

UNIVERSITY OF WALES TRINITY SAINT DAVID

BACHELOR OF ARTS: BUSINESS MANAGEMENT, LEVEL 5

MODULE TITLE: DATA HANDLING AND BUSINESS INTELLIGENCE

MODULE CODE: LCBB5000

ASSESSMENT 2

REPORT

By: Georgi Georgiev

Student number: 2011297

Lecturer Name:

Joy Botu

London, 2022

Table of contents

1. Title page	1
2. Content list	2
3. Introduction	3
4. Main body	4
5. Microsoft Excel	4
6. IBM SPSS	8
7. Conclusion	15
8. References	16

Data Handling and Business Intelligence

Introduction

This report will discuss the collection and processing of data through various software programs. The main functions of Excel will be described, with which data can be easily processed and analysed. Data analysis will be performed and results will be calculated through graphical visualization. The second part will use specialized software for data analysis. Conclusions will be made about the importance of business intelligence, data management. Two data processing software will be compared.

MAIN BODY

1. EXCEL

Excel is a spreadsheet tool developed by Microsoft and part of office suite Microsoft Office 365. From the largest software company - Microsoft offer the use of the office suite of all operating systems such as - Windows, Mac OS, Linux, Android and iOS. A very useful fact is that all applications run on a cloud platform from anywhere and anytime. To organise and manage data, spreadsheet applications like MS Excel employ a collection of cells organised into rows and columns. It includes, among other things, the ability to make simple calculations, use graphing tools, construct pivot tables, and create macros. They can also use charts, histograms, and line graphs to present data. MS Excel allows users to organise data in order to see various elements from numerous angles. Microsoft Visual Basic is a programming language that may be used to develop a range of advanced numerical algorithms in Excel. Programmers have the option of creating code directly in the Visual Basic Editor, which includes Windows for debugging and organising code modules.

In business, Excel is the most extensively used spreadsheet application. Examples include business analysis, human resource management, operations management, and performance reporting. Excel uses a large array of organised cells to organise and edit data and complete mathematical computations. Spreadsheet data may be organised using graphing tools, pivot tables, and formulae. The spreadsheet application comes with a macro programming language called Visual Basic for Applications.

Organizations use Microsoft Excel for the following: business analysis, data entry and storage, strategic analysis, accounting and budgeting, data analysis, account management, collection and verification of business data, performance reporting, office administration, administrative and managerial management, project management.

The most commonly used by business are the excel functions:

IF function

The IF function, which allows you to create logical comparisons between a number and what you anticipate. As a result, an IF statement can have two outcomes. If your comparison is True, the first result is True; if your comparison is False, the second result is

False.

For example, =IF(A1="Yes",1,2) says IF(A1 = Yes, then return a 1, otherwise return a 2).

Formula	Result
=IF('Crime_data '!H2="0","Male","Female")	Female
=IF('Crime_data '!B2="0","Northern","Southern")	Southern

LOOKUP

When you need to search in a single row or column and find a value from the same point in a second row or column, use LOOKUP, one of the lookup and reference functions. Consider the following scenario: you know the component number for a car part but not the price. When you enter the vehicle component number in cell H1, you may use the LOOKUP function to get the price in cell H2. There are two types of lookup – vector form vlookup and Array form. The vector form of LOOKUP looks in a one-row or one-column range (known as a vector) for a value and returns a value from the same position in a second one-row or one-column range.

Question	Formula	Result
Which region has a lowest crime rate (45.5) ?	=VLOOKUP(45.5,'Crime_data '!A1:C48,3,0)	0
Which region has a highest crime rate (161.8) ?	=VLOOKUP(161.8,'Crime_data '!A2:C49,3,0)	0

PIVOT TABLES

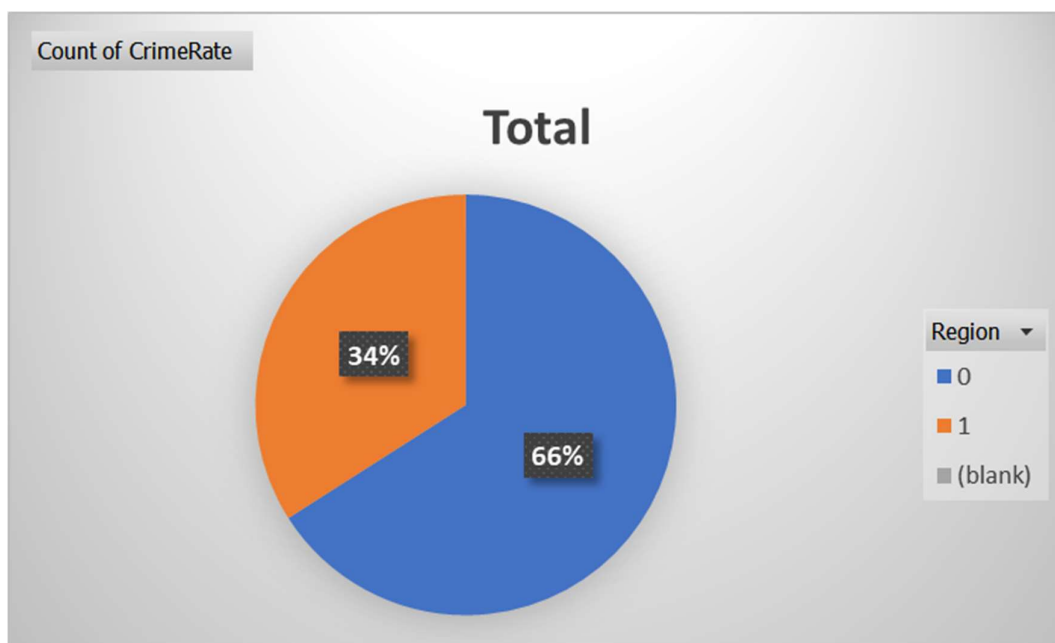
For data processing and ultimate presentation, data tabulation is necessary. One of Excel's most powerful is pivot tables, which are tabulation routines that substantially improve descriptive statistics. Users may use pivot tables to summarise and construct tables in several dimensions, generate a range of summary statistics, and alter them interactively with instant results. Because huge and extensive datasets may be easily changed, pivot tables are unquestionably the best way to study, summarise, and show data from several angles.

Determining the crime rate using the Crime data set provided:

Region	1
Sum of CrimeRate	
1610.9	0.001611

Determining the count of crime rate by regions:

Row Labels	Count of CrimeRate
0	31
1	16
(blank)	
Grand Total	47



The useful features of Microsoft Excel in handling numbers are a necessary skill for any management team. Unlimited storage for millions of rows or columns, can summarize data for the whole company is even a single file. Presenting the data in graphs by formatting the desired information is a solid basis for business analysis. Data analysis is a complete examination of data with the goal of drawing findings that will help a firm or organisation to

make a decision. All this can be summed up as business intelligence. Business intelligence assists firms in analysing historical and current data in order to quickly identify actionable insights for strategic decision making. **Business intelligence** solutions enable this by analysing massive data sets from numerous sources and providing results in visually appealing ways that are simple to understanding and use. Here are the fundamental stages that business intelligence takes to convert raw data into simple insights that everyone in the organisation can utilise. Data gathering, analysis, and visualization—these are the steps that set the stage for the decision process. Firms had to undertake much of their analysis manually but Business Intelligence technologies automate many of the procedures, saving businesses time and effort. Data visualisations are used in business intelligence reporting to make results easier to understand and disseminate. Interactive data dashboards, charts, graphs, and maps allow people understand what's going on in the business right now. Business intelligence technologies are useful in helping firms eliminate inefficiencies, discover possible problems, locate new income streams, and identify areas of future development since they speed up information processing and performance review.

Business intelligence has a direct link to the quality of economic performance. Business intelligence helps management in the environment of marketing, management, productivity, the impact of internal and external factors, customer evaluation, market demand, service and revenue. After preparing a database with all the above factors - business intelligence moves on to analysis. Strategies and planning can be prepared on the basis of previous business periods. Deciding on the future is important and is at the heart of business intelligence. Business intelligence has a strategic impact on the company and economic recovery, social policy and sustainability.

A data warehouse is a sort of data management system that enables and supports business intelligence operations, particularly analytics. Data warehouses exist exclusively to do queries and analysis on enormous volumes of historical data. A data warehouse's data is often generated from a variety of sources, including application log files and transaction programmes. A data warehouse consolidates and centralises vast volumes of data from several sources. Its analytical skills enable firms to gain useful business insights from their data and make better decisions. He has amassed a collection of historical records that data scientists and business analysts may rely on. Because of these qualities, the data warehouse might be called the organization's "sole source of truth."

Spreadsheets are one of the oldest types of data analysis. Spreadsheets, which are frequently used by individuals who lack high technical abilities and may be used for relatively simple analysis that does not require extensive expertise, sophisticated and massive amounts of data, and management databases. Excel cannot be categorized as a group of specialized analysis software, but it is still widely used around the world. One of the reasons is that this software is present in all Windows computers and everyone has had to work with it. Excel is a flexible tool for manipulating rows and columns from which analysis can be created. Excel works as an electronic accounting sheet and can also be used as a reporting tool. Excel comes with a wide range of functionalities, from arranging to manipulating, calculating, and evaluating quantitative data to building complex calculations and using pivot tables, conditional formatting, adding multiple rows, and creating charts and graphs – Excel has definitely earned its place in traditional data management.

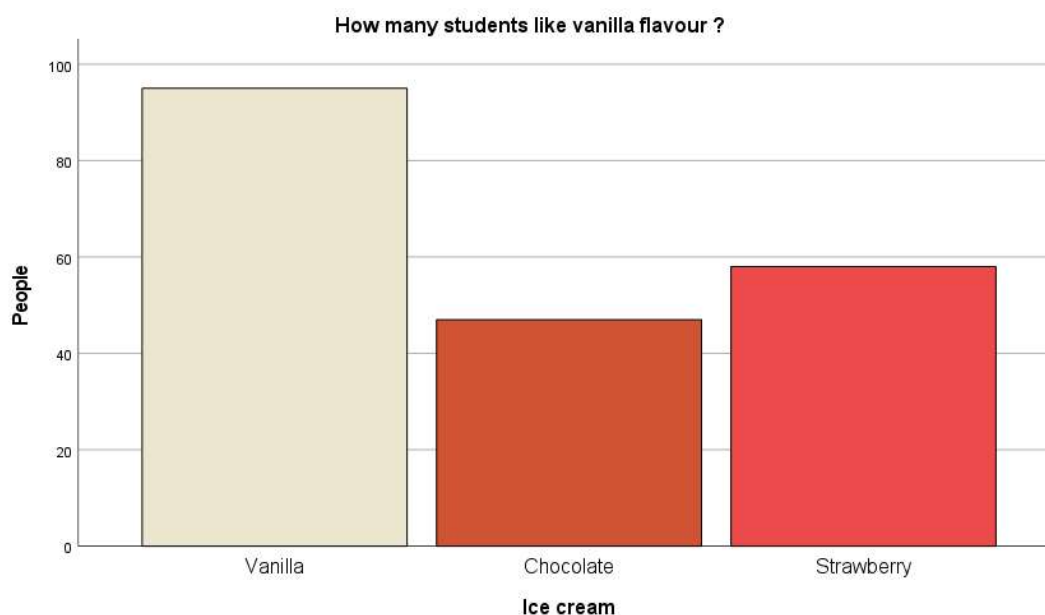
However, Excel is not designed for data analysis and is not claimed to be. As a negative feature for Microsoft Excel can be pointed out the lack of analysis functions, the complexity of working with formulas, errors in command execution, data formatting takes a long time, Excel is focused on the calculation. Yes excel is a great tool in the office, but there is better data analysis software - one of them is IBM SPSS.

2. **IBM® SPSS®** Statistics software provides a robust collection of statistical tools that allow your company to make the most of the important information provided by your data. You may uncover insights to enhance decision making by diving deeper into your data—for example, growing markets, enhancing research outputs, maintaining regulatory compliance, controlling risk, and optimising ROI.

- 1) How many students like vanilla flavour?

From the data generated by the knowledge of frequencies it is clear that students who prefer vanilla ice cream for most of the group. They are 47.5% or 95 people.

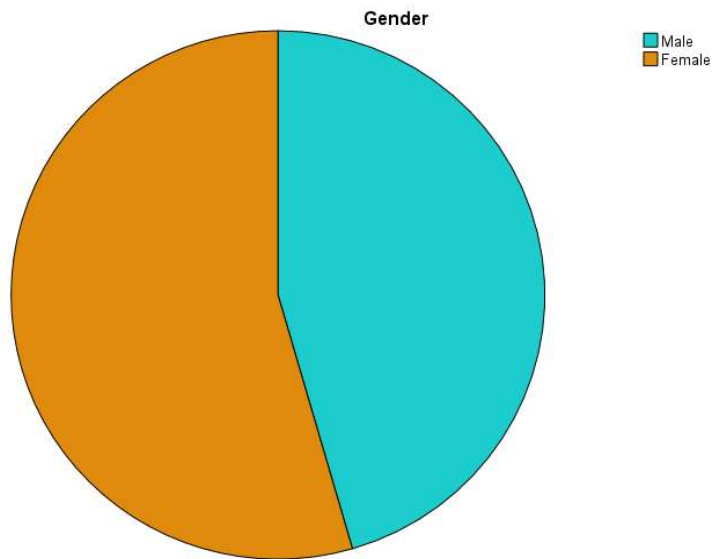
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Vanilla	95	47.5	47.5	47.5
	Chocolate	47	23.5	23.5	71.0
	Strawberry	58	29.0	29.0	100.0
	Total	200	100.0	100.0	



2) How many students are Male and Female?

From the generated information it can be seen that in the study group there is a large number of women 109 and respectively 91 men. In percentage terms, women are 54.5% and 45.5% men.

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	91	45.5	45.5	45.5
	1	109	54.5	54.5	100.0
	Total	200	100.0	100.0	



3) What are the Mean and Median of participants who likes chocolate and strawberry flavour?

Favorite Flavor (1: Vanilla, 2: Chocolate, 3: Strawberry)

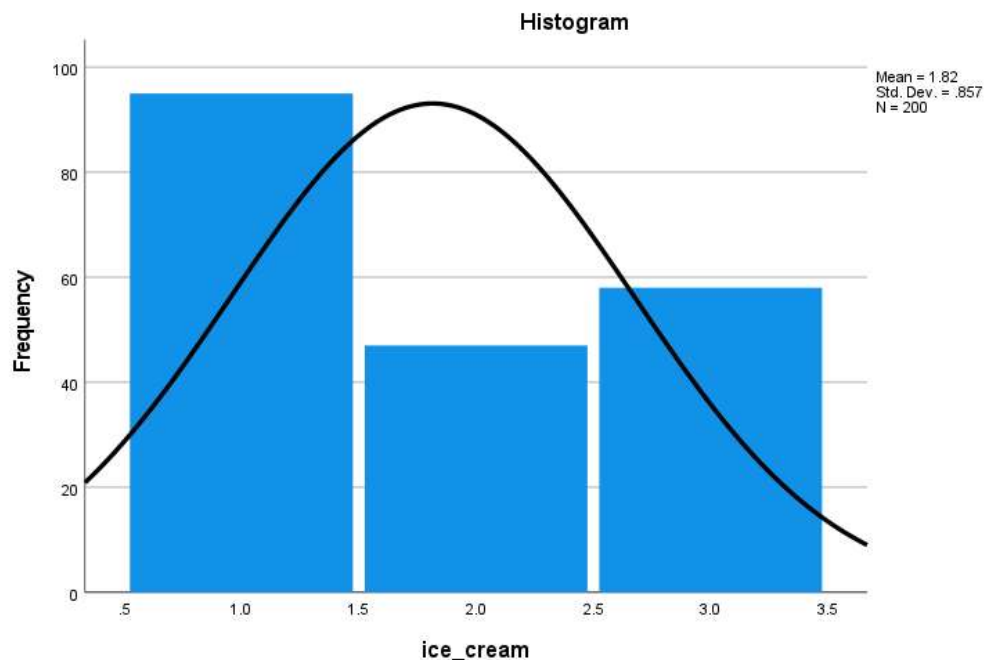
Statistics

ice_cream

N	Valid	200
	Missing	0
Mean		1.82
Median		2.00
Sum		363
Percentiles	25	1.00
	50	2.00
	75	3.00

ice_cream

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	95	47.5	47.5	47.5
	2	47	23.5	23.5	71.0
	3	58	29.0	29.0	100.0
	Total	200	100.0	100.0	

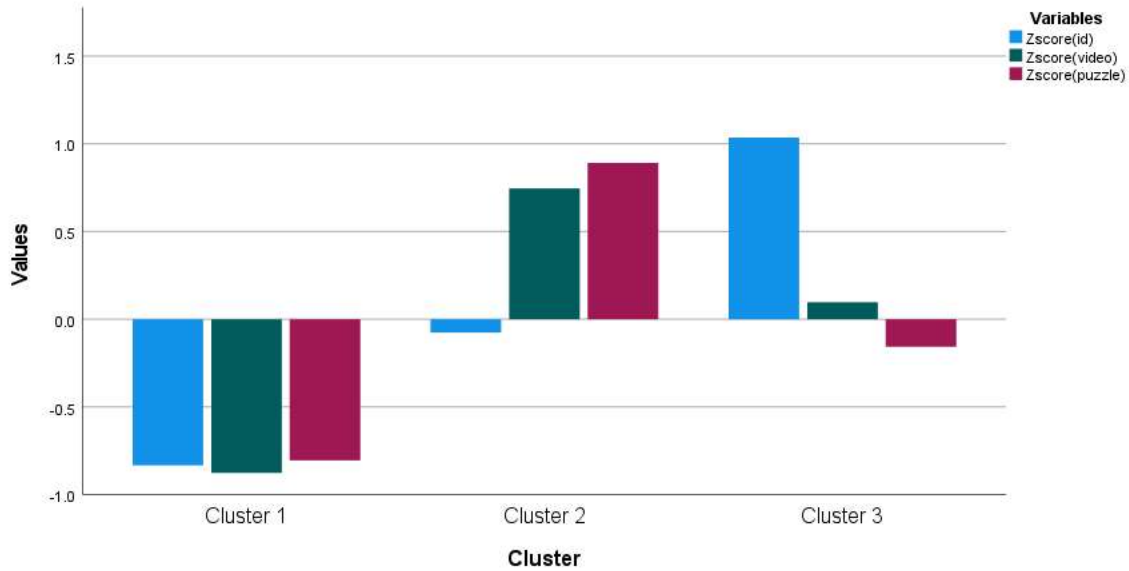


a) Using the data set Ice_cream_csv provided in conjunction with SPSS give a specific example of clustering (ex. K-means). Show your workings with screenshots and clearly explain your results.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(id)	-.83287	-.07631	1.03549
Zscore(video)	-.87604	.74600	.09763
Zscore(puzzle)	-.80481	.89115	-.15726

Final Cluster Centers



File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help																
1: Visible: 10 of 10 Variables																
	id	female	ice_cream	video	puzzle	Zid	Zvideo	Zpuzzle	OCL_1	OCL_2	var	var	var	var	var	var
1	70	0	2	47	57	-.52696	-.48985	.42801	1	1						
2	121	1	1	63	61	.35419	1.12616	.80059	2	2						
3	86	0	3	58	31	-.25052	.62116	-.199380	1	1						
4	141	0	3	53	56	.69973	.11615	.33486	2	3						
5	172	0	1	53	61	1.23533	.11615	.80059	2	3						
6	113	0	1	63	61	.21597	1.12616	.80059	2	2						
7	50	0	1	53	61	-.87251	.11615	.80059	2	2						
8	11	0	1	39	36	-.154632	-.129786	-.152807	1	1						
9	84	0	1	58	51	-.28508	.62116	-.13087	2	2						
10	48	0	1	50	51	-.90706	-.18685	-.13087	1	1						
11	75	0	1	53	61	-.44057	.11615	.80059	2	2						
12	60	0	1	63	61	-.69973	1.12616	.80059	2	2						
13	95	0	3	61	71	-.09503	.92416	1.73206	2	2						
14	104	0	3	55	46	.06047	.31815	-.59660	2	3						
15	38	0	2	31	56	-.107964	-.210587	.33486	1	1						
16	115	0	2	50	56	.25052	-.18685	.33486	2	3						
17	76	0	3	50	56	-.42330	-.18685	.33486	2	2						
18	195	0	1	58	56	1.63271	.62116	.33486	2	3						
19	114	0	3	55	61	.23324	.31815	.80059	2	2						
20	85	0	1	53	46	-.26780	.11615	-.59660	1	1						
21	167	0	1	66	41	1.14895	1.42916	-.106233	2	3						
22	143	0	1	72	66	.73429	2.03517	1.26632	2	2						
23	41	0	1	55	56	-.102800	.31815	.33486	2	2						
24	20	0	3	61	61	-.139083	.92416	.80059	2	2						
25	12	0	1	39	46	-.152905	-.129786	-.59660	1	1						
26	53	0	1	39	31	-.82068	-.129786	-.199380	1	1						
27	154	0	3	61	66	.92434	.92416	1.26632	2	2						
28	178	0	1	58	46	1.33900	.62116	-.59660	2	3						
29	196	0	3	39	46	1.64999	-.129786	-.59660	1	3						
30	29	0	2	55	41	-.123533	.31815	-.106233	1	1						
31	126	0	1	47	51	.44057	-.48985	-.13087	2	3						
32	103	0	3	64	61	.04319	1.22716	.80059	2	2						
33	192	0	3	66	71	1.58088	1.42916	1.73206	2	2						
34	150	0	1	72	31	.85523	2.03517	-.199380	2	3						
35	199	0	3	61	61	1.70182	.92416	.80059	2	3						
36	144	0	3	61	66	.75157	.92416	1.26632	2	2						
37	200	0	1	66	66	1.71910	1.42916	1.26632	2	2						
38	80	0	3	66	66	-.35419	1.42916	1.26632	2	2						
39	16	0	2	36	36	-.145994	-.160087	-.152807	1	1						
40	153	0	1	39	51	.90706	-.129786	-.13087	1	3						
41	176	0	1	42	51	1.30444	-.99486	-.13087	2	3						

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

Text mining is the act of converting unstructured text into a structured format in order to uncover relevant patterns and fresh insights. Text mining employs natural language processing (NLP), which enables machines to interpret and process human language automatically. Natural language processing employs approaches from a variety of disciplines, including computer science, artificial intelligence, linguistics, and data science, to assist computers in understanding human language in both written and spoken form. The most popular software tools for text mining are IBM SPSS Modler, SAS Enterprise Miner, Statistical Data Miner, Rapid Miner, and GATE.

The benefits of text retrieval are clear in sectors where huge amounts of text data are created, such as academic research, finance, medicine, biology, technology, marketing, Google search engine, email client, etc. The following are some of the most common text mining application areas:

- Information gathering. Key word recognition for individuals and organisations, location names, and temporal expressions
- Tracking of topics. Text mining can forecast further papers of interest to a user based on a user profile and documents seen by the user.
- Recapitulation. Summarizing a document in order to save the reader's time.
- Classification. Identifying a document's primary topics and then categorising the document based on those themes.
- Grouping. Similar papers are grouped without a preset set of Concept connecting.
- Answering questions. Using knowledge-driven pattern matching to get the best solution to a given query.

Data mining is to identifying useful information in huge data sets. Data mining tools are strong statistical, mathematical, and analytical tools that sift through enormous volumes of data to uncover trends, patterns, and linkages to enable informed decision-making and planning. Because of the advancement of data warehousing technologies and the rise of big

data, the use of data mining techniques has accelerated during the last several decades, supporting businesses in translating raw data into valuable knowledge.

Data mining may be performed using a number of approaches, including classification, regression, clustering, and association. For each of these strategies, most data mining software products incorporate more than one strategy.

Data retrieval is becoming increasingly important and as much as possible is a set of as much information extracted from them will be more accurate. Data mining offers the tools to take full advantage of big data and transform it into applied intelligence. The process of data mining is divided into several steps: 1. Identify the problem - it is important to ask the right question. 2. Data collection - all available information on the problem. 3 Data processing and reading - formatting, grouping, modeling. 4. Presentation of data - summarizing the most important information, the easiest way is through graphs and tables.

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

SPSS Stands for statistical packages for Social Science is a specialized software for statistical analysis, which offers faster and more secure operation. With a ready-made database, graphical analysis can be generated with just a few clicks. SPss offers in-depth software solutions for professional use - analysis.

Excel is a powerful spreadsheet program with unlimited data storage and processing capabilities. By clearing the system interface from the cells it provides easy operation. Built-in features such as formulas and automatic calculations can save time and money. In the case of a large data set, formatting them can provide specific information. Then you can easily make analyzes and comments, which in turn give the right direction for successful decisions.

At first glance, both software programs may perform the same tasks, but it does matter who will use those programs. If the person who is a professional date analyst and this is his main profession he would use SPSS. A person working in the field of business management, human resources, warehouses, office workers would use Excel, which is even part of his reinstalled office suite. The difference is big, moreover, the two software should not be compared because they perform different activities. Each of them has its own

environment. In conclusion, SPSS is specialized software for professional use, and excel is a cellular program for spreadsheets with widespread use.

CONCLUSION:

The successful business model is built on information management, which is business data for the past and future. Nowadays, data processing and storage is done digitally and is extremely important. There are many software solutions, but the most prominent are Excel and SPSS. With their help, it can put business intelligence into action so that business can grow and pay dividends. Data analysis is a comprehensive study of a set of information that aims to draw conclusions that allow a company or organization to make a decision. Every company has its own business intelligence as part of its business culture. Business management is directly related to information management on the basis of which decision management is performed. A successful business model with a machine with many parts and in order for it to work well, all parts must work for themselves and for everyone else.

References:

1. Technopedia (2020) *Microsoft excel*. Available at: <https://www.techopedia.com/definition/5430/microsoft-excel> (Accessed: 13/05/2022).
2. Gillis A. (2021) *Definition Excel*. Available at: <https://www.techtarget.com/searchenterprisedesktop/definition/Excel> (Accessed: 13/05/2022).
3. Microsoft (2022) *If Function*. Available at: <https://support.microsoft.com/en-us/office/if-function-69aed7c9-4e8a-4755-a9bc-aa8bbff73be2> (Accessed: 13/05/2022).
4. Microsoft (2022) *Look up Function*. Available at: <https://support.microsoft.com/en-us/office/if-function-69aed7c9-4e8a-4755-a9bc-aa8bbff73be2> (Accessed: 13/05/2022).
5. Grech V. (2018) *WASP (Write a Scientific Paper) using Excel – 2: Pivot tables*. Available at: <https://www.sciencedirect-com.ezproxy.uwtsd.ac.uk/science/article/pii/S0378378218300070> (Accessed: 13/05/2022).
6. Microsoft (2022) *What is business intelligence?* Available at: <https://powerbi.microsoft.com/en-us/what-is-business-intelligence/> (Accessed: 13/05/2022).
7. Datapine (2022) *Essential Data Analyst Tools*. Available at: <https://www.datapine.com/articles/data-analyst-tools-software> (Accessed: 13/05/2022).
8. Oracle (2022) *What Is a Data Warehouse?* Available at: <https://www.oracle.com/uk/database/what-is-a-data-warehouse/> (Accessed: 13/05/2022).
9. IBM (2020) *Text Mining*. Available at: <https://www.ibm.com/cloud/learn/text-mining> (Accessed: 13/05/2022).
10. Azevedo A. (2015) *Integration of data mining in business intelligence systems*. Available at: <https://eu.alma.exlibrisgroup.com/leganto/readinglist/citation/6685773080002421?auth=SAML> (Accessed: 13/05/2022).
11. Sharda R. et. Al. (2019) *BUSINESS INTELLIGENCE AND ANALYTICS*. Available at: <https://www.google.bg/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK EwjS3JrywN33AhXNesAKHRKhAq0QFnoECAcQAQ&url=http%3A%2F%2Fseul.org%2Ffiles%2Flevel8%2FIT445%2FIT445%2520BOOK%2520EDIT.pdf&usg=AO>

[vVaw3bJXUzHQAkrUBNAOb2cFu7](#)
(Accessed: 13/05/2022).

12. SAP (2022) *What is data mining?* Available at: <https://www.sap.com/insights/what-is-data-mining.html>
(Accessed: 13/05/2022).