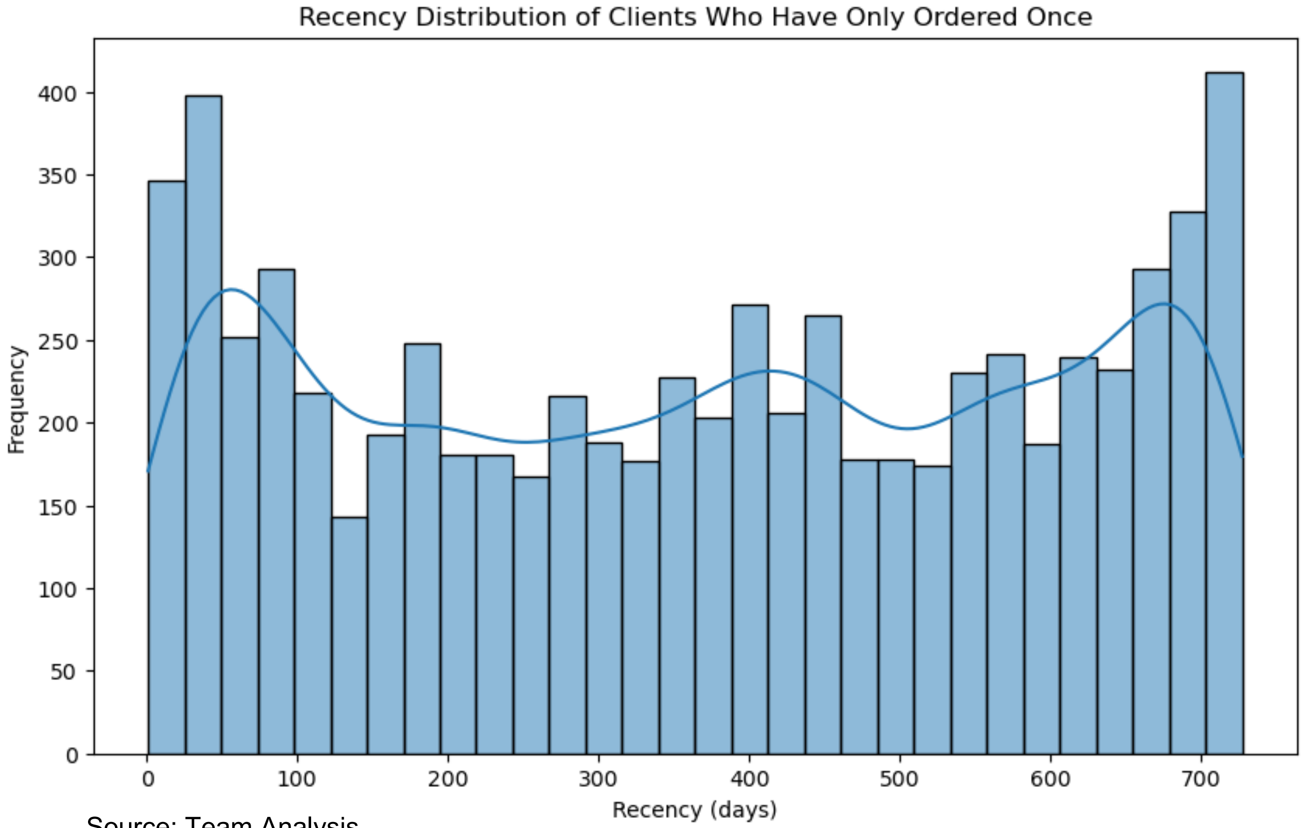
But Why?

- We cannot calculate the Maximum Days Between Orders for customer that only bought once.
- Since we have two years of data or 730 days, we can create a proxy on 365 days.
- Client Length and Recency is the same for this type of customers.



We summed both buckets to use both approaches:

…since it allows to build a robust and dynamic proxy for frequent and occasional buyers, being the later less engaged but key to customer acquisition.

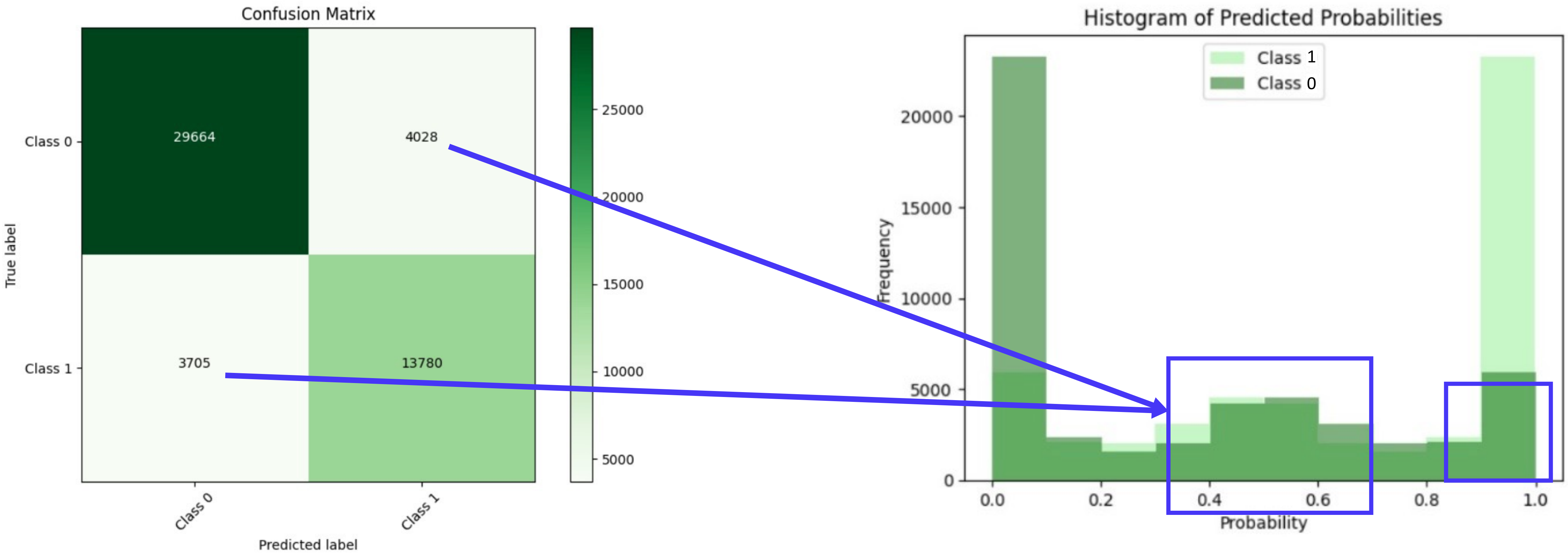


Source: Team Analysis

Take a look at this diagram. Recency for occasional buyers almost follows a Uniform Distribution! Very Random!
Thus, we cannot directly drop occasional customers. There are too many and should be treated differently as we explained in the former slide.
In this context, High Recency means Churn while Low Recency means a need for further engagement.

FOR DISCUSSION ONLY

Once churn was defined, we could develop a predictive model using Random Forest Classifier. These are the results:



Error type B (false negative rate) accounts for the lowest error in our confusion matrix while Error type A (false positive rate) is very close to it. The reason for this errors are calculated probabilities close to 0.5. We also have significant cases where the model misclassifies. This is not particularly bad, even if making mistakes we would need to evaluate the model on the 450K clients and not on just the 170K provided by the dataset

FOR DISCUSSION ONLY

Brainstorming an increasing number of calculated fields was key to define and build a churn model
The Team defined the following variables for each client_id:

Recency

Refunds per ticket → (N_neg_invoices / N_invoices)

Refunds % → (sum_invoices / sum_returns)

% Ch. Sales | % Ch. Quantity → (Period Sum 2/ Period Sum 1) – 1

Total_Period/2

mean_order_delay → (invoice_date – order_date)

Total_period → (First_order_date – Last_order_day)

Monetary

total_sales | total_quantity | n_invoices | n_returns | sum_invoices + | sum_returns -

sales_percentage → (Client's Sales / Total Sales) x 100

avg_ticket_quantity → (N_items / N_transactions)

avg_sale_value → (Sum Sales / Total Sales)

Frequency

avg_days_btw_orders | std_days_btw_orders | monthly_frequency

Other – Customer Care

channel | most_product_id | most_branch_id | unique_branches

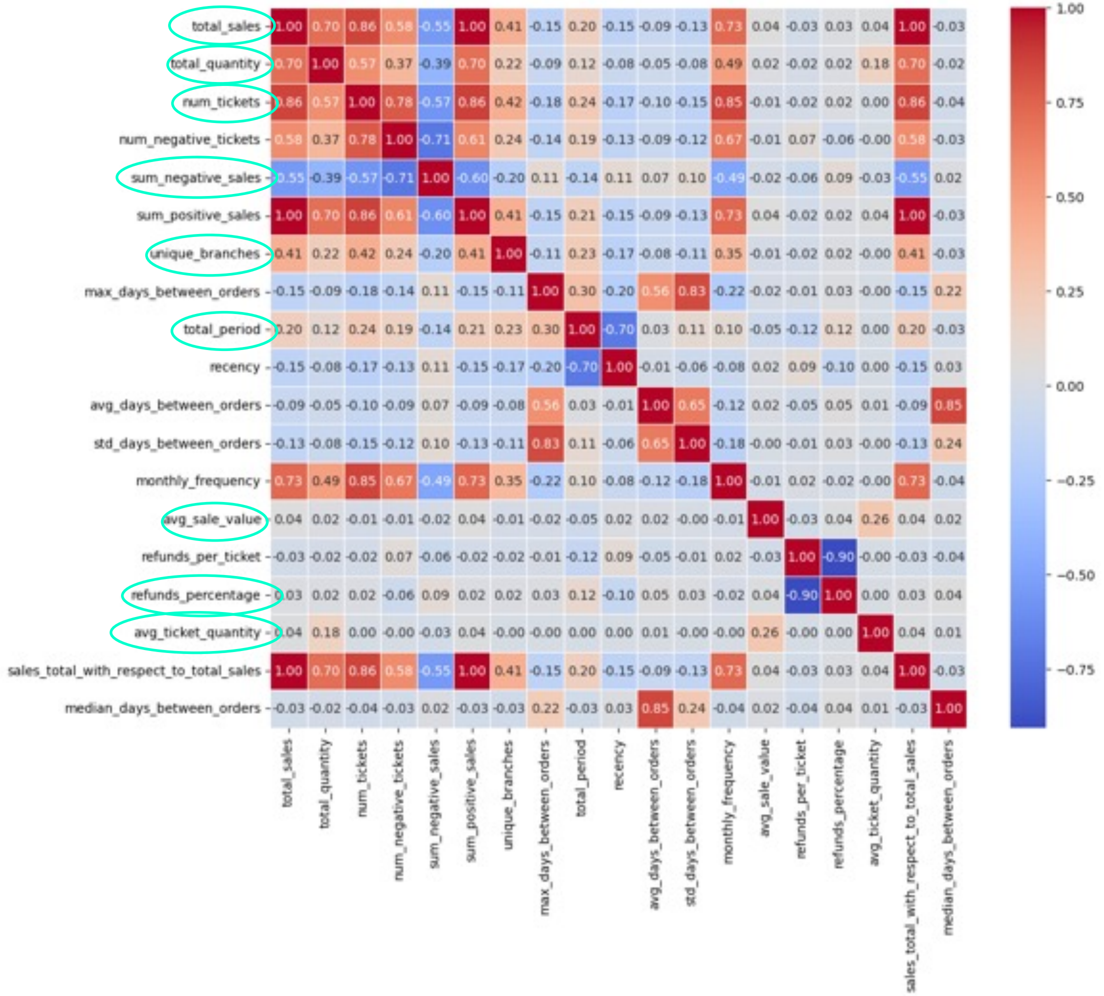
FOR DISCUSSION ONLY

In total, the team created 19 variables to measure, but many of them were highly correlated. To build the churn model only 9 variables were kept, yet they track customer behavior in a very controlled manner.

The idea of this model is to have recency, maximum purchase lag for defining churn, as well as the median days between purchase as a regularization parameter.

All other variables that measure the engagement of a client will be considered as explanatory variables.

In essence, based on customer engagement, we will be able to know if the client has churned or not, based on the number of tickets, its total sales, the average sale value and many more.



FOR DISCUSSION ONLY

We iterated through different models and analyzed the feature importance to secure the best model for robustness and churn recognition. Also, we dropped all the variables used to define churn, in order to avoid data leakage.

```
recency                       0.485585
total_period                  0.163582
max_days_between_orders       0.114335
num_tickets                   0.066815
std_days_between_orders       0.064804
avg_days_between_orders       0.035265
total_sales                   0.027370
total_quantity                0.018974
avg_sale_value                0.006148
avg_ticket_quantity           0.005570
unique_branches               0.004480
sum_negative_sales            0.003553
refunds_percentage            0.003520
dtype: float64
```

```
total_period           0.478026
num_tickets            0.117006
total_sales            0.106288
total_quantity         0.079405
avg_sale_value         0.077873
avg_ticket_quantity    0.056311
refunds_percentage     0.031767
sum_negative_sales     0.031608
unique_branches        0.021717
dtype: float64
```

The best hyperparameters for the model found by Optuna were:

1. Number of estimators = 178
2. Maximum depth = 12
3. Minimum samples split = 3
4. Samples leaf = 9

These allowed us to achieve a f1 score of 0.78

FOR DISCUSSION ONLY

Looking closely at the results of our model per cluster we find an extremely good performance in labelling churn in Low value and Lost customers but there is still room for improvement

The model has a similar amount of predicted churn and actual churn. Across segments the accuracy seems to differ:

The model performs well when labeling low value and lost customers, but for top, high value and medium value customers it underestimates the churn. This is quite unfortunate since these customers are the most valuable.

The reason for the performance difference on the top three segments likely comes down to the number of clients in the segments. The sample of the dataset we worked with is unbalanced. Solutions could be a balanced sample or training separate models

Customer_segment	number of clients	number of churned	pred	true_positive_percentage	false_negative_percentage
Top Customers	14136	181	81	0.056593	0.516412
High value Customer	16363	645	348	0.886146	1.240604
Medium Value Customer	35589	4796	2975	4.602546	3.756779
Low Value Customers	51756	15359	15868	20.252724	10.406523
Lost Customers	52742	37300	40263	66.711539	9.628000

Customer_segment	false_positive_percentage	true_negative_percentage	accuracy	recall	precision	F1
Top Customers	1.223826	98.203169	0.982598	0.098765	0.044199	0.061069
High value Customer	3.055674	94.817576	0.957037	0.416667	0.224806	0.292044
Medium Value Customer	8.873528	82.767147	0.873697	0.550588	0.341535	0.421567
Low Value Customers	9.423062	59.917691	0.801704	0.660575	0.682466	0.671342
Lost Customers	4.010087	19.650374	0.863619	0.873879	0.943298	0.907262

FOR DISCUSSION ONLY

Valuable insights were given by looking in depth at the product analysis where we discovered an extreme example of Pareto principle

Product group A

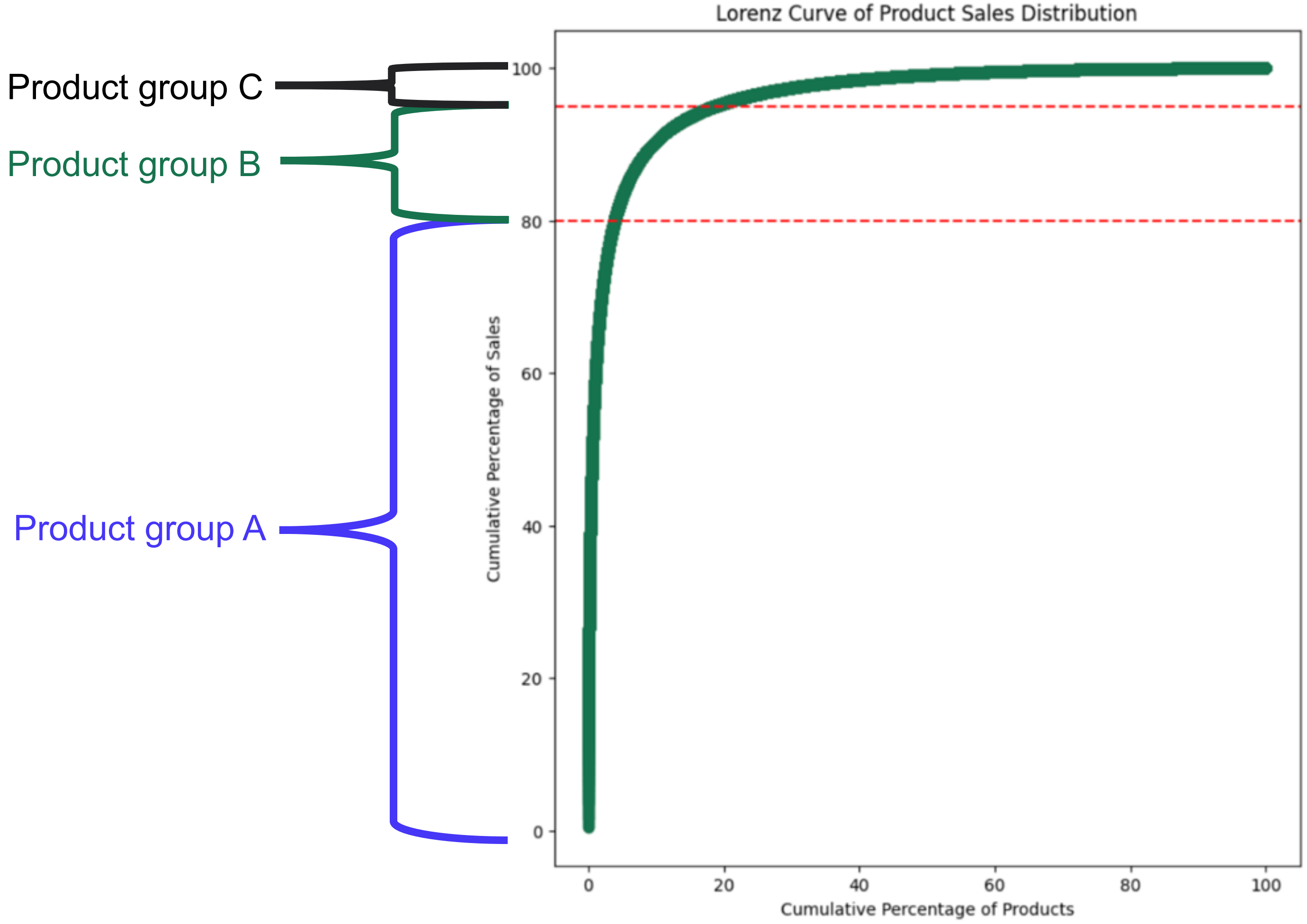
3.8% of products (23 884)
80% of total sales (7.9 billion)
49 million transactions
161 average transaction value

Product group B

14.8% of products (90 364)
15% total sales (1.5 billion)
7.4 million transactions
203 average transaction value

Product group C

81,4% of products (491 327)
5% total sales (500 million)
2.4 million transactions
208 average transaction value

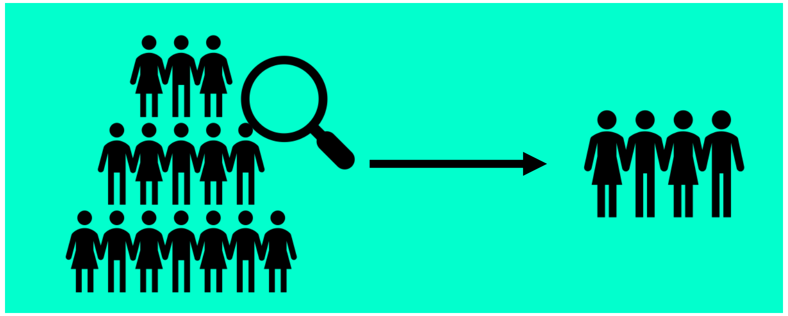


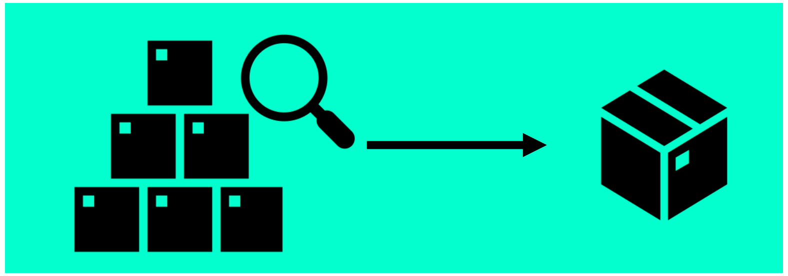
Source: Team Analysis

FOR DISCUSSION ONLY

By sampling the P&C we managed to identify potential cross-selling opportunities from which ClientCo could benefit from a 2.4% increase in sales

1





By sampling the 100 top clients and the top 50 products we got a sample of 8604 transactions, that created **191,000** association rules..

2

We then applied the following limitations:

- Confidence > 0.5
- Lift > 1.2
- Support of the consequent < 0.2

And identified **31,334** opportunities

3

We calculated the expected return as follows:

$$\textit{Expected return} = \textit{confidence} \times \textit{price of consequent}$$

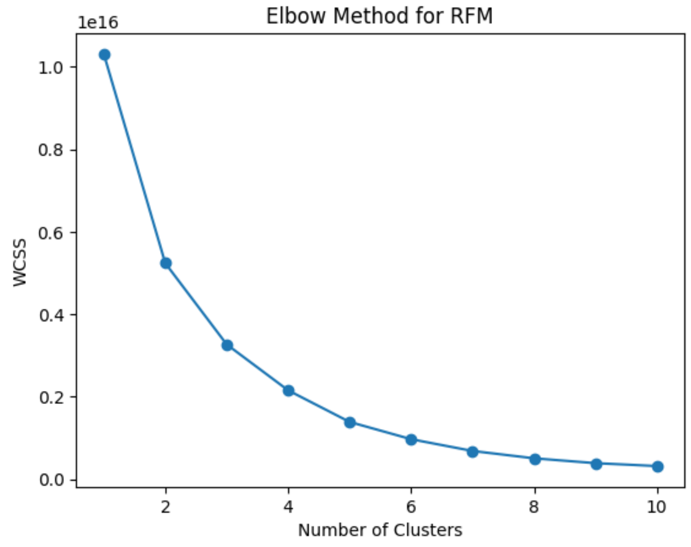
By then looking at the highest return consequent for each product, we reduced the number of opportunties to **37…**

...which on the sample transactions increases the sales by

2.4% !

FOR DISCUSSION ONLY

Customer Segmentation on weighted RFM is key to calculate the Customer Lifetime Value by cluster. The team will use this to select which customers to target.



Top-Clients: 8%
High-Value Clients: 9.5%
Medium-Value Clients: 20.8%
Low-Value Clients: 30%
Lost Clients: 31%

Our RFM is weighted, giving three times the importance to monetary value when compared to frequency or recency.

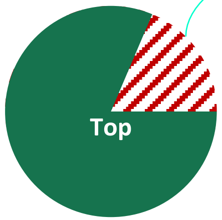
The number of ideal clusters based on the Elbow plot that looks at the percentage of total variance explained indicates that the optimal number is around 4 to 6 clusters.

When partitioning and looking at the centroids of each cluster, we realized we could classify them into Top, High, Medium, Low value and Lost customers.

Customer_segment	number of clients	recency	frequency	monetary_value
Top Customers	14136	4.72	2,755.72	447,561.73
High value Customer	16363	14.17	558.25	101,964.87
Medium Value Customer	35589	52.11	184.99	33,928.92
Low Value Customers	51756	126.96	41.85	8,413.37
Lost Customers	52742	325.64	6.25	1,822.40

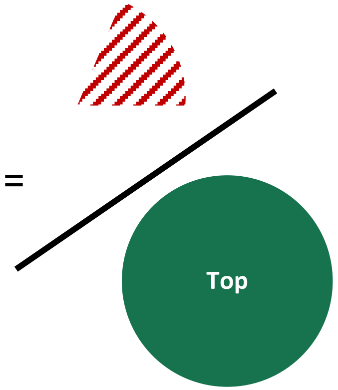
The team followed a very simple and effective method to calculate the CLV based on the churn rate in a way such that difference between clusters are shown..

First of all we looked at the clustered clients



Churned customers within the cluster

Churn rate =



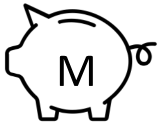
By analyzing the churned customers within each cluster we calculate churn rate and ...

Retention factor =

1 – churn rate / churn rate

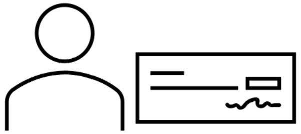
Retention Factor

X



Monetary value from RFM

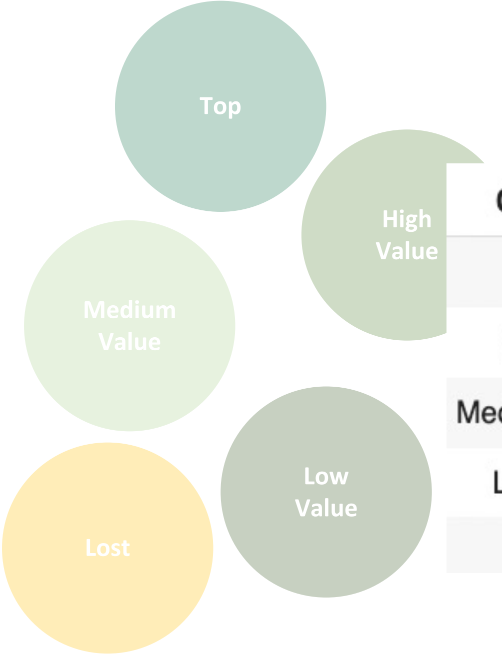
=



...Retention rate, which we multiply by the previously calculated monetary value and obtain the CLV per cluster !

FOR DISCUSSION ONLY

The team followed a very simple and effective method to calculate the CLV based on the churn rate in a way such that difference between clusters are shown..



Retention factor =
1 – churn rate / churn rate

Customer_segment	monetary_value	churn_rate	retention_factor	CLV
Top Customers	447,561.73	0.01	75.92	33,980,263.87
High value Customer	101,964.87	0.04	24.64	2,512,518.90
Medium Value Customer	33,928.92	0.14	6.41	217,396.44
Low Value Customers	8,413.37	0.30	2.37	19,914.48
Lost Customers	1,822.40	0.71	0.41	755.25

X M

Monetary value from RFM

=

First of all we looked at the clustered clients

By analyzing the churned customers within each cluster we calculate churn rate and ...

...Retention rate, which we multiply by the previously calculated monetary value and obtain the CLV per cluster !

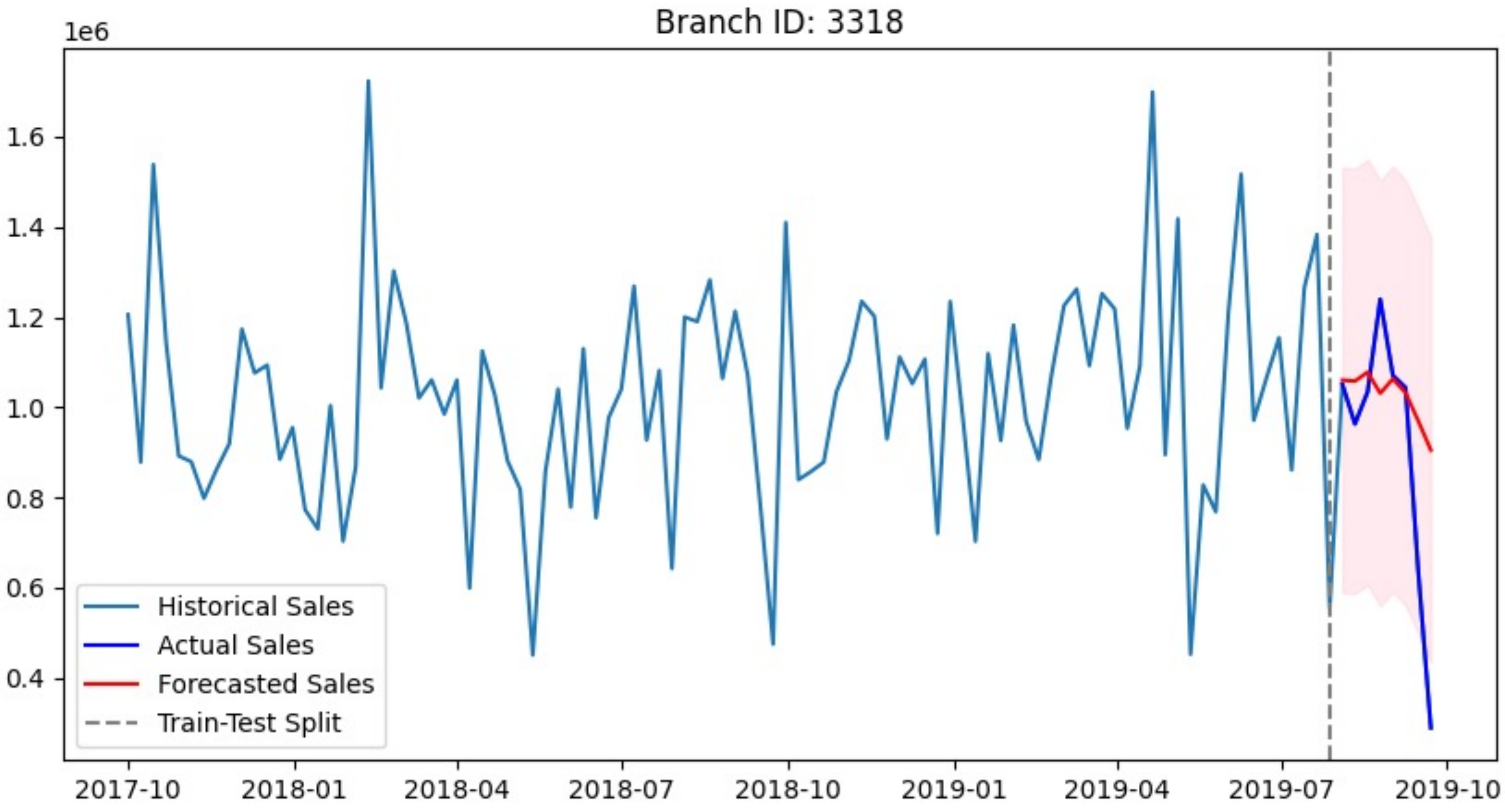
SARIMA predicted a given level of stock needed for the given Branch. This could help ClientCo. identify which level of stock is required as a percentage out of the total number of branches.

As we are able to see, branches exhibit a seasonal behaviour with May having a drop in both years.

The branch was chosen based on sales criteria. It is the top-selling branch out of 582 as the time of this analysis.

The model can predict with a good confidence, but it is limited by sharp increases and decreases. Nevertheless, the confidence interval is within the borders most of the time.

Let's see how do individual products perform, we promise, you will like it!



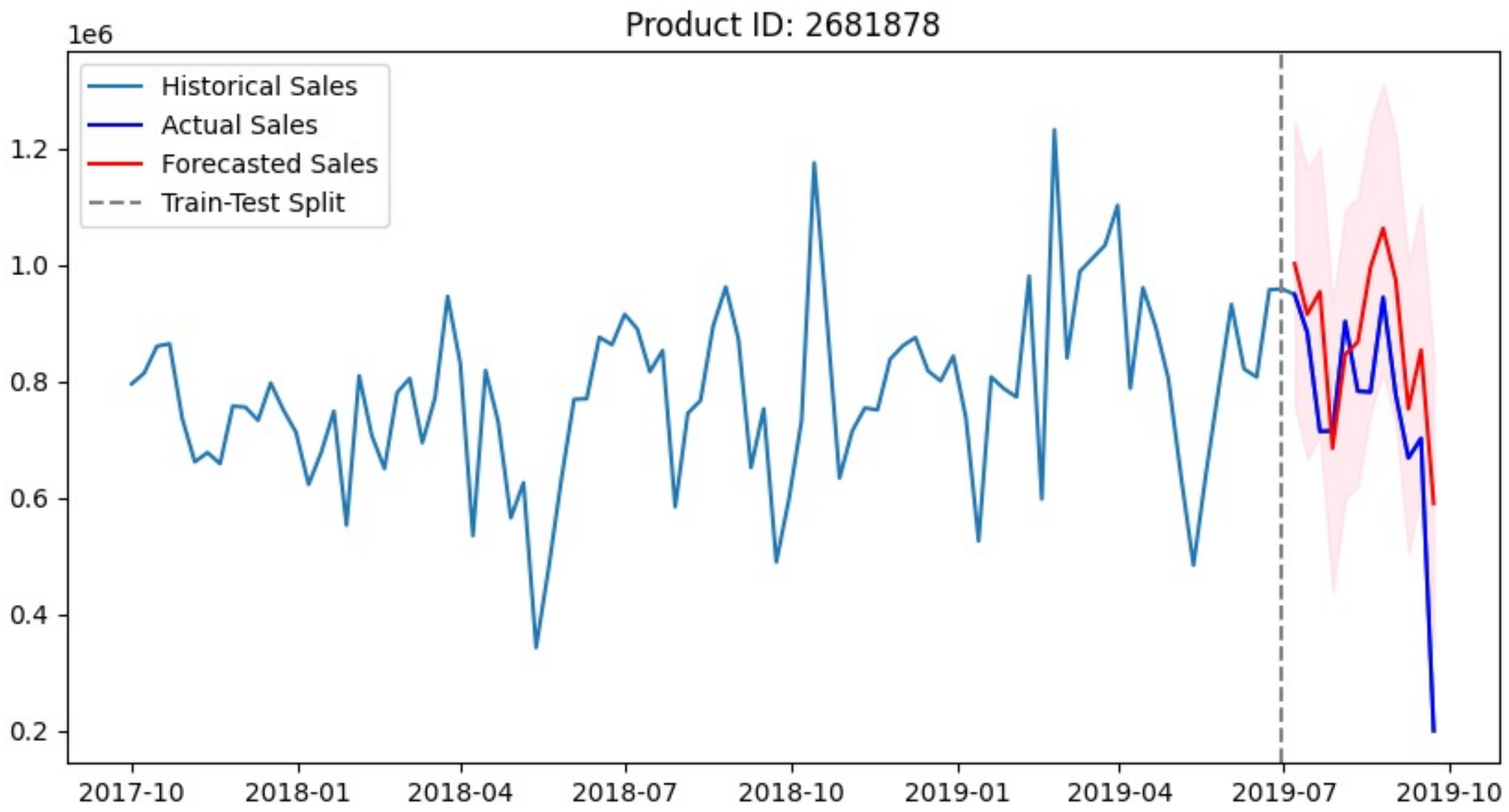
FOR DISCUSSION ONLY

Forecasting products in a monthly basis had a better performance. We recommend to implement SARIMA for global products and redistribute to branches based on the Upper Confidence Interval level.

The Team tried with products from Group A, and it gave significant performance when deploying a weekly basis SARIMA.

We recommend to implement this model for the global demand of a given product. Then quantify the percentage of quantity to be sent for a given branch.

Due to possible stockout, we do not recommend to send just the amount given by the red line, but the amount given by the upper confidence interval.



FOR DISCUSSION ONLY

The most important part is to use all this knowledge and analyses in a combined manner. Here is the real value of bringing Churn, Customer Lifetime Value and RFM Segmentation together.

client_id	Customer_segment	churn_max	churn_proba	CLV	risk
647406	Top Customers	0	0.746600	3.398026e+07	2.536967e+07
1816983	Top Customers	0	0.728655	3.398026e+07	2.475989e+07
376530	Top Customers	0	0.722670	3.398026e+07	2.455651e+07
330321	Top Customers	0	0.693708	3.398026e+07	2.357236e+07
268020	Top Customers	0	0.681810	3.398026e+07	2.316807e+07
49546	Top Customers	0	0.642172	3.398026e+07	2.182118e+07
1425521	Top Customers	0	0.614281	3.398026e+07	2.087344e+07
1483475	Top Customers	0	0.585591	3.398026e+07	1.989852e+07
861755	Top Customers	0	0.579236	3.398026e+07	1.968258e+07
545446	Top Customers	0	0.578082	3.398026e+07	1.964339e+07
2257695	Top Customers	0	0.565225	3.398026e+07	1.920651e+07
1438544	Top Customers	0	0.564882	3.398026e+07	1.919483e+07
836161	Top Customers	0	0.564179	3.398026e+07	1.917096e+07
1273854	Top Customers	0	0.563964	3.398026e+07	1.916366e+07
1201752	Top Customers	0	0.563848	3.398026e+07	1.915971e+07
1631107	Top Customers	0	0.563202	3.398026e+07	1.913776e+07
995015	Top Customers	0	0.561549	3.398026e+07	1.908160e+07
1284203	Top Customers	0	0.559501	3.398026e+07	1.901200e+07
2096102	Top Customers	0	0.558283	3.398026e+07	1.897060e+07
1709865	Top Customers	0	0.554956	3.398026e+07	1.885755e+07

1 Prioritizing Top Customers (calculated by our weighted RFM segmentation)…

2 … using our Churn Probability with value 0 (not churn) but ordered by highest probability…

3 …calculating Customer Lifetime Value for that given client group…

4 …and calculating the associated risk of that customer leaving = churn probability x expected monetary CLV

= We can target these specific clients that are the most valuable and retain through customized actions (eg. calling them directly and offer big batch discounts).

FOR DISCUSSION ONLY

Implementing this approach with Trigger Events in ClientCo's CRM system is another dynamic solution, based on model retraining and subsequent probability increases

Churn
0
1
0
0.7
0
0
0.2

Churn
0
1
0
0.7
0
0
0.5

When a new row (client_id) in the churn column changes to a probability level of 50%

twilio

Apache Airflow

It triggers a process that sends out alerts in 2 ways

Sales/Care Team

- Visualization of clients probable of churning
- Ranking for sales reps based on importance of customer
- Targeted information to the branch related to the client

Client

- Tailored Campaigns
- Product Analysis and Recomendations
- More or less aggressive Discounts based on probability of churn

FOR DISCUSSION ONLY

The future of ClientCo. looks promising. Based on feedback received from the stakeholders and the team analyses, a forecast of the P&L/IS is provided.

The numbers have been forecasted on the basis of the average growth of the years *2016 – 2019*, period of governance of the current CEO

The *1%* revenue growth was suggested by the new CDO as a potential effect of the churn implementation plan. On top of it, we believe we can further improve churn prevention by 0.2% based on the CRM solution. An increase in sales of 2.4% is also expected.

COGS will decrease based on removing low-income earning products. Stock prediction improvements will allow for a 0.5% improvement.

R&D invested will increase costs by 2%. Hiring a greater data team will increase cost by another 2%

ClientCo. Income Statement/Profit & Loss in millions (USD)	2017 31/12/2017	2018 31/12/2018	LTM 30/06/2019	AS-IS 2020 31/12/2019	TO-BE 2020 31/12/2019
Decrease in Churn Assumption by CDO					*1.0%*
Decrease in Churn by CRM Churn Dynamics					*0.2%*
Increase in Sales due to Cross-Selling Efforts					*2.4%*
YoY Rev. Growth %		*0.4%*	*1.8%*	*1.8%*	*1.8%*
Revenue	15,440	15,504	15,787	16,070	16,358
Decrease in COGS due to ZBB Product Analysis					*1.0%*
Better Stock Prediction Inventory with SARIMA					*0.5%*
YoY COGS. Growth %		*0.1%*	*1.7%*	*1.7%*	*1.7%*
- Cost of Revenue	11,670	11,682	11,882	12,081	11,900
Gross Profit	3,770	3,822	3,905	3,989	4,459
Gross Profit Margin %	*24.42%*	*24.65%*	*24.74%*	*24.82%*	*27.26%*
R&D Deployment					*-2.0%*
Hiring a team of data scientists and MLOps Engineers					*-2.0%*
YoY SGA. Growth %		*0.5%*	*1.4%*	*1.4%*	*1.4%*
- Selling, General & Admin Expense	3,130	3,144	3,187	3,231	3,405
- Other Operating Items	2	3	5	included in SGA	included in SGA
Operating Profit (EBIT*)	638	675	714	758	1,053
				Potential Upside	**39.03%**

FOR DISCUSSION ONLY

The Team outlines the next steps and future LoW for ClientCo in this timeline:

Hire a team of around 20 to 50 people for the Data and AI Team as it is currently very small

Connect IT Billing/Transaction platform (SAP BBOO) with CRM to push segmentation to Sales teams

Contract a Cloud Provider, we recommend Google Cloud, as it has the edge in Machine Learning Modelling over other providers. Push the SQL tables to BigQuery

Implementing our Churn Model into IT with a MLOps Pipeline using Jenkins/GitLab, Airflow and GCP

Connect Logistic platforms to Transactions for Better Stock Demand Forecasting. Leverage new barcode system to build database table that would help gather qualitative data. Push it to GCP.

Increase Online Sales with a proprietary website and discounts for Top Customer so there is less dependency on branches minimizing churn. Maximize Sales by Cross-Selling bundles

Make constant retraining on the models and monitor data drift. Conduct analysis on the poor selling products and see if it makes sense to discontinue them

Predict Stock for products on a weekly basis to improve stockout probability

…and also make a unique model for Top customers due to the unbalanced dataset!

FOR DISCUSSION ONLY

The Team was given a seemingly impossible task by ClientCo and BCG X, but nothing is truly beyond reach.

Ignacio Alonso López-Linares

MBD Student, Accenture, PwC

i.alonso.linares@student.ie.edu

Paolo Brambillasca

MBD Student

paolo.brambillasca@student.ie.edu

Emilio Rodríguez Capitán

MBD Student, Danfoss

emilio.capitan@student.ie.edu

Morten Aas-Lyngby

MBD Student, Sparebanken Sør, StudConsult AS

morten.aas.lyngby@student.ie.edu

Alberto Donà

MBD Student, CrossFund, Perris e Associati

alberto.dona@student.ie.edu

BCG X

I have 65M transactions in this dataset. Can you extract any value & build a churn model?

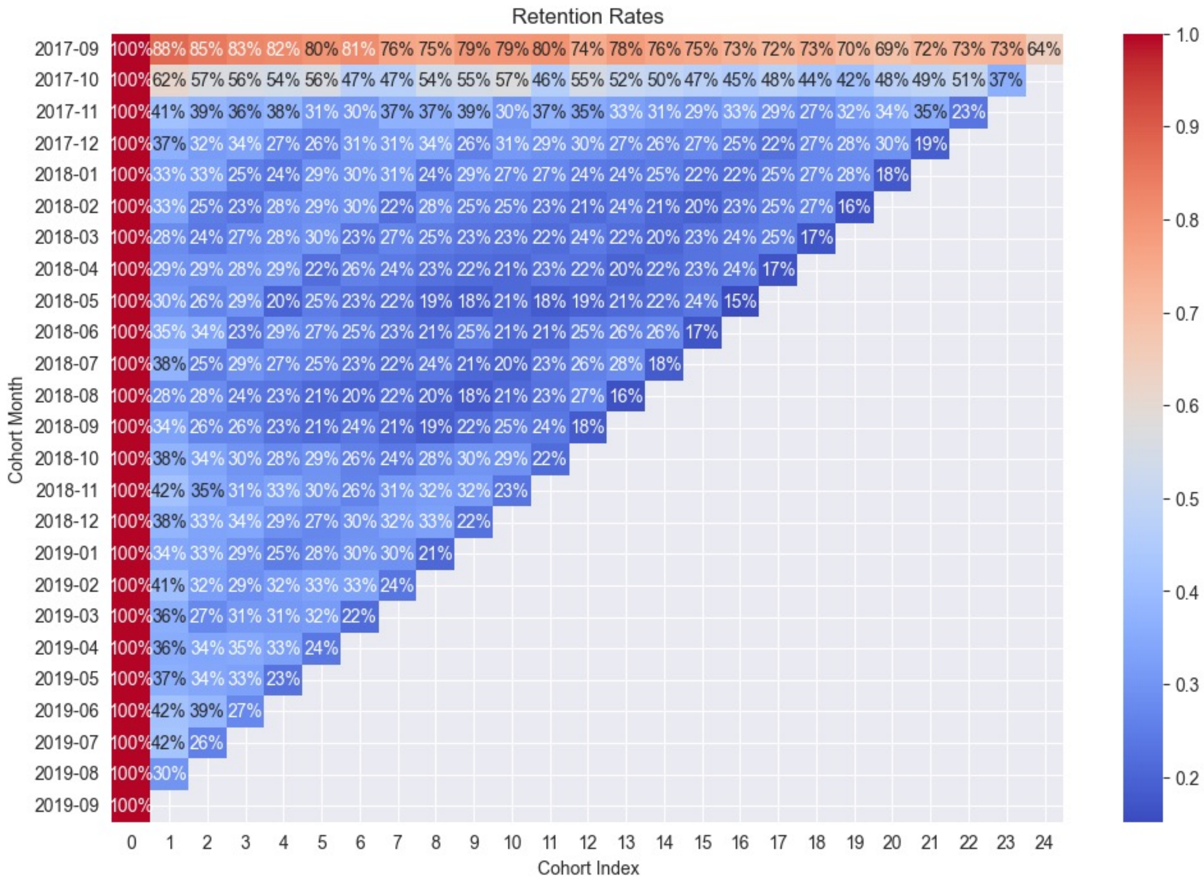
The Team

Consider it done

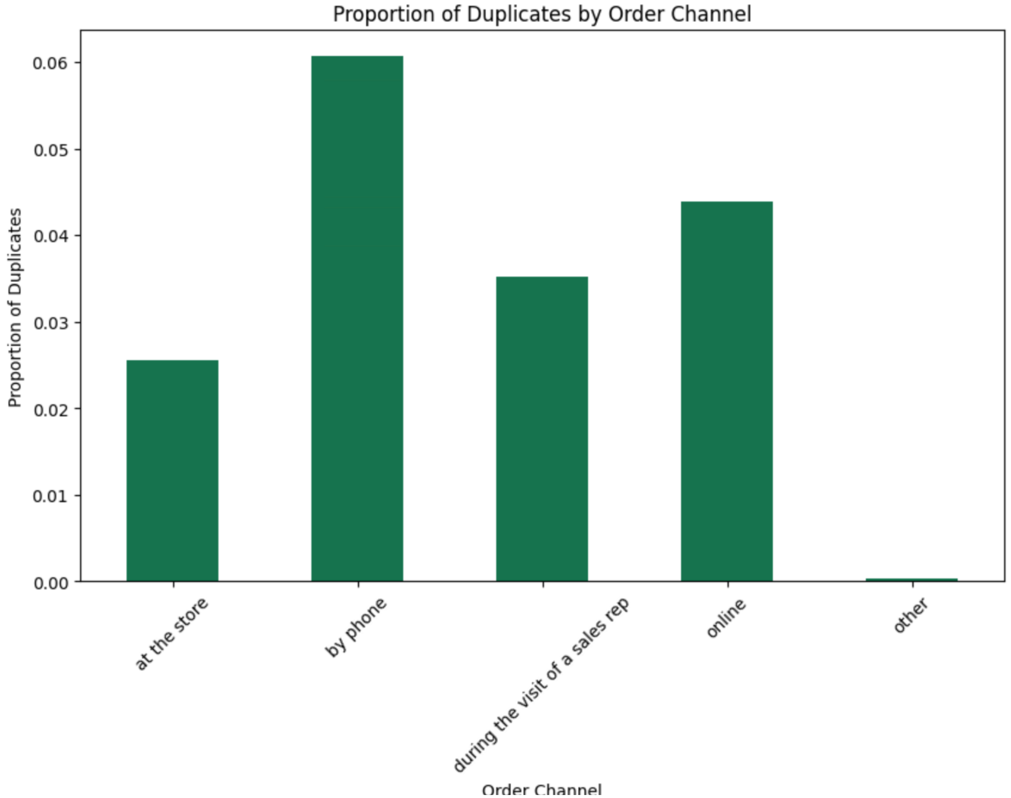
Thank you for listening
Q&A Time

Appendix I

Retention Curve by Cohorts Matrix

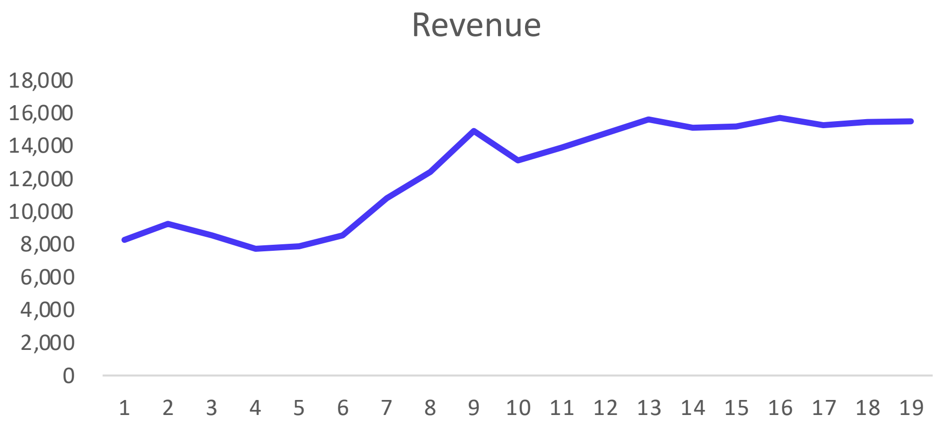


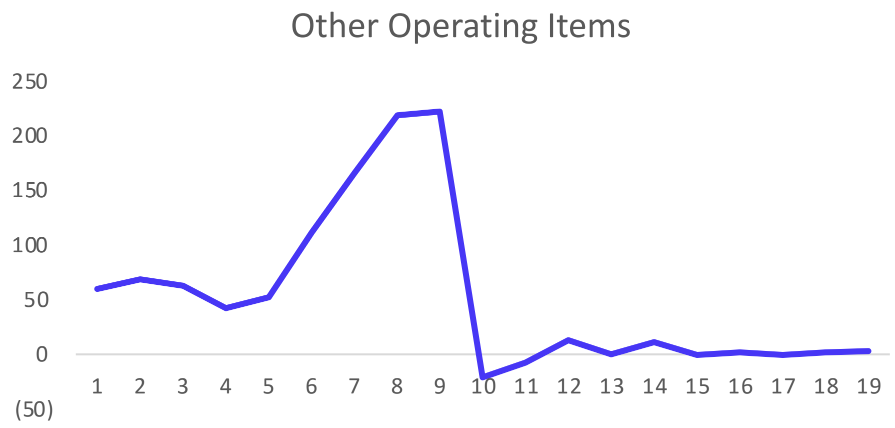
Duplicates Analysis by Channel

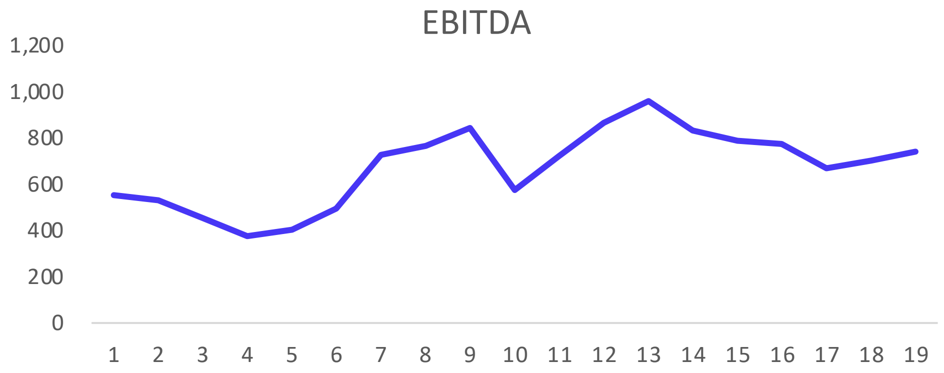


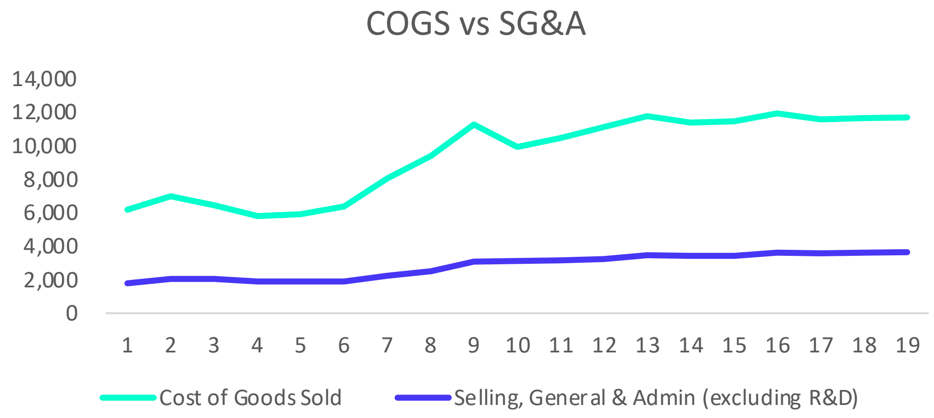
Appendix II

Comparison charts of key metrics (Revenue, Other Operating Items, EBITDA, COGS v SG&A









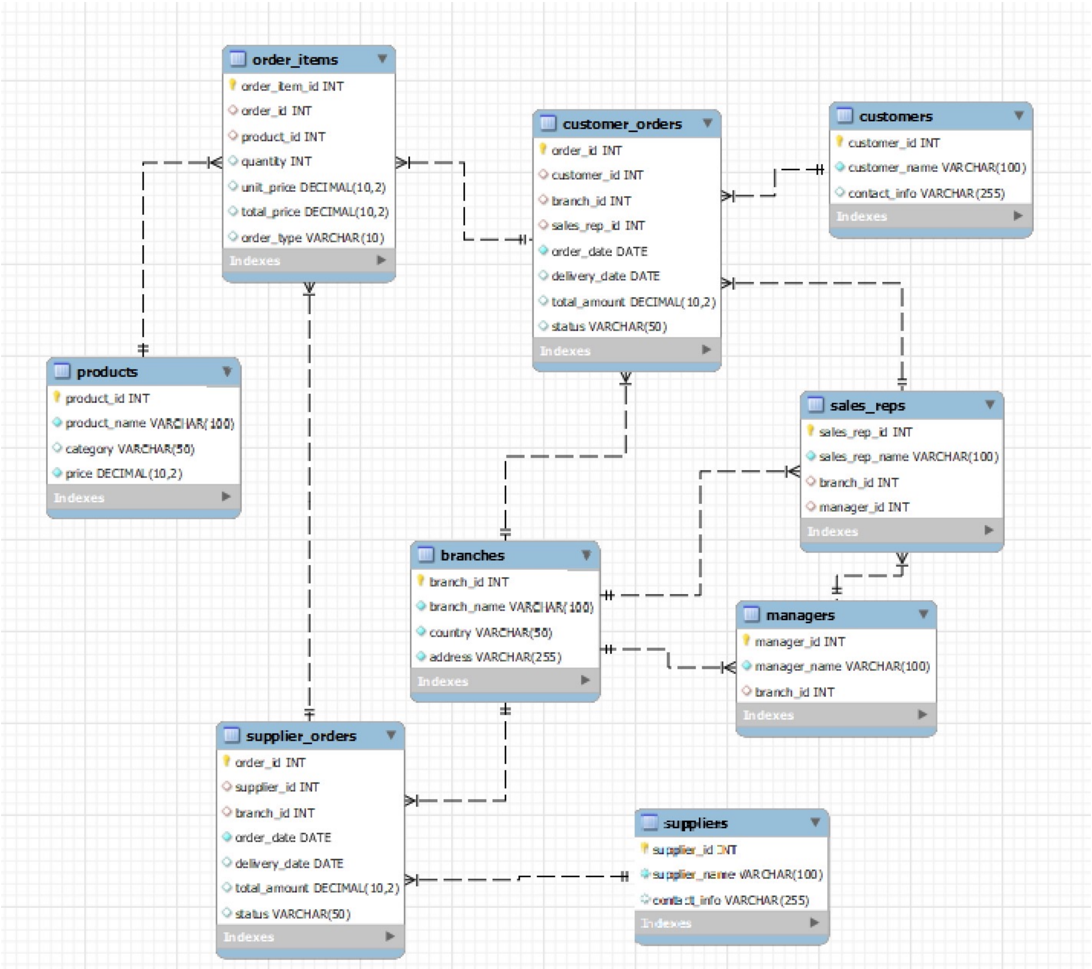
Appendix III

Branch Profitability Analysis

Period of current CEO leadership (2016 – LTM)

Branch	Period Date	Currency	Exchange Rate to USD	2000 31/12/2000	2001 31/12/2001	2002 31/12/2002	2003 31/12/2003	2004 31/12/2004	2005 31/12/2005	2006 31/12/2006	2007 31/12/2007	2008 31/12/2008	2009 31/12/2009	2010 31/12/2010	2011 31/12/2011	2012 31/12/2012	2013 31/12/2013	2014 31/12/2014	2015 31/12/2015	2016 31/12/2016	2017 31/12/2017	2018 31/12/2018	LTM 30/06/2019
A	Revenue	USD	1	4,640	4,286	3,538	3,384	3,631	3,816	4,458	4,680	4,806	3,896	4,109	4,784	4,818	5,037	5,321	5,438	5,683	5,902	6,410	6,638
	Profit			66	32	11	8	14	17	57	83	87	37	42	82	87	81	88	91	93	72	144	159
B	Revenue	USD	1	3,971	3,553	2,848	2,966	3,701	4,348	5,581	6,614	6,934	5,630	6,183	6,946	7,066	7,036	7,283	6,995	8,614	8,958	9,492	9,796
	Profit			79	30	43	42	78	90	209	254	196	-29	109	188	125	201	195	128	121	109	156	192
C	Revenue	USD	1	4,386	4,134	3,758	3,714	4,228	4,996	6,012	6,784	6,905	5,225	5,722	6,922	7,435	8,490	8,915	8,496	8,290	8,677	9,240	9,255
	Profit			33	20	23	30	65	104	217	241	213	105	115	196	202	277	275	208	101	163	225	230
D	Revenue	GBP	1.27	2,596	2,996	2,784	2,841	3,037	3,045	3,471	3,730	4,350	4,841	5,030	5,321	5,581	6,350	6,411	6,758	7,737	8,936	9,455	9,648
	Profit			138	150	160	160	163	158	164	165	181	189	202	157	248	263	268	296	338	394	415	422
E	Revenue	EUR	0.847	12,138	12,304	11,343	10,992	12,976	14,621	17,188	21,669	22,924	19,771	24,512	28,026	29,978	29,483	31,221	33,351	30,913	30,976	25,488	33,307
	Profit			692	-1,031	478	486	636	1,102	1,441	1,735	1,844	957	1,922	2,037	2,062	2,103	2,205	1,608	1,938	2,365	2,059	2,581
F	Revenue	EUR	0.847	3,855	2,692	2,415	2,898	3,171	7,719	9,682	9,166	9,131	9,502	9,964	10,531	12,261	12,993	13,495	16,221	14,761	15,169	14,887	21,411
	Profit			163	-15	82	128	144	-268	22	49	-208	700	241	532	1,419	1,514	-73	-403	2,098	5,191	1,136	2,397
G	Revenue	GBP	1.33	5,653	6,499	7,251	7,481	9,217	10,245	12,884	14,761	15,060	13,141	12,015	12,338	12,213	11,970	11,948	12,132	13,131	13,854	18,884	19,631
	Profit			194	185	288	298	397	461	537	474	74	-1,173	-340	271	57	305	504	213	650	783	1,267	1,395
H	Revenue	GBP	1.33	1,260	1,365	1,512	1,790	1,951	2,817	3,039	3,400	3,391	3,127	3,363	5,098	5,169	5,493	5,953	6,338	6,633	6,863	7,530	6,558
	Profit			74	97	117	138	165	179	212	235	129	200	179	270	330	336	329	213	18	297	-111	97
I	Revenue	JPY	0.0091	3,225	2,810	2,427	2,342	2,458	2,616	2,829	2,804	2,551	1,851	2,130	2,374	2,396	2,647	2,644	2,646	2,671	2,764	4,267	2,991
	Profit			29	-91	17	13	15	29	25	15	10	-34	21	26	34	35	47	45	49	52	82	58
J	Revenue	JPY	0.0091	1,125	1,080	1,075	1,131	1,193	1,266	1,416	1,452	1,330	1,179	1,317	1,395	1,529	1,811	1,855	1,938	1,871	2,000	3,118	2,204
	Profit			20	18	17	21	23	28	37	38	27	24	23	28	31	42	55	50	51	56	86	64
K	Revenue	AUD	0.75	348	310	372	433	510	589	649	763	838	879	876	911	885	895	1,035	1,215	1,327	1,416	1,754	3,185
	Profit			26	19	28	31	41	50	58	69	76	65	77	79	76	80	82	111	129	142	168	135
L	Revenue	JPY	0.0063	167	181	189	198	217	241	262	268	245	232	241	220	218	234	228	222	219	218	222	222
	Profit			6	7	9	10	12	16	17	17	14	12	13	11	12	13	12	11	11	13	13	13
M	Revenue	CNY	0.151	89	120	206	139	235	304	261	237	221	147	201	317	408	374	526	926	1,063	1,229	1,775	1,852
	Profit			8	4	-4	-16	0	3	1	-17	1	-21	-11	10	5	7	9	17	3	-16	4	8
N	Revenue	DKK	0.1582	na	na	na	na	na	na	na	na	na	na	na	na	na	na	1,349	1,393	1,457	1,461	1,650	1,494
	Profit			na	na	na	na	na	na	na	na	na	na	na	na	na	na	-33	23	18	3	21	12

Appendix IV

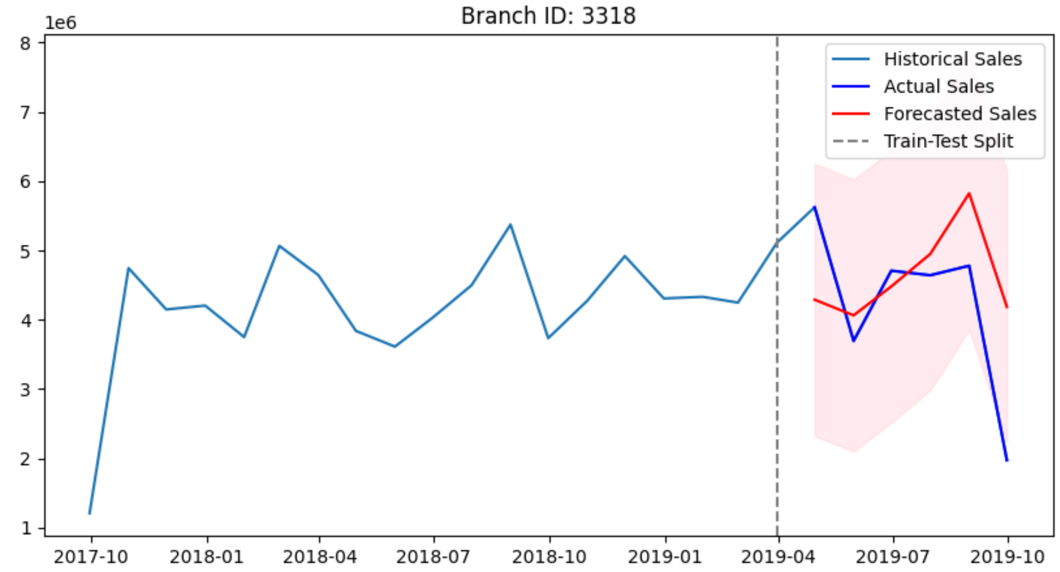


ERD Diagram for Inventory Management System with SQL

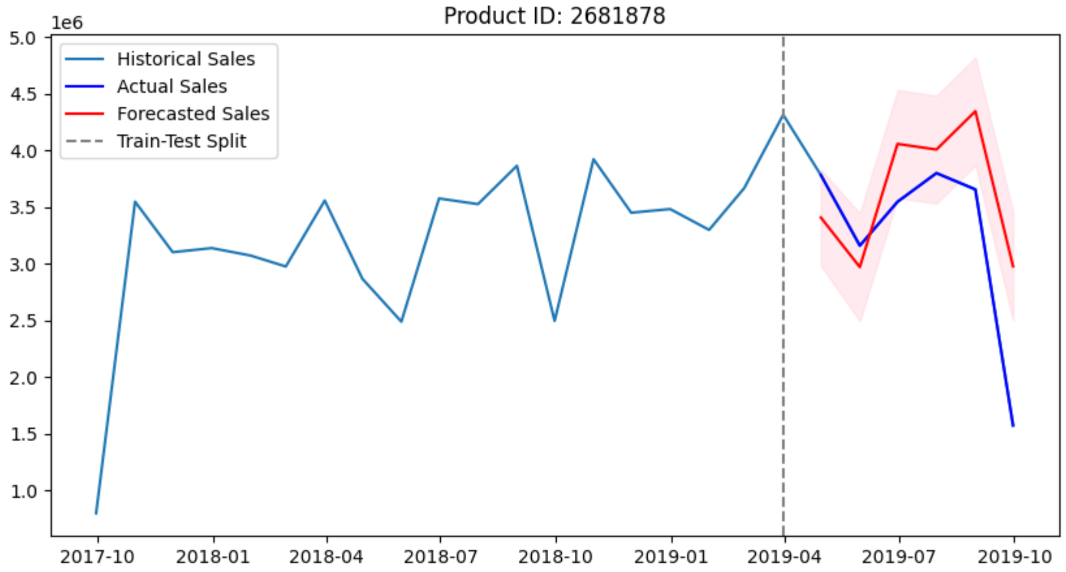
Appendix V

Additional Stock demand forecasting - Branches

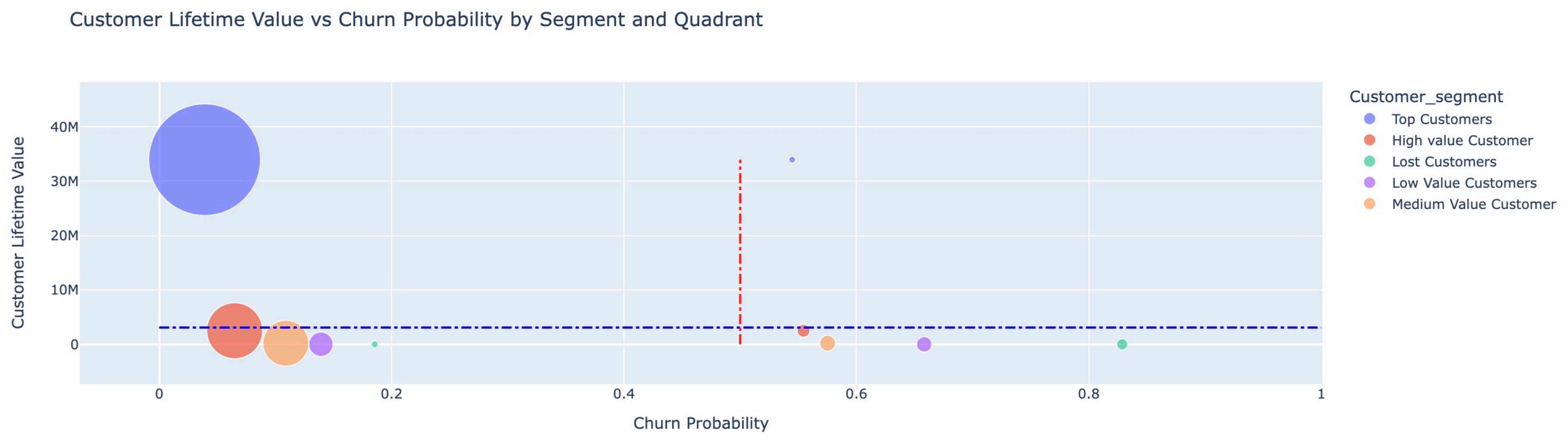
Monthly SARIMA – Branch 3318



Monthly SARIMA – Product 2681878



Appendix VI


Customer Lifetime Value vs Churn Probability by Segment and Quadrant
Customer Lifetime Value
40M
30M
20M
10M
0
0
0.2
0.4
0.6
0.8
1
Churn Probability
Customer_segment
Top Customers
High value Customer
Lost Customers
Low Value Customers
Medium Value Customer

Appendix VII

Association rules

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
(2734140)	(1390401)	0.164726	0.268743	0.107288	0.651311	2.423548	0.063019	2.097159	0.703220
(1390401)	(2734140)	0.268743	0.164726	0.107288	0.399221	2.423548	0.063019	1.390318	0.803249
(653092)	(1461750)	0.136726	0.147717	0.092765	0.678469	4.593036	0.072568	2.650702	0.906177
(1461750)	(653092)	0.147717	0.136726	0.092765	0.627989	4.593036	0.072568	2.320562	0.917863
(1180925)	(1438570)	0.088447	0.097344	0.081643	0.923077	9.482630	0.073034	11.734528	0.981341

Rules ordered by confidence

antecedents	consequents	confidence
2732769	1438570	1.00
2732769	1862474	1.00
915256	1180925	1.00
2995747	1862474	1.00
1461750	517859	1.00
915256	1862474	1.00
2564840	1438570	1.00
1745137	1390401	1.00
2564840	1862474	1.00
1390401	1862474	1.00

Expected return

antecedents	consequents	expected_return
11100	2008648	35.21
59473	2198494	3.16
378845	517859	6.96
417321	2008648	45.48
463156	572469	21.52

Output of cross-selling on sample

antecedent	consequent	potential_revenue_increase
1390401.00	572469.00	23270.28
653092.00	11100.00	13612.92
2734140.00	572469.00	13011.69
2008648.00	11100.00	8024.46
1325810.00	572469.00	7829.59
2542409.00	2077753.00	7662.53
517859.00	572469.00	5244.01
463156.00	572469.00	4518.98
417321.00	2008648.00	4366.34
2564840.00	2077753.00	4266.64