



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022

BA Business Management

Data Handling & Business Intelligence (LCBB5000)

Dr Xavier Matieni

CONTENTS

PART ONE.....	2
Data utilisation for business success in a competitive environment	2
Introduction	2
Monthly sales trends.....	3
Product trends	4
Product purchase per transaction.....	6
Customer segmentation by profit.....	6
Recommended strategies for increasing profitability	8
PART 2 COMPARISON OF ADVANTAGES AND DISADVANTAGES OF EXCEL AND SPSS	10
Student gender distribution	10
Number of students (not)qualifying for free/ reduced lunch	11
Mean and median of students (not)qualifying for reduced/ free lunch	13
Clustering (K-Means Clustering).....	14
Advantages and disadvantages of using Excel and SPSS	16
References.....	18
Figure 1 November Sales Trend.....	3
Figure 2 December Sales Trend.....	4
Figure 3 Twenty most frequently purchased items.....	5
Figure 4 Regular customers.....	6
Figure 5 Geographical profitability segmentation	8

PART ONE

DATA UTILISATION FOR BUSINESS SUCCESS IN A COMPETITIVE ENVIRONMENT

INTRODUCTION

In today's competitive business climate, data utilisation is critical for making choices, increasing operational efficiencies, improving customer experiences, and gaining a competitive advantage. Businesses can use data effectively in various ways, including data-driven strategies, predictive analytics, personalisation, operational efficiency, risk management, innovation, competitive advantage, real-time insights, benchmarking, revenue maximisation, and cloud computing services (Maheshwari, 2015).

Informed decision-making entails analysing historical data and market trends to improve product development and marketing strategies (Vercellis, 2009). Businesses can use predictive analytics to predict future trends and behaviours, allowing them to anticipate market demand better and manage inventory. Customising marketing messages and product suggestions based on client behaviour, preferences, and previous purchases is one way to improve the customer experience. Process optimisation, risk assessment, compliance monitoring, and innovation can all contribute to increased operational efficiency.

Data may help identify potential risks in investment and operational initiatives, as well as ensure regulatory compliance. Market trend research, research and development optimisation, and real-time data can all help to promote innovation. Real-time analytics enable organisations to make swift decisions that outperform competition, whereas benchmarking helps them grasp industry standards and aim to exceed them. Dynamic pricing techniques, cross-selling, and upselling can all help you maximise your revenue (Idemudia, 2020). To efficiently exploit data, firms frequently use tools and technology like data management systems, business intelligence and analytics tools, customer relationship management systems, advanced analytics, and cloud computing services.

Good data utilisation has become a must for companies working in competitive markets. Companies that harness the power of data not only streamline their operations and reduce costs but also enhance customer experiences and innovate more effectively, thus achieving sustainable growth and competitive advantage (Anubha, Sharma, 2023). As businesses continue to evolve in their data journey, the focus should also be on ethical data use and privacy considerations to maintain trust and comply with regulations.

In this report, we will analyse the sales metrics of a London-based e-commerce business selling homeware and gifts worldwide through its website using the analytic tools of MS Excel.

Microsoft Excel is a versatile spreadsheet application for data preprocessing, analysis, and visualisation. It is easy to use and provides simple data transformations such as pivot tables and formula-based calculations. Excel includes built-in statistical, logical, and financial capabilities, as well as strong pivot tables and an Analysis ToolPak for more advanced statistical work. It also offers a variety of customisable visualisation tools, such as charts and histograms (Venzin, 2019). However, it is limited in data quantity and lacks automatic documentation, making replication difficult. Excel is not as well-suited to complex statistical analysis as specialised tools like SPSS or Python, and it may not support advanced visualisations like Tableau (Pavolini, 2023). To summarise, Excel is useful for managing small to medium-sized datasets because of its accessibility and ease of use.

MONTHLY SALES TRENDS

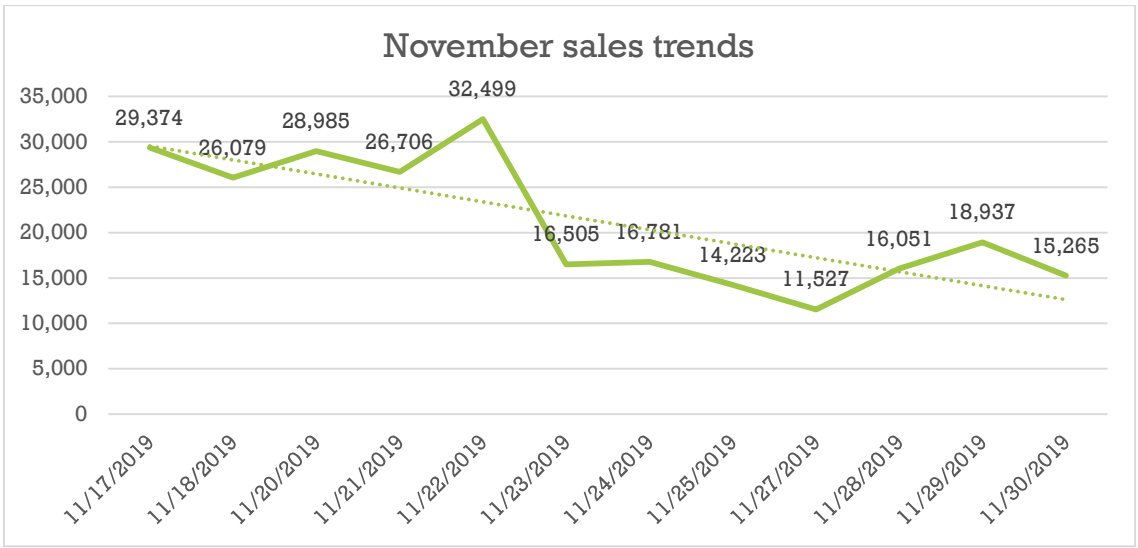


Figure 1 November Sales Trend

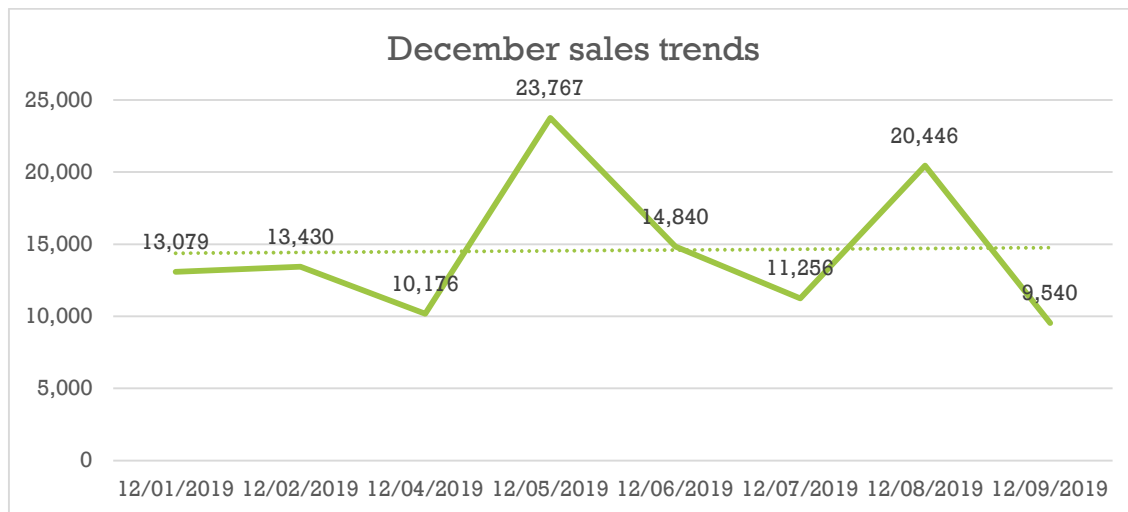


Figure 2 December Sales Trend

Following data preparation, cleaning and organising analysis of the data supplied in the "Sales Transactions" file listing sales records for the period from 17th to 30th of November 2019 and 1st to 9th of December the following could be said:

- **Trend direction** - during the November period a decreasing trend can be noted. Following a steady start in the beginning, a sharp decline can be seen throughout the rest of the period. Following the drop in sales during the second half of the November period, the lower trend persists during the first three days of December, followed by a sharp increase in sales revenue;
- **Volatility** – due to the short time span of provided data, volatility can be analysed on a daily, rather than monthly basis.
Minor volatility can be noted throughout November increasing during December where wider fluctuations in day-to-day revenue can be seen. This volatile trend is likely caused by economic changes for instance the outbreak of COVID in December 2019. Despite this, taking into consideration the time of the year as well as the activity of the business (homeware and gift sales), introducing seasonal marketing and sales campaigns could lead to a positive financial outcome (Jain, N/A).
- **Seasonality** - Sales trends might reflect seasonal patterns, showing peaks during certain times of the year and troughs during others. For example, retail businesses often see higher sales during holiday seasons. However, due to limited time set of the data, it is not possible to analyse the seasonal trends in sales margins.
- **Comparative Performance** - the current sales trend might also be compared to previous years or benchmarks to evaluate whether the trend is positive, negative, or neutral in a broader context (Indeed, 2023). Again, not sufficient data is available for such analysis.

PRODUCT TRENDS



Analysing the dataset by creating a Pivot table depicting the number of purchases per product over the whole period it can be said that a seasonal trend is observed.

- The top 20 most purchased items consist of mainly Christmas and season-related products;
- Total number of transactions for the period – 46160;
- Top selling product "Rabbit night light" appeared in 206 transactions (with an overseas bulk order of over 80 000 units in a single transaction).

The below chart shows the number of transactions per product without calculating the number of units sold.

Figure 3 Twenty most frequently purchased items

PRODUCT PURCHASE PER TRANSACTION

The metric "product purchase per transaction" indicates the average number of products bought in each sales transaction, providing insights into customer purchasing behaviours and business performance. It helps assess the effectiveness of sales strategies, reveals cross-selling opportunities, assists in inventory management, and can enhance revenue by understanding and adapting to customer buying patterns (Mohd Suki, 2020). Additionally, it offers clues about customer satisfaction and helps in market segmentation based on purchasing behaviour, making it a critical indicator for optimizing sales strategies and improving the overall shopping experience.

Quantity	Count of Transactions
1-50	45224
51-100	582
101-150	159
151-200	77
201-250	30
251-300	34
301-350	5
351-400	5
401-450	3
451-500	8
551-600	5
601-650	8
651-700	1
751-800	5
851-900	1
951-1000	3
1101-1150	1
1151-1200	2
1401-1450	2
1451-1500	2
2001-2050	1
12501-12550	1
80951-81000	1
Grand Total	46160

Table 1 Purchases per transaction

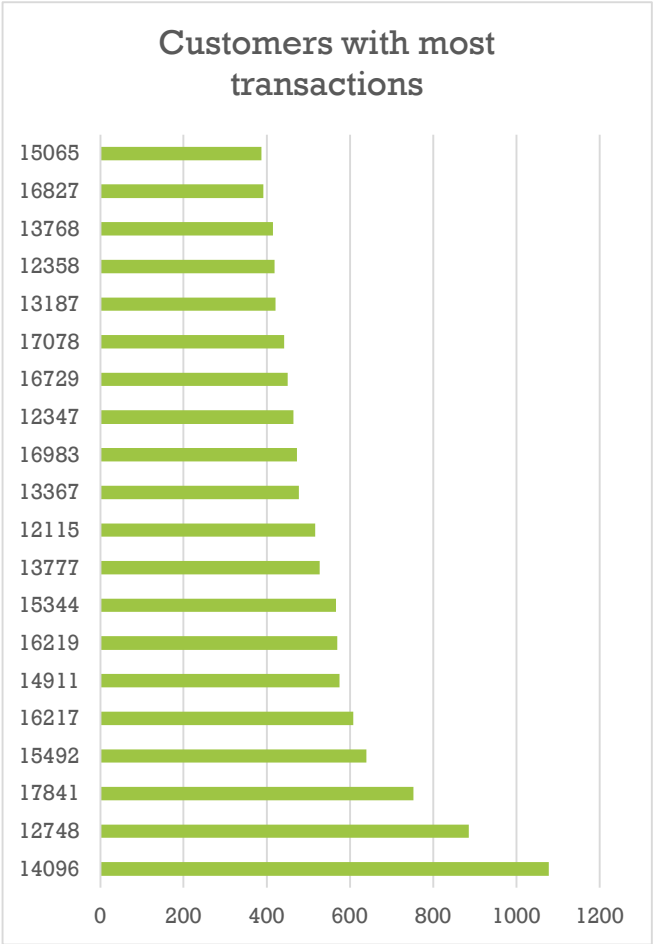
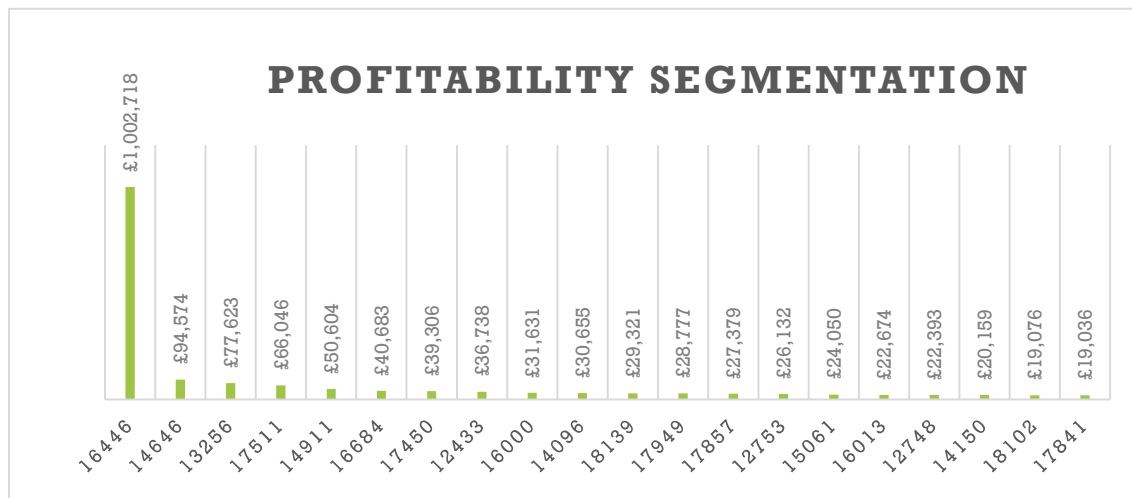


Figure 4 Regular customers

CUSTOMER SEGMENTATION BY PROFIT



Customer segmentation by profit involves categorising customers based on the profitability they bring to your business (Kotler et al., 2020). This approach allows companies to focus their resources more effectively and tailor their marketing strategies to maximise returns.

From the available dataset, profitability could be assessed by two factors:

- Profitability level – high, medium and low
 - high volume buyers who purchase large volumes of stock regularly;
 - medium profitability customers are those who's profit generation is moderate but consistent;
 - low profitability customers with low volume, irregular purchases;

- Geographical area – as the company trades internationally, profits and purchasing behaviour should be assessed for different countries. Profitability rates, potential and risks should be determined for the creation of strategies for sales enhancement. Such strategies can include seasonal sales offers, individually targeted marketing depending on items of interest, effective inventory management – use of data analysis to predict trends and ensure desirable stock is available when needed (McDonald, Dunbar, 2010);
- Figure 5 below illustrates revenue per country, where United Kingdom is identified as most profitable generating approximately 90% of the company's turnover.

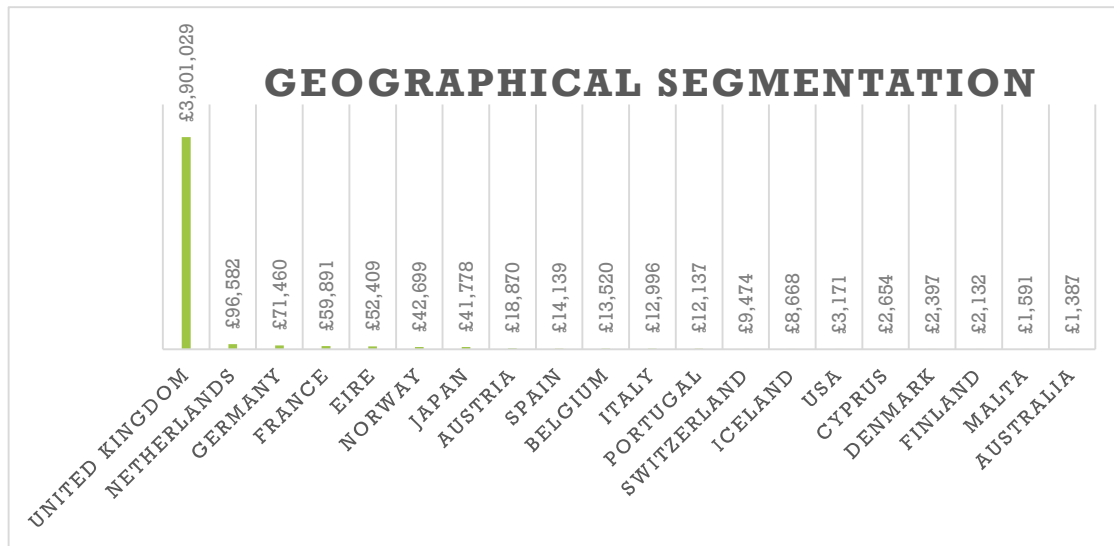


Figure 5 Geographical profitability segmentation

RECOMMENDED STRATEGIES FOR INCREASING PROFITABILITY

Implementing a comprehensive strategy that covers many parts of operations, marketing, and customer involvement is critical boosting profitability (Kotler et al., 2020). Here are some strategic tips that can help an online retailer increase profitability:

1. Improve User Experience

- Optimise website for speed, mobile compatibility, and ease of navigation. A flawless shopping experience can drastically lower cart abandonment rates.
- Checkout Simplification: Reduce the processes required to execute a purchase. Provide guest check-out alternatives and save customer data for future usage.

2. Utilise Data Analytics

- Track Consumer Behaviour: Use analytics tools to monitor consumer behaviour on the website. Determine what generates purchases and optimise accordingly.
- Personalisation: Implement personalised marketing techniques based on user behaviour, preferences, and previous purchases to enhance conversion.

3. Expanding Product Range and Cross-Sell:

- Considering extending inventory to include complimentary products or accessories. This encourages larger purchases (Payne, 2005).
- Cross-marketing and Upselling: Using product recommendations to display comparable products or higher-end alternatives, increasing average order value.

4. Optimising Pricing Strategy

- Competitive Analysis: analysing rivals' pricing regularly to ensure your prices are both competitive and profitable.
- Dynamic Pricing (Rajagopal, 2013): Use dynamic pricing strategies to alter prices in response to demand, inventory levels, and market conditions. (Example: order No 581483 where 80995 units of "Papercraft Little Birdie" were ordered with original price of £12.38 and a 50% discount was given)

5. Enhance Customer Loyalty through Loyalty Programmes: Set up a loyalty programme that rewards repeat consumers with discounts, special deals, or early access to new products. Excellent customer service is essential for recurring business. Ensure that the customer service is efficient and helpful.

6. Utilising Social Media and Content Marketing to communicate with customers, run promotions, and display new items.

- Creating informative content to attract customers to the company's website. This might include how-to manuals, and films about items sold.

8. Implementing Subscription Models - Considering delivering products on a subscription basis to ensure repeat sales and a consistent cash stream (McAfee, 2009).

9. Streamline logistics and operations.

- Employing an effective inventory management system to cut holding costs and avoid stockouts or overstocks.
- Offering a variety of shipping alternatives, including expedited shipping for clients willing to pay extra for faster delivery.

Implementing these strategies requires a balanced approach and ongoing evaluation to determine what is the best course of action for this specific online retail business. Continuous testing, learning from customer feedback, and adapting to market changes are key to driving long-term profitability.

STUDENT GENDER DISTRIBUTION

Gender Data Overview

Statistics

Gender

N	Valid	2133
	Missing	0

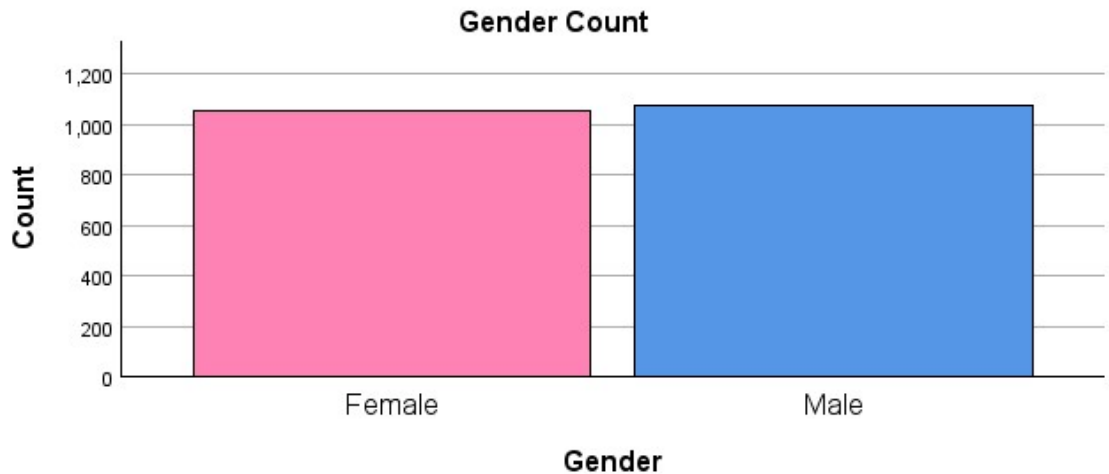
Total Responses:

- **Valid Responses:** 2133
- **Missing Responses:** 0

Data Completeness:

- The dataset is complete with no missing entries for the gender variable, indicating high data quality and completeness for this attribute.

		Gender			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1056	49.5	49.5	49.5
	2	1077	50.5	50.5	100.0
	Total	2133	100.0	100.0	



- Based on provided dataset there are 2133 participants in the survey. 2133 of 2133 responses are valid and can be used for further analysis.
- Analysis shows that there are 1056 or 49.5% female and 1077 or 50.5% male students.

NUMBER OF STUDENTS (NOT) QUALIFYING FOR FREE/ REDUCED LUNCH

Lunch Distribution Report

Dataset Overview:

- **Total Responses:** 2133
- **Data Completeness:** No missing data on the "Lunch" variable, ensuring full utilization of the dataset for analysis.

Lunch Statistics:

- **Categories:**
 - **Category 0 ('No Lunch'):** 1215 participants (57.0% of total)
 - **Category 1 ('Lunch Provided'):** 918 participants (43.0% of total)

Percentages:

- **Valid Percent:** Each category accurately represents its share of the total dataset, with Category 0 at 57.0% and Category 1 at 43.0%.

Detailed Analysis:

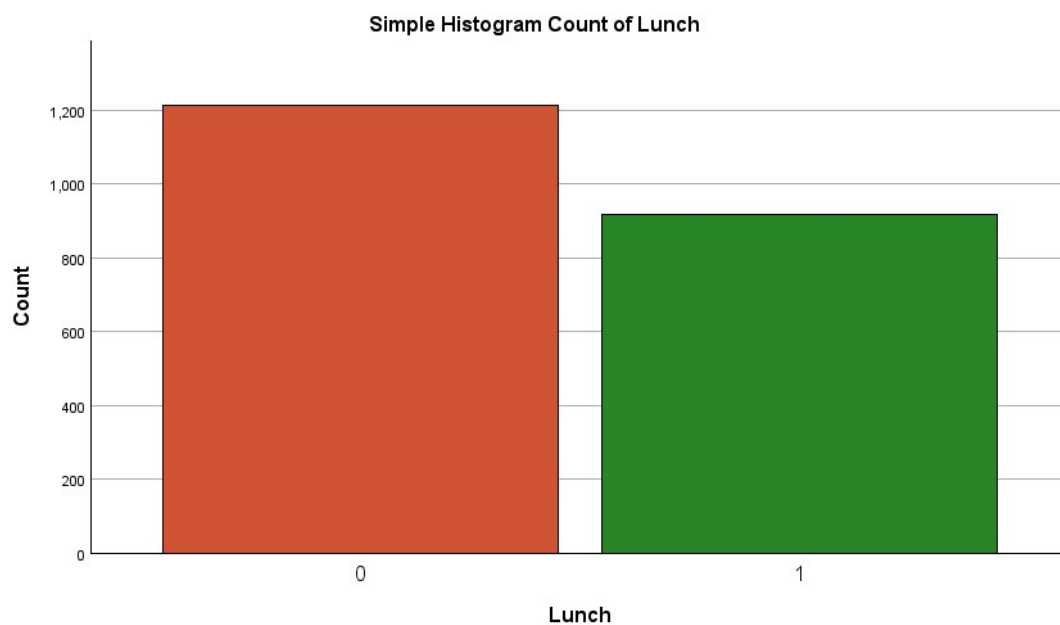
- **Representation Balance:** The data shows a majority of responses falling under Category 0, which represents those not provided lunch (57.0%). The remainder, 43.0%, represent those who were provided lunch.
- **Implications of Distribution:** This balance indicates a significant division in the dataset, which could reflect policy implementations, event settings, or organisational provisions depending on the context in which the data was gathered.

Recommendations:

1. **Contextual Clarity:** Further clarification about what 'Lunch' signifies in this dataset is crucial for meaningful analysis.
2. **Cross-Variable Analysis:** Examine relationships between lunch provision and other variables such as attendance, performance, or satisfaction to uncover potential impacts or benefits.
3. **Policy Evaluation:** If this data relates to a policy, the results could inform discussions about its continuation or modification.

Conclusion:

The "Lunch" variable shows a clear division between participants with and without lunch provision, with over half not receiving lunch. Understanding and interpreting this data depends significantly on the context and the implications of these provisions, which should be investigated further to draw actionable insights and conclusions.



		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 OF STUDENTS (NOT)QUALIFYING FOR REDUCED/ FREE LUNCH

Statistics

Lunch		
N	Valid	2133
	Missing	0
Mean		.43
Median		.00
Variance		.245

The provided statistics describe a dataset related to the "Lunch" variable, consisting of 2133 valid responses with no missing data. Here's a detailed analysis of these statistics:

Statistical Analysis

1. **Mean (Average):** The mean value of 0.43 suggests that, on average, 43% of the responses fall into the category associated with "1" (Lunch Provided).
2. **Median:** The median of 0.00 indicates that more than half of the dataset is categorized as "0" (No Lunch Provided). This suggests a skew in the distribution toward "0".
3. **Variance:** A variance of 0.245 indicates a moderate level of spread in the data around the mean. In the context of a binary variable (0 or 1), this suggests a significant mix of "0" and "1" responses, but with "0" being more prevalent.

Interpretation

- **Distribution:** The mean being less than 0.5 and the median being 0 reflects that a larger proportion of the dataset consists of the "No Lunch Provided" category. This difference between

the mean and median indicates a right-skewed distribution, where the majority of the data points are clustered at the lower end (0).

- **Implication of Unevenness:** The unevenness towards "No Lunch Provided" could imply several scenarios depending on the context (such as a default policy against providing lunch in a educational setting).
- **Data Balance:**
- Although there is a clear dominance of the "No Lunch" category, the presence of a substantial proportion of "Lunch Provided" responses (43%) cannot be ignored, suggesting that the policy or condition for providing lunch is significant but not universally applied.

CLUSTERING (K-MEANS CLUSTERING)

Initial Cluster Centers

	Cluster	
	1	2
Pretest	23	93
Posttest	35	100

Iteration History^a

Iteration	Change in Cluster Centers	
	1	2
1	30.191	33.606
2	.885	.791
3	.611	.532
4	.212	.181
5	.227	.200
6	.150	.128
7	.000	.000

- a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 7. The minimum distance between initial centers is 95.525.

Final Cluster Centers		
	Cluster	
	1	2
Pretest	43	65
Posttest	55	78

Number of Cases in each Cluster		
Cluster	1	
	2	1140.000
Valid		2133.000
Missing		.000

Analyzing the movement from “Initial Cluster Centers” to “Final Cluster Centers” can provide significant insights into the dynamics and adjustments made by the clustering algorithm (K-Means), reflecting how well the algorithm has adapted to the dataset's distribution to optimize the clustering solution

Analysis of Changes

- Cluster 1 (Lower Scores to Mid-Range Scores):
 - Pretest - Increased from 23 to 43. This is a significant upward adjustment, suggesting that the initial center was perhaps too low compared to the actual data points it needed to represent.
 - Posttest - Also rose significantly from 35 to 55. This aligns with the pretest adjustment and indicates a consistent upward shift in the dataset's representation for this cluster.
- Cluster 2 (High Scores to Mid-High Scores):
 - Pretest - Decreased markedly from 93 to 65. This substantial decrease suggests that the initial estimation was too high and needed to be adjusted to more accurately reflect the central tendency of the high-performing data points.
 - Posttest - Decreased from 100 to 78. This change, like that seen in the pretest scores, points to an initial overestimation of the final outcomes for this cluster.

The movement from initial to final cluster centers highlights the iterative nature of k-means clustering and its effectiveness in adapting to the actual data distribution. The adjustments made provide a more accurate segmentation of the dataset, crucial for precise and reliable analysis in various applications. This shift toward a central range also hints at less extreme differentiation among the data points, which may impact how strategies are tailored for different groups identified through this clustering process.

ADVANTAGES AND DISADVANTAGES OF USING EXCEL AND SPSS

SPSS and Excel are two popular data analysis programmes, however they perform somewhat different functions and perform in different situations. The following table compares the benefits and negatives of using SPSS versus Excel for data analysis (StatAnalytica, 2023).

Criteria	SPSS	Excel
User Interface	User-friendly, menu-driven with a graphical interface. Good for users familiar with statistical analysis.	Highly intuitive and widely familiar spreadsheet interface, great for general data manipulation and visualization.
Statistical Analysis	Offers comprehensive and advanced statistical tests and analyses, suitable for complex research needs.	Basic statistical functions are available, but more complex analyses require additional add-ons or tools.
Data Management	Excellent at handling large datasets and complex data transformations.	Good for smaller datasets and basic data manipulation tasks. Can become cumbersome with very large datasets.
Cost	Expensive, with costs associated with licensing fees. Not ideal for individuals or small organizations with limited budgets.	Relatively inexpensive, comes as part of Microsoft Office suite or standalone. More accessible for personal use.
Flexibility and Customization	Less flexible for advanced custom statistical scripting compared to programming-based software.	High flexibility in data manipulation and presentation but limited in advanced statistical customization.
Output and Reporting	Generates outputs that are ready for presentations and reports, easy to navigate and format.	Flexible reporting and visualization tools, but requires more manual setup for complex reporting.
Integration and Expansion	Supports integration with Python and R for enhanced functionality.	Extensive functionalities with add-ons like Power Query and Power Pivot, and can integrate with other Microsoft tools.

Support and Resources	Extensive documentation and support available, large community for troubleshooting.	Ubiquitous usage leads to vast resources, tutorials, and community support.
Performance with Large Data	Handles large datasets well but can be slower than more specialized statistical software.	Struggles with very large datasets or complex, data-intensive operations.
Learning Curve	Easy to use for basic functions, but advanced features require knowledge of syntax and statistical methods.	Easy to start with basic features, but advanced features like VBA scripting have a steep learning curve.

Practical Examples

1. **Academic Research:** For a researcher conducting complex statistical analysis on survey data, SPSS provides powerful tools to run descriptive analyses, regressions, and more, which would be bulky and less straightforward in Excel.
2. **Business Reporting:** A business analyst needs to create weekly sales reports and dashboards from store data. Excel would be preferable here, especially with its PivotTables, charting tools, and the ability to update spreadsheets and graphs easily as new sales data comes in.
3. **Data Cleaning:** SPSS has robust data management tools that can handle large datasets with complex requirements more efficiently than Excel, which might require significant manual intervention for cleaning large or messy data sets.
4. **Budget Management:** For managing a small business's budget, Excel is more suitable due to its spreadsheet format, ease of use, and sufficient capabilities for calculating expenditures, forecasting, and financial reporting.

Conclusion

Choosing between SPSS and Excel depends heavily on specific data analysis needs, budget constraints, and personal or organisational preferences. SPSS is more suited for specialised statistical analyses, while Excel excels in general data management and simpler statistical tasks. Both tools have their place in data analysis and can be extremely effective when used in appropriate contexts.

REFERENCES

- Anubha and Sharma, H. (eds) (2023) *Data-driven approaches for effective managerial decision making*. Hershey, PA: IGI Global, Business Science Reference, an imprint of IGI Global.
- Idemudia, E.C. (2020) *Handbook of research on IT applications for strategic competitive advantage and decision making*. Edited by E.C. (Efosa C. Idemudia. Hershey, PA: IGI Global.
- Indeed Editorial Team (2023) How To Write a Sales Report (With Template and Example), Business and sales operation analyst, available at: <https://www.indeed.com/career-advice/career-development/example-of-a-sales-report> (accessed: 30 April 2024)
- Jain, R., (N/A) 16 Sales reports examples you can use for daily, weekly or monthly reports, Business Services, available at: <https://www.salestrendz.com/sales-reports-examples/> (accessed: 30 April 2024)
- Kotler, P. *et al.* (2020) *Principles of marketing*. Eighth European edition. Harlow, England: Pearson.
- Maheshwari, A.K. (2015) *Business intelligence and data mining*. New York :, Business Expert Press.
- McAfee, R.P. (2009) *Competitive Solutions : The Strategist's Toolkit*. Princeton, NJ: Princeton University Press,. Available at: <https://doi.org/10.1515/9781400828531>.
- McDonald, Malcolm. and Dunbar, I. (2010) *Market segmentation how to do it, how to profit from it*. 4th ed. Oxford: Goodfellow Publishers.
- Mohd Suki, N. (2020) *Handbook of research on technology applications for effective customer engagement*. Edited by N. Mohd Suki. Hershey, Pennsylvania: IGI Global.
- Pavolini, J., (2023) Why Specialized Data Analysis Software is Better than Spreadsheets, available at: <https://www.linkedin.com/pulse/why-specialized-data-analysis-software-better-than-giovanni-pavolini/>, (Accessed:30 May 2024)
- Payne, A. (2005) *Handbook of CRM: Achieving Excellence through Customer Management*. 1st ed. New York: Routledge.
- Rajagopal (2013) *Marketing decision making and the management of pricing successful business tools*. Hershey, Pa: IGI Global.
- Stat Anlytica (2023) SPSS vs Excel: Which One is The Best Tool For Statistics, available at: <https://statanalytica.com/blog/spss-vs-excel/> (accessed: 1 May 2024)
- Venzin, M. (2019) "Make Use of Microsoft Excel as an Analytic Tool," *The Major Gifts Report*, 21(6), pp. 1–1. Available at: <https://doi.org/10.1002/mgr.31225>.
- Vercellis, Carlo. (2009) *Business intelligence [electronic resource] : data mining and optimization for decision making*. Chichester, U.K. , Wiley.