Wholesale customers: The impact of region on sales of products

Artur Bruev
Mentors: Dr. Christelle Scharff

Seidenberg Undergraduate Research
Pace University, Seidenberg School of CSIS

Abstract

This study analyzes a dataset comprising information on product purchasing behavior in the context of two distinct sales channels across multiple regions. The dataset includes various product categories, such as 'Fresh,' 'Milk,' 'Grocery,' 'Frozen,' 'Detergents_Paper,' and 'Delicassen,' which are examined for their relationship to the channels of distribution. The analysis employs a combination of exploratory data visualization techniques to uncover patterns and relationships within the data.

Scatter plots and bar charts were utilized to assess the spending behavior across different product categories and to examine the distribution of purchases between the two sales channels. Through these visualizations, significant differences and patterns were observed in customer spending preferences.

A box plot and swarm plot were used to provide further insight into the distributions of specific product categories within each sales channel, allowing for the identification of potential outliers and variations. Violin plots helped visualize the distribution of 'Grocery' spending within each channel.

The findings from these visualizations provide valuable insights into the characteristics of customers in different sales channels. This analysis aims to enhance our understanding of purchasing behavior and its implications for marketing and distribution strategies.

In conclusion, this study leverages visualizations to explore the relationship between product categories and sales channels. The results offer actionable insights for businesses looking to optimize their marketing and distribution strategies based on customer spending behavior. Further detailed analysis and segmentation can be conducted to refine these insights for practical implementation.

Research Question

How ‘region’ (column) of trading affect on categories of products purchased?

Related Work

Several Papers and books was used during a work. Logistic Regression was performed as in paper [2-3]. Several books [5-8] was connect to correct data visualization. Neo4j App was used as a part of EDA for better understand of data [4]. SubCluster alghoritm was tried [1] on that dataset.

Dataset

The dataset consists of a collection of data points representing various customers and their purchasing behavior. It includes information about two key features:**Channel**: This categorical variable indicates the sales channel through which the customer makes their purchases. There are two distinct channels represented as '1' and '2.'**Region**: Another categorical variable that specifies the region in which the customer is located. It includes three different regions, denoted by '1,' '2,' and '3.'

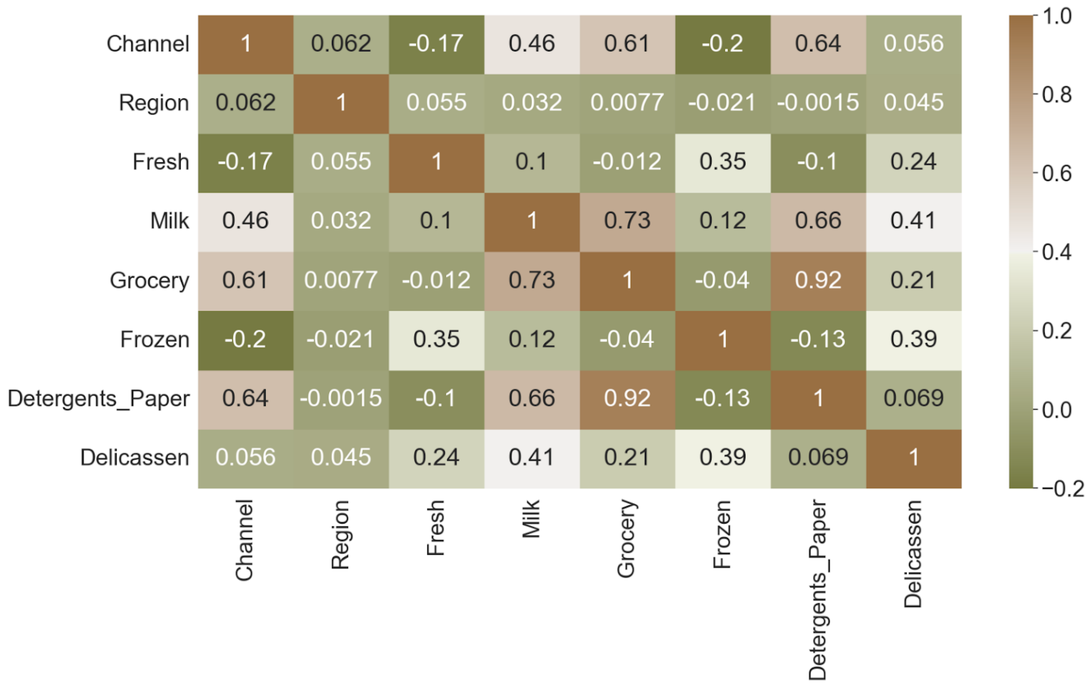
In addition to these categorical features, the dataset provides data on the following product categories:

Fresh: The amount of spending on fresh products.
Milk: The amount of spending on dairy and milk products.
Grocery: The amount of spending on grocery items.
Frozen: The amount of spending on frozen products.
Detergents_Paper: The amount of spending on detergents and paper products.
Delicassen: The amount of spending on delicatessen items.

The dataset offers insights into customer spending patterns and how they vary across different regions and sales channels. It is valuable for various types of analyses, including customer segmentation, market research, and the development of marketing and distribution strategies tailored to different customer groups.

Methodology

Dataset was downloaded from online resource and contains 8 different columns. It don’t have any Null values, so next step was CorrelationMap. All develop was complete on python notebook. Heatmap showed interesting connections, which next were used to find insights. I found interesting to use Linear Regression model on data, to predict values of channel, for better understanding of spending on items.



Results

The results of the manipulations and analysis conducted on the customer purchasing behavior dataset are summarized below:

Product Category Insights:

•Visualizations, such as box plots and swarm plots, provided detailed insights into spending distributions within each sales channel.
•Variations in spending patterns across product categories, particularly 'Grocery,' were identified, indicating distinct customer behaviors.

Regional Influence:

•The 'Region' feature classifies customers into three regions ('Region 1,' 'Region 2,' and 'Region 3').
•The dataset serves as a foundation for further regional analyses, though these specific analyses were not conducted.

Marketing and Distribution Strategies:

1. Businesses can use the insights to tailor marketing and distribution strategies based on customer preferences within each sales channel.
2. The data-driven approach helps optimize product offerings, stock management, and regional targeting.

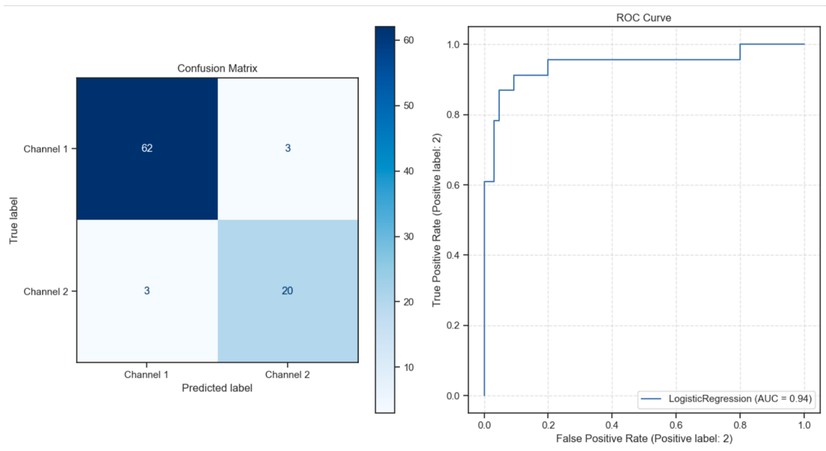
Machine Learning Model:

3. A logistic regression model was applied to predict sales channel behavior based on customer spending patterns.
4. The model's performance, assessed through accuracy and visualizations (e.g., confusion matrices and ROC curves), provides a tool for channel prediction and customer segmentation.

Recommendations:

5. The dataset, combined with visualizations and modeling, offers actionable recommendations for businesses to enhance marketing and distribution strategies.
6. Tailored approaches can be adopted to meet the unique needs of customers in each sales channel.

In summary, the results of these manipulations provide insights into customer purchasing behavior, aiding businesses in developing more effective marketing and distribution strategies. The combination of exploratory data analysis, visualizations, and modeling facilitates data-driven decision-making for improving customer satisfaction and operational efficiency.



Conclusions

In conclusion, the analysis of the customer purchasing behavior dataset has unveiled valuable insights into the dynamics of customer spending across different sales channels and product categories. Visualizations and statistical techniques have exposed variations in preferences and patterns that are pivotal for businesses seeking to optimize their marketing and distribution strategies. By identifying distinct behaviors within each sales channel and product category, companies can tailor their approaches to better meet customer needs.

As a variable, Region (which present theoretical area of human living) has a low Heatmap coefficient to another variables. Despite that, it can be used to characterize a set of columns, which will present preferences of people in area. Geographical location, size of territory, nature of the relief and climate are columns, which is not present in dataset, but can be important to review Region and understand people’s preferences of products.

Potential future research may involve deeper regional analyses, advanced machine learning models for customer segmentation, and exploration of additional factors influencing customer behavior.

References

[1] B.Jaya Lakshmi, K.B.Madhuri, M.Shashi, "An Efficient Algorithm for Density Based Subspace Clustering with Dynamic Parameter Setting", International Journal of Information Technology and Computer Science(IJITCS), Vol.9, No.6, pp.27-33, 2017. DOI:10.5815/ijitcs.2017.06.04
[2] Y. Yang and N. Li, "Research on Residents’ Travel Behavior Based on Multiple Logistic Regression Model," in *IEEE Access*, vol. 11, pp. 74759-74767, 2023, doi: 10.1109/ACCESS.2023.3297497.
[3] Y. -X. Wu, Z. -N. Hu, Y. -Y. Wang and F. Min, "Rare Potential Poor Household Identification With a Focus Embedded Logistic Regression," in *IEEE Access*, vol. 10, pp. 32954-32972, 2022, doi: 10.1109/ACCESS.2022.3161574.
[4] Neo4J in Action Author: Jonas Partner, Aleksa Vukotic, and Nicki Watt
[5] Storytelling with data: a data visualization guide for business professionals,Author: Cole Nussbaumer Knaflic
[6]Fundamentals of Data Visualization: A primer on making Informative and Compelling Figures, Author: Claus O. Wilke
[7] Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks, Author: Jonathan Schwabish
[8] The visual display of quantitative information, Author: Edward R. Tufte

printed by MegaPrint Inc. www.postersession.com