

DATA ANALYSIS & VISUALISATION
AMITAVE BANIK
DUMITRITA CIUDIN (2030848)

Executive Summary

Concerning the provided case study, it is identified that different types of excel tools has been utilized in the paper to determine different product category and the declining margin of revenue and profit. It is further evaluated that with the IF function it is evaluated that by swelling 38 commodities the Superstores sales margin would decline by **8361 margins**. The overall research has represented that 47.5% of total students mostly like vanilla flavor in ice cream. In addition, frequencies show that 45.5% of total students are male and 54.5% of total students are female. The mean value stands at 1.82 and the median value stands at 2.00. Through the recognition of the mean value, it is stated by this study that the center point of this data stands at 1.82. Furthermore, the median value when a set of data is arranged from least to greatest shows the middle value of the dataset which stands at 2.00. This research has also discussed both advantages and disadvantages of *Excel and SPSS*.

Table of Contents

Part 1:	4
Introduction.....	4
Main body	4
Use of excel for pre-processing, analyzing, and visualizing data.....	5
The sale of products concerning the category.....	6
Concerning the units the sales margin of the commodity	7
Pre-profit from the sale of commodity.....	9
Pre-processing data	11
Performance concerning the product	11
Conclusion	12
Part 2	12
2.1 Identification of number of students who like the vanilla flavor of ice-cream.....	12
2.2 Recognition of number of students as male and female	13
2.3 Mean and Median value of participants.....	15
2.4 Clustering using K-Means	15
2.5 Explanation regarding data mining and text mining.....	16
2.6 Advantages and disadvantages of Excel and SPSS.....	17
Reference list	19

Part 1:

Introduction

Concerning the provided case study, it is determined that data visualization is also referred to as business intelligence and it is also referred to as an important factor in business intelligence as it allows business organizations to avail a large amount of business data. Apart from that it also helps to gain relevant and accurate data. It is further articulated that data visualization further plays a significant role to recognize the different patterns of development. Further, it is perceived that this method helps to create a huge amount of data related to business. Along with the extraction of reliable data, it is also important to appoint a skillful employee so that the data can be handled efficiently and effectively and productivity can also be enhanced easily. It is further identified that data handling is an authenticated and reliable source to ensure the integrity of the business with the help of research and mitigate the identified data handling issue.

Main body

It is further articulated that the fundamental part of the paper mainly creates an emphasis on different factors and elements that create an impact on the revenue margin of the business and its production level. The reason behind the negative impact of the decreasing business revenue is the operating of business activities on a large scale. Moreover, concerning the provided scenario it is also identified that the sales and profit margin have declined to a great extent over the years. The categorization of products into different categories of products further determined the reason behind the decreasing margin of profit. Business growth mainly gets evaluated by the help of future results projection and present financial condition (Nariswari and Nugraha, 2020, p.2).

Further, the provided data is used to identify the interest of customers concerning the sale of different categories of commodities. The identification of the interests of customers and their tastes and preferences has further been identified by an implication of data visualization techniques. On the other hand, by the utilization of data interpretation or analysis techniques the different category commodities are identified. It is further perceived that the utilization or implication of different Excel functions has created a huge impact on the evaluation and

estimation of products. Excel is an application program that helps in creating various advanced software helpful in creation of database (Qarkaxhija, 2018, p.2).

Use of excel for pre-processing, analyzing, and visualizing data

Microsoft Excel has played a significant role in preprocessing as well as handling data. V lookup mainly is used for checking the codes according to the worksheet (Meng *et al.* 2020, p.2). This data helps to identify that Excel is one of the important tools by which a company can pre-process, analyze, and visualize data in a significant way. There are a few techniques included in Excel that help to pre-process data such as IF function, IFERROR, COUNTA, COUNTBLANK, and many others. Through the IF function, an individual has become able to check the condition as well as return a particular value accordingly. In addition, the IFERROR function helps with error handling by which common errors can be found. Through this analysis, it is identified that the use of different functions of excel is helpful in pre-processing data.

According to the research of Wiyono *et al.* (2021, p.1), Microsoft Excel is one of the popular tools that helps in understanding the relation between variables. In addition, this research has also stated that it is software for managing data that allows an individual for easily importing, exploring, cleaning, analyzing, and visualising data. There are several tools that are used to analyze data in an effective way such as What-if analysis, Data model, Power Pivot, and many others.

Microsoft Excel is an important tool that helps in the visualization of data. This analysis helps to identify that the use of Excel can be beneficial for an individual in order to visualize data. There are a few steps included in this software which are used for the visualization of data such as the creation of organized spreadsheets through labeling, finalizing dataset, highlighting data, selecting of “insert” option, and selecting the type of charts or graphs as it will be helpful in visualizing data.

Based on the provided description and the overall observation it is identified that the utilization of different functions of excel is also determined as an innovative technique of data visualization. With the help of such functions, relevant and authenticated information can be driven out easily. The functions that are mainly used to identify the categorized

commodities are the PIVOT table and the function of IF. The Pivot table is further determined to be a bilateral way through which the large amount of data set can be characterized easily in a very short span. By the help of pivot table to text generation is done and then based on notations overview of the model is given (Ma *et al.* 2019, p.2). In order to find out the detailed report, the pivot table function of excel is implemented. Therefore, it can be said that the pivot table function is mainly designed to evaluate the large amount of data set in different user-friendly ways. On the other hand, the IF function is further identified as an essential or crucial excel function that further helps to make a logical difference between the expected and calculated values. It is further advocated that it determines two results: the first result involves necessary data in context to the comparison and determines whether the comparison is true or not. However, the second result helps to evaluate whether the compared data is false or not.

The sale of products concerning the category

It is articulated that the different commodities or goods of the firm are identified by an estimation of different categories of the commodity. It is further identified that the commodities comprise the demand and supply of technology. According to further estimation, it is perceived that the company has sold around 14,920 commodities previously. On the other hand, the overall evaluation of the sales margin further involves different commodities such as office suppliers, furniture and others. The comparison between the sales margin of furniture and office suppliers indicates that the sales amount of furniture is more as compared to that of office suppliers.

Row Labels	Sum of Sales
Furniture	5178590.542
Office Supplies	3752762.1
Technology	5984248.182
Grand Total	14915600.82

Figure 1: computation of sales margin concerning different category of commodity

(Source: Self-developed)

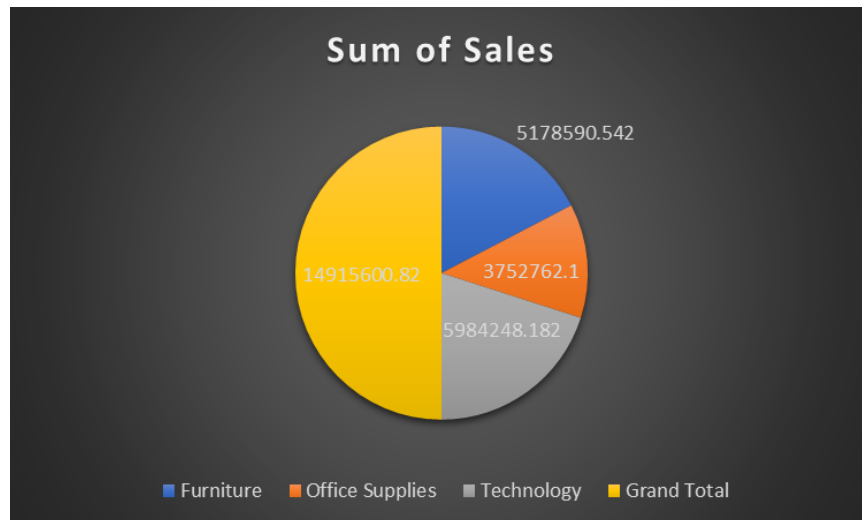
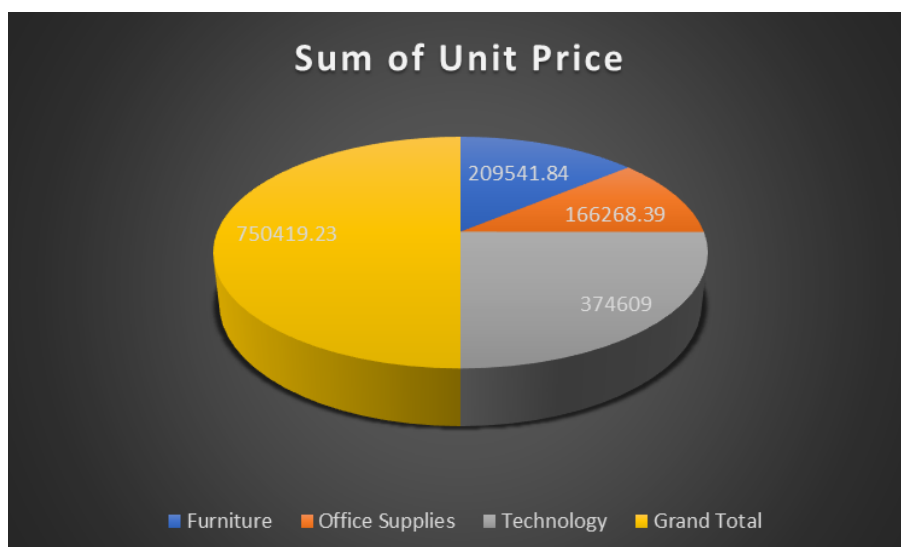


Figure 2: Graphical representation of the category of commodity

(Source: Self-developed)

Concerning the provided data and the image, it is identified that the sales margin is highly affected due to the different requirements of the commodity. Therefore, it can be said that the technological factor is the common factor in the fluctuation of the production margin and the operational cost.

Concerning the units the sales margin of the commodity



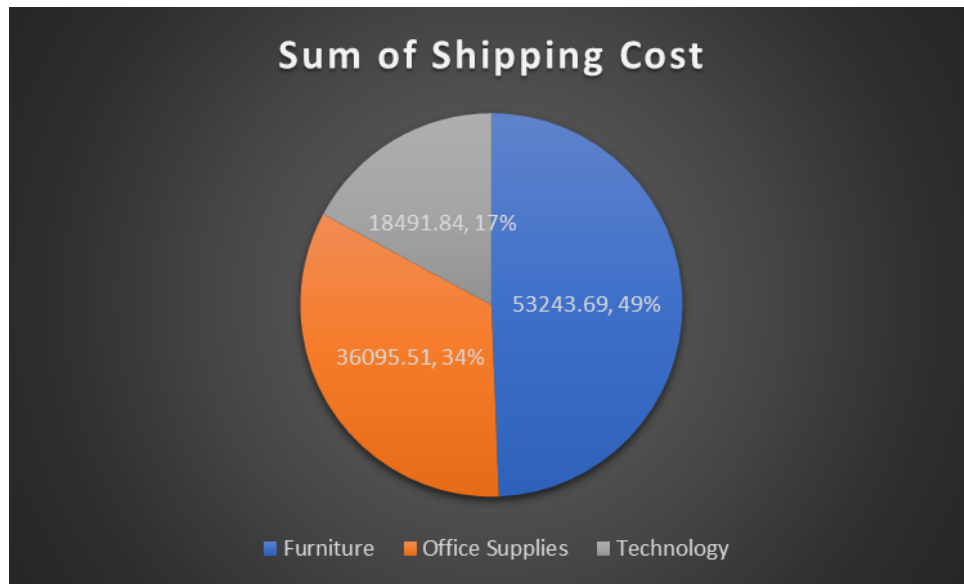


Figure 3: Computation of unit price of the commodity

(Source: Self-developed)

According to the provided data, it is identified that the above-provided figure helps to evaluate the unit price of the products by implementing the data visualization technique. Moreover, by the implication of the price evaluation technique, the enhanced price of the commodity is identified to be 750419. Moreover, it is articulated that the excessive price of the commodity is further identified by comparing the prices and range of two different commodities. Based on the provided results and the overall evaluation it is determined that the authenticated data results can be easily driven out by applying smart technological or data visualization tools. Apart from that, it is also notified that the data intelligence techniques create an impact on the cost of shipping and the price of the commodity. By the help of data analysis imaging tool power as well as reshaping the workflow could be done (Shi *et al.* 2020, p.1).

PIVOT TABLE			
Row Labels ▼	Sum of Order Quantity	Sum of Unit Price	
Furniture	44324	209541.84	
Office Supplies	118281	166268.39	
Technology	52175	374609	
Grand Total	214780	750419.23	
PIVOT TABLE			
Row Labels ▼	Sum of Shipping Cost	Sum of Sales	Sum of Profit
Furniture	53243.69	5178590.542	117433.03
Office Supplies	36095.51	3752762.1	518021.43
Technology	18491.84	5984248.182	886313.52
Grand Total	107831.04	14915600.82	1521767.98

Figure 4: Estimation of the PIVOT table

(Source: Self-developed)

Concerning the provided data set and the previous illustration, it is articulated that the PIVOT table tool of excel is implemented to visualize the data and to drive out the authenticated and reliable information. It is further perceived that the tools can also be utilized to summarize large data sets. Through the PIVOT table tool, an unanticipated question can be answered easily. It is further articulated according to the provided image that office supplies data are huge as compared to that of other categorized commodities. Therefore, the PIVOT table function is implemented to determine the cost of shipping, the overall margin of revenue and the profit margin of the commodity.

Pre-profit from the sale of commodity

The major goal of every organization is to maximize profit margin as through it the firm can easily make essential and data manipulation decisions. By the help of Markov decision process complexity in the action and state spaces could be recovered. (Wang *et al.* 2018, p.1). Further, it is perceived that by implementing the IF function and different big data techniques the profit concerning different commodities have been evaluated.

Row Labels	Sum of Profit
Furniture	117433.03
Office Supplies	518021.43
Technology	886313.52
Grand Total	1521767.98

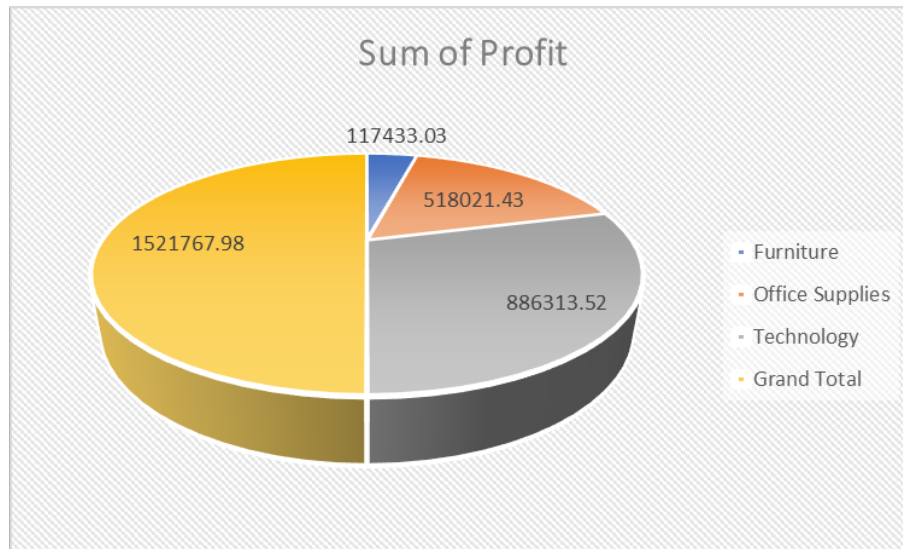


Figure 5: Sales of commodities based on the profit margin

(Source: Self-developed)

According to the provided image and the graphical representation it is perceived that the data concerning the commodity helps the business to enhance its profit margin. The image further indicates that the profit margin has been identified to be £ 1,521,767. It is further observed that by developing its profit-enhancing strategy the company can easily enhance its sales margin. It is further opined that 60% of the sales margin indicates low growth of income.

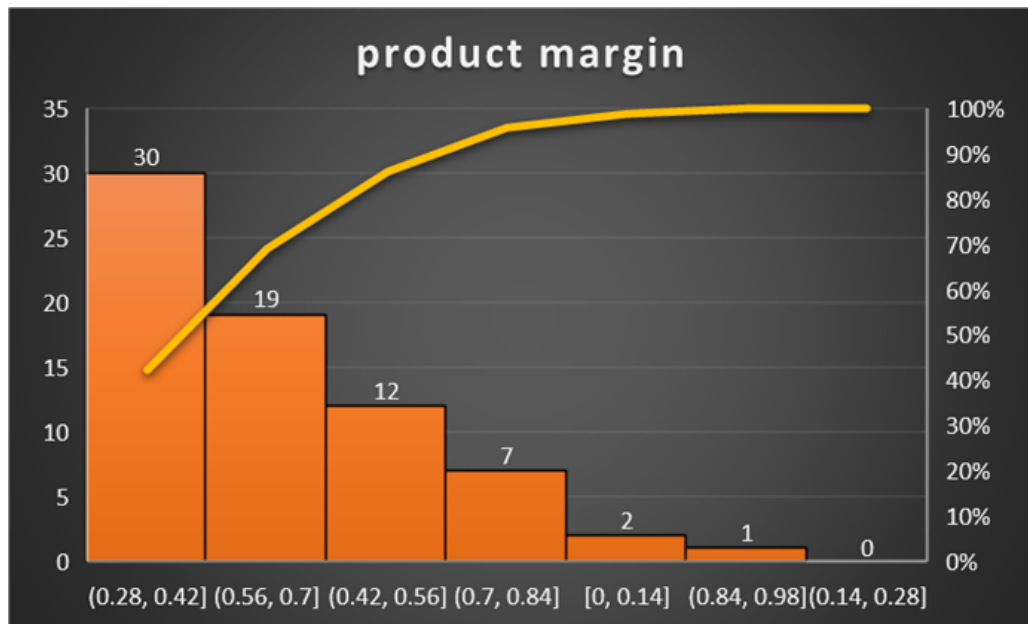


Figure 6: Sales margin concerning the production

(Source: Self-developed)

According to the provided image, it is opined that the data is represented to identify the production margin and different products range. It is further observed that the firm emphasized the profit maximization strategy and commodities categorization. Moreover, the implementation of IF obliges the suppliers to evaluate the profit margin and the demand and supply of the commodity.

Pre-processing data

It is identified as an essential tool that obliges the company to enhance its manufacturing activity. Apart from that it also helps in business transformation and appropriate utilization of data sets. According to the provided description by appropriate utilization of excel functions the performance of the stock in the market can be evaluated easily.

Performance concerning the product

It is also identified that data procession helps to maintain the overall business data and to evaluate the market performance of the organization. On the other hand, it is perceived that PIVOT Table is utilized to identify the categories of the commodity. On the other hand, the IF

function is implemented to analyze the performance of the company. The performance of production reflects the profits have declined by 8361.

Conclusion

Concerning the Superstore data set, it is determined that the company deals with the sales activity of different commodities. The different commodities which the activity of the company deals is technology, office supplies, furniture and different other commodities. Moreover, the utilization of the Pivot table helps to evaluate the different types of commodities. On the other hand, in the case of the IF function, it is articulated that the tool is implemented to evaluate and determine the diminishing sales and profit margin. It is further observed that the sales of 38 commodities reflect that the profit margin has declined by 8361 products. Moreover, it is articulated that with the help of different data visualization techniques the different costs of the products and the cost of shipping can be evaluated easily.

Part 2

2.1 Identification of number of students who like the vanilla flavor of ice-cream

ice_cream					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Vanilla	95	47.5	47.5	47.5
	Chocolate	47	23.5	23.5	71.0
	Strawberry	58	29.0	29.0	100.0
	Total	200	100.0	100.0	

Figure 7: Ice cream flavors

(Source: Self-created)

As per the information provided in the case study, it is observed that there is a total of 200 students who likes ice cream. This research is about the identification of a number of students who like the vanilla flavor of ice cream. In this regard, this research has conducted frequencies of flavors of ice cream to identify the number of students who mostly like the vanilla flavor of ice cream.

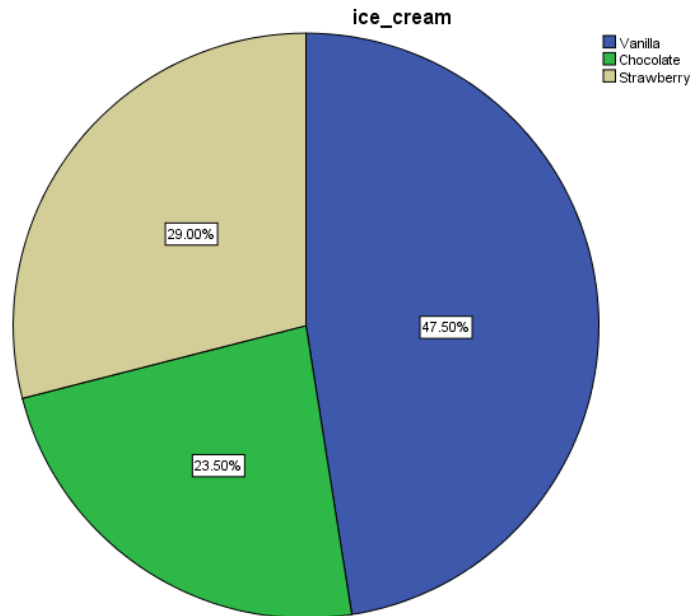


Figure 8: Different flavors of ice cream

(Source: Self-created)

This figure has represented that 47.5% of total students mostly like vanilla flavor in ice cream. In addition to this, it is also identified that 29% of total students like strawberry flavor and 23.50% of total students like chocolate flavor in ice cream. As it is recognized that there are a total of 200 students, therefore, the determination of frequencies shows that 95 students like vanilla flavor, 47 students like chocolate flavor, and 58 students like chocolate flavor. Cumulative per cent of ice cream flavor stands at 47.5, 71.0, and 100.0 in the context of Vanilla, Chocolate, and Strawberry respectively. This analysis helps to identify the maximum number of students who like vanilla flavor in ice cream among chocolate and strawberry.

2.2 Recognition of number of students as male and female

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	91	45.5	45.5	45.5
	Female	109	54.5	54.5	100.0
	Total	200	100.0	100.0	

Figure 9: Gender

(Source: Self-created)

This research covers the number of students as males and females among total students. In this way, it can be identified whether there are a maximum number of females or males. Due to this reason, this research has been conducted with frequencies regarding the gender of total participants.

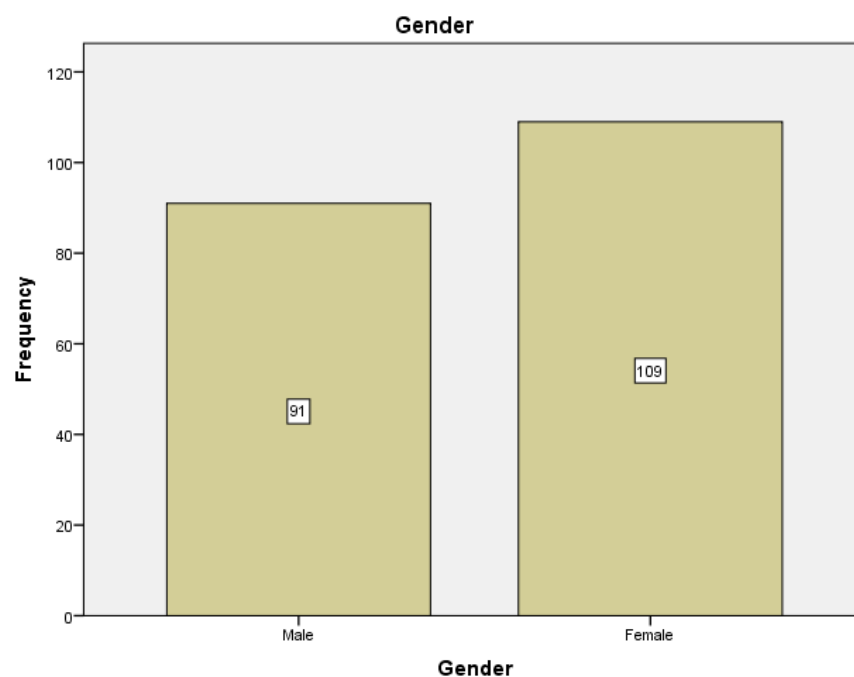


Figure 10: Gender

(Source: Self-created)

This figure has shown that 45.5% of total students are male and 54.5% of total students are female. In addition to this, it is also identified that 91 males and 109 females are there among the total number of students. This research helps to recognize that there is a maximum number of students who are female. Therefore, it is stated that female students are more interested in this process in comparison with male students. In addition, the cumulative

percentage stands at 45.5 for male and 100.0 for female students. However, there is no missing value in the provided dataset.

2.3 Mean and Median value of participants

As per the research of Chakrabarty (2021, p.17324), four formulations of the central tendency involves AM, HM and GM. Central tendency is referred to as the measurement of statistics that helps to identify the single value as representative of the entire distribution. This analysis helps to identify that the use of central tendency can be helpful in identifying both mean and median values of the students. This is because central tendency has included four measures such as mean, median, mode, and midrange.

This research has identified that there is a valid value that stands at 200 which represented that total of 200 students have participated in this process. In addition, the missing value is 0 which shows the organization of data in an effective way. The mean value stands at 1.82 and the median value stands at 2.00. Through the recognition of mean value, it is stated by this study that the centre point of this data stands at 1.82. Furthermore, the median value when a set of data is arranged from least to greatest shows the middle value of the dataset which stands at 2.00.

2.4 Clustering using K-Means

According to the research of Xu and Lange, (2019, p.8), K- Means requires initial power but don't need careful tuning. K-means is referred to as the statistical method that helps to average the data which represents the identification of the centroid. On the basis of this identification, it is stated that this research has become able to identify the average of data regarding ice cream flavors.

Initial Cluster Centers		
	Cluster	
	1	2
female	0	1
ice_cream	3	1

Figure 11: K-means

(Source: Self-created)

This research has considered two different variables for the conduction of K-means such as females and ice cream. It is observed from the determination of K-means that female stands at 0 in cluster 1 and 1 in cluster 2. On the other hand, ice cream stands at 3 in cluster 1 and 1 in cluster 2. In addition, through this research, it is also identified that the minimum distance between initial centres stands at 2.236. As K is greater than one, therefore, it is stated that the value of K will be high as well. This analysis represents that there are more students than flavors of ice cream and therefore a greater concentration of equilibrium among the students.

2.5 Explanation regarding data mining and text mining

Data mining is considered the process to sort data through a large set of data for the purpose of identifying patterns as well as relationships that can be helpful in solving issues of a business through the analysis of data. Data mining can be used for dealing with transaction graph and data (Wang *et al.* 2020, p.1). This analysis helps to recognize that the use of data mining can be helpful for this organization to identify issues of this business by the analysis of data provided in this research. Moreover, it is observed that there are different types of tools and techniques regarding data mining that provide support to the organization for predicting future trends and help to make decisions regarding business with more information.

Text mining represents the process in which an individual can transform text that is in an unstructured format for the purpose of identifying meaningful patterns as well as new insights. This analysis helps to identify that the use of text mining can be beneficial for this company to transform all the text into a meaningful pattern that can lead to effective processing of text. In this way, this company will be able to identify the number of students who like different flavors of ice cream.

2.6 Advantages and disadvantages of Excel and SPSS

Excel and SPSS both are effective tools and techniques that help to store, analyze, and interpret data in an effective way as per the viewpoint of Weathers and Aragón (2019). However, both these tools have advantages and disadvantages which are presented below:

Tools and Techniques	Advantages	Disadvantages
Excel	● It is one of the best ways for storing data ● An individual can performing calculations with the use of this tool. ● It helps to transform and clean data effectively. ● It is an easy method to visualize data along with charts.	● It may lead to errors in calculation. ● Excel has included different functional tools which require different steps to calculate represent increased consumption of time ● As Excel has included different tools which require a price, therefore, it is stated that Excel is a cost-effective
SPSS	● This tool offers reliable outcomes to the users ● It is one of the effective management of data. ● This type of software can be used for both qualitative and quantitative data.	● It is the most expensive software, therefore, it is stated that students might not be able to use it. ● It is very similar to Excel ● It has included limited functionality.

Table 1: Advantages and disadvantages of excel and SPSS

(Source: Self-developed)

Reference list

- Chakrabarty, D., 2021. Arithmetic-harmonic mean: a measure of central tendency of ratio-type data. *International Journal of Advanced Research in Science, Engineering and Technology*, 8(5), pp.17324-17333.
- Ma, S., Yang, P., Liu, T., Li, P., Zhou, J. and Sun, X., 2019. Key fact as pivot: A two-stage model for low resource table-to-text generation. *arXiv preprint arXiv:1908.03067*.
- Meng, X., Cui, Y. and Cai, X., 2020, April. Check ID Cards with an Excel Function. In *International Conference on Education, Economics and Information Management (ICEEIM 2019)* (pp. 251-253). Atlantis Press.
- Nariswari, T.N. and Nugraha, N.M., 2020. Profit growth: impact of net profit margin, gross profit margin and total assests turnover. *International Journal of Finance & Banking Studies* (2147-4486), 9(4), pp.87-96.
- Qarkaxhija, J., 2018. Database creation through if, vlookup, sum and sumif functions of Microsoft Excel. *Prizren Social Science Journal*, 2(2), pp.27-34.
- Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y. and Shen, D., 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE reviews in biomedical engineering*, 14, pp.4-15.
- Wang, S., Bi, S. and Zhang, Y.J.A., 2018. Demand response management for profit maximizing energy loads in real-time electricity market. *IEEE Transactions on Power Systems*, 33(6), pp.6387-6396.
- Wang, S., Cao, J. and Yu, P., 2020. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*.
- Wiyono, W., Christian, P.Y. and Abidin, A., 2021. SIMPLE LINIER CORRELATION AND REGRESSION ANALYSIS USING MICROSOFT EXCEL AND ITS APPLICATION IN MARKETING MANAGEMENT. *Akselerator: Jurnal Sains Terapan dan Teknologi*, 2(1), pp.29-40.
- Xu, J. and Lange, K., 2019, May. Power k-means clustering. In *International conference on machine learning* (pp. 6921-6931). PMLR.