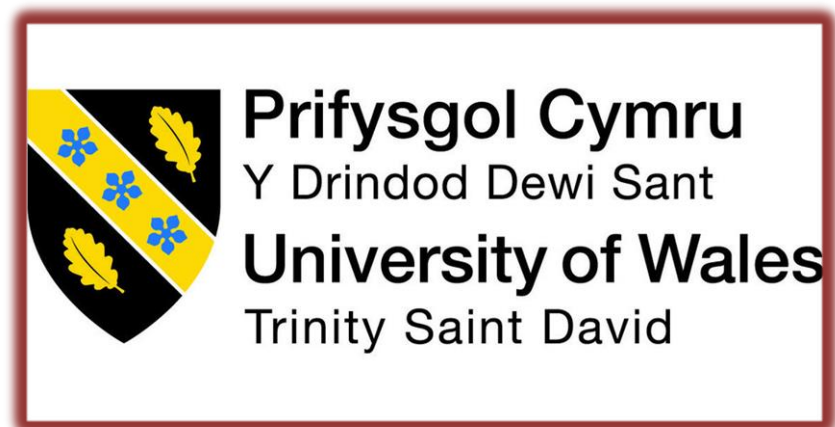




University of Wales, Trinity Saint David

Essay

Module Name: Data Handling & Business Intelligence

Module Code: LCBB5000

Student Name: Mahu Maria

Student ID 2115144

Lecturer: Xavier Matieni

Table of Contents

Part 1	3
Content-Evaluation of the use of Microsoft Excel	3
Analysing the data	4
Visualizing the data	4
Inspiration-Sales trends over the months	5
Most frequently purchased products	7
Most profitable customer segments.....	8
Recommendations	9
Part 2	10
Data Analysis.....	10
Data-mining and text-mining methods.....	14
Advantages and disadvantages of using SPSS and Excel.....	15
Conclusion.....	16
Reference List (Part 1)	17
Reference List (Part 2)	18

Part 1

Content-Evaluation of the use of Microsoft Excel

According to Vidal et al. (2020), even for those who are not familiar with coding, Microsoft Excel makes it easy to manipulate and analyse organised data. The features and tools of Excel might be very helpful if it needs to process structured data. What follows is an explanation and demonstration of the functioning of a few of the many possible ways. An important part of any data analysis process is getting the data ready for use. There are a variety of options available for transforming raw data into a computer-friendly format. The data must be valid and usable before analysis may begin (Díaz de los Ríos and Hernández Ramos, 2020). By checking for inconsistencies and making sure the data is compatible with other sources, preprocessing improves data quality.

- If a certain condition is satisfied, the IF function will return a specific amount. Assuming the line "is 2 greater than 3" to be true, the method assesses the condition. It yields the string "2 > 3" when the condition is true and the string "2 < 3" when it is false.
- To fix any problems, users may want to use the IFERROR code. #N/A, #VALUE!, #DIV/0!, #REF!, #NAME?, and #NUM! are the most typical Excel mistakes. Allow me to illustrate how to fix the mistake by using an example where we try to divide 1 by 0.
- Using the LEFT method, one might get the first 'n' characters of a given text. As an example: Some may look to illustrations for the first three letters of the word "Excel" to easily remember (Hossain and Hossain, 2021). As a string, "Exc" is what the function will give back.
- To get the last "n" characters out of a string, Excel could utilise the RIGHT function. Discovering the last three letters of the word "Excel" is an easy assignment in this situation. The output of the function will be "cel".
- The user has to provide the beginning location and length of the substring to utilise the MID method to extract it from a word (Salkind and Frey, 2021). Studying a sample that separates the word "Excel" into its component string "xce" Starts at 2 and delivers an output of no more than 3 lines.

Analysing the data

As per Campbell et al. (2020), Excel has a stellar reputation for dependability and remarkable adaptability. To make it easier to find certain cell values, the software comes with pre-installed tools including MIN, MAX, and AVERAGE. Users may get a thorough grasp of the data with the help of the Analyse Data in Excel tool by asking questions phrased in ordinary language. With these searches, anybody may ask questions about the data without having to do complex maths. Spreadsheets, data organisation, and analysis are all within Microsoft Excel's purview. Individuals, including analysts, may use spreadsheet programs like Excel to organise data in a way that makes it easier to spot changes or additions (Sara et al. 2023). Several cells organised in rows and columns are standard in Excel.

The properties of the cells are as follows. To analyse data, spot patterns, and create high-level visual reports, users may use the Analyse Data tool. They may move equations to other cells and manage the pointer using the application's drag tool. Users may arrange data into many categories by using the filter feature of the Pivot Table (Nugraha et al.2023). For efficient research, it is essential to learn how to utilise the Pivot Table function. In this case, it breathes new life into the previously inexplicable raw data. Excel is an advanced piece of software with many features that need practice to fully understand.

Visualising the data

Data visualisation in Excel involves making good use of charts, graphs, and other visual components to display data clearly and understandably (Handayani et al. 2021). By drawing attention to underlying trends, relationships, and patterns, data visualisation makes otherwise difficult data easier to understand and analyse. Any company, no matter how big or little, may use Excel to gather and evaluate data. Users can efficiently and effortlessly visualise data because to the software's powerful features and intuitive UI.

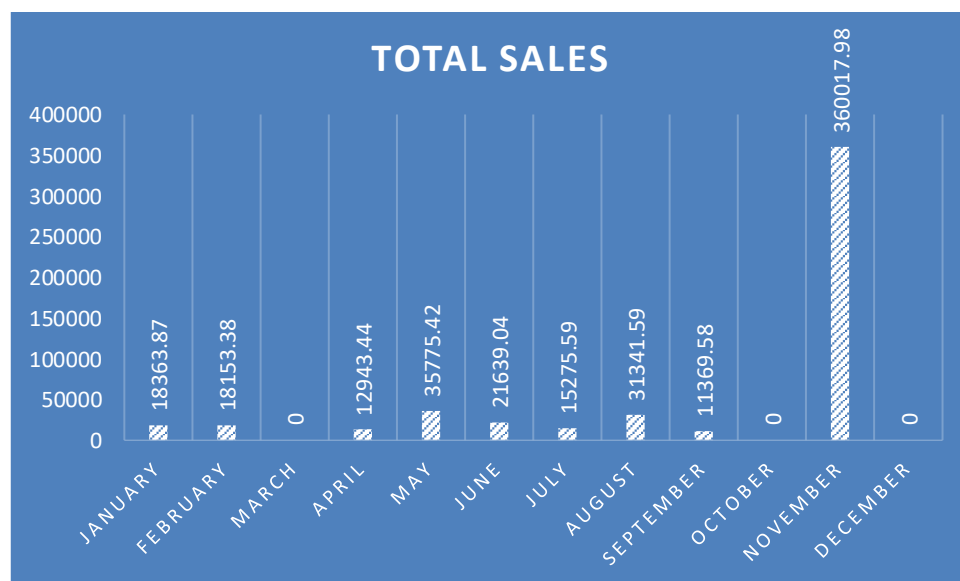
According to Mahato (2023), individuals can produce, edit, organise, analyse, and summarise data with Excel, in addition to working with numbers. Because of its mathematical and scientific computation capabilities, Microsoft Excel is an invaluable tool for resolving logical and mathematical problems.

When used to turn plain text into aesthetically pleasing charts, graphs, and maps, Excel's principal data visualisation tool works wonders. It makes difficult ideas more accessible by breaking them down into their parts. Rather than working with raw data, Excel's data visualisation features analyse it and then display it visually via charts, graphs, and maps to

highlight its importance (Joseph et al. 2023). Data visualization tools make use of a variety of graphic components, including charts, graphs, and maps, to display large datasets in a way that allows users to quickly see patterns, outliers, and trends. For business owners or employees who have to explain complex ideas to those who don't know anything about technology, it is a great tool to have on hand.

Inspiration-Sales trends over the months

Month	Total Sales
January	18363.87
February	18153.38
March	0
April	12943.44
May	35775.42
June	21639.04
July	15275.59
August	31341.59
September	11369.58
October	0
November	360018
December	0



The UK sales had significant fluctuations during the year, with distinct peaks and valleys in several months. Consumers made substantial purchases totaling £18363.87 during the first

month of the year. Consequently, the firm was thriving. However, in February, there was a little deceleration in the increase of sales, as they reached a total of £18153.38. This essentially maintained the performance at the same level.

In March, they saw a complete absence of sales, which deviated from their typical pattern. The rapid decline might be attributed to any changes in the corporate landscape, weather conditions, or market dynamics, and their subsequent impact on human behaviour.

The sales figures for both April and February were identical, amounting to £18153.38. The situation has returned to its usual state. May was a significant month. The sales amounted to £35775.42, representing an increase of almost 100% compared to the previous month. People's preferences may vary due to factors such as sales or seasonal changes, increasing this figure.

In June, the sales amounted to £21639.04, a little lower than the peak reached in May. The current situation remains favorable, notwithstanding a decrease in sales seen last month. What may have caused the decline in sales in July? The amount was £15275.59, which is lower than the amount in the prior month.

In August, the sales figures saw a significant increase, reaching £31,415 once again. This significant surge might be attributed to back-to-school shopping, summer bargains, or other factors that stimulate consumer demand.

In September, sales saw a further decline, resulting in a total revenue of just £11,395.58 for the firm. These declines may be attributed to either natural occurrences or external forces that are now influencing people's behaviour.

October saw zero sales, the same as March. If this recurring event persists, it may indicate that market dynamics are occurring this month that are impacting sales.

In November, sales skyrocketed and reached a staggering £3,06,018. The primary cause for this substantial increase may be attributed to typical factors such as Black Friday and the Christmas shopping season. Attendees of these events tend to increase their expenditure, resulting in a further boost in sales.

December, March, and October all had zero sales. In November, there were none as well. Currently, it is plausible that this occurred due to the imminent conclusion of Christmas shopping. Market dynamics are subject to change, resulting in a decline in sales.

Purchasing patterns in the UK saw significant changes over the year. Various factors contributed to the fluctuations, such as shifts in the industry, seasonal variations, and changes in market dynamics. These factors influence people's behaviour and their expenditure.

Most frequently purchased products

One of the most-bought items all year was the "Paper Craft Little Birdie," which was a customer favourite. That's probably because it was unique, well-made, and useful in many ways. A lot of people wanted and looked for it.

The "Paper Craft Little Birdie" probably became popular because it was fun for a lot of different types of people, from craft and do-it-yourself fans to skilled artists and makers. A lot of people liked it because it added creativity and charm to their projects with its artistic style, attention to detail, and complicated design.

It can use it for a lot of different art projects, like making cards, scrapbooks, home decor, and more. This is another reason why it is so popular and frequently purchased. Despite being useful in many ways, the "Paper Craft Little Birdie" stayed a choice among people of all ages and hobbies.

Because of good reviews, good marketing, and word of mouth, people may have become even more interested in the product. Because of this, it might have become one of the most well-known things in its field.

Number of products the customer purchases in each transaction

Total number of transactions = 65536

Total Quantity of sold products = 553873

Hence, customers purchase in each transaction = $553873 / 65536 = 8.45$

Learning the average number of items bought in a transaction could help shed light on customer buying habits. The total number of things bought was 55,379 pieces. We divide the total number of things by the total number of transactions to obtain the mean amount of products bought per transaction. Based on my estimates, every transaction typically includes around 8.45 products. This number shows the normal amount of goods that a buyer will buy when they shop. Businesses rely on this numerical value when making decisions about product offers, pricing, and marketing tactics.

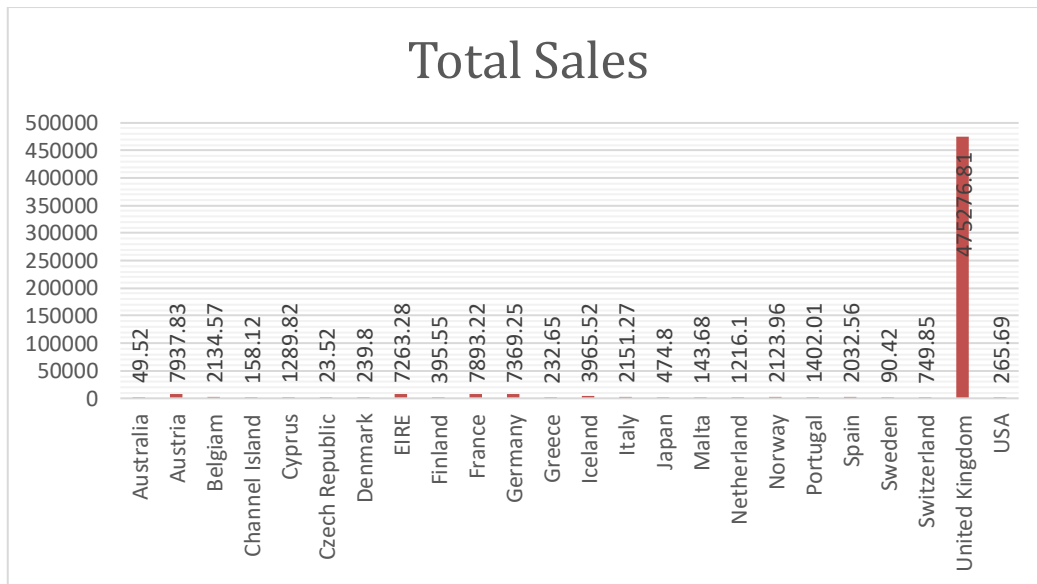
Consumers may be taking advantage of bundled deals, which allow them to buy more things per transaction, or they may be buying more items all at once. Retailers that have access to this data may utilise it to their advantage by boosting sales and coming up with enticing group offers that customers are more likely to purchase.

The other side of this is that a lower average might mean there's room to upsell or cross-sell related products, which would increase the customer's bargain and make their shopping experience better overall. Reviewing the average number of things bought at every event could

help businesses make sure their objectives and client preferences are in sync. This might potentially boost their revenue and encourage consumer loyalty. From an empirical standpoint, this method helps them optimize their retail operations and stay ahead of the competition in a constantly changing market.

Most profitable customer segments

Location	Total Sales
Australia	49.52
Austria	7937.83
Belgium	2134.57
Channel Island	158.12
Cyprus	1289.82
Czech Republic	23.52
Denmark	239.8
EIRE	7263.28
Finland	395.55
France	7893.22
Germany	7369.25
Greece	232.65
Iceland	3965.52
Italy	2151.27
Japan	474.8
Malta	143.68
Netherland	1216.1
Norway	2123.96
Portugal	1402.01
Spain	2032.56
Sweden	90.42
Switzerland	749.85
United Kingdom	475276.8
USA	265.69



Customers who make purchases from the shop with the highest overall sales will have the greatest likelihood of achieving success. With purchases of £47,5276.8, it is evident that individuals in the UK consider this information very valuable. The substantial number of purchases indicates that the majority of the business's revenue is derived from clients in the United Kingdom. While it is possible that there were sales made outside of the UK, they were not substantial in size. Despite the favourable outlook for the UK market, maintaining and expanding market share will need effort.

Identifying the most valuable consumers is of utmost significance for a company. This might include conducting customer satisfaction surveys with consumers in the UK to assess their level of satisfaction with the marketing strategies, goods, and customer service they have received. Utilising data analytics and market research may further enhance its understanding of the preferences and interests of this demographic. Despite the strong sales in Austria, France, and Germany, the United Kingdom surpasses the combined sales of these three nations. These figures unequivocally demonstrate that the United Kingdom is the highest-performing nation.

Recommendations

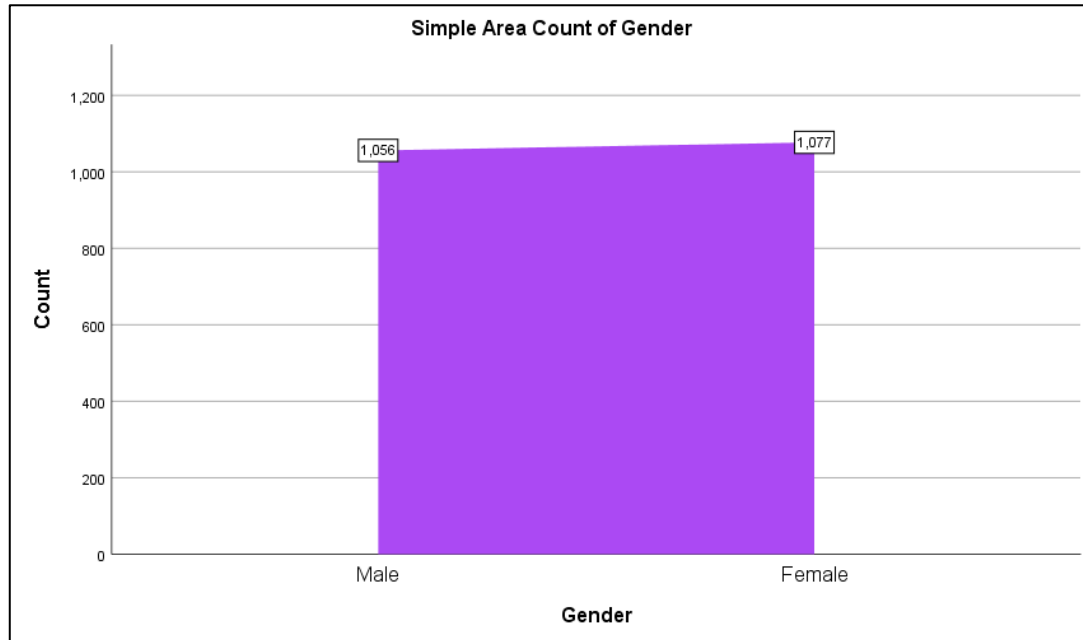
Overall, the United Kingdom market has the utmost significance. Therefore, the corporation must augment its investment in this sector. Customization in business, targeted advertising, and the launch of innovative goods and services to the British population are all achievable. The customer would have an enhanced experience with all of these alternatives. An effective strategy to enhance profitability and meet the high demand in the UK is to optimise and enhance the supply chain. By equipping this staff with the necessary resources, the firm will flourish and earn boundless riches.

Part 2

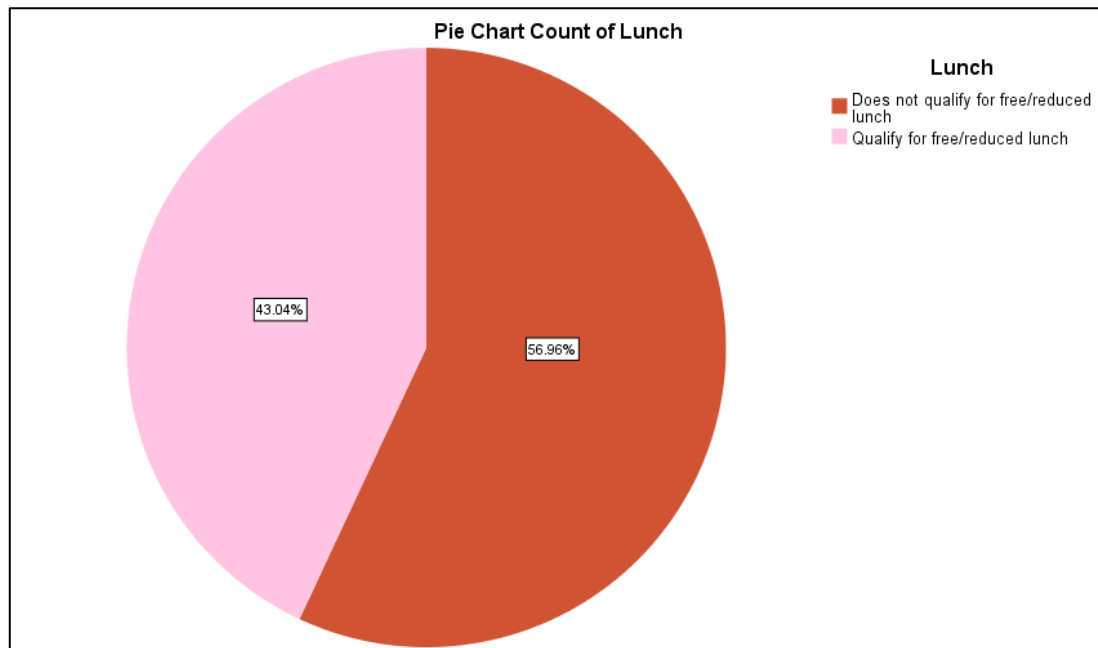
Data Analysis

1. How many students are male and female?

Males – 1056, Females - 1077



2. How many student qualifies and does not qualify for reduced/free lunch?



3. What are the mean and median of students who qualifies and does not qualify for reduced/free lunch?

Students who doesn't qualify for reduced/free lunch

Mean 21.44774

Median 21

Student who qualifies for reduced/free lunch

Mean 24.5817

Median 25

4. Example of Clustering

Example of 3 clusters

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(Gender)	-1.00966	-1.00966	.98997
Zscore(Lunch)	1.15018	-.86902	1.15018
Zscore(Pretest)	-2.35609	2.65751	.07698
Zscore(Posttest)	-2.29518	2.20907	.49317

Iteration History^a

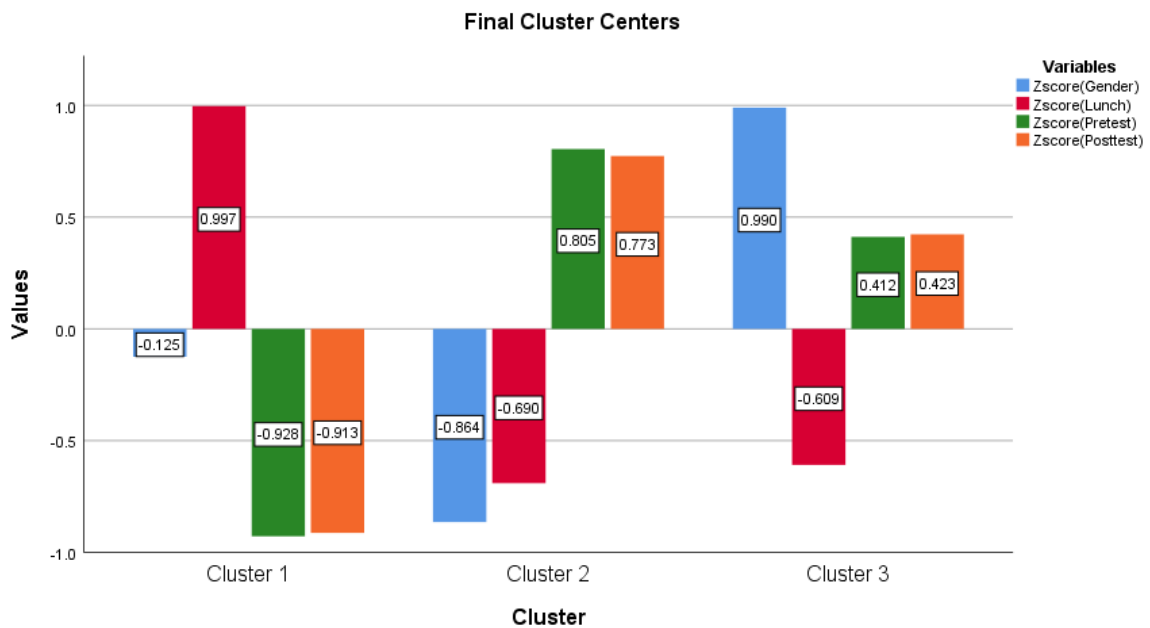
	Change in Cluster Centers		
Iteration	1	2	3
1	1.575	1.762	1.382
2	.399	.261	.189
3	.155	.068	.135
4	.093	.063	.145
5	.062	.058	.109
6	.038	.057	.073
7	.035	.054	.064
8	.017	.043	.050
9	.011	.055	.057

10	.036	.099	.128
----	------	------	------

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is .088. The current iteration is 10. The minimum distance between initial centers is 4.205.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(Gender)	-.12516	-.86434	.98997
Zscore(Lunch)	.99652	-.69039	-.60858
Zscore(Pretest)	-.92823	.80472	.41161
Zscore(Posttest)	-.91329	.77328	.42275



ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Gender)	565.966	2	.470	2130	1205.424	.000
Zscore(Lunch)	690.468	2	.353	2130	1958.153	.000
Zscore(Pretest)	623.093	2	.416	2130	1498.269	.000

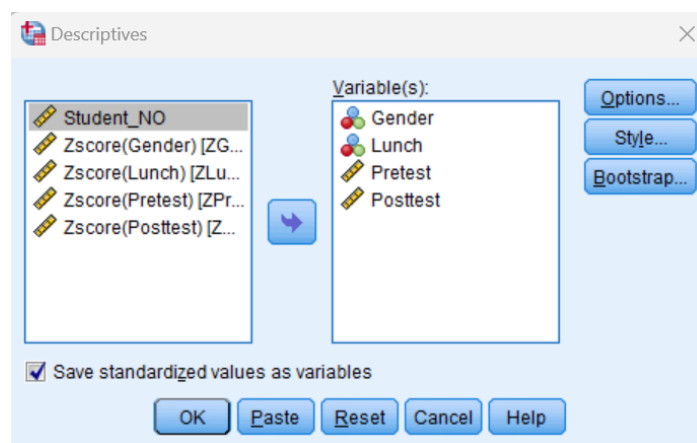
Zscore(Posttest)	598.882	2	.439	2130	1365.412	.000
------------------	---------	---	------	------	----------	------

Number of Cases in each Cluster

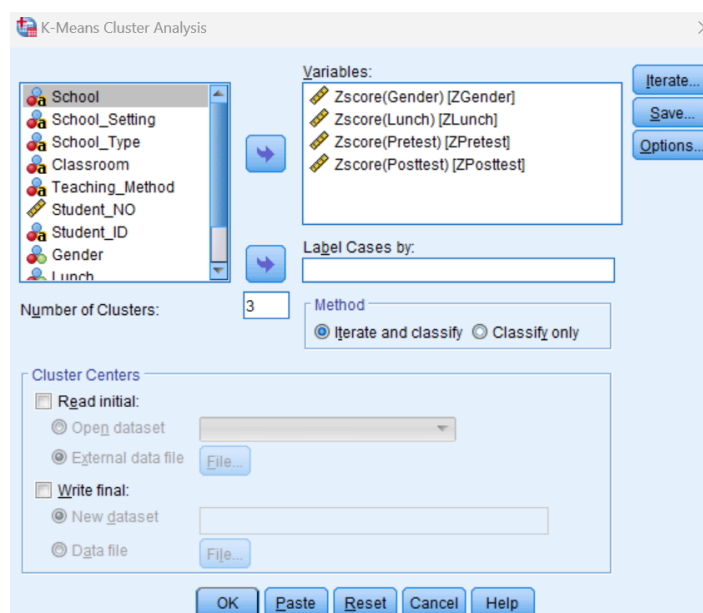
Cluster	1	841.000
	2	633.000
	3	659.000
Valid		2133.000
Missing		.000

Steps to get the outcomes

Descriptive statistics of four variables



K-means cluster with three clusters



Here, the significance of all four variables is zero and it means that the importance of all of those is equal.

Data-mining and text-mining methods

Data mining is the process of observing and analyzing large data sets to identify the patterns within them and extract useful information. There are multiple data mining tools and software for creating data models and testing them. Some of the popular methods for doing so are clustering, classification, regression, etc. All of these have different purposes to fulfil and the data sets are also needed to be compatible with different methods before the implementation (Tandel et al. 2019). For example, clustering is used for grouping similar data points according to the characteristics and features of those. In the business environment, companies use this technique for segmenting their market according to the characteristics and behaviour of customers. Regression is used for making predictions of numeric values under a certain context (Vijayarani et al. 2015). For example, even if an organisation has available data regarding its past sales performance then it can identify specific patterns from those and predict future sales under the same scenarios. Classification is a simple technique for making an existing data structure compatible with new data sources. It is also used for making predictions for specific categories after developing knowledge about those.

The version of data mining is **text mining**; the unstructured text is transformed into a structured format through text mining. Text mining is also known as natural language processing which is necessary for the machines to understand human language and process that automatically. There are multiple types of text mining such as text clustering, entity extraction, text categorization, granular taxonomy, entity relation modeling, sentiment analysis, etc (Alwidian et al. 2015). Now, the application of all these approaches requires the help of specific methods. Some of the major methods are sentiment analytics topic modeling named entity recognition (NER) etc. Sentiment analytics is used for producing structured data to make it easy for machines to understand human language. Some of the areas for it to be implemented are managing interaction with customers, performance appraisal for the employees, etc (Rybchak and Basystiuk, 2017). Topic modeling is used to analyse multiple data sets or texts to summarise them. Most of the applications of artificial intelligence are based on topic modeling these days. NER has multiple applications like legal document analysis, news aggregation etc. as the major purpose of it is to retrieve information through machine translation for answering specific questions.

Advantages and disadvantages of using SPSS and Excel

	Advantages	Disadvantages
SPSS	Computing capabilities and data recording are effective features of it.	There is less scope for handling long question texts using SPSS.
	SPSS is highly effective for doing cluster analysis (Lee and Peters, 2015)	Statistical analysis is good but the generation of graphs and charts is not that flexible and though all the required options are there
	The availability of regression, correlation and descriptive statistics in the same place is beneficial.	It can be expensive to manage certain statistical problems
	This tool is available for free and the free version has enough features for all major operating systems like Windows, Mac, etc.	The input and output files are software-dependent and can't be opened anywhere else
	Advantages	Disadvantages
Excel	It is very easy to use for the formation of general graphs, charts and other visual representations	In terms of statistical analysis, it is not as good as SPSS because of lack of features
	The software is available freely with most of its features from different external sources (Shen et al. 2013)	The options for data analysis are not available in all the MS Excel packages rather the user has to download it
	It is highly beneficial for developing financial plans such as cash flow statements, budgets, profit loss, etc.	There is no automatic update for statistical output (Kaurav et al. 2023).
	Many features in Excel can be added according to the requirements of the users. It reduces the pressure from the application as a user can uninstall a particular feature if it is not necessary to him.	Many different rules and formulas for different calculations are tough to remember (Almquist et al. 2020).

Conclusion

Data analysis is crucial in today's digital world, analyzing data, patterns and trends can predict future outcomes. Data analysis provides valuable insights into customer behavior, preferences and market trends. also helps to stay ahead in today's data-driven economy.

Reference List (Part 1)

Almquist, Y. B., Kvart, S., and Brännström, L. (2020). A practical guide to quantitative methods with SPSS.

Alwidian, S. A. A., Bani-Salameh, H. A., and Alslaity, A. A. N. (2015). Text data mining: a proposed framework and future perspectives. *International Journal of Business Information Systems*, 18(2), 127-140.

Kaurav, R. P. S., Gursoy, D., and Prakash, M. (2020). Why do we need SPSS?. In *An SPSS guide for tourism, hospitality and events researchers* (pp. 20-34). Routledge.

Lee, N., and Peters, M. (2015). *Business statistics using EXCEL and SPSS*. Sage.

Rybchak, Z., and Basystiuk, O. (2017). Analysis of methods and means of text mining. *ECONTECHMOD: An International Quarterly Journal on Economics of Technology and Modelling Processes*, 6(2), 73-78.

Shen, C., Klein, R., Holguin, P., and Kulaylat, A. N. (2023). Statistical software. In *Translational Surgery* (pp. 467-469). Academic Press.

Tandel, S. S., Jamadar, A., and Dudugu, S. (2019). A survey on text mining techniques. In *2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1022-1026). IEEE.

Vijayarani, S., Ilamathi, M. J., and Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science and Communication Networks*, 5(1), 7-16.

Reference List (Part 2)

- Campbell, C.D., Smallwood, Z.M. and Stewart, M.I., 2020. A self-directed workshop for developing advanced data processing and analysis skills in chemistry using Microsoft Excel. *Journal of Chemical Education*, 97(9), pp.2635-2642.
- Díaz de los Ríos, M. and Hernández Ramos, E., 2020. Determination of the Hansen solubility parameters and the Hansen sphere radius with the aid of the solver add-in of Microsoft Excel. *SN Applied Sciences*, 2, pp.1-7.
- Handayani, I.P., Wardani, H.W. and Rifa'i, A., 2021. Pelatihan Penggunaan Microsoft Excel Bagi Pengurus Posyandu. *Jurnal Pengabdian Masyarakat TEKNO*, 2(2), pp.98-103.
- Hossain, E. and Hossain, E., 2021. Introduction to Microsoft Excel. *Excel Crash Course for Engineers*, pp.1-18.
- Joseph, K.A., Ajiboye, F.C., Seun, O.O., Oluwaseun, A. and Babatunde, F., 2023. Analysis and Design of Building's Structural Members (Slabs and Beams) using Microsoft Excel and AUTOCAD Software: A Case Study of MARCGSO'Building. *Suan Sunandha Science and Technology Journal*, 10(2), pp.255-273.
- Mahato, T.K., 2023. Biostatistics: Simple practical way to use Microsoft excel to find mean, median, mode, standard deviation and correlation. *World Journal of Biology Pharmacy and Health Sciences*, 13(3), pp.150-155.
- Nugraha, D., Thursina, F., Devina, P., Irma, A. and Deti, N., 2023. Improving the Efficiency of MSME Financial Statements through Microsoft Excel-Based Applications. *West Science Social and Humanities Studies*, 1(01), pp.10-18.
- Salkind, N.J. and Frey, B.B., 2021. *Statistics for people who (think they) hate statistics: Using Microsoft Excel*. Sage publications.
- Sara, K., Mude, A., Togo, R., Sala, E.E. and Radja, M., 2023. Pelatihan Penggunaan Microsoft Excel Bagi Guru-Guru SDK Bokogo Dalam Mengolah Nilai: Microsoft Excel. *Jurnal Pengabdian kepada Masyarakat Nusantara*, 4(3), pp.2911-2914.
- Vidal, N.P., Manful, C.F., Pham, T.H., Stewart, P., Keough, D. and Thomas, R., 2020. The use of XLSTAT in conducting principal component analysis (PCA) when evaluating the relationships between sensory and quality attributes in grilled foods. *MethodsX*, 7, p.100835.