Programme: BA (HONS) BUSINESS MANAGEMENT

Module: Data Handling and Business Intelligence

Module Code: LCBB5000

Name: Simranjeet Kaur

Student id: 2225768

Lecturer: Dr Anupam Mazumdar

Table of Contents

Part 1 3

Introduction 3

Change of Sales Trend Over the Months 3

Most Frequently Purchased Products 8

Products purchased by the customers in each transaction 9

Most Profitable Segment Customers 23

Recommendation to the business based on findings to gain more profit 24

Gender 25

Number of students qualified and does not qualify for reduced/free lunch 27

Mean and median of students who qualifies and does not qualify for reduced/free lunch 28

K-Means Clustering 28

Advantages and Disadvantages of using SPSS and Excel using relevant examples and based on different criteria *31*

Overall Summary 32

Reference List 33

Part 1

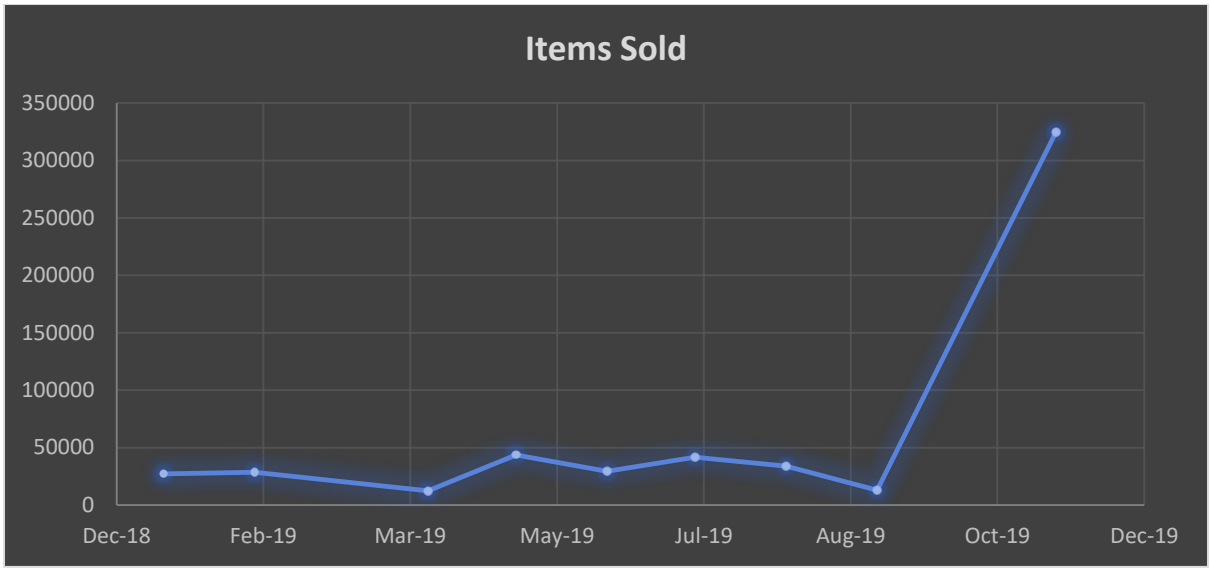
Introduction

The section has helped in understanding the ways in which transactions are done and the most sold products of the e-commerce business operating since 2007. There is identification of the sales change that has happened over the months and also the most frequently purchased products. There is identification of the number of transactions that have been done by customers in each such transaction along with identification of the most profitable target segment customers. This has been done by identifying the regions from which the highest number of sales has happened. It has been done by identifying the regions from which highest number of sales has occurred. There is change in the sales percentage observed in each month and it has been identified by use of the MS EXCEL functions. Data analysis has been done with help of MS Excel analytical tool.

Change of Sales Trend Over the Months

Months	Items Sold
Jan-19	27252
Feb-19	28483
Apr-19	12381
May-19	43626
Jun-19	29329
Jul-19	41743
Aug-19	33859
Sep-19	12901
Nov-19	324753

In January, a total of 27252 units are sold and it has increased in February to 28483 units followed a decrease to less than half the amount. **In April**, the sales decreased to that of 12381 units and then it increased by more than 3.5 times to 43626 units in May, 2019. **In June**, the sales then decreased to that of 29329 units and further increased to 41743 units in July. This shows that the sell done by the e-commerce business has been erratic. The sales the decreased further to 33859 and 12901 before increasing by a huge amount in November and accounted to 324753 units.



Column1	*January*
Mean	9.44610052
Standard Error	0.534267091
Median	4
Mode	1
Standard Deviation	28.6966587
Sample Variance	823.4982207
Kurtosis	1038.935879
Skewness	26.5904535
Range	1199
Minimum	1
Maximum	1200
Sum	27252
Count	2885

February	
Mean	9.983526113
Standard Error	0.837562417

Median	3
Mode	1
Standard Deviation	44.7371246
Sample Variance	2001.410318
Kurtosis	653.4860228
Skewness	23.14176147
Range	1550
Minimum	50
Maximum	1500
Sum	28483
Count	2853

April	
Mean	9.98353
Standard Error	0.83756
Median	3
Mode	1
Standard Deviation	44.7371
Sample Variance	2001.41
Kurtosis	653.486
Skewness	23.1418
Range	1550
Minimum	50
Maximum	1500
Sum	28483
Count	2853

May	
Mean	8.232874
Standard Error	0.348361

Median	2
Mode	1
Standard Deviation	25.35865
Sample Variance	643.0613
Kurtosis	287.8966
Skewness	13.83174
Range	840
Minimum	40
Maximum	800
Sum	43626
Count	5299

June	
Mean	8.765391512
Standard Error	0.359424207
Median	3
Mode	1
Standard Deviation	20.79075649
Sample Variance	432.2555553
Kurtosis	57.07133776
Skewness	5.207082701
Range	538
Minimum	250
Maximum	288
Sum	29329
Count	3346
July	
Mean	17.3568
Standard Error	1.23813
Median	6
Mode	1

Standard Deviation	60.719
Sample Variance	3686.79
Kurtosis	275.6
Skewness	14.4895
Range	1478
Minimum	38
Maximum	1440
Sum	41743
Count	2405

August	
Mean	6.87772
Standard Error	0.50618
Median	2
Mode	1
Standard Deviation	35.5158
Sample Variance	1261.37
Kurtosis	1032.13
Skewness	28.6631
Range	1800
Minimum	300
Maximum	1500
Sum	33859
Count	4923

September	
Mean	7.95376
Standard Error	70.642
Median	3
Mode	1

Standard Deviation	2845.04
Sample Variance	8094259
Kurtosis	810.449
Skewness	-0.0084
Range	161990
Minimum	80995
Maximum	80995
Sum	12901
Count	1622

November	
Mean	8.08366
Standard Error	0.33374
Median	3
Mode	1
Standard Deviation	66.8934
Sample Variance	4474.73
Kurtosis	30689.9
Skewness	164.903
Range	12828
Minimum	288
Maximum	12540
Sum	324753
Count	40174

Most Frequently Purchased Products

Product Name	Price	Quantity
Paper Craft Little Birdie	6.19	80995
Polyester Filler Pad 40x40cm	6.19	300
Boys Alphabet Iron on Patches	6.19	288
Victorian Metal Postcard Spring	6.13	288
Office Mug Warmer Pink	6.19	250

The section has helped in understanding the most frequently purchased products and the top five product items that have been sold the highest number of times from the e-commerce website is identified. It has been observed that the Paper Craft Little Birdie has been bought 80995 number of times and is the maximum selling product followed by Polyester Filler Pad 40*40 cm which has 300 units sold followed by Boys Alphabet Iron on Patches getting sold 288 times along with Victorian Meta postcard spring. Office Mug Warner Pink has been bought 250 number of times by the customers.

Products purchased by the customers in each transaction

Transaction No	Quantity
578459	96.00
578459	20.00
578459	20.00
578459	20.00
578459	24.00
578459	20.00
578459	12.00
578459	12.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00

577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	5.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	10.00
577358	4.00
577358	3.00
577358	3.00
577358	2.00
577358	1.00
577358	2.00
577358	4.00
577358	2.00

577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	4.00
577358	3.00
577358	3.00
577358	9.00
577358	15.00
577358	1.00
577358	5.00
577358	2.00
577358	3.00
577358	1.00
577358	2.00
577358	2.00
577358	10.00
577358	1.00
577358	3.00
577358	1.00
577358	6.00
577358	4.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	5.00
577358	1.00
577358	4.00
577358	10.00

577358	5.00
577358	2.00
577358	1.00
577358	10.00
577358	1.00
577358	7.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00

577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	14.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	2.00
577358	4.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00

577358	1.00
577358	1.00
577358	2.00
577358	4.00
577358	6.00
577358	9.00
577358	3.00
577358	10.00
577358	6.00
577358	1.00
577358	2.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	3.00
577358	8.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	6.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	18.00
577358	4.00
577358	2.00
577358	1.00
577358	1.00

577358	1.00
577358	4.00
577358	3.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	3.00
577358	2.00
577358	3.00
577358	6.00
577358	6.00
577358	12.00
577358	4.00
577358	2.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	4.00
577358	3.00
577358	2.00
577358	4.00
577358	4.00
577358	2.00
577358	30.00
577358	1.00
577358	3.00

577358	4.00
577358	8.00
577358	2.00
577358	4.00
577358	1.00
577358	2.00
577358	11.00
577358	7.00
577358	3.00
577358	3.00
577358	1.00
577358	4.00
577358	1.00
577358	2.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	7.00
577358	2.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	4.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00

577358	3.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	3.00
577358	3.00
577358	1.00
577358	2.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	4.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00

577358	1.00
577358	2.00
577358	2.00
577358	3.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	3.00
577358	279.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	10.00
577358	2.00
577358	1.00
577358	1.00
577358	3.00
577358	5.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00

577358	2.00
577358	2.00
577358	3.00
577358	5.00
577358	1.00
577358	2.00
577358	3.00
577358	2.00
577358	2.00
577358	2.00
577358	3.00
577358	2.00
577358	1.00
577358	3.00
577358	1.00
577358	1.00
577358	4.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	7.00
577358	1.00
577358	7.00
577358	4.00
577358	2.00
577358	6.00
577358	1.00
577358	1.00
577358	19.00
577358	4.00

577358	1.00
577358	1.00
577358	2.00
577358	6.00
577358	1.00
577358	1.00
577358	7.00
577358	3.00
577358	3.00
577358	4.00
577358	8.00
577358	1.00
577358	2.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	14.00
577358	8.00
577358	1.00
577358	1.00
577358	3.00
577358	5.00
577358	1.00
577358	6.00

577358	4.00
577358	3.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	4.00
577358	3.00
577358	3.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	3.00
577358	2.00
577358	8.00
577358	4.00
577358	2.00
577358	3.00
577358	2.00
577358	1.00
577358	3.00
577358	1.00
577358	2.00
577358	1.00
577358	2.00
577358	1.00
577358	4.00
577358	5.00
577358	2.00
577358	1.00
577358	1.00
577358	4.00

577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	4.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	4.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	6.00
577358	5.00
577358	16.00
577358	1.00
577358	7.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	3.00
577358	5.00

577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	3.00
577358	1.00
577358	2.00
577358	2.00
577358	7.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	6.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	2.00
577358	2.00
577358	1.00
577358	1.00
577358	8.00

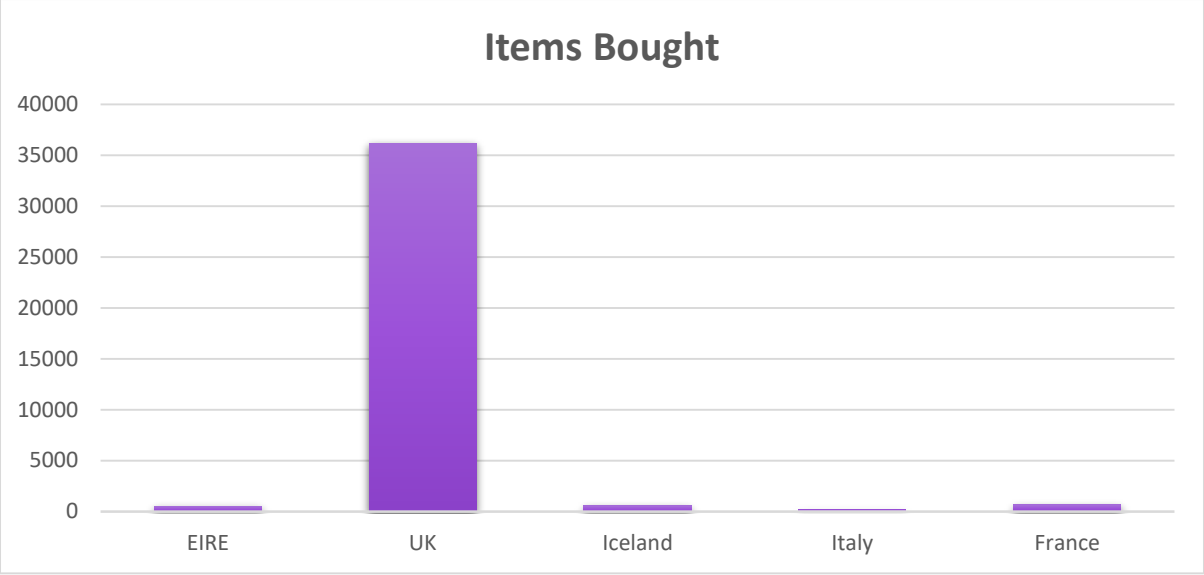
577358	1.00
577358	5.00
577358	2.00
577358	1.00
577358	1.00
577358	2.00
577358	3.00
577358	2.00
577358	1.00
577358	1.00
577358	1.00
577358	1.00
577358	5.00
577358	1.00
577358	2.00
577358	5.00
577358	1.00
577358	4.00
577358	4.00

The above table shows the maximum number of times that the customers have engaged in doing transactions and bought the items from the e-commerce platform.

Most Profitable Segment Customers

Country-customer segment	Items Bought
EIRE	516
UK	36205
Iceland	604
Italy	167
France	674

From the table, it can be stated that the most profitable target segment is UK as 36,205 number of people has bought the products followed by France where there are sells done to 674 number of people. 604 people of Iceland bought the products which is third profitable target segment followed by EIRE accounting to 516 and then Italy (167).



Recommendation to the business based on findings to gain more profit

There is recommendation done to improve upon the marketing campaigns and invest more in doing marketing campaigns in regions where there have been low sales, however, a huge demand has been observed. Due to less awareness among customers, there can be less generation of revenue from those regions by the e-commerce business along with improving the distribution channel efficiency. It helps in reaching to a wide range of customers **(Purnomo, 2023).** In addition to this, there is need of increasing the sales percentage by improving upon the product quality. It will help the business to make sure that more demand is generated in the market (**Lockett, 2018).**

PART 2-

INTRODUCTION-The section has made use of **SPSS analytical** to do statistical analysis. It has been prepared for assisting the school to get an understanding of the male and female students as well as the number of student who has qualifies and not qualified for reduced/free lunch. Data is collected from a total of 2133 school students. There is also mean and media identified of the students who qualifies and does not qualify for reduced/free lunch. The report has produced an evaluation of the student demographics of the school and also discussed the advantages of using SPSS as well as Excel for the data analysis.

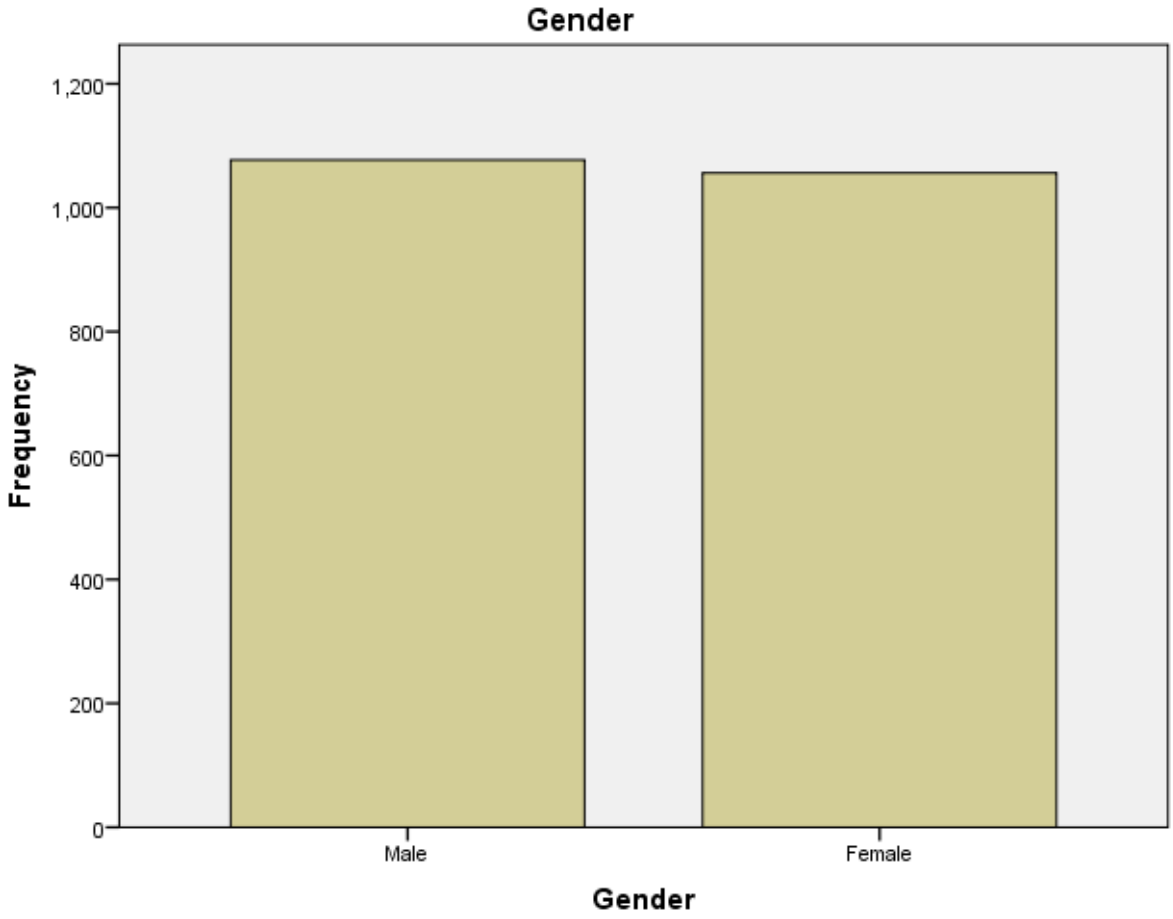
Gender

Statistics		
Gender		
N	Valid	2133
	Missing	0
Mean		1.50
Std. Error of Mean		.011
Median		1.00
Mode		1
Std. Deviation		.500
Variance		.250
Skewness		.020
Std. Error of Skewness		.053
Minimum		1
Maximum		2

Gender

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

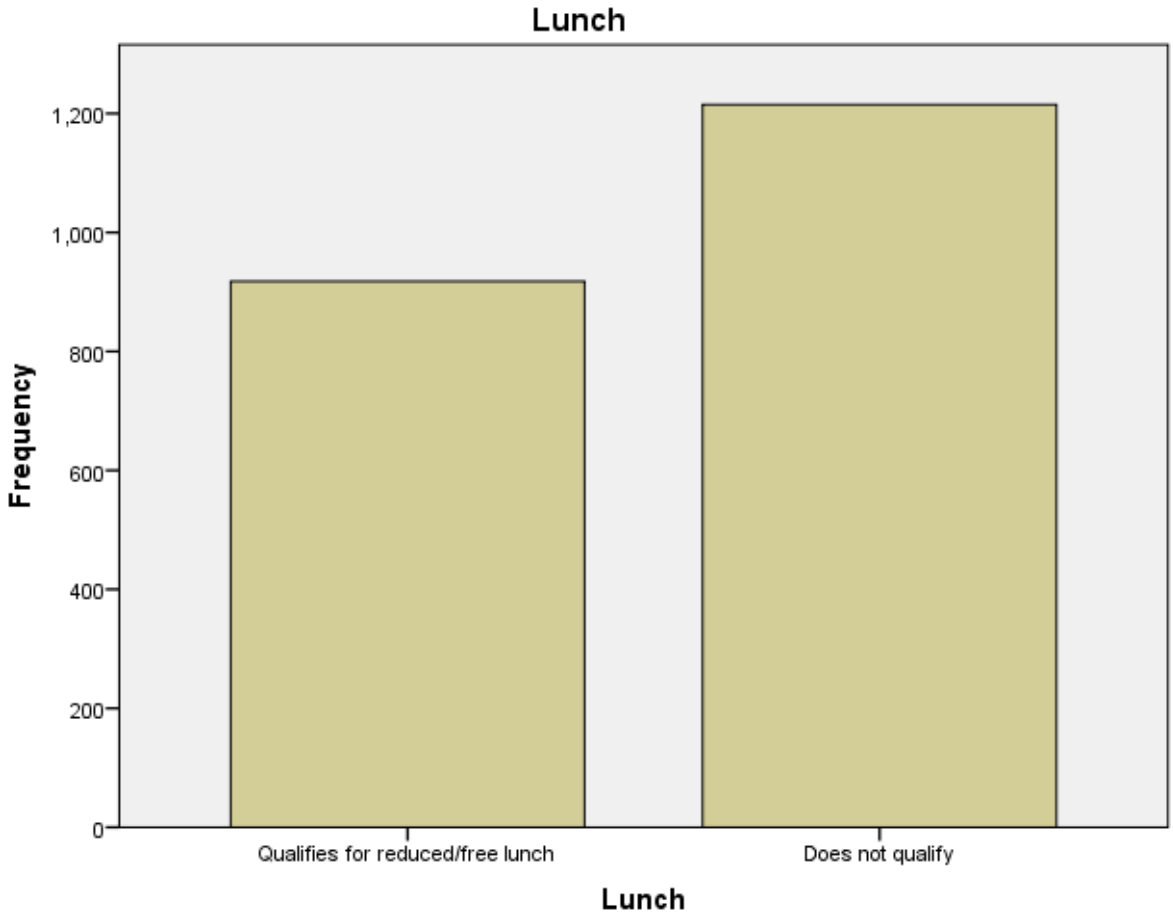
The mean value is 1.50 and median is 0.50 for the gender demographics. It has been identified that out of 2133 school students, majority of them are male and account to 50.5%. Remaining of the students are female and accounts to 49.5%. 1077 of the students are male and 1056 are female.



Number of students qualified and does not qualify for reduced/free lunch

918 of the students have qualified for reduced or free lunch and rest 1215 of them do not qualify for free or reduced lunch. This accounts to 57% of the students studying in the school who has not qualified for lunch for free or at a reduced rate. 43% of the students qualifies for reduced/free lunch.

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



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

The mean value has been identified to be 1.57 and median is 2. Mode is 2 indicating that in majority of the cases the students did not qualify for free or reduced lunch. The Std. Deviation is estimated to be 0.495. The variance value has been identified to be 0.245. The Std. Error of Mean value is estimated to be 0.011 and the variance value has been evaluated to be 0.245.

Statistics		
Lunch		
N	Valid	2133
	Missing	0
Mean		1.57
Std. Error of Mean		.011
Median		2.00
Mode		2
Std. Deviation		.495
Variance		.245
Skewness		.281
Std. Error of Skewness		.053
Minimum		1
Maximum		2

K-Means Clustering

K-Means clustering is a technique used for vector quantization and the objective of doing it is to divide a set of n observations into k clusters. It has been considered as unsupervised machine learning that groups all the unlabelled dataset selected into different clusters. K means clustering assigns data points to one of the K clusters based on the distance from centre of the clusters. It begins by doing random assigning of the cluster centroid. The main goal of clustering is to divide the respondent population or a set of data points into number of groups so that the data points within each group are more comparable. It is considered as grouping of the things based on how similar or different they are to one another.

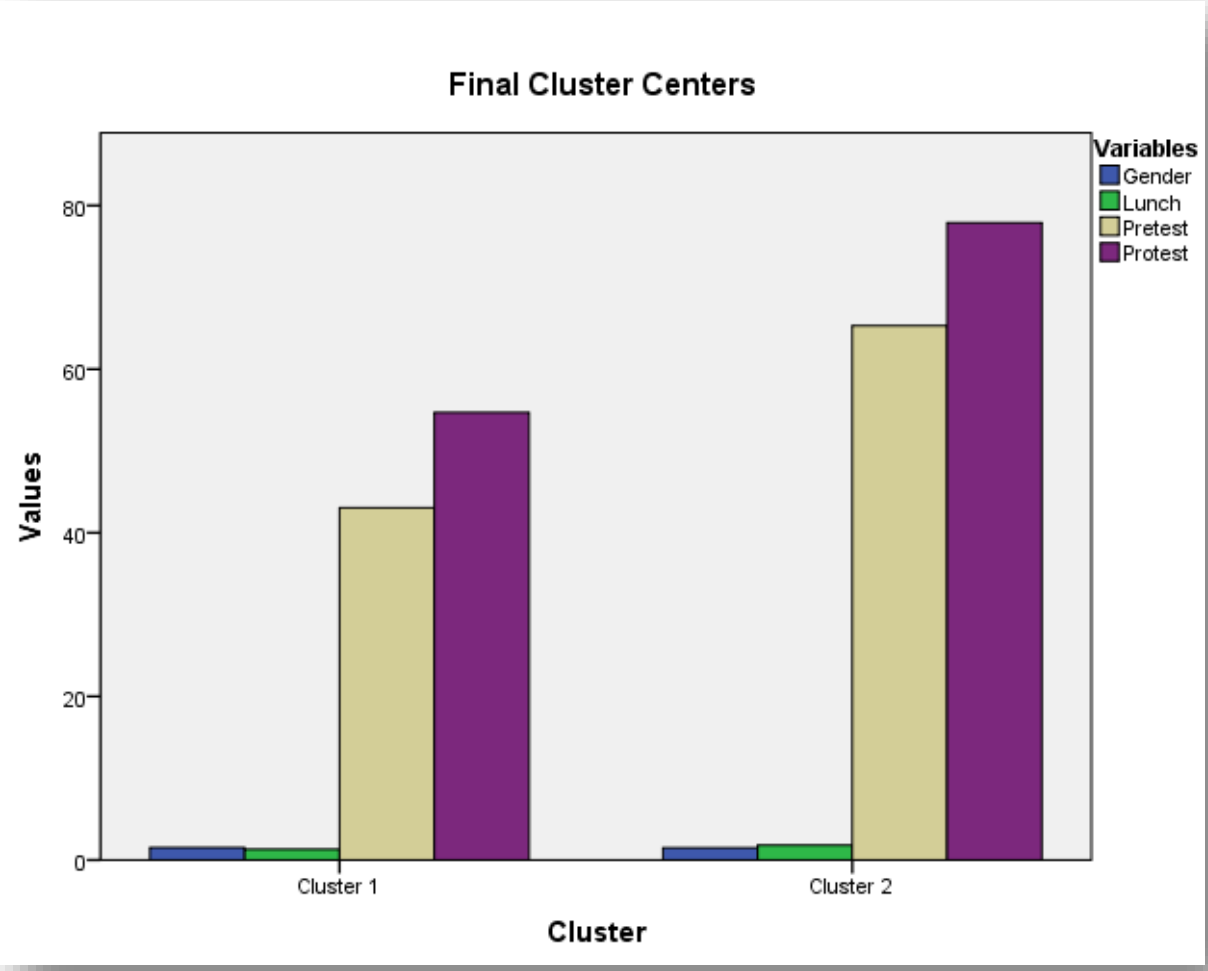
Initial Cluster Centres		
	Cluster	
	1	2
Gender	2	1
Lunch	1	2
Pretest	23	93
Protest	35	100

Iteration History		
Iteration	Change in Cluster Centres	
	1	2
1	30.197	33.610
2	.885	.791
3	.546	.476
4	.276	.237
5	.227	.200
6	.150	.128
7	.000	.000

a. Convergence achieved due to no or small change in cluster centres. The maximum absolute coordinate change for any centre is .000. The current iteration is 7. The minimum distance between initial canters is 95.535.

Final Cluster canters		
	Cluster	
	1	2
Gender	2	1
Lunch	1	2
Pretest	43	65
Protest	55	78

The table shows the final cluster centres and change in cluster centre. In the pretest and free/reduced lunch cluster there are 43 units as compared to 55 for protest. For male gender and not having reduced/free lunch, the pretest value is 65 as compared to 78 for protest.



Advantages and Disadvantages of using SPSS and Excel using relevant examples and based on different criteria-

Advantages of using SPSS	**Disadvantages of using SPSS**
SPSS analytical tool has been used in interpreting the data and was used in market research. The advantage of using SPSS is that it provides a user-friendly interface and helps in doing the statistical analysis without any programming language **(Pallant, 2020).** Through help of graphical user interface, the data analysis can be done **(Stehlik-Barry and Babinec, 2017)** There is also wide use of SPSS analytical tool and it can be used for doing regression analysis, inferential statistics, chi-square testing, logistic regression as well as contingency table analysis **(Kafle, 2019).** It can also be done to integrate with other programs such as Access, Excel, SQL databases (**McCormick and Salcedo, 2017).**	Disadvantage of using SPSS is that it is a commercial software and is expensive to install. There is also restriction in customizability and are sometimes constrained.
Advantages of using EXCEL	**Disadvantages of using Excel**
Excel has been used frequently for doing data analysis and is relatively easy to make use of. There is option of provider of filtering data that provides advantage of retrieving those data only that are required **(Hollman., 2017).** There are various Excel functions such as COUNTIF (), COUNT () with which data analysis becomes an easier process **(Sousa, 2021).** The MS Excel tool us widely available and is already installed in most of the computers **(Ferrari and Russo, 2017).** There is more familiarity with the MS Excel tool and can be used for extensive training as well as data analysis **(Salkind and Frey, 2021).** Through Excel, there can be quick as well as ad-hoc analysis of data along with providing different chart types and formatting options allowing to have clear visibility of the data, graphs **(Guerrero, Guerrero and Rauscher, 2019).**	There are certain limitations in MS Excel in managing large data sets. At times, the MS Excel software can become slow and inefficient for working with extensive data. There is also disadvantage of the Excel files getting prone to errors and inconsistencies. There are also limitations in having the advanced analysis techniques done in Excel tool. Complicated modelling cannot be done with help of MS Excel. It has been identified that Excel requires manual input and manipulation issues happens with data. It becomes a time-consuming process and also does not provide an audit trial.

Overall Summary

From the study done, it can be summarized that United Kingdom has been the region which has the maximum sales from the e-commerce business and the sales percentage has been erratic. There is need of doing marketing campaign to engage the customers and make them aware in buying the products. Moreover, there has been K-means clustering done along with understanding of the fact that more students have not qualified for free or reduced meals.

Reference List

- Ferrari, A. and Russo, M., 2017. *Analyzing data with power BI and power pivot for excel*. microsoft press.
- Guerrero, H., Guerrero, R. and Rauscher, 2019. *Excel data analysis*. Springer International Publishing.
- Hollman, C., Paulden, M., Pechlivanoglou, P. and McCabe, C., 2017. A comparison of four software programs for implementing decision analytic cost-effectiveness models. *Pharmacoeconomics*, *35*(8), pp.817-830.
- Kafle, S.C., 2019. Correlation and regression analysis using SPSS. *Management, Technology & Social Sciences*, *126*.
- Lockett, A., 2018. *Online marketing strategies for increasing sales revenues of small retail businesses*
- McCormick, K. and Salcedo, J., 2017. *SPSS statistics for data analysis and visualization*. John Wiley & Sons.
- Pallant, J., 2020. *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
- Purnomo, Y.J., 2023. Digital marketing strategy to increase sales conversion on e-commerce platforms. *Journal of Contemporary Administration and Management (ADMAN)*, *1*(2), pp.54-62.
- Salkind, N.J. and Frey, B.B., 2021. *Statistics for people who (think they) hate statistics: Using Microsoft Excel*. Sage publications.
- Sousa, R., Miranda, R., Moreira, A., Alves, C., Lori, N. and Machado, J., 2021. Software tools for conducting real-time information processing and visualization in industry: An up-to-date review. *Applied Sciences*, *11*(11), p.4800.
- Stehlik-Barry, K. and Babinec, A.J., 2017. *Data analysis with IBM SPSS statistics*. Packt Publishing Ltd.