

Data Handling and Business Intelligence

- STUDENT NAME:VERONICA-DANIELA FLORIA
- STUDENT ID:2115140
- MODULE CODE:LCBB5000
- LECTURER NAME:XAVIER MATIENI

Table of Contents

Introduction.....	3
Task 1.....	3
Sales Trend Analysis.....	3
Most Frequently Purchased Products.....	4
Products per Transaction.....	5
Most Profitable Segment Customers.....	5
Recommendations for Increasing Profitability	6
Task 2 IBM SPSS	7
Descriptive Statistics	7
Gender Frequencies.....	8
Lunch Qualification.....	9
Cluster analysis	11
Comparison of SPSS and Excel	15
Conclusion	16
References.....	17
Appendix. Task 1 Excel Workbook.....	19

Introduction

Across the contemporary information-driven business environment, effective analysis of data plays a pivotal role in decision-making processes as well as organizational success. Two extensively used tools for information analysis are SPSS (Statistical Package for the Social Sciences) and Excel. While each application has diverse features and capabilities, both applications have their own advantages and limitations and must be applied appropriately in order to get the best out of them. The following report seeks to explore the use of SPSS and Excel in analysing data, providing insights into their respective strengths and limitations.

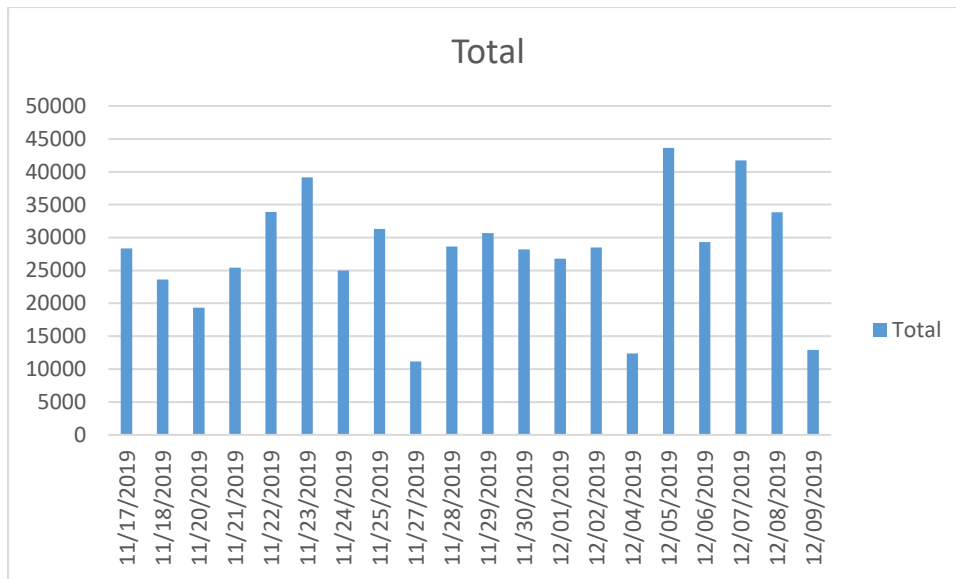
SPSS is renowned for its advanced statistical analysis capabilities, making it a preferred choice for researchers and analysts engaging in complex statistical analyses. With an extensive range of statistical tests and procedures, SPSS allows users to analyse large datasets, carry out regression analysis, conduct hypothesis testing, and generate graphical representations of data. Its specialised capabilities make it especially well-suited for research, survey analysis, and social science research. On the other hand, Excel, as a part of the Microsoft Office suite, is a spreadsheet utilised by experts across numerous industries for statistics analysis, reporting, and visualisation. While Excel might also lack the superior statistical functions of SPSS, it gives a user-friendly interface and intuitive data manipulation tool. Excel's pivot tables, charts, and functions make it ideal for basic statistics analysis, easy calculations, and creating visually appealing charts and graphs. The following report leverages both SPSS and Excel in analysing sales and student data sets, providing valuable insights on the data that can be used in informed decision making process.

Task 1

Sales Trend Analysis

The sales trend over the months showcases fluctuations, increase and decrease from November 17th to December 30th 2019, peaking on November 23rd and 5th of December. This pattern implies heightened income activity around late November, likely attributed to factors such as Black Friday and Cyber Monday promotions, followed by waning sales post these events. As such, it is clear that the trend highlights a surge in income succeeded through a gradual downturn, emphasizing the importance of seasonal advertising and marketing techniques and promotions to keep sales momentum all through the period. Adjusting advertising strategies to align with top purchasing periods can capitalize on client enthusiasm, mitigating the put up-promotion sales dip.

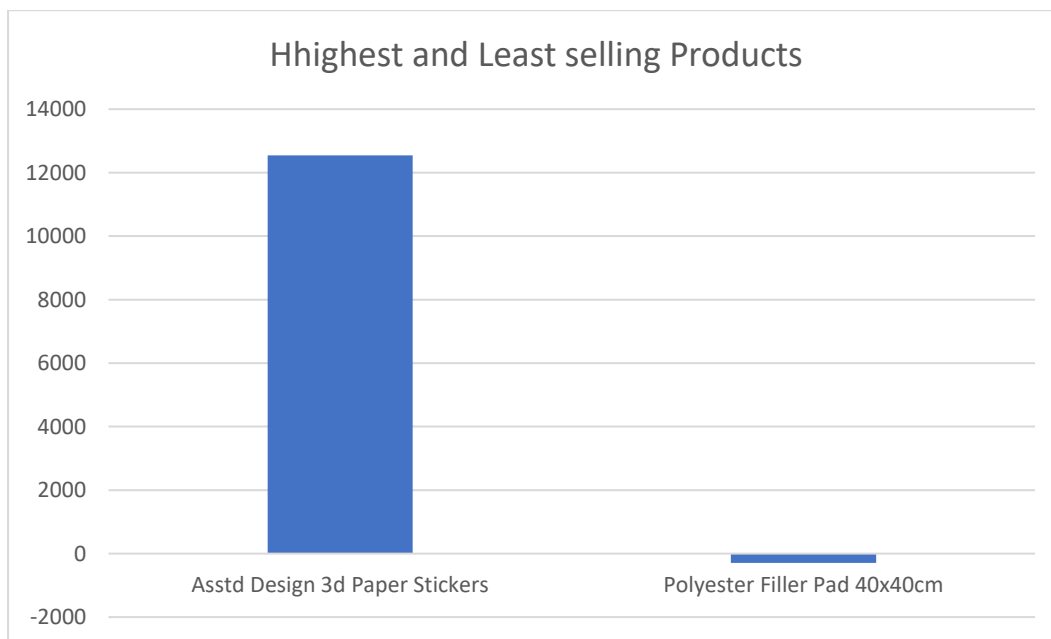
Figure1 sales trends



Most Frequently Purchased Products

Based at the analysed statistics from Excel, the highest selling product is "Asstd Design 3d Paper Stickers" with total quantity sold of 12,540 units. Conversely, the least selling product is "Polyester Filler Pad 40x40cm" with a negative total amount of -296 units, indicating a decrease in inventory as opposed to sales. This shows that "Asstd Design 3d Paper Stickers" is the most regularly bought product, probably due to its recognition amongst clients. Conversely, "Polyester Filler Pad 40x40cm" can also have encountered challenges in demand or faced issues with product quality.

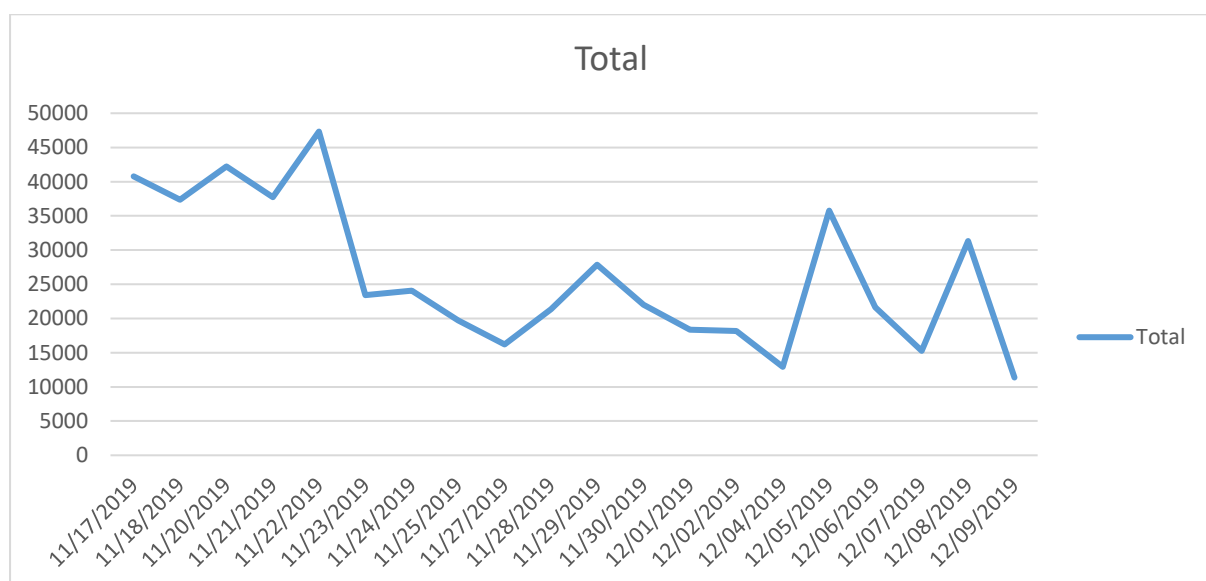
Figure 2: most and least frequently purchased products



Products per Transaction

The data reveals insights into the average number of products purchased per transaction and the total amount of products bought on every date. Similarly, by comparing the metrics revealed by the “count of product No by quantity” highlighted in the Excel Data Sheet in the appendix section, one can discern patterns and fluctuations in consumer behaviour over the period. For instance, on November 23, 2019, despite the fact that the entire amount of products sold was relatively high, the count of product numbers in per transaction was comparatively lower. This shows that whilst customers may have bought a large quantity of products collectively, individual transactions consisted of fewer products. Such trends may be influenced by factors such as promotional offers, seasonal trends, or changes in customer preferences. Analysing these trends can assist corporations understand customer tastes and preferences as well as their shopping behaviour and tailor their promotional techniques to maximise sales and enhance customer satisfaction.

Figure 3: Sales per transaction

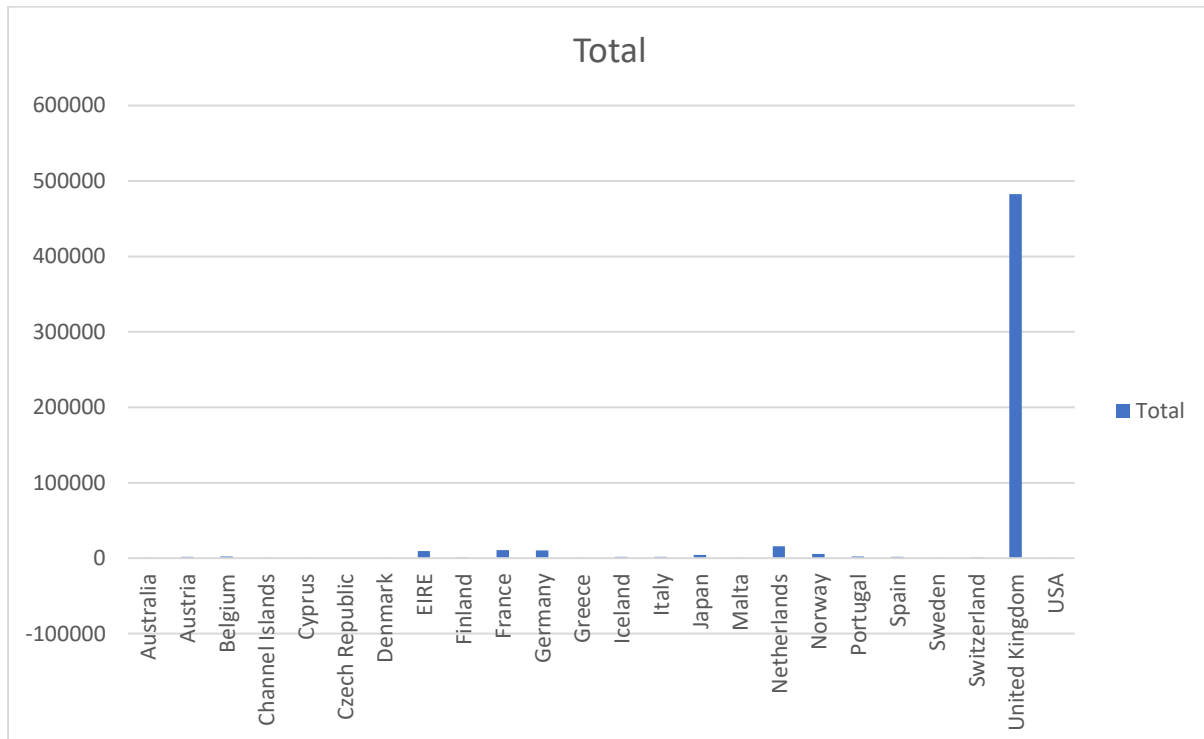


Most Profitable Segment Customers

Analysing total revenue generated by each country as well as customer groups reveals the most profitable segments of customers. Based on the analysed data, the UK emerges as the top contributor to overall sales. The kingdom is followed by EIRE, France, Germany, and the Netherlands. Segmenting customers based totally on purchasing behaviour, order frequency, and average order price can enhance the accuracy of identifying profitable segments. By targeting these segments with tailored advertising strategies and promotions, companies can maximize revenue (Kushwaha, 2022). Additionally, understanding the unique preferences and

customer behaviours of segment can allow the organisation to tailor customer experiences, fostering loyalty and long-term profitability. Therefore, focus on high value segments significantly impacts the overall success and profitability of the business (Kushwaha, 2022).

Figure 4: quantity sold by country



Recommendations for Increasing Profitability

To enhance profitability, organizations have to adopt a multidimensional approach that addresses various aspects of their operations. By focusing on strategic areas consisting of customer segmentation, product optimization, customer experience, supply chain management, and international expansion, companies identify growth opportunities and enhance sales.

Customer Segmentation and Targeted Marketing

Identifying and prioritizing high value customer segments is crucial for effective advertising. By analysing customer data, companies will be able to pinpoint segments with the best potential for profitability. This entails those from the United Kingdom, EIRE, France, Germany, and the Netherlands. Tailoring marketing campaigns and promotions to appeal to the preferences and behaviours of these segments can bring about better conversion rates and consumer retention.

Product Analysis and Optimization

Analysing sales data can enable the businesses to become aware of bestselling merchandise and optimize their availability, pricing, and advertising strategies. By understanding which products resonate maximum with clients, agencies can allocate resources efficiently, improve inventory management, and maximize income (Watkins, 2021). Diversifying product offerings primarily on market tendencies and customer preferences can also appeal to a much wider consumer base and increase general revenue.

Enhanced Customer Experience

Investing in enhancing customer support, satisfaction, and loyalty programs is critical for retaining present customers and attracting new ones (Watkins, 2021). Personalising the shopping experience through targeted recommendations enhances customer engagement and fosters emblem loyalty. By prioritising customer experience, businesses can be able to build strong relationships with customers, leading to repeat purchases and positive word-of-mouth referrals.

International Expansion

Exploring the possibilities for expanding into new markets or increasing market share in current ones is key to driving revenue growth. Profitable regions across Europe offer opportunities for growth. Conducting thorough market studies to identify market opportunities, cultural nuances, and competitive landscapes is critical for success (Watkins, 2021). Adapting advertising techniques and product offerings to match contemporary market conditions can help the companies gain competitive advantage as well as maximize profitability.

By implementing these techniques, businesses can position themselves for achievement in today's competitive market. Regular tracking and analysis of key overall performance indicators can be vital to track the effectiveness of these strategies and make necessary adjustments over the years. By prioritising strategic projects that align with their commercial enterprise targets, companies can enhance sales, increase profitability, and maintain a competitive advantage in the market (Okagbue et al., 2021).

Task 2 IBM SPSS

Descriptive Statistics

		Minimu	Maximu		
N	Range	m	m	Sum	Mean

	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Pretest	2133	71	22	93	117221	54.96	.294
Posttest	2133	68	32	100	143129	67.10	.303
Valid (listwise)	N 2133						

The descriptive statistics offer valuable insights into the distribution of pretest and posttest rankings among the sample populace. The range, minimum and maximum values indicate the spread of scores, while the mean represents the central tendency of the data. The standard errors gives an estimate of the precision of the mean (Harrison et al., 2020). Overall, these data offer a complete overview of the overall performance levels of the students within the pretest and posttest tests, facilitating better interpretation of the dataset.

Gender Frequencies

		Gender			
	Frequency	Percent	Valid Percent	Cumulative Percent	
Male	1056	49.5	49.5	49.5	
Female	1077	50.5	50.5	100.0	
Total	2133	100.0	100.0		

The table shows the frequency and percentage of individuals categorised through gender. Out of 2133 people: 1056 (49.5%) are male and 1077 (50.5%) are females. The valid percentage

and cumulative percent columns affirm that each class's chances are calculated based on the total valid cases.

Lunch Qualification

Lunch

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Does not qualify	1215	57.0	57.0	57.0
	Qualifies for reduced/free lunch	918	43.0	43.0	100.0
	Total	2133	100.0	100.0	

The table shows the frequency and percentage of individuals categorised based on their lunch qualification status. Out of 2133 people, 1215 (57.0%) do not qualify for reduced/free lunch. 918 as such, (43.0%) qualify for reduced/free lunch.

The frequency tables provide a clear overview of the distribution of students by using gender and lunch qualification status. The data indicates a relatively balanced representation of genders, with slightly more males than female student. Regarding lunch qualification, a larger percentage of students do not qualify for reduced/free lunch as compared to people who do. These findings can be valuable for additional analysis or decision making-making processes, such as assessing the demographics of a population or comparing the effectiveness of lunch assistance programs. Additionally, the tables facilitate easy comparison and interpretation of the distribution of people throughout distinct classes, thus assisting in understanding the traits of the dataset.

The mean and median of students who qualifies and does not qualify for reduced/free lunch

Lunch = Does not qualify for free/reduced lunch**Statistics^a**

Lunch

N	Valid	1215
	Missing	0
Mean		.00
Median		.00

a. Lunch = Does not
qualify for free/reduced
lunch

Lunch = Qualify for free/reduced lunch**Statistics^a**

Lunch

N	Valid	918
	Missing	0
Mean		1.00
Median		1.00

a. Lunch = Qualify for
free/reduced lunch

Cluster analysis

The screenshot displays the IBM SPSS Statistics Data Editor and Output Viewer. The Data Editor shows a dataset with variables 'Teaching_method' and 'Gender'. The Output Viewer shows the results of a cluster analysis, including the ANOVA table and the Number of Cases in each Cluster.

ANOVA Table:

	Cluster	Mean Square	df	Error	Mean Square	df	F	Sig.
Student_No	1	10397.273	1	13.013	2131	798.999		.000
Pretest	1	262906.434	1	60.672	2131	4333.262		.000
Posttest	1	285692.567	1	61.657	2131	4633.576		.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	Number of Cases
1	993.000
2	1140.000
Valid	2133.000
Missing	.000

Initial Cluster Centers

	Cluster	
	1	2
Student_No	28	17
Pretest	23	93
Posttest	35	100

Iteration History^a

Change in Cluster Centers

Iteration	1	2
1	30.372	33.776

2	1.005	.898
3	.570	.494
4	.280	.242
5	.183	.158
6	.066	.057
7	.018	.016
8	.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 8. The minimum distance between initial centers is 96.156.

Final Cluster Centers

	Cluster	
	1	2
Student_No	25	21
Pretest	43	65
Posttest	55	78

Distances between Final Cluster Centers

Cluster	1	2
1		32.454
2	32.454	

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Student_No	10397.273	1	13.013	2131	798.999	.000
Pretest	262906.434	1	60.672	2131	4333.262	.000
Posttest	285692.567	1	61.657	2131	4633.576	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	993.000
	2	1140.000
Valid		2133.000
Missing		.000

Final Cluster Centers:

The tables displays the final cluster centers for 2 clusters. Each row represents a variable, and every column represents a cluster. For example, in Cluster 1, the mean value of Student No is 25, the mean value of Pretest is 43, and the mean value of Posttest is 55. In Cluster 2, the mean values for those variables are 21, 65, and 78, respectively.

Distances between Final Cluster Centers

The distances between the final cluster centers are provided in the table. Both clusters have the identical distance of approximately 32.454 units. This indicates a balanced distribution of cases between the clusters.

ANOVA Results

The ANOVA results highlights the F statistics, degrees of freedom, and significance levels for Student_No, Pretest, and Posttest variables. The F tests suggest sizeable differences between clusters for all variables ($p < 0.001$). However, it's critical to note that the tests are primarily for descriptive functions because of the nature of cluster analysis.

The found significance levels are not corrected for the fact that clusters were selected to maximise variations among cases, so they cannot be interpreted as tests of the hypothesis that cluster means are equal.

Number of Cases in Each Cluster:

The table affords the variety of instances in every cluster. Cluster 1 has 993 instances, even as Cluster 2 has 1140 instances, totalling 2133 valid cases. There are no missing cases.

While the clustering analysis successfully grouped cases into two distinct groups based on the variables, the ANOVA outcomes suggest large differences among clusters for all variables. This suggests that the clusters are meaningful in distinguishing between groups (Abu-Bader, 2021). However, it is vital to recognize that cluster analysis is exploratory, and the observed significance need to be interpreted cautiously due to the nature of cluster selection. Additionally, the balanced distribution of cases among clusters suggests that the clustering algorithm correctly separated cases into two relatively equal groups. However, further analysis and interpretation, through the use of additional statistical methods or considering domain knowledge, can provide deeper insights into the characteristics and implications of those clusters for the dataset (Abu-Bader, 2021).

Comparison of SPSS and Excel

Criteria	SPSS	Excel
Data Analysis Capabilities	SPSS offers superior statistical analysis tool, making it suitable for complex analyses such as regression, ANOVA, factor analysis, and cluster analysis. It presents a wider range of alternatives for data manipulation and transformation (Wasito and Birowo, 2023).	Excel provides statistical functions and data analysis tools, suitable for less complex calculations, charts, and pivot tables. However, It lacks the extensive statistical capabilities of SPSS and often requires manual data manipulation complex analyses.
Ease of Use	SPSS has a steeper learning curve, especially for novices, because of its specialised statistical features and interface (Salcedo and McCormick, 2020). However, once familiarised, it offers a user-friendly environment for data analysis.	Excel is extensively used and familiar to many users, offering a more intuitive interface for records access and simple analysis. However, complicated operations may additionally require advanced competencies or formulas (Wasito and Birowo, 2023).
Visualisation	SPSS offers limited options for facts visualisation as compared to Excel. While it offers basic charting features, Excel excels in developing visually attractive and customizable charts, graphs, and dashboards (Salcedo and McCormick, 2020).	Excel is renowned for its effective visualization abilities, allowing users to create diverse chart types, sparklines, and interactive dashboards. Users can without difficulty customise visuals to provide information correctly.
Data Handling	SPSS is designed for handling large datasets successfully, with functions for records cleaning, transformation, and merging. It can deal with complex data systems and formats,	Excel has limitations in coping with large datasets, as it could slow down or crash with extensive information. Similarly, It lacks advanced data management features in comparison

	making it suitable for research and educational functions (Salcedo and McCormick, 2020).	to SPSS, making it much less suitable for complex data manipulation tasks (Wasito and Birowo, 2023).
Collaboration	SPSS files are proprietary and may require specific software to open, limiting collaboration possibilities. However, SPSS allows users to export results to common formats like Excel for sharing and collaboration.	Excel files are generally compatible and can be effortlessly shared and collaborated on with others. This is because the tool integrates seamlessly with different Microsoft Office programs, facilitating collaborative work within groups (Kushwaha, 2022).
Cost	SPSS is a paid software program, requiring a license for ordinary use. While it gives superior features for statistical analysis, the price may be prohibitive for individual users or small businesses (Salcedo and McCormick, 2020).	Excel is included in the Microsoft Office suite, making it more cost-effective for users who already have access to Office applications. However, advanced functions may also require a subscription to Microsoft 365 (Kushwaha, 2022).

Conclusion

Data analysis is one of the most important aspect of decision making within contemporary business organisations. This is for the reason that most businesses rely on fact based decision making that is founded on insights gathered from statistical analysis of various aspects of the business. Both SPSS and Excel are valuable tools for data analysis, with each tool offering specific benefits and limitations. SPSS excels in undertaking complex statistical analyses and coping with huge datasets, making it vital for researchers and analysts in educational and research settings. Its specialised features and strong statistical functions allow users to carry out complex analyses comfortably and generate comprehensive insights from data. On the other hand, Excel's user-friendly interface, versatility, and availability makes it a popular choice for experts across various industries. While Excel may also lack the advanced statistical abilities of SPSS, it is well-suited for basic data analysis, reporting, and visualisation tasks. Its pivot tables, charts, and functions offer users with powerful equipment for organizing and studying

facts effectively. This report has successfully leveraged both tools in analysing two distinct data sets In order to draw valuable insights that can be used in decision making.

References

- Abu-Bader, S. H. (2021). *Using statistical methods in social science research: With a complete SPSS guide*. Oxford University Press, USA.
- Harrison, V., Kemp, R., Brace, N., and Snelgar, R. (2020). *SPSS for Psychologists*. Bloomsbury Publishing.
- Kushwaha, S. (2022). *Statistics and Data Handling through Excel and SPSS* (Doctoral dissertation, Banaras Hindu University (India)).
- Okagbue, H. I., Oguntunde, P. E., Obasi, E. C., and Akhmetshin, E. M. (2021). Trends and usage pattern of SPSS and Minitab Software in Scientific research. In *Journal of Physics: Conference Series* (Vol. 1734, No. 1, p. 012017). IOP Publishing.
- Roni, S. M., and Djajadikerta, H. G. (2021). *Data analysis with SPSS for survey-based research*. Singapore: Springer.
- Salcedo, J., and McCormick, K. (2020). *SPSS Statistics for dummies*. John Wiley and Sons.
- Wasito, B., and Birowo, S. (2023). Pelatihan Aplikasi Komputer Microsoft Word, Microsoft Excel, SPSS Bagi Siswa-Siswa SMA Kristen Tunas Bangsa Cakung. *Bakti Sekawan: Jurnal Pengabdian Masyarakat*, 3(2), 118-120.

Watkins, M. W. (2021). *A step-by-step guide to exploratory factor analysis with SPSS*.
Routledge.

Appendix. Task 1 Excel Workbook



QRK0bSales
Transaction - For A2