Buying Channels strategy:

Advanced data analytics for procurement efficiency

by

Foyinsola Tolulope Adeyemi
MSc, Civil Engineering, University of Lagos

and

Fabrizio Umberto Boaron
MSc, Business Administration, Università Bocconi

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Signature of Author: ____

Foyinsola Tolulope Adeyemi
Department of Supply Chain Management
May 10, 2024

Signature of Author: ____

Fabrizio Umberto Boaron
Department of Supply Chain Management
May 10, 2024

Certified by: ____

Dr Maria Jesus Saenz
Executive Director, MIT Supply Chain Management Master’s Program
Director, MIT Digital Supply Chain Transformation Lab
Capstone Advisor

Accepted by: ____

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Buying Channels strategy:
Advanced data analytics for procurement efficiency

by

Foyinsola Tolulope Adeyemi

and

Fabrizio Umberto Boaron

Submitted to the Program in Supply Chain Management
on May 10, 2024 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

The procurement organization, which manages up to 80% of total organizational spend, is increasingly recognizing the importance of leveraging data to enhance its processes. A buying channel in the procurement function is the end-to-end series of steps to request, approve, purchase, receive, and pay for goods and services. This capstone project explores the procurement buying channels processes in the sponsor organization, an S&P500 company operating in the pharmaceutical sector, using advanced data analytics to provide insights on variables that influence the use of channels.

Our analysis included analyzing the behavior of the stakeholders in the buying channels processes through clustering of the data using a K-Means algorithm. Clusters were added to existing and newly engineered features in the dataset. Machine learning models for regression and classification were then developed, to identify key variables impacting the use of buying channels and to predict buying channel utilization. The results show that there is a confusion in channel usage due to lack of clarity in the classification of channels. Also, predicting the completion time of purchases through the channels had low accuracy due to lack of granular information on transactions.

Our recommendations include a new framework articulated over three pillars: enterprise-wide structured and unstructured data integration, a Co-Pilot architecture to support stakeholders through generative AI applications and performance incentives for ongoing learning. Efficiency will be driven by a feedback loop within the buying channels process to incorporate insights gained from each completed transaction to inform future transactions.

Capstone Advisor: Dr Maria Jesus Saenz
Title: Executive Director, MIT Supply Chain Management Master’s Program
Director, MIT Digital Supply Chain Transformation Lab

ACKNOWLEDGMENTS

We would like to express our gratitude to the following people for the support we received to make this capstone project successful.

Dr. Maria Jesus Saenz, Executive Director, MIT Supply Chain Management Master's Program, Director, MIT Digital Supply Chain Transformation Lab, and Capstone Advisor, for her support, valuable guidance, and feedback throughout the project.

Dr. Devadrita Nair who patiently helped with insights on the machine learning algorithms.

Toby Gooley, Writing Coach, Writing Department at MIT, for her patience and for carefully reviewing the many versions of the report to ensure a great outcome.

The sponsor company, for the people who dedicated time to provide information and support, and those who gave guidance and feedback to make the project meaningful.

I would like to express my deepest gratitude to everyone who has supported and encouraged me throughout this journey, and to my family and friends, whose unwavering belief in me have been a constant source of strength. Your encouragement has made this achievement possible. A special thanks to my beloved children, Iacopo, Nicolò, and Rodrigo. Your patience, love, and joyful spirits have been my greatest motivation and inspiration. This work is dedicated to you, with all my love.

Fabrizio U. Boaron

I am thankful to the One who made this possible. My special appreciation goes to my husband, Damilola, for his constant love and support. I would also like to thank my children Adeife, Adeiyi and our baby who came along with me to study, thank you all for your patience throughout the journey. I am grateful to my siblings, parents, in-laws, dear friends, and to everyone who encouraged and supported me in the journey towards the successful completion of this project.

Foyinsola T. Adeyemi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Procurement in organizations is tasked with taking on increasingly strategic roles. The consequence is the need for effective utilization of data both to make purchasing decisions and to harness available resources and tools, to ensure maximum benefits are derived from the procurement processes. The procurement organization must build, maintain, and improve its purchasing processes to satisfy the internal customers and ensure that the outcomes of these process improvements fit perfectly with the aims of the organization, its customers, and its prospects.

A key area for achieving this business objective is efficient utilization of buying channels. A buying channel is the end-to-end series of steps to request, approve, purchase, receive, and pay for goods and services. This customer-oriented concept identifies customers' purchasing needs and adapts the resources of the organization to deliver need-satisfying goods and services. Buying channels are composed of different mixes of service offerings—including requisitioning (i.e., tracking the procurement request), buying (choosing a supplier), new supplier onboarding (if needed), contracting (across different types of agreements, as required), and invoicing and payments.

1.1 Background and motivation

The sponsor for this project is an S&P500 company operating in the pharmaceutical sector. In regard to buying channels, the sponsor company defines customers as those employees who are authorized to initiate purchase requests to the procurement organization on behalf of business units. The business units represent different functions of the business, for example Commercial or Manufacturing, where the need for a material or service originates. The critical stakeholders are therefore the procurement function, suppliers, internal customers, and business units.

The procurement taxonomy in the sponsor company is set up to define every purchase that a customer can request from the procurement function, classified into family, categories, subcategories, and commodities. A purchase request is initiated for goods or services through an online catalogue. However, if the item is unavailable in the catalogue, a purchase request form is chosen from several possible purchase request forms, filled, and passed on to the procurement organization. In some cases, the customer chooses the appropriate form, and the purchase is routed through the right channel, but in other cases, the purchase request is redirected to another buying channel based on the discretion of the procurement organization, to ensure efficiency and compliance with company policies, among other reasons.

These purchase request steps by the internal customer influence what buying channel will be used, with each channel having a different sequence of processes and approval levels. On one extreme are highly automated and standardized channels (“Quick Pass’’ with “No Touch Processes”), where transactions happen quickly, and a purchase order automatically goes to a vendor. On the other extreme are traditional manual channels, requiring human intervention on multiple levels (“End-to-End Procurement” with “High Touch Processes”). In such cases, first, sourcing will try to match the need to an existing supplier based on sourcing knowledge and experience of the category. If no suitable supplier exists, or if the customer insists, sourcing will initiate the process to find a new supplier, or at the very least add a specific supplier that the customer requires.

Since 2018, the sponsor company has been restructuring and rationalizing its procurement processes through the introduction of new buying channels and its procurement organization is motivated to understand the variables that influence the use of these buying channels. The focus has been to improve the company's internal customers’, procurement associates’, and suppliers’ experiences in managing transactions, with the result of freeing resources for more strategic work and increasing efficiency. Efficiency in the context of this capstone project describes purchase requests going through highly automated channels – reducing the number of “touches” by operators rather than the traditional manual channels, therefore reducing costs and time in the procurement process.

1.2 Problem statement and research questions

With more robust buying channels in place, the procurement organization is motivated to understand what influences the use of each channel. Hence, the sponsor company’s key objective is to:

1. Develop a methodology to understand the tendency of customers to utilize certain channels for various categories of purchases and which independent variables influence the adoption of buying channels with the aim of improving channel planning process.
2. Understand what the independent variables that drive cycle time are, once a transaction is in a buying channel.

1.3 Project outcomes

The dataset analyzed refers to North America, where the sponsor company offers a full range of buying channel services. The exploratory analysis includes understanding the types of buying channels, scope of channels, and channel trends. In addition, we hypothesized that it will be possible, through advanced analytical models (prediction models in machine learning), to identify which independent channel variables are likely to be predictors of buying channel utilization and to test the accuracy of those models. Therefore, the deliverables to the sponsor company include:

1. A descriptive analysis of current buying channels planning processes and utilization.
2. Exploratory analysis, looking at data for variability and relationships among the variables.
3. Clustering of variables to understand the characteristics of relevant stakeholders in the buying channels process.
4. Prediction of channel utilization by exploring the data with different machine learning models, identifying the most influential variables on the performance of the machine learning models.
5. Managerial insights and recommendations, together with a framework, with potential opportunities to improve efficiency and improve internal customers' experience.

2 State of the Practice

In this chapter we study the state of the practice, and in particular critical areas of focus to evaluate the role of buying channels analytics in increasing efficiency in the procurement process:

- Procurement planning and organizational spend
- Segmentation and category management in procurement
- Buying Channels: utilization, user adoption and efficiency
- Buying Channels in the sponsor company
- Advanced data analytics to aid buying channel strategies

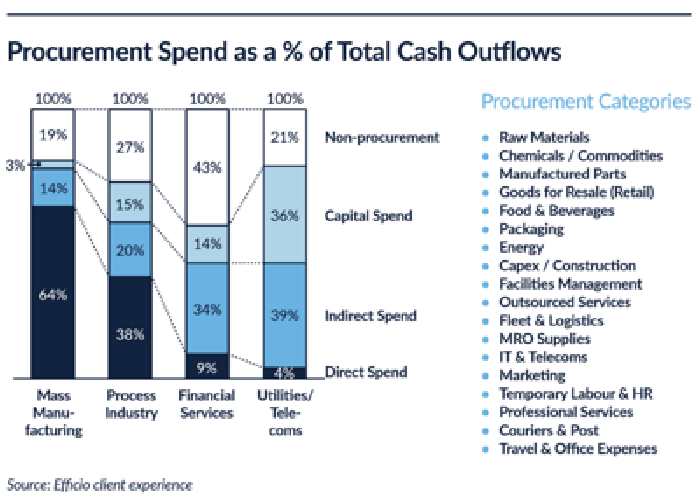
2.1 Procurement planning and organizational spend

Procurement is an essential business function that involves various activities across the organization to buy goods and services required by various business units. It also serves as an important source of competitive advantage to organizations (Novak & Simco, 1991). Where purchasing functions align with the business strategy, are well trained, and achieve planned outcomes in relevant areas, they present significant contributions to the overall business performance (Gonzalez-Benito, 2007).

A key contribution of procurement is the visibility and control of organizational spend. Cost reduction and optimization are recurrent in literature as pillars of procurement function transformation. According to BCG (2021), cost visibility and control are the first steps in driving transformation of procurement. Cost management by the procurement function can be an important driver of competitive advantages for organizations, as procurement is responsible for up to 50% to 80% of a company's total spend across several categories, business units, geographies, and budget lines. Figure 2.1 illustrates the magnitude of spend that the procurement function oversees across several industries. This presents a huge opportunity for cost reduction and optimization (Klein et al., 2001).

Figure 2.1

Spend of a typical manufacturing company, with revenues indexed to 100



From "Profit from Procurement," by A. Klein, S.Whatson, and J.Oliveira, 2021, Wiley, (https://learning.oreilly.com/library/view/profit-from-procurement/ 9781119784739/).

The sponsor company has initiated a procurement transformation program since 2018 by introduction of buying channels, and further seeks advanced planning and analytics methodologies to improve this important function. Given the wide range of products and customers with different requirements, a one-size-fits-all approach to drive improvement in procurement can lead to high costs and poor service. Differentiated approaches are therefore necessary (Protopappa-Sieke & Thonemann, 2017).

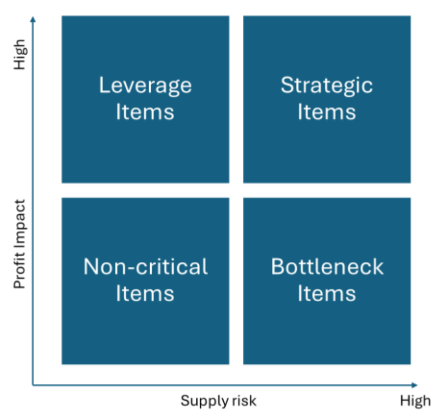
2.2 Segmentation and category management in procurement

Segmentation is the process of grouping a combination of channels, customers, or products that have similar requirements, patterns, and characteristics (Nicoletti, 2018). Understanding supplier segmentation, category management, and process segmentation provides insights on design and standardization of processes, by simplifying the complex procurement matrix of a company.

Supplier segmentation has many benefits including cost and process design efficiencies, and an improved capability to innovate products and services, as it is not practical to dedicate the same amount of time and resources with every transaction (CIPS, 2023b). The most common methodology for supplier segmentation is the Kraljic matrix (Figure 2.2), plotting Suppliers across value and amount of spend.

Figure 2.2

The Kraljic matrix



Adapted From "Purchasing Must Become Supply Management," by P.Kraljic, 1983, Harvard Business Review (https://hbr.org/1983/09/purchasing-must-become-supply-management).

Category management in procurement is defined as "The practice of segmenting the main areas of organizational spend on bought-in goods and services into discrete groups of products and services according to the function of those goods or services and, most importantly, to mirror how individual marketplaces are organized" (O'Brien, 2019). The ultimate objective of category management is to aggregate similar categories of spend to create value for the organization (Toikka, 2023). This has several advantages as it provides clarity on what is bought, in what quantities, from whom and at what price. It also aids understanding, articulation, influence and optimization of business requirements (Klein et al., 2001).

2.3 Buying Channels: utilization, user adoption, and efficiency

Procurement activities have a reach across various functions within the organization. It is therefore important to identify the internal customers and stakeholders necessary to develop comprehensive purchasing solutions. When internal customers do not comply with the standard systems, processes, and preferred suppliers' setup up by procurement, there is a resultant leakage of value and lost savings (Karumsi & Prokopets, 2021). Engaging them right from the beginning of processes, ensuring completeness of feedback and securing appropriate buy-in is key to the success of the procurement function (CIPS, 2023).

In Section 1.1, we define buying channels as the end-to-end series of steps to request, approve, purchase, receive, and pay for goods and services. It is a customer-oriented concept which identifies the wants, needs, and desires of internal customers and adapts the resources of the organization to deliver the required goods and services. Business to Consumer (B2C) retail principles, which are user-centric and focus on customer behaviors, can be adapted by procurement into the Business to Business (B2B) space to set up buying channels that can ensure purchases are aligned with business objectives while providing solutions that meet the needs of the customer (Karumsi & Prokopets, 2021). The use of electronic commerce, which includes e-procurement software, e-catalogues and ultimately virtual catalogues that are intelligent, dynamic, active, and capable of learning, can be deployed for companies to benefit from opportunities presented by adopting B2C principles (Baronet al., 2000). Also, a review of the organizational goals can help procurement find the most optimized buying channel for each of its categories. Factors such as pre- and post-purchase visibility, cost per transaction and category characteristics may influence the determination of the ideal channel strategy, after which the technology needed to execute the channel strategy can be put in place (Karumsi & Prokopets, 2021).

In Table 2.1, SAP (2023) lists the state of the practice metrics for evaluating and measuring the efficiency and impact of buying channels on procurement. Spend analysis, measured for example by the amount of categorized and uncategorized spending, is a metric for assessing the difference between planned and realized transactions. Those differences could be due to ease of use of the channel or to unauthorized spending. Key metrics used by the sponsor company in assessing buying channels are cycle time, and level of resource engagement ("Touch" / "No Touch").

Table 2.1

State of the practice procurement buying channels metrics

Increased Spend Visibility and Tracking	Provide visibility and real-time monitoring of spend data influence future sourcing and category management decisions.
Efficiency in Transactional Processes	Drive efficient processes through automation of low- and no-touch engagements and invoice auto matching.
Enhanced User Experience	Offer better user experience for each stakeholder involved in the transactions and promotes process adoption and adherence.
Compliance	Confirm the right vendors, spend category, and prices are used through procurement policy enforcement that drives compliance in the purchasing process and mandates relying on contract and line-item pricing.
Governance and Control	Through clearly laid-out requisition and approval processes, buying channels allow control on spend while standardized processes increase governance and helps reduce financial, process, and product risks.

From "Buying Channels: The Key to Frictionless Procurement," by SAP, 2023 (https://sapinsider.org/articles/buying-channels-the-key-to-frictionless-procurement/).

2.4 Buying Channels in the sponsor company

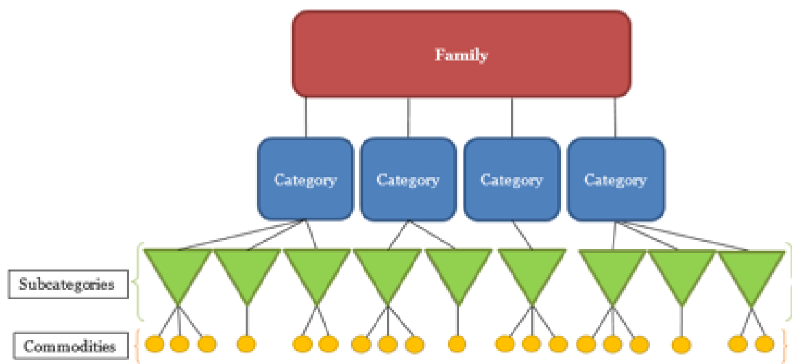
Here we describe the structure and process of in-house buying channels as obtained through interviews with stakeholders and internal documents made available by the sponsor company.

2.4.1 Taxonomy

Requests for purchase of items or services commence in the in-house online marketplace which is embedded in the organization's Enterprise Resource Planning (ERP) system. To understand where requests for purchases originate and what is being requested, the taxonomy in the ERP system is grouped into different levels of granularity, beginning with "Commodity," which is the most granular of the descriptions of items being requested. The Commodity group is then aggregated into increasingly less granular groups, i.e., Subcategories, Categories, and Family, in that order as shown in Figure 2.3.

Figure 2.3

Taxonomy of goods and services



The procurement process in the sponsor company involves several key stakeholders:

1. Requester: The individual who identifies the need within a business unit and owns the purchase request.
2. Preparer: Has access to the ERP and goes into it to initiate the requisition on behalf of the Requester. Requester and Preparer could be the same person.
3. R2P (Request to Pay): The procurement team which supports the business unit Requesters and Preparers through interfacing with Suppliers and other stakeholders to ensure Purchase Orders are created and requests are delivered in accordance with planned timelines. This is also referred to as the triaging team. They redirect requests through the most efficient channels and identify the right stakeholders to get involved.

4. Sourcing and Contracting: The sourcing and contracting teams are involved in the process when items requested are not available in the ERP or within the existing supplier base and would need to be obtained from new sources.
5. Supplier: The individual or entity who provides goods or services at a fee to the organization.

2.4.2 Process flow

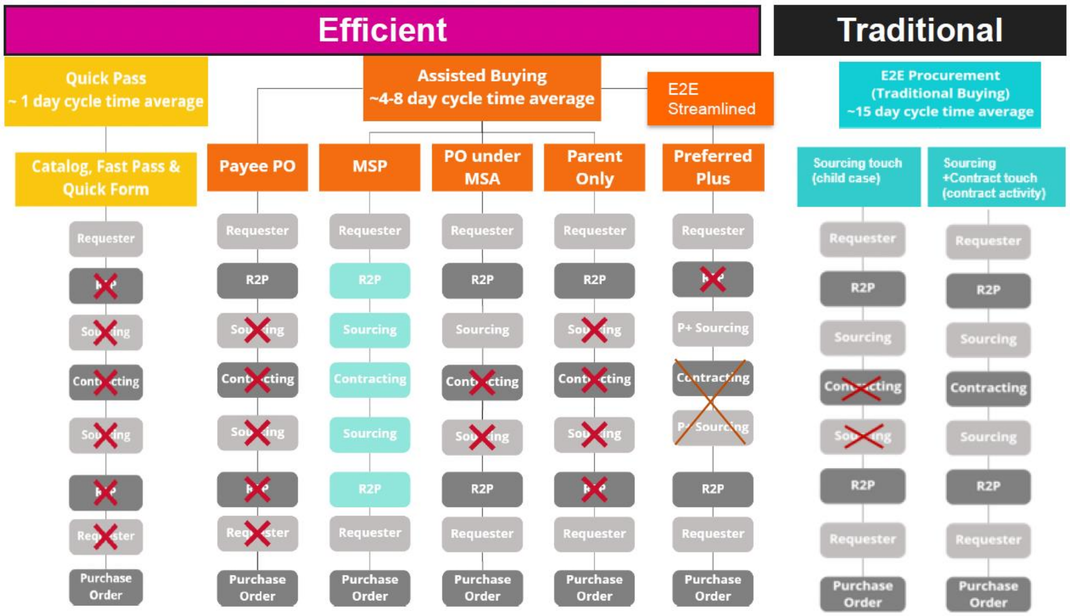
The process begins when a need is identified within a business unit. The Requester or Preparer, as the case applies, then goes into the online marketplace to select the item from the catalogue hosted on the ERP system, searching the catalogue using a keyword. The item can either be found in the hosted catalogue or the Requester will be redirected to an external catalogue. If the items are not available in the catalogue, an automated "shopping assistant" helps the buyer find the right intake form so that the request can be passed on to the next stakeholder. The different intake form types inform what buying channels will be used, making filling of forms a key step in the process. A process map showing the process flow is shown in Appendix A.

2.4.3 How efficient and traditional channels are defined within the sponsor company

Generally, purchases are regarded to have been handled through an efficient buying channel if they go through a less complex process, with minimal to no touch from the R2P, Sourcing, and/or Contracting teams. Channels within the sponsor organization are therefore classified as Efficient or Traditional as shown in Figure 2.4. Traditional channels require a broader set of End-to-End (E2E) procurement processes, which results in the use of more time and resources.

Figure 2.4

How are Efficient & Traditional Buying Channels defined?



From "Efficient Buying Channels," sponsor company, internal document

Preparers make decisions on the correct form to use by considering specified criteria, including commodity code, spend threshold, availability of quotes, supplier status (if the supplier is registered already in the company database), regulatory risk impacts, and compliance with company policies. Buying channels based on granularity are grouped into Level 1 consisting of 5 channels, Level 2 with 13 channels and Level 3 with 30 channels, and are classified after the transaction has been completed.

When applying machine learning techniques, we observed that Level 2 had the best balance between descriptive power and accuracy and therefore focused the analysis on predictions for Level 2 channels described in Table 2.2. From this point onward, we will refer to those channels as L1, L2 and L3.

Table 2.2

Description of Level 2 (L2) Buying channels

Classification	L2 Channel	Description
Efficient	Catalog	Items have clear specifications and pricing and are available in the in-house online marketplace
Efficient	Fast Pass	Transaction is low-value and low-risk, contracts are in place, supplier is registered and has given a quote
Efficient	Quick Form	Request is with selected preferred suppliers using a proposal/quote obtained upfront
Efficient	Payee-PO	A transaction with a non-commercial supplier such as health care professional, government institution, etc.
Efficient	Managed Service Provider	Authorized service providers who may carry out end-to- end transactions on behalf of the organization
Efficient	Parent Only	Specific to some categories where R2P touches the transaction but does not carry out sourcing or contracting
Efficient	Preferred Plus	Used for recurring and well-defined services with known suppliers and agreed pricing.
Efficient	E2E Streamlined	Used for recurring and well-defined services with known suppliers and agreed pricing.
Traditional	Spot Buy	Low value but still requiring sourcing and/or contracting activities
Traditional	eMP Exception	Misclassification
Traditional	Non-eMP Import	Involves strategic procurement and importation
Traditional	Child Case	Requires sourcing
Traditional	Child Case and Contract Activity	Requires sourcing and contracting

2.5 Use of advanced data analytics to aid Buying Channel strategies

Data analytics is often connected with several disciplines including artificial intelligence, machine learning and deep learning (Mandl, 2023). Advanced data analytics plays a vital role in helping to discover meaningful information from available raw data. Analytics and state-of-the-art technologies are therefore central to enabling procurement organizations to focus on how to optimize enterprise spend rather than on time-consuming approvals and reporting. Four common categories of advanced data analytics are: descriptive, diagnostic, predictive and prescriptive analytics.

According to PWC (2022), machine learning algorithms can support decisions on buying channel strategies to meet the unique needs and patterns of the customer. Additionally, through assessing past performance, cost savings, delivery times, and supplier reliability, supervised machine learning algorithms can make predictions based on labelled data on the best channels for specific procurement needs. Classification algorithms assign data to categories, simplifying the analysis of large datasets by clustering data in groups with similar features. Regression algorithms highlight the relationship between dependent and independent variables.

In cases where there are large datasets with numerous characteristics, two especially successful algorithms are (The Tech Spark, 2023):

- *Random Forest.* It is an ensemble technique which aggregates the results of random trees to create predictions. Random Forest can be used with both categorical and continuous data and are especially used for large datasets with numerous characteristics.
- *Gradient Boosting.* Similarly to Random Forest, Gradient Boosting is another ensemble technique, and it can handle both classification and regression.

Random Forest and Extreme Gradient Boosting (XGBoost) are especially effective for large datasets with numerous features, such as the case of this project, due to their ability to handle high dimensionality, reduce overfitting, provide intuitive outputs for feature importance, and handle missing data in the dataset as well as and their scalability and efficiency.

Advanced data analytics can be combined with user friendly interfaces. For example, those powered by large language models (LLMs) can improve data access and analysis, provide recommendations and insights, and overall improve the customer's experience, which is one of the objectives of the sponsor company for this project. Deloitte (2023) highlights the transformative impact of LLMs and Generative AI, which support the streamlining of processing, analyzing extensive datasets and simplifying intricate manual operations. As will be discussed in section 6.1.2, our proposed framework includes the use of LLMs to access large amounts of structured and unstructured data.

3 Methodology

After a review of the state of the practice for buying channels and interviews with internal company stakeholders, we discuss in this section the methodology employed to achieve the project's key objectives. The objectives include understanding which customers make use of specific channels, identifying variables that influence channel adoption, and predicting the use of buying channels.

3.1 Method overview

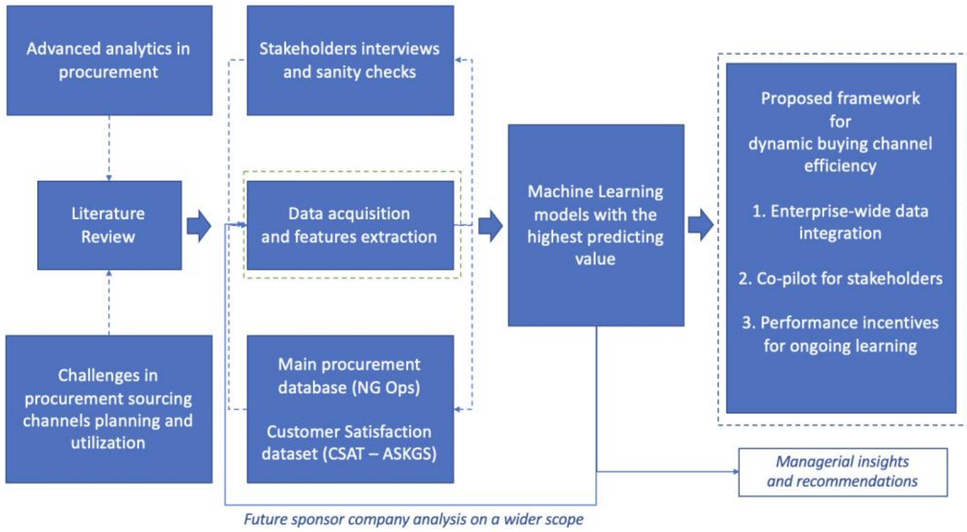
We conducted an analysis of available data on the use of buying channels, leveraging an extract from the sponsor company's primary procurement database known as NG Ops, which includes detailed transaction records. Additionally, we examined the Customer Satisfaction Dataset (CSAT) database, which contains satisfaction surveys completed by system users who encountered challenges while navigating purchase requests. Following stakeholder interviews, we identified key variables to focus our analysis and cleaned the data to remove duplicates and outliers. This resulted in a descriptive analysis and mapping of the current structure of buying channels within the organization.

Relevant features (independent variables) were then identified and engineered from the dataset, and these, along with the categorical (buying channels) and numerical (cycle time) dependent target variables, were used in our models. Next, we established preprocessing pipelines and using machine learning techniques such as Random Forest and XGBoost, which we previously identified as especially effective for large datasets with numerous features, we analyzed the data. This allowed us to pinpoint variables with the most significant impact on the use of buying channels and purchase completion time (cycle time). Subsequently, we modeled the influence of these variables, predicting which channels purchase requests would be routed through, how these channels interact, and the expected turnaround times of the purchase process for requests routed through each channel (cycle time).

Finally, we evaluated the models based on predetermined metrics and summarized our findings, providing a proposed framework for dynamic channel efficiency, together with insights and recommendations based on our analysis as evidenced in Figure 3.1.

Figure 3.1

Methodology



3.2 Data extraction and preprocessing

In this section, we will look at the available data, and steps taken to prepare the data for further analysis.

3.2.1 Data sources

Data was provided (as described in Section 3.1) in two comma-separated values (CSV) files, detailing five years of records of purchase transactions, which were imported using the Pandas data analysis library into Python programming language data frames:

- NG Ops [1,387,704 rows × 171 columns]
- Customer Satisfaction Dataset (CSAT) [133,289 rows x 13 columns]

We have been supported in the interpretation of the data by a limited number of interviews with stakeholders.

3.2.2 Data selection

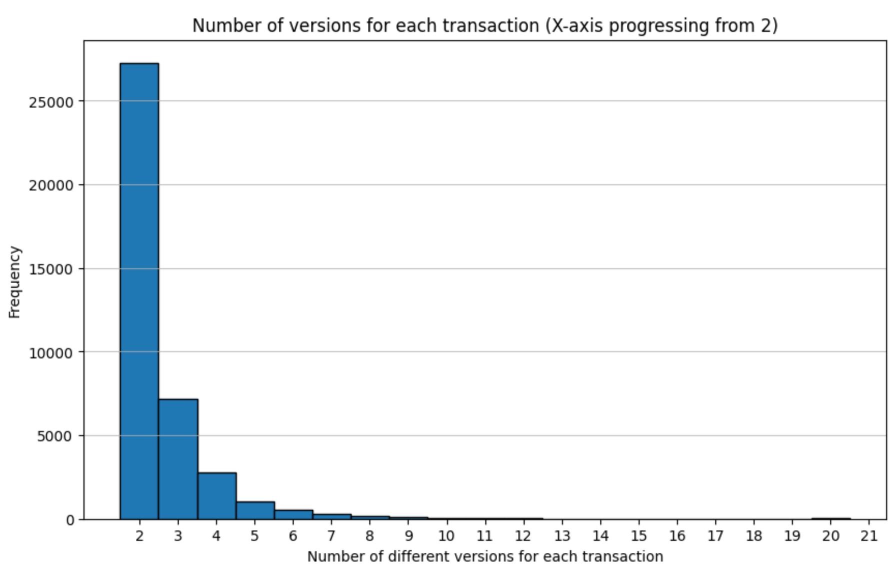
After a high-level data exploration, we identified columns that were relevant to the use of buying channels as listed in Appendix B along with the data types and descriptions. These two data frames were then merged into one by looking up the average Net Promoter Score (a metric used to measure the user satisfaction) of Preparers into NG Ops data frame to create a foundational database for further analysis.

3.2.3 Data cleaning and exploration

The procurement service cycle time keeps a record of timestamps for purchase requisitions routed through R2P/Sourcing/Contracting, which is the focus of this capstone project. Hence, entries with procurement service cycle time as zero or null, were taken out of the data. Cycle time was then converted to its log form. Rows containing duplicate values for versions of purchase requisitions were dropped. Several transactions had repeated versions for the same request, and these were filtered out to get more insights on the trends for unique requests. Figure 3.2 shows details of the number of versions created for transactions with greater than one version.

Figure 3.2

Histogram of number of versions for each purchase requisition



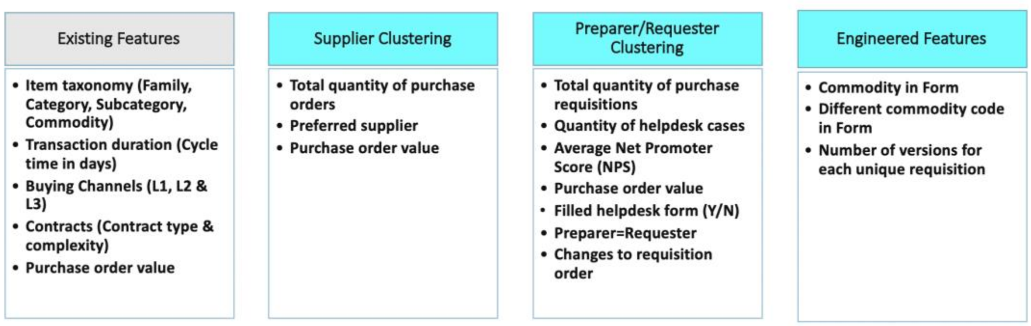
Over 15% of requests exhibited some form of modification, indicating initial mistakes in how the transactions were instructed. All intermediate versions of the requests were removed from the dataset, keeping only the last version. The final data frame used in the analysis consists of approximately 200,000 transactions.

3.2.4 Features

We have identified several existing features in the data that we found could explain cycle time and buying channel utilization and that could be used in the machine learning models. However, we added several other variables through feature engineering, both through clustering and through extraction of features from other existing features as described in Figure 3.3.

Figure 3.3

Data features



3.2.5 Clustering

Clustering was done primarily through the unsupervised machine learning method of K-Means clustering and segmentation to create the following features: Supplier clusters, Preparers clusters and Requester clusters.

3.2.6 Target outcome variables

We hypothesized in Section 1.3, that several independent variables have influenced the use of buying channels and could result in variability of buying channel properties such as cycle time. To evaluate the predicting value of those independent variables, we focused the models on two major target outcome variables: buying channel utilization and cycle time.

3.3 Machine learning models

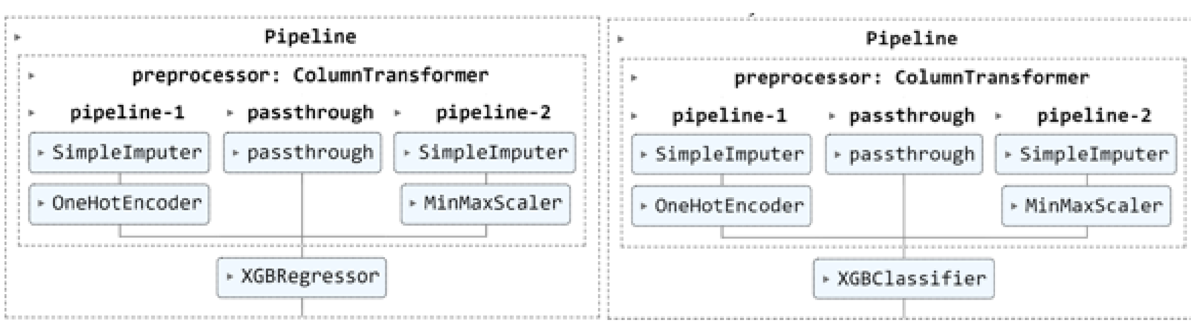
Here we provide a description of the machine learning models and the methods for evaluating the performance of the models.

3.3.1 Model preprocessing pipeline

On completion of problem definition and review of the state-of-the-art methodologies in machine learning and analytics, we examined different models that are useful in prescriptive analysis. After testing a number of models, including Random Forest and Neural Networks, we have focused our analysis on Extreme Gradient Boosting (XGBoost) with the model preprocessing pipeline as shown in Figure 3.4. The algorithm was designed to split the data set into two, the training and testing splits: the training split to learn properties on the data and the testing split to be the reserved dataset on which to test the learned properties.

Figure 3.4

Model preprocessing pipelines



3.3.2 Model evaluation measures

In order to evaluate the performance of the models we have identified specific metrics for both regression and classification tasks. For regression tasks, key metrics include (Mandl, 2023):

- Mean Absolute Error (MAE): It measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.
- Mean Squared Error (MSE): It measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. MSE is more sensitive to outliers than MAE as it squares the differences.
- Root Mean Squared Error (RMSE): This is the square root of the mean of the squared errors. RMSE is even more sensitive to outliers than MSE and provides a measure in the same units as the response variable.
- R-squared (R^2): It provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance.

For classification tasks, key metrics include (Mandl, 2023):

- Accuracy: It measures the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined. It is useful when the target classes are well balanced.
- Precision: It is the ratio of correctly predicted positive observations to the total predicted positives. High precision relates to a low rate of false positives.
- Recall (Sensitivity): It is the ratio of correctly predicted positive observations to all observations in actual class. It relates to the ability of a model to find all the relevant cases.
- F1 Score: The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is particularly useful when the classes are imbalanced.

Confusion matrices are widely used in machine learning for evaluating the performance of classification models. The confusion matrix function evaluates classification accuracy by comparing the actual classes versus predicted classes of all the instances of the prediction. By definition, entry i & j in a confusion matrix is the number of observations actually in group i, but predicted to be in group j (Pedregosa et al., 2011). We have used confusion matrices to better visualize the above classification metrics.

4 Results

In this chapter, the results obtained from clustering of Suppliers, Preparers and Requesters are presented. This is followed by the prediction models for cycle time and buying channels based on an evaluation of the confusion matrix, gain plots, and accuracy of the models.

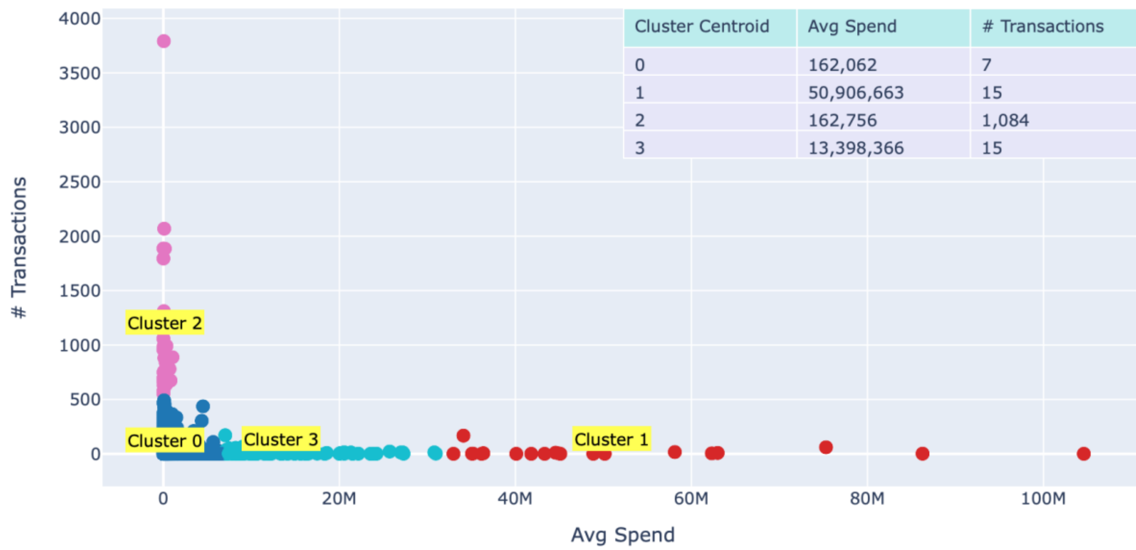
4.1 Clustering

Using a K-Means algorithm, we created clusters of three groups of stakeholders: Suppliers, Preparers and Requesters. Data was clustered with two variables: total number of transactions and average spend per transaction for everyone in each of the groups. We calculated the Silhouette Score for 2 to 20 clusters for each of the groups and noticed that the score monotonically decreases as the number of clusters increases, to between 0.98 to 0.75 for Suppliers and between 0.98 and 0.65 for Preparers. We have decided to use four clusters for all three groups rather than two due to the better descriptive power of the results and the slightly improved evaluation metrics in the models.

Supplier clustering produced four clusters, as shown in Figure 4.1. Cluster 0 with low average spend and low volume, Cluster 1 showed high average spend and low number of transactions. Cluster 2 Suppliers on the other hand, had a relatively high number of transactions with low value spend per transaction and Cluster 3 had medium spend values and low number of transactions.

Figure 4.1

Descriptive Analysis: Suppliers Clustering, with centroids and clusters descriptions

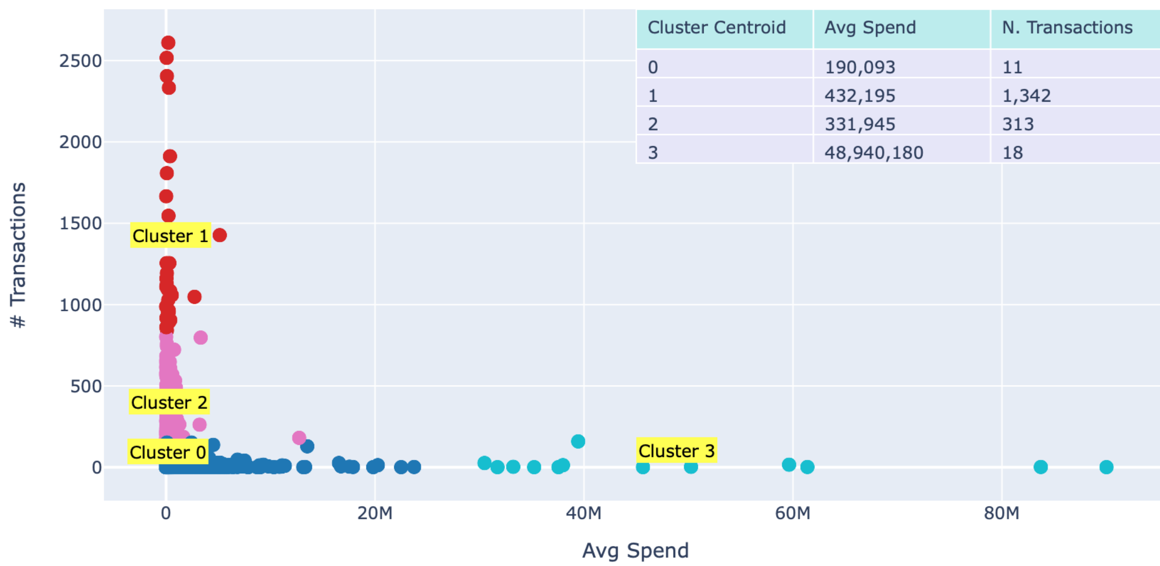


Supplier Cluster	# Transactions	Averange Spend	Suppliers
0	194,754	162,062	26,244
1	298	50,906,662	20
2	30,359	162,756	28
3	1,217	13,398,365	83

Preparer Clustering, as visible in Figure 4.2, showed most transactions in Cluster 0, with low to medium spend value and low number of transactions per Preparer. Cluster 1 showed a high number of transactions with low average spend, while Preparers in Cluster 2 created a moderate quantity of requisitions having equally low dollar values. Cluster 3 Preparers fell into the category of few transactions with relatively high average spend value. Clustering of Requesters showed patterns compared to Preparer clustering.

Figure 4.2

Descriptive Analysis: Preparers Clustering, with centroids and clusters descriptions



Cluster Centroid	Avg Spend	N. Transactions
0	190,093	11
1	432,195	1,342
2	331,945	313
3	48,940,180	18

Preparer Cluster	# Transactions	Average Spend	Preparers
0	114,176	190,092	10791
1	38,917	432,195	29
2	73,306	331,944	234
3	229	48,940,180	13

Appendices C and D show details of the spread of Requesters and Suppliers across the various L3 channels.

4.2 Machine learning models: Buying Channels

Extreme Gradient Boosting (XGBoost) Classifier was the final model employed in the classification models for Buying Channels. The features were used for training the model on 80% of the cleaned data, then the trained model was used to predict the clusters on the remaining 20% of the clean data (the test data). The confusion matrix shows the true labels on the vertical axis, with the horizontal axis showing the predicted labels. The confusion matrix for Level 1 and Level 2 channels are shown in Figure 4.3 and Figure 4.4 respectively. For each square, the number represents the transactions in the test data which were classified as the combination of a specific true versus predicted channel.

As evidenced in Figure 4.3, in the case of L1 channels, the overall model accuracy was 74.81% and the model correctly predicted a significant number in the first two channels (E2E Procurement and Assisted Buying). However, the model is not able to identify eMP Exception and non-eMP Strategic, and has a limited capability of predicting Quick Pass, with only 43.13% of the transactions correctly predicted. We concluded that the features may not be good predictors of the channels, and the aggregation methodology may not be appropriate for effectively distinguishing among these specific categories effectively. It may also be possible that additional and relevant process data that is available has not been provided to us, or that the process is not collecting as much information from each transaction as it should to effectively map transactions to channels. To improve performance, it might be necessary to redesign the channel definitions or incorporate more granular data that can capture the nuances between different transaction types within these channels. It is also possible that the model is identifying transactions that are being miscategorized, e.g. mapping is incorrect in the original data file.

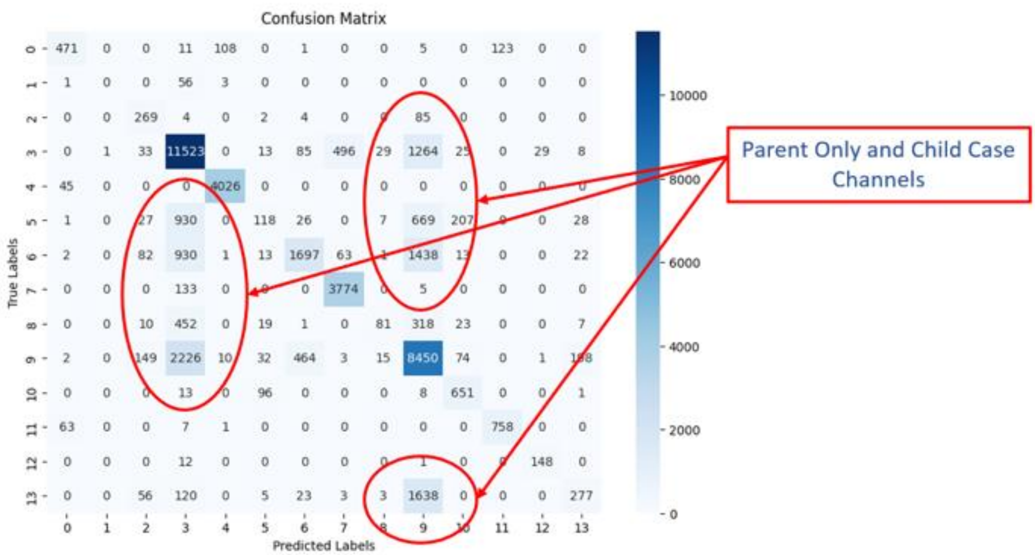
Figure 4.3

Confusion matrix for L1 Buying Channels

Level 2 channels were predicted with model accuracy of 71.13%. Here we encounter the same problem as for Level 1 with Parent Only and Child Case (Figure 4.4). Parent Only indicates transactions which did not involve Sourcing or Contracting functions. Child Case on the contrary indicates transactions where there has been some involvement of Sourcing and Contracting. Buying channels definitions, being too generic and with overlapping boundaries, have limited descriptive power because they do not represent groups of homogeneous transactions.

Figure 4.4

Confusion matrix for L2 Buying Channels



Given the imbalanced nature of the channels, with wide differences in number of transactions per channel, we tried to rebalance and improve results using a Synthetic Minority Over-sampling Technique (SMOTE) algorithm to reduce the issue of classes significantly underrepresented compared to others. The algorithm works by increasing the number of instances in the minority class by artificially generating new examples rather than simply oversampling the existing instances. This approach did not impact the model positively as the accuracy was reduced to 61.44% with SMOTE from 70.66% without SMOTE.

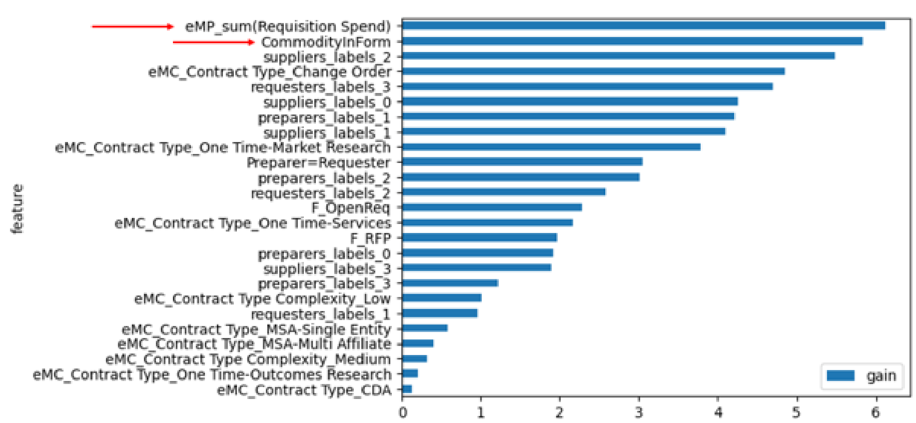
Both original and engineered features have been combined in the models to understand which are most important when explaining buying channel utilization. From interviews with stakeholders, as explained in Section 2.4, we have identified forms as the starting point for requesting the input of the R2P team for making purchases outside of the catalog. Forms are a critical part of the process as an incorrect filling of the forms leads to revisions and delays. Form features are derived from form codes in the dataset and appear to be important in predicting buying channels. Form features improve the model accuracy by more than 10%.

Gain is the improvement in accuracy for tree-based models brought by a feature to the branches. It measures the contribution of each feature in the model by calculating the change in the performance metric each time a feature is used to split the data in a tree. When a node is split on a particular feature, the algorithm measures the improvement in splitting criterion before and after the split. The gain for a particular feature is accumulated every time it is used to make a split in the tree across the ensemble of trees. Higher gain values indicate that a particular feature significantly impacts the model, implying that changes to this feature would significantly affect the model output.

The Gain plot showing top features that impacted model accuracy has "Commodity in Form" (whether the correct commodity code was in the original form), as the second most important feature as shown in Figure 4.5. "Spend" appears as the most important variable to explain buying channels, consistently with criteria which allow transactions with low spend to go through expedited channels.

Figure 4.5

Gain features: L2 Buying Channels

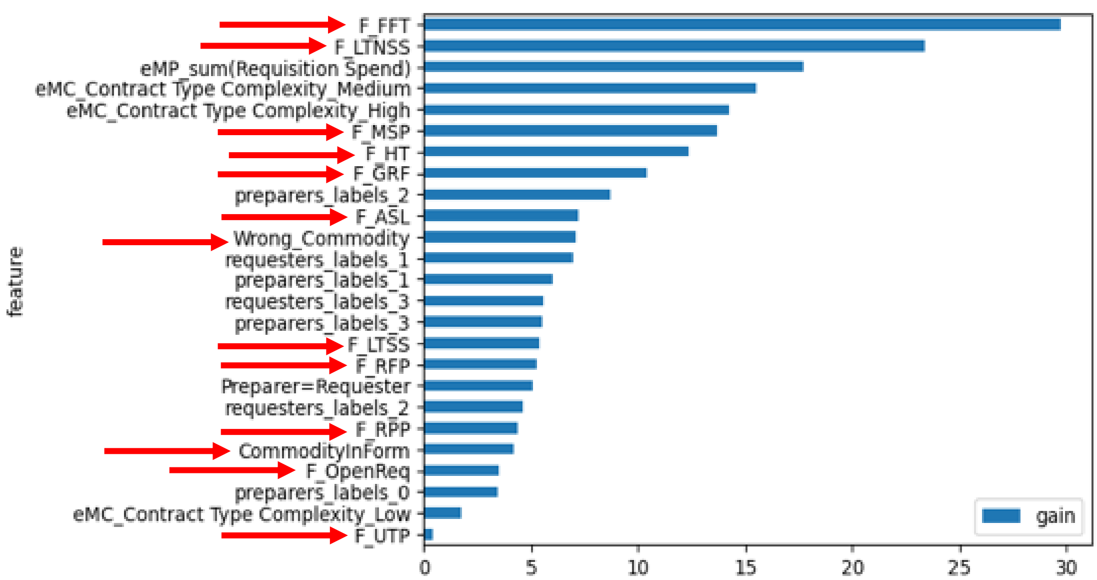


4.3 Machine learning models: cycle time

In the modelling of cycle time predictions, the Extreme Gradient Boosting (XGBoost) Regressor was employed. Overall low R^2 values when running the model with different feature combinations has led us to conclude that the features were not good predictors of cycle time. However, the highest values were obtained with the inclusion of form features which increase R^2 value from 0.28 to 0.40. Increase in the R^2 value means an increase in capacity of the model to predict and explain variability in cycle time. This is in line with the results obtained in Figure 4.5 for prediction of buying channels given the high ranking of the influence of "Commodity in Form" on the model. Figure 4.6 shows in this respect with red arrows all form features highlighted as important in a Gain histogram for cycle time. It is clear from Figure 4.6 how most of the gain derives from Form features. Further analysis of the trend of cycle time weighted average is shown in Appendices E, F and G.

Figure 4.6

Gain features: cycle time



5 Discussion

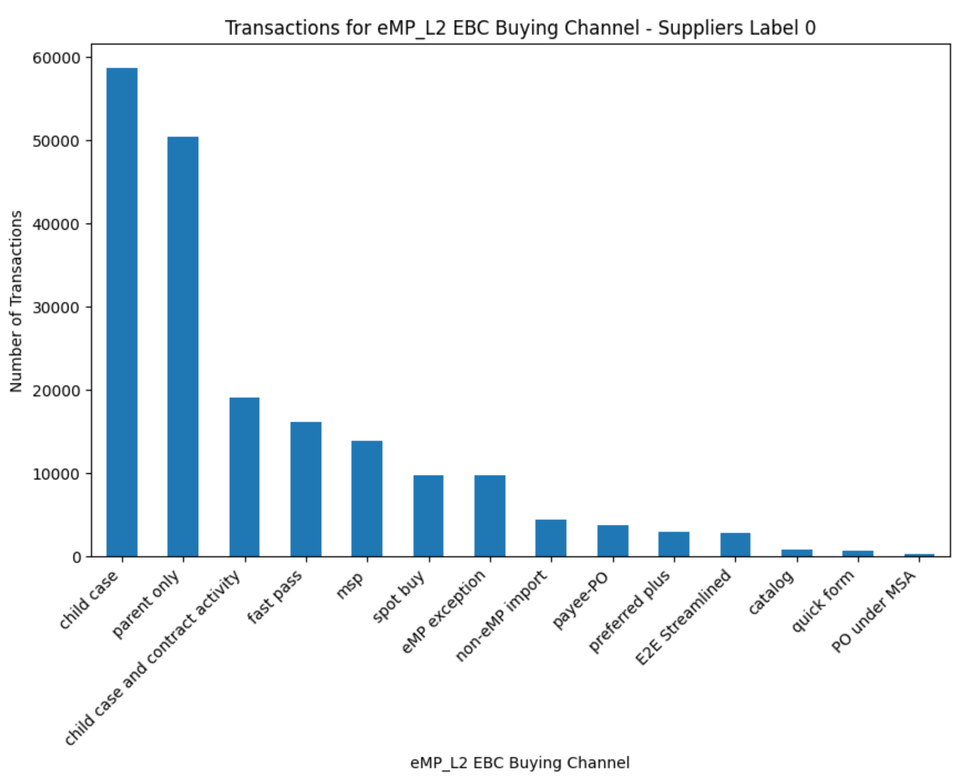
In this chapter, we draw insights from the data analysis and modeling to address the hypothesis and key research questions for this project.

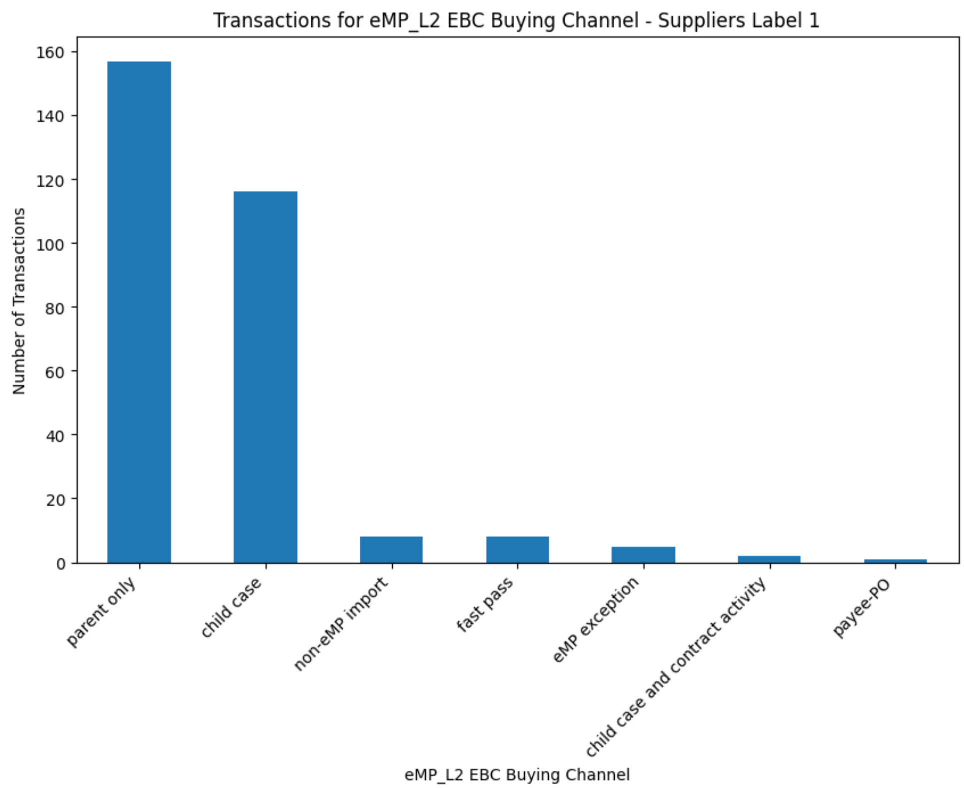
5.1 Use of Buying Channels across clusters

We have looked at how different clusters for Suppliers, Preparers and Requesters use buying channels. The most used channels for R2P related transactions are the Child case and Parent Only channels, reinforcing the previously observed overlapping nature of those two channels and the potentially incorrect classification. Figure 5.1 shows the spread of Suppliers by clusters across L2 Channels.

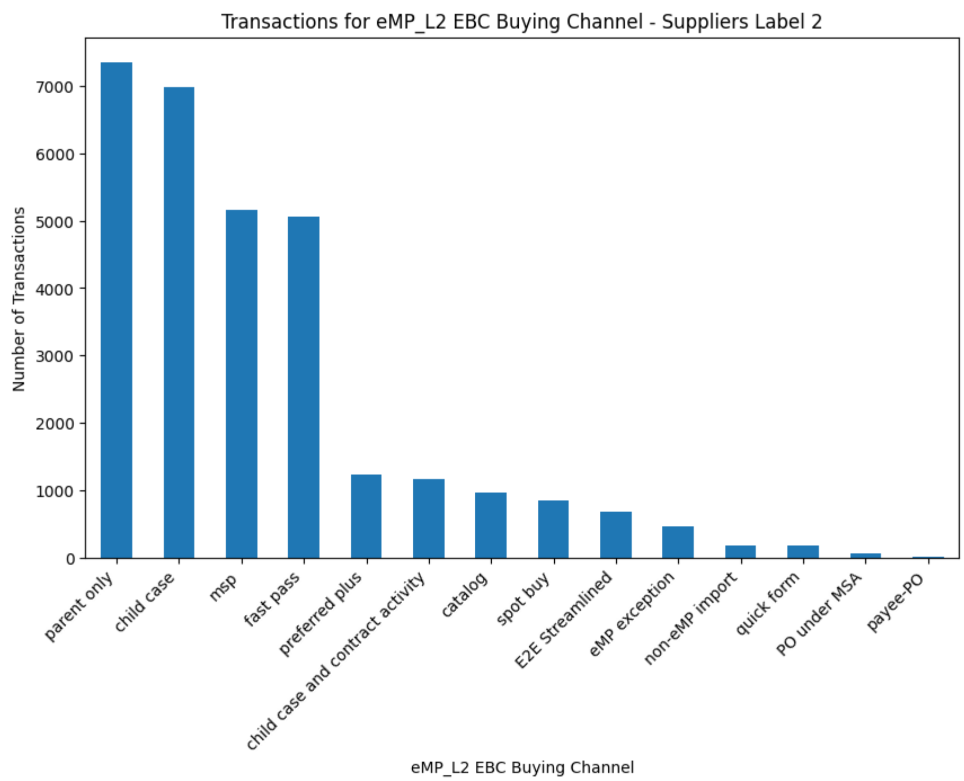
Figure 5.1

Spread of Suppliers by clusters across L2 Channels

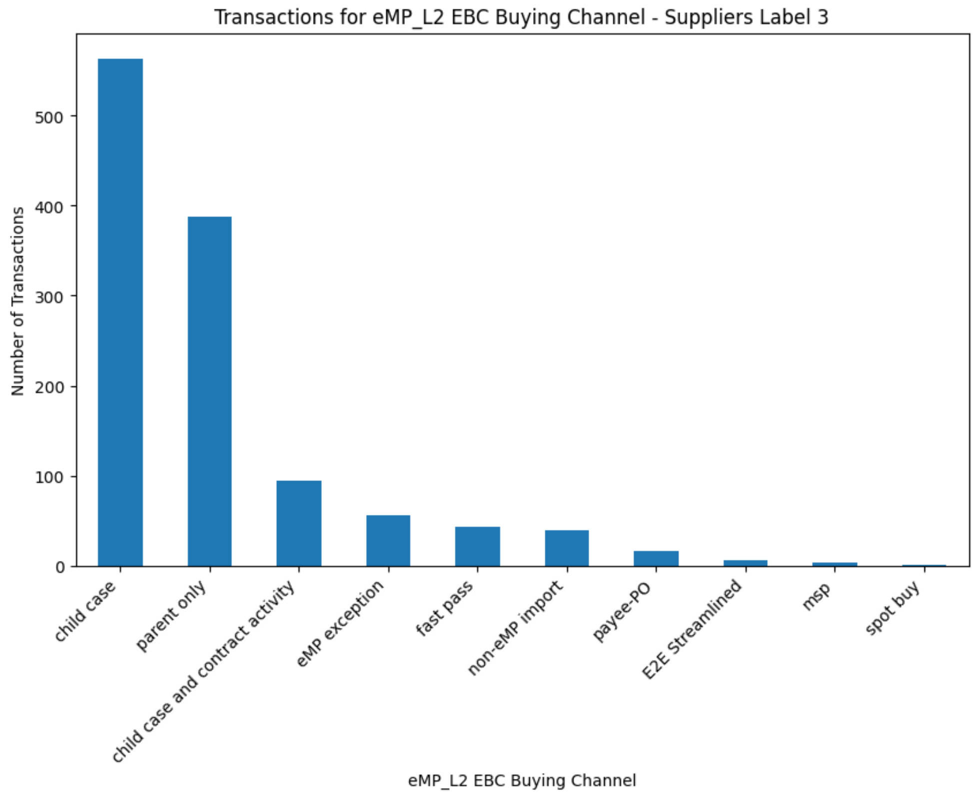




Transactions for eMP_L2 EBC Buying Channel - Suppliers Label 1
Number of Transactions
0
20
40
60
80
100
120
140
160
parent only
child case
non-eMP import
fast pass
eMP exception
child case and contract activity
payee-PO
eMP_L2 EBC Buying Channel



Transactions for eMP_L2 EBC Buying Channel - Suppliers Label 2
Number of Transactions
0
1000
2000
3000
4000
5000
6000
7000
parent only
child case
msp
fast pass
preferred plus
child case and contract activity
catalog
spot buy
E2E Streamlined
eMP exception
non-eMP import
quick form
PO under MSA
payee-PO
eMP_L2 EBC Buying Channel

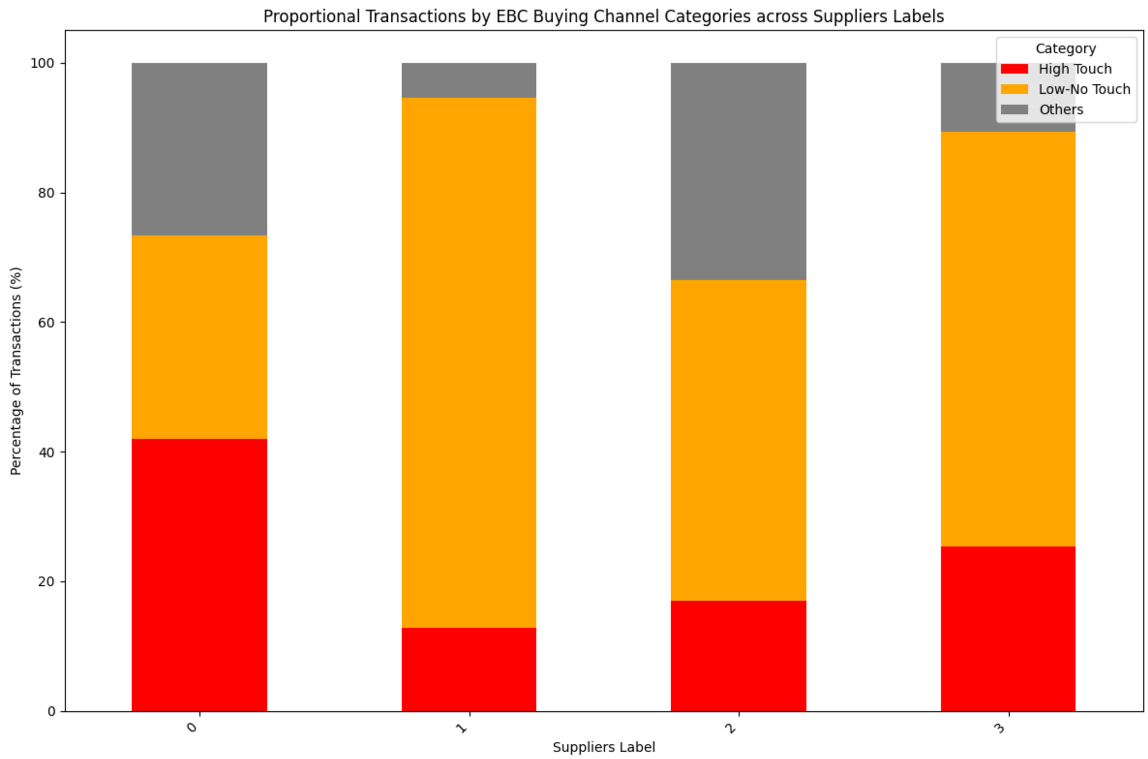


The different channels are present across all clusters of Suppliers with more Child Case and Parent Only cases than any other channel going through the R2P team. The exception to this spread is Cluster 1 (Suppliers with high value purchase orders and low number of transactions) having fewer channels and predominantly Child Case (requiring sourcing and contracts).

As evidenced in Figure 5.2, across the Suppliers' Cluster 0 (low volume and low spend) high touch channels are significant in the number of requests to R2P, which suggests either a misclassification or that there are still several relatively simple transactions not going through efficient buying channels.

Figure 5.2

Percentage of type of transactions by Suppliers clusters across L3 Channels

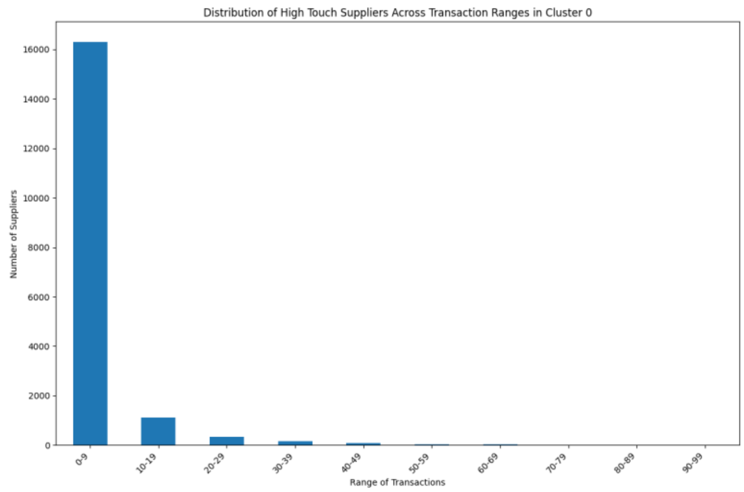


In terms of understanding why the high number of high touch transactions for cluster 0, it looks like most of those Suppliers have only a few transactions as shown in Figure 5.3. In short, it would be useful to understand if for all those transactions it is possible to reduce the number of Suppliers. To do this it would be necessary to have more specific data on individual transactions and then aggregate to fewer Suppliers.

We also observed in the spread of L3 channels across Preparers that all categories of Preparers made use of both properly classified and unclassified channels with Parent Only and Child Case channels being the majority (see Appendix I).

Figure 5.3

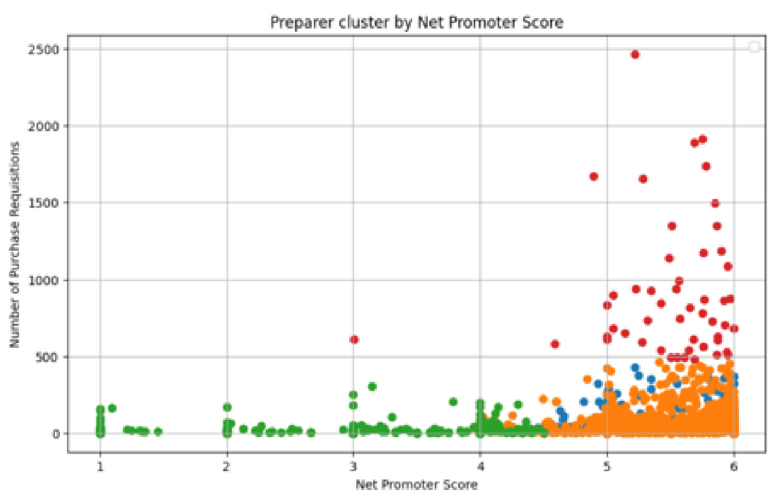
Number of transactions with High Touch for Suppliers in Cluster 0



Net Promoter Scores did not have a significant influence on the performance of the models as most of the Preparers fall within the band width of high NPS, which is 4 to 5 for those who have filled the CSAT surveys as per Figure 5.4.

Figure 5.4

Preparers cluster by Net Promoter Score



Score 6 relates to those who did not fill any customer satisfaction form. Worthy of note is that Requesters with low NPS scores tend to have a low volume of purchase requests as well suggesting that there is some lack of clarity on navigating the process among those who do not use the procurement system often.

5.2 Prediction of Buying Channel utilization

The confusion matrices in buying channels modeling show that the distinction between channels may not be properly understood by stakeholders. This can be observed by the misclassification of several channels with Parent Only and Child Case channels; i. e., channels requiring sourcing and contracting being the most misclassified. The impact of choosing the right commodity at the start when filling request forms to the R2P team plays a significant role in the outcome of the models meaning that buying channels can be better classified when there is a better understanding by the requester on how to go about the request at the onset.

Improved R^2 by 12% (as explained in Section 4.3) on the XGB Regressor model for cycle time with the inclusion of form features and subcategories reflects a moderate influence of the type of category being requested, and the request form types selected by Preparers on the performance of cycle time. The gain plot (Figure 4.6) further shows that value of spend and contract complexity play significant roles in the amount of time taken to conclude R2P transactions. The average weighted age of cycle time fluctuated in the year-on-year evolution graph of the cycle time; however, this was calculated with the combination of buying channels. Evolution of cycle time by individual L1, L2 and L3 channels may show a different trend.

6 Conclusion

In addition to supplier segmentation, which is widely acknowledged as an effective method to enhance procurement processes, segmenting procurement purchases based on buying channels proves effective in aiding the procurement organization to gain a clearer understanding of areas for strategic improvement. Advanced data analysis provides significant insight into how buying channels are used by stakeholders such as the internal customer and can be invaluable in finetuning the process for efficiency and user satisfaction.

We initially framed our work as trying to understand if we could predict buying channels utilization and cycle time, and if procurement could improve its utilization of the efficient versus traditional channels. Our conclusion is that the current mapping of buying channels is misleading with regards to the understanding of the efficiency of the transactions, and that a system based on clustering comparable transactions and a continuous review on how those clusters could be improved in terms of efficiency will be a better fit for the buying channel process. We propose a new framework that could address the above issues while providing a robust platform for ongoing learning by the organization.

6.1 Recommendations

A remapping of buying channels to provide clear distinctions between each channel can aid Preparers in selecting the right channels from the start of the requisition process. In addition, streamlining the number of buying channels would be advantageous, particularly for Preparers who interact with the system infrequently, to make the requisitioning experience less overwhelming.

Furthermore, the current definition of efficient buying channels may be insufficient. While metrics such as cycle time and number of touches (which address responsiveness), as well as adherence to procedures may offer valuable snapshots of transactional performance, they fail to capture the dynamic nature of the customer's needs and do not emphasize the ongoing learning that the organization undertakes and deploys when processing similar transactions in future. They therefore do not reflect the evolving landscape of procurement processes.

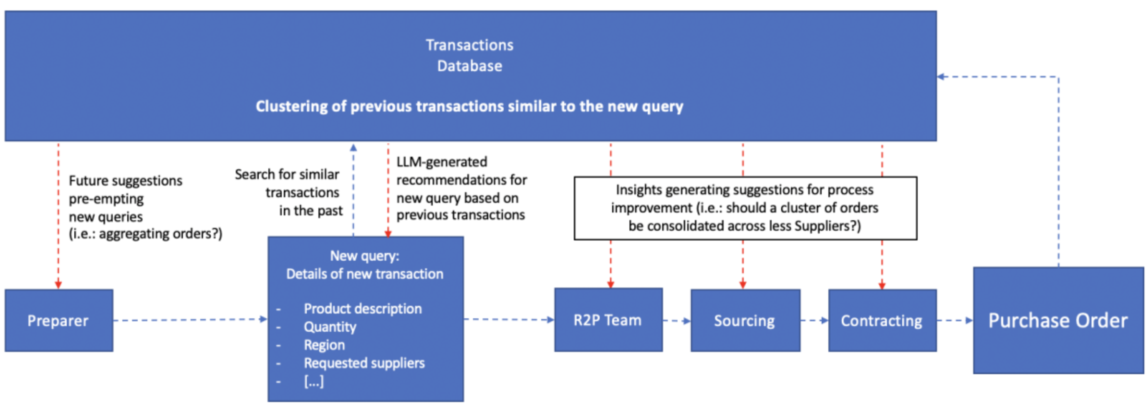
The variability of the data show that even at the most granular product level, "Commodity", demonstrates how very different types of transactions are currently categorized in a way that does not capture similarities in features that are useful to the procurement process but rather groups requests by market-driven approaches. Clustering similar transactions at a more specific level would allow all stakeholders to be aware of the current best channel to be used for each specific product given past experiences and the current organizational set up for that product. Additionally, a specific type of transaction going through an efficient buying channel may still have room for improvement. Insights given across the organization can allow the various stakeholders to contribute to ideas on how to improve on those specific transactions.

A number of metrics against which stakeholders can take decisions to further improve, and that may also mean moving to a different efficient buying channel, should be agreed across the organization.

Proposed metrics should incorporate cycle time, number of touches and waiting times for each "touch" but they should not be a measure of efficiency, merely dimensions across which the stakeholders decide to move. Information collected on past transactions with a view of what type of questions stakeholders (Preparers in particular) are likely to have when buying a new product can then provide guidance for future transactions. Structured and unstructured information can be collected in a Transactions Database. Clustering transactions with such information can then form the basis for insights through a clear and demonstrable feedback loop as shown in Figure 6.1, improving past experiences on how to manage such clusters.

Figure 6.1

Feedback loop of information across stakeholders



A Preparer looking to instruct a new transaction, for instance, by receiving insights into how similar products have been purchased in the past, could make a decision that would lead to preparing the requisition form in a more informed way. Another example is using a supplier that already has a master service agreement (MSA) with the company. Also, insights on clusters of similar transactions given to the R2P, Sourcing and Contracting teams could empower them with the necessary information to take a different approach with those types of transactions in the future, for instance, entering an MSA with a specific supplier.

6.1.2 Proposed framework for dynamic Buying Channel efficiency

We propose a new framework, described in this section, which includes three main lines of intervention (pillars) to modify the structure of the buying channels processes.

Pillar 1: Enterprise-wide data integration

At the core of our recommended framework is the creation of a network of interconnected databases, encompassing both structured and unstructured data repositories. Previously, projects were limited by the availability of structured data, making it challenging to successfully gather and integrate diverse datasets. However, the emergence of generative AI has enabled the utilization of previously inaccessible unstructured data sources. A combination of both structured and unstructured databases allows for systematic mining and cataloguing of data to extract valuable insights and support informed decision-making. Structured data will house key variables identified by the procurement function as integral to the requisitioning process such as prices, quantities, and lead times, while unstructured data, accessed through generative AI, offers insights from reports, chats, emails, and process documents. Variables that should be taken into consideration when collecting data for each transaction would be variables to support the metrics as stated in Table 2.1.

This comprehensive data ecosystem could serve as the foundation for a more extensive use of data for the purpose of restructuring buying channels. The availability of large amount of data facilitates automated adjustments that can then be introduced to dynamically redefine channels. For example, the current business risk threshold as shown in Appendix A necessitates more “touches” for compliance checks. However, using automated feedback loops, a “credit score” or “social credit score” rating system can be developed and tracked to evaluate as a Key Performance Indicator (KPI) how reliably a party has conducted itself in the past (for example tracking the number of mistakes or missed deadlines). This could allow the organization to fast-track their transactions without static standard controls. This type of innovation could potentially accelerate transactions for verified, reliable parties, streamlining processes and increasing efficiency.

Pillar 2: Co-Pilot for stakeholders

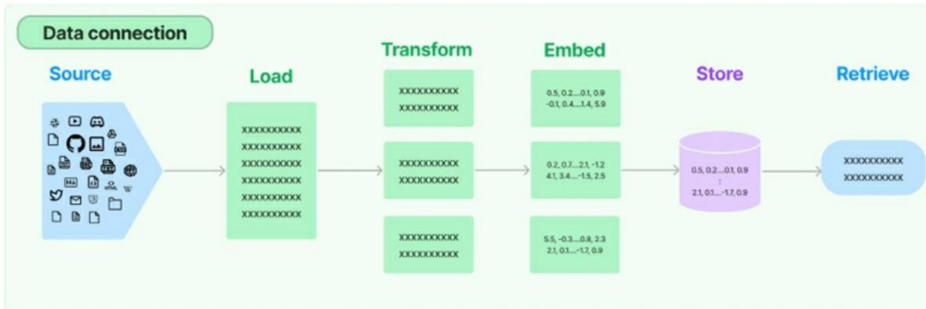
Collecting and analyzing data is not sufficient to make an impact in buying channel efficiency. Data needs to be integrated with an efficient retrieval system, a reasoning engine for recommendations and an easy-to-use interface. Stakeholders need to be able to rapidly access information and insights in order to solve problems that require collecting information disseminated in various parts of the organization. We have gathered from interviews with key stakeholders that in many cases, delays occur due to the difficulty in knowing where and how to access the necessary information. Leveraging sophisticated models, stakeholders will be empowered to query the database using intuitive language commands, eliminating barriers to access and time lags, thus enabling rapid knowledge dissemination across the organization.

We provided to the sponsor company a demonstration with LangChain, a new Python package (see Appendix J), on how to link LLMs (open or private) as reasoning engines, in this case Open AI's GPT 4, with multiple data sources that the LLM uses as "tools". Those tools are accessed through embedding and retrieving algorithms to identify the portion of information that should be sent to the LLM for processing through each prompt. This makes it possible to overcome the limitations of LLMs context window by selecting only the relevant information for the specific query. Language models then generate text based on the given prompt, processing it against previously learned parameters and memory. This can create unstable results, without the necessary precision for the final users. Adding Retrieval-Augmented Generation techniques such as the one previously mentioned can reduce this type of problem by employing semantic search algorithms to retrieve pertinent information from specific databases derived from multiple sources. Data is initially loaded in raw format and then transformed into chunks of variable sizes. Embeddings are then created, mapping the chunks of raw data into a multi-dimensional vector space, encapsulating its semantic content while excluding non-relevant information. Those embeddings can then be stored. Efficient retrieval of information, comparing these embeddings to find the closest matches to a query vector by the user, facilitates fast similarity searches through techniques like approximate nearest neighbor search.

This retrieval process conditions the language model to refine its output, leading to text that is not only more nuanced but also factually correct. This capability significantly reduces the incidence of erroneous or "hallucinated" responses often associated with LLMs.

Figure 6.2

Retrieval-Augmented Generation (RAG) pipeline



From *Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT, and other LLMs* by B. Auffarth, 2023, Packt Publishing, p. 135

We believe that an architecture based on the above steps, as illustrated in Figure 6.2, rather than a newly trained model on company data, would be more precise in accessing information and would be better able to adapt to revisions of past incorrect data, benefitting all stakeholders.

Pillar 3: Performance incentives for ongoing learning

Central to our framework is the concept of iterative learning and adaptation, whereby insights obtained from past transactions serve as the foundation for ongoing process refinement and optimization. Through the establishment of feedback loops, stakeholders can systematically incorporate lessons learned from previous transactions into future decision-making processes, reducing instances of repeated errors and enhancing overall process efficiency.

As the additional information is continuously fed into the process, the stakeholders have a better way to connect the dots and every new transaction benefits from past experience and from the knowledge of past problems and how they have been addressed. As the process operates and accumulates additional actionable information over time, stakeholders gain a clearer understanding of the interconnectedness between various data points. Furthermore, the framework facilitates cross-functional collaboration by providing a common platform for sharing insights and best practices, fostering a culture of collective learning and collaboration.

We identify different types of learnings:

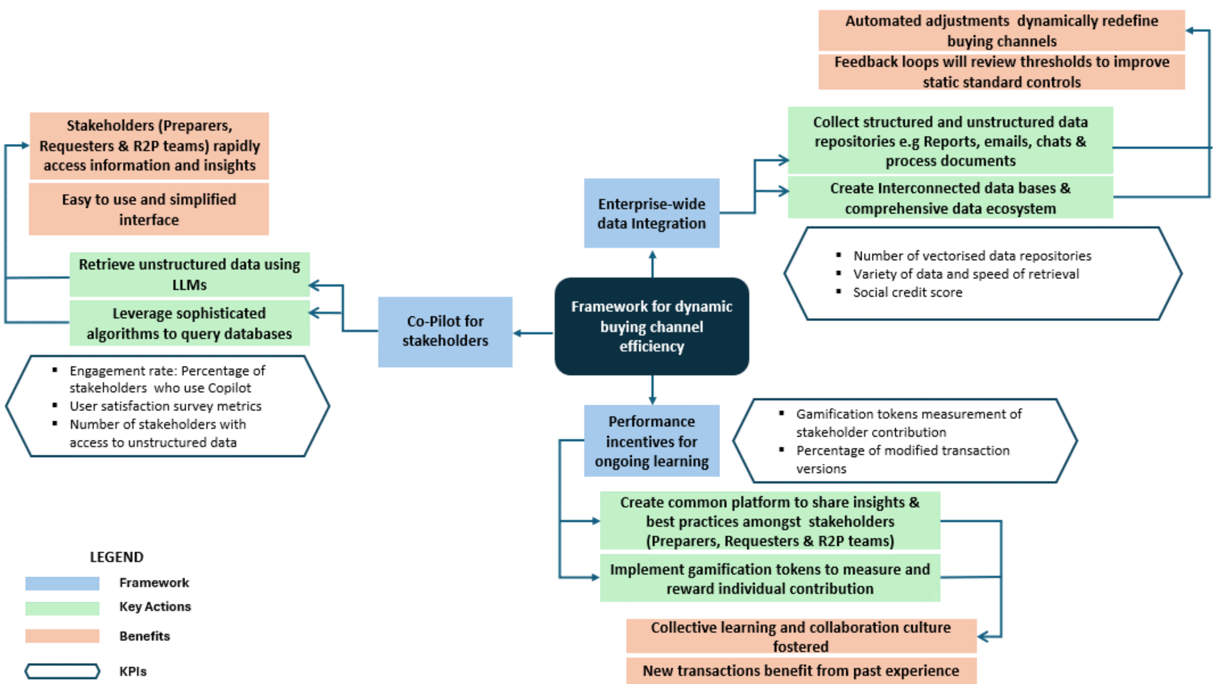
Customers' perspective: Is the customer happy with the process and with the final result? The process for collecting satisfaction feedback from transactions can be incorporated in the new clustering of transactions, highlighting issues where low scores across transactions in the same cluster are repeated over time, and potentially flagging to other stakeholders the need to address structural issues. Even unstructured email exchanges can be incorporated in this learning process. Engagement is a sign of attention to feedback, and the lack of it a reason for concern, so the system should encourage customers' feedback by highlighting the fact that all feedback information will be incorporated into future decisions. Key Performance Indicators could be the percentage of stakeholders that use the co-Pilot for every transaction, number of questions asked to Co-pilot vs questions asked to other stakeholders, number of stakeholders giving feedback with Net Promoter Score (NPS) and the NPS itself.

Process perspective: Have improvements in the process for similar products or services been put in place? Here is where the metrics of cycle time and number of "touches" are useful as potential KPIs, looking at % of new transactions which have been improved from previous similar transactions. We are not measuring in the abstract but rather, the actual improvement compared to previous transactions the actual improvement, recognizing therefore that the efficiency is not in the static value but in the delta from one similar transaction to another. Additional KPIs from a process perspective could be the number of stakeholders that have access to data or that have been trained to access the data.

Recognizing the pivotal role of stakeholders in driving process improvements and fostering a culture of innovation, in line with the culture of the sponsor company, we propose the implementation of a performance incentive mechanism which recognizes both the customer and the process perspectives. By establishing a reward system that tracks and quantifies into "tokens", the improvements in process efficiency for specific transaction types, organizations can incentivize proactive engagement and knowledge sharing among employees through gamification. Those 'tokens', reflecting improvements, would represent a performance score for stakeholders contributing to the efficiency improvements. Such a score can be normalized across the organization and then integrated into internal human resources policies, where individuals receive part of their variable compensation linked to such measures, at both individual and team levels. Other forms of incentives such as internal competitions can further enhance the scope of such a mechanism, serving not only as a tangible measure of individual contributions but also as a catalyst for driving collective learning and continuous improvement within the procurement function. Tokens being assigned to stakeholders and number of stakeholders actively participating in the tokens program could be additional KPIs to be considered.

Figure 6.3

Proposed framework for Buying Channels



6.2 Limitations

While commodity was the lowest level of taxonomy provided to us, a more granular view of items within each commodity would have proven useful in gaining more understanding of the dataset. In some instances, descriptions of taxonomy in individual requisitions show some generalization that may contain very different items requiring separate processing steps in order to be managed efficiently.

The data provided did not incorporate "touches" within a process, making it difficult to understand if there are additional dimensions of inefficiency, as a different baseline cycle time should be identified for items which are very different in nature.

6.3 Next steps

To explore future research opportunities, we recommend qualitative research on the experience of Preparers and Requesters to identify pain points on the use of the buying channels process.

Also, future research opportunities include a more in-depth analysis on the use of tokenization and gamification of positive feedback loops within the procurement function. This research could include other mechanisms which increase stakeholders' engagement and participation, such as the "reliability", or "social credit score", as well as reward mechanisms that could enhance positive feedback loops.

The use of advanced data analytics and the proposed framework represent a natural next step for the sponsor company in their journey to improve efficiency in buying channels processes. Our contribution in understanding the variables that influence buying channel utilization has led us to rethink the process structure: we envisage that by implementing the suggested actions the sponsor company will be able to accelerate the rate at which transactions are efficiently processed through newly defined buying channels.

References

Auffarth, Ben., (2023), Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT, and other LLMs (p. 135). Packt Publishing.

Baron, J. P., Shaw, M.J., & Bailey, A.D. (2000). Web-Based e-Catalog Systems in B2B Procurement. *Association for Computing Machinery*, 43(5). Retrieved from: https://doi.org/10.1145/332833.332845

BCG. (2021, January 06). *The Future of Pharma Company Procurement Function.* Retrieved from: https://www.bcg.com/publications/2021/future-of-pharma-company-procurement-function

CIPS. (2023a*). Procurement Process. Retrieved from*: https://www.cips.org/intelligence-hub/procurement/procurement-process

CIPS. (2023b). *Big Data in Procurement.* Retrieved from: https://www.cips.org/intelligence-hub/procurement/big-data-in-procurement

Deloitte. (2023, June 27). *Generative AI in Procurement*. Retrieved from: https://www2.deloitte.com/us/en/blog/business-operations-room-blog/2023/generative-ai-in-procurement.html

Fischer, N.F., Roux, A., Haas, F., Platsch, A., Carradine, I., Tue, F., Schäfer, & L., Meintrup, J. (2022). PWC Global Digital Procurement Survey. Retrieved from: https://www.pwc.com/gx/en/services/consulting/digital-operations/digital-procurement-survey.html

González-Benito, J. (2007). A theory of purchasing's contribution to business performance. Journal of Operations Management, 25(4), 901-917. Retrieved from: https://doi.org/10.1016/J.JOM.2007.02.001

Karumsi, D., & Prokopets, L. (2021). *Buying Channel Strategy: The Secret to Ending Procurement. Value Leakage. KPMG Advisory Services*. Retrieved from: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2020/buying-channel-strategy.pdf

Klein, A., Whatson, S., & Oliveira, J. (2021). Profit from Procurement: Add 30% to Your Bottom Line by Breaking Down Silos. John Wiley & Sons. Retrieved from: https://learning.oreilly.com/library/view/profit-from-procurement

Kraljic,P.,(1983). Purchasing Must Become Supply Management Retrieved from: https://hbr.org/1983/09/purchasing-must-become-supply-management

Mandl ,C. (2023). Procurement Analytics. Springer Series in Supply Chain Management. Retrieved fom: https://doi.org/10.1007/978-3-031-43281-1

Novack, R. A., & Simco, S. W. (1991). The Industrial Procurement Process: A Supply Chain Perspective. Journal of Business Logistics, 12(1), 145. Retrieved from: https://www.proquest.com/scholarly-journals/industrial-procurement-process-supply-chain/docview/212644925/se-2?accountid=12492

O'Brien, J. (2019). Category Management in Purchasing: a strategic approach to maximize business profitability. Kogan Page Limited. Retrieved from: https://books.google.co.uk/books?hl=en&lr=&id=C3iJDwAAQBAJ&oi=fnd&pg=PP1&dq=definition+of+category+management+in+procurement&ots=YQV8GHiDS6&sig=M4SdM55cnYm5GUOj3uBJ5aWgcXM#v=onepage&q&f=false

Pedregosa *et al.*,(2011) Journal of Machine Learning Research, 12, pp. 2825-2830, 2011. Retrieved from: Scikit-learn: Machine Learning in Python

Protopappa-Sieke & Thonemann (2017) . *Supply Chain Segmentation: Best-in-Class Cases, Practical Insights and Foundations.* Germany: Springer International Publishing. Retrieved from: https://link.springer.com/book/10.1007/978-3-319-54133-4?utm_medium=referral&utm_source=google_books&utm_campaign=3_pier05_buy_print&utm_content=en_08082017

SAP. (2022, May 2022). *Make the Case for Change in Procurement*. Retrieved from: https://www.sap.com/cmp/dg/make-the-case-for-change/typ.html?pdf-asset=98de70d0-237e-0010-bca6-c68f7e60039b&page=1

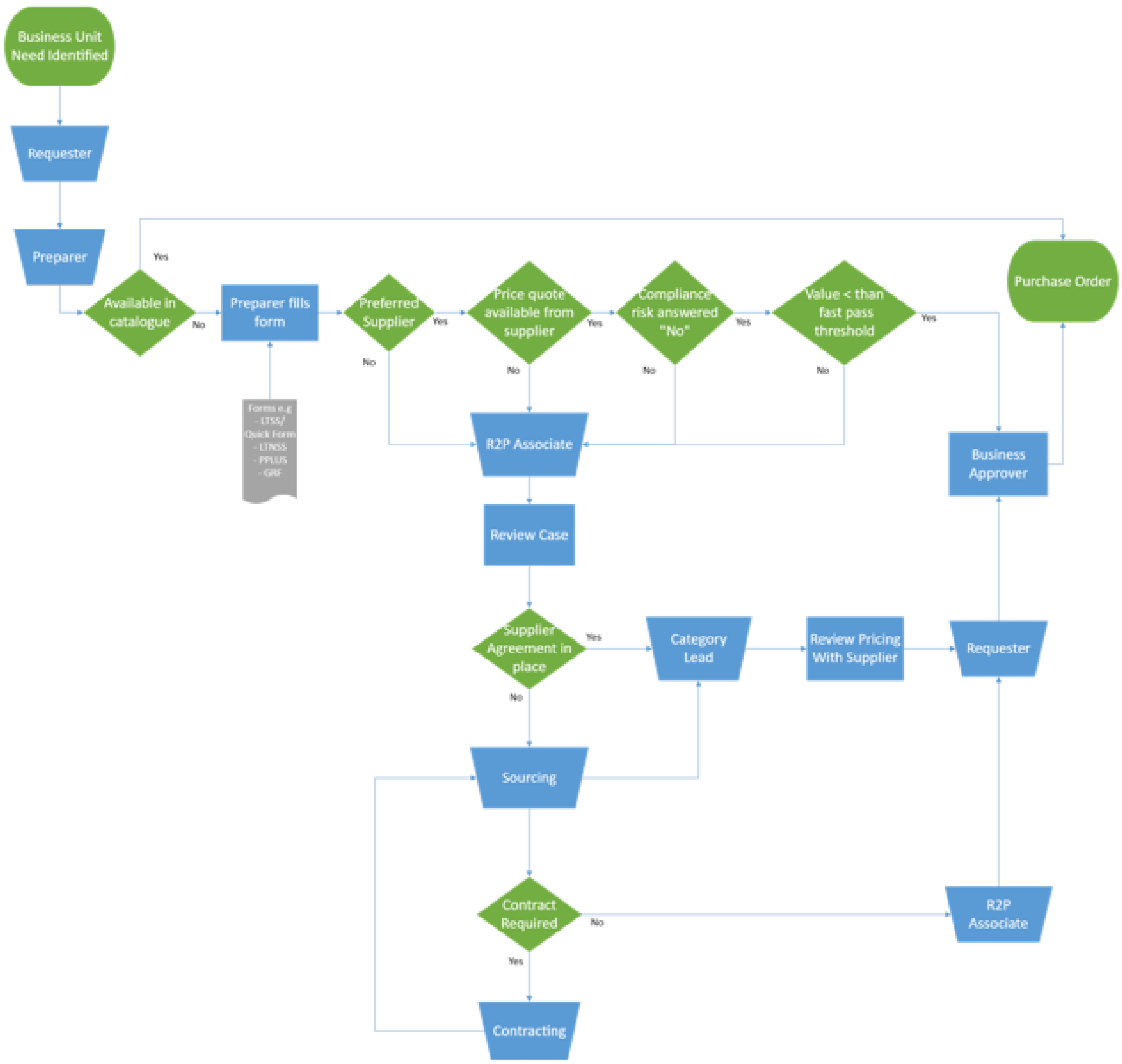
SAP. (2023). *Buying Channels: The Key to Frictionless Procurement.* Retrieved from: https://sapinsider.org/articles/buying-channels-the-key-to-frictionless-procurement/

The Tech Spark. (2023, March 22). *Best prediction algorithms for Machine Learning.* Retrieved from: https://thetechspark.com/best-prediction-algorithms-for-machine-learning/

Toikka, E. (2023, August 22). *Category Management: A Comprehensive Guide*. Retrieved from: https://sievo.com/blog/category-management

Appendix

Appendix A
Process map for Buying Channels



Appendix B
Data selection

Relevant columns in NG Ops data frame and their corresponding datatypes.

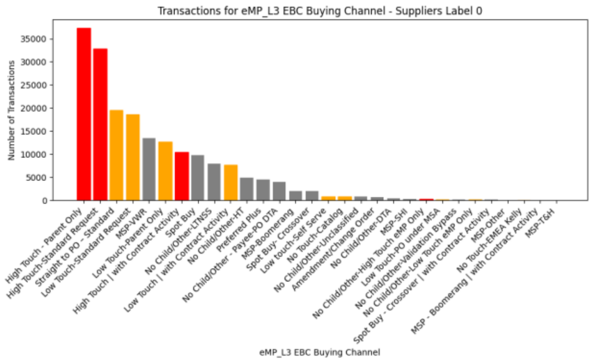
S/N	Column	Data type	Description
1	E2E Country Code	object	Geographic location of the transaction; this is based on purchasing unit from which PR is submitted
2	E2E Sector	object	Sector of the transaction; this is based on purchasing unit from which PR is submitted
3	E2E Family	object	Categorization of good/service within Procurement taxonomy for the transaction (family-category-subcategory-commodity)
4	E2E Category	object	Categorization of good/service within Procurement taxonomy for the transaction (family-category-subcategory-commodity)
5	E2E SubCategory	object	Categorization of good/service within Procurement taxonomy for the transaction (family-category-subcategory-commodity)
6	eMP_(REQ)ERP Commodity (ERP Commodity ID)	object	Commodity identifier that represents the nature of the good or service being purchased in the PR
7	eMP_(REQ)ERP Commodity (ERP Commodity)	object	Commodity that represents the nature of the good or service being purchased in the PR
8	E2E Purchasing Unit	object	Procurement (purchasing) group from which a PR is submitted
9	eMP Preparer ID	object	Unique identifier for preparer of PR (person who physically input and submitted the PR in ERP)
10	eMP Requester ID	object	Unique identifier for requester on record for PR (PR owner)
11	eMP_(PO) Order Id	object	PO number of the purchase order that's created upon full approval of the PR
12	eMP_(REQ) Request Form ID	object	Unique identifier for items that are not available in the catalogue and referred to R2P
13	eMP_(REQ) Requisition ID	object	PR number of the purchase requisition
14	eMP_(REQ)Supplier (ERP Supplier ID)	object	Unique identifier of the vendor from which the good/service will be delivered for the PR
15	eMP_Dateof Requisition	datetime64[ns]	Date when PR was submitted
16	eMP_Is Catalog Item	object	Indicates whether good/service being purchased in PR was from a catalog in ERP
17	eMP_Category Card Ind	object	Indicates if request with supplier is not identified in Category card (meaning it is not in preferred supplier list maintained by Procurement)

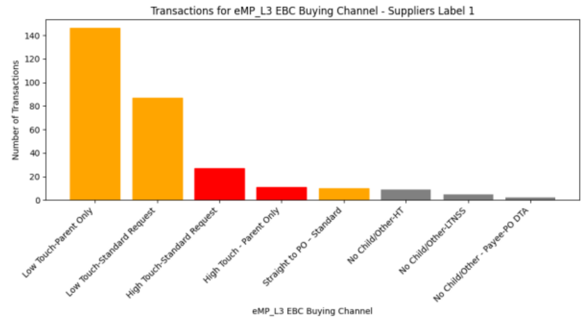
S/N	Column	Data type	Description
18	eMP_L1 EBC Buying Channel	object	Highest level of aggregation for buy channel designation
19	eMP_L2 EBC Buying Channel	object	Second-highest level of aggregation for buy channel designation
20	eMP_L3 EBC Buying Channel	object	Lowest level of aggregation for buy channel designation
21	eMP_sum(Requisition Spend)	float64	Total ($) value of good/service being purchased in PR
22	SFDC Parent_Case Age (Days)	float64	Provides information on cycle time (Time taken to complete the transaction)
23	SFDC Parent_Case Number	float64	Unique identifier of the parent case request (request made via dynamic request form)
24	Versionless PR num	object	PR number without version suffix for the purchase requisition
25	SFDC Child_Case Number	float64	Unique identifier of the child case request (request made for contracting)
26	eMC_Contract Type Complexity	object	Indicates how complex a contract request is

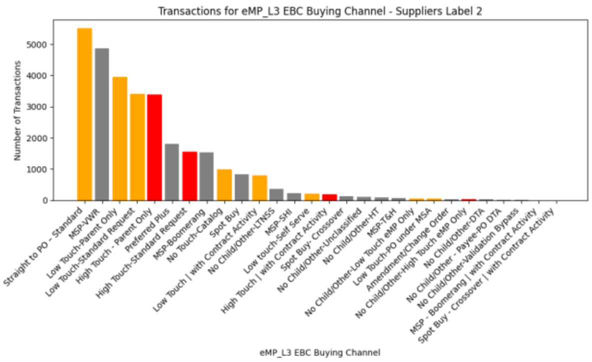
Relevant columns in CSAT data frame and their corresponding datatypes.

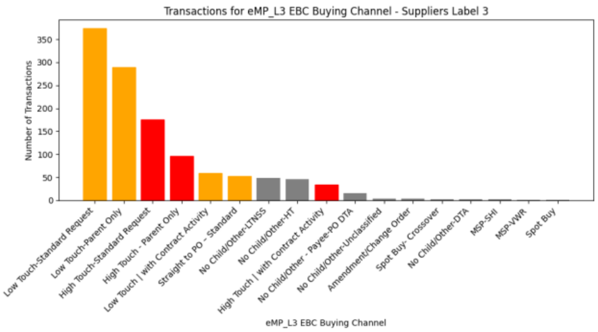
S/N	Column	Data type	Description
1	Contact WWID	float64	Unique identifier to connect "eMP Requester ID" from NG-Ops
2	Net Promoter Score	float64	Satisfaction survey form filled by requester

Appendix C
Descriptive analysis: Suppliers by L3 Channel

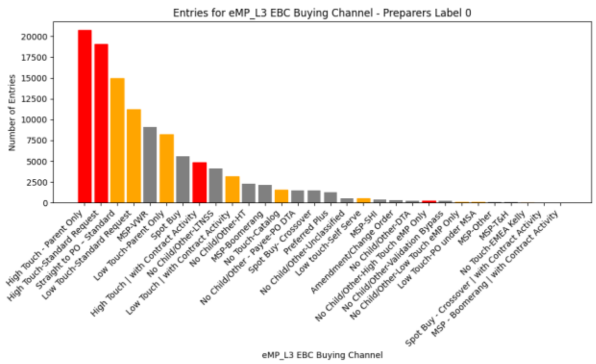


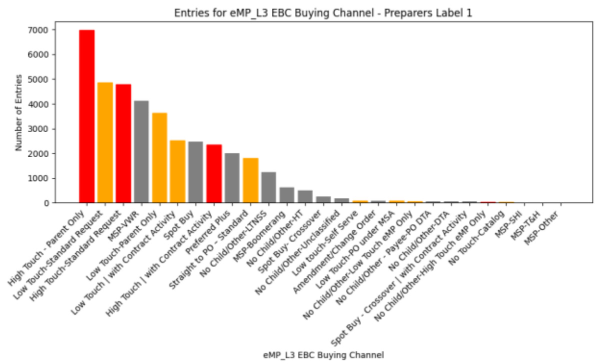


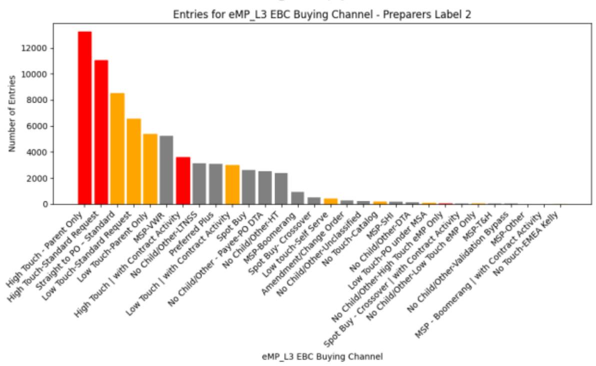


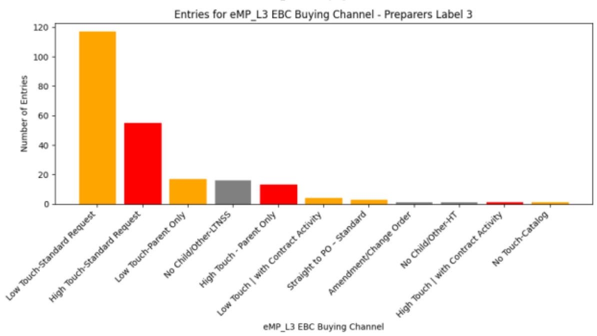


Appendix D
Descriptive analysis: Preparers by L3 Channel

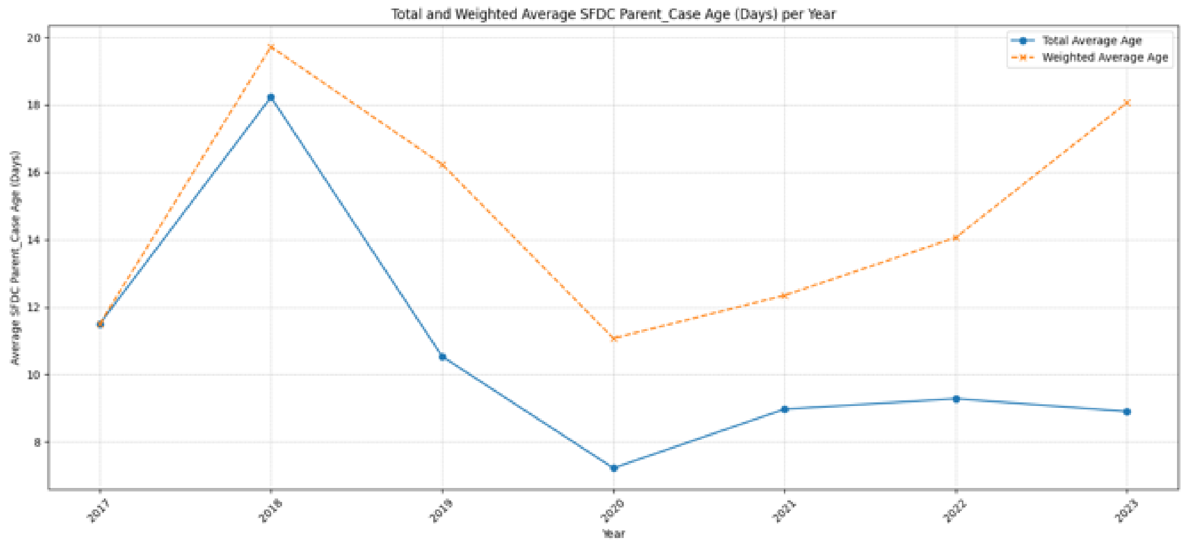




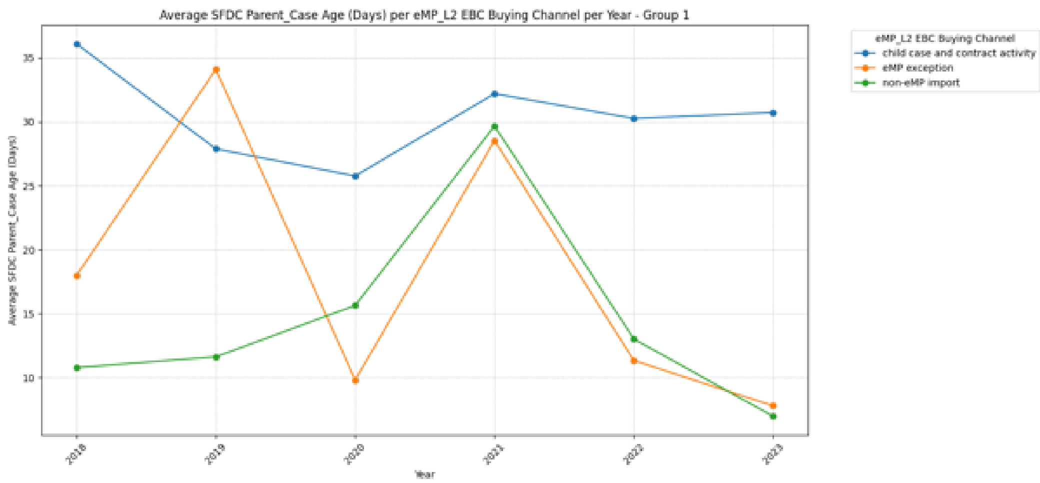


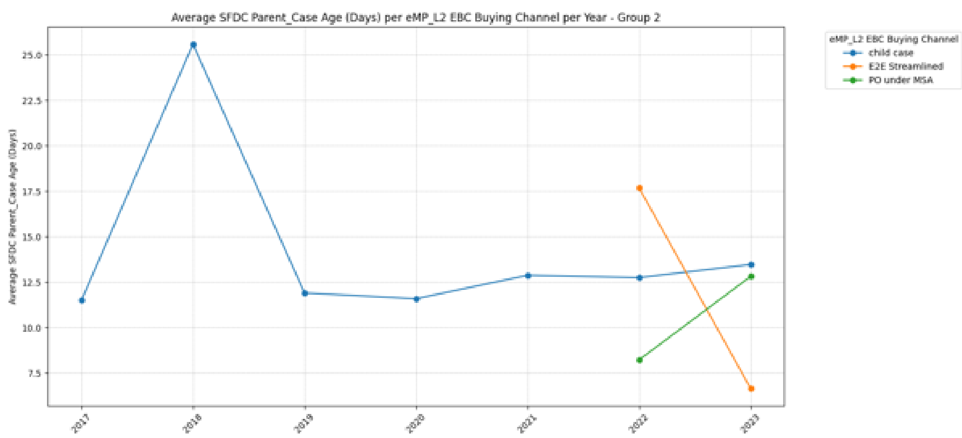


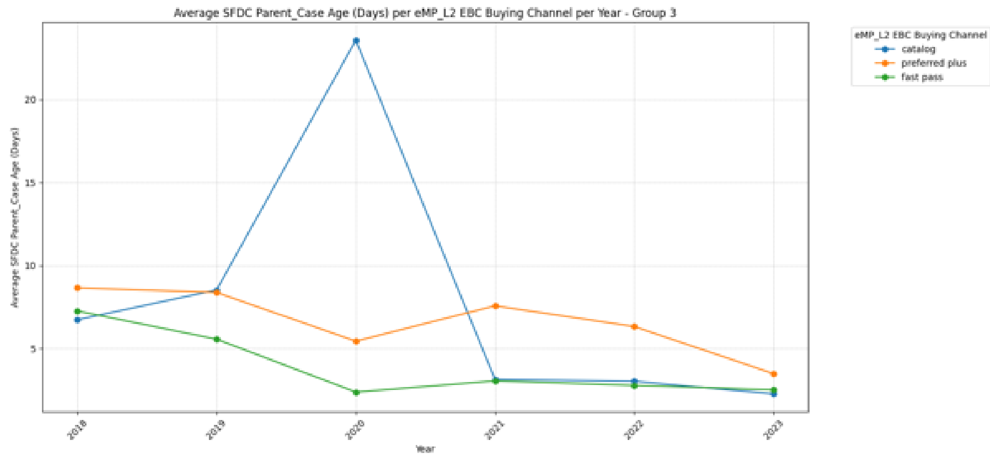
Appendix E
Evolution of aggregate cycle time per year

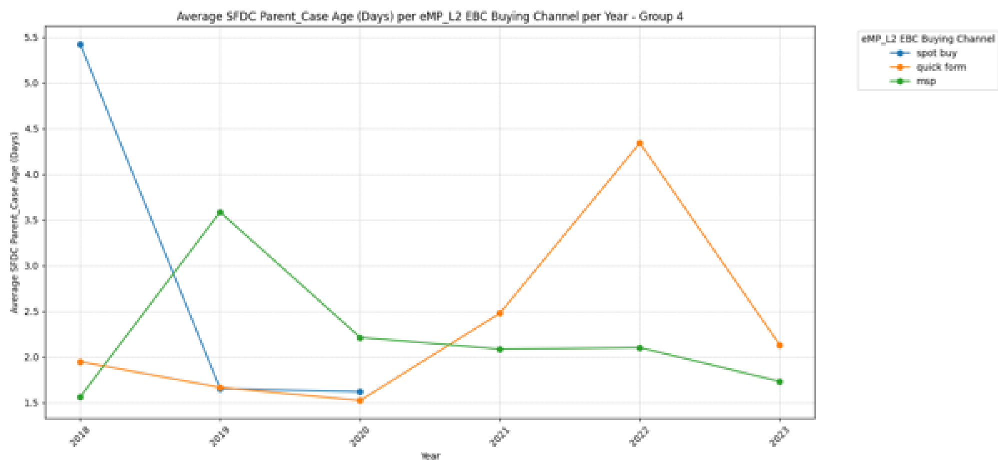


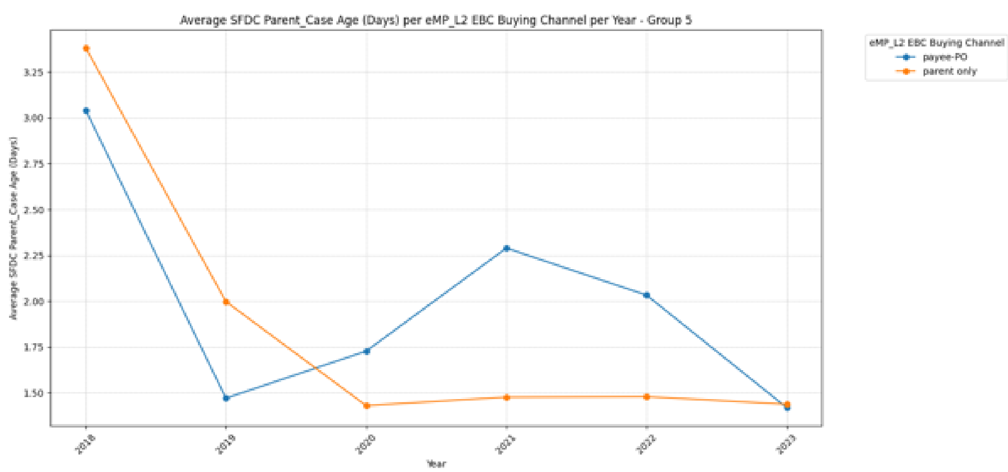
Appendix F
Evolution of cycle time by Buying Channel



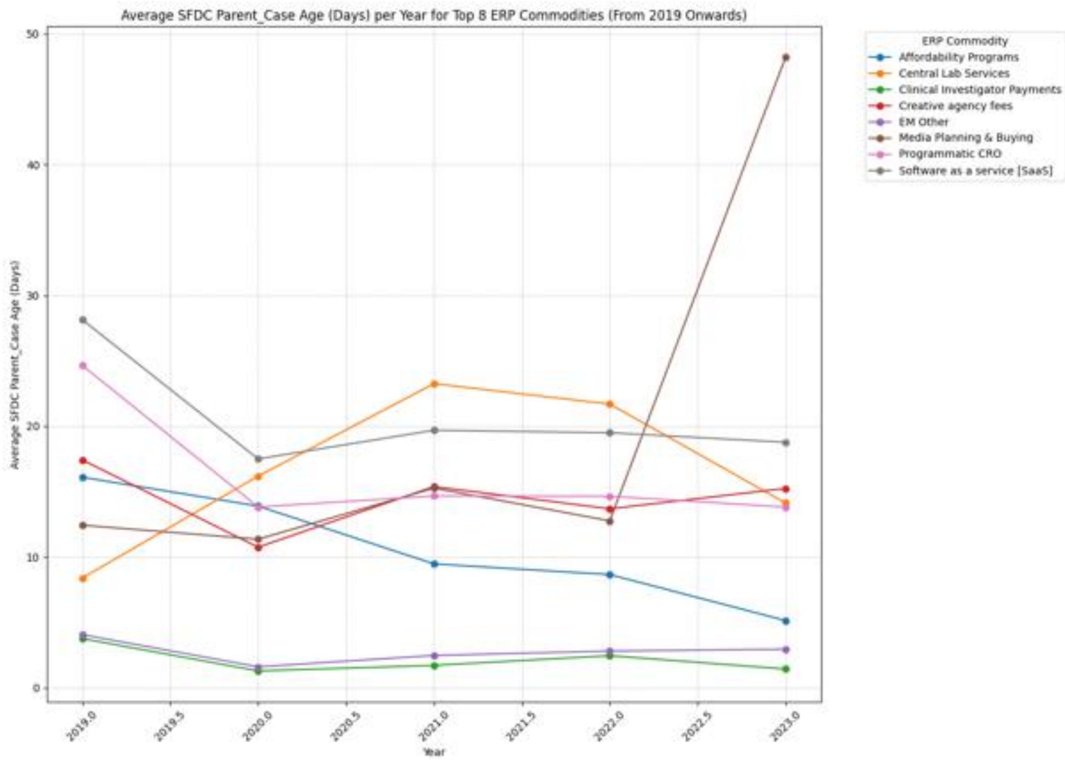



Average SFDC Parent_Case Age (Days) per eMP_L2 EBC Buying Channel per Year - Group 3
Average SFDC Parent_Case Age (Days)
Year
eMP_L2 EBC Buying Channel
catalog
preferred plus
fast pass

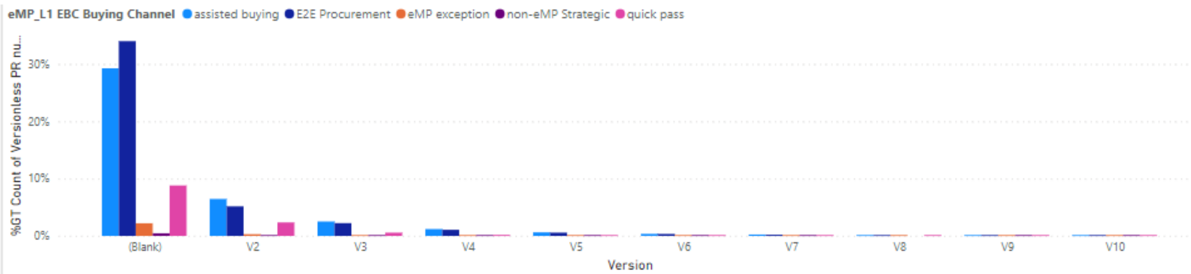

Average SFDC Parent_Case Age (Days) per eMP_L2 EBC Buying Channel per Year - Group 4
Average SFDC Parent_Case Age (Days)
Year
eMP_L2 EBC Buying Channel
spot buy
quick form
msp


Average SFDC Parent_Case Age (Days) per eMP_L2 EBC Buying Channel per Year - Group 5
Average SFDC Parent_Case Age (Days)
Year
eMP_L2 EBC Buying Channel
payee-PO
parent only

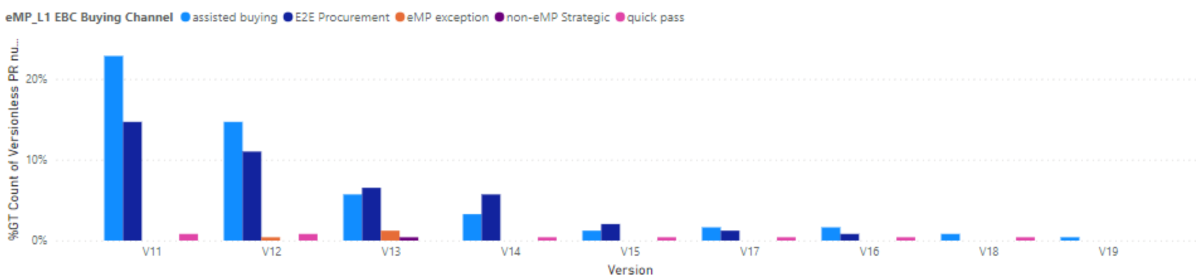
Appendix G
Evolution of cycle time for the ‘Commodity’ classes most used



Appendix H
PR Versions by L1 Buying Channel

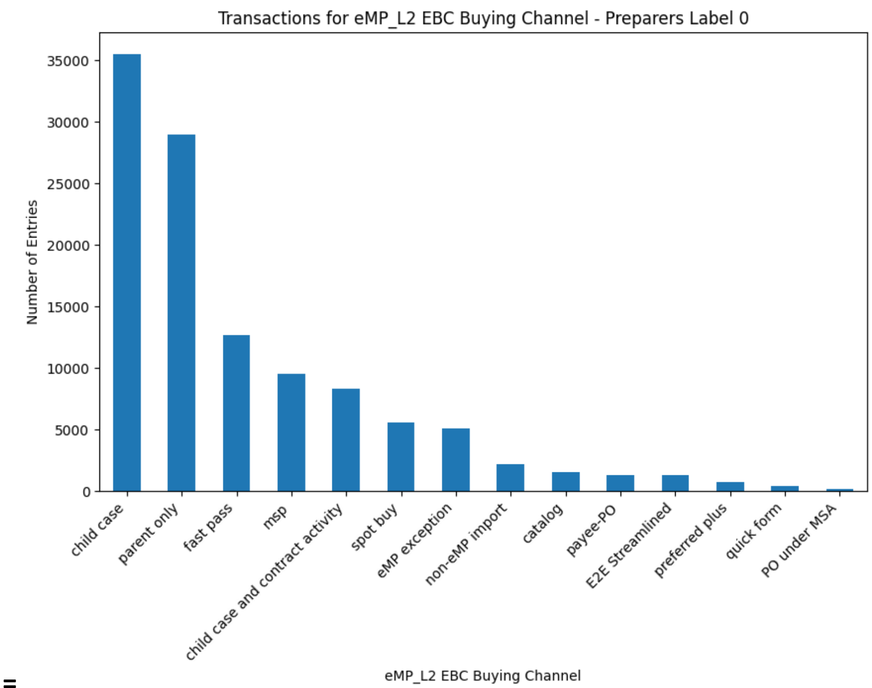


Version	assisted buying	E2E Procurement	eMP exception	non-eMP Strategic	quick pass
	29.26%	34.01%	2.20%	0.46%	8.80%
V2	6.45%	5.16%	0.29%	0.06%	2.35%
V3	2.51%	2.23%	0.12%	0.02%	0.56%
V4	1.19%	1.07%	0.05%	0.01%	0.19%
V5	0.62%	0.57%	0.02%	0.00%	0.06%
V6	0.35%	0.34%	0.01%	0.00%	0.04%
V7	0.23%	0.21%	0.01%	0.00%	0.01%
V8	0.14%	0.12%	0.00%		0.01%
V9	0.08%	0.08%	0.00%	0.00%	0.01%
V10	0.05%	0.05%	0.00%	0.00%	0.00%

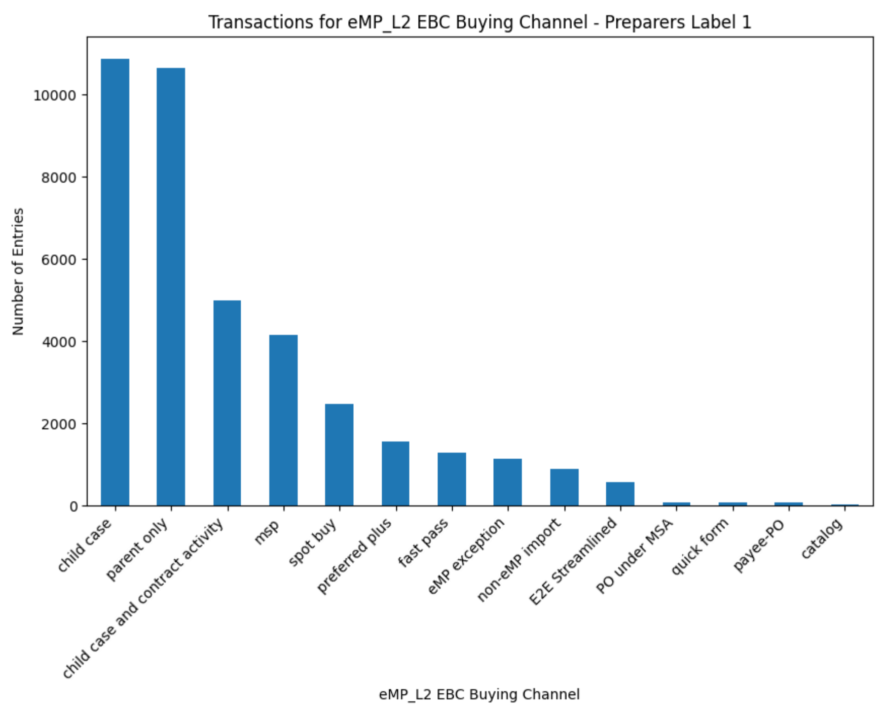


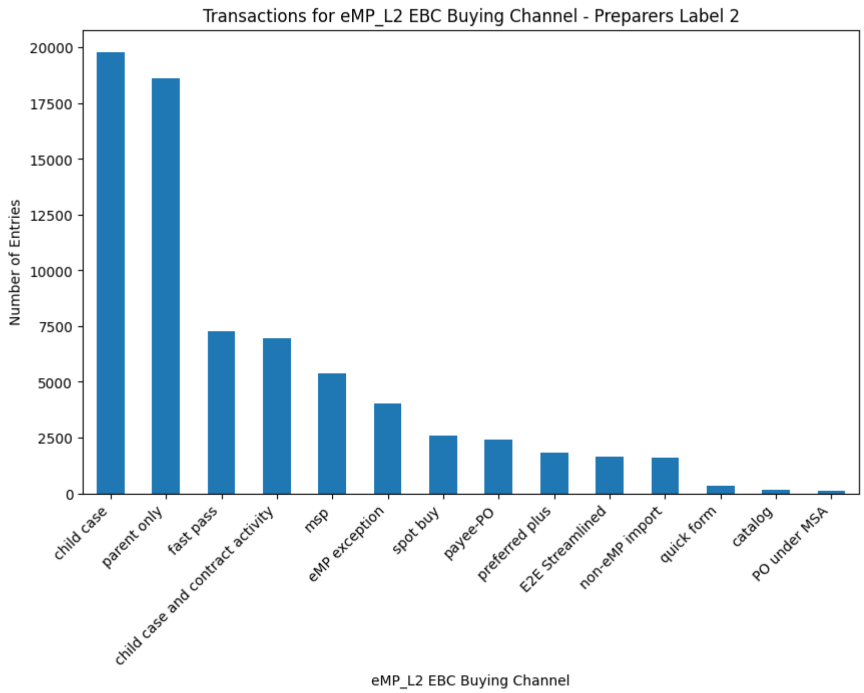
Version	assisted buying	E2E Procurement	eMP exception	non-eMP Strategic	quick pass
V11	22.86%	14.69%			0.82%
V12	14.69%	11.02%	0.41%		0.82%
V13	5.71%	6.53%	1.22%	0.41%	
V14	3.27%	5.71%			0.41%
V15	1.22%	2.04%			0.41%
V17	1.63%	1.22%			0.41%
V16	1.63%	0.82%			0.41%
V18	0.82%				0.41%
V19	0.41%				

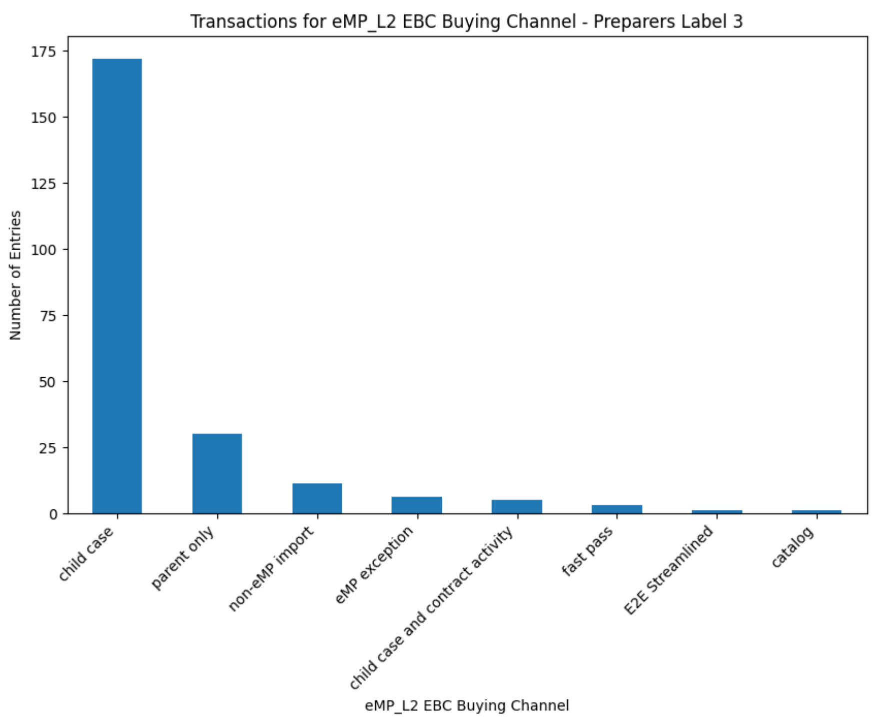
Appendix I
Spread of Preparers by clusters across L2 Channels



=




Transactions for eMP_L2 EBC Buying Channel - Preparers Label 2
Number of Entries
20000
17500
15000
12500
10000
7500
5000
2500
0
child case
parent only
fast pass
child case and contract activity
msp
eMP exception
spot buy
payee-PO
preferred plus
E2E Streamlined
non-eMP import
quick form
catalog
PO under MSA
eMP_L2 EBC Buying Channel


Transactions for eMP_L2 EBC Buying Channel - Preparers Label 3
Number of Entries
175
150
125
100
75
50
25
0
child case
parent only
non-eMP import
eMP exception
child case and contract activity
fast pass
E2E Streamlined
catalog
eMP_L2 EBC Buying Channel

Appendix J
Co-Pilot demonstration on a randomly generated dataset including structured and unstructured data

Preparer Co-Pilot

Hello! Enter your query about a new requisition:

Do we have any recent email about problems for deliveries in Boston?

Yes, we do have a recent email about delivery problems in Boston. The email is from Bob the Preparer and it mentions that the supplier Cloud Castles has been having issues with deliveries in Boston. They are currently investigating the situation.

Preparer Co-Pilot

Hello! Enter your query about a new requisition:

I want to buy a hard drive. Who is the preferred supplier in Boston?

The preferred supplier for external hard drives in Boston is Gadget Gurus. The purchase price is approximately $448.62 and the delivery time is 2 days. If you need any support, you can contact Hannah the Helper at (212) 123 4567.

Preparer Co-Pilot

Hello! Enter your query about a new requisition:

What is the supplier rating and reliability of Gadget Gurus?

The supplier rating for Gadget Gurus is 4 out of 5. They are listed as a preferred supplier with a low complexity factor. However, their reliability is rated as medium.

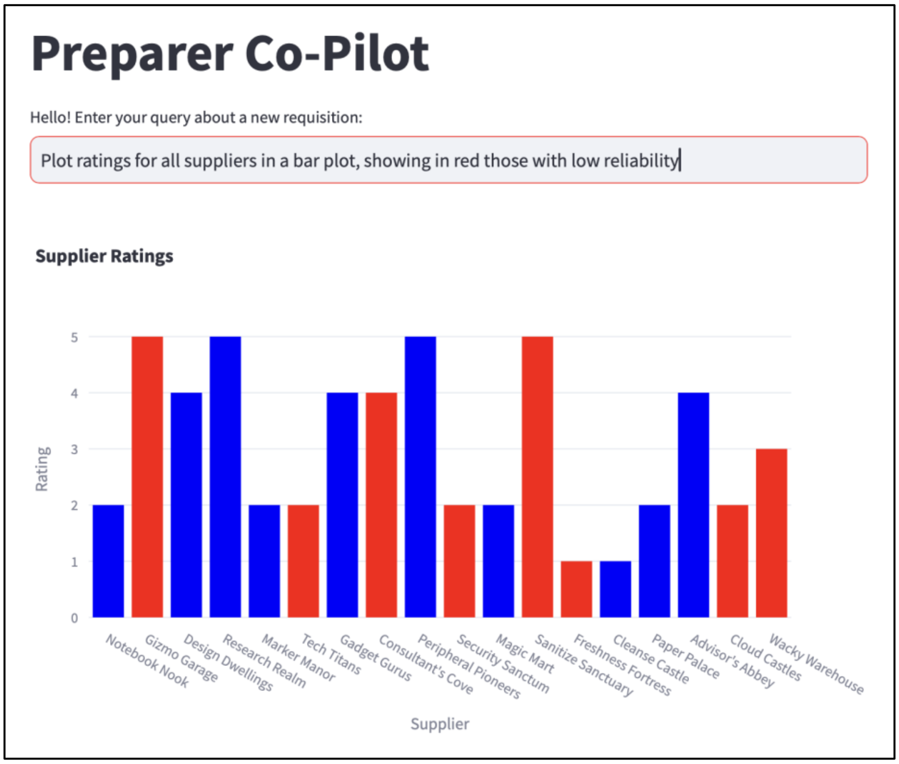
Preparer Co-Pilot

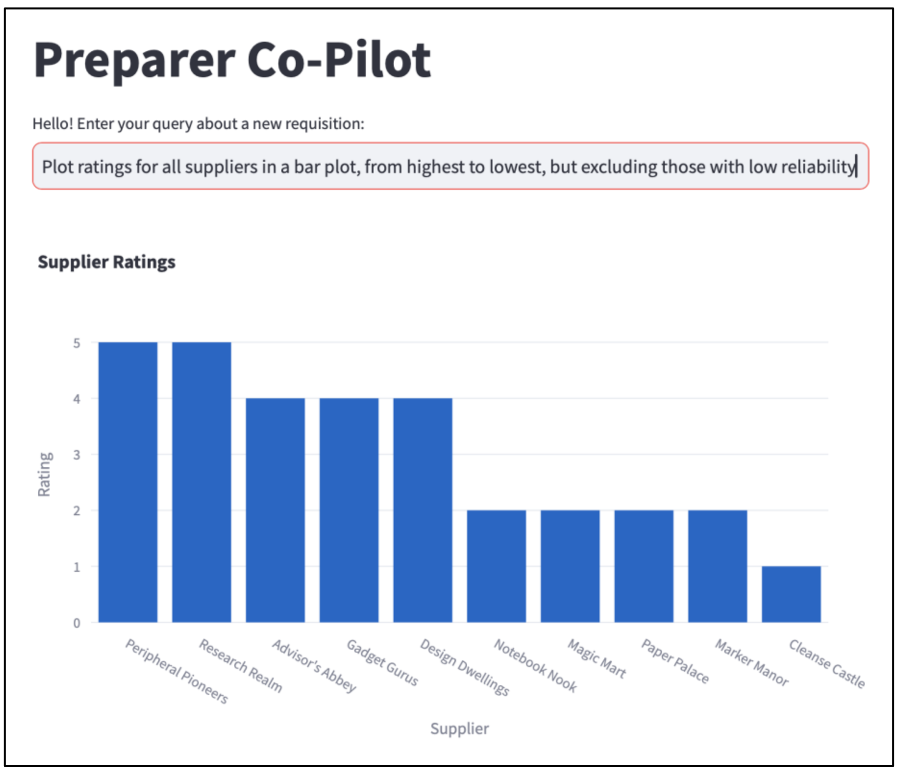
Hello! Enter your query about a new requisition:

What is the company's policy for suppliers responsibility? Reply without mentioning the name of the co

The policy for suppliers' responsibility includes several key areas:

1. Environmental Responsibility: Suppliers should optimize the use of natural resources, establish greenhouse gas emissions reduction goals, improve energy efficiency, manage waste responsibly, and ensure products do not contain restricted or banned materials. They should also have policies and management systems to identify, prevent, mitigate and account for impacts on the environment in their own operations and their supply base.
2. Quality: Suppliers must meet established minimum quality requirements, including compliance with regulations. They are expected to establish and maintain quality requirements to protect the integrity of the goods and services provided, continuously improve the quality and efficiency of processes, and support quality audits.
3. Laws and Regulations: Suppliers are expected to operate in compliance with all applicable laws and regulations of the countries, states and localities in which they operate. This includes laws and regulations related to ethical business practices, quality, labor and employment practices, as well as health, safety and environmental protection.
4. Ethics and Business Conduct: Suppliers are expected to behave ethically and with integrity in all business transactions. This includes upholding standards for fair business practices, prohibiting the payment of bribes or kickbacks, ensuring workers can report concerns without threat of reprisal, and implementing safeguards to prevent improper use of intellectual property.


Preparer Co-Pilot
Hello! Enter your query about a new requisition:
Plot ratings for all suppliers in a bar plot, showing in red those with low reliability
Supplier Ratings
Rating
0
1
2
3
4
5
Notebook Nook
Gizmo Garage
Design Dwellings
Research Realm
Marker Manor
Tech Titans
Gadget Gurus
Consultant's Cove
Peripheral Pioneers
Security Sanctum
Magic Mart
Sanitize Sanctuary
Freshness Fortress
Cleanse Castle
Paper Palace
Advisor's Abbey
Cloud Castles
Wacky Warehouse
Supplier


Preparer Co-Pilot
Hello! Enter your query about a new requisition:
Plot ratings for all suppliers in a bar plot, from highest to lowest, but excluding those with low reliability
Supplier Ratings
Rating
0
1
2
3
4
5
Peripheral Pioneers
Research Realm
Advisor's Abbey
Gadget Gurus
Design Dwellings
Notebook Nook
Magic Mart
Paper Palace
Marker Manor
Cleanse Castle
Supplier