



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

Programme: BA Business Management

Module code: LCBB5000

Module: Data Handling & Business Intelligence

Assignment Title: Assessment 2

Name: IVO NESHEV

Student ID: 2130931

An Essay on Data Handling and Business Intelligence

Introduction:

The term "**data handling**" describes the steps taken to collect, organise, and display data for the benefit of others. Privacy, security, and long-term storage concerns must be addressed during data management if research data are to be kept in pristine condition (Christian and Winston, 2016). Data storage, retrieval, and deletion may all be optimised via careful design. Business intelligence is a technology-driven strategy used to help upper-level management, middle managers, and front-line workers make informed business decisions (Foreman, 2014).

This essay has two parts. In the first part of this essay, whether the reservation price of a room has been declined or increased will be analysed using Lookup and Pivot table. Besides, the use of Microsoft Excel for pre-processing, analysing, and visualising the data will be addressed.

After the second part of this essay, using tables and charts, different questions will be answered about the students of a school.

Part 1:

Excel, a spreadsheet programme developed by Microsoft, is part of the company's Office suite. Spreadsheet data may be formatted, organised, and calculated with the help of Microsoft Excel (Fischer, 2016). Cells, seen in abundance in Excel, are organised in rows and columns. The contents of these cells are data.

The use of Microsoft Excel for pre-processing the data, analysing the data and visualising the data:

Excel for pre-Processing and analysing the data:

Microsoft Excel is excellent for cleaning and analysing tabular data. Excel's in-built tools and methods for cleaning structured data make this task much more manageable. A ready-made formula that does calculations or computations using a particular value in a particular sequence can be called a function. In Microsoft Excel, built-in functions are used for mathematical calculations such as mean, median, mode and others. Besides that, many functions are used to manipulate or prepare the dataset before conducting calculations (Winston, 2019). Here some of the functions of Microsoft Excel that are used are demonstrated below:

- IF Function

The IF function is one of the most essential functions used in Microsoft Excel. This helps users to compare a dataset or numbers logically and provides two outcomes based on its computation. If the comparison is accurate, it gives the first result; otherwise, it gives the second result. This function is precious for business analysts as it can provide accurate results in less time, accelerating the decision-making process. Besides that, for enhancing the productivity of a business or process, this function helps as one can search data from a small data set manually, eliminating the chances of mistakes. However, they may make mistakes if they can not realise the logic (Fischer, 2016). This function performs a test and then returns the result. Checking if "is two greater than 3" is the condition tested by the function in the code snippet below. The function will return "2 > 3" if the condition evaluates to TRUE (which it does not) and "2 < 3" otherwise.

```
=IF (2>3,"2 > 3","2 < 3")
```

Given that two does not exceed 3, the result will be "2 < 3" as the condition is false.

- IF ERROR:

Error handling is easier with this feature. #N/A, #VALUE!, #DIV/0!, #REF!, #Mention?, #NUM!, to mention a few, are among Excel's most often encountered errors (Humphrey, 2017). Let us try dividing one by 0 as an example of managing errors.

=IFERROR (1/0, -1000)

One has been subtracted from zero in the preceding example. The result of this computation is an invalid value (#DIV/0!). If the supplied computation or formula does not produce an error, the IFERROR function will return the calculated value. In the event of an error, the value we provided will be returned (-1000 in the given example) (Alexander, 2011).

Errors would be returned by any equations that used the flawed computation if the problem were not addressed. For instance, if we used the SUM function to add ten integers together, and #DIV/0! was one of those numbers, the SUM function would also return #DIV/0!

- **LEFT:**

This method will return the first 'n' characters for a given string. Here is a sample that takes the first three characters of the string "Excel" and returns them. The value "Exc" will be returned by the function.

=LEFT("Excel",3)

- **RIGHT:**

A string's final 'n' characters are returned by this method. The final three characters of the string "Excel" need to be extracted, so let us see how it works with an example. "cel" will result from the function. =RIGHT("Excel",3)

- **Look UP Function:**

The lookup feature is a part of the Microsoft Excel reference and lookup functions. It allows a user to search for an appropriate or approximate match within a single column or row and gets value from another row or column radius. If a business analyst needs to evaluate two different rows or columns of financial data, they use this function. Two variants of Lookup functions are VLOOKUP and HLOOKUP. This function helps users to save time by providing data searching the whole table automatically. As a result, business analysts can get important information quickly, saving time and reducing error chances (Humphrey, 2017). Therefore, they can make errorless decisions in less time. Besides that, one of the main positive aspects of that function is its easy use. Thus, analysts can avoid complex calculations. However, this function also has some disadvantages. Among them, the main disadvantage is that if the user needs to understand the concept of this function, the user can make errors that can result in wrong results, which can lead to wrong choices while making decisions. Below the figure represents that function:

- **FIND:**

Input a substring or character; this function will return the index of the first occurrence in the string. Determine where in the string "Excel" the substring "ce" begins by looking at a sample. The function always returns 3, as that is where "ce" appears once (and only once) in the input.

=FIND ("ce", Excel") (McFedries, 2019)

A user can also set the search's initial position with the FIND function. Find the location of the second "a" location in the string "aba" by using the example below. Since we do not want to look for the first 'a' in the string, we will provide a beginning point parameter 2. It is a 4-returning function.

=FIND("a", "abca",2)

The search in FIND is case-sensitive. However, another function called SEARCH is identical to FIND except that it ignores the case. If the given substring or character is absent in the string, the #VALUE! The FIND and SEARCH functions return errors.

- **PIVOT Table:**

A pivot table serves the purpose of organising, sorting, reorganising, grouping, calculating totals, and finding the average data recorded in a table. It helps users to turn rows into columns and columns into rows and allows them to create groups and computations. This has some advantages, which include management. This table helps to manage data by sorting, reorganising and other, which help to manage the data quickly and gather data whenever needed easily. Planning is an essential part of any business, and Pivot table can help that business to organise its data effectively, which helps for easy management of the firm. This function can help users create personalised tables from large datasets by altering data so that current trends in the data can be found. This helps users by predicting data (McFedries, 2019). Another advantage of the PIVOT table is that it can help users build efficient reports or presentations by allowing them to report automatically. Thus, users can save time and effort. These are a few main functions used in Excel for pre-processing, cleaning, and analysing the data in Excel.

Usage of Excel for presenting the data:

Charts and graphs are the visual representation of data that can help to understand the information quickly. This can represent data as pictures which can be challenging to understand just by looking at the numbers. They provide a visual way to analyse and present data from an Excel sheet (Kusleika, 2021). The following represents the charts that have been used in Excel:

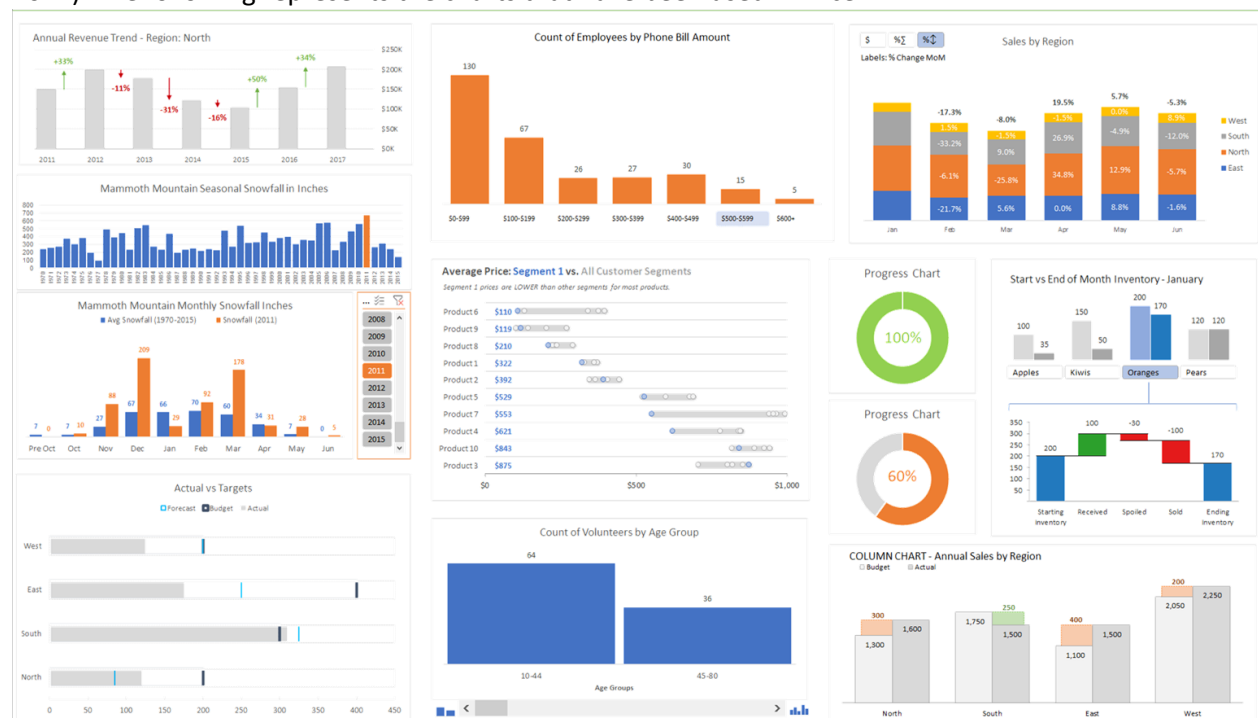


Figure 1: Charts for Pre-processing the data

Source: (Kusleika, 2021)

Application of the PIVOT Table using hotel reservation data:

PIVOT Table:

The screenshot shows an Excel spreadsheet with a PivotTable summarizing hotel reservation data. The PivotTable is located in the range A3:B10. The PivotTable Fields task pane is visible on the right side of the screen.

Arrival_year	Total
2017	586095.9
2018	3164525.6
2019	425.78
2020	325.6
2021	316
Grand Total	3751688.88

The PivotTable Fields task pane shows the following configuration:

- Choose fields to add to report:**
 - ☒ Arrival_year
 - ☐ Arrival_month
 - ☐ Arrival_date
 - ☐ Market_type
 - ☐ Repeated_guest
- Drag fields between areas below:**
 - Filters:** (Empty)
 - Columns:** (Empty)
 - Rows:** Arrival_year
 - Σ Values:** Sum of Average price

Presentation of the data:

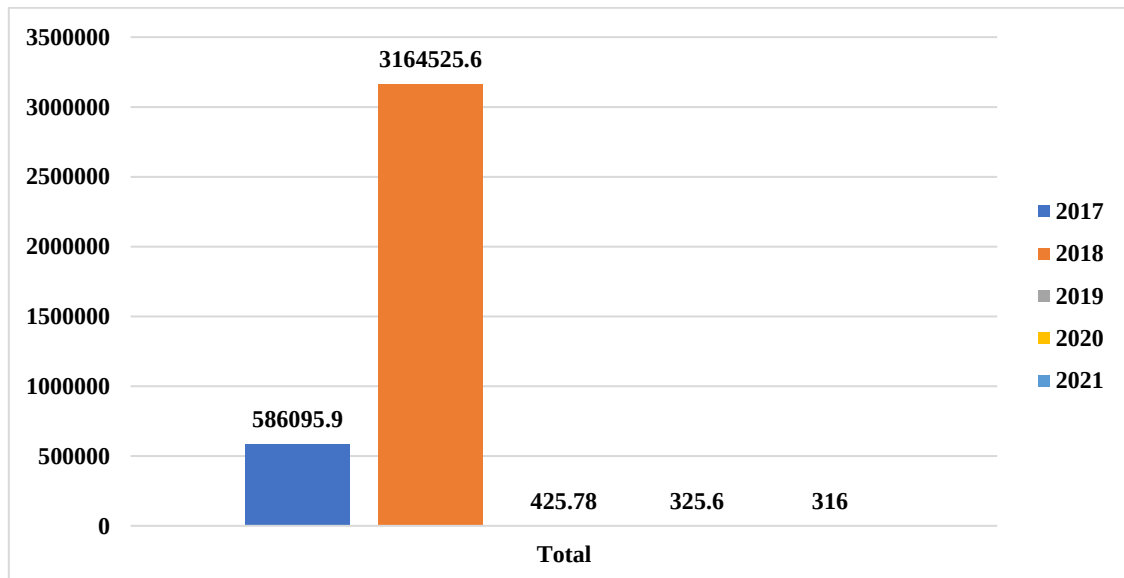


Figure 2: Presentation of the sum of the average price

Source: (Self-Made)

By looking into the data, it can be seen that the average price of the hotel room declined after 2018 but increased after 2017. This is because the covid-19 pandemic had been spread all over the world.

Task 2:

Male and Female students in the class:

C4																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Gender																		
2	Female																		
3	Female	Female	1056																
4	Male	Male	1077																
5	Female																		
6	Male																		
7	Female																		
8	Male																		
9	Female																		
10	Male																		
11	Female																		
12	Male																		
13	Male																		
14	Male																		
15	Female																		
16	Female																		
17	Male																		
18	Male																		
19	Male																		
20	Female																		
21	Male																		

Figure 3: Male and Female student

Source: (Self-Made)

Using COUNTIF Formula, Male and Female students in the class has been determined, and it has been found that Female student is 1056 and male student is 1077.

A student qualifies and does not qualify for reduced/free lunch:

C5																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O			
1	Does not qualify																	
2	Does not qualify																	
3	Does not qualify	Does not qualify	1215															
4	Does not qualify	Qualifies for reduced/free lunch	918															
5	Does not qualify																	
6	Does not qualify																	
7	Does not qualify																	
8	Does not qualify																	
9	Does not qualify																	
10	Does not qualify																	
11	Does not qualify																	
12	Does not qualify																	
13	Does not qualify																	
14	Does not qualify																	
15	Does not qualify																	
16	Does not qualify																	
17	Does not qualify																	
18	Does not qualify																	
19	Does not qualify																	
20	Does not qualify																	
21	Does not qualify																	

Figure 4: Student qualifies and does not qualify for reduced/free lunch

Source: (Self-Made)

Using the Count IF function, the student who qualified for free lunch and the student who did not qualify has been calculated. The following figure represents it:

C3 =COUNTIF(A1:A2133,A1)															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Does not qualify														
2	Does not qualify														
3	Does not qualify	Does not qualify	1215												
4	Does not qualify	Qualifies for reduced/free lunch	918												
5	Does not qualify														
6	Does not qualify														
7	Does not qualify														
8	Does not qualify														
9	Does not qualify														
10	Does not qualify														
11	Does not qualify														
12	Does not qualify														
13	Does not qualify														
14	Does not qualify														
15	Does not qualify														
16	Does not qualify														
17	Does not qualify														
18	Does not qualify														
19	Does not qualify														
20	Does not qualify														
21	Does not qualify														

Figure 5: Application of Count IF

Source: (Self-Made)

The number of students who do not qualify for reduced/ free lunch is 1215, and those who qualify for reduced/ free lunch is 918.

The mean and median of students who qualify and do not qualify for reduced/free lunch:

The mean of students who qualify and do not qualify for reduced/free lunch:

H13													
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Lunch	Pretest	Posttest										
2	Does not qualify	62	72										
3	Does not qualify	66	79					Data					
4	Does not qualify	64	76				Lunch	Average of Pretest	Average of Posttest				
5	Does not qualify	61	77				Does not qualify	62.29218107	74.37530864				
6	Does not qualify	64	76				Qualifies for reduced/free lunch	45.24618736	57.47603486				
7	Does not qualify	66	74										
8	Does not qualify	63	75										
9	Does not qualify	63	72										
10	Does not qualify	64	77										
11	Does not qualify	61	72										
12	Does not qualify	61	73										
13	Does not qualify	64	74										
14	Does not qualify	66	78										
15	Does not qualify	60	71										
16	Does not qualify	64	77										
17	Does not qualify	64	73										
18	Does not qualify	63	70										
19	Does not qualify	67	73										
20	Does not qualify	63	71										
21	Does not qualify	64	73										

Figure 6: Mean of the student who qualifies and does not qualify for reduced/free lunch

Source: (Self-Made)

PIVOT Table has been used for calculating this. The following figure represents it:

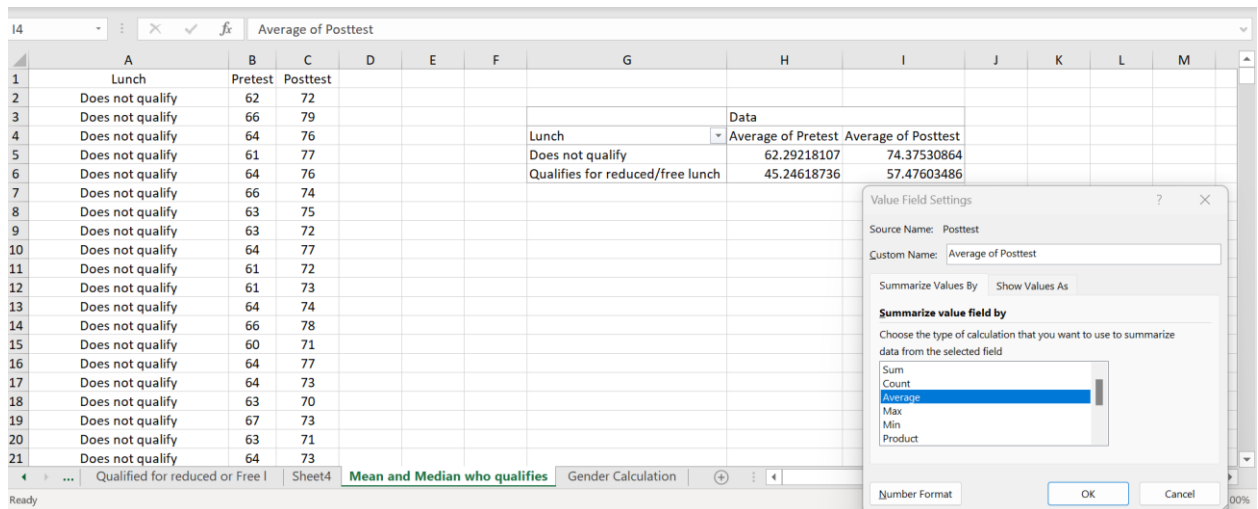


Figure 7: Application of the PIVOT table for calculating the mean of the student who qualifies and does not qualify for reduced/free lunch

Source: (Self-Made)

The median of students who qualify and do not qualify for reduced/free lunch:

The median of the student who does not qualify for reduced/ free lunch is 65 and 74 for the pretest and post-test, and the median formula has been used for calculating this. The following figure represents it:

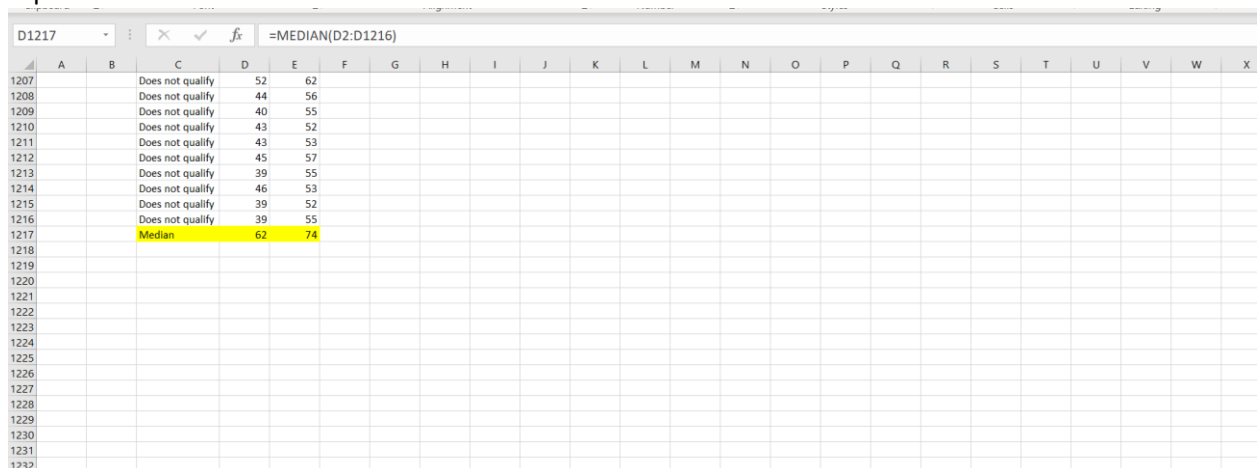


Figure 8: Median of students who do not qualify

Source: (Self-Made)

The median of the student who qualifies for reduced/ free lunch is 44 and 55 for the pretest and post-test, and the median formula has been used for calculating this. The following figure represents it:

Qualifies for reduced/free lunch	40	47		
Qualifies for reduced/free lunch	41	48		
Qualifies for reduced/free lunch	38	48		
Qualifies for reduced/free lunch	42	55		
Qualifies for reduced/free lunch	41	49		
Qualifies for reduced/free lunch	42	50		
Qualifies for reduced/free lunch	38	46		
Qualifies for reduced/free lunch	40	46		
Qualifies for reduced/free lunch	41	50		
Qualifies for reduced/free lunch	45	47		
Qualifies for reduced/free lunch	38	51		
Qualifies for reduced/free lunch	39	50		
Qualifies for reduced/free lunch	38	46		
Qualifies for reduced/free lunch	45	51		
Qualifies for reduced/free lunch	46	53		
Qualifies for reduced/free lunch	41	48		
	44	55		

Figure 9: Median of students who qualifies

Source: (Self-Made)

For sorting students who qualify and do not qualify have been sorted by using the filter function.

K means clustering:

K-means in data mining and statistics, clustering is a typical analysis method. The purpose of this method is to use Voronoi cells to split a set of observations into many clusters (k). It is a method for establishing the correct classification of an item.

It has found its most common use in statistics, although its potential applications are limitless. For faster and more efficient marketing, it might be used to categorise people from different demographic backgrounds into manageable chunks (Foster, Barkus and Yavorsky, 2006). Researchers in the field of astronomy use it to sift through mountains of data in order to find interesting objects or phenomena to study in greater depth. The following figure represents an example of the cluster:

Initial Cluster centers			Iteration History			Final cluster centers			Number of Cases in each cluster		
	Cluster		Iteration	Change in Cluster Centers			Cluster		Cluster		
	1	2		1	2		1	2	1	49	
Customer Code	4	100	1	23.812	24.499	Customer	26	76	Valid	49	
Gender	1	2	2	0.511	0.515	Gender	1.51	1.49	Missing	0	
Age	13	13				Age	20.67	20.02			
Eat Rice	1	1				Eat Rice	59	61			

Figure 10: Example of K-means clustering

Source: (Warner, 2020)

Convergence was achieved with minimal or no shifts in cluster centres. The most considerable absolute coordinate change for any given centre is.000. This is the third revision. The iminium distance between the original centres is 96.005.

Data Mining and text mining methods that can be used in contemporary real business:

Data Mining Methods:

One of the most common data mining uses is correlation analysis, which determines how closely related two variables are using a linear distance metric. This method of data analysis determines the extent to which one variable has changed due to a change in another variable. When there is a high degree of correlation between two variables, they are more likely related. Changes in advertising costs can affect the volume of sales, and seasonal weather might affect ice cream demand (Christian and Winston, 2016).

In data mining, classification is gathering numerous qualities and organising them into meaningful groups for subsequent study. Businesses may place a consumer in one of three categories: low, medium, or high, depending on their purchasing habits and level of disposable income. Using these groups, we can learn more about our patrons.

In contrast to other data mining methods, which strive to unearth previously unseen patterns and trends, outlier identification aims to isolate what makes a single data point different from the rest of the sample. Errors in data entry or sampling are only two examples of mistakes that might be uncovered using outlier detection (Pang-Ning Tan, Steinbach and Vipin Kumar, 2005). The inherent irregularities of raw data can reveal interesting insights.

Cluster analysis is the process of categorising diverse data points into groups with similar characteristics. This allows data miners to make more informed judgements about large groups of people and their behaviours (such as consumers or users) by breaking down enormous data into smaller subgroups.

With the use of computers, artificial neural networks attempt to mimic the human brain's performance. Using many processors (mimicking the brain's neuronal structure), neural networks attempt to learn, reason, and analyse data as humans do. There is a wide range of applications for neural networks. Marketers might use them to zero in on specific demographic groups and anticipate consumer spending patterns. Merchants may also gain insight into dynamic pricing and sales forecasting with their help (Linoff and Berry, 2011). They also contribute to enhanced patient diagnosis and treatment methods, leading to better outcomes.

Computational marketing is a relatively young discipline that applies data mining and computational techniques to marketing. Information chunks too large for human analysis can be given structure and meaning through the computation process.

Individuals expose themselves when they engage in activities on the internet. Their identity is revealed through the data they provide on social media and the websites they visit. Marketers want this data because it helps them better understand consumer habits and tastes (Pang-Ning Tan, Steinbach and Vipin Kumar, 2005).

Computational marketing enhances the consumer buying journey by applying analytical tools, including algorithms, mathematical equations, software, and more.

Text Mining Method:

"text mining" refers to extracting useful information from structured textual data. Texts, emails, and files written in a widely spoken language supply this data. Its primary function is to help users get insight from accumulated data.

Businesses may use word frequency to determine which phrases or ideas are most prevalent in the data. Finding the most frequently used terms might help analyse consumer feedback, social media interactions, and reviews. Reviews that frequently include phrases like "expensive," "overpriced," and "overrated" may indicate the need to adjust their rates (Aggarwal, 2015).

Data such as keywords, entity names, addresses, email addresses, and more may be extracted from a text using a text analytics technique called text extraction. Businesses may save time and effort by employing text extraction software instead of manually going through data (Peter, 2020).

When analysing the substance of a piece of writing, keywords are the most valuable tools for businesses. They may develop tag clouds, index searchable data, and generate summaries of text with the help of a keyword extractor.

Businesses may use it to find and remove company, organisation, or person names from any text.

It helps identify details about a service or product in a large data set. Information such as a product's colour, brand, model number, etc., may be readily extracted from the provided data.

Advantages and disadvantages of Excel and SPSS in businesses:

Advantages and Disadvantages of SPSS:

Advantages	Disadvantages
▪ It provides accurate and quick results. ▪ The tables and graphs are interactive and helpful. ▪ Many more individuals can use it because it is available in many languages. ▪ Managing information effectively ▪ The learning curve for this programme is low. ▪ It works well with both quantitative and qualitative information. Companies are looking for information that will help them take decisive action. Businesses create massive amounts of data, and a human search could be more efficient. ▪ SPSS allows market researchers to use consumer data by doing complex statistical studies. Practical tools for analysing survey data provide reliable insights into market tendencies.	One of its significant drawbacks is SPSS's inability to handle a large data set. In specific disciplines, a veritable gold mine of information already exists. It is possible that SPSS is not the optimal choice in such fields. The results of any statistical study are suspect if the data were gathered using unreliable or biased procedures. There should not be a problem if the discrepancy between the two is negligible. However, if the disparity is significant, the results will be unreliable.

Table 1: Advantages and Disadvantages of SPSS

Source: (Field, 2020)

Advantages and disadvantages of Excel:

Advantages	Disadvantages
▪ Microsoft Excel is widely used for such tasks as data entry and analysis because of how quickly and easily it can be used to input as much data as needed. Is. Filtering information is an easy and helpful feature of Excel. ▪ It is simpler to retrieve data: Compared to an electronic Excel spreadsheet, searching for information printed on paper might be time-consuming. The	In terms of data and expansion, one of the significant areas for improvement of Microsoft Excel is that it becomes unusable after adding a few thousand rows. The maximum allowed file size is likewise limited. Nonetheless, much like any other app or software tool problem, there are workarounds.

data is easily accessible and may be found. ▪ Useful, Straightforward Mathematical Formulas: The Formulas Tab in Microsoft Excel has made calculating a breeze. ▪ Highly dependable: Protected by a laptop or desktop computer password, these spreadsheets are much less likely to go missing than hard copies of records or slips of paper containing data. Microsoft Excel is a computer program that uses spreadsheets to analyse and organise information. Excel is used in business for various purposes, including budgeting, analysis, forecasting, trend detection, and reporting.	
--	--

Table 2: Advantages and Disadvantages of Excel

Source: (Fischer, 2016)

Conclusion:

Data management refers to keeping track of and accounting for research data before, during, and after a study is completed. With careful preparation, data storage, retrieval, and deletion may be optimised for speed and economy. The sole purpose of business intelligence is to facilitate and support improved decision-making inside an organisation. Businesses may improve their operations in many different areas thanks to the information made available to them by BI systems. Since they help with data analysis and improved decision-making, data mining and business intelligence have a promising future in business analytics (Christian and Winston, 2016). The essay stated that Excel is crucial for data analysis since it reveals market trends. This makes it straightforward to identify trends in the company's financial data, which can be used to generate educated guesses about its financial health.

References:

- Aggarwal, C. C. (2015) *Data Mining: The Textbook*. Cham: Springer.
- Alexander, M. (2011) *Excel 2007 Dashboards and Reports for Dummies*. John Wiley & Sons.
- Christian, S. and Winston, W. L. (2016) *Business Analytics: Data Analysis and Decision Making*. Boston, Ma: Cengage Publishing.
- Field, A. P. (2020) *Discovering Statistics Using IBM SPSS Statistics*. New Delhi, India: Sage Publications.
- Fischer, W. (2016) *Excel*. CreateSpace Independent Publishing Platform.
- Foreman, J. W. (2014) *Data Smart: Using Data Science to Transform Information into Insight*. Hoboken, New Jersey: John Wiley & Sons.
- Foster, J. J., Barkus, E. and Yavorsky, C. (2006) *Understanding and Using Advanced Statistics*. London; Thousand Oaks, Calif.: Sage Publications.
- Humphrey, M. L. (2017) *Excel for beginners*. 1st ed. CreateSpace Independent Publishing Platform.
- Kusleika, D. (2021) *Data Visualisation with Excel Dashboards and Reports*. John Wiley & Sons.
- Linoff, G. and Berry, M. (2011) *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons Inc.
- McFedries, P. (2019) *Excel Data Analysis for Dummies*. Hoboken, NJ: John Wiley & Sons, Inc.
- Pang-Ning Tan, Steinbach, M. and Vipin Kumar (2005) *Intro to Data Mining*. New Jersey: P. Ed Australia.
- Peter, B. (2020) *DATA MINING FOR BUSINESS ANALYTICS: Concepts, Techniques and Applications in Python*. SL: Wiley-Blackwell.
- Warner, R. M. (2020) *APPLIED STATISTICS: Basic Bivariate techniques*. S.L.: Sage Publications.
- Winston, W. L. (2019) *Microsoft Excel 2019: Data Analysis and Business Modeling*. New York, New York: Published with The Authorization Of Microsoft Corporation By Pearson Education, Inc.