Interim Report

Researcher Name: Musadig Aliyev

Research Title: Implementations of Product Bundling and Pricing Strategies using Machine Learning Techniques in the Banking Industry

Project Objectives/Description:

The banking industry is constantly evolving due to technological advancements, changing customer demands and trends. Banks need to offer innovative products and services to their customers to stay competitive and attractive. Currently, banks in Azerbaijan compete with each other in terms of the amount of cashback they offer to customers. They offer cashback ranging from 2% to 6% for each transaction in specific merchant categories chosen by customers. In the banking applications, you can primarily choose one or two categories. Many customers have purchased cards from multiple banks and chosen various merchant services to receive cashback for their spending on different services or categories.

As ABB bank, our goal is to encourage customers to consolidate all their spending with us, which would help us increase our total turnover and profit. So we decided to implement an innovative strategy called product bundling and pricing, which entails combining multiple merchant categories into a single package or bundle at an optimal price.

With the implementation of this strategy, our bank can distinguish itself from other banks by offering ease, customization, and an enhanced customer experience. It can also influence the spending habits of existing customers, encouraging them to use our credit/debit cards more frequently. Furthermore, this innovative solution has the potential to drive the sale of a significant number of new cards and to bring new customers.

To implement the new strategy, I have decided to adopt a data-driven approach instead of creating bundles manually by domain experts. So, I have applied machine learning and statistical techniques to create optimized and customized bundles which lead to increased customer satisfaction, activation and loyalty.

This is an example bundle planned to be prepared as a result of this project.

Bundle 1	Percentage
Entertainment	3%
Restaurants	2%
Hotels	3%
Grocery	2%

Research Methodology:

My research topic involves the collection and analysis of numerical data, the application of machine learning techniques, the utilization of statistical analysis, and the measurement of the effectiveness of models. This approach can be considered a quantitative way of conducting research.

My research consisted of two main parts: Firstly, exploring bundling techniques and implementing them across different merchant categories. Secondly, conducting research on pricing techniques for bundles and implementing them accordingly.

In the first phase of the project, I explored techniques such as market basket analysis and clustering for bundling. Then, I applied the market basket analysis technique to transactional data.

For the second part of the project, I plan to apply clustering techniques. Additionally, I will continue my research on optimal pricing techniques for bundles and apply them to different merchant categories in bundles.

During the project, the following steps have been or will be undertaken to complete the implementation:

A. Research:

Research has been conducted in two ways: Firstly, Azerbaijani market research was performed, which involved analyzing the cashback systems of other banks. Secondly, papers were read to identify appropriate techniques for bundling and pricing.

B. Data Collection:

Since I am using bank's data I obtained necessary permissions and ensure that I adhere to legal and ethical standards. I have anonymized or encrypted the data where required. Then, I identified the specific tables or databases within the bank that contain the required data for my research.

In the Bundling model, I extracted raw transactional data for each service (merchant category) that I'm interested in for a period of 3 months.

Initially, the data had 7 columns as below:

customer_no: A unique identifier for each customer.

trn_id: A unique identifier for each transaction.

mcc: A unique identifier for each merchant.

trn_dt: The date when the transaction occurred.

mcc_description: The description of merchant category code.

acc_amt_lcy: The monetary value of the transaction.

card_product_name: The name of card which transaction have been made.

Data Collection for Pricing Model: I will use the raw data to create feature tables for the pricing model. These tables will contain information on customer characteristics such as transactional behavior, personal information, balance status and etc.

C. Data Preprocessing:

During the data collection phase, I encountered several challenges and issues that required data cleansing. Since I am using transactional data for the bundling model, there are some customers who do not have transactions on certain services (merchant categories). To handle this, I will fill the missing values with 0, as it means no transactions. Furthermore, I noticed that the data exhibits inconsistencies in formatting. These inconsistencies include variations in date formats and inconsistent data types. To address this issue, I standardized the formatting across the dataset. In addition, I removed outliers within my dataset which I

detected using visualization methods like box plots and statistical measures such as z-scores. After performing the initial preprocessing, I created a 'Merchant Category' column from the 'Merchant Category Code' using the international convention. For example, the code 5411 corresponds to the category 'Restaurants'.

D. Statistical Analysis:

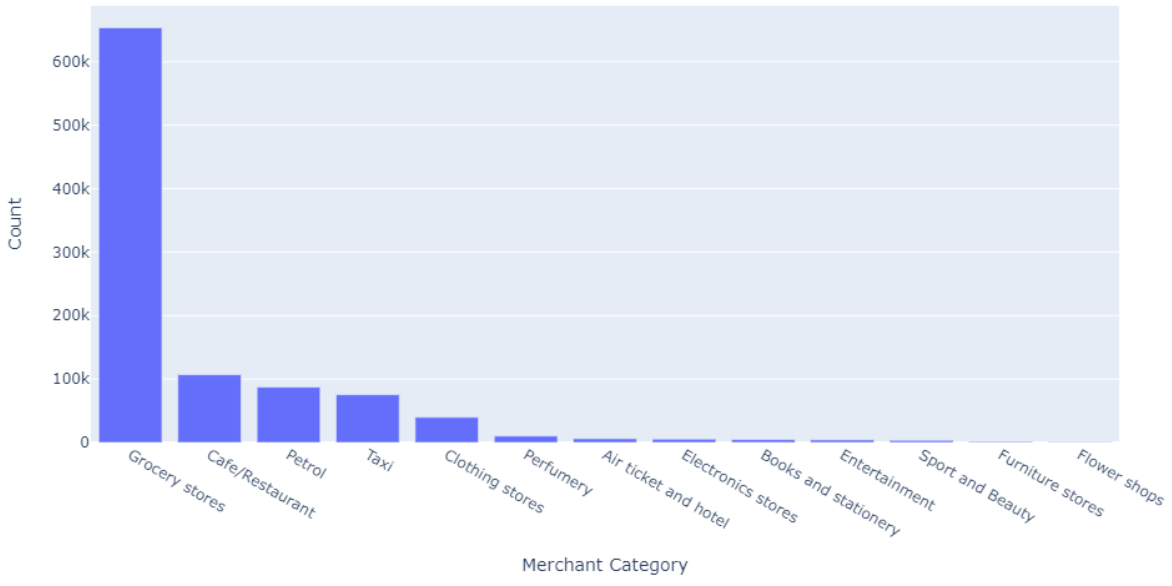
I have used several statistical techniques such as sampling, hypothesis testing, central tendency analysis and etc to do statistical analysis. I used moments to describe various characteristics of a probability distribution or a set of data.

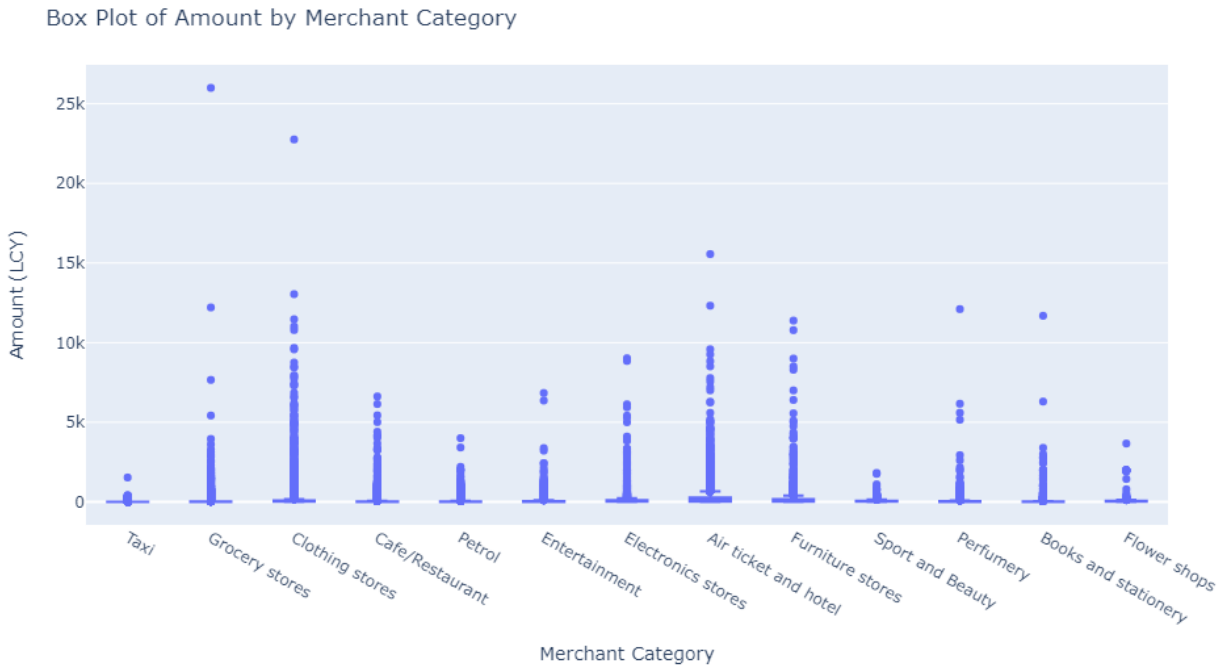
Mean: 26.76968894244116
Variance: 2267.748840113119
Skewness: 28.11798809811283
Kurtosis: 1612.2083578016152

Note: The details regarding statistical analysis were documented in Report 3.

E. Exploratory Data Analysis:

During the development phase of the project, I utilized various visualization techniques to gain insights and make informed decisions based on the data. For categorical data, I predominantly employed bar charts, while for numerical data, I relied on boxplots. These visualizations enabled me to better comprehend the information and take appropriate actions accordingly.





Note: The details regarding exploratory data analysis were documented in Report 3.

F. Baseline Model Development:

I used market basket analysis, also known as affinity analysis as a technique to understand the relationships between frequently used merchant services by customers. It is commonly applied in the retail industry to analyze transactional data to identify associations between transactions. The goal is to find patterns for product bundling. The output of market basket analysis is typically represented in the form of association rules, which consist of an antecedent (the items already present in the basket) and a consequent (the items that tend to be added to the basket). These rules are quantified by metrics such as support, confidence, and lift, which indicate the strength and significance of the associations.

Initial Model Results:

	Item 1	Item 2	Item 3	Item 4
0	category_Cafe/Restaurant	category_Grocery stores	category_Clothing stores	category_Air ticket and hotel
1	category_Cafe/Restaurant	category_Grocery stores	category_Air ticket and hotel	category_Petrol
2	category_Cafe/Restaurant	category_Grocery stores	category_Clothing stores	category_Books and stationery
3	category_Clothing stores	category_Grocery stores	category_Cafe/Restaurant	category_Entertainment
4	category_Clothing stores	category_Grocery stores	category_Perfumery	category_Cafe/Restaurant
5	category_Clothing stores	category_Grocery stores	category_Cafe/Restaurant	category_Petrol
6	category_Sport and Beauty	category_Grocery stores	category_Clothing stores	category_Cafe/Restaurant
7	category_Taxi	category_Clothing stores	category_Grocery stores	category_Cafe/Restaurant
8	category_Grocery stores	category_Perfumery	category_Cafe/Restaurant	category_Petrol
9	category_Taxi	category_Grocery stores	category_Perfumery	category_Cafe/Restaurant
10	category_Taxi	category_Grocery stores	category_Cafe/Restaurant	category_Petrol
11	category_Clothing stores	category_Grocery stores	category_Perfumery	category_Petrol

G. Measurement/Evaluation:

Measuring the correctness or accuracy of market basket analysis typically involves evaluating the performance of association rules generated by the analysis. Here are a few commonly used measures to assess the correctness of market basket analysis:

Support: Support measures the frequency of occurrence of a particular itemset (combination of items) in the dataset. It indicates how frequently a specific itemset appears in the transactions. Higher support values indicate a higher occurrence of the itemset, implying greater importance.

Confidence: Confidence measures the reliability or strength of the association between items in the form of an "if-then" rule. It calculates the conditional probability of finding an item (consequent) given the presence of another item (antecedent). Higher confidence values indicate a stronger association between the items.

Lift: Lift measures the degree of dependency between the antecedent and the consequent in an association rule. It compares the observed support of the rule with the expected support if the items were independent. A lift value greater than 1 indicates a positive association, while a value less than 1 suggests a negative association.

Initial Results:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
75	(category_Grocery stores, category_Perfumery, ...	(category_Cafe/Restaurant)	0	0	0	1	3	0	3
76	(category_Perfumery, category_Cafe/Restaurant,...	(category_Grocery stores)	0	1	0	1	1	0	6
77	(category_Perfumery, category_Petrol)	(category_Grocery stores, category_Cafe/Restau...	0	0	0	1	3	0	3
78	(category_Taxi, category_Perfumery, category_G...	(category_Cafe/Restaurant)	0	0	0	1	3	0	6
79	(category_Taxi, category_Perfumery, category_C...	(category_Grocery stores)	0	1	0	1	1	0	4
80	(category_Taxi, category_Perfumery)	(category_Grocery stores, category_Cafe/Restau...	0	0	0	1	3	0	5
81	(category_Taxi, category_Grocery stores, categ...	(category_Cafe/Restaurant)	0	0	0	1	3	0	3
82	(category_Taxi, category_Cafe/Restaurant, cate...	(category_Grocery stores)	0	1	0	1	1	0	5
83	(category_Taxi, category_Petrol)	(category_Grocery stores, category_Cafe/Restau...	0	0	0	1	3	0	3
84	(category_Clothing stores, category_Perfumery,...	(category_Grocery stores)	0	1	0	1	1	0	7

Future Works:

Firstly, I will apply clustering techniques to the bundling model and compare the results with the aim of determining which one to use as the final model. Additionally, I plan to apply the same workflow structure to the pricing component of the project. Then, I will use historical data to simulate different bundling scenarios and evaluate the effectiveness of the approach in increasing turnover and revenue. Finally, I aim to demonstrate new and innovative solutions for cashback systems by leveraging machine learning techniques in final presentation.