E-Commerce Customer Behavior Analysis

Introduction

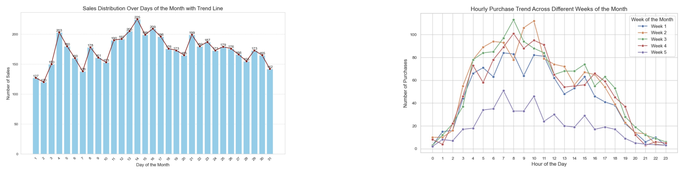
Online shopping has been becoming more common nowadays and a growing amount of consumers are making use of e-commerce platforms as a result. Differences in customer behavior have resulted due to the ever-changing usage of internet-based purchasing. As a result, our focus switched to a particular e-commerce site where consumers engage in a variety of actions, such as product browsing, cart addition, and purchase completion. We sought to extract valuable data and offer recommendations **to optimize website functionality and improve customer interaction, ultimately leading to potential revenue increases**. We analyzed user activities to uncover insights into purchasing behavior, the information gained from these interactions holds the capacity to provide details on user preferences and behavior. Our primary goal is to utilize data analytics and machine learning to address key research questions: **examining sales tendencies over time, pricing strategies analysis, conversion analysis, customer segmentation, association rule searching**.

Data Collection & Pre-processing

In this project, we collected a substantial dataset from Kaggle, specifically focusing on data from October, amounting to 5.6GB and randomly sampled 300k records. Data preparation involved **converting data types, eliminating duplicates, handling missing values, and ensuring values consistency**. We addressed nulls in 'category code' and 'brand' and removed zero-priced items. Outliers analysis on 'price' column revealed the presence of outlying high prices; we made sure that such items didn't cause any abnormal tendencies in customer behavior and decided to stay with them. Feature engineering included extracting date-related attributes. After these pre-processing steps, our dataset was ready for EDA and Model Building.

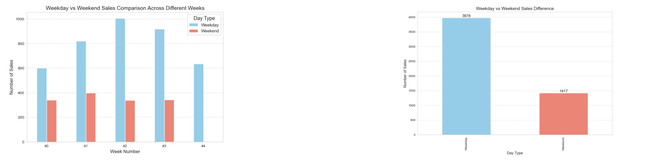
Time Series Analysis of Sales

In this part of our research, we tried to understand how sales are trending over time (Figure 1a). We tried to find out the most sales that happened per day, and for that, we summed up the total sales per day; We found the **highest number of purchases happened during 14th day**. In the same way, we tried to find out the sales trend based on the hour of the day for different weeks (Figure 1b in 1). **On average, the highest sales are between 7 am and 11 am**. We found out the sales differences between weekdays and weekends (Figure 2). These findings guide us in tailoring marketing campaigns, and fine-tuning inventory management. By aligning promotions and services with **peak periods on weekdays**, we enhance the overall customer experience and drive business efficiency. Our adaptive approach to the online shopping experience, coupled with targeted marketing messages and supply chain optimization, positions our store for sustained growth in the dynamic e-commerce landscape.



(a) Purchases Trend over Days (b) Hourly Purchases of Different Weeks

Figure 1. Time Series Analysis



(a) Weekly Analysis of Sales (b) Overall Weekdays vs Weekends

Figure 2. Purchases Time Series Analysis between Weekdays and Weekends

Conversion Analysis

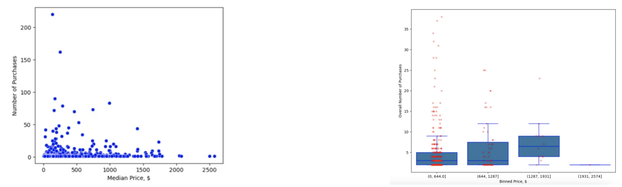
For this problem, next lists were created: top 20 items by views, top 20 by purchases, and top 20 by **conversion**. To make the top 20 items by conversion ratio, we filtered out items that have less than the 98th percentile of views, such that we could work with eye-attractive items only and make a fair comparison.

It was found, that 90% of top 20 viewed are included into top 20 purchased and that only 5% of top 20 viewed took place in top 20 by conversion. It means **sight-attractive items yield high purchases but low conversion**. So, sellers of top viewed product could think on how encourage viewers click 'buy'.

Analyzing the Pricing Strategies

In this section, one aimed to see if items with lower prices have higher purchase rates. The items with at least 1 purchase were examined, we binned the prices into 4 bins, filtered out extreme purchase rate values (1st and 99th percentile), and created the plot attached to Figure 3. Based on the findings, there is no interesting tendency except the 3rd bin.

However, this sharp increase of purchase rate mean in the 3rd bin can be explained through next: dealers of high-price products can't stay with bad-sellers on market unlike cheap-goods dealers. The reason behind it might be **markup**. For example, according to [1], average markup of HDMI Cables is 1000%, whereas average markup of smartphones is 79%. In addition, reinforcing this explanation, it was explored that all product allocated in the 3rd price bin were smartphones. That is, **the price doesn't define the purchase rate of an item**.



(a) Before Data Cleaning (b) After Data Cleaning

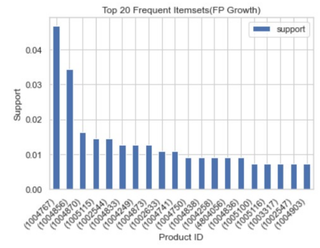
Figure 3. Purchase rate VS price

Market Basket Analysis

We conducted MBA to reveal the associations between the products. This analysis is conducted on the first 300k rows because we are missing the regular customers when selected randomly. For our project, we considered two algorithms which were FP-growth and Apriori [2]. Using both algorithms, we tried to find out the frequency(support) of each item-sets and the association rules (Figure 4). Based on the output of the association rules, we can conclude that **Samsung products are mostly bought together**. Knowing which products are frequently bought together, we can maximize the revenue by capitalizing on product placements. Based on our findings we can suggest complementary items to customers which can potentially increase the average order value.

Product ID	Category Code	Brand
1004856	electronics.smartphone	samsung
1004833	electronics.smartphone	samsung
1004870	electronics.smartphone	samsung
1004767	electronics.smartphone	samsung
1004856	electronics.smartphone	samsung
1004767	electronics.smartphone	samsung
1002544	electronics.smartphone	apple
1004767	electronics.smartphone	samsung

(a) Association Rules Pair-Wise



(b) Support(frequency) of top 20 Products

Figure 4. FP- Growth Algorithm Outcomes

Key concepts

Customer Value (CV), a part of the Customer Lifetime Value [3], aims to demonstrate long-term value of a client and it is defined by the following formula:

$$\text{CV} = \text{Average Transaction Price} \cdot \text{Average Number Of Transactions}$$

Conversion is widely used for evaluating the efficiency of selling and marketing campaigns.

$$\text{Conversion} = \frac{\text{Number Of Purchases}}{\text{Number Of Views}}$$

Markup aims to demonstrate the difference between how much a seller spent on a product and how much they made out of its sell.

$$\text{Markup} = \frac{\text{Gross Profit}}{\text{Sales Price}}$$

Customer Segmentation

Three dimensions were calculated for customer k-means clustering [2]: total spending, conversion rate, and **customer value**. Only regular customers (those who made at least 2 purchases) were included in the segmentation, and extreme values of spendings (1st and 99th percentile) were filtered out. The optimal number of clusters was defined by using the Elbow Method and Silhouette Analysis. In the Table 1, one can find the Inertia Values and Silhouette Scores per different number of clusters.

Number of Clusters	Inertia Value	Silhouette Score
2	182.50	0.49
3	110.77	0.55
4	**58.81**	**0.60**
5	46.35	0.52
6	36.32	0.50

Table 1. Measures per different number of clusters.

Four customer clusters associated with distinctly different purchase patterns were formulated. Their titles: **'Rock Stars', 'Dedicated', 'Know What They Want', and 'Least Active Customers'**. The visual description of these clusters can be found in the Figure 5. The e-commerce platform can take advantage of this model to treat customers differently in terms of **UX, ads, and recommendation system**.

Figure 5. Customers Clusters Description

References

[1] Johnson. 20 items with way higher than average product markup. `https://moneygenius.ca/blog/average-product-markup`. accessed: 12.09.2023.

[2] Han Kamber Pei. *"Data Mining Concepts and Techniques, Third Edition"*. Elsevier Inc., 2012.

[3] Sauro. *"Customer Analytics For Dummies"*. John Wiley Sons., 2015.