

# Project E-Commerce: Product Range Analysis

## Project Decomposition: Oleg Izenberg

### Project Goal Breakdown

#### Who's is our customers?

The primary reason why individuals choose to shop online instead of in-person is either to save time or due to limited spare time. It is important to note that the majority of household goods are considered necessities rather than luxuries.

As a result, it is essential to understand the buyer persona of the target audience, which typically consists of busy homeowners who require essential goods but have limited time to shop. However, there may be some differences in the circumstances of these customers, such as requiring several items after moving, replacing worn-out items at a more leisurely pace, or urgently needing to replace an item.

#### What needs to be achieved and why?

To maximize customer satisfaction and drive sales growth, it is crucial to evaluate the performance of each product in a logical and unbiased manner to ensure they meet customer needs. This evaluation should include comparing individual product performance with that of the entire product range.

To achieve this, it is important to obtain an accurate understanding of the current state of the product range and identify any trends. This analysis should begin by examining the product range at a high-level categorization and then drill down to specific products. Throughout this process, it is crucial to consistently compare the average metric value for the group of products with the corresponding metric value at the individual product level.

#### What decisions will be made based on the analysis?

The analysis can aid in making decisions regarding keeping, replacing, or removing a product based on its performance. By identifying the best-performing products and removing or replacing the underperforming ones, businesses can optimize their product offerings, increase revenue, and enhance customer satisfaction.

Furthermore, we can offer substitute products when a product is out of stock, prepare for seasonal boosts in sales during specific times and dates, and make other data-driven decisions to improve their overall performance.

## Project Goal

The goal is to understand the target audience, evaluate each product's performance, and optimize the product range to increase customer satisfaction and drive sales growth.

The analysis will help make data-driven decisions to keep, replace, or remove products and offer substitutes when products are out of stock. Seasonal sales boosts will also be considered.

## Tasks Breakdown:

### Data overview and data Preprocessing

- Initial overview of the data in order to identify preprocessing tasks
- Checking N/A
- Checking Duplicates
- Checking data types
- Standardize columns
- Adding sparse columns for 'date' and 'time'
- Total price aggregation column
- Making a category column for the products using lemmatization

### Exploratory Data Analysis

- Summary statistics
- Unique values and counts in each column
- Product category distribution breakdown
- Product distribution breakdown
- Dealing with outliers on category and product level
- Summary statistics and distribution breakdown W/O outliers

### Product Range Analysis

- Product category breakdown and top performing categories by the number of units sold and total price
- Product category breakdown and top performing categories by the number of units sold and total price ( 'date' and 'time' )
- Top performing Products within a category by the number of units sold and total price
- Top performing Products within a category by the number of units sold and total price ( 'date' and 'time' )
- Top performing Products by the number of units sold and total price
- Top performing Products by the number of units sold and total price ( 'date' and 'time' )

Additionally:

- During what time period are there more orders, including the date and time of day?
- which items/categories are ordered together

## Statistical Hypothesis testing

1. What categories or products generate the highest revenue?

- Hypothesis:
  - Products with a higher price-point per unit may generate more revenue, even if they have lower sales volume.
  - Even if they have a higher sales volume, products with a lower price per unit may not necessarily generate more revenue.

The Genral approach for this Hypothesis testing will be:

1. analyzing revenue data to identify which categories or products generate the highest and lowest revenue
  2. analyzing price per unit data to identify which categories or products generate the highest and lowest price per unit
  3. Compere the total revenue and count for items with low price point and high price points
- Use Kmeans to cluster products by price points and performance

1. Does seasonality affect revenue ?

- Hypothesis:
  - Seasonal products may generate higher profits during specific times of the year

The Genral approach for this Hypothesis testing will be:

1. Check the trends of total revenue per item/category by date
2. Check the trends of total revenue per item/category by time

1. Does top selling items that are sold with less sold items affect sales and profitabiliy?

- Hypothesis:
  - Top selling items that are sold with less sold items may lead to an increase in overall sales and profitability due to the "halo effect" where the popularity of the top selling item positively influences the less sold item.
  - Offering less sold items as a bundle or package deal with the top selling item may encourage customers to purchase both and increase overall sales and profitability.

The Genral approach for this Hypothesis testing will be:

1. Check which items are the top sellers and least sold.
2. Use Apriori methode to identifying an associations between products

## 1. Is there Mutually substitutive products?

- Hypothesis:
  - Customers may consider some products in the product assortment as interchangeable, such that if one product is out of stock they are willing to substitute it with another product

The Genral approach for this Hypothesis testing will be:

1. Analyzing product sales data to identify which products are frequently sold together.
2. Examining purchase patterns and determine which products are commonly substituted for one another. (if customers frequently purchase either product A or product B in the same transaction, it suggests that these products are interchangeable or substitutive)

## conclusions

I am confident that by conducting the project as per this plan, I will achieve the project's goals and provide an analysis that can help make decisions to optimize product offerings, increase revenue, and enhance customer satisfaction.

## References

<https://www.dotactiv.com/blog/product-range-review>

<https://www.beuniqueness.co.uk/solutions/household-items>

<https://mipler.com/blog/ecommerce-product-performance-analysis/>

<https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>

<https://retalon.com/blog/product-clustering#:~:text=Product%20clusters%20are%20groups%20of,brand%2C%20function%2C%20and>

