



**Prifysgol Cymru**  
Y Drindod Dewi Sant  
**University of Wales**  
Trinity Saint David

Module: Data Handling and Business Intelligence

Module code: LCBB5000

George Daniel Varzaru

2128103

## Table of Contents

|                                                                     |    |
|---------------------------------------------------------------------|----|
| Introduction.....                                                   | 3  |
| Database Analysis of Hotel Reservation.....                         | 3  |
| Observation.....                                                    | 5  |
| Evaluation of Data Using Microsoft Excel.....                       | 6  |
| The Utilization of LOOKUP and PIVOT TABLE in Use of Data .....      | 7  |
| LOOKUP .....                                                        | 7  |
| PIVOT TABLE.....                                                    | 8  |
| Dataset Analysis of Students Based on The Score .....               | 9  |
| Number of Male and Female Students.....                             | 9  |
| Qualification of students for Reduced/Free Lunch .....              | 10 |
| Mean and Median of the students .....                               | 11 |
| Pre-Test and Post-Test scores of Students for Differentiation ..... | 11 |
| Use of K-Means Clustering in Analysing SPSS .....                   | 12 |
| Methods Use For Common Data Mining and Text Mining.....             | 13 |
| Conclusion: .....                                                   | 15 |
| References:.....                                                    | 16 |

## Introduction

Excel is a great tool and the most widely used software for preparing and working data. For preparing data it can be used in various ways like through the use of Excel one can clean, and organize, remove needless space from the dataset, fix errors, and many more things in an effective way (Bree and Gallagher, 2016). Excel also helps to fix wrong numbers with the right numbers. It helps a person to format his or her data in many ways for example to show the dates, make percentages, helps to modify the layouts of data cells and column height and width so that one person can make format and show the Excel data in a better way. Excel helps a person manipulate data through its built-in functions (Hvitfeldt and Silge, 2021). Excel help to manipulate and modify data through sorting, filtering, and aggregating data options. Among many tools, the Pivot table is one of the most powerful tools. Because this tool helps a person summarize and analyze the data in a more easy and better way. The nice thing about Excel is that it is a very friendly tool and a person does not have to be a computer expert to operate Excel sheets. Even if a person is nascence and starts using Excel with using some knowledge can capable of operating Excel sheets because it is very handy and online many resources help a beginner if the person is stuck in operating Excel (Carlberg, 2014). Though Excel is a very wonderful tool, it has some limitations and in that case, a person needs more specialized tools for better work. But otherwise, Excel is very handy and a great tool for data analysis and data modification and also use for doing great research.

## Database Analysis of Hotel Reservation

For hotel reservations often need to analyze the data of the hotel and for this, there is a need to calculate the medium amount for calculating the medium in Excel one has to use the “AVERAGE” function. It is a great tool for analyzing a big amount of datasets and using this one can save a great amount of time (Berenson et al., 2019).

So some steps followed in the time of calculation “AVERAGE”. The first step is to select the column where the desired data contain and wish to average. The second step is to select the cell where a person gets the desired result by typing “=AVERAGE”. The third step is to click and get the desired result. Suppose a person wants to find the average amount of cells from the B20 cell to the B30 cells. So for this select a cell where the person wants the answer and use the formula which is “=AVERAGE(B20:B30)”.

## Data Charts

[illegible]

### Figure 1 Finding the decline or increase in average price

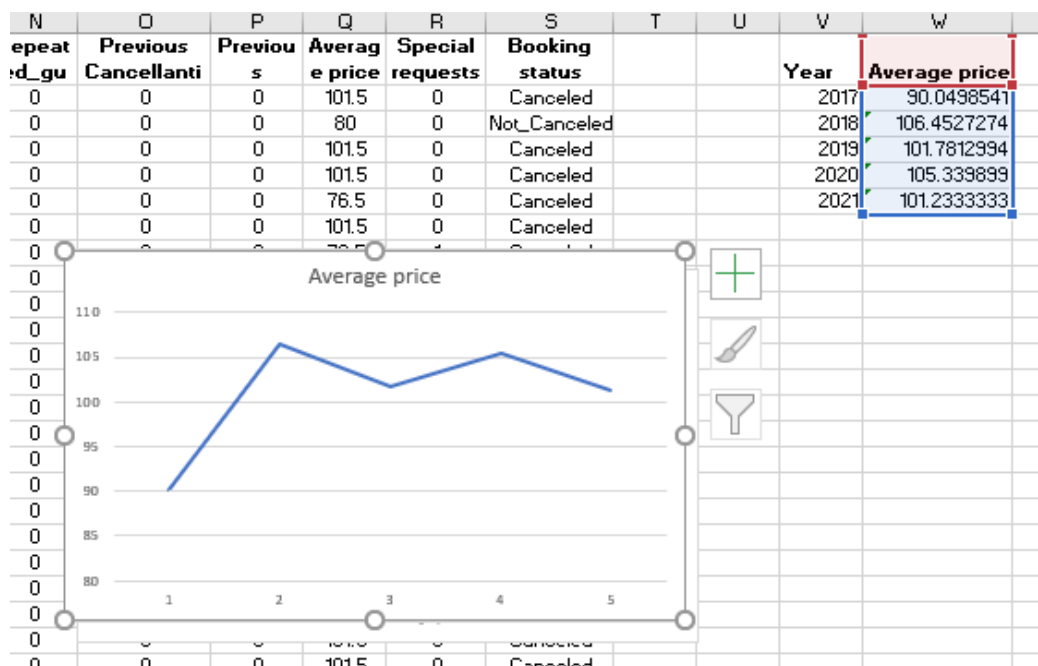


**Figure 2: Finding the decline or increase in average price using the chart (Source: Author)**

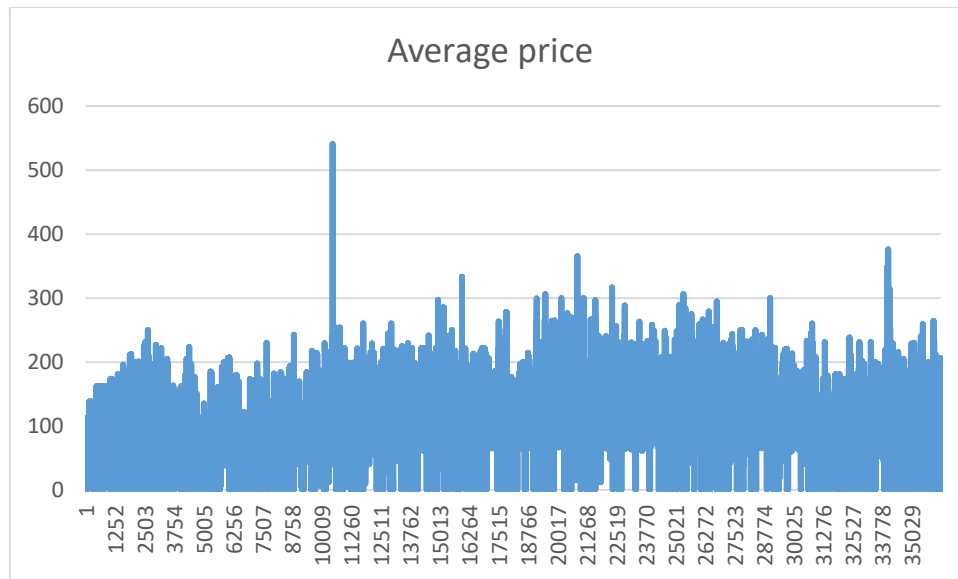


**Figure 3: Finding the decline or increase in average price using all the values in a chart**

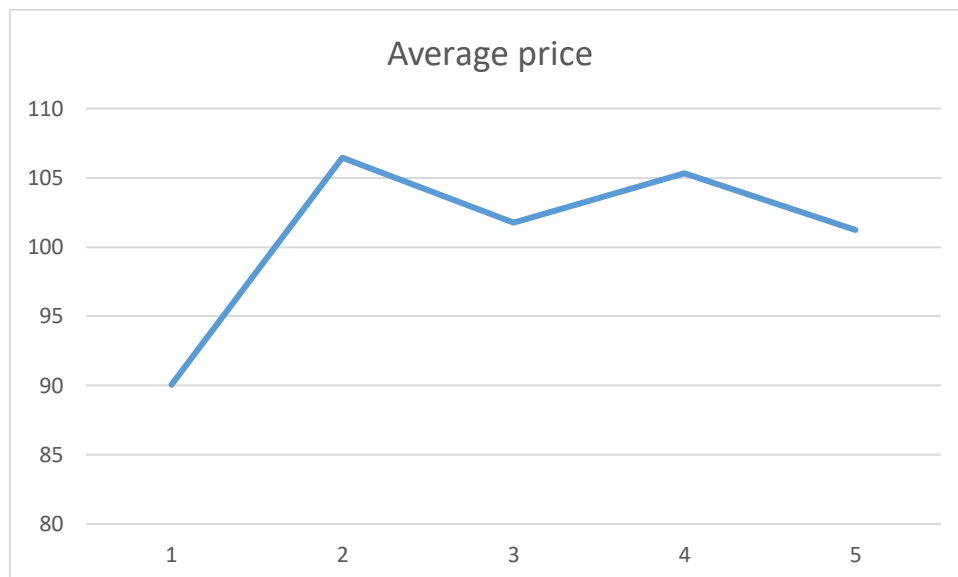


**Figure 4:Chart of the decline or increase in average price**

### Observation

In this observation, one can see that in the past few years, and in especially in the year 2021 the trend of price is declining in great amount. The price is quite low at the beginning of the period but things are getting changed and the prices gradually began to rise. After this, the prices are sometimes going higher and sometimes going lower.

## Evaluation of Data Using Microsoft Excel

Microsoft is a proficient, powerful, and versatile tool that can be used for data or information preprocessing, and analysis and is a great presentation tool. Microsoft Excel is used extensively across various organizations and industries due to its flexibility, wide range of uses, and user-friendly design (Bree and Gallagher, 2016). This article will look at the capabilities of Excel for data processing, analysis, and representation of data.

The first thing in this article is to discuss the preprocessing power of Microsoft Excel. Because of its simplicity, Microsoft Excel is an excellent choice for data preprocessing. The comfortable and easy interface of Microsoft Excel makes information or data organizing and management and analyzing and manipulation of data very easy. It provides a built function that helps users for organizing and clean up the dataset effectively and this is the reason users can easily access a big amount of data and sorted the data from a messy database. It helps the users to minimize the chances of mistakes and other abnormal things (Berenson et al., 2019).

The second thing that comes in the processing of data is the Pivot Table of Microsoft Excel. It is one of the great functions of Microsoft Excel. It allows users to organize the data, classify the data into different categories and make the data easy form to understand. The users and clients can apply different filters to display different kinds of certain data in a group form which gives access to more focused analysis (Bree and Gallagher, 2016). Besides, Excel provides pivot tables with a variety of layout options and chart style that is a great benefit to the users, and the users can customize in a different way to improve the data's visual appeal.

The third good thing about Excel is its “Formula and Functions”. The true potential of Excel lies in its formula and functions. Statistical analysis does not matter whether the analysis is simple or in complex form, excel can solve it like a pro in simple steps. With this help, anyone can make calculations faster and quicker by utilizing the functions (Carlberg, 2014). Investigate and explore the new functions of Excel that assist and tackle specific tasks with ease and quicker.

Excel offers some great built-in capabilities such as SUM, AVERAGE, and COUNT functions that help to determine and identify the fundamental statistics for example the mode, median, mean, average, time series analysis, and standard deviation etc. (Berenson et al., 2019). Excel regression analysis is another helpful tool to analyze data. Through these, users can identify and analyze the metrics between any numbers and can build patterns in the datasets.

For visualizing information in Excel Sparkline is another effective tool. It helps to show the trends data using miniature graphs inside cells. It helps the users to understand how quickly the data is changing frequently over time. Every tool has its limitations so as this tools. One person faces problems dealing with a large amount of data. For this excel is not considered the best choice because as the amount of data increases the charts and graphs are become so messy that it can very hard to identify and understand (Carlberg, 2014).

In short, Excel is a fantastic and effective tool for analyzing the relations between numbers and visualizing data through different formats like sparklines, bar charts, pie charts, etc. Though it is a great tool and many things can be done with the help of Excel but it has also some limitations. It may not be eligible for many sophisticated visualizations. But when it comes to the choice of data representation excel is a very friendly and most popular alternative because of its different tools and functions.

## The Utilization of LOOKUP and PIVOT TABLE in Use of Data

### LOOKUP

There are many reasons to use the LOOKUP function but mainly this function is used for searching and obtaining values from different groups.

The first thing is to open a new Excel spreadsheet and then insert the desired data that want to analyze. For example, the company inserts the desired product data at different times in multiple sections. After this step, determine the product's normal price and use the Excel AVERAGE function to measure. To get the value select a cell of an Excel spreadsheet and apply the function. One has to do this process many times until the person comes achieved each time frame data (Berenson et al., 2019). Once get the time frame data for each period it is ready to use LOOKUP. By applying the LOOKUP function a person can calculate and compare the medians and after that, can easily determine if the price has increased or decreased over time. Through this process, one can easily find out the trends data and make his own decisions (Jelen, 2022).

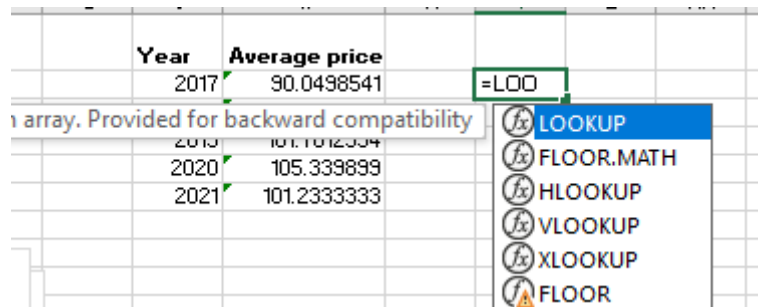


Figure 5: Performing task using LOOKUP

To implement the LOOKUP function the users must enter “=LOOKUP(lookup value, lookup array)” in the selected cell and then select the lookup array which contains the average value of each quarter lookup value means the average value price calculated in the first quarter,

Now it is time to know the result if the price will rise or will fall. If the LOOKUP method is less than the price of the median for the following period, it is considered that the average price has decreased and if the function value is more than the average cost in the next quarter then the average price has also fallen. For a specific time, a person who chose to study has to go through this process again and again.

## PIVOT TABLE

A pivot table is a great tool for data visualization in Excel. This tool is used to figure out if the price will up or down (Jelen, 2022). TO work with this tool first open the new spreadsheet in Excel. Put the input that wish to study. Here insert the data on the costs of an item in different periods for different categories. After that, click the Insert tab and the data area, and from there choose the Pivot Table option. Once created the Pivot Table is done put the value of the prices in the values section and the Period in the rows sections and create filters for filtering the data (Berenson et al., 2019).

After creating the first period within the pivot table choose: show value as” and from, this section select “% difference from”. By doing this a person can measure the growth and fall and through this same process, can apply for the second period as defined as the reference period.

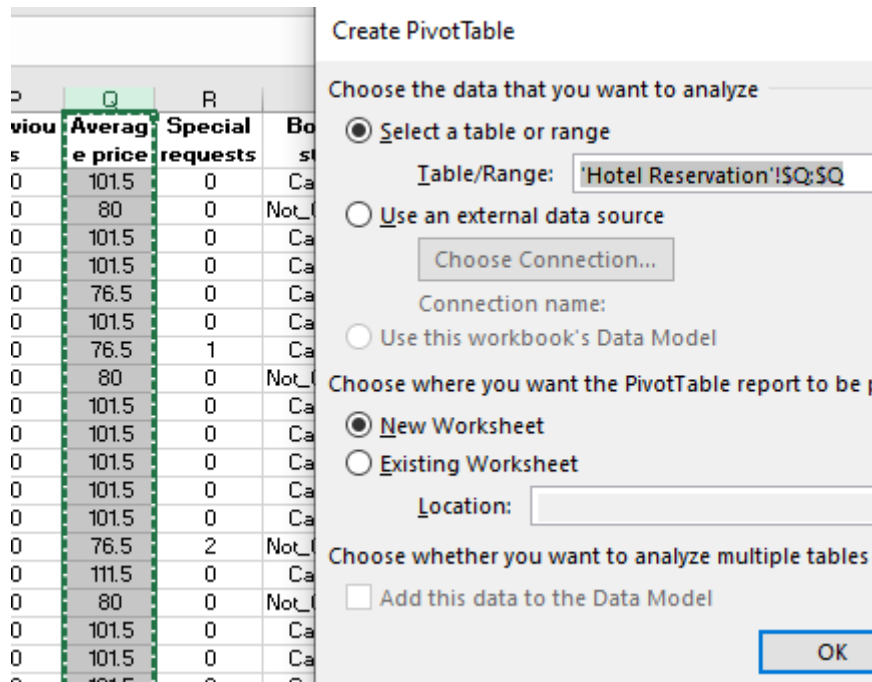


**Figure 6: Performing task using PIVOT TABLE**

The Pivot chart is going to show the percentage difference between the two different periods. In this table, if the percentage is positive then the standard price will be increased and if the is negative then the price will reduce.

Repeat this process at some period wish to analyze, In short, Microsoft Excel's pivot table is a very useful and also very handy tool for measuring, analyzing, and presenting information, and by using of % difference it is easy to figure out the changes in an average price or medium.

## Dataset Analysis of Students Based on The Score

### Number of Male and Female Students

In Microsoft Excel, the users can use the SUMIF function to add numbers in specific cells that meet the particular requirement of the function. The users can follow some steps for using the SUMIF (Zhou, 2020).

The first thing needed is to select the column in which the users wish to show the sumif results. The second thing is that in that column type the “=sumif()” function and select the cell ranges in which the condition will apply. while doing this the users must keep in mind that the condition must meet its needs according to its cells. For example, the users want to total the integers bigger than 5 in cells A1 to A10, so the users use the condition which is “>5”. The number of cells selected by the users that he wants to add. After that, the program will

summation all of the integers that fall inside the range of cells with a specified condition. Once the users finish their work can press enter.

| =SUMIF(H2:H2134, "Female", F2:F2134) |           |                 |            |            |        |          |         |          |   |            |              |   |
|--------------------------------------|-----------|-----------------|------------|------------|--------|----------|---------|----------|---|------------|--------------|---|
|                                      | D         | E               | F          | G          | H      | I        | J       | K        | L | M          | N            | O |
| type                                 | Classroom | Teaching_method | Student_No | Student_ID | Gender | Lunch    | Pretest | Posttest |   |            |              |   |
| public                               | 60L       | Standard        | 20         | 2FHT3      | Female | Does not | 62      | 72       |   | male       | female       |   |
| public                               | 60L       | Standard        | 20         | 3JIVH      | Female | Does not | 66      | 79       |   | 1077       | 1056         |   |
| public                               | 60L       | Standard        | 20         | 3XOWE      | Male   | Does not | 64      | 76       |   |            |              |   |
| public                               | 60L       | Standard        | 20         | 55600      | Female | Does not | 61      | 77       |   | total male | total female |   |
| public                               | 60L       | Standard        | 20         | 74LOE      | Male   | Does not | 64      | 76       |   | 24160      | 24465        |   |
| public                               | 60L       | Standard        | 20         | 7YZO8      | Female | Does not | 66      | 74       |   |            |              |   |
| public                               | 60L       | Standard        | 20         | 9KMZD      | Male   | Does not | 63      | 75       |   |            |              |   |
| public                               | 60L       | Standard        | 20         | 9USQK      | Female | Does not | 63      | 72       |   |            |              |   |
| public                               | 60L       | Standard        | 20         | CS5QP      | Male   | Does not | 64      | 77       |   |            |              |   |
| public                               | 60L       | Standard        | 20         | D6HT8      | Female | Does not | 61      | 72       |   |            |              |   |

Figure 7:: Number of male and female students

### Qualification of students for Reduced/Free Lunch

The users can apply a variety of conditions with the sumif methods and can produce the desired result. This is the reason the users should use the SUMIFS methods. This SUMIFS function gives access to users to provide multiple criteria for aggregating the data and satisfy each criterion according to the user's requirements.

It is important to keep in mind for the users that the SUMIFS function only adds values in a combination that compiles with a specified standard (Salihoun, 2020). If the users want just the summation of all values in a given range and if it does not matter to the users whether it fulfils the condition or not in that time the users should use the SUM function or method.

As the SUMIFS function in Microsoft Excel is a more powerful tool so it can use for measuring and analyzing complex datasets and computing the data based on certain characteristics. This function can reduce time in terms of gathering data on a large scale yet give relevant data.

| =SUMIF(I2:I2134, "Qualifies for reduced/free lunch", F2:F2134) |           |                 |            |            |        |                  |         |          |   |                        |                  |  |
|----------------------------------------------------------------|-----------|-----------------|------------|------------|--------|------------------|---------|----------|---|------------------------|------------------|--|
| C                                                              | D         | E               | F          | G          | H      | I                | J       | K        | L | M                      | N                |  |
| ol_type                                                        | Classroom | Teaching_method | Student_No | Student_ID | Gender | Lunch            | Pretest | Posttest |   |                        |                  |  |
| public                                                         | 60L       | Standard        | 20         | 2FHT3      | Female | Does not qualify | 62      | 72       |   | male                   | female           |  |
| public                                                         | 60L       | Standard        | 20         | 3JIVH      | Female | Does not qualify | 66      | 79       |   | 1077                   | 1056             |  |
| public                                                         | 60L       | Standard        | 20         | 3XOWE      | Male   | Does not qualify | 64      | 76       |   |                        |                  |  |
| public                                                         | 60L       | Standard        | 20         | 55600      | Female | Does not qualify | 61      | 77       |   | total male             | total female     |  |
| public                                                         | 60L       | Standard        | 20         | 74LOE      | Male   | Does not qualify | 64      | 76       |   | 24160                  | 24465            |  |
| public                                                         | 60L       | Standard        | 20         | 7YZO8      | Female | Does not qualify | 66      | 74       |   |                        |                  |  |
| public                                                         | 60L       | Standard        | 20         | 9KMZD      | Male   | Does not qualify | 63      | 75       |   | qualify for free lunch | does not qualify |  |
| public                                                         | 60L       | Standard        | 20         | 9USQK      | Female | Does not qualify | 63      | 72       |   | 22566                  | 26059            |  |
| public                                                         | 60L       | Standard        | 20         | CS5QP      | Male   | Does not qualify | 64      | 77       |   |                        |                  |  |
| public                                                         | 60L       | Standard        | 20         | D6HT8      | Female | Does not qualify | 61      | 72       |   |                        |                  |  |

Figure 8: Number of students qualifying for reduced/free lunch

The users of Microsoft Excel through SUMIFS and Pivot Table can summarise, maintain and analyze a big volume of data.

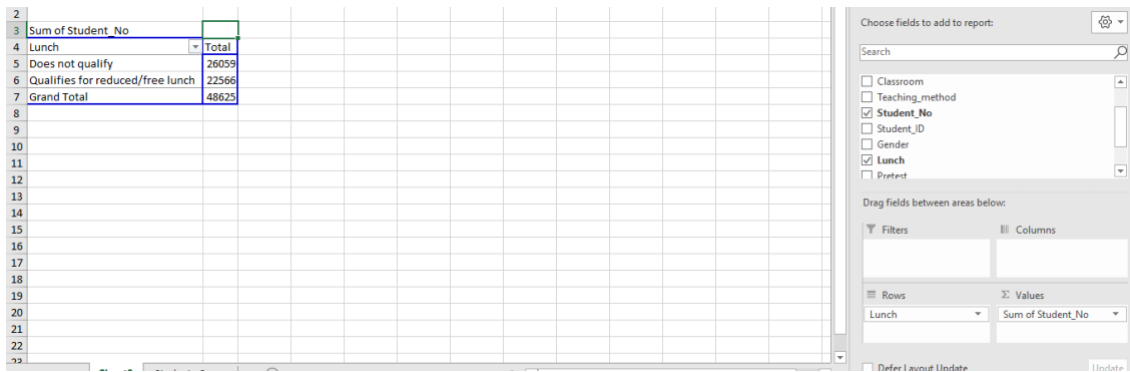


Figure 9:: Number of students qualifying for reduced/free lunch using pivot table

## Mean and Median of the students

The mean, median, and pivot table functions are used simply for measuring the mean, and median of the students.

| Data  |                |                 |                                  |                                    |                                  |                                    |
|-------|----------------|-----------------|----------------------------------|------------------------------------|----------------------------------|------------------------------------|
| ID    | Sum of Pretest | Sum of Posttest | Mean of not qualifying (PRETEST) | Median of not qualifying (PRETEST) | Mean of not qualifying (POSTEST) | Median of not qualifying (POSTEST) |
| 15087 | 61             | 70              | 62.29218107                      | 62                                 | 74.37530864                      | 74                                 |
| 30219 | 66             | 82              |                                  |                                    |                                  |                                    |
|       | 67             | 80              |                                  |                                    |                                  |                                    |
|       | 64             | 76              |                                  |                                    |                                  |                                    |
|       | 40             | 55              |                                  |                                    |                                  |                                    |
|       | 72             | 79              | Mean of qualifying (PRETEST)     | Median of qualifying (PRETEST)     | Mean of qualifying (PRETEST)     | Median of qualifying (PRETEST)     |
|       | 68             | 72              | 45.24618736                      | 44                                 | 57.47603486                      | 55                                 |
|       | 57             | 79              |                                  |                                    |                                  |                                    |
|       | 70             | 79              |                                  |                                    |                                  |                                    |
|       | 59             | 66              |                                  |                                    |                                  |                                    |
|       | 60             | 68              |                                  |                                    |                                  |                                    |
|       | 68             | 76              |                                  |                                    |                                  |                                    |
|       | 43             | 52              |                                  |                                    |                                  |                                    |

Figure 10: Mean and Median of Students

## Pre-Test and Post-Test scores of Students for Differentiation

The mean, median, pivot table, and subtraction or elimination functions are used simply for measuring the mean, and median of the students (Zhou, 2020).

| Data        |                |                 |          |          |          |          |  |
|-------------|----------------|-----------------|----------|----------|----------|----------|--|
| Student_ID  | Sum of Pretest | Sum of Posttest | pretest  |          | posttest |          |  |
| 15087       | 61             | 70              | mean     | median   | mean     | median   |  |
| 30219       | 66             | 82              | 54.95593 | 56       | 67.1022  | 68       |  |
| 59212       | 40             | 48              |          |          |          |          |  |
| 80000000000 | 52             | 68              |          |          |          |          |  |
| 02WIH       | 37             | 56              | DIF      | 1.044069 | DIF      | 0.897797 |  |
| 038FI       | 67             | 80              |          |          |          |          |  |
| 03T0E       | 64             | 76              |          |          |          |          |  |
| 03TIL       | 46             | 49              |          |          |          |          |  |
| 04AKE       | 48             | 54              |          |          |          |          |  |
| 04DG5       | 63             | 84              |          |          |          |          |  |

Figure 11:Differences in students' pre-test and post-test scores

## Use of K-Means Clustering in Analysing SPSS

There is a difference between means and K-means. K means is a popular clustering algorithm method and it is used mainly in cluster analysis in SPSS. In this approach, it divides its given dataset into K groups or K sections where K is the total quantity of clusters which is supplied by the users. In K means the main goal is to reduce the distance between the centre value of the cluster and the value of given datasets items. For comparing the relation the k means clustering method divides its datasets items into a predetermined number of items so that it can easily to comparable with the data sets (Zhou, 2020).

This method distributes each data point to the closest cluster centre once the users select the total number of groups that needs to be created. Once the cluster has been set, the project will continue to improve the cluster centres (Jelen, 2022). The users can utilize SPSS to do k-means clustering and personalize the initial plans of action, and the number of organizations and companies between in this.

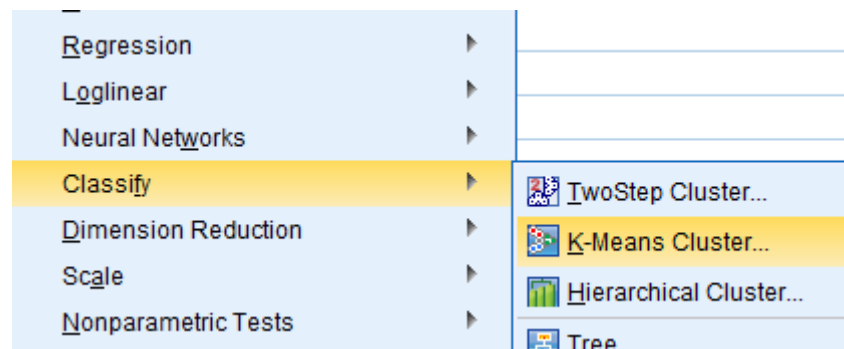


Figure 12: K-Means Clustering in SPSS

### Initial Cluster Centers

|            | Cluster |     |
|------------|---------|-----|
|            | 1       | 2   |
| Student_No | 28      | 17  |
| Pretest    | 23      | 93  |
| Posttest   | 35      | 100 |

### Iteration History

| Iteration | Change in Cluster Centers |        |
|-----------|---------------------------|--------|
|           | 1                         | 2      |
| 1         | 30.372                    | 33.776 |
| 2         | 1.005                     | .898   |

|   |      |      |
|---|------|------|
| 3 | .570 | .494 |
| 4 | .280 | .242 |
| 5 | .183 | .158 |
| 6 | .066 | .057 |
| 7 | .018 | .016 |
| 8 | .000 | .000 |

a. The locations of the cluster centres shifted minimally or not at all in order to achieve convergence. The maximum coordinate change that can occur at any given centre is .000. Currently, there are eight iterations in use. 96.156 kilometres is the shortest distance between two points of origin.

|            | Cluster |    |
|------------|---------|----|
|            | 1       | 2  |
| Student_No | 25      | 21 |
| Pretest    | 43      | 65 |
| Posttest   | 55      | 78 |

**Table 1: K-Means Clustering of data of students SPSS**

| Number of Cases in Each |          |          |
|-------------------------|----------|----------|
| Cluster                 |          |          |
| Cluster                 | 1        | 2        |
|                         | 993.000  | 1140.000 |
| Valid                   | 2133.000 |          |
| Missing                 | .000     |          |

### Methods Use For Common Data Mining and Text Mining

In the modern world and business world there are massive amounts of valuable information, and a big portion of these data are unorganized. This unorganized and big portion of the dataset can be extracted and can be drawn insightful conclusions through the text analysis and some of the data analysis techniques that help to make crucial decisions (McCormick, and Salcedo, 2017). So here discussed several models and methods that are used to mine valuable data in today's business perspective.

**Classification:** By using this method, one can classify and mine data and categories in several points of fresh data (Stine and Foster, 2018). Suppose a shop can try to predict a customer that the customer make an investment or not, so to make this decision he analyzes the traits of that customer, the previous purchases of the consumer, the amount of purchasing of the customer, and the other various factors. By doing this research, the shop may be able to decide whether it gives a discount to the customer and enhance the interaction label or not.

Many companies can use this classification method to identify whether the transaction is fraudulent or not, the companies make their decision by looking at the trends of transaction.

**Clustering: In this method,** similar types of data are identified and organized using the mining techniques or using K-means and it is based on the characteristics of the data (Adriaans and Zantinge, 2019). This method is widely used in the banking system and financial institutions. It divides its data and clustering into several categories based on many factors like the spending habits of the consumers, how they invest, and other essential things (Salihoun, 2020).

This method is also widely used by online retailers. they categorize the consumer's behaviour through different sections and try to identify the pattern and consumers' similarities and after that based on this, they make their decision (Stine and Foster, 2018).

**Association rule mining:** In this data mining system, try to discover and identify the database and the data mining methods that are associated with a particular purpose (Adriaans and Zantinge, 2019). A good example of this is the retailer trying to identify the goods and the services that customers commonly purchase from the store. This mining method helps retailers cross-marketing and position their strategies (Linoff and Berry, 2011). Telecommunication providers also use these methods.

**Text Classification:** This data mining process is commonly used for classifying and organizing textual datasets. That is the reason of called this dataset is Text Classification. The main use of this mining system is to identify any data which are in the form of text (Bluttman, 2022). The main users of this process are the news organization. This type of data mining helps businesses to grow more through its content and make better decisions (Stine and Foster, 2018).

## Conclusion:

This article mainly described the process of a data mining system and how it handles the business and based on the data the business how can make decisions. This is a very important topic. This article discussed how the hotel analysis its data for its reservation systems and sees some charts and graphs of the hotel reservations systems. After this, this article discussed the use of Excel in the process of data mining and management and the little history of Microsoft Excel. Here discussed various functions and tools of Microsoft Excel such as the pivotal table, LOOKUP value, mean and median, k-means, and the charts of these methods. Here also discussed some data mining tools like classifications, clustering, association rule mining, and text classification.

## References:

- Adriaans, P. and Zantinge, D. (2019) *Data mining*. New Delhi: Pearson Education.
- Berenson, M.L., Levine, D.M., Szabat, K.A. and Stephan, D. (2019) *Statistics for managers using Microsoft Excel*. Hoboken: Pearson.
- Bluttman, K. (2022) *Excel Formulas And Functions For Dummies*. S.L.: John Wiley & Sons.
- Bree, R.T. and Gallagher, G. (2016) Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach. *All Ireland Journal of Higher Education*, 8(2).
- Carlberg, C. (2014) *Statistical analysis: Microsoft excel 2013*. Que Publishing.
- Hvitfeldt, E. and Silge, J. (2021) *Supervised Machine Learning for Text Analysis in R*. CRC Press.
- Jelen, B. (2022) *Microsoft Excel pivot table data crunching (Office 2021 and Microsoft 365)*. New York: Pearson Education, Inc.
- Linoff, G.S. and Berry, M.J. (2011) *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- McCormick, K. and Salcedo, J. (2017) *SPSS Statistics for Data Analysis and Visualization*. John Wiley & Sons.
- Salihoun, M. (2020) State of art of data mining and learning analytics tools in higher education. *International Journal of Emerging Technologies in Learning (iJET)*, 15(21), pp.58-76.
- Stine, R.A. and Foster, D.P. (2018) *Statistics for business: decision making and analysis*. Boston: Pearson.
- Zhou, H. (2020) *Learn data mining through Excel: a step-by-step approach for understanding machine learning methods*. Berkeley, CA: Apress L.P.