

University of Wales Trinity Saint David



Module name: Data Handling & Business Intelligence

Module code: LCBB 5000

Assignment Title: Assessment 2-Data Analysis & Visualisation

Student name: Yordan Aleksandrov Yordanov

Student ID: 2028612

Submission date: 10/01/2023

Lecturer name: Gasso Wilson

Contents

INTRODUCTION 1

MAIN BODY 1

Big data and its characteristics 1

CONCLUSION 1

REFERENCES 2

INTRODUCTION

Data analysis is the term given to the whole process that involves analysis of data to end up to a meaningful decision from the raw data derived from various sources. This brings from to unorganised data. The main tool used for this is MS Excel , in which Pivot table is one of the favourites of various users , this table helps extract required data from a vast set and facilitates calculations therein.(Nizam, Zhang and Tian, 2018)

In this report , the provided data set – Superstore has been analysed to answer questions which would've been impossible without the use of analytical tools which have been briefly described in this report. In the second part , the next data set has been visualised with the help of tools like bar graph and line chart.

MAIN BODY

Multiple sheets can be prepared in excel. It has multiple functions and programs for presentation and calculation of data.

Pre-Processing of data – The User-Interface of excel is comprehensive to use. Data in raw form is of no use. That is why after collecting data it is arrayed so that similar data segments are compiled together to form multiple segments of data. Data is arrayed, cleaned, transformed, modified and reduced in excel. (Rohman, Sata, Othman and et.al., 2021) For example information about students in a university was collected regarding

their names, courses, contact no. , fees paid, guardian etc. if these info' remains like such a form of compilation or such bulk of unarranged form it is of no use. So different segments of fees, courses, guardian will be prepared regarding every student. So that it is easy to find information of any particular student.

Data Analysis – As the name sounds data analysis is scrutinizing a certain set of info. Their are various functions in excel like IF, IFERROR, LEFT, MID, RIGHT, FIND and many more. These helps in forming fast decision making and proper conclusion drawing.(Zhou, Chen, Huangyang and et.al., 2022)

Visualising Data – Presenting data in fast understand way , excel allows the user to prepare graphs charts , tables and many more. To make data representation effective.

VLOOKUP Function – It is a important data analysis function in Microsoft Excel. Information in Microsoft excel is stored in rows and columns in sheet. Sometimes data around any particular topic can be quite large. Implying that number of rows and columns in excel are quite large. In this scenario if a single value or small piece of information is to be found. That becomes quite tough and lengthy for the information seeker.(Hristova, Bogdanova and Netov, 2022) It is used in two forms Vector form and Array form. Their are other functions similar to it HLOOKUP and LOOKUP.

Formula of VOOKUP Function – VLOOKUP(lookup_vale, lookup_vector, [result_vector])

The Profit of Superstore extracted through VLOOKUP fuction -

VLOOKUP	
Enter Row ID	80

get profit-	1198.97
-------------	---------

Decrease or increase in value of sales and profit for each year	
sales	profit
0	0
-649810	-69178.7
-130142	15393.17
285727	-37866.4

PIVOT table – It is a user-oriented function of Microsoft Excel. Many a time data set around a subject are quite large for example profits of last ten years, students in a university, guests in a large event, products in departmental store etc. In these type of large bulk of information, it becomes very hard to extract a small size of information. Towards these problem of data extraction the function of pivot table helps. It is used to draw out data of some rows or columns from a large excel sheet.(Trivedi and Chen, 2018)

Given the information of Superstore following the below process it was prepared Pivot Table for sales and profit.

Open the data file of Superstore. Then choose the set of cells that are having sales and profit values. After selecting cells, click on insert pivot table.

Select the Table range for which data is required.

Select whether Pivot table is to be placed in existing sheet or in new sheet.

Then place a final click on OK.

Pivot Table			
Row Labels	Sum of Sales	Sum of Profit	Sum of Order Quantity

2009	4209896.846	434096.02	54508
2010	3560087.045	364917.33	54379
2011	3429944.981	380310.5	51413
2012	3715671.953	342444.13	54480
Grand Total	14915600.82	1521767.98	214780

Analysing the Pivot Table -

Sales revenue – The trading revenue was 4209897 in 2009 then severely decreased to 3560087 in 2010. In the year 2011 it fallen more to 3429945. Then the sales revenue escalated to 3715672 in 2012.

Profit – The store a good profit of 434096 in 2009, which fell considerably in 2010 to 364917. The Profit then increased a little in the year 2011 to 380310 then again fell to 1521768 in 2012.

Trend – From the pivot table we computed decline and increase in profit for every year. As 2009 was the base year their was no year to compare it with so the change in sales and profit is 0. then fell with large amount of 649810 in 2010, then by 130142. then from 2011 to 2012 it increased with 285727. (PylaeV, Vanzha, Avdeeva and et.al., 2019)looking at the profit trend it is found out that profit fall with 69178, then increased by 15393 then again fall by 37866.

Chart -

Charts refers to plotting, expressing and presenting information on a graph. They are used to increase the understanding and reading of data fast. Chart are very helpful when amounts or digits of data and variables are quite large Some types of charts are histogram, pie chart, bar chart, flow chart, Open-high-low-close chart and many more. From the Superstore data about for the sales made. Their was prepared table for the orders placed from different regions of the world. Then was prepared Pie chart from the table

prepared for preparing the chart following steps were followed up to construct Chart in Microsoft excel -

Click on Insert icon in the excel sheet.

Then click on charts and select the pie chart from the listed charts. Select the data of various regions given in the tabular form as shown in the excel.

Plot the data of table in the Pie chart.(Tonetto, Fogliatto and Anzanello, 2020)(

The chart can further modified using the options of Chart elements, Chart filters, Chart styles etc.

Description of Chart – The highest amount of orders came from Ontario, secondly from West, thirdly from Prairie and fourthly from Atlantic .

Line Chart representing data from the pivot table

Region	Total number of orders
Atlantic	1080
North Carolina	79
Northwest Territories	59
Ontario	2161
Prarie	1706
Quebec	781
West	1991
Yukon	542

Graph -

Graph refers to elaboration and presentation of data in a way that it can be understood through optical senses. Some type of graphs are bar/column graph, pie graphs, line chart, scatter plot etc.

Graph are used in a variety for representing profits, sales, revenue, clients data, calculations, performance of organizations. They assist in data visualisation and relationship establishment of data variables.

Using the pivot table made it was made Line Graph for the Superstore – The graph very comprehensively described and showed the revenue and gaining of the enterprise.

Process of using Graph in the excel sheet -

Click on Insert option in the spreadsheet.

Click on the type of that that serves the purpose of survey or data analysis.

Select the data of data that you want plot in the selected graph. Otake,)(Kouotou, Ghalit, El Ouahabi and et.al., 2021)

The style, font, elements, shape and other characteristics can be modified through other option.

Lastly press the OK icon to the final form of graph.

2.1 How many students like vanilla flavour of ice-cream

2.1	
COUNT	FAV. FLAVOR
95	Vanilla
47	Chocolate
58	Strawberry

2.2 How many students are Male and Female?

2.2	
91	Male
109	Female
200	total

2.3 What are the Mean and Median of participants who likes chocolate and strawberry flavour ice-cream?

2.3	
52.5	Mean
63	Median of Chocolate
101.5	Median of Strawberry

2.5 Explain the most common data mining and text mining methods that can be used in business with real world examples.

Data mining and text mining methods used in businesses

The word given to the automation of the process of selection and analysis of huge amounts of data or text to asses in searching, discovering relationships, screening patterns and finding meaningful insights from the data id Data and Text Mining (DTM). The tools mentioned below are widely used- **SAS data mining**- Statistical Analysis System is a tool which helps in management of various data sets from ample sources by helping the user in changing , mining , and managing the data. This also offers a graphical user interface specially designed to help non-technical users. The tool is used to reach to decisions quickly by serving the purpose of mining and optimising the data too , it can also remove errors , detect duplication , clean or fill null values in the sheet.

Orange data mining- this tool is the perfect data mining software as it is based on components written in Python computing language (Nakayama and et.al., 2020). The various

components of it are called 'widgets' , widgets can be used for various tasks like – reading data , visualising data elements , training and comparing learning outcomes , displaying etc.

Aylien- this is one of the most used tools as it has automated the process of comprehending human made forms and documents with the help of AI , language processors and machine learning. Aylien's text mining tools- categorisation , topic discovery and extraction etc. with the help of its text analysers make it easy to use and fruitful.

Thematic- Powered by AI , this tool serves the purpose of transforming feedbacks received from customers into data that is understandable and actionable. Thematic intelligence and Thematic catalyst are 2 products wherein the first aids in unveiling the true sense of the feedback and then proceeds to group similar phrases , in the end it segregates them into positive or negative categories respectively. This is done by forming relationships and patterns among large number of feedbacks which then are analysed by creation of sub themes , applying metrics and grouping customers to get a deep insight about the problems and their solutions. The second product assists in formation of dashboards to represent the data uncovered which in turn helps in convincing the management and other decision makers involved to take quick and proper actions based on the customers complaints.

Comprehend by Amazon- This tool aids the user in quick detection of sentiments , creating model topics from ambiguous message content which in turn helps in revealing useful insights in terms of emotions about people , events and places by finding relationships between various texts , phrases etc.

Google cloud NLP- this platform is used to extract insights from varied texts with the help of AI and machine learning. Meaningful information about places , people , events etc. can be derived by analysing and integrating texts with document storage on Cloud Storage. Semantic analysis and syntax are the techniques mainly used along side NLP to arrange words in an order that makes

sense grammatically and to assess meaning from the syntax assembled.

2.6 Discuss the advantages and disadvantages of using SPSS and Excel using relevant practical examples and present your discussions in a table.

SPSS

SPSS stands for Statistical Package for the Social Sciences. It is software or tool which was created by the IBM company. It helps with the research, experiments, scrutinizing of data. It assists in efficient management of finance and administration. SPSS is a tool or say software which is developed by IBM which enable to perform various models starting from very minimal to even a complex one. SPSS is data extraction tool for finding data from large piece of information. Under it numerical estimation of data through using formulas, trigonometry is performed.

Below are listed Advantages and Disadvantages of SPSS -

Advantages	Disadvantages
The users of data gets their time saved in these software, as tasks are performed in few seconds and minutes.	It can not detect and resolve personal bias and human mistakes made by the users of SPSS.
SPSS brings precise and exact results in computations made.	SPSS is not that useful for extracting info in very large information bulk.
Extraction can be made from big size of data set.	This somewhere questions the use and implementation of the software. When a person uses these software it just give the end result in form of computations it doesn't establishes relationship and reasons.
It save time consumed in tasks thus helping in efficient functioning of organisation.	SPSS is not that well-grounded software, implying that it cannot be totally dependent on.

Tranferring data in SPSS is quite facilitating for the users and thus considered more suitable.	Extensive training needs to be given for making people use of these software which increases human resource management cost. It is not that easy to use and user-friendly.
It doesn't convert values in approximate form.	Costly especially for low income groups and learning students.
All forms of data economical, techinal, environmental are equally given importance.	Preparing charts and graphs in SPSS is quite difficult.

Excel

Microsoft Excel is a component of Microsoft office that is majorly used for quantitative analysis. It is widely used software for numerical computation. It encompasses a lot mathematical formulas that are very helpful when data regarding various topics like sales, expenditures, profits, percentage is stored. The functioning is so facilitating that calculations for a large set of data sets can be made simultaneously in excel. Some widely used functions are sum, average, count, modulus, power, if, multiplication, divide etc. approx 90 % of organization in the world uses these software.

Advantages and disadvantages of MS EXCEL.

Advantages	Disadvantages
A person can use his full creativity in excel, through using the large number of templates, features and options provided.	MS Excel lacks behind in terms of security and other controls . MS excel cannot prevent errors that are caused through the user as their is no system of Artificial Intelligence.
MS Excel encompasses many formulas for efficient execution of work.	It doesn't have system safety features and cyber security practices.
Visual represents of data can be easily prepared in Excel through the Pie chart, Graphs, Histograms, Pivot Tables etc.	Plotting of large set of data is difficult.

The function of Macro in excel helps in completing the same file many times.	Excel has failed in proving itself to be compatible for performing tasks in case of agile business operations.
The program is easy to learn and use.	It doesn't help in providing assistance the fast decision making process.
The user can access to multiple numbers of spreadsheets while using excel.	The platform is also not suitable for collaborating and working altogether .
A person can insert variety in tables in excel as when required in literally no time for better representation of information.	The advance pricing rules are bit difficult to manage .

CONCLUSION

To conclude in this report is a mix of analysis and visualisation of the 2 different data sets provided. In the first set – Superstore , information like profits of each year and sales have been analysed after which it was ascertained that sales have been declining since 2009. In the second report too , visualisation tools have been used and results regarding the number of people who prefer vanilla , strawberry or chocolate icecreams have been attained and mean and median has also been calculated.

REFERENCES

Books and Journals:

Nizam, R.S., Zhang, C. and Tian, L., 2018. A BIM based tool for assessing embodied energy for buildings. *Energy and Buildings*. 170, pp.1-14.

Pylaev, T., Vanzha, E., Avdeeva, E., and et.al., 2019. A novel cell transfection platform based on laser optoporation mediated by Au nanostar layers. *Journal of biophotonics*. 12(1), p.e201800166.

Zhou, S., Chen, D., Huangyang, B., and et.al., 2022, March. Influence of Psychological Factors on Evacuation in Fire Scene Based on SPSS Data Analysis. In *The International Conference on Cyber Security Intelligence and Analytics* (pp. 527-534). Springer, Cham.

Hristova, G., Bogdanova, B. and Netov, N., 2022, September. Data mining of public opinion: An overview. In *AIP Conference Proceedings* (Vol. 2505, No. 1, p. 020004). AIP Publishing LLC.

Tonetto, L.M., Fogliatto, F.S. and Anzanello, M.J., 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*. 83, p.101760.

Ollikainen, A.C., 2021. Suitability of B-18 Allowables to Additive Manufactured Materials. In *AIAA Scitech 2021 Forum* (p. 1489).

Wee, M., Scheepers, H. and Tian, X., 2022. The role of leadership skills in the adoption of business intelligence and analytics by SMEs. *Information Technology & People*, (ahead-of-print).

Veisi, E., Noormohammadi, M. and BABASHAHI, M., 2021. The Association of Sociocultural Attitude towards Body Image and Eating Disorders among College Student Girls in Tehran City, Iran.

Rohman, F.S., Sata, S.A., Othman, M.R., et.al., 2021. Dynamic Optimization of Autocatalytic Esterification in a Semi-batch Reactor. *Chemical Engineering & Technology*. 44(4), pp.648-660.

Nett, N.S., Arroyo, R.X., Nguyen, T., Mashak, B.W., and et.al., 2018. IBM POWER9 systems designed for commercial, cognitive, and cloud. *IBM Journal of Research and Development*. 62(4/5), pp.7-1.

Trivedi, M. and Chen, Z., 2018, August. Lessons learned from the industry's first TPC benchmark DS (TPC-DS). In *Technology Conference on Performance Evaluation and Benchmarking* (pp. 140-154). Springer, Cham.

Kouotou, D., Ghalit, M., El Ouahabi, M. and et.al., 2021. Removal of formaldehyde and methylene blue from aqueous solution by adsorption on cerium-doped activated carbon: a comparative study. *International Journal of Environmental Analytical Chemistry*. pp.1-19.

Otake, T., Nakayama, S.F., and et.al., 2020. Development of Artificial Urine Certified Reference Material for Quantification of Neonicotinoid

Insecticides. Journal of AOAC International. 103(6), pp.1469-1477.