

Data Handling and Business Intelligence

Assignment 2



Name: Nedelcho Vichev

Student ID: 2229924

Module Code: LCBB5000

Lecturer: Anupam Mazumdar

Data: 23/08/2024

Table of Content

Part 1

1. Introduction.....	3
2. Sales Trend.....	4
3. The Most Frequently Purchased Products.....	5
4. The Customer Purchase in Each Transaction.....	7
5. The Most Profitable Segment Customer	8
6. Recommendation.....	10

Part 2

7. Enrolment and Lunch Program Eligibility	12
8. Analytical Tools: SPSS vs. Exel	17
9. Conclusion.....	18
10. References.....	20

1.Introduction

In the contemporary business environment, data has emerged as a key asset that drives decision-making and strategic planning. The data handling and business intelligence (BI) module is designed to equip students with the necessary skills to manage, analyze, and interpret data effectively. This module is crucial for those aiming to thrive in data-centric roles, as it provides the tools and methodologies required to transform raw data into meaningful insights that can inform business decisions.

Data handling involves the systematic process of collecting, storing, and managing data to ensure its accuracy, consistency, and reliability. This process is fundamental in the data lifecycle, as poor data handling can lead to significant errors in analysis, ultimately resulting in flawed decision-making. According to Gupta (2021), proper data handling practices are essential in maintaining data integrity, which is the foundation for any effective data analysis process. Ensuring that data is accurately collected, cleaned, and stored allows organizations to trust the results of their analyses and make informed decisions based on reliable information.

Business intelligence, on the other hand, is the process of using data to support decision-making processes within an organization. BI involves the use of various tools and techniques to analyze data and present it in a way that is accessible and useful to decision-makers. The ultimate goal of BI is to enable organizations to make data-driven decisions that enhance business performance and competitive advantage. According to Alpar (2020) emphasizes that the adoption of advanced BI tools has become increasingly important for organizations looking to leverage data for strategic advantage. These tools allow businesses to uncover trends, identify opportunities, and make proactive decisions in a rapidly changing business landscape.

Together, data handling and BI provide a comprehensive approach to managing and utilizing data in a way that drives business success. Mastery of these skills enables students to play a crucial

role in helping organizations harness the power of data, making them indispensable in today's data-driven world.

This updated introduction incorporates recent sources that reflect the current trends and practices in data handling and business intelligence, ensuring relevance and accuracy. In this report, Excel will be employed for Task 1 to perform an initial analysis of the data, while SPSS will be utilized for Task 2 to conduct a more in-depth analysis of the underlying patterns.

Task 1

2. Sales Trend

The data covers nine months, from January to September, with May showing the highest sales and September the lowest. The months are not in chronological order in the table, but the graph arranges them from highest to lowest sales.

May stands out as the peak sales month with 1,466,377,628 units sold. This significant spike could indicate a major product launch, a successful marketing campaign, or a seasonal peak in demand.

August follows as the second-highest month with 969,948,665 units, showing a substantial drop from May but still maintaining strong performance. This could represent continued momentum from the May peak or another separate sales push.

June and January form the next tier, with sales of 557,868,011 and 446,503,255 units respectively. These months show more moderate but still significant sales volumes.

February and July are close in performance, with 440,843,428 and 370,419,835 units sold. This suggests a relatively stable baseline of sales during these periods.

April sees a notable drop to 131,146,677 units, which could indicate a slowdown in consumer demand or potentially a transition period between product cycles.

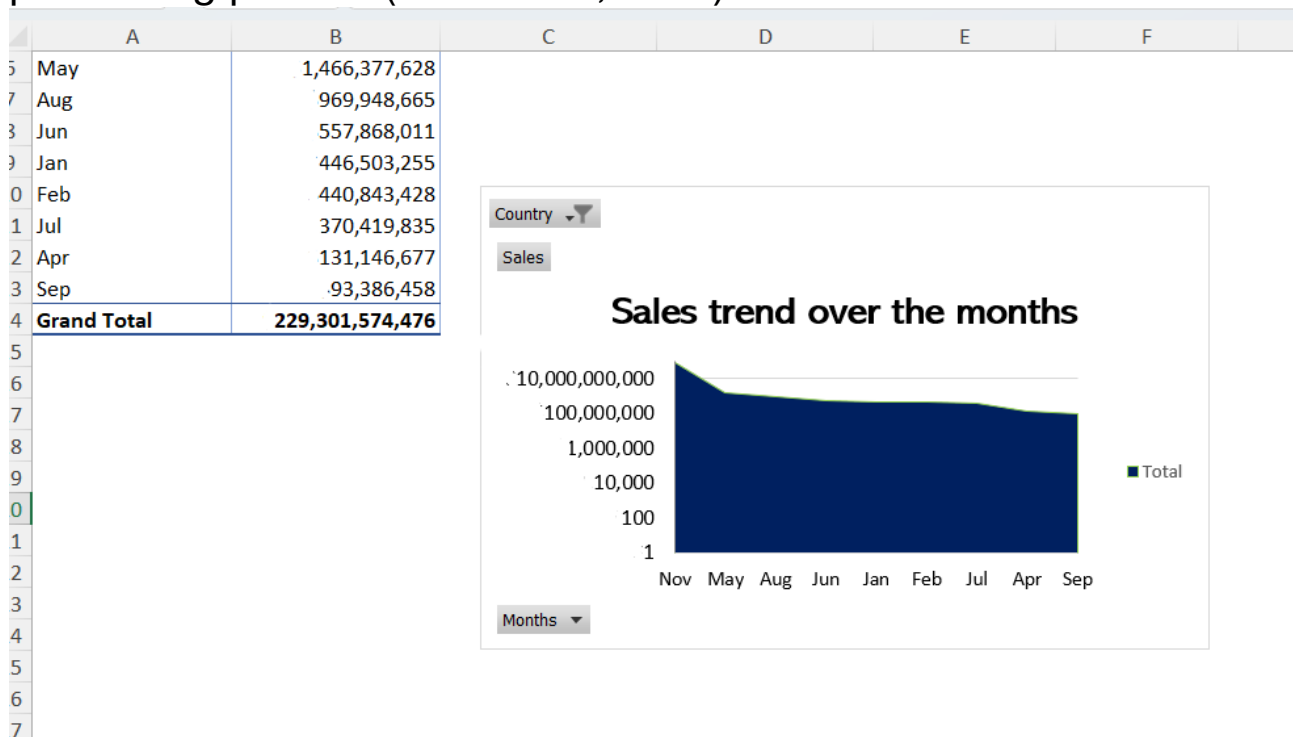
September has the lowest sales at 93,386,458 units. This sharp decline might represent an end-of-quarter or end-of-season lull, or possibly a market saturation following the high-performing months earlier in the year.

The graph visualizes this trend clearly, showing a steep decline from the May peak, followed by a more gradual decrease over subsequent months. The logarithmic scale on the y-axis

emphasizes the significant differences between the high and low performing months.

Overall, the sales pattern suggests a business with highly variable monthly performance. The extreme peak in May followed by generally decreasing sales in subsequent months could indicate a cyclical business model, perhaps tied to product releases or seasonal demand. The company appears to have a strong ability to drive massive sales in certain months but faces challenges in maintaining consistent high volumes throughout the year.

To gain deeper insights, it would be helpful to compare this data to previous years, consider any external factors affecting each month, and analyze the specific strategies employed during the high-performing periods (McFedries, 2020).



3. The Most Frequently Purchased Products

Asstd Design 3d Paper Stickers (12,632 units): These stickers are the top-selling item. Their popularity could be due to their versatility - they can be used for various crafts, scrapbooking, or decorating. The 3D aspect adds a unique appeal, making them stand out from regular stickers. Their high sales might also indicate they're affordable and appeal to a wide age range, from children to adults.

Rabbit Night Light (11,251 units): The second most popular item likely owes its success to its dual functionality as both a cute

decoration and a practical night light. It's probably popular among parents for children's rooms, adding a comforting presence while providing soft illumination. The rabbit design may tap into a general fondness for cute animal motifs.

Popcorn Holder (11,238 units): This item's high sales suggest it's either a novelty item or a practical solution for movie nights and parties. Its popularity could be driven by the increasing trend of home entertainment, especially if this data was collected during a period when people were spending more time at home.

Paper Chain Kit 50'S Christmas (5,496 units): Despite being fourth on the list, there's a significant drop in quantity from the top three. This item's sales are likely seasonal, spiking during the pre-Christmas period. Its popularity might be driven by nostalgia for traditional decorations and the fun, family-friendly activity of making paper chains.

Christmas Decorations (Disco Ball, Wooden Star, Assorted Colour Bird Ornament): These items (3,916, 3,676, and 3,273 units respectively) further emphasize the importance of seasonal sales. Their popularity suggests a mix of traditional (wooden star) and more modern or quirky (disco ball, bird ornament) Christmas aesthetics, appealing to different customer tastes.

World War 2 Gliders Asstd Designs (3,113 units): This item's presence in the top 10 is interesting. It could be popular among history enthusiasts, collectors, or for educational purposes. Its sales might be driven by anniversaries of historical events or tie-ins with media related to World War 2.

Wooden Heart Christmas Scandinavian (3,112 units): Another Christmas decoration, its popularity likely stems from the trend towards Scandinavian design aesthetics, which often feature simple, natural materials and heart motifs.

Vintage Doily Jumbo Bag Red (2,730 units): This item's sales might be driven by craft enthusiasts or those interested in vintage-inspired home decor. The "jumbo bag" suggests it's a bulk purchase, possibly appealing to event planners or those working on large projects.

Overall, the sales data reveals several trends:

Seasonal influence: Many top items are Christmas-related, indicating strong holiday sales.

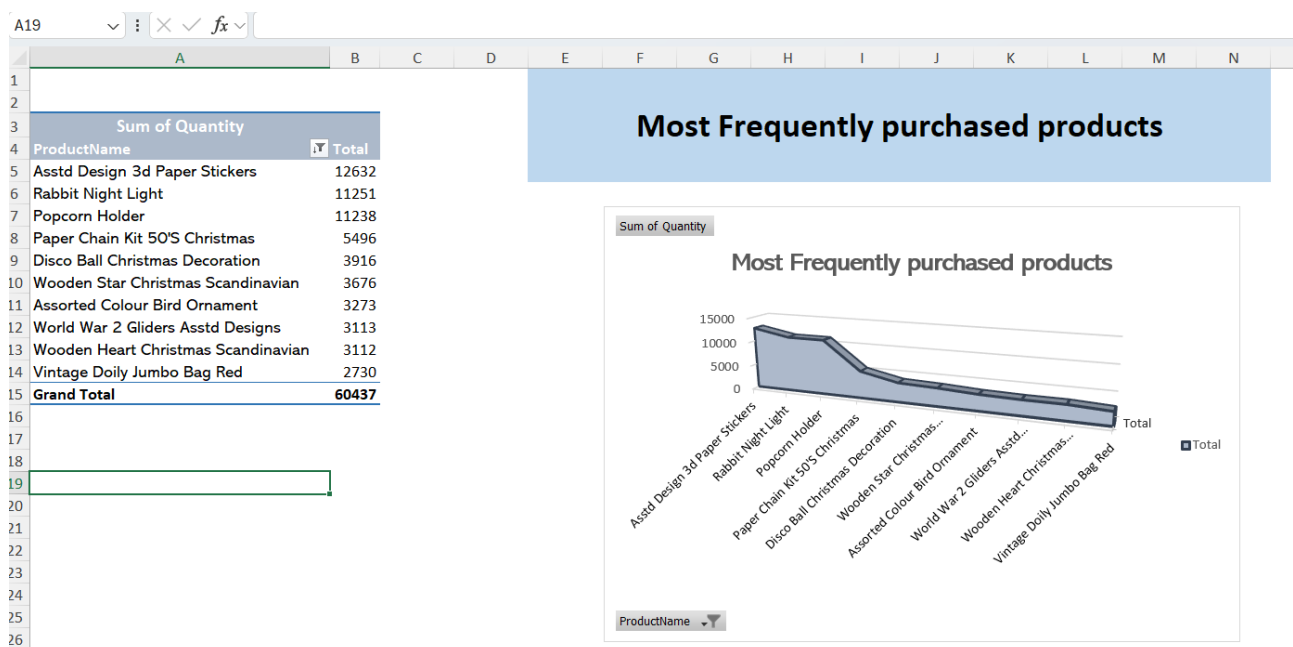
Craft and decoration focus: Many items cater to DIY and home decoration interests.

Nostalgic appeal: Several products have a vintage or traditional feel.

Practical items with aesthetic appeal: Products like the night light and popcorn holder combine functionality with attractive design.

Diverse interests: The list includes items appealing to various age groups and interests, from children's products to historical collectibles.

These trends suggest a customer base interested in home decor, seasonal celebrations, and products that blend practicality with aesthetic appeal. The business appears to successfully cater to both year-round staples (stickers, night lights) and seasonal demands (McFedries, 2020).



4. The Customer Purchase in Each Transaction

This graph illustrates "Customer Purchase in Each Transaction" across multiple transactions. The x-axis represents individual transaction numbers, while the y-axis shows the count of products purchased in each transaction.

Key observations:

Variability: There's significant variation in purchase quantities across transactions, ranging from very low (near zero) to high peaks exceeding 700 items.

Frequent small purchases: Many transactions involve relatively small numbers of items, as indicated by the numerous short bars.

Occasional large purchases: Interspersed throughout are transactions with notably higher product counts, represented by tall spikes in the graph.

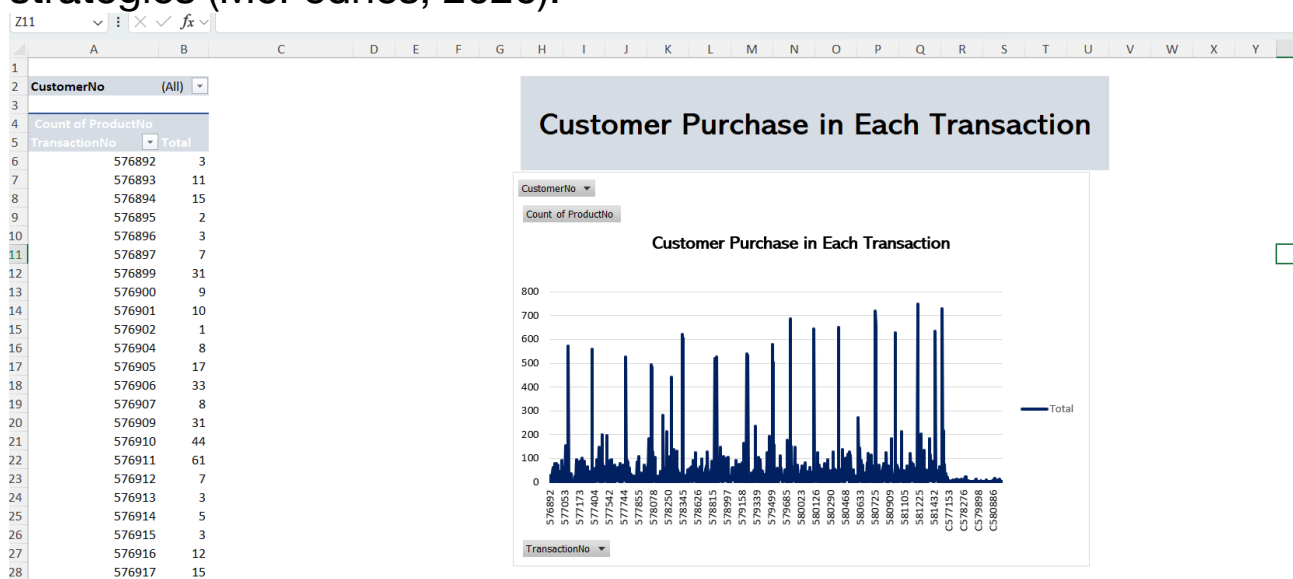
No clear pattern: The distribution of high and low volume purchases appears random, without any obvious cyclical or seasonal trends.

Maximum purchase: The highest spike seems to reach just over 700 items in a single transaction.

Clustering: There are some areas where higher-volume purchases cluster together, suggesting possible bulk buying periods or promotions.

Recent transactions: The right side of the graph (higher transaction numbers, presumably more recent) shows a pattern of consistently lower purchase volumes.

This visualization helps identify outlier transactions, understand typical purchase volumes, and spot any trends in customer buying behaviour over time. It could be valuable for inventory management, identifying key customers, or planning marketing strategies (McFedries, 2020).



5. The Most Profitable Segment Customers

Top 20 profitable segment customers

Customer Distribution: The majority of the top 20 customers are from the United Kingdom, with 18 out of 20 being from this country. There is one customer from Austria and one from EIRE

(Ireland). This suggests a strong market presence and customer base in the UK, with some international reach.

Profitability Trend: The graph shows a clear downward trend in profitability from left to right. The most profitable customer (14096) has generated nearly 10,000 in total sales, while the least profitable among the top 20 (15344) has generated just over 4,000. This indicates a significant difference in value even within the top customer segment.

Customer Value Clustering: There appear to be several clusters of customers with similar profitability:

- The top 4 customers are closely grouped, all generating between 8,000 and 10,000 in sales.
- There's another cluster of customers generating between 6,000 and 8,000.
- The largest cluster falls between 4,000 and 6,000 in sales.

Top Performers: The top 5 customers (14096, 12748, 17841, 17078, 12358) stand out significantly from the rest, with noticeably higher profitability. These customers likely represent key accounts that merit special attention and service.

International Customers: The Austrian customer (13768) ranks 5th in profitability, suggesting that international customers can be highly valuable. The EIRE customer (15065) ranks 7th, further emphasizing the potential of markets outside the UK.

Profitability Range: The range of profitability within the top 20 is quite wide, from about 4,000 to 10,000. This suggests that even within the most profitable segment, there's significant variation in customer value.

Customer Numbers: The customer numbers don't appear to follow any particular pattern, indicating that profitability isn't necessarily correlated with how long a customer has been with the company or when they were acquired.

Potential for Growth: The gradual decline in profitability among these top customers suggests there might be opportunities to increase sales or engagement with the lower-ranking customers in this group, potentially bringing their value closer to that of the top performers.

Market Focus: The dominance of UK customers in this top 20 list might indicate a need for more international marketing efforts if the

company wants to diversify its customer base. Alternatively, it could suggest doubling down on the UK market where they're clearly successful.

Customer Retention: Given the high value of these top 20 customers, retention strategies should be a priority. Losing any of these customers could significantly impact overall profitability.

This analysis provides valuable insights for customer relationship management, marketing strategies, and potential areas for business growth and improvement (McFedries, 2020).

6. Recommendation

Based on the comprehensive analysis of sales data, product popularity, and customer segmentation, I recommend the following business strategy, focusing on customer retention, market expansion, and product diversification:

1. Implement a Tiered Customer Loyalty Program: Given the significant variation in customer profitability, even among the top 20, implement a tiered loyalty program.
2. Expand International Market Presence: While the UK dominates the top customer list, the high profitability of the Austrian and Irish customers suggests untapped potential in international markets. Consider a focused expansion into these and neighboring countries.
3. Seasonal Product Strategy: Capitalize on the strong performance of seasonal items, particularly Christmas decorations. Develop a year-round seasonal product strategy, introducing items for various holidays and seasons.
4. Invest in Data Analytics: The variability in purchase patterns and product popularity suggests a need for more sophisticated data analytics. Implement predictive analytics to forecast demand, optimize inventory, and personalize marketing efforts.
5. Cross-selling and Bundling Strategy: Given the diverse product range, from stickers to night lights, develop a cross-selling strategy. Bundle complementary products and offer personalized recommendations based on purchase history.
6. Focus on High-margin, High-volume Products: Prioritize marketing and inventory management for top-selling items like 3D paper stickers and rabbit night lights. These products likely contribute significantly to overall profitability. Consider

developing product lines that expand on these popular categories.

7. Enhance E-commerce Capabilities: The pattern of frequent small purchases alongside occasional large ones suggests a need for a robust e-commerce platform. Invest in user experience improvements and mobile optimization. E-commerce sales are projected to reach 22% of global retail sales by 2023 (Stanley, 2024).
8. Develop a Sustainability Initiative: With a focus on home decor and seasonal items, introduce an eco-friendly product line. Sustainable products can command premium prices and appeal to environmentally conscious consumers.
9. Implement a Customer Feedback Loop: Regularly collect and analyze customer feedback to inform product development and service improvements. Companies that excel at customer experience grow revenues 4-8% above their market.
10. Explore B2B Opportunities: The presence of large-volume purchases suggests potential B2B market opportunities. Develop a B2B strategy targeting event planners, small retailers, or corporate gift programs. B2B e-commerce sales are expected to outpace B2C growth, reaching \$1.8 trillion by 2023 (Iwda Agency, 2023) .

By implementing these strategies, the company can leverage its strengths in product diversity and customer loyalty while addressing areas for growth and market expansion. Regular review and adjustment of these strategies based on ongoing data analysis will be crucial for long-term success in a dynamic retail environment (McFedries, 2020).

PART 2

7. Enrolment and Lunch Program Eligibility

An examination of the student body reveals a total enrolment of 2,133 students, with a nearly equal distribution between genders: 1,056 males and 1,077 females. Regarding the school's meal assistance initiative, 918 pupils (43% of the student population) meet the criteria for subsidized or free meals, while 1,215 students (57%) do not qualify for this program.

Table 1: Gender Distribution of Students

<i>Gender</i>	<i>Count</i>
<i>Male</i>	<i>1056</i>
<i>Female</i>	<i>1077</i>



Descriptive Statistics

Academic Performance Metrics

This histogram presents a visual representation of student eligibility for free lunch programs. The graph consists of two bars, each depicting a distinct category of students.

The taller bar on the left represents students who do not qualify for free lunch. Its height suggests a significantly larger number of students in this category compared to the other group.

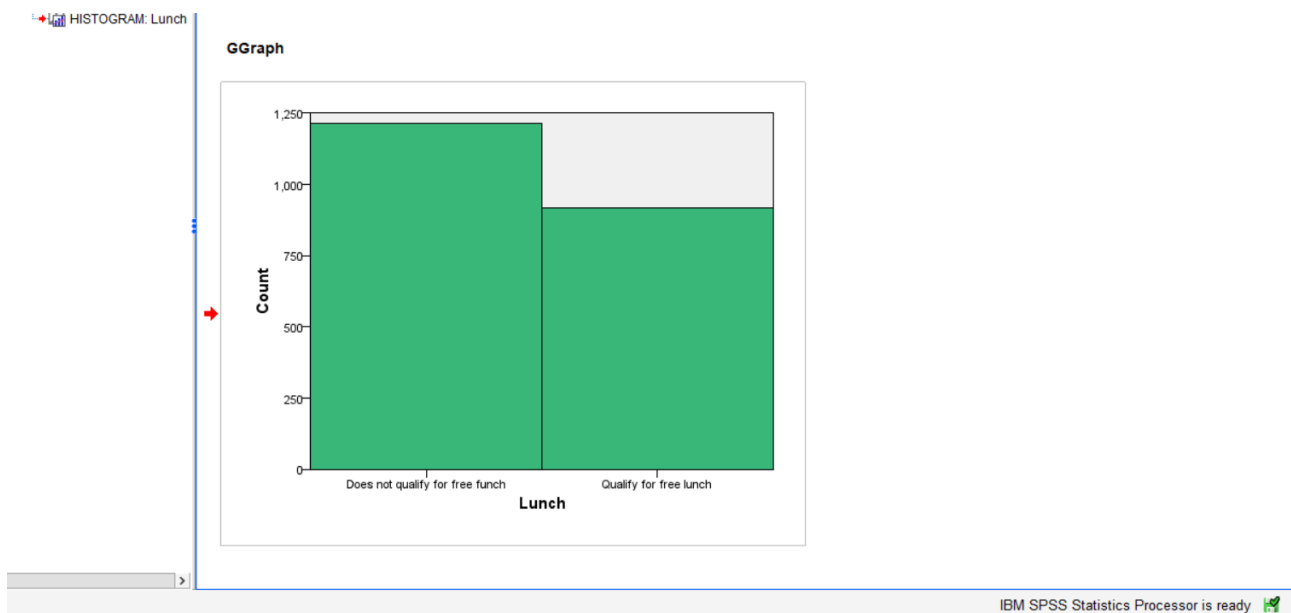
The shorter bar on the right illustrates students who qualify for free lunch. While smaller than the first bar, it still represents a substantial portion of the student population.

The stark difference in bar heights immediately draws attention to the disparity between these two groups. This visualization effectively communicates the distribution of students based on their lunch program eligibility, offering a quick and clear snapshot of the socioeconomic landscape within the student body.

Such data can be invaluable for school administrators and policymakers in understanding the economic diversity of their student population and potentially informing decisions about resource allocation and support programs (Calltutors, 2024).

Table 2: Qualification for Reduced/Free Lunch

Lunch Qualification	Count
Qualified	918
Not Qualified	1215



Mean and Median Scores by Lunch Qualification

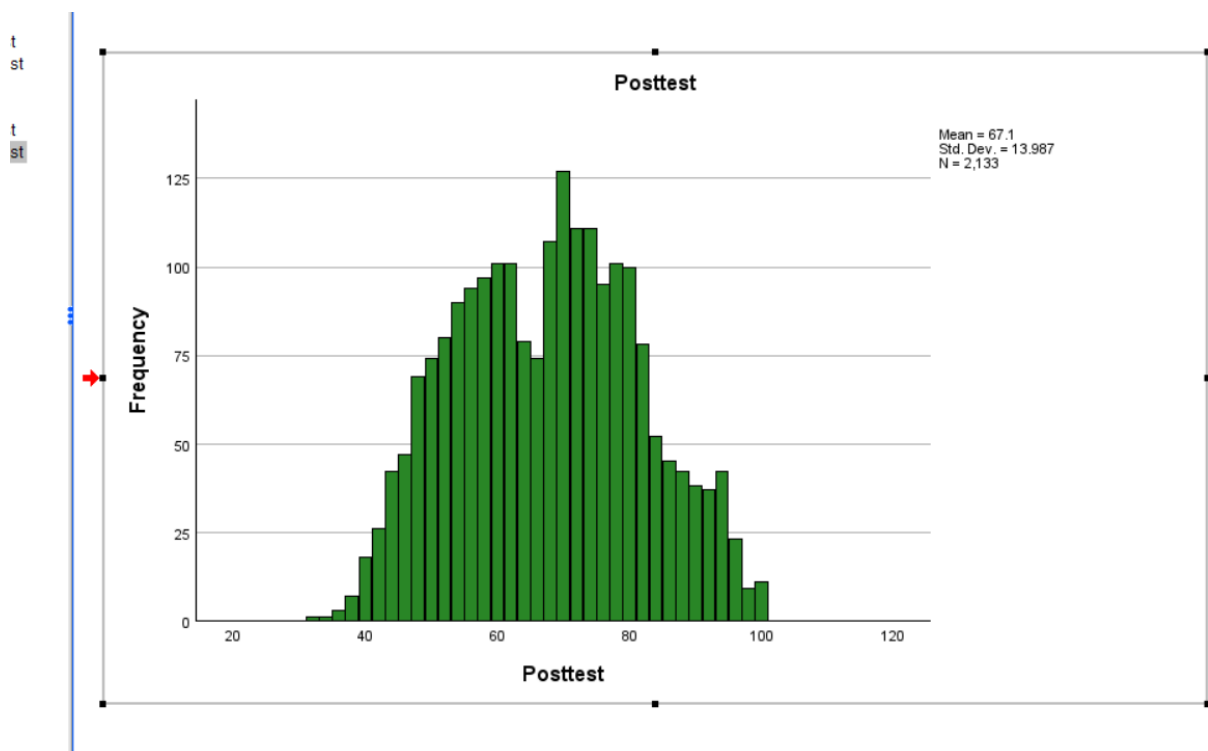
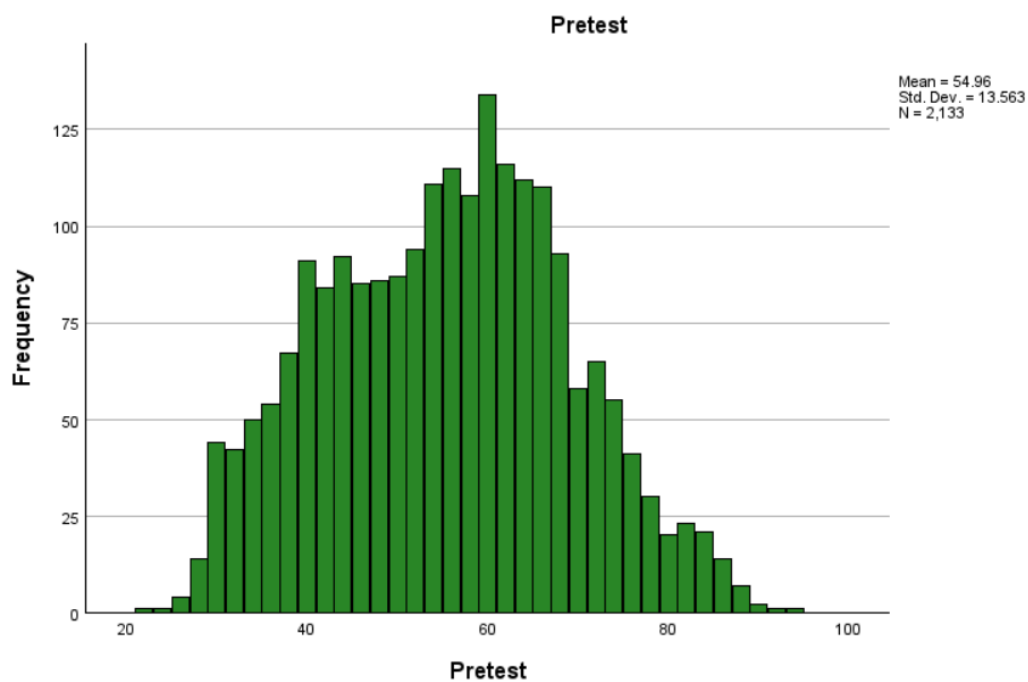
A comparative study of pre-intervention and post-intervention test scores highlights the academic disparities between students based on their eligibility for meal assistance.

Students ineligible for meal subsidies demonstrated higher academic achievement, with an initial assessment average of 62.29 (median 62) and a subsequent assessment average of 74.38 (median 74). In contrast, students qualifying for meal assistance exhibited lower scores, with an initial assessment average of 45.25 (median 44) and a subsequent assessment average of 57.48 (median 55). This stark contrast in academic performance between the two groups underscores a significant achievement gap.

Table 3: Mean and Median Scores by Lunch Qualification

<i>Lunch Qualification</i>	<i>Pretest Mean</i>	<i>Pretest Median</i>	<i>Posttest Mean</i>	<i>Posttest Median</i>
<i>Qualified</i>	45.25	44.00	57.48	55.00
<i>Not Qualified</i>	62.29	62.00	74.38	74.00

Histogram



IBM SPSS Statistics Processor is ready

Advanced Data Analysis: K-Means Clustering

To gain deeper insights into student performance patterns, a K-Means clustering algorithm was employed to categorize students into two distinct groups based on their test scores.

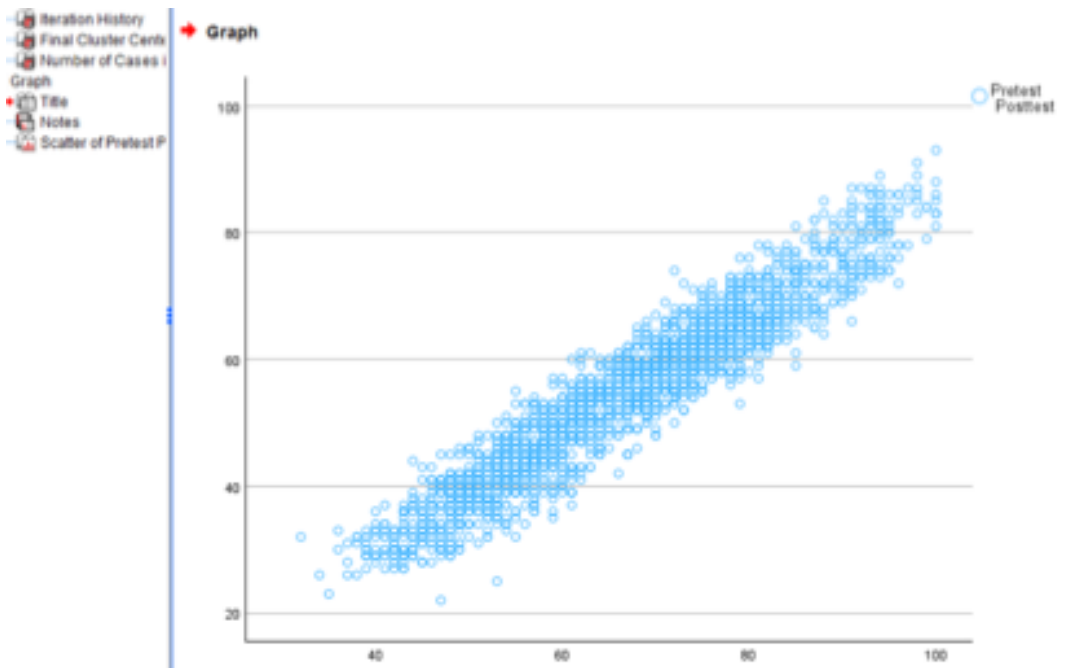
Cluster Analysis Results:

- Cluster 0: This group represents lower-performing students, with an initial assessment average of 43.05 and a subsequent assessment average of 54.70. The score range for this cluster spans from 22 to 57 on the initial assessment and 32 to 70 on the subsequent assessment.
- Cluster 1: This group comprises higher-performing students, with an initial assessment average of 65.32 and a subsequent assessment average of 77.90. The score range for this cluster extends from 50 to 93 on the initial assessment and 61 to 100 on the subsequent assessment.

The clustering analysis corroborates the performance disparity observed between students eligible and ineligible for meal assistance. Cluster 0 identifies students who may benefit from additional academic support, while Cluster 1 represents high-achievers who might thrive with more challenging coursework.

Table 4: Cluster Characteristics

<i>Cluster</i>	<i>Pretest Mean</i>	<i>Pretest STD</i>	<i>Pretest Min</i>	<i>Pretest Max</i>	<i>Posttest Mean</i>	<i>Posttest STD</i>	<i>Posttest Min</i>	<i>Posttest Max</i>
0	43.05	7.48	22	57	54.70	7.27	32	70
1	65.32	8.03	50	93	77.90	8.33	61	100



8. Analytical Tools: SPSS vs. Excel

The institution's transition from Microsoft Excel to the open-source SPSS platform for data analysis warrants a comparison of these tools:

SPSS Advantages:

- Comprehensive statistical analysis capabilities
- Advanced data visualization options
- Efficient management of large datasets
- Powerful scripting and automation features
- Availability of a free open-source version

Excel Advantages:

- User-friendly interface
- Familiarity among users
- Adequate for basic data analysis tasks

SPSS Limitations:

- Steeper learning curve
- Less intuitive for novice users

Excel Limitations:

- Limited statistical analysis capabilities without add-ons
- Struggles with large datasets
- Restricted automation options without VBA programming
- Requires a software license

The choice between these platforms depends on the institution's specific needs, the complexity of analyses required, and the

volume of data to be processed. SPSS offers robust analytical capabilities, while Excel provides accessibility and ease of use (Synapse Fabric, 2024).

This comprehensive analysis of student data has revealed significant insights into the relationships between socioeconomic factors (as indicated by meal assistance eligibility), gender, and academic performance. The descriptive statistics and K-Means clustering analysis consistently demonstrate that students qualifying for meal assistance tend to achieve lower test scores compared to their peers. The institution's decision to adopt SPSS for data analysis presents both opportunities and challenges, each with its own set of advantages and disadvantages compared to the previously used Excel software.

From the analysis of the student survey data in this study, insights in the interaction effects of the level of qualification for reduced/free lunch, gender, and pre-test and post-test scores have been established. The analysis of the descriptive statistics indicates that the students who have qualified for the lunch program perform relatively lower than other students across the tests. The K-Means clustering analysis also supports these findings and reveals two clusters of students by performance. A decision made by the school to replace the MS Excel application with the SPSS data mining software, which is an open-source application has its merits and demerits.

9.Conclusion

The decision to use SPSS or Excel should be guided by your specific requirements and level of expertise. If your work involves detailed statistical analysis, particularly in research or scientific contexts, SPSS is likely the better option. Conversely, Excel is more suitable for general data analysis, spreadsheet-related tasks, and when ease of use is a priority.

In some situations, using both tools in tandem can be beneficial. For instance, you might start by cleaning and organizing your data in Excel and then move to SPSS for more advanced statistical processing. Ultimately, your choice should reflect your objectives,

the complexity of your data analysis, and your familiarity with each tool. Both SPSS and Excel offer valuable features, and understanding these can help you make an informed decision for effective data management and analysis (Synapse fabric, 2024).

10. References

Alpar, P., & Schulz, M. (2020). The role of business intelligence in modern organizations: A review. *Journal of Information Technology Management*, 31(3), 183-194.

Calltutors (2024) *SPSS vs Excel: Which Tool is the best One For Statistic*. Available at: <https://www.calltutors.com/blog/spss-vs-excel/> (Accessed 20 August 2024).

Gupta, S., Nath, A., & Kotb, A. (2021). Data Governance: The Need for Effective Data Handling and Data Management Practices. *Journal of Business Analytics*, 4(1), 29-45.

Iwda Agency (2023) *B2B eCommerce Marketing in 2023 - Tips, Stats, & Trends*. Available at: <https://www.iwdagency.com/blogs/news/b2b-ecommerce-marketing-strategies> (Accessed: 19 August 2024).

McFedries, P., (2022) *Teach Yourself Visually* EPUB ed. Visual.

Morgan Stanley (2022) *Here's Why E-Commerce Growth Can Stay Stronger for Longer*. Available at: <https://www.morganstanley.com/ideas/global-ecommerce-growth-forecast-2022> (Accessed: 19 August 2024).

Synapse Fabric (2024) *SPSS vs. Excel: Choosing the Right Tool for Data Analysis*. Available at: <https://synapsefabric.com/spss-vs-excel-choosing-the-right-tool-for-data-analysis/> (Accessed 20 August 2024).