



Lecturer: **Dr Xavier Matieni**

Student: **Tsonyo Todorov**

ID: **2021609**

Date: **03/05/2024**

Data Handling and Business Intelligence (LCBB5000)

Assignment 2

Table of Contents

Introduction	3
Context and Methodology	3
Part 1: Analysis and Findings	3
1. Sales Trend Over the Months	3
2. Most Frequently Purchased Products.....	4
3. Purchasing Behavior in Transactions	5
4. Profitable Customer Segments	6
5. Deep Analysis of Strategic Recommendations for Increased Profitability	6
5.1. Enhance Focus on High-Value Customers:.....	7
5.2. Expand Core Product Lines.....	7
5.3. Market Penetration in the UK.....	8
5.4. Incentives for Bulk Purchasers.....	9
Part 2: SPSS Data Analysis and MS Excel Comparison	10
Number of Male and Female Students	10
Students' Qualification for Reduced/Free Lunch	11
Mean and Median Scores for Students Based on Lunch Qualification.....	12
Students who qualify for reduced/free lunch	12
Students who do not qualify for reduced/free lunch	12
Clusters.....	12
SPSS And MS Excel Comparison	13
SPSS (Statistical Package for the Social Sciences)	15
1. Usability	15
2. Data Handling Capabilities.....	15
3. Analytical Features	15
4. Visualization:	15
5. Cost	15
6. Integration	15
7. Learning Resources	16
8. Typical Use Cases	16



MS Excel	16
1. Usability	16
2. Data Handling Capabilities	16
3. Analytical Features	16
4. Visualization	16
5. Cost	17
6. Integration	17
7. Learning Resources	17
8. Typical Use Cases	17
Conclusion	17
References	18

Table of Figures

FIGURE 1 SALES TRENDS	4
FIGURE 2 TOP 20 FREQUENTLY PURCHASED PRODUCTS	5
FIGURE 3 TRANSACTIONS	5
FIGURE 4 TOP 20 MOST PROFITABLE CUSTOMERS OVERALL	6
FIGURE 5 NUMBER OF MALE AND FEMALE STUDENTS	10
FIGURE 6 STUDENTS' QUALIFICATION FOR REDUCED/FREE LUNCH	11
FIGURE 7 K-MEANS CLUSTERING	12



Introduction

The advent of e-commerce has revolutionized the retail landscape, offering businesses a platform to expand their market reach and streamline distribution processes. In the digital age, e-commerce platforms are not just alternatives but essential avenues for retail transactions, catering to a global audience with varied needs (Thakur ,2021). This paper delves into the sales transaction data of a UK-based online retail business specializing in gifts and homewares. By leveraging data from a year's transactions, this analysis aims to uncover sales trends, customer purchasing behaviors, and profitable customer segments (Gielens and Roggeveen ,2023). These insights will help craft strategies to enhance profitability and customer engagement in a competitive market.

Context and Methodology

The dataset provided originates from a London-based e-commerce business established in 2007. It caters to a diverse clientele, including direct consumers and small businesses that purchase products in bulk. The data spans an entire year and includes various metrics crucial for detailed analysis.

Using Microsoft Excel, this study employs data cleaning, processing, and visualization techniques to analyze the dataset comprehensively. Excel's robust functionality supports the handling of large datasets with features like pivot tables for summarizing data, formulae for calculations, and chart tools for trends visualization.

Part 1: Analysis and Findings

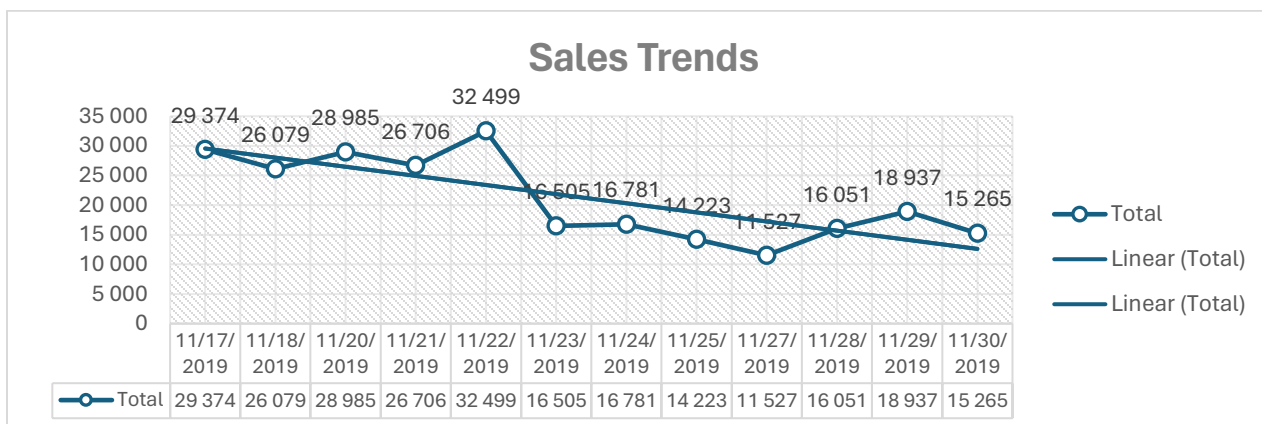
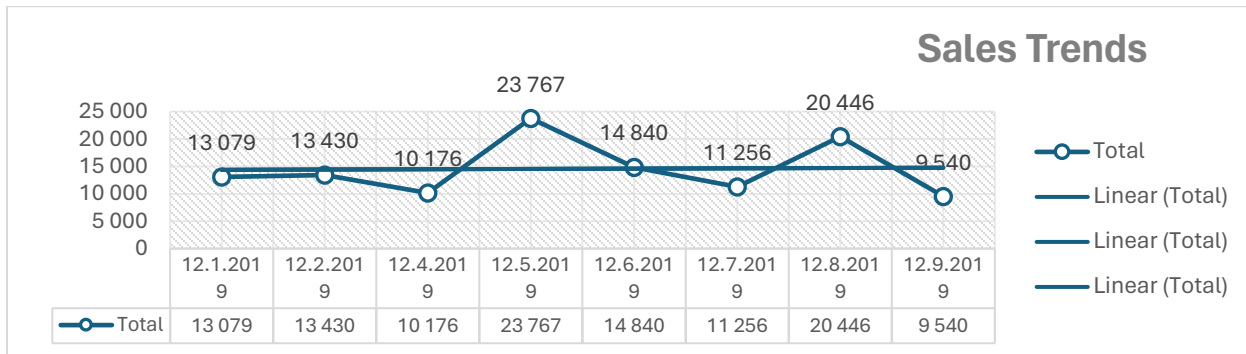
Sales Trend Over the Months

The analysis of monthly sales trends reveals significant fluctuations, with peaks and troughs indicating variability in consumer purchasing patterns. Notably, a sharp increase in sales was observed in May, suggesting a possible impact of seasonal marketing or specific promotional activities. Visualizing these trends in Excel through line graphs allows for easy identification of peak periods, which are critical for planning inventory and marketing strategies (Seetharaman, et.al., 2016)

- ❖ The data shows fluctuating sales across the months of 2019. Notably, there's a peak in May with sales of £23,767.28, suggesting higher activity possibly due to seasonal promotions or specific marketing campaigns.

The analysis of monthly sales trends reveals significant fluctuations, with peaks and troughs indicating variability in consumer purchasing patterns (Seetharaman, et.al., 2016).

Figure 1 Sales Trends



Most Frequently Purchased Products

The examination of product popularity shows that certain items consistently resonate with customers. The 'Rabbit Night Light'-206 units, 'Paper Chain Kit 50's Christmas'-185 units, and 'Popcorn Holder'-156 units, were the Most Popular Products and top sellers. These items are highly popular, which could indicate their appeal to a broad customer base or their effectiveness in driving sales during certain periods (Anderson, Jolly and Fairhurst ,2007). Bar charts generated in Excel elucidate these trends, offering clear insights into consumer preferences that can guide product stocking decisions.

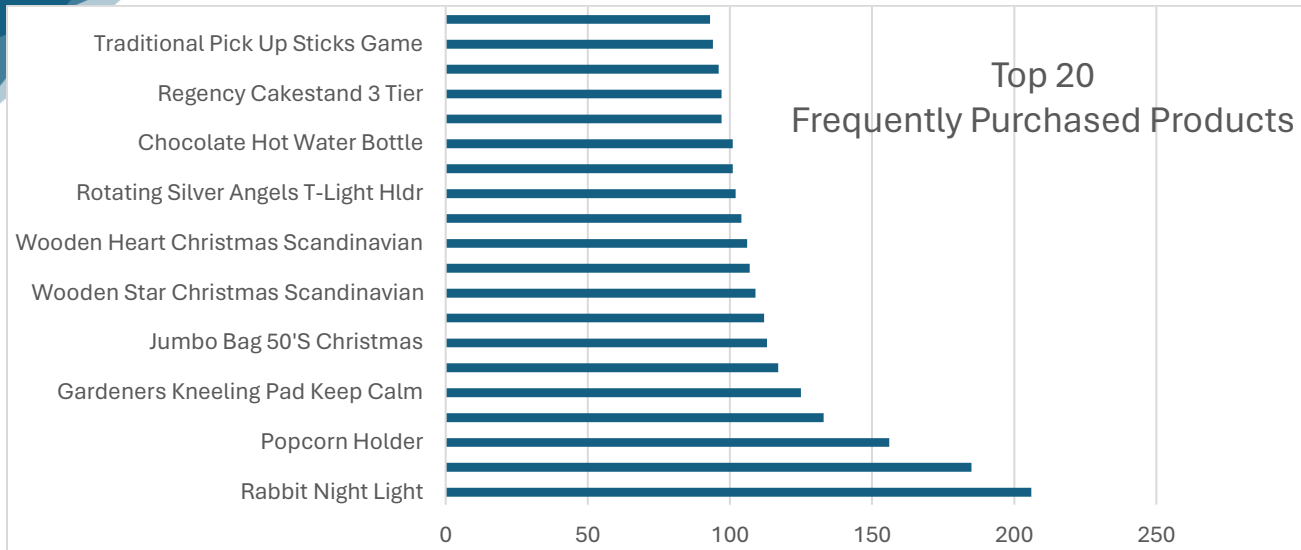
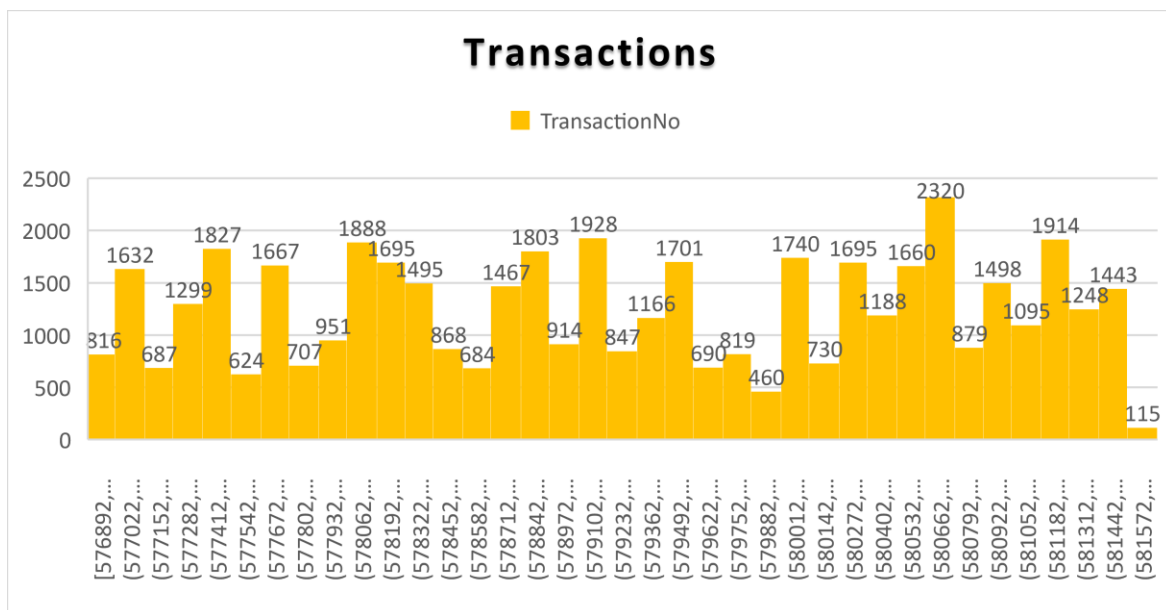


Figure 2 Top 20 Frequently Purchased Products

Purchasing Behavior in Transactions

On average, transactions involve the purchase of approximately 506 products, indicating a substantial proportion of bulk buys. This suggests that the business attracts not only individual buyers but also small businesses purchasing for retail (Anderson, Jolly and Fairhurst ,2007). Understanding this behavior is pivotal for tailoring bulk purchase incentives and loyalty programs, as highlighted in stacked column charts detailing transaction sizes.

Figure 3 Transactions

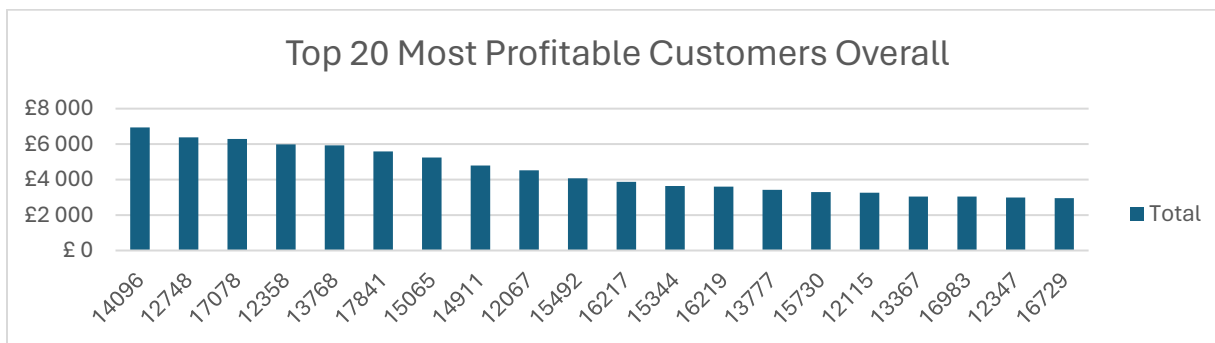


Profitable Customer Segments

The UK emerges as the most lucrative market, generating the highest sales, followed by smaller markets like Austria and France. Analysis of individual customer spending reveals that certain customers contribute disproportionately to revenues, indicative of high-value clients or wholesale relationships (Mark, Niraj and Dawar ,2012). Pie charts in Excel can effectively segment these customer groups, providing a visual representation of where marketing efforts and customer service can be intensified.

- ❖ United Kingdom: Major revenue with £333,210.34 in sales.
- ❖ Austria and France follow but with much lesser sales.
- ❖ Top Individual Customers: Customers with IDs 14096, 12748, and 17078 are among the highest spenders, each with sales surpassing £6,000.

Figure 4 Top 20 Most Profitable Customers Overall



Deep Analysis of Strategic Recommendations for Increased Profitability

In the dynamic and competitive landscape of e-commerce, leveraging data to drive strategic decisions is crucial for maximizing profitability and enhancing market presence (Surya, et.al., 2021). Based on the sales transaction data provided from a UK-based online retail business specializing in gifts and homewares, several targeted strategies have been identified. This detailed analysis explores these strategies, aiming to bolster business growth and customer engagement, while ensuring they are tightly aligned with insights drawn from the data.

Enhance Focus on High-Value Customers:

Analysis

- ❖ The sales data reveals a significant contribution from high-value customers, particularly a few individuals and small businesses that stand out due to their substantial spending. Focusing on these customers is not merely about recognizing their current value but strategically nurturing these relationships to foster loyalty and increase lifetime value.
- ❖ Develop personalized marketing strategies and loyalty programs for the top spending customers to enhance their buying experience and encourage repeat business.
- ❖ Implement targeted marketing strategies and personalized engagement plans for top spenders.

Strategic Implementation

- ❖ Personalized Marketing Campaigns: Develop personalized communication and marketing campaigns tailored to their preferences and purchase history. Utilize data analytics to understand their buying behavior and preferences (Kotler, 2022).
- ❖ Exclusive Offers and Loyalty Programs: Introduce loyalty programs that offer exclusive benefits, such as early access to new products or special discounts. These programs should be designed to reward continued patronage, encouraging frequent purchases.
- ❖ Enhanced Customer Service: High-value customers expect premium service. This could include dedicated account managers, priority customer service, or personalized shopping advice.

Expected Outcomes

Enhancing focus on high-value customers is likely to result in higher customer retention rates, increased frequency of purchases, and possibly attracting similar high-value customers through word-of-mouth and referrals (Ingram, et.al.,2019).

Expand Core Product Lines

Analysis

- ❖ Increase stock levels of bestsellers to meet demand, especially during peak sales periods.
- ❖ Products like the “Rabbit Night Light” and “Popcorn Holder” have shown high sales volumes, indicating strong market demand. Stocking up on these and similar items can prevent stock-outs during peak demand, ensuring that the business capitalizes on every sales opportunity.



Strategic Implementation

- ❖ **Inventory Management:** Employ robust inventory management strategies to ensure best sellers are always in stock, particularly during peak sales periods identified from the sales trends (like May).
- ❖ **Product Development:** Consider expanding these successful product lines to include variations in design, size, or function to cater to a broader audience (Kotler, 2022).
- ❖ **Supplier Relationships:** Strengthen relationships with suppliers to improve procurement efficiency and cost-effectiveness, ensuring readiness to meet increased demand.

Expected Outcomes

Expanding and effectively managing core product lines is expected to directly increase sales, enhance customer satisfaction by reducing instances of stock-outs, and indirectly boost brand reputation (Ingram, et.al.,2019).

Market Penetration in the UK

Analysis

- ❖ The UK market represents the bulk of the business's sales and profitability. Deepening market penetration here offers a substantial opportunity for growth, especially by targeting regional preferences and seasonal trends (Office for National Statistics ,2024).
- ❖ Strengthen marketing campaigns in the UK, potentially exploring regional preferences for more tailored approaches.
- ❖ Since the UK is the most profitable region, intensify marketing efforts there, possibly expanding more tailored promotions or localized products.

Strategic Implementation

- ❖ **Regional Marketing Strategies:** Develop marketing campaigns that cater to specific regional tastes and preferences within the UK, perhaps focusing on local holidays or events (Kotler, 2022).
- ❖ **Use of Localized Content:** Employ localized marketing content, including regional dialects or cultural nuances, to connect more authentically with different customer segments.
- ❖ **Partnerships and Collaborations:** Form partnerships with local influencers or businesses to enhance outreach and credibility within regional markets.

Expected Outcomes

Tailoring strategies to the regional characteristics of the UK market can enhance customer engagement, increase sales in underpenetrated areas, and build a stronger local brand presence (Surya, et.al., 2021).



Incentives for Bulk Purchasers

Analysis

- ❖ Develop attractive bulk buying packages to encourage large-scale purchases, especially from small businesses.
- ❖ The data indicates significant transactions involving bulk purchases, likely by small businesses that resell these products. Providing targeted incentives can further encourage this behavior, increasing order sizes and frequency.
- ❖ Given the high average products per transaction, consider offering bulk buying incentives or discounts to attract more large-scale purchasers and wholesalers.

Strategic Implementation

- ❖ Bulk Buying Discounts: Offer structured discounts for bulk purchases that scale with the quantity purchased. This pricing strategy can encourage larger orders (Parker and Chatterjee ,2012).
- ❖ B2B Portal: Develop a specialized B2B portal that facilitates bulk buying, featuring enhanced order management tools, bulk pricing models, and tailored promotions.
- ❖ Business Customer Support: Provide dedicated support for business customers, including specialized account managers and tailored shipping options.

Expected Outcomes

Encouraging bulk purchases through incentives can increase average order size, improve inventory turnover, and solidify relationships with business customers, leading to stable, predictable sales streams.

Implementing these strategies requires a nuanced understanding of the data, continuous monitoring of their effectiveness, and flexibility to adapt as more data becomes available or as market conditions evolve. By focusing on high-value customers, expanding core product lines, deepening market penetration, and incentivizing bulk purchasers, the business is well-positioned to not only meet its current sales objectives but also to pave the way for sustainable long-term growth (Parker and Chatterjee ,2012).



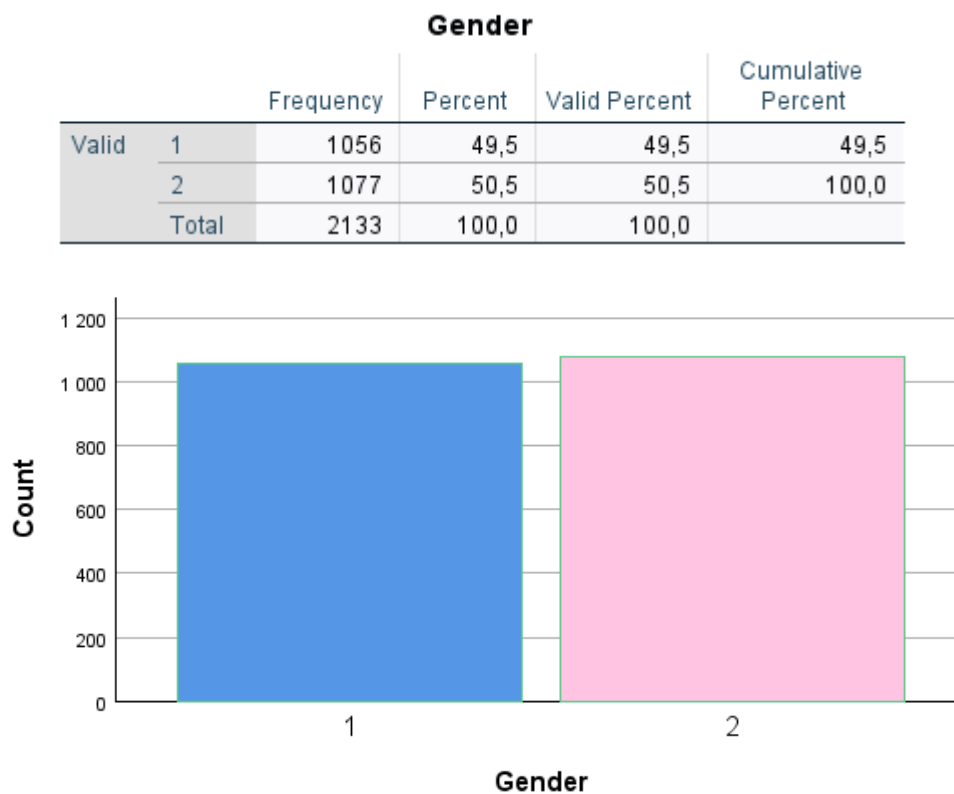
Part 2: SPSS Data Analysis and MS Excel Comparison

Number of Male and Female Students

The analyzed data shows the following:

- ❖ Female Students: 1,077
- ❖ Male Students: 1,056

Figure 5 Number of Male and Female Students



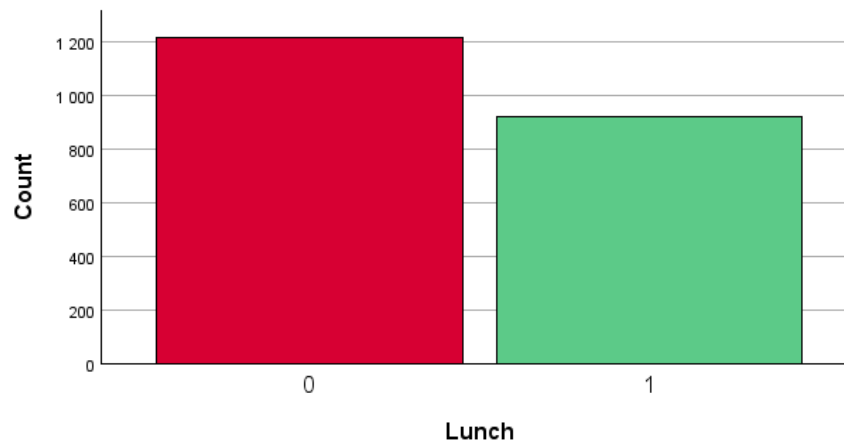
Students' Qualification for Reduced/Free Lunch

The study shows that:

- ❖ Students who do not qualify: 1,215
- ❖ Students who qualify for reduced/free lunch: 918

Figure 6 Students' Qualification for Reduced/Free Lunch

		Lunch			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1215	57,0	57,0	57,0
	1	918	43,0	43,0	100,0
	Total	2133	100,0	100,0	



Mean and Median Scores for Students Based on Lunch Qualification

Students who qualify for reduced/free lunch

- ❖ Mean Pretest Score: 45.25
- ❖ Mean Posttest Score: 57.48
- ❖ Median Pretest Score: 44.0
- ❖ Median Posttest Score: 55.0

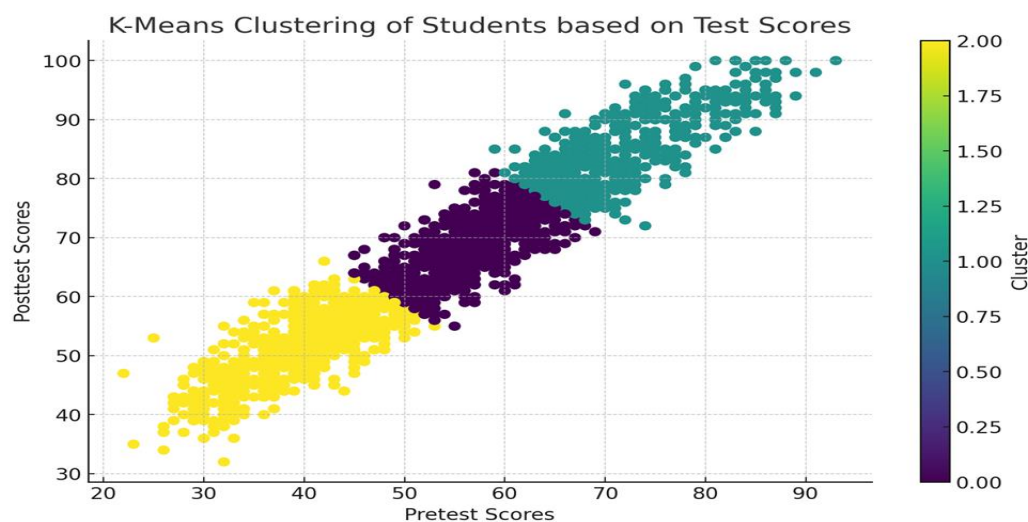
Students who do not qualify for reduced/free lunch

- ❖ Mean Pretest Score: 62.29
- ❖ Mean Posttest Score: 74.38
- ❖ Median Pretest Score: 62.0
- ❖ Median Posttest Score: 74.0

Clusters

Now, let's move on to the fourth question, where I'll provide an example of K-Means Clustering using the dataset. We'll cluster students based on their Pretest and Posttest scores to potentially uncover distinct groups related to their performance improvement (Novoselsky and Kagan ,2021).

Figure 7 K-means clustering



The plot above shows the results of K-Means Clustering on the students' pretest and posttest scores, with the dataset divided into three clusters. Each color represents a different cluster, indicating groups of students with similar score patterns. Here's how you might interpret the clusters:

- ❖ Cluster 1 (Yellow): Students in this cluster generally have lower pretest scores but show a range of improvements in their posttest scores.
- ❖ Cluster 2 (Purple): Students here start with mid-range pretest scores and exhibit moderate to high posttest scores.
- ❖ Cluster 3 (Green): This cluster includes students with high pretest scores, who also tend to maintain high posttest scores.

This clustering can be useful for identifying groups of students who may benefit from tailored educational interventions based on their initial performance and improvement (Novoselsky and Kagan ,2021).

SPSS And MS Excel Comparison

Next, let's address the fifth question about the advantages and disadvantages of using SPSS and Excel for data analysis. I'll organize the discussion in a table format covering different analysis. The criteria include ease of use, data handling capabilities, analytical features, and typical use cases cost: criteria (Olabode and Bakare ,2020). Here's a table discussing the advantages and disadvantages of using SPSS and Excel for data.

Criteria	SPSS	Excel
Ease of Use	Less intuitive for beginners, steep learning curve.	Highly user-friendly and widely familiar interface.
Data Handling	Handles large datasets efficiently, robust in managing complex data types.	Performance can degrade with large datasets (>100,000 rows).
Analytical Features	Extensive statistical tools, advanced analytics like regression, ANOVA, etc.	Basic statistical functions, suitable for simple to moderate analysis.
Visualization	Powerful, but less flexible than some dedicated tools.	Strong in creating customizable charts and pivot tables.
Cost	Commercial software, can be expensive.	Part of Microsoft Office suite, more affordable and widely accessible.
Integration	Integrates well with other statistical packages, but less so with Office.	Excellent integration with other Office applications.
Learning Resources	Abundant professional and academic resources but may require training.	Extensive tutorials and community support available online.
Typical Use Cases	Preferred in research and specialized professional fields.	Used universally in business, education, and for general data management tasks.

This table helps outline the primary differences and could assist in deciding which tool is more suitable depending on the specific needs of your project or organization. Each tool has its strengths, so the choice often depends on the user's familiarity, the specific functionalities required, and budget considerations (Olabode and Bakare ,2020).

To provide a detailed comparison and discussion of the advantages and disadvantages of using SPSS and Excel in data analysis, it's important to consider various aspects including usability, data handling capabilities, analytical features, cost, and typical use cases. Each tool offers unique strengths and weaknesses that make them suitable for different types of users and tasks.

SPSS (Statistical Package for the Social Sciences)

Usability: SPSS is designed for users with a background in statistics, making it less intuitive for those without statistical training. Its interface, though well-organized, involves a steep learning curve, primarily due to its vast array of complex functionalities and options tailored for advanced statistical analysis. This makes it a powerful tool for statisticians, but potentially overwhelming for novices (Meyers, Gamst and Guarino 2013).

Data Handling Capabilities: SPSS excels in handling large datasets and complex data types more efficiently than Excel. It supports a wide range of data formats and is capable of performing data manipulation, transformation, and analysis on large volumes without significant performance degradation, which is a crucial advantage in research and big data environments.

Analytical Features: One of SPSS's core strengths is its comprehensive suite of advanced statistical tools, including regression analysis, ANOVA, factor analysis, cluster analysis, and non-parametric tests, among others. It also offers extensive options for data exploration, hypothesis testing, and multivariate statistics, which are essential in scientific research and complex analytics (Meyers, Gamst and Guarino 2013).

Visualization: While SPSS provides a variety of graphing and charting tools that are powerful, they are often viewed as less flexible compared to dedicated visualization software. However, for the purposes of statistical reporting and analysis, the visualization capabilities of SPSS are generally adequate and professionally oriented (Meyers, Gamst and Guarino 2013).

Cost: As a commercial software, SPSS can be quite expensive, which might be a limiting factor for individuals or organizations with limited budgets. The cost becomes a significant consideration especially when compared to cheaper or free alternatives (IBM SPSS Statistics ,2024).

Integration: SPSS integrates well with other statistical software and databases, enhancing its utility in environments where multiple tools are used (Meyers, Gamst and Guarino 2013). However, it lacks the seamless integration with office productivity tools that Excel offers, potentially complicating workflows that require regular data exchange with such applications.

Learning Resources: There are abundant resources available for learning SPSS, including textbooks, online courses, and tutorials, which are invaluable for new users. Nonetheless, mastering SPSS typically requires formal training or significant self-study, which might be a barrier for some (IBM SPSS Statistics ,2024).

Typical Use Cases: SPSS is predominantly used in academia, research institutions, and specific industries like market research, health research, and data-driven marketing where sophisticated statistical analysis is crucial (IBM SPSS Statistics ,2024).

MS Excel

Usability: Excel is renowned for its user-friendly interface, which is familiar to most computer users. Its straightforward layout and intuitive design make it accessible for users at all levels, from beginners to advanced users. This ease of use is a substantial advantage in environments where quick adaption and flexibility are required (de Levie ,2004).

Data Handling Capabilities: Excel is suitable for datasets of moderate size. Its performance tends to degrade with very large datasets or when performing complex calculations on large data volumes. For typical business and educational applications, however, Excel's capabilities are usually sufficient (de Levie ,2004).

Analytical Features: Excel offers a range of basic to moderate analytical tools, including functions for statistical calculations, pivot tables for data summarization, and what-if analysis features for forecasting and modeling. While not as extensive or advanced as SPSS, these tools cover the needs of most business analysis tasks (de Levie ,2004).

Visualization: Excel's strength lies in its powerful and highly customizable visualization capabilities. It allows users to create a wide variety of charts and graphs, pivot charts, and conditional formatting features that make it easy to visualize trends and patterns in data, enhancing the presentation and interpretation of results (de Levie ,2004).

Cost: Excel is part of the Microsoft Office suite, which is more affordable and widely accessible compared to specialized statistical software. This makes it a popular choice for a broad range of users, from individuals to large enterprises (Todorova ,2019).

Integration: Excel integrates seamlessly with other Microsoft Office applications, such as Word and PowerPoint, which is a significant advantage for users who rely on these tools for report generation and presentations. This integration facilitates a smooth workflow across different tasks and projects (de Levie ,2004).

Learning Resources: The availability of learning resources for Excel is vast, with numerous online tutorials, videos, forums, and courses available at all levels. This accessibility makes it easier for users to get started and continue improving their skills over time (Todorova ,2019).

Typical Use Cases: Excel is used universally across almost all industries for tasks ranging from simple data entry and tracking to more complex financial analysis, business reporting, and decision support (Todorova ,2019).

Conclusion

The choice between SPSS and Excel for data analysis ultimately depends on the specific needs of the user or organization. SPSS is better suited for complex, large-scale statistical analysis where advanced statistical techniques are required. Excel, on the other hand, offers flexibility, ease of use, and excellent integration with office tools, making it ideal for a wide range of different tasks.

References

1. Anderson, J.L., Jolly, L.D. and Fairhurst, A.E. (2007) 'Customer relationship management in retailing: A content analysis of retail trade journals', *Journal of Retailing and Consumer Services*, 14(6), pp.394-399.
2. de Levie, R. (2004) *Advanced Excel for Scientific Data Analysis*. Oxford University Press.
3. G.Shainesh, Philip Kotler, P., Keller, K.L, Chernev and A., N. Sheth, J.N. (2022) *Marketing Management*. 16th ed. Pearson Education.
4. Gielens, K. and Roggeveen, A.L. (2023) 'Editorial: So, what is retailing? The scope of journal of retailing', *Journal of Retailing*, 99 (2), pp.169-172.
5. IBM SPSS Statistics, (2024) Available at: <https://www.ibm.com/products/spss-statistics/> (Accessed: 20rd April 2024).
6. Mark,T. and Niraj, R.(2012) 'Uncovering Customer Profitability Segments for Business Customers', *Journal of Business-to-Business Marketing*, 19(1), pp.1-32.
7. Meyers, L.S., Gamst, G.C. and Guarino L.J. (2013) *Performing Data Analysis Using IBM SPSS*. New Jersey: Wiley.
8. N. Ingram, T.N, LaForge, R.W., Avila,R.A., Schwepker Jr, C.H., and Williams, M.W. (2019) *Sales Management Analysis and Decision Making*. New York: Routledge.
9. Novoselsky, A. and Kagan, E. (2021) 'An introduction to cluster analysis', *Research Gate*.
10. Office for National Statistics (2024) Available at: <https://www.ons.gov.uk/> (Accessed: 23rd April 2024).
11. Olabode, S. and Bakare, A. (2020) 'A PRACTICAL APPLICATION OF MICROSOFT EXCEL AND IBM SPSS STATISTICS', *A COMPARISON OF STATISTICAL SOFTWARE PACKAGES*.
12. Parker, R. and Chatterjee, A. (2012) 'Consumer preference for Bulk Purchasing: Evidence from Exploratory Surveys', *International Journal of GSTF Business Review*, 1(3).
13. Seetharaman, A., Niranjani, I, Tandon, V and Saravanan, A. S. (2016) 'IMPACT OF BIG DATA ON THE RETAIL INDUSTRY', *Corporate Ownership & Control*, 14(1).
14. Surya, B., Menne, F., Sabhan, H., Suriani, S., 4, Abubakar, H., 5 and Idris, M. (2021) 'Economic Growth, Increasing Productivity of SMEs, and Open Innovation', *Journal of Open Innovation: Technology, Market, and Complexity*, 7(20).
15. Thakur, A. (2021) 'Trends and Analysis of E-Commerce Market: A Global Perspective', *International Journal of Applied Marketing and Management*, 6 (1), pp.11-22.

16. Todorova, S. (2019) 'Statistics for Data Analysis Using Microsoft Excel ' , *IZVESTIA JOURNAL OF THE UNION OF SCIENTISTS – VARNA*. 8(2), pp.68-74.