

**Awarding Body:** University of Wales Trinity Saint David

**Programme Name:** BA Business Management

**Module Name:** Data Handling & Business Intelligence (LCBB5000)

**Assignment Title:** Assessment 2

**Student Name and Number:** Anita Varga / 2012895

**Tutor Name:** Joy Botu

**Word Count:** 3000

# **TABLE OF CONTENTS**

## **PART 1**

### **1.1; USE AND BENEFITS OF EXCEL**

1.1.1 Table 1. Statistics for sales and profit

### **1.2; HISTOGRAMS, BOXPLOTS, ANALYSIS**

1.2.1. Figure 1. The sales histogram

1.2.2. Figure 2. The boxplot outliers

1.2.3. Figure 3. The histogram of the profits

1.2.4. Figure 4. The boxplot for the profits

1.2.5. Table 2: The pivot analysis of the superstore sales

1.2.6. Figure 5. The graphical presentation of the changes in the sales from the superstore

1.2.7. Table 3: The pivot analysis for the profit

1.2.8. Figure 6. The changes in profits from the superstore

### **1.3; CONCLUSION ABOUT EXCEL**

## **PART 2**

2.1. How Many Students Like Vanilla Flavour of Ice-Cream?

2.2. How Many Students are Male and Female?

2.3. What Are the Mean and Median of Participants Who Likes Chocolate and Strawberry Flavour Ice-Cream?

2.4. Using the Dataset Ice\_cream\_csv Provided in Conjunction with SPSS, Give a Specific Example of Clustering (ex. K-Means)

2.5. The Most Common Data Mining and Text Mining Methods that Can Be Used in Business

2.6. Advantages and Disadvantages of SPSS and Excel

## **REFERENCES**

## **PART 1**

### **1.1; USE AND BENEFITS OF EXCEL**

Microsoft Excel is one of the most widely used programmes for storing, organising and analysing data. It is a powerful tool, which can help one assess large datasets and uncover insights to make better decisions. In this study, the researcher will explore the best practices for working with Excel data for Superstore Sales. The activities would include pre-processing, analysing and visualising data to determine the decline in sales and profits. The study will also discuss the importance of pivot charts, analysis of sales and profits, and the use of excel forecasting for sales and profits.

Excel is a comprehensive programme that can be used for data manipulation, analysis and visualisation. Data analysts, business intelligence professionals and marketers often use it to work with large datasets. Albright and Winston (2014) suggest that Excel can quickly analyse and visualise data, which can be used to identify trends, discover correlations and relationships, and make predictions. The benefits of using Excel for data analysis include cost saving, speed, flexibility, visualisation and automation. Excel is a cost-effective programme; it is free if one uses a laptop or a PC. Excel is very fast and can handle large datasets quickly and easily. Excel can be customised to fit their specific needs (Bernhardt et al., 2018). One can use formulas, macros and pivot tables quickly and accurately analyse data. It can be used to create charts and graphs to visualise data. It can be automated quickly and accurately analyse data.

When working with Excel, it was important to follow best practices to ensure accurate data analysis and visualisation. The researcher made sure to back up their work. It is advisable always to save different versions of the work to ensure that one does not lose any data. The researcher also made sure to use the correct data types. Excel has built-in functions allowing one to convert data types to ensure accuracy. There was the use of formulas to

automate tasks. Excel formulas were used quickly and accurately to analyse data. The researcher used pivot tables to analyse large datasets quickly. Pivot tables were an effective way to analyse large datasets quickly. There was the use of data validation to ensure accuracy. Data validation was used to ensure that data was entered correctly and accurately. The researcher also used the data filters to find specific data quickly (Alexander, Decker, and Wehbe, 2014). Data filters were used to quickly and easily find specific data in a large dataset. Charts and graphs are a great way to visualise data quickly and accurately.

When working with large superstore sales datasets, it was important to follow best practices to ensure accuracy and speed. The researcher used data validation to ensure accuracy. Data validation was used to ensure that data was entered correctly and accurately. Used data filters to find specific data quickly (Alexander et al., 2014). There was the use of macros to automate repetitive tasks. Macros were used to automate repetitive tasks, such as data entry, analysis, and formatting. Excel was used to quickly analyse and visualise data, identify trends, discover correlations and relationships, and make predictions.

The superstore sales data required pre-processing of data. Pre-processing data in Excel is a crucial step in the data analysis process. Pre-processing is organising and cleaning up data before it is analysed. Pre-processing helped the researcher optimise the data for analysis and make it easier to understand (Bernhardt et al., 2018). When pre-processing data for sales and profits, it was important to ensure that it was accurate and complete. It can be done by removing duplicate or inaccurate data and ensuring it is properly formatted. It is also important to ensure that the data is up to date. In addition, pre-processing data in Excel for sales and profits can also include adding additional columns to the dataset. It can help organise the data and make it easier to analyse. For example, the researcher added columns for customer information, product information and sales and profit figures.

Descriptive statistics were used to describe data and summarise its characteristics. They provided a quick and easy way to understand the data. Using descriptive statistics in Excel helped to quickly and accurately analyse data. When analysing data in Excel, the researcher used descriptive statistics to summarise data, identify trends, and make predictions. Excel's most commonly used descriptive statistics included measures of central tendency (mean, median and mode) and measures of variability (range, standard deviation, and variance). Using descriptive statistics in Excel helped the researcher quickly identify data patterns, trends and relationships. It also helped the researcher make informed decisions and predictions.

**1.1.1; Table 1:** Descriptive statistics for sales and profit

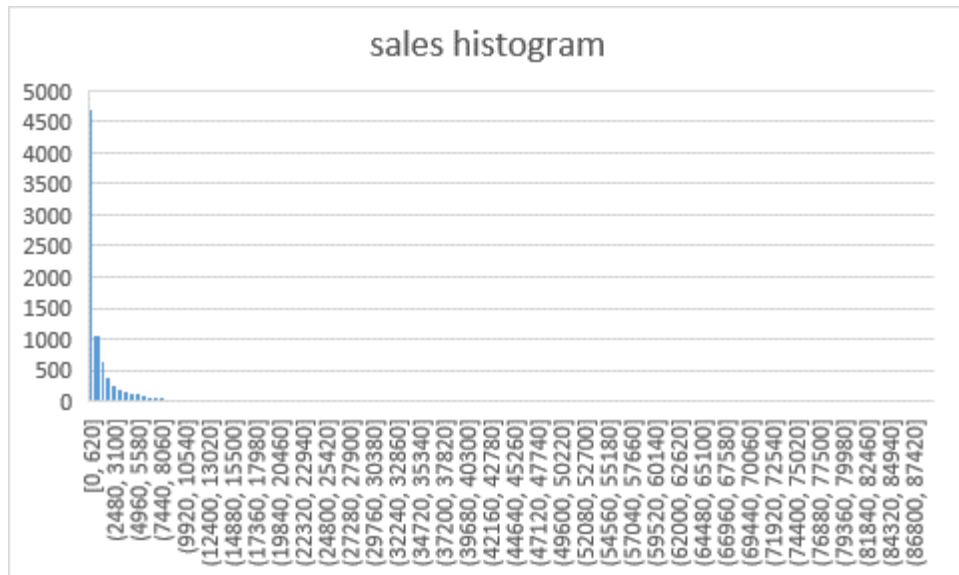
|                    | <i>Sales</i> | <i>Profit</i> |
|--------------------|--------------|---------------|
| Mean               | 1775.88      | 181.18        |
| Standard Error     | 39.12        | 13.06         |
| Median             | 449.42       | -1.50         |
| Mode               | 75.19        | -969.05       |
| Standard Deviation | 3585.05      | 1196.65       |
| Sample Variance    | 12852587.27  | 1431979.29    |

Table 1 above shows that the mean for sales is 1775.88 units, and the profit means 181.18. The standard deviation for both variables is higher than the mean, which means that there is a very big range of the minimum and maximum values for both variables.

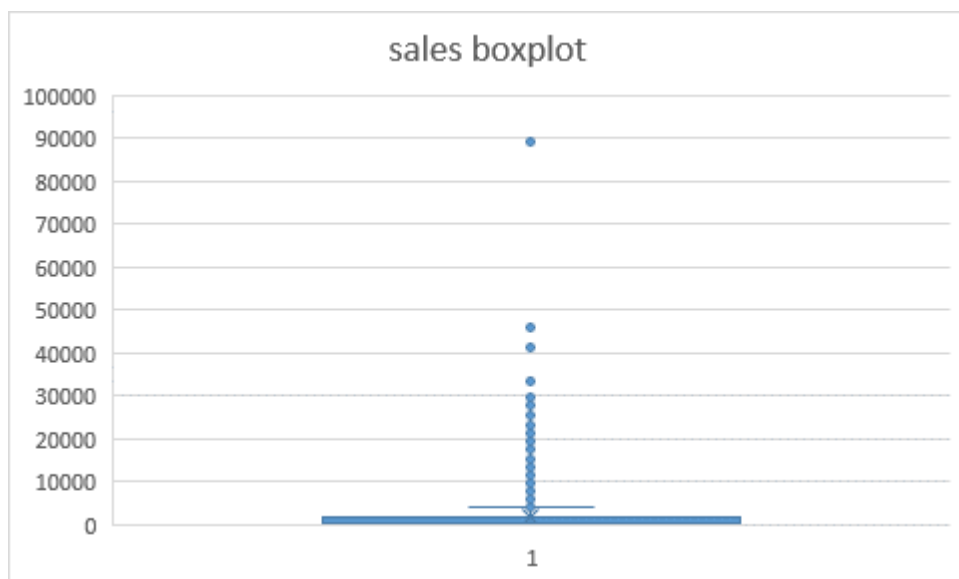
## **1.2; HISTOGRAMS, BOXPLOT, ANALYSIS**

Histograms and boxplots are two of the most commonly used tools for data visualisation. Histograms were used to visualise the data distribution by plotting the frequency of each value in a dataset. For example, the researcher used to analyse sales and

profits by plotting the frequency of each sales value and each profit value. Boxplots were used to visualise the data distribution by plotting the median, quartiles and outliers. Using histograms and boxplots in Excel helped quickly and accurately analyse sales and profits.

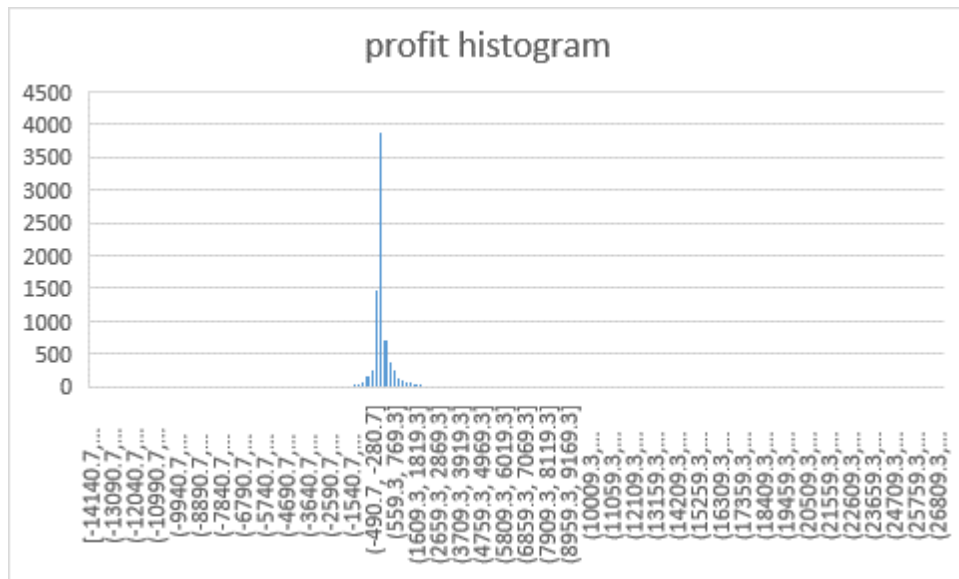


**1.2.1. Figure 1.** The sales histogram

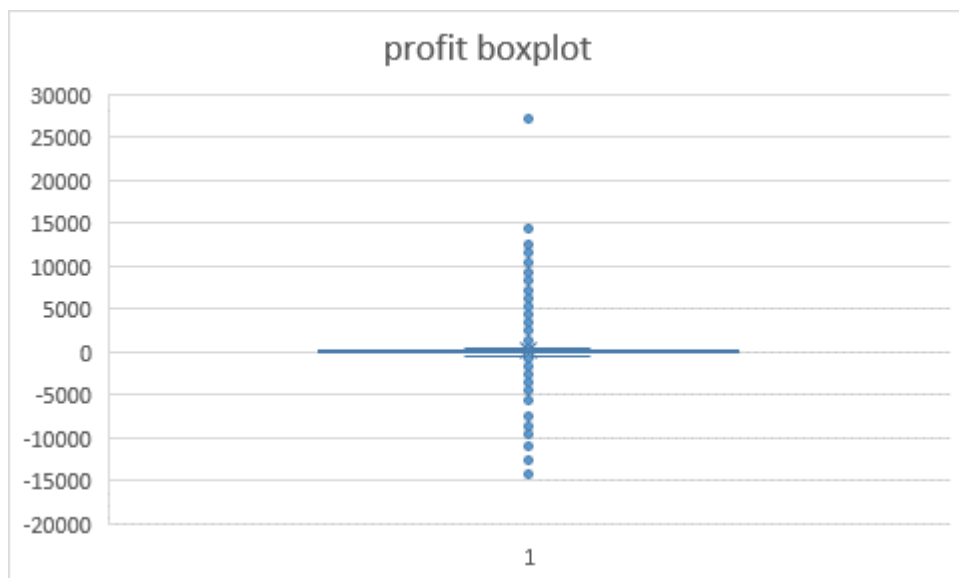


**1.2.2. Figure 2.** The boxplot outliers

Both figure 1 and 2 above show the presence of outliers in the data. The outliers exist from the value of 500 onwards.



**1.2.3. Figure 3.** The histogram of the profits



**1.2.4. Figure 4.** The boxplot for the profits

From the two figures shown above (fig. 1 and fig. 2), it is evident that outliers exist in the variable profit from the dataset. The box plot also shows that the outliers exist above and below the median.

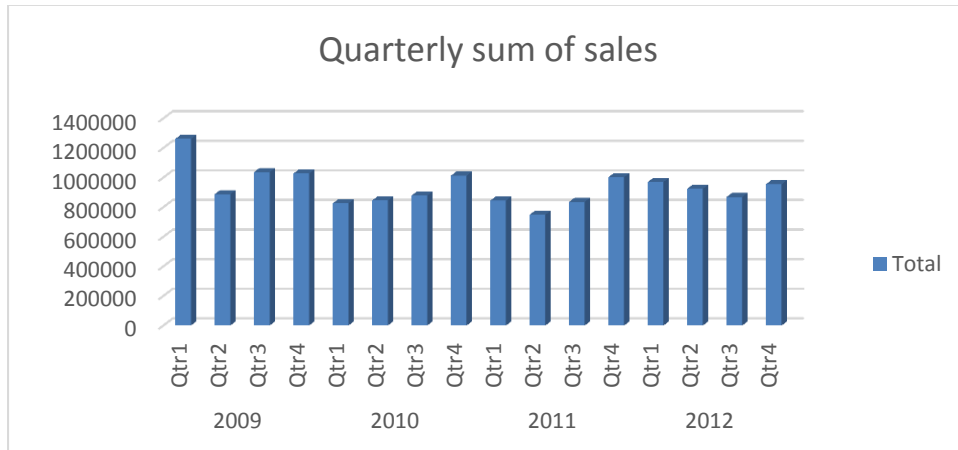
Pivot charts as an effective way to quickly analyse the decline and changes in sales and profits. Pivot charts were used to analyse sales and profits by plotting the sales and profits for each product and period. It can also help one make informed decisions and predictions. For example, by analysing sales and profits for the superstore product, the

researcher used pivot charts to identify how there was a decline, which year was generating the most sales, and which year was generating the most profits (Tanev, Liotta, and Kleismantas, 2015). This information helped the researchers make informed decisions about optimising sales and profits.

**1.2.5. Table 2:** The pivot analysis of the superstore sales

| Row Labels | Sum of Sales |
|------------|--------------|
| 2009       | 4209896.846  |
| Qtr1       | 1260505.865  |
| Qtr2       | 886878.088   |
| Qtr3       | 1036090.29   |
| Qtr4       | 1026422.603  |
| 2010       | 3560087.045  |
| Qtr1       | 825530.275   |
| Qtr2       | 844439.909   |
| Qtr3       | 877873.556   |
| Qtr4       | 1012243.305  |
| 2011       | 3429944.981  |
| Qtr1       | 843994.75    |
| Qtr2       | 748576.579   |
| Qtr3       | 835066.3905  |
| Qtr4       | 1002307.262  |
| 2012       | 3715671.953  |
| Qtr1       | 969332.0665  |
| Qtr2       | 923686.9725  |

|                    |                    |
|--------------------|--------------------|
| Qtr3               | 868312.0995        |
| Qtr4               | 954340.8145        |
| <b>Grand Total</b> | <b>14915600.82</b> |



**1.2.6. Figure 5.** The graphical presentation of the changes in the sales from the superstore

From the analysis of the pivot charts, it is visible that the units made in sales are significantly reducing from 1,260,506 to 954,341 (see table 2 and fig 5 above). The first quarter in 2009 had the highest sales, but the level decreased significantly.

**1.2.7. Table 3:** The pivot analysis for the profit

| Row Labels  | Sum of Profit |
|-------------|---------------|
| <b>2009</b> |               |
| Qtr1        | 115498        |
| Qtr2        | 91808.63      |
| Qtr3        | 102709.98     |
| Qtr4        | 124079.41     |
| <b>2010</b> |               |
| Qtr1        | 75498.29      |
| Qtr2        | 95896.19      |

|                    |                   |
|--------------------|-------------------|
| Qtr3               | 87673.02          |
| Qtr4               | 105849.83         |
| <b>2011</b>        |                   |
| Qtr1               | 82860.15          |
| Qtr2               | 78653             |
| Qtr3               | 92689.09          |
| Qtr4               | 126108.26         |
| <b>2012</b>        |                   |
| Qtr1               | 108373.19         |
| Qtr2               | 90635.68          |
| Qtr3               | 74929.57          |
| Qtr4               | 68505.69          |
| <b>Grand Total</b> | <b>1521767.98</b> |



**1.2.8. Figure 6.** The changes in profits from the superstore

Like the sales, the profit for the superstore also declines with time. The first quarter in 2009 had a profit of 115,498, which decreased to 68,505.69 in the final quarter of 2012 (see table 3 and fig. 6 above).

### **1.3; CONCLUSION ABOUT EXCEL**

Excel can be used to quickly and accurately forecast sales and profits. Excel can be used to analyse past data and make predictions. When forecasting sales and profits with Excel, it was important to use the correct data types. Excel has in-built functions allowing the researcher to convert data types to ensure accuracy. It was also important to use the correct chart type. For example, the researcher used a line chart while forecasting sales and profits (Ragsdale, 2021). The forecast was used to analyse past data and make predictions. When forecasting sales and profits for a particular product, the researcher used Excel to identify which year was most likely for the product to increase and when the product will most likely generate the most sales. This information helps make informed decisions about optimising sales and profits.

In conclusion, Excel is a powerful tool that can be used for data manipulation, analysis and visualisation. The last name of the first author of each work should alphabetise reference list entries in this study, and the researcher explored the best practices for working with excel data, pre-processing data, analysing and visualising data, using pivot charts and analysis on sales and profits, and using Excel forecasting for sales and profits. Using Excel for data analysis and visualisation helped the researcher identify data patterns, trends and relationships promptly and accurately. The programme also help make informed decisions and predictions. With the power of Excel, the researcher unlocked the potential of the data and can now make better decisions.

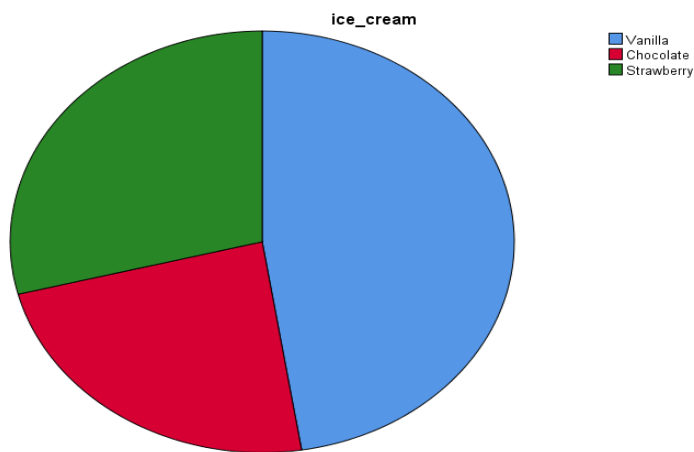
## **PART 2**

### **2.1. How Many Students Like Vanilla Flavour of Ice-Cream?**

The number of students with vanilla flavour is given as follows:

**ice\_cream**

|       |                | Frequency | Percent | Valid<br>Percent | Cumulative<br>Percent |
|-------|----------------|-----------|---------|------------------|-----------------------|
| Valid | Vanilla        | 95        | 47.5    | 47.5             | 47.5                  |
|       | Chocolate      | 47        | 23.5    | 23.5             | 71.0                  |
|       | Strawberr<br>y | 58        | 29.0    | 29.0             | 100.0                 |
|       | Total          | 200       | 100.0   | 100.0            |                       |



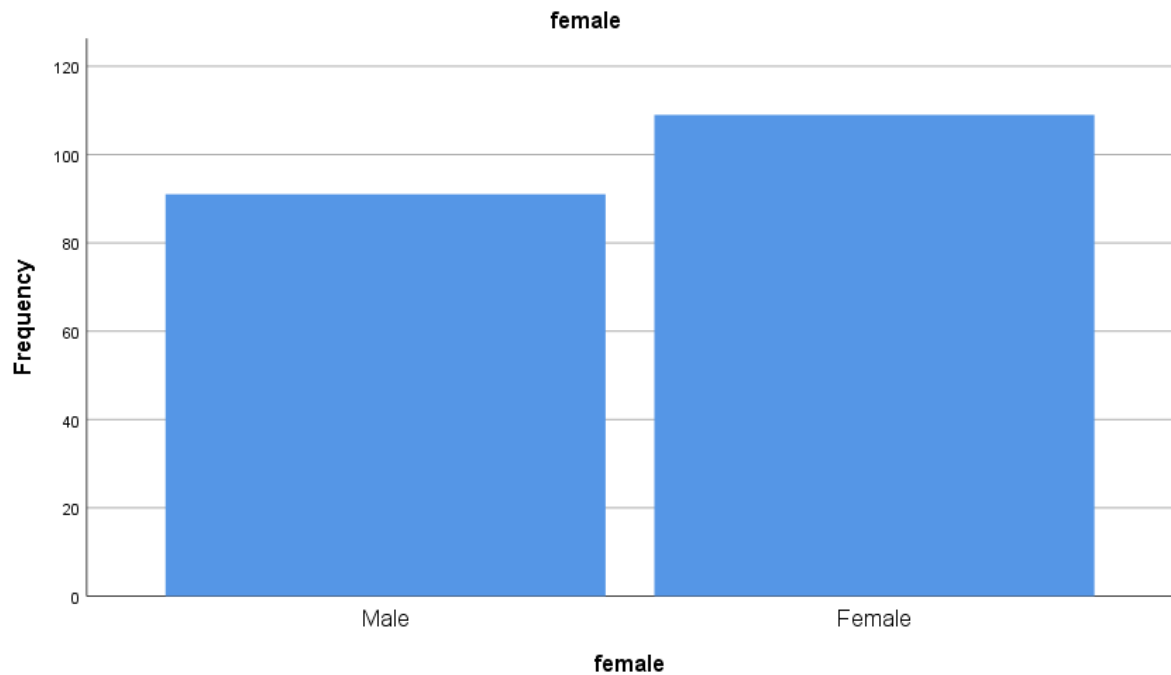
The number of students with vanilla flavour is 95, representing 47.5%.

## **2.2. How Many Students are Male and Female?**

The number of male and female students is given as follows:

### **female**

|       |        | Frequency | Percent | Valid<br>Percent | Cumulative<br>Percent |
|-------|--------|-----------|---------|------------------|-----------------------|
| Valid | Male   | 91        | 45.5    | 45.5             | 45.5                  |
|       | Female | 109       | 54.5    | 54.5             | 100.0                 |
|       | Total  | 200       | 100.0   | 100.0            |                       |



The number of female students is 109, with a percentage of 54.5%, while the number of male students from the analysis is 91, with 45.5%. This summary is represented in the bar graph shown above.

### **2.3. What Are the Mean and Median of Participants Who Likes Chocolate and Strawberry Flavour Ice-Cream?**

The mean and median of participants are given as follows:

#### **Descriptives**

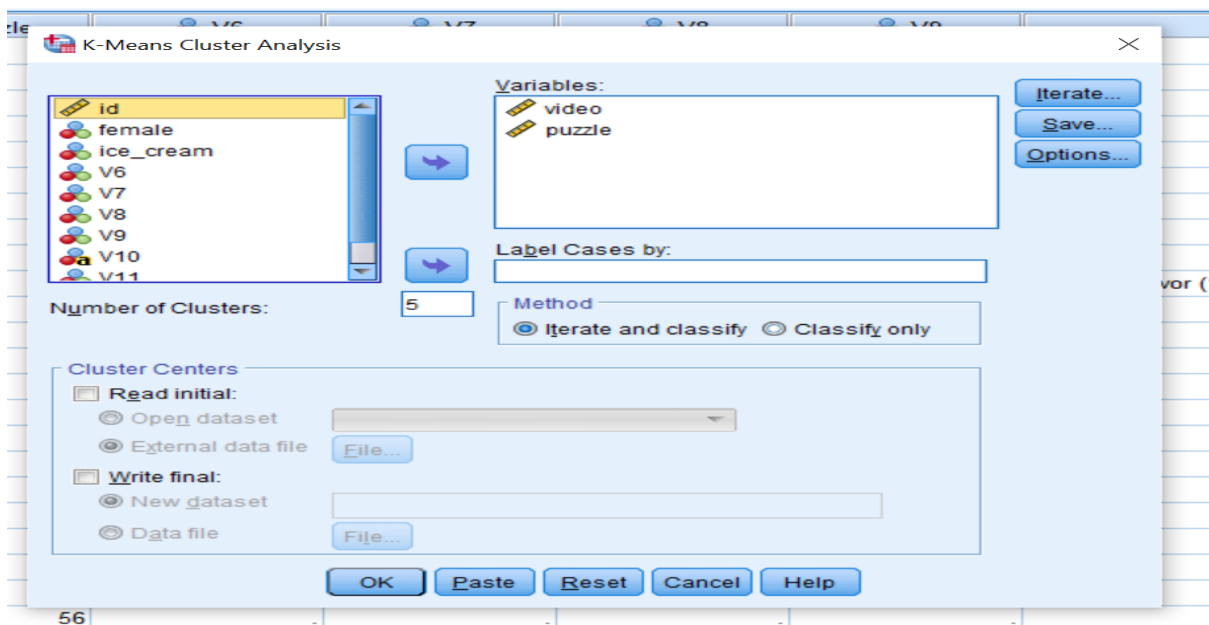
| ice_cream  |        | Statistic | Std. Error |
|------------|--------|-----------|------------|
| Chocolate  | Mean   | 47.70     | 1.519      |
|            | Median | 47.00     |            |
| Strawberry | Mean   | 55.45     | 1.286      |
|            | Median | 51.00     |            |
| Chocolate  | Mean   | 47.32     | 1.581      |
|            | Median | 46.00     |            |

|            |        |       |       |
|------------|--------|-------|-------|
| Strawberry | Mean   | 57.14 | 1.308 |
|            | Median | 61.00 |       |

The mean video for chocolate and strawberry is 47.70 and 55.45, respectively, while the median is 47.00 and 51. The mean puzzle for chocolate and strawberry are 47.32 and 57.14, respectively, while the median is 46 and 61.

#### **2.4. Using the Dataset Ice\_cream\_csv Provided in Conjunction with SPSS, Give a Specific Example of Clustering (ex. K-Means)**

K-Means is a method for solving the clustering problem in which many data points are clustered. K-Means aims to minimise errors between observations from each cluster and those from others (Govender and Sivakumar, 2020). The cluster centres tend to have values similar to those in any given cluster, so the algorithm tries to minimise squared errors between these two distances (Grenci, 2022). For example, the researcher can aim to group the data into five clusters depending on the video and puzzle values. To do this, the following steps are necessary:



The first step of the SPSS is to analyse and then classify. There is an option for k-means to get the results from the above screenshot. The next step is choosing which variables

are required to be clustered. In this case, the researcher has chosen a video and a puzzle. It is important to choose the number of iterations. The results from the clustering are as follows:

**Initial Cluster Centers**

|        | Cluster |    |    |    |    |
|--------|---------|----|----|----|----|
|        | 1       | 2  | 3  | 4  | 5  |
| video  | 31      | 69 | 72 | 29 | 55 |
| puzzle | 56      | 71 | 31 | 26 | 51 |

**Iteration History<sup>a</sup>**

|           | Change in Cluster Centers |           |           |           |           |
|-----------|---------------------------|-----------|-----------|-----------|-----------|
| Iteration | 1                         | 2         | 3         | 4         | 5         |
| 1         | 9.294                     | 9.593     | 13.438    | 14.019    | 3.108     |
| 2         | .450                      | .060      | 2.320     | 1.527     | 1.005     |
| 3         | .478                      | .001      | .110      | .053      | .216      |
| 4         | .015                      | 2.320E-5  | .005      | .002      | .003      |
| 5         | .000                      | 4.549E-7  | .000      | 6.259E-5  | 4.162E-5  |
| 6         | 1.459E-5                  | 8.919E-9  | 1.193E-5  | 2.158E-6  | 5.780E-7  |
| 7         | 4.560E-7                  | 1.749E-10 | 5.681E-7  | 7.443E-8  | 8.028E-9  |
| 8         | 1.425E-8                  | 3.468E-12 | 2.705E-8  | 2.567E-9  | 1.115E-10 |
| 9         | 4.453E-10                 | 2.842E-14 | 1.288E-9  | 8.850E-11 | 1.555E-12 |
| 10        | 1.392E-11                 | .000      | 6.134E-11 | 3.049E-12 | 1.589E-14 |

**Final Cluster Centers**

|        | Cluster |    |    |    |    |
|--------|---------|----|----|----|----|
|        | 1       | 2  | 3  | 4  | 5  |
| video  | 40      | 62 | 58 | 39 | 53 |
| puzzle | 51      | 64 | 40 | 36 | 55 |

**Number of Cases in each Cluster**

| Cluster | 1 | 33.000  |
|---------|---|---------|
|         | 2 | 50.000  |
|         | 3 | 25.000  |
|         | 4 | 27.000  |
|         | 5 | 65.000  |
| Valid   |   | 200.000 |
| Missing |   | .000    |

The clustering of the results above has been used as a natural method of organising data into groups. It was useful in both the video and puzzle values, especially when there were important relationships between the variables. The purpose of clustering is to identify which values belong together and make their group together for each cluster separately. If a certain object falls into more than one cluster, it has called a poly cluster.

## **2.5. The Most Common Data Mining and Text Mining Methods that Can Be Used in Business**

Association mining is a data mining technique that maps a relationship between variables and Boolean values. A hospital example can show links between the responses and the covariates: the relationship between hospital admission for heart conditions, alcohol use, and employment status (Ledolter, 2013). Thus, association methods focus on pairs or multiples of variables (covariates) that lead to a common outcome variable.

Outlier detection is commonly used in data mining to identify values that may be spurious, such as the result of incorrect calibration. For example, if one is trying to analyse the success of a marketing campaign by analysing monthly sales data, one might find that sales tended to spike during the spring months (Gupta and Jiwani, 2021). In this case, the spike occurring after March happens at a time when it is unlikely to happen naturally.

Data warehousing is a process of extracting vital information from the original database. It improves business intelligence (BI), enables faster decision-making, and expands the scope of data to be analysed. Data warehouses are designed to store transactional and time-variant data from multiple sources such as raw sales records, financial transactions, customer accounts and other systems requiring high-level analysis.

## **2.6. Advantages and Disadvantages of SPSS and Excel**

| Advantages                                                                                                                                                                                                                                            | Disadvantages                                                                                            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| One of the main advantages of using SPSS and Excel is the ability to manipulate data quickly and easily. Both SPSS and Excel allow the user to perform simple calculations on data sets and present the results in an easy-to-understand format (Ali, | Excel also has some features not available in SPSS, such as data validation and automatic recalculation. |

|                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Crvenkovski, and Johnson, 2016).</p> <p>Additionally, SPSS and Excel offer a range of statistical analysis tools that allow the user to analyse data and draw meaningful conclusions. SPSS and Excel also have various graphical tools that allow the user to visualise data and make it easier to interpret.</p>                                                                                               |                                                                                                                                                                                                                                                                                                                                                                           |
| <p>Another advantage of using SPSS and Excel is the ability to access and analyse large amounts of data. SPSS and Excel both can store and manipulate large datasets (Ali et al., 2016). This is especially useful for researchers who may have to deal with large amounts of data regularly. Additionally, both SPSS and Excel can create meaningful reports allowing the user to interpret data efficiently.</p> | <p>If one works with large amounts of data and does not have time to apply corrections manually, both programmes can be frustrating platforms if users are unfamiliar with them.</p>                                                                                                                                                                                      |
| <p>A third advantage of using SPSS and Excel is that they are both widely available. SPSS and Excel are both widely used in many industries and academic settings, making them easy to access and use (Ragsdale, 2021). Additionally, SPSS and Excel are both relatively affordable, making them accessible to many users.</p>                                                                                     | <p>If one decides to print out the data, there are disadvantages to this method. First, printing the data can be time-consuming because it requires large data storage space. Second, printed reports may not be as accurate or detailed as those created using other formats (Pisarek, Akpınar, and Hızıroğlu, 2017). Thirdly, the printed output is static and does</p> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                             |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | not include interactivity features such as sorting and filtering functions. |
| <p>For example, university researchers often use SPSS and Excel to analyse survey data. These surveys can collect data about student experiences, such as their satisfaction with a particular course or their feelings about a particular professor (Gupta and Jiwani, 2021). Researchers can then use SPSS and Excel to analyse this data, draw meaningful conclusions, and present their results in an easy-to-understand format. This makes it easier for universities to decide based on the survey results.</p> |                                                                             |

## REFERENCES

- Albright, S.C. and Winston, W.L., 2014. *Business analytics: data analysis & decision making*. Boston: Cengage Learning.
- Alexander, M., Decker, J. and Wehbe, B., 2014. *Microsoft business intelligence tools for Excel analysts*. Hoboken: John Wiley & Sons.
- Ali, O., Crvenkovski, P. and Johnson, H., 2016. Using a business intelligence data analytics solution in healthcare. In: *2016 IEEE 7th annual information technology, electronics and mobile communication conference (IEMCON)*. Vancouver: IEEE, pp.1-6.
- Bernhardt, B.R., Hinds, L.H., Meyer, L. and Strauch, K.P., 2018. *What's past is prologue: Charleston conference proceedings, 2017*. Purdue: Purdue University Press.
- Govender, P. and Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). *Atmospheric Pollution Research*, **11**(1), pp.40–56.
- Grenci, R.T., 2022. Introducing business analytics in context: an Excel project to analyze product and sales data for inventory purchase decisions. *Decision Sciences Journal of Innovative Education*, **20**(4), pp.190–200.
- Gupta, K. and Jiwani, N., 2021. A systematic overview of fundamentals and methods of business intelligence. *International Journal of Sustainable Development in Computing Science*, **3**(3), pp.31–46.
- Ledolter, J., 2013. *Data mining and business analytics with R*. Hoboken: John Wiley & Sons.
- Pisarek, R., Akpinar, M.T. and Hızıroğlu, A., 2017. Data mining application in air transportation – the case of Turkish airlines. *Logistics and Transport*, **36**, pp.79–88.
- Qiao, J. and Zhang, Y., 2015. Study on K-means method based on data-mining. In: *2015 Chinese automation congress (CAC)*. Wuhan: IEEE, pp.51-54.

Ragsdale, C., 2021. *Spreadsheet modeling and decision analysis: a practical introduction to business analytics*. Boston: Cengage Learning.

Tanev, S., Liotta, G. and Kleismantas, A., 2015. A business intelligence approach using web search tools and online data reduction techniques to examine the value of product-enabled services. *Expert System*