



Data Handling and Business Intelligence

Assessment N.2

Essay

LCBB5000

Lecturer :Dr.Anupam Mazumdar

Student Name : Luca Gerardi

ID Number: 2219573

Table Of Contents

PART 1 Introduction	3
Microsoft Excel in Pre Processing the data.....	4
Microsoft Excel in Analysing the Data.....	5
Microsoft Excel in Visualizing the Data.....	6
Order of cancellation	7/8
Sales trend over the month of September.....	9/10
Most purchased product	11/12
How many products does the customer purchase in Each transaction.....	13
Most profitable segment customer.....	14
Based on your findings, what strategy could you recommend to the business to gain more profit?.....	15
Part 2/Introduction	16
Rate of male and female Students.....	17/18
How many students qualify and do not qualify for reduced/free lunch?.....	19
What are the mean and median of students who qualifies and does not qualify for reduced/free lunch?.....	20
Using the dataset Student_Score.csv provided in conjunction with SPSS, give a specific example of Clustering (ex. K-Means Clustering).....	21
Advantages/Disadvantages.....	22
Conclusion Resume	23
List of reference.....	24

Part 1

Introduction

Due to its ability to influence consumer purchasing behavior and provide low-cost, effective distribution channels, e-commerce has completely changed the way businesses develop. The dataset from a UK-based online retailer highlights how crucial data analysis is to understanding why purchases are canceled.

The process of methodically utilizing logical and statistical tools to efficiently demonstrate, illustrate, evaluate, and appraise data is known as data analysis. The data analysis of a London-based shop's product and sales trends as well as consumer segmentation will be the main topics of this assignment. This task will be more dependable because it will be conducted using a pivot table and bar chart.

Data is frequently referred to as the new oil in the digital age. However, until it is treated and refined, raw data has limited value, much like crude oil. This is where data preparation, an essential stage in the pipeline for data analysis that guarantees data is clean, consistent, and prepared for additional analysis, comes into play. Because of its many features and tools, Microsoft Excel is still a popular option for preprocessing data, especially for professionals who need to get their data ready for analysis but aren't data scientists.

The efficiency and efficacy of the entire data analytics process, from data pre-processing to analysis and visualization, can be greatly impacted by the tools selected. One of the most used tools in a data analyst's toolbox, Microsoft Excel, has long been an essential component of these procedures. In the context of pre-processing, analyzing, and visualizing data, this essay assesses the benefits and drawbacks of Microsoft Excel, emphasizing its applicability and possible difficulties in each of these stages.

Using of Excel in Pre Processing Data

Microsoft Excel is a fantastic tool for handling and pre-processing structured data when it comes to data pre-processing. Excel contains tools and techniques that make cleaning up the data structure much simpler. Before we go any further, the IF function, IFERROR, LEFT, RIGHT, FIND, COUNT, COUNTIF, and so on are some of the fundamental functions that help with processing vast amounts of data.

The process of converting unprocessed data into an analysis-ready format is known as data preparation. This covers operations such as feature selection, data cleansing, normalization, and transformation. Even though Python and SPSS are two of the more advanced programming languages and tools available for these kinds of jobs, Excel is still a strong and user-friendly choice, particularly for small to medium-sized datasets.

It is also evident that the Excel function is mostly used to check a particular condition and return a particular type of value in accordance. Aside from this, PivotTables are used as an interactive means of rapidly summarizing vast volumes of data. The use of pivot tables is more advantageous in this case as well since they can aid in the detailed examination of numerical data and provide unexpected answers to concerns about the information. This is because the store provides a huge deal of information, both small and precise.

Use of Excel for Analysing the Data

Importing, examining, analyzing, cleaning, and visualizing data may become easier with the use of Microsoft Excel, an essential tool for data analysis. When analyzing data in Excel, users can ask questions about the data without needing to utilize complex calculations because the data is presented in natural language. Consequently, by recognizing patterns, trends, and visualizing the data, Excel aids in its simplicity. Thus, by the efficient application of data in various forms, the Excel file influences how numerical values are interpreted. Additionally, Excel gives users access to

more than 400 operations that may be simplified and analyzed using formulae. As a result, Excel is unquestionably a useful tool for conducting essential data analysis. Excel's extensive function and formula library, which enables users to easily execute intricate computations and data manipulation, is one of its most potent features. Common uses for functions like SUM, AVERAGE, VLOOKUP, INDEX, and MATCH include data aggregation, information finding, and task automation. More sophisticated formulas, such as IF, SUMIF, and array functions, can also be used by advanced users to carry out conditional computations and operate more quickly with huge datasets.

Use of Excel for visualizing the Data

Excel gives users access to a multitude of methods for visualizing data that improves communication and makes it possible to comprehend the story the data is telling. For example, Excel provides a range of charts that can be used for data analysis and interpretation, from intricate 2D charts to 3D models. As a result, Excel offers additional formatting and graphic options that can be used to analyze the datasets contained in the Excel page. Various graphical charts, such as vertical and horizontal bar charts, have been employed in this study to illustrate the data and provide insight into the changes noted in the datasets. As a result, the study effectively makes use of Excel's visualisation capabilities for the dataset analysis. By conducting a comparison analysis over a predetermined time period, the

bar chart utilized in the study aids in reflecting on values and helps provide an effective comprehension of trends and patterns in the data. Because of this, the graphical representation plays a crucial role in simplifying complex numerical values so that the reader may understand them more easily. Excel's visualisation feature is very relevant in educating users about new methods and presenting abilities that enhance the quality of a presentation and give them a deeper comprehension of the data. With its strong data visualization features, Excel is a must-have tool for anyone dealing with data. Because of its accessibility, adaptability, and sophisticated features, users may construct insightful visual representations of their data, which improves comprehension and facilitates improved decision-making. Excel gives you the tools you need to transform unstructured data into meaningful insights, whether you're a researcher, business analyst, instructor, or student.

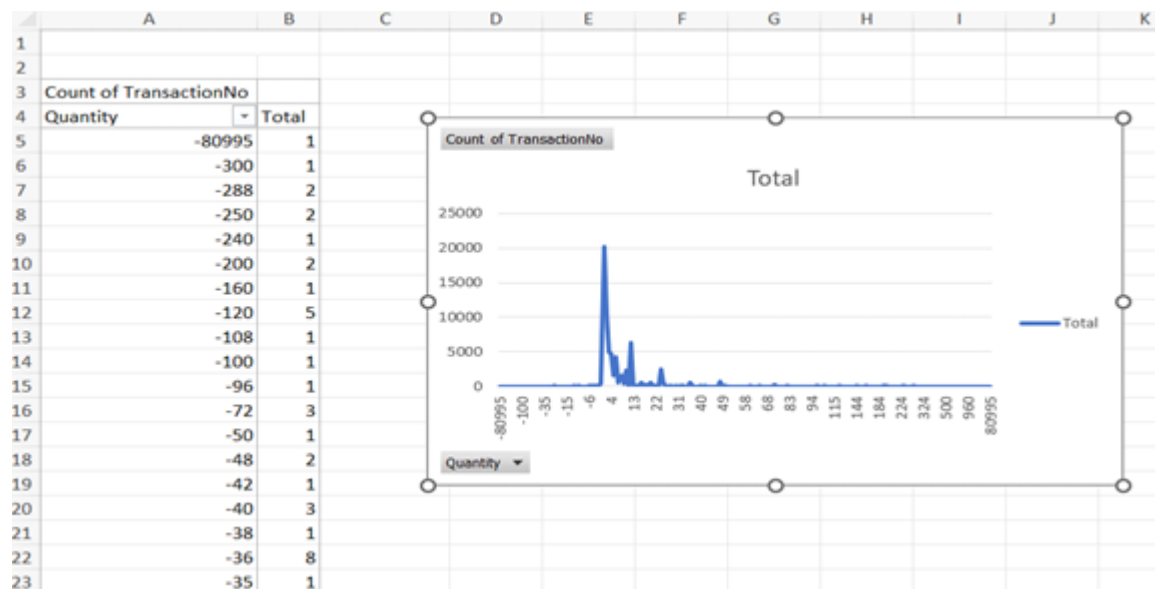
ORDER OF CANCELLATION

In order to determine the negative quantities, first pick the quantity column. Next, click Filter at the bottom of the Data tab. Finally, select the Filtering dropdown in the quantity cell and set a filter that reads "less than 0." The dataset results for locating the cancellation orders are as follows:

1	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country
1618	C581484	12/09/2019	23843	Paper Craft Little Birdie	6.19	-80995	16446	United Kingdom
1619	C581490	12/09/2019	22178	Victorian Glass Hanging T-Light	6.19	-12	14397	United Kingdom
1620	C581490	12/09/2019	23144	Zinc T-Light Holder Stars Small	6.04	-11	14397	United Kingdom
1621	C581568	12/09/2019	21258	Victorian Sewing Box Large	6.19	-5	15311	United Kingdom
1622	C581569	12/09/2019	84978	Hanging Heart Jar T-Light Holder	6.19	-1	17315	United Kingdom
1623	C581569	12/09/2019	20979	36 Pencils Tube Red Retrosport	6.19	-5	17315	United Kingdom
6478	C581228	12/08/2019	22423	Regency Cakestand 3 Tier	6.19	-6	16019	United Kingdom
6479	C581228	12/08/2019	23210	White Rocking Horse Hand Painted	6.19	-12	16019	United Kingdom
6480	C581228	12/08/2019	82494L	Wooden Frame Antique White	6.19	-6	16019	United Kingdom
6481	C581228	12/08/2019	22781	Gumball Magazine Rack	6.19	-24	16019	United Kingdom
6482	C581228	12/08/2019	22467	Gumball Coat Rack	6.19	-36	16019	United Kingdom
6483	C581229	12/08/2019	84828	Jungle Popsicles Ice Lolly Moulds	6.19	-24	12558	USA
6484	C581229	12/08/2019	79191C	Retro Plastic Elephant Tray	6.19	-24	12558	USA
6485	C581229	12/08/2019	23392	Spaceboy Rocket Lolly Makers	6.19	-24	12558	USA
6486	C581229	12/08/2019	23158	Set Of 5 Lucky Cat Magnets	6.19	-36	12558	USA
6487	C581229	12/08/2019	22712	Card Dolly Girl	6.19	-12	12558	USA
6488	C581229	12/08/2019	22027	Tea Party Birthday Card	6.19	-12	12558	USA
6489	C581229	12/08/2019	21508	Vintage Kid Dolly Card	6.19	-12	12558	USA
6490	C581229	12/08/2019	21507	Elephant Birthday Card	6.19	-12	12558	USA
6491	C581229	12/08/2019	85099B	Jumbo Bag Red Retrosport	6.19	-10	12558	USA
6492	C581229	12/08/2019	23344	Jumbo Bag 50'S Christmas	6.19	-10	12558	USA
6493	C581229	12/08/2019	23209	Lunch Bag Vintage Doily	6.19	-20	12558	USA
6494	C581235	12/08/2019	23243	Set Of Tea Coffee Sugar Tins Pantry	6.19	-1	15950	United Kingdom
6495	C581235	12/08/2019	22840	Round Cake Tin Vintage Red	6.19	-2	15950	United Kingdom
6496	C581235	12/08/2019	23459	Dolly Cabinet 2 Drawers	6.19	-2	15950	United Kingdom
6497	C581235	12/08/2019	23273	Heart T-Light Holder Willie Winkie	6.19	-1	15950	United Kingdom

Than in order to visualize singularly

insert a Pivot Table new Worsheet and ok than and drag the fields labeled "quantity" and "transactionNo" into the Rows and Values sections, respectively. Then, view the results using the pivot chart:



Sales Trend over the Month / September

Create new Columns and

Sales for total sales using the formula `'=E2*F2'` and dragged down to the list.

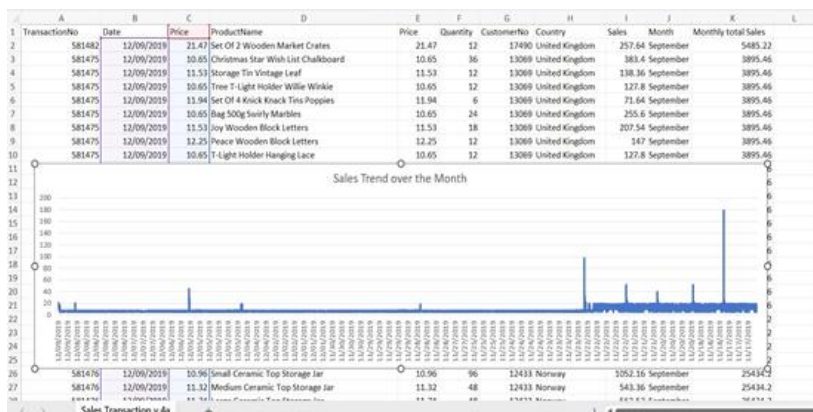
Month using the formula `'=TEXT (B2, "mmmm")'`, to get the month from data and dragged down to the list.

Monthly total Sales with the formula: `'=SUMIFS (I:I, G:G, G2)'`, to calculate the sum of sales for each month.

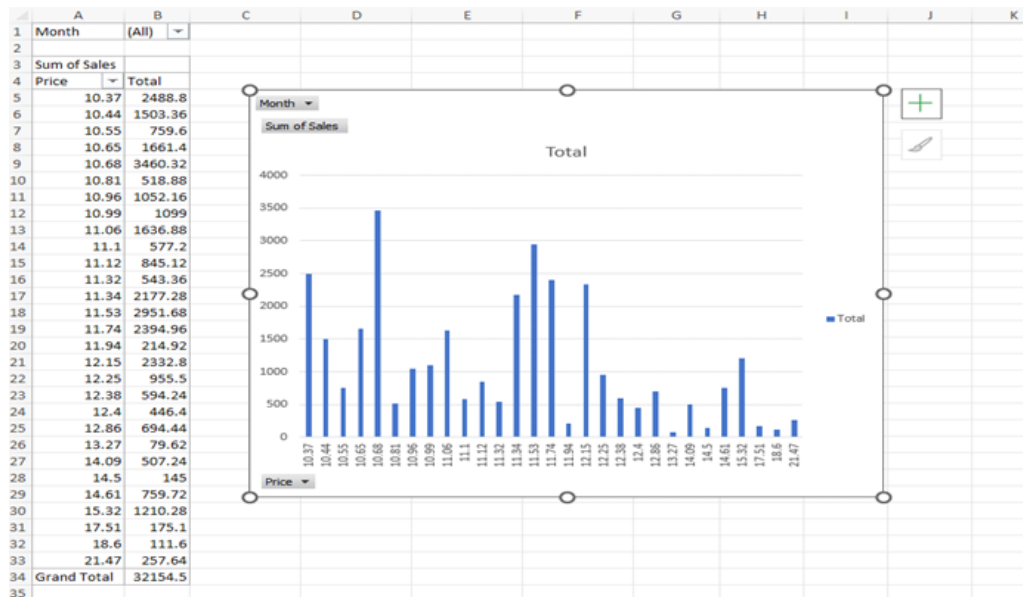
	A	B	C	D	E	F	G	H	I	J	K
1	TransactionNo	Date	ProductNo	ProductName	Price	Quantity	CustomerNo	Country	Sales	Month	Monthly total Sales
2	581482	12/09/2019	22485	Set Of 2 Wooden Market Crates	21.47	12	17490	United Kingdom	257.64	September	5485.22
3	581475	12/09/2019	22596	Christmas Star Wish List Chalkboard	10.65	36	13069	United Kingdom	383.4	September	3895.46
4	581475	12/09/2019	23235	Storage Tin Vintage Leaf	11.53	12	13069	United Kingdom	138.36	September	3895.46
5	581475	12/09/2019	23272	Tree T-Light Holder Willie Winkie	10.65	12	13069	United Kingdom	127.8	September	3895.46
6	581475	12/09/2019	23239	Set Of 4 Knick Knack Tins Poppies	11.94	6	13069	United Kingdom	71.64	September	3895.46
7	581475	12/09/2019	21705	Bag 500g Swirly Marbles	10.65	24	13069	United Kingdom	255.6	September	3895.46
8	581475	12/09/2019	22118	Joy Wooden Block Letters	11.53	18	13069	United Kingdom	207.54	September	3895.46
9	581475	12/09/2019	22119	Peace Wooden Block Letters	12.25	12	13069	United Kingdom	147	September	3895.46
10	581475	12/09/2019	22217	T-Light Holder Hanging Lace	10.65	12	13069	United Kingdom	127.8	September	3895.46
11	581475	12/09/2019	22216	T-Light Holder White Lace	10.55	24	13069	United Kingdom	253.2	September	3895.46
12	581475	12/09/2019	22380	Toy Tidy Spaceboy	11.06	20	13069	United Kingdom	221.2	September	3895.46
13	581475	12/09/2019	22442	Grow Your Own Flowers Set Of 3	12.25	12	13069	United Kingdom	147	September	3895.46
14	581475	12/09/2019	22664	Toy Tidy Dolly Girl Design	11.06	20	13069	United Kingdom	221.2	September	3895.46
15	581475	12/09/2019	22721	Set Of 3 Cake Tins Sketchbook	12.25	12	13069	United Kingdom	147	September	3895.46
16	581475	12/09/2019	22723	Set Of 6 Herb Tins Sketchbook	11.53	12	13069	United Kingdom	138.36	September	3895.46
17	581475	12/09/2019	22785	Squarecushion Cover Pink Union Jack	11.53	12	13069	United Kingdom	138.36	September	3895.46
18	581475	12/09/2019	22955	36 Foil Star Cake Cases	11.06	24	13069	United Kingdom	265.44	September	3895.46
19	581475	12/09/2019	23141	Triple Wire Hook Pink Heart	11.06	12	13069	United Kingdom	132.72	September	3895.46
20	581475	12/09/2019	22956	36 Foil Heart Cake Cases	11.06	24	13069	United Kingdom	265.44	September	3895.46
21	581475	12/09/2019	22581	Wood Stocking Christmas Scandispot	10.55	48	13069	United Kingdom	506.4	September	3895.46
22	581476	12/09/2019	23198	Pantry Magnetic Shopping List	11.53	48	12433	Norway	553.44	September	25434.2
23	581476	12/09/2019	23197	Sketchbook Magnetic Shopping List	11.74	24	12433	Norway	281.76	September	25434.2
24	581476	12/09/2019	23184	Bull Dog Bottle Opener	15.32	8	12433	Norway	122.56	September	25434.2
25	581476	12/09/2019	23168	Classic Cafe Sugar Dispenser	11.53	12	12433	Norway	138.36	September	25434.2
26	581476	12/09/2019	23167	Small Ceramic Top Storage Jar	10.96	96	12433	Norway	1052.16	September	25434.2
27	581476	12/09/2019	23166	Medium Ceramic Top Storage Jar	11.32	48	12433	Norway	543.36	September	25434.2

Only sales for are included in the chosen rows from the dataset that is provided.

September, thus it is impossible to determine a pattern over a number of months. Insert a Line Chart and choose the ranges that contain the Date and the Price that we wanna identify in order to see the Sales Trend over the Month, in this case, September , display:



Create a Pivot Table, drag the Month field to Rows and the Sales field to Values to examine sales over the course of the month. Then, attach a Bar Chart to display the results. 32154.5 was the grand total of the first 100 rows chosen.



Most Frequently Purchased Product

First, create new columns to identify the Unique Product and calculate the Occurrences of each in order to identify the most commonly purchased goods from the dataset (100 rows picked from the provided dataset). Unique Product by counting the number of times the product name appears in the initial "ProductName" column using the Countif function.

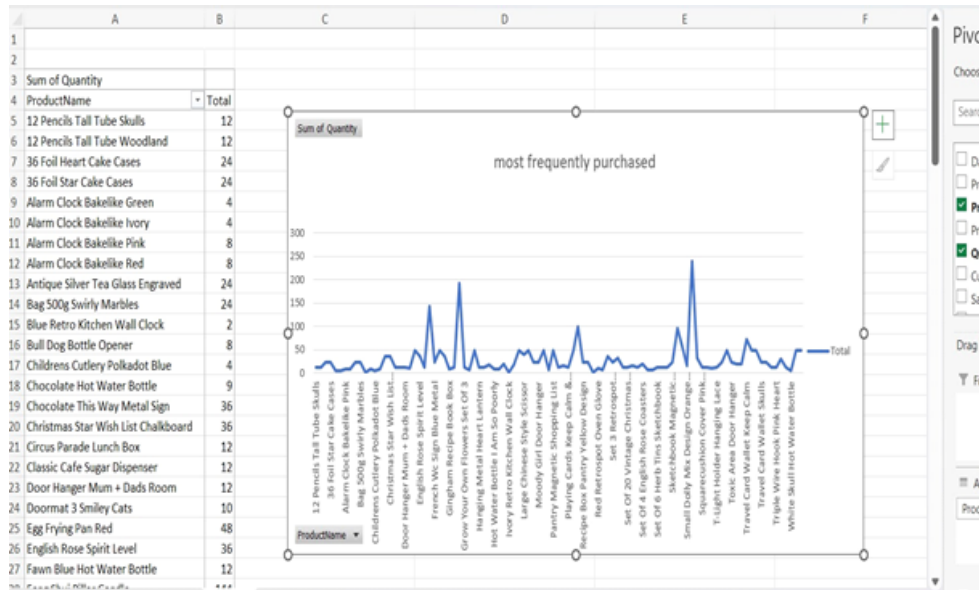
Here following Formulas:

Unique Product =UNIQUE (D2:D100)

Occurrence =COUNTIF (D:D, L2)

Than click Enter and drag down.

Create a new Pivot Table, drag the ProductName and Quantity fields to the Rows and Values areas, and you will be able to see and analyze the most popular products from the selected dataset. Next, include a line chart to see which product is most frequently bought. This shows that, with 240 purchases, the "Small Folding Scissor (Pointed Edge)" is the most popular item:



	A	B	C
79	Small Chinese Style Scissor	60	
80	Small Dolly Mix Design Orange Bowl	16	
81	Small Folding Scissor(Pointed Edge)	240	
82	Small Marshmallows Pink Bowl	32	
83	Squarecushion Cover Pink Union Jack	12	
84	Storage Tin Vintage Leaf	12	
85	Tea Time Oven Glove	10	
86	T-Light Holder Hanging Lace	12	
87	T-Light Holder White Lace	24	
88	Toilet Metal Sign	48	
89	Toxic Area Door Hanger	24	
90	Toy Tidy Dolly Girl Design	20	
91	Toy Tidy Spaceboy	20	
92	Travel Card Wallet Keep Calm	72	
93	Travel Card Wallet Pantry	48	
94	Travel Card Wallet Retrosport	48	
95	Travel Card Wallet Skulls	24	
96	Travel Card Wallet Transport	24	
97	Tree T-Light Holder Willie Winkie	12	
98	Triple Wire Hook Pink Heart	12	
99	Vintage Doily Travel Sewing Kit	30	
100	Wall Art Only One Person	12	
101	White Skull Hot Water Bottle	4	
102	Wood Stocking Christmas Scandisport	48	
103	You're Confusing Me Metal Sign	48	
104	Grand Total	2793	

How many products does the customer purchase in each transaction?

Create two additional new columns with the names "Unique TransactionNo" and "Products per Transaction" to calculate the number of products purchased in each transaction. Then, use the Countif function to count the number of times that transaction number appears, which is the number of products that were purchased in that transaction. Formulas applied:

Unique TransactionNo =Unique (A2: A100)

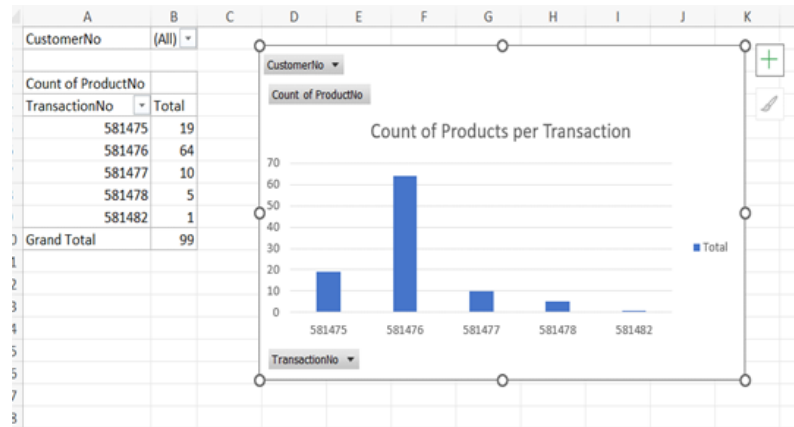
Products per Transaction =COUNTIFS(A: A, N2) to find the count for each, drag this formula down to the list of unique transaction numbers.

Create than a:

“Pie charts, or circular graphs as they are also known, are a popular way to show how much individual amounts or percentages contribute to the total.”

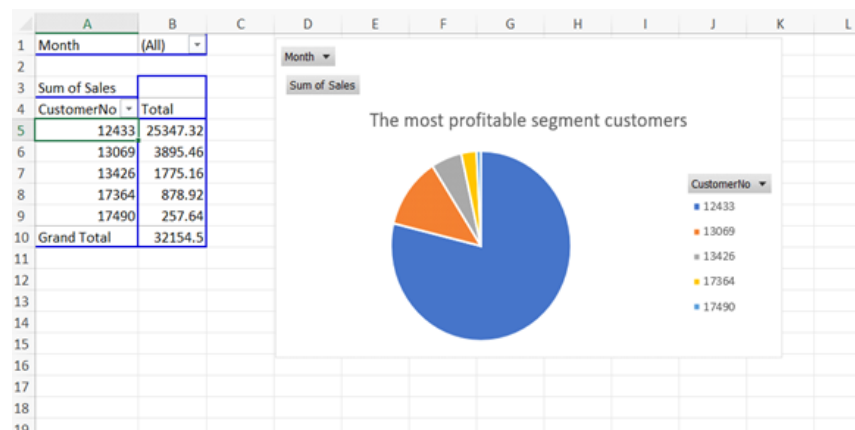
(Ablebits).

To examine customer purchases in each transaction, a new pivot table was made by dragging the values for "ProductName," "TransactionNo," and "CustomerNo" into the Rows, Filters, and Values, respectively. The findings indicated that whereas only one product was purchased in "TransactionNo." 581482, 64 products were purchased in "TransactionNo." 58476):



What are the most profitable segment customers?

Create a pivot table and customize it by dragging the "CustomerNo" column to the Rows and the "Sales" field to the Values from the selected dataset (the first 100 rows). With a total of 25347.32, this study reveals that customer 12433 from Norway has concealed sales, suggesting potential profitability:



Based on your findings, what strategy could you recommend to the business to gain more profit?

You can better manage your sales staff and plan future marketing tactics based on past performance by utilizing spreadsheets to list customer and sales targets with tools like Excel and Solver which is a Microsoft Excel add-in program (Microsoft.com,2024).

With the the benefit that there targeted searches offer,and from the datas that we analyzed and found on the above sections,we could apply or

implement some marketing strategies to the business in order to let it growth and be more rationalized.

Some of the strategies might be for example Market Expansion, investigate untapped markets by examining sales patterns and demographic information.

Fit pricing strategy: use sales data to modify prices in response to customer price sensitivity, competition, and demand.

Customer segmentation: review past purchases made by the customer to distinguish different sections. To boost sales, provide each section with a customized campaign.

Applying these strategies or implementing them can growth or boom the business, by periodically check and control those data analysis based on specifications produced of and by the product itself.

PART 2

Introduction

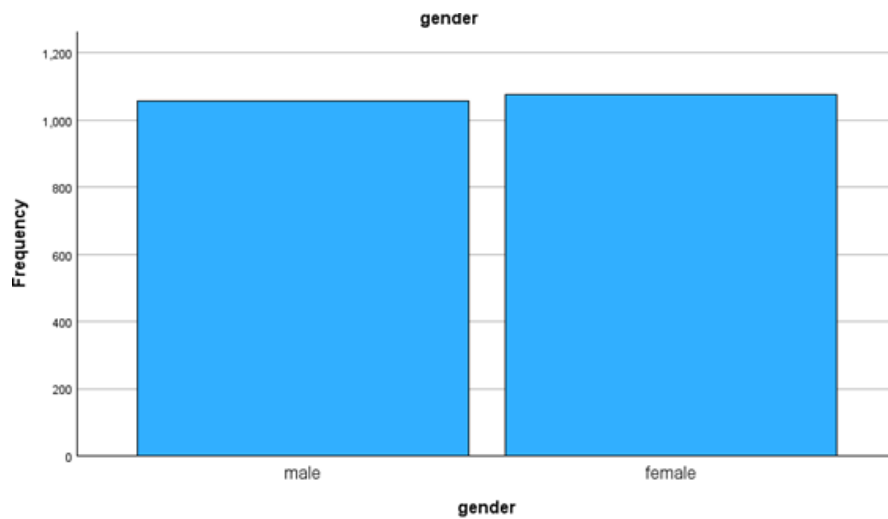
In Part 2 we will face to assist a local school in analysing survey information that was gathered from 2,133 students. This analysis's main goal is to investigate the connections between pre- and post-test results, gender, and students' eligibility for reduced or free lunch programs. The school hopes to gain a better understanding of the possible effects of socioeconomic status on gender disparities and academic achievement by looking at these linkages. The institution has traditionally used Microsoft Excel for data analysis. They are eager to use the open-source SPSS data mining program's sophisticated features to extract more meaningful insights from the data, though, as they recently acquired it. As I lead them through this shift, I will make sure that SPSS is used to reliably and efficiently analyze the data, enabling the school to make well-informed decisions to support the needs of their children.

Rate of male and female students

Analyze, click on *Descriptive Statistics*, then *Frequencies*. Select the *Gender* and click on the arrow from the middle, then on *OK*:

Gender					
Frequency			Percent	Valid Percent	Cumulative Percent
Valid	male	1056	49.5	49.5	49.5
	female	1077	50.5	50.5	100.0
	Total	2133	100.0	100.0	

And then using the bar chart



How many students qualify and do not qualify for reduced/free lunch?

Analyze, click on Descriptive Statistic, and select Lunch to move to the right side under Variable(s) in Frequencies.

lunch		Frequ e n c y	Perce nt	Vali d Perce nt	Cumulati ve Percent
Valid	does not qualify for	1215	57.0	57.0	57.0

	free/r educ d lunch				
	qualify for free/reduc ed lunch	918	43.0	43.0	100.0
	Total	2133	100.0	100.0	

An analysis of meal qualification reveals that 918 students or 43% of the total student body qualified for free or reduced lunch, while 1,215 students or 57% of the total student body do not qualify for either.

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

Analyze, then Compare Means and Proportions, select Means, drag Gender into Dependent List, and Lunch into Layer 1 of 1, and click OK.

does not qualify for free/reduced lunch	1.51	2.00
qualify for free/reduced lunch	1.50	1.00
Total	1.50	2.00

Using the dataset Student_Score.csv provided in conjunction with SPSS, give a specific example of Clustering (ex. K-Means Clustering). Show your workings with screenshots and clearly explain your results.

Analyze, select Classify, K-Means Cluster Analysis. Select 'Student- No' and include into 'Variables', complete the Number of Clusters with 20, Iteration and select 3 Maximum Iterations, Continue and click OK .

Initial Cluster Centers							
	1	2	3	Cluster 4	5	6	7
Student_No	20	21	18	15	16	19	17
Initial Cluster Centers							
	8	9	10	Cluster 11	12	13	14
Student_No	28	27	24	14	22	23	31
Initial Cluster Centers							
	15	16	Cluster 17	18	19	20	
Student_No	25	26	29	30	20	20	
Final Cluster Centers							
	1	2	3	Cluster 4	5	6	7
Student_No	20	21	18	15	16	19	17
Final Cluster Centers							
	8	9	Cluster 10	11	12	13	14
Student_No	28	27	24	14	22	23	31
Final Cluster Centers							
	15	Cluster 16	17	18	19	20	
Student_No	25	26	29	30	.	.	

5. Discuss the advantages and disadvantage of using SPSS and Excel using relevant practical examples and present your discussions in a table by using different criteria

Excel	SPSS
Excel has a plethora of capabilities that can be leveraged to effortlessly satisfy	Strong tools are available in SPSS Statistics, as opposed to spreadsheets, for handling

data analysis requirements. For example, Excel's Analyze Data tool, which effectively expedites the process to obtain higher value(Microsoft,2024).	missing data and looking for outliers that could impede analysis.(IBM,2018)
---	---

Disadvantages	
Excel	SPSS
Accidentally inserting a number into a cell already containing a formula will overwrite the equation and turn the cell content into a constant.(IBM,2018)	- Large data sets provide a number of important issues when employing SPSS analysis. - Using SPSS tools demands a high level of proficiency.

Conclusion Resume

To sum up, the study has been comprehensive in finding and critically assessing data to identify the primary patterns. A pivot table has been analyzed using the information provided about a UK-based online retailer to ascertain the possibility of the firm expanding on its own. Furthermore, the student data was also analyzed using the SPSS analytical tool in order to identify the various factors that can be useful in determining a student's eligibility for lunch as well as their performance on the pre- and

post-test. The impact in the data set is ascertained using the mean and median values.

Reference List

Ablebits, (2015). *Make Pie Chart Excel*, Available at: <https://www.ablebits.com/office-addins-blog/make-pie-chart-excel/#:~:text=Right%20click%20any%20slice%20on,the%20entire%20pie%20representing%20100%25>. (Accessed: 19 August 2024)

Microsoft, (2021). *Define and solve a problem by using Solver, Microsoft* . Available at: <https://support.microsoft.com/en-gb/office/define-and-solve-a-problem-by-using-solver-5d1a388f-079d-43ac-a7eb-f63e45925040> (Accessed: 18 August 2024).

IBM, (2024) *spss-statistics* Available at: <https://www.ibm.com/products/spss-statistics> (Accessed: 10 June 2024).

IBM,(2018) *The risk of using spreadsheets for statical analysis* Available at: <https://www.ibm.com/downloads/cas/7YEX9BKK> (Accessed: 19 august 2024).

MEDIUM(2024) *Data Analysis* Available at <https://medium.com/tag/data-analysis> (Accessed: 19 august 2024).