

**Name: Bandac Gabriela**

**Student Number:2014067**

**Module Title: Data Handling and Business Intelligence**

**Assignment Title: Assessment 2**

**Module Code : LCBB5000**

**Module Lecturer's Name : Joy Botu**

### **Executive summary**

The present report shed light on the use of Excel as well as SPSS software for deriving the results and outcomes of the data set. The functions of excel are important for the visualisation of data and the functions of SPSS are important for quick estimation of results based on the data set. Moreover, the study also involves the representation of several graphs based on the PIVOT table and LOOKUP function, which is important to performance in excel to summarise a large data set. Lastly, several tests in SPSS have also been performed in this report for understanding the outcomes and results of the survey conducted on 200 school students.

## Table of Contents

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| Introduction .....                                                                    | 4  |
| Part 1 .....                                                                          | 4  |
| Determination of crime rate .....                                                     | 4  |
| Use of LOOKUP function.....                                                           | 4  |
| Use of PIVOT TABLE function.....                                                      | 8  |
| Evaluation of the use of excel for pre-processing, analyzing, and visualization ..... | 11 |
| Part 2 .....                                                                          | 12 |
| Evaluation of the following .....                                                     | 12 |
| i. Number of students who like vanilla flavour .....                                  | 12 |
| ii. No. of students are male and female .....                                         | 13 |
| iii. Mean median of participants .....                                                | 14 |
| Evaluation of the data set.....                                                       | 16 |
| A. Specific example of clustering.....                                                | 16 |
| B. Data mining and text mining methods.....                                           | 17 |
| C. Advantages and disadvantages of using SPSS and Excel .....                         | 17 |
| Conclusion.....                                                                       | 18 |
| References .....                                                                      | 20 |

## Introduction

Data handling is the process of strong data which plays a significant role in the excellent evaluation of data for getting an outcome. Data can be handled by collecting, storing, organising, and presenting the data in a way so that one can understand the advantages and disadvantages of the collection of data. The study mainly aims to analyse and visualise data outcomes achieved after the use of excel functions such as ***IF, LOOKUP, and PIVOT TABLE***.

## Part 1

This part contains the use of excel functions for deriving a reliable and accurate result of crime rate from the provided data set. The data set involves a range of variables, some of which depict the crime rate in every aspect of life such as “***youth, education, males, youth unemployment, mature unemployment, crimerate10, and its associated variables***”. The data set contains the data of crime rate and crime rate 10 and based on which several excel functions will be performed so that one can understand the number of offences per million populations (Abdel-Rahman, 2020, p. 91). Moreover, the use of excel functions is most important for attaining the possible outcome from the data set. This is because the functions of excel like ***IF, LOOKUP, and PIVOT TABLE*** allow the user to manipulate the data as per their preference for chaining the positive results from the computations. Therefore, the excel functions are performed below by using the aforementioned excel functions:

## Determination of crime rate

The determination of crime rate can become possible by simultaneous use of the data of crime rate and its associated variables. The data set contains 47 different variables based on which the crime rate can be performed by which one would understand the crime rate per million populations and crime rate 10 per million populations (Ahmed *et al.* 2018, p. 14). Hence, the calculations of crime rate using LOOKUP and PIVOT TABLE function are provided below:

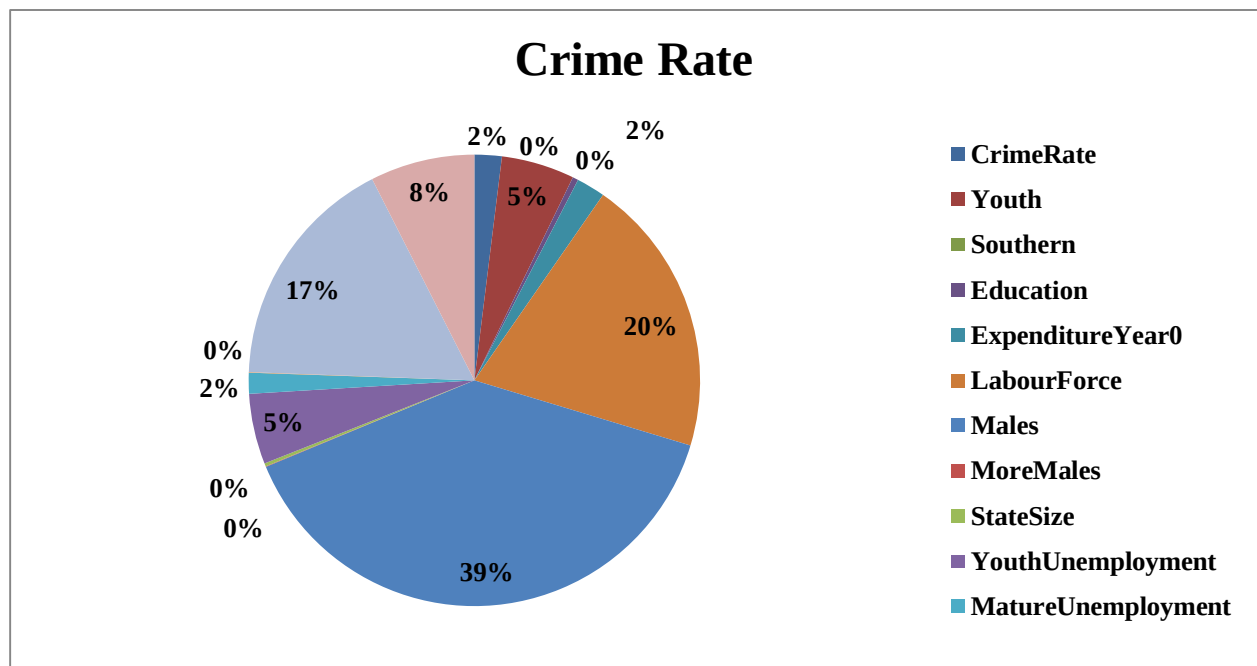
## Use of LOOKUP function

***Crime rate using LOOKUP***

| Crime Rate | Youth | Southern | Education | Expenditure Year0 | Labour Force | Males | More Males | State Size | Youth Unemployment | Mature Unemployment | High Youth Unemployment | Wage | Below Wage |
|------------|-------|----------|-----------|-------------------|--------------|-------|------------|------------|--------------------|---------------------|-------------------------|------|------------|
| 52.3       | 140   | 0        | 10.9      | 55                | 535          | 1045  | 1          | 6          | 135                | 40                  | 1                       | 453  | 200        |

**Table 1: LOOKUP**

(Source: Created by the researcher)



## Figure 1: LOOKUP

(Source: Created by the researcher)

### Discussion

LOOKUP function is one of the excellent functions of excel in which one can derive the result of one cell based on the referencing of other cells simultaneously. The LOOKUP function is helpful in deriving the result of one cell as well as multiple cells, which creates the ease of understanding the outputs of the results (Ali *et al.* 2018, p. 5). In the same way, the LOOKUP function is applied in a data set to derive the results of the crime rate per million populations. The estimation of LOOKUP required two important things that are “*lookup value and table array*”. The lookup value is the cell-based on which the outcomes will be derived and the table array is the referencing of the overall table based on which results will be depicted on the selected cell range. Moreover, the LOOKUP function contains the variables of crime rate per million in which data of other variables have also been considered while estimating the results.

Additionally, it can be illustrated that the function of data validation has also been done in the crime rate cell so that one can choose the input for deriving the expected output. The data validation contains the values of the overall crime rate hence; by selecting one value of the crime rate the results of other variables will automatically be changed (Band *et al.* 2020, p. 75). Due to the help of this function, one can check the crime rate based on youth, education, males, state size, etc since it will show the correct values of other variables by changing the input variable of the crime rate cell. Moreover, from the above table it is found that when the crime rate is 52.3 the rate of youth is 140, education is 10.9 and so on. However, when the crime rate value will change from 52.3 to 56.6 the rate of youth will become 157, education will be 11.2 and so on.

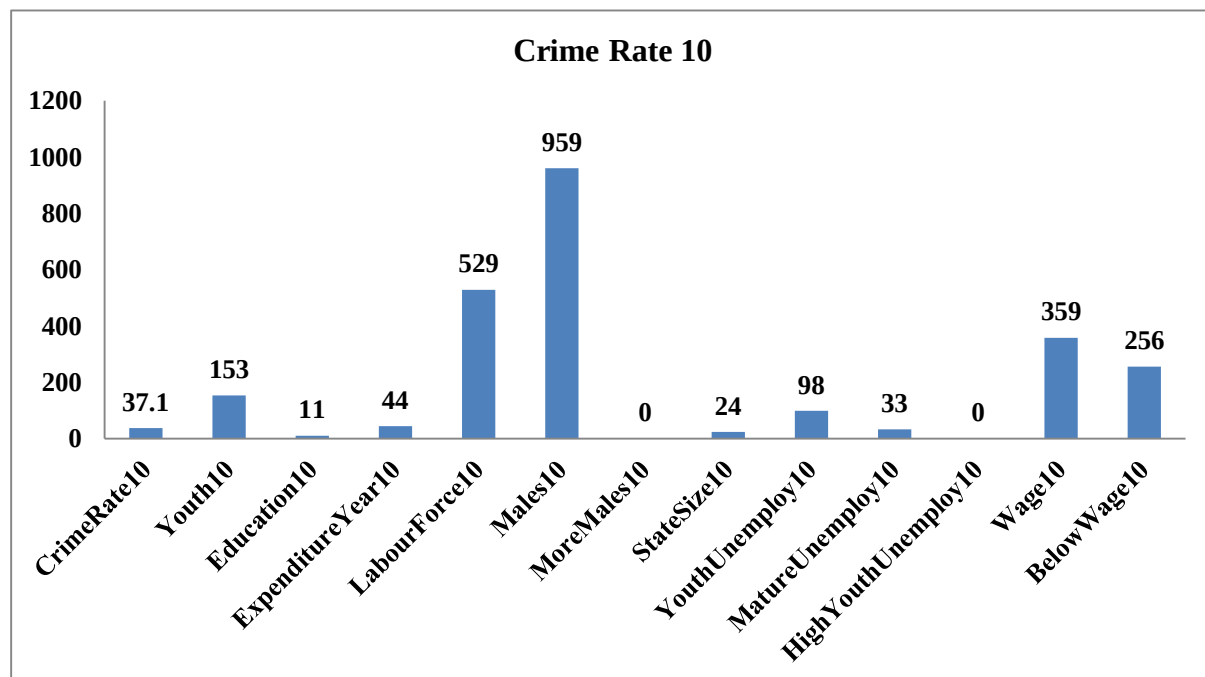
Therefore, this clarifies that the change of input in the crime rate cell will change the overall results of the other cells (Hamidzadeh and Moradi, 2020, p. 74). This means the other cells are interdependent upon the crime rate since a change in this cell can change the outcomes of the crime rate variables.

### Crime rate 10 using LOOKUP

| Cri<br>meR<br>ate1<br>0 | Yo<br>uth<br>10 | Edu<br>cati<br>on1<br>0 | Expen<br>diture<br>Year1<br>0 | Labo<br>urFor<br>ce10 | M<br>ale<br>s1<br>0 | Mor<br>eMal<br>es10 | Stat<br>eSiz<br>e10 | Youth<br>Unem<br>ploy10 | Mature<br>Unemp<br>loy10 | HighYo<br>uthUne<br>mploy10 | W<br>ag<br>e1<br>0 | Belo<br>wWa<br>ge10 |
|-------------------------|-----------------|-------------------------|-------------------------------|-----------------------|---------------------|---------------------|---------------------|-------------------------|--------------------------|-----------------------------|--------------------|---------------------|
| 37.1                    | 153             | 11                      | 44                            | 529                   | 959                 | 0                   | 24                  | 98                      | 33                       | 0                           | 359                | 256                 |

**Table 2: LOOKUP**

(Source: Created by the researcher)



## Figure 2: LOOKUP

(Source: Created by the researcher)

### Discussion

The above table and graph have been derived based on the use of an excel function i.e. LOOKUP function. The difference between the above table and in this table is that it contains the variables of crime rate 10 and the above table has associated the variables of crime rate. The data of crime rate and crime rate 10 is also different which depicts a different result. Moreover, the value of crime rate 10 is lower than the crime rate which means the associated variables of the crime rate 10 will also be low in comparison to the above table and graph (Kadirhodjaev *et al.* 2020, p. 56). The main reason behind the estimation of both the variables in different forms is to make a significant comparison to the crime rate in both the variables. Therefore, the LOOKUP function is quite helpful in segregating each variable of crime rate 10 in comparison to the crime rate.

Apart from that, the data validation is also done in crime rate 10 cells so that one can choose the preferred rate to see the results of other variables. This is quite satisfactory for depicting the results of all the variables since complete information of variables is required to estimate the number of offences per million populations. A bar graph is also made for the illustration of crime rate 10 since the results of charts will also be changed by changing the input rate of crime (Kumar and Khatri, 2017, p. 181). For instance, when the crime rate is 10 is 37.1 the rate of youth is 153, education is 11, expenditure is 44 and so on. However, when the crime rate cell changed from 37.1 to 42.7 it depicts different results than the aforementioned results. Due to this reason, the use of the LOOKUP function is important to oversee the complete result with one click on the lookup value.

### Use of PIVOT TABLE function

|                    |           |           |           |            |            |           |           |           |            |            |            |           |           |
|--------------------|-----------|-----------|-----------|------------|------------|-----------|-----------|-----------|------------|------------|------------|-----------|-----------|
| <b>Val<br/>ues</b> |           |           |           |            |            |           |           |           |            |            |            |           |           |
| <b>Su</b>          | <b>Su</b> | <b>Su</b> | <b>Su</b> | <b>Sum</b> | <b>Sum</b> | <b>Su</b> | <b>Su</b> | <b>Su</b> | <b>Sum</b> | <b>Sum</b> | <b>Sum</b> | <b>Su</b> | <b>Su</b> |

| m<br>of<br>Crime<br>Rate | m<br>of<br>Youth | m<br>of<br>Southern | m<br>of<br>Education | of<br>ExpenditureYear0 | of<br>Labour<br>Force | m<br>of<br>Males | m<br>of<br>More<br>Males | m<br>of<br>State<br>Size | of<br>Youth<br>Unemployment | of<br>Mature<br>Unemployment | of<br>High<br>Youth<br>Unemployment | m<br>of<br>Wage | m<br>of<br>Below<br>Wage |
|--------------------------|------------------|---------------------|----------------------|------------------------|-----------------------|------------------|--------------------------|--------------------------|-----------------------------|------------------------------|-------------------------------------|-----------------|--------------------------|
| 483<br>2                 | 65<br>13         | 16                  | 582<br>.4            | 3995                   | 263<br>76             | 46<br>20<br>2    | 9                        | 172<br>1                 | 4487                        | 1597                         | 15                                  | 24<br>69<br>3   | 911<br>8                 |

| Sum of<br>CrimeRate10 | Sum<br>of<br>Youth<br>10 | Sum of<br>Education10 | Sum of<br>ExpenditureYear10 | Sum of<br>LabourForce10 | Sum<br>of<br>Males<br>10 | Sum of<br>MoreMales10 | Sum of<br>StateSize10 | Sum of<br>YouthUnemployment10 |
|-----------------------|--------------------------|-----------------------|-----------------------------|-------------------------|--------------------------|-----------------------|-----------------------|-------------------------------|
| 4797.3                | 6605                     | 582.9                 | 3771                        | 26580                   | 46383                    | 10                    | 1772                  | 4580                          |

**Table 3: Pivot table**

(Source: Created by the researcher)

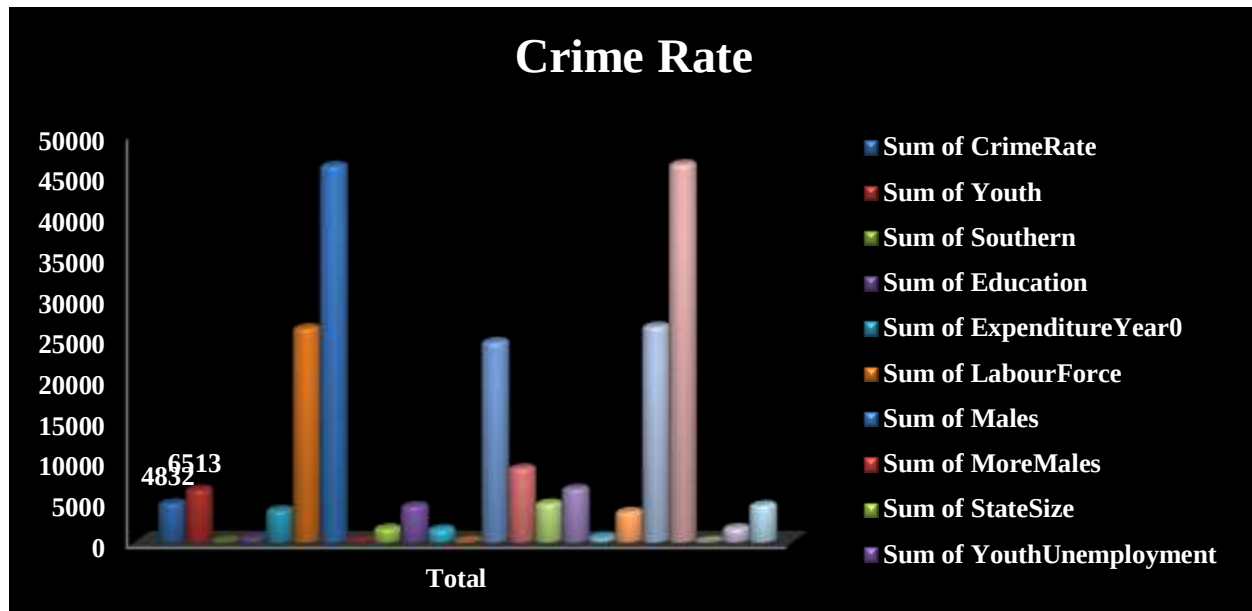


**Figure 3: No. of offences**

(Source: Self-made)

## Discussion

A pivot table can be defined as the efficient excel function to summarise the large data set into a smaller data set so that one can understand the overall facts of the data. In the same way, the crime rate data contains different variables and their values of percentages which are difficult to understand instantly. Hence, a pivot table and a graph are prepared below based on the crime rate and crime rate 10 (Lanza-Cruz *et al.* 2018, p. 5). Both the variables have been used simultaneously to depict the no of offences per million populations and the pivot table is one of the easy methods of excel to represent the large crime rate data into a smaller set of data. The sum of all the variables is estimated in the above table whether it is crime rate and its associate variables or crime rate 10 and its associate variables.

After measuring the above pivot table it can be illustrated that the no of offences is lower in crime rate 10 in comparison to the crime rate. This is because the sum of the crime rate is 4832 and the sum of the crime rate of 10 is 4797.3, which clearly depicts that the offence per million populations is higher in crime rate and its associated variables. Moreover, a graph has also been presented above which is prepared based on the pivot table (Mohd and Biswas, 2018, p. 27). The

graph depicts that the number of offences in different variables is different which can be changed due to a change in the crime rate significantly. Moreover, the no of offences among males is higher in crime rate 10 in comparison to the crime rate. This is because the sum of males is 46,2020 and the sum of males 10 is 46,383, which means the male 10 has made high no of offences in comparison to male.

In addition, the use of both the excel functions is helpful in sorting the crime rate data and is helpful in estimating the number of offences per million population (Mrowczynska and Sztubecki, 2019, p. 7). Lastly, it can be said that the use of graphs and charts is helpful in presenting the table in a graphical form so that by seeing the graph one can understand the trends and changes in the crime rate and its associated variables.

### **Evaluation of the use of excel for pre-processing, analyzing, and visualization**

Pre-processing of data is the manipulation of raw data in order to attain a favourable outcome. The pre-processing of data needs analysis for illustrating all the aspects of the data processing so that one can properly understand the reason behind the collection and analysis of data (Paradza and Daramola, 2021, p. 11). The most important thing in pre-processing of data is its visualisation since the entire representation of data is dependent upon the visualisation of data. The visualisation of data involves the use of graphs and charts for representing the tables and data in a graphical format so that everyone can understand the changes and trends of the data. Moreover, the visualisation of data is the key to success in data handling and data mining by which one can evaluate every aspect of the collection. For instance, several graphs have been used for the above tables of crime rates so that one can easily understand the number of offences per million populations (Pearman *et al.* 2020, p. 1). Due to this reason, the representation or visualisation of data and its analysis is significantly important to achieve the required outcome from the data set.

## Part 2

### Evaluation of the following

Data analysis based on the survey responses is required to be done in this part so that one can understand the preferences of different students. 200 students are taken for the survey in which the preferences of all the students can be understood based on video games and puzzles (Polaka *et al.* 2021, p. 1). Moreover, the responses of students are also gathered based on their favourite flavoured ice cream. Hence, these responses of students can be analyzed and performed based on statistical software named SPSS and the outcomes or results are provided below:

#### i. Number of students who like vanilla flavour

| ice-cream |            |           |          |               |                    |
|-----------|------------|-----------|----------|---------------|--------------------|
|           |            | Frequency | Per cent | 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         |                    |

**Table 4: Students like vanilla**

(Source: Created by the researcher)

## Discussion

SPSS software has been used to derive the above table which depicts the no. of students who like ice cream flavour. Hence, from the above estimation, it is found that out of 200 students 95 students like vanilla flavour, 47 students like chocolate flavour, and 58 students like strawberry flavour. Therefore, the majority of students like vanilla flavour as the percentage of liking vanilla flavour are 47.5%, which is higher than the percentage of other flavours.

**ii. No. of students are male and female**

| <b>Gender</b> |        |           |          |               |                    |
|---------------|--------|-----------|----------|---------------|--------------------|
|               |        | Frequency | Per cent | 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         |                    |

**Table 5: Male and female**

(Source: Created by the researcher)

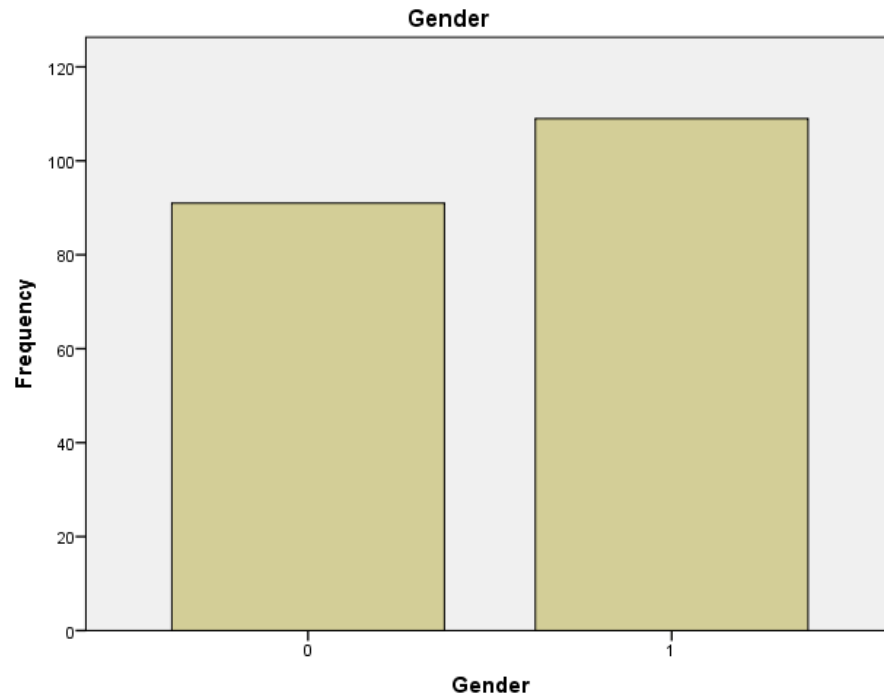


**Figure 4: Gender**

(Source: Created by the researcher)

## Discussion

The above table and graph depict the no. of male and female students in the survey containing the data of 200 high school students. The table depicts that the no. of female students is higher than that of male students since the no. of female students is 109 and male students are 91. This is also shown in the percentage form in which the male students are 45.5% and female students are 54.5%. Lastly, it can be said that the ratio of males and females is not the same due to which the majority of preferences will be different in the survey.

### iii. Mean median of participants

## Report

| Gender     |      |     |                |        |
|------------|------|-----|----------------|--------|
| ice-cream  | Mean | N   | Std. Deviation | Median |
| Vanilla    | .51  | 95  | .503           | 1.00   |
| Chocolate  | .68  | 47  | .471           | 1.00   |
| Strawberry | .50  | 58  | .504           | .50    |
| Total      | .54  | 200 | .499           | 1.00   |

**Table 6: Mean and median**

(Source: Created by the researcher)

## Discussion

The above table contains the information of mean and median along with the ice cream flavour, which is derived with the help of SPSS. Therefore are a total of 200 students provided their responses on ice cream flavour based on which the table has been derived. Hence, based on the table, it is illustrating that the mean is the average value of the frequency and 0.68 students like chocolate flavour and 0.50 students like strawberry flavour (Ratia *et al.* 2019, p. 39). Hence, approximately 47 and 58 students are those who like the chocolate flavour and strawberry flavour. Apart from that, the median is the middle value of the distribution which depicts the same result in chocolate flavour in comparison to vanilla flavour.

## Evaluation of the data set

### A. Specific example of clustering

| ANOVA     |             |    |             |     |         |      |
|-----------|-------------|----|-------------|-----|---------|------|
|           | Cluster     |    |             |     | F       | Sig. |
|           | Mean Square | df | Mean Square | df  |         |      |
| Gender    | .030        | 1  | .250        | 198 | .120    | .729 |
| ice-cream | 3.095       | 1  | .723        | 198 | 4.284   | .040 |
| video     | 8700.783    | 1  | 54.579      | 198 | 159.415 | .000 |
| puzzle    | 12815.244   | 1  | 51.116      | 198 | 250.709 | .000 |

**Table 7: K-means**

(Source: Created by the researcher)

## Discussion

The clustered table is derived above with the help of SPSS which depicts the K-means of the data set gathered from the responses of 200 students. The K-means is important to derive in the above table since it depicts the concentration of the identified variable of collected responses. Therefore, an ANOVA test has been done which depicts the mean square, significance and frequency of the responses of students. This significance of gender is 0.729 which is higher than

ice cream, video, and puzzles. This is because the significance of ice cream is 0.040, the video is 0.000, and the puzzle is 0.000. Lastly, it is illustrating that the df of all the variables are the same, which means the degree of freedom of gender, ice cream, video, and puzzle are the same.

## **B. Data mining and text mining methods**

Data mining can be defined as the process of understanding the patterns and correlations among a large data set for predicting the possible outcome and results. On the other hand, text mining is the method or technique to obtain quality information from a range of texts. Hence, the process of data mining and text mining is simultaneously important for a business for the prediction of outcomes based on their results. It is quite evident that large firms possess a large amount of data and use the data they want to predict the future outcomes but due to having a heavy collection of data; it is quite difficult for them to predict the outcome (Supriyati and Abdillah, 2020, p. 87). In this situation, they use the methods or techniques of data mining for predicting the outcomes from the large data set. Hence, some of the methods of data mining are provided below:

1. Classification analysis
2. Clustering analysis
3. Regression analysis

The above three analyses are the vital methods of data mining and are commonly used by large companies in their business for correlating the relationship of their large data set. For instance, **Amazon** is one of the multinational companies that use text mining in business for attaining a reliable price for their products. Moreover, **Netflix** is also using data mining in their business for estimating the customer's reviews on movies and web series.

## **C. Advantages and disadvantages of using SPSS and Excel**

Both the SPSS and Excel is an excellent function for analysing and visualising the complex data set so that one can understand all the outcomes or results of the data. Hence, every function has the positive and negative results hence, the **advantages and disadvantages of both SPSS and excel** are provided below:

| <b><i>Particulars</i></b>   | <b><i>SPSS</i></b>                                                                                                                                                                                                                                                                | <b><i>Excel</i></b>                                                                                                                                                                                                                                                   |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>Advantages</i></b>    | <ol style="list-style-type: none"> <li>1. It provides quick results of the data set.</li> <li>2. It can handle a large amount of data and text simultaneously.</li> <li>3. With the help of this one can easily handle the large data and information of the business.</li> </ol> | <ol style="list-style-type: none"> <li>1. It can be used for strong data and information about the business.</li> <li>2. The estimation of various calculations can be done in excel.</li> <li>3. Data visualization will become very easy by using excel.</li> </ol> |
| <b><i>Disadvantages</i></b> | <ol style="list-style-type: none"> <li>1. The estimation using this software is time-consuming.</li> <li>2. The complexity level is very high in this software.</li> </ol>                                                                                                        | <ol style="list-style-type: none"> <li>1. The risk of human error is high in excel.</li> <li>2. Excel is not designed for collaborative work.</li> </ol>                                                                                                              |

**Table 8: SPSS and excel**

(Source: Waqar *et al.* 2020, p. 12)

## **Conclusion**

The above discussion focuses on the use of SPSS for producing an excellent data analysis of the data set containing the information about different variables. It is found from the above statements that the use of excel has been helpful in excellent data representation of visualization by using several graphs and charts. Lastly, it is also concluded that the use of SPSS is helpful in

estimating the positive and negative outcome of the survey responses of students and based on which the means, median, and ANOVA test has been done.

## References

### Journals

Abdel-Rahman, A.H., 2020. The Use of the Traditional Tools in Consecutive Interpreting: Advantages and Disadvantages. *International Journal of Applied Linguistics & English Literature*, **9**(6), pp. 91-98.

Ahmed, V., Aziz, Z., Tezel, A. and Riaz, Z., 2018. Challenges and drivers for data mining in the AEC sector. *Engineering, Construction and Architectural Management*, **25**(11), pp. 1436-1453.

Ali, M.S., Shah, J.M. and Khan, S., 2018. Antecedents of Business Intelligence Implementation for Addressing Organizational Agility in Small Business Context. *Pacific Asia Journal of the Association for Information Systems*, **10**(1), pp. 5.

Band, S.S., Al-Shourbaji, I., Karami, H., Karimi, S., Esfandiari, J. and Mosavi, A., 2020. Combination of Group Method of Data Handling (GMDH) and Computational Fluid Dynamics (CFD) for Prediction of Velocity in Channel Intake. *Applied Sciences*, **10**(21), pp. 7521.

Bitour: A Business Intelligence Platform for Tourism Analysis. 2020. *ISPRS International Journal of Geo-Information*, **9**(11), pp. 671.

Hamidzadeh, J. and Moradi, M., 2020. Enhancing data analysis: uncertainty-resistance method for handling incomplete data. *Applied Intelligence*, **50**(1), pp. 74-86.

Kadirhodjaev, A., Rezaie, F., Moungh-Jin, L. and Lee, S., 2020. Landslide Susceptibility Assessment Using an Optimized Group Method of Data Handling Model. *ISPRS International Journal of Geo-Information*, **9**(10), pp. 566.

Kumar, N. and Khatri, S., 2017. Significance of Data Mining in Disease Classification and Prediction for Mining Clinical Data : A Review. *International Journal of Advanced Research in Computer Science*, **8**(5), pp. 181-186.

Lanza-Cruz, I., Berlanga, R. and Aramburu, M.J., 2018. Modeling Analytical Streams for Social Business Intelligence. *Informatics*, **5**(3),.

- Mohd, A.A. and Biswas, R., 2018. Sorted r-Train: An Improved Dynamic Data Structure for Handling Big Data. *International Journal of Intelligent Systems and Applications*, **10**(11), pp. 27.
- Mrowczynska, M. and Sztubecki, J., 2019. Group Method of Data Handling as a Tool to Determine Vertical Displacements. *IOP Conference Series. Materials Science and Engineering*, **471**(5),.
- Paradza, D. and Daramola, O., 2021. Business Intelligence and Business Value in Organisations: A Systematic Literature Review. *Sustainability*, **13**(20), pp. 11382.
- Pearman, W.S., Freed, N.E. and Silander, O.K., 2020. Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. *BMC Bioinformatics*, **21**, pp. 1-15.
- Polaka, I., Razuka-Ebela, D., Jin, Y.P. and Leja, M., 2021. Taxonomy-based data representation for data mining: an example of the magnitude of risk associated with 0RW1S34RfeSDcfkexd09rT2H. pylori1RW1S34RfeSDcfkexd09rT2 infection. *Biodata Mining*, **14**, pp. 1-21.
- Ratia, M., Myllärniemi, J. and Helander, N., 2019. The potential beyond IC 4.0: the evolution of business intelligence towards advanced business analytics. *Measuring Business Excellence*, **23**(4), pp. 396-410.
- Review of Wind Models at a Local Scale: Advantages and Disadvantages. 2021. *Journal of Marine Science and Engineering*, **9**(3), pp. 318.
- Supriyati and Abdillah, S.R., 2020. Data Mining in Sales Data Grouping. *IOP Conference Series. Materials Science and Engineering*, **879**(1),.
- Waqar, A.K., Chung, S.H., Awan, M.U. and Wen, X., 2020. Machine learning facilitated business intelligence (Part II): Neural networks optimization techniques and applications. *Industrial Management & Data Systems*, **120**(1), pp. 128-163.

