GEORGIAN COLLEGE

COURSE: 23S Math for Data Analytics - 02 (BDAT1005-23S-30736)

TOPIC: Data Set Exploration Part 4

SUBTOPIC: Analysis Report: Retail Customer Behavior and Sales Trends And Forecast

DATE: 13/08/2023

LECTURER: Ehsan Pourjavad

STUDENT: Anthony Ofordum

STUDENT ID: BDAT200559474

1.0 Introduction

1.1 Background

1.2 Objectives

1.3 Scope

2.0 Dataset Description

2.1 Overview

2.2 Data Collection

2.3 Data Dictionary

3.0 Data Limitations

4.0 Charts

4.1 Price Range

4.2 Customer Payment Methods

4.3 Gender Distribution

4.4 Shopping Mall Popularity

4.5 Age Group Counts

5.0 Research Questions and Hypotheses

5.1 Research Question 1: Most Popular Products/Categories

5.2 Research Question 2: Preferred Payment Methods

5.3 Research Question 3: Age Distribution Across Product Categories

5.4 Research Question 4: Variation in Sales Revenue Across shopping malls

5.5 Research Question 5: Sales Revenue Distribution Across Product Categories

5.6 Research Question 6: Differences in Shopping Behavior between Genders

6.0 Additional Analysis

6.1 Hypothesis 1: Gender Impact on Total Amount Spent

6.2 Hypothesis 2: Association between Age and Category of Items Purchased

6.3 Hypothesis 3: Relationship between shopping mall Choice and Payment Method

7.0 Key Takeaway

8.0 Recommendations

9.0 Conclusion

1.0 Introduction

1.1 Background:

In the dynamic realm of retail, comprehending customer actions and market trends is crucial for businesses to remain competitive and relevant. With the emergence of technology and data analysis, retailers have access to a wealth of information that can guide their strategies, refine operations, and elevate customer experiences. This report seeks to delve into a diverse array of data-driven insights, spanning customer payment preferences, gender distribution, shopping center popularity, age group inclinations, and sales patterns. Through the analysis of these aspects, retailers can deepen their understanding of their target audience, make well-informed choices, and adjust their approaches to cater to evolving consumer needs.

1.2 Objectives:

The main aims of this report are as follows:

Exploring Data: To scrutinize and interpret a variety of data insights related to customer behavior, payment methods, gender distribution, shopping center preferences, age-linked tendencies, and sales trends.

Strategic Guidance: To provide actionable guidance for retailers, enabling them to fine-tune their strategies, tailor marketing initiatives, and optimize operations based on the identified trends and patterns.

Customer-Centric Focus: To underscore the importance of data-driven decision-making in enriching customer experiences and cultivating stronger connections with distinct consumer segments.

Projection and Readiness: To demonstrate the value of analyzing historical sales trends and anomalies in predicting forthcoming challenges and opportunities, empowering retailers to develop proactive strategies.

1.3 Scope:

This report's scope encompasses a comprehensive analysis of diverse data-driven insights pertinent to the retail industry. It encompasses the following key domains:

Customer Payment Preferences: An assessment of the distribution of customer payment methods, highlighting the prevalence of cash, credit cards, and debit cards, along with their implications for retailers.

Gender Dispersion: An exploration of the gender distribution among customers, illuminating the dominant gender group and its implications for marketing strategies.

Shopping Center Popularity: An evaluation of the popularity of different shopping centers based on sales and customer visits, offering insights into strategic locations for retailers to target.

Age Group Inclinations: A scrutiny of age-related shopping behaviors and preferences, equipping retailers with insights into the target audience for various product categories.

Sales Patterns: An analysis of historical sales trends and irregularities across distinct months and years, aiding retailers in understanding customer behaviors and planning for the future.

2.0 Dataset Description

2.1 Overview

The dataset consists of retail customer shopping transactions, providing information about consumer behavior and purchasing patterns.

2.2 Data Collection

The data was obtained from an online repository specializing in retail-related datasets. Relevant datasets related to customer shopping transactions were identified using specific search criteria.

2.3 Data Dictionary

The data dictionary includes variables such as invoice number, customer ID, gender, age, category, quantity, price, payment method, invoice date, and shopping mall.

3.0 Data Limitations

The dataset has limitations, including sampling bias, data quality issues, time-bound findings, and the absence of external factors influencing customer behavior.

4.0 Charts

The charts provide insights into customer payment methods, gender distribution, shopping mall popularity, and age group counts.

A graph showing a number of values

Description automatically generated

Price Range: The dataset includes prices that vary from a minimum of $5.23 to a maximum of $5250.00, demonstrating a broad spectrum of price points for the items that were purchased.

Central Tendency: The median price is $203.30, which represents the middle value in the dataset. The mean price is higher at $689.26, indicating that there are some higher-priced items that contribute to the average. These measures provide insights into the typical or average price range of the items.

Variability: The standard deviation of approximately 941.1845672 suggests a significant amount of variability or spread in the prices. This indicates that there is a wide range of prices observed in the dataset, with some items priced significantly higher or lower than the average.

A graph of a distribution of payment

Description automatically generated

The chart depicts the breakdown of customer payment methods, encompassing cash, credit card, and debit card. It demonstrates that cash is the most prevalent payment option, having the highest frequency count. In contrast, the usage of debit cards is relatively infrequent, with the lowest count. These findings shed light on customer preferences and underscore the dominant payment mode observed in the dataset.

A graph with numbers and a bar chart

Description automatically generated with medium confidence

The chart illustrates the distribution of customer gender in terms of percentages, showcasing the differing proportions of males and females. It reveals that females make up most customers with the highest percentage distribution. In contrast, males have the lowest percentage, indicating a relatively smaller presence within the dataset. This data sheds light on the gender composition of the customer base, highlighting the dominant gender group among the customers.

A graph of a number of visitors

Description automatically generated

The bar chart displays the frequency of shopping malls in the dataset. According to the summary provided, "Mall of Istanbul" and "Kanyon" have the highest occurrence, with approximately 19,943 and 19,823 instances, respectively. They are followed by "Metrocity" with around 15,011 occurrences and "Metropol AVM" with approximately 10,161. The remaining malls, including "Emaar Square Mall," "Cevahir AVM," "Forum Istanbul," "Viaport Outlet," and "Zorlu Center," have lower counts ranging from 4,811 to 5,075.

The key takeaway from this chart is that "Mall of Istanbul" and "Kanyon" are the most prevalent shopping malls in the dataset, indicated by their high frequencies. "Metrocity" and "Metropol AVM" also appear, albeit with slightly lower occurrences. In contrast, the lower counts for "Emaar Square Mall," "Cevahir AVM," "Forum Istanbul," "Viaport Outlet," and "Zorlu Center" suggest that they may be less popular or less frequently represented in the dataset.A graph of a graph with numbers and a bar

Description automatically generated with medium confidence

The graph depicts the number of people in various age groups. The ages range from 18 to 27, 28 to 37, 38 to 47, 48 to 57, 58 to 67, and 68 to 77. Each age group has a different count: 18-27 (19,205), 28-37 (19,224), 38-47 (19,216), 48-57 (19,020), 58-67 (18,998), and 68-77 (3,794).

The most important takeaway from this chart is that the age groups 28-37, 18-27, and 38-47 have the highest visitor counts, indicating that these age groups visit shopping malls most frequently. The age group 68-77, on the other hand, has the lowest count, indicating that people in this age range visit shopping malls less frequently than others. This information can be useful for understanding the target audience and tailoring marketing or promotional strategies to attract visitors of various ages to shopping malls.

5.0 Research Questions and Hypotheses

5.1 Research Question 1:

What are the most popular products or categories among customers?

By analyzing the sales data, one can identify the top-selling products or categories based on the number of purchases or revenue generated.

5.2 Research Question 2:

What are the preferred payment methods among customers?

By examining the payment information, one can determine the most used payment methods (credit cards, cash, online wallets, etc.) among customers.

5.3 Research Question 3:

What is the distribution of customer age across different product categories?

By examining the data, one can gain insights into the distribution of customer age within each product category, highlighting which age groups are more inclined to purchase certain types of products.

5.4 Research Question 4:

What is the variation in sales revenue across different shopping malls?

By examining the sales revenue data, it is possible to analyze the variation in revenue generated across different shopping malls. This analysis can provide insights into the performance and attractiveness of each shopping mall in terms of generating sales.

5.5 Research Question 5:

What is the distribution of sales revenue (PRICE) across different product categories?

By examining the dataset, one can analyze the sales revenue distribution across various product categories. This analysis will provide insights into which product categories generate higher or lower sales revenue.

5.6 Research Question 6:

What are the differences in shopping behavior between male and female customers in terms of the number of items purchased?

6.0 Additional Analysis

Hypothesis 1: Gender has an impact on the total amount spent:



The regression analysis results indicate a very weak relationship between the predictor variable(s) and the dependent variable, Total Amount Spent. The multiple correlation coefficient (Multiple R) of 0.001 suggests a negligible association. The coefficient of determination (R Square) indicates that only an extremely small proportion (0.0000021) of the variance in the dependent variable is explained by the predictors.

The adjusted R Square is negative, indicating that the model does not fit the data well. This suggests that the chosen predictor variable(s) may not be sufficient in explaining the variation in the dependent variable.

The ANOVA table assesses the overall significance of the regression model. The F-statistic of 0.211 compares the variability explained by the model to the unexplained variability. The significance F-value of 0.646 indicates that the regression model is not statistically significant at the conventional significance level.

Examining the coefficients, the intercept (0.5986) represents the estimated log-odds of the dependent variable when all predictors are zero. The provided predictor variable labeled as "1500.4" lacks clear interpretation and significance, as its coefficient estimate (-7.59E-07) is very small and not statistically significant (p = 0.646).

Overall, the regression results suggest that the given predictor variable(s) do not strongly influence the Total Amount Spent. The model does not adequately explain the variation in the dependent variable. Further analysis, including additional predictors or considering alternative models, may be needed to gain deeper insights into the factors influencing the Total Amount Spent.

Hypothesis 2: Age is associated with the category of items purchased:



The objective of this analysis was to examine the relationship between age and the types of items purchased. To accomplish this, a statistical test called chi-squared test of independence was conducted. A contingency table was created to summarize the number of items in each category for different age groups. By performing the chi-squared test, a p-value of 0.435569447 was obtained. Considering the common significance level of 0.05, the obtained p-value suggests that there is insufficient evidence to reject the null hypothesis. Hence, no significant association between age and the category of items purchased was found. It is important to acknowledge that this analysis was based on the available data and the specific categories considered in the study. Other factors or variables that were not taken into account may influence the relationship between age and item categories. Further research or exploration of alternative statistical tests may be necessary to gain a deeper understanding of this relationship.

Hypothesis 3: Shopping mall choice is related to the preferred payment method.



This analysis aimed to investigate the relationship between shopping mall choice and payment method. A chi-squared test of independence was conducted using the provided data on payment method counts for each shopping mall. The test assessed if there was a significant association between the variables.

The results of the chi-squared test indicated that there was no significant relationship between shopping mall choice and payment method. The calculated p-value of 0.723464654 was greater than the chosen significance level, suggesting that the null hypothesis of independence cannot be rejected.

While the findings suggest no significant association in this particular dataset, it is important to consider the limitations of the analysis. The results are specific to the dataset and may not generalize to other populations or contexts. Additionally, the assumptions of the chi-squared test should be taken into account when interpreting the results.

Further research with larger sample sizes or exploring additional variables could provide a more comprehensive understanding of the relationship between shopping mall choice and payment method. These findings contribute to the broader understanding of consumer behavior in retail settings and highlight the need for considering various factors in studying consumer decision-making.

Sales Trend Overview (2021-2023)

Annual Highlights:

2021: The year saw sustained high sales with January leading the charge at 7,932,626.63, while November lagged with a figure of 4,995,141.11. The cumulative sales for 2021 summed up to 31,383,675.29.

2022: The sales trajectory mirrored that of 2021. January once again recorded the top sales of 7,932,626.63, while November was at the lower end, registering 4,995,141.11. The year's overall sales amounted to 31,383,675.29.

2023 (Until March): While January and February's sales were in line with the previous years, March experienced a sharp fall, registering just 683,721.31 – a significant drop from its past records.

Monthly Breakdown:

January: Steadily, this month topped the sales chart across the three years, consistently nearing the 8 million marks.

March: Sales in 2021 and 2022 were robust, exceeding 2.5 million. However, 2023 marked a drastic decline, plunging below 700,000.

February, April, & May: These months recorded sales between 5 and 7 million, showcasing only slight shifts over the years.

June through October: Sales in these months remained steady, averaging about 5.5 million, without notable changes between 2021 and 2022.

November: Traditionally, this month's sales were modest, hovering around the 5 million thresholds.

December: Consistently, the sales during this month settled around the 5.3 million marks.

Sales Forecast Analysis for January 2024

Historical Data (2021-2023):

2021: January sales were robust, recording a high value of 5,712,613.83.

2022: There was a noticeable decline in sales to 2,609,692.05, marking almost a 50% decrease from the previous year.

2023: The decrease trended slightly further, with sales settling at 2,559,847.27, though this dip wasn't as pronounced as between 2021 and 2022.

Forecast for 2024:

Projected Sales: The forecast predicts a drastic slump in January 2024 sales, with an expected value of just 589,007.97. This is significantly lower than any of the past years, suggesting potential challenges or market shifts affecting sales in 2024.

Confidence Bounds: The model's confidence interval is wide, ranging from -889,105.36 to 2,067,121.30. This wide range indicates high uncertainty in the forecasted value. In the worst-case scenario (based on the lower bound), sales could even be negative, which might mean returns or massive cancellations. On the optimistic end (upper bound), sales could bounce back to 2022 levels.

7.0 Key Takeaway:

Cash is the most prevalent payment method among customers, while debit cards are used less frequently. Retailers should ensure they have sufficient cash-handling processes in place to cater to customer preferences.

Females comprise the majority of customers, indicating the dominant gender group in the dataset. Retailers can tailor marketing strategies and product assortments to cater to the preferences of female customers.

"Mall of Istanbul" and "Kanyon" are the most popular shopping malls, followed by "Metrocity" and "Metropol AVM." Retailers should consider these malls as key locations for targeting customers and expanding their presence.

Age groups 28-37, 18-27, and 38-47 visit shopping malls most frequently, while the age group 68-77 has the lowest count. Retailers should focus on marketing efforts and promotions that appeal to the age groups that visit shopping malls more frequently.

The regression analysis suggests that gender does not have a significant impact on the total amount spent by customers. Therefore, retailers should focus on other factors, such as product assortment, pricing strategies, and customer service, to drive sales and increase the total amount spent.

The analysis does not find a significant association between age and the category of items purchased. Retailers should consider a broader range of factors, such as personal preferences, marketing campaigns, and product availability, when determining product categories to target different age groups.

The analysis does not find a significant relationship between shopping mall choice and the preferred payment method. Retailers should prioritize providing a variety of payment options in all shopping malls to accommodate different customer preferences.

Sales patterns were remarkably consistent from 2021 to 2022, with minor monthly fluctuations.

The sales dip in March 2023 is of concern, warranting a deeper examination. This could be a result of various external influences or possible inconsistencies in data recording.

In spite of varying monthly performances, the annual sales figures for 2021 and 2022 were stable, hinting at a dependable market demand and solid business operations.

Forecast for January 2024:

Projected Sales: The forecast predicts a drastic slump in January 2024 sales, with an expected value of just 589,007.97. This is significantly lower than any of the past years, suggesting potential challenges or market shifts affecting sales in 2024.

Confidence Bounds: The model's confidence interval is wide, ranging from -889,105.36 to 2,067,121.30. This wide range indicates high uncertainty in the forecasted value. In the worst-case scenario (based on the lower bound), sales could even be negative, which might mean returns or massive cancellations. On the optimistic end (upper bound), sales could bounce back to 2022 levels.

8.0 Recommendations:

Diversify Payment Methods: While cash is the dominant payment method, retailers should also consider expanding their payment options to accommodate customer preferences. Integrating popular digital payment methods and offering convenient and secure payment solutions can enhance the shopping experience and attract a broader customer base.

Gender-Specific Marketing Strategies: Given the higher presence of female customers, retailers should develop targeted marketing campaigns that appeal to the preferences and needs of this demographic. This may involve product selection, visual merchandising, and tailored promotions that resonate with female customers.

Strategic Partnerships with Popular Malls: To capitalize on the popularity of "Mall of Istanbul," "Kanyon," "Metrocity," and "Metropol AVM," retailers should explore strategic partnerships and collaborations to increase brand visibility and drive foot traffic. This can include sponsorships, exclusive offers, or collaborative events that attract customers to both the mall and the retail store.

Age-Appropriate Marketing Initiatives: Understanding the age groups that visit shopping malls more frequently enables retailers to develop targeted marketing initiatives. Retailers can customize their communication channels, promotions, and product offerings to align with the preferences and interests of each age group. This may involve social media campaigns, loyalty programs, and in-store experiences tailored to specific age demographics.

Continuous Data Analysis: Retailers should continue analyzing customer behavior, payment trends, and sales data to identify evolving patterns and adapt their strategies accordingly. Regularly updating and expanding the dataset will provide a more comprehensive understanding of customer preferences and allow for data-driven decision-making in the retail industry.

March 2023 Anomaly: A deep dive is needed into March 2023's sales figures to determine the root cause of the unexpected decline.

January's Strength: Capitalize on the consistently strong sales of January by reinforcing successful strategies and introducing new campaigns.

November Optimization: Investigate and address the consistent sales dip in November, considering special promotions or marketing campaigns to bolster sales.

Leverage Stability: The stability from 2021 to 2022 is an asset. Focus on sustaining this stability by ensuring product quality, efficient operations, and high customer satisfaction.

Anticipate & Prepare: Develop contingency plans to handle future disruptions, especially if March 2023's decline was due to an external factor.

Time-Series Forecasting: Adopt forecasting tools for better sales predictions, facilitating proactive decision-making.

Customer Engagement: Foster a stronger connection with the customers to gain insights into their preferences and potential challenges.

Seasonal Strategies: Introduce promotions or product launches during consistently low-sales months to boost revenue.

Competitive Monitoring: Stay updated with competitors' activities to devise effective counter-strategies.

9.0 Conclusion:

In an era of rapid consumer shifts, leveraging data-driven insights is vital for retailers seeking a competitive edge in the market. This report has unveiled essential facets of consumer behavior, illuminating favored payment modes, prominent gender groups, favored shopping malls, and age-linked shopping tendencies. Equipped with this knowledge, retailers can make well-informed choices, from enhancing payment options to crafting targeted marketing strategies that resonate with specific age segments. Additionally, recognizing consistent sales trends and anomalies across months empowers businesses to anticipate challenges and capitalize on opportunities for enduring success. As the retail landscape continues to evolve, a commitment to ongoing analysis and adaptability remains pivotal for sustaining resilience and delivering exceptional shopping experiences.