1. INTRODUCTION TO SPREADSHEET ANALYSIS

This chapter is devoted to the most important issues related to data analysis using an Excel spreadsheet, which can also be used to analyse logistic data. It contains:

- basic definitions,
- data analysis,
- the importance of data for logistics,
- spreadsheet and its application.

1.1. Introduction

The analysis and management of information in an enterprise covers all organizational units at subsequent levels of the economic system, i.e. operational, tactical, strategic, as well as knowledge management. Data analysis and management include the following activities (Szymonik, 2010):

- constituting the information function of the enterprise, i.e. acquiring, storing, processing, sharing and using information,
- within the planes (technological, organizational, human resources) influencing the implementation of this function.

The guarantee of the success of any undertaking is achieving an information advantage, defined as the ability to acquire, process and disseminate information, which will enable, for example, dominating competitors or improving the logistics process. Information advantage can be gained, among others, by: by meeting the expectations of specific users, e.g. supply chain participants, by providing qualitative information characteristics from which to exchange (Szymonik, 2015):

- relativity – information meets the needs and is important for the recipient,
- accuracy – information is adequate to the level of knowledge represented by the recipient, precisely and exactly reflects and defines the topic,
- up-to-date – the update cycle is consistent with the content and the pace of change,
- completeness – information contains an optimal and sufficient amount of data to transform the information into specific knowledge, and its level of detail depends on the needs of the recipient,
- consistency – individual data harmonizes with each other, the form corresponds to the content, data updating is consistent with the goals,
- appropriateness – appropriate presentation of information and description for presentation, enabling correct interpretation,
- availability – information is available from anywhere and at any time,
- credibility – the information confirms the truthfulness of the data and contains elements ensuring the reliability of the message,
- congruence – information is consistent with other information, interpreted in the appropriate context, functioning in a familiar communication system.

1.2. Data analysis

Data is a representation of raw, unstructured facts, concepts, instructions or results gathered from observations or records about phenomena, objects or people that can be shaped and formed to create information in a form that can be communicated, interpreted, deduced, inferred or processed by humans or automatic devices.

Data analysis is the process of examining, interpreting and presenting information collected from various sources. Using a variety of techniques and tools, data scientists turn raw data into actionable information that helps enterprises make decisions, identify trends, and solve problems. In today's world, where companies generate huge amounts of data, learning data analysis is becoming more and more important, and its skill is becoming more

and more desirable on the labour market (www_1.1). Data analytics is a key process in science and business to transform raw data into useful and valuable information through various methods and analytical techniques.

The use of modern analytical techniques in logistics allows data to be transformed into valuable resources, supporting innovation and the development of business intelligence, which is a key element of the modern business approach (Zhang & Shao, 2020; www_1.2).

In the context of logistics and supply chain, the key challenges in managing this data fall into three main areas. Firstly, there is the problem of data pre-processing and compression. Secondly, logistics data management faces difficulties due to the fragmentation of companies in the supply network, such as missing data or disruptions in network equipment, which increases the risk for both suppliers and customers. Thirdly, there is an insufficient level of sophistication in data analysis and decision support. Deficiencies in modelling technology, data mining methods and decision support systems limit the ability to provide valuable information for logistics operations.

1.3. MS Excel Spreadsheet and its application

Excel is one of the most widely used programmes on company computers, as it is where most corporate reports are prepared. In addition, many company systems export data in formats that are compatible with Excel, making it easier to organise and view information in a clear and structured manner. With the VBA programming language, which is an integral part of Excel, the programme gains even wider application possibilities, such as automating routine tasks, creating more advanced tools or developing functions. VBA is a key tool for spreadsheet automation, enabling the creation of macros for repetitive tasks and integration with other elements of Microsoft Office, as well as with programmes such as AutoCAD (Shinsato Jr et al., 2023). Excel Spreadsheet is a spreadsheet programme in the Microsoft Office group of applications. MS Excel provides features such as calculations, charting tools, pivot tables and a macro programming language called Visual Basic for Applications. It also offers a set of statistical analysis functions and other tools that can be used to run descriptive statistics and perform several different statistical tests.

A spreadsheet is a computer programme used to perform various types of calculations, often very complex ones. In spreadsheets, we can present data, mainly numerical, in the form of a set of tables that allow automatic processing of this data, its analysis and presentation in various ways, e.g. in the form of various types of charts, ranging from simple line charts, through pie charts and bar charts, to eye-catching bubble charts. The most important capabilities that spreadsheets provide to the user are (www_1.2): (1) data analysis, (2) performing calculations, (3) preparing offers, (4) presentation of results, (5) creating charts, (6) creating reports and summaries.

In each cell of a spreadsheet, you can enter numerical data, text data or a formula in the sheet called a **formula**, which allows you to calculate a given value based on the contents of the cells. It can include in its content the **addresses** of these cells, mathematical symbols and more advanced operations such as **functions** – not only mathematical, but also statistical, financial, date and time or database functions, which are the most important and frequently used tools that a spreadsheet provides. A **function** in a spreadsheet, on the other hand, is an algorithm specially designed by the programme's creators, ready-to-use formulas that allow specialized calculations or searches for specific values. Examples include the Average function, which calculates the arithmetic mean of given numbers, or the Maximum function, which searches for the largest of given numbers, and many others. With the help of these functions, data entered into the programme is processed automatically and can be used to create simulations. Formulas in the worksheet are built using standard rules for creating mathematical expressions. The entry of a formula should always be preceded by an equals sign, e.g. =A8+C11 or =(F14-E10)*12 etc. Formulas are used to calculate and analyse data in a spreadsheet. If a number in the formula is changed, the programme will make the changes automatically and still display the correct result. This way, you do not have to change everything manually. Good spreadsheets, such as Excel, for example, have ready-made functions built in (www_1.2).

The superiority of spreadsheets over other types of software also lies largely in the ability to perform a very large number of calculations with a lot of data, without having to

manually confirm each individual action. Performing calculations in such an automated manner significantly reduces working time and requires incomparably less effort from the employee.

In addition, as already mentioned, spreadsheets make it possible to illustrate the data collected and the results of the calculations in a way that is clear and attractive to the recipient, such as various types of charts and diagrams. Sophisticated spreadsheet programmes are able to generate many different types of graphs, which can be used for statistical purposes, optimisation of a particular process or visualisation of changes to be implemented in an organisation. For this reason, they are very often used in various types of presentations of planned projects, where they are used to show the results obtained or predicted for the future. Both charts and pivot tables make it easier to see interdependencies and trends, and thus to better determine the effectiveness of particular activities or tools (www_1.2).

Spreadsheets are often used as a multifunctional tool for data entry, storage, analysis and visualisation. Most spreadsheet software allows users to perform all these tasks, but spreadsheets are best suited for data entry and storage, while analysis and visualisation should be done separately. Analysing and visualising data in a separateor at least in a separate copy of the data file reduces the risk of contamination or destruction of the raw data in the spreadsheet (Broman & Woo, 2018).

1.4. The most important tools provided by spreadsheets

The spreadsheet offers a wide range of tools that can be used to analyse both logistics data and other types of data. Among the tools provided by the MS Excel spreadsheet, the following should be mentioned:

- data filtering,
- data sorting,
- statistical analysis functions,
- linear regression analysis tool,
- correlation diagram,
- pivot tables,

- Solver,
- macros,
- Power Query,
- 3D maps.

Information gathering often involves large data sets, characterised by redundancy. Before any further analysis of them can be carried out, only those data that meet specific criteria, tailored to the information needs of decision-makers, need to be extracted from the database. In Excel, two filtering methods are available, located on the Data ribbon (Fig. 1.1): auto filter (Filter command) and Advanced filter (Advanced command).

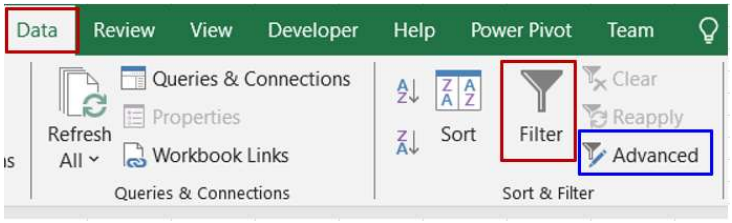


Figure 1.1. View of the Data ribbon with filter commands

Source: own study

Filtering data by format (auto filter, Filter by Colour option) allows you to select values with a specific font colour, cell fill colour or that contain a specific cell icon, inserted via Conditional Formatting (Fig. 1.2).

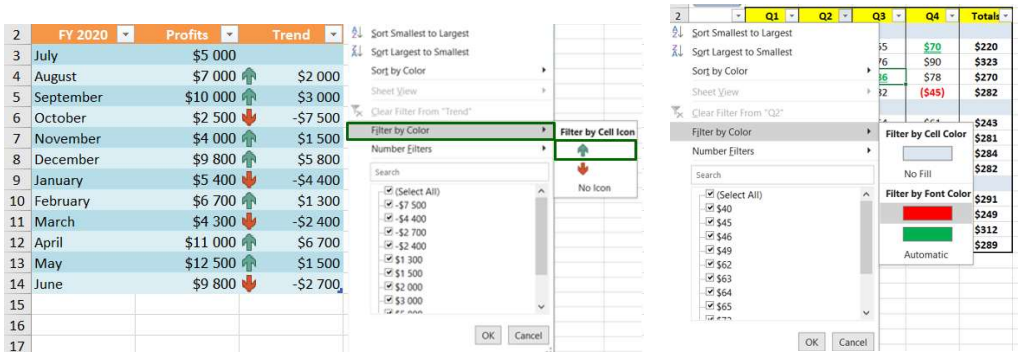


Figure 1.2. Example of the application of the auto filter by format (by cell icons and font colour)

Source: own study

If more criteria need to be defined, the advanced filter should be used. When applying the advanced filter, a so-called filter criterion must be defined.

When working with databases, there is often a need to organise data in a specific order according to user-defined criteria. This process can be achieved by **sorting**. The simplest form of sorting is simple sorting, that is, by a single criterion. There is also the possibility of multi-level sorting, during which the database is sorted according to two or more criteria (Fig. 1.3).

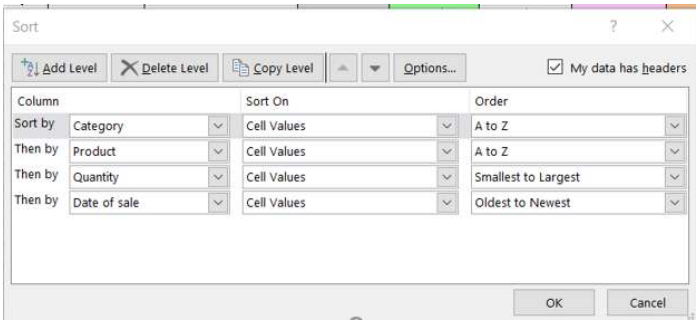


Figure 1.3. View of the Sorting window with set criteria for multi-level sorting

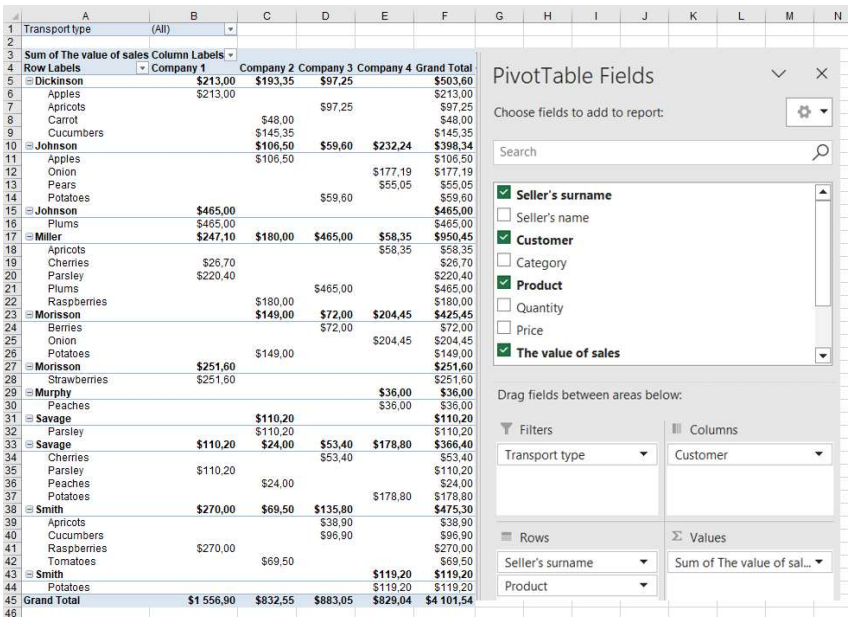
Source: own study

Evaluating datasets for meaning can be difficult, especially as the volume of data increases. Sifting through rows of raw data in spreadsheets can be virtually impossible with any hope of seeing deeper meaning. Numerical summaries can be helpful, but they can still be inadequate. Transforming these numerical summaries into pivot tables and pivot charts can often make them more comprehensible with excellent layout and visual representation.

Excel gives users the ability to create pivot tables and associated pivot charts. These beneficial tools contribute to the automation of the data analysis process and allow almost instantaneous changes to the patterns in which the data is organised, as well as to the parts of the data that are viewed. Reports that meet any need can be generated instantly to answer questions that arise about the data. Pivot tables allow individual data points to be highlighted for immediate comparison with other points, allowing for easy comparison of many different variables.

Pivot tables are an analytical tool in which, as the name suggests, you can freely rearrange the information contained in them. By using a pivot table, you can freely redesign rows and columns so that the resulting table form is clearer or more clearly indicates specific data that the user wants to emphasize. Knowledge of this spreadsheet function is essential when creating summaries and reports (www_1.2). With a pivot table, it is possible to invert data that is in rows. Data can be moved into columns running right across the spreadsheet, which can help the data take a more useful form when converted to a visual chart. The automation of data manipulation contributes to speeding up the process and eliminates potential human error resulting from manual data manipulation. Pivot tables and pivot charts are dynamic in nature and allow their content to be changed instantly to answer specific data-related questions, whereas significant effort would need to be put into rearranging the data to answer the same questions with traditional tables (Miller, 2014).

An example of a pivot table report showing the total value (Values area) of products sold by each vendor (Rows area) to individual contractors (Columns area) by a given mode of transport (Filters area) is presented in Figure 1.4.

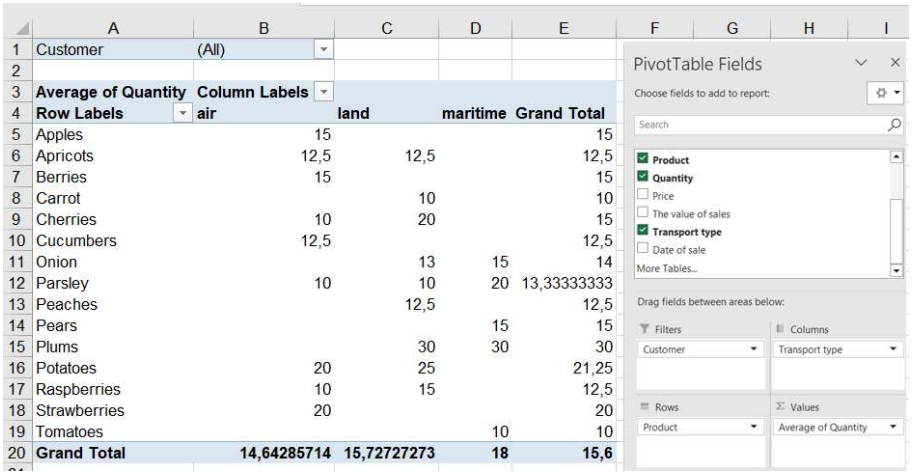


Transport type	(All)				
Sum of The value of sales	Column Labels				
Row Labels	Company 1	Company 2	Company 3	Company 4	Grand Total
Dickinson	$213,00	$193,35	$97,25		$503,60
Apples	$213,00				$213,00
Apricots			$97,25		$97,25
Carrot		$48,00			$48,00
Cucumbers		$145,35			$145,35
Johnson		$106,50	$59,60	$232,24	$398,34
Apples		$106,50			$106,50
Onion				$177,19	$177,19
Pears				$55,05	$55,05
Potatoes			$59,60		$59,60
Johnson	$465,00				$465,00
Plums	$465,00				$465,00
Miller	$247,10	$180,00	$465,00	$58,35	$950,45
Apricots				$58,35	$58,35
Cherries	$26,70				$26,70
Parsley	$220,40				$220,40
Plums			$465,00		$465,00
Raspberries		$180,00			$180,00
Morisson		$149,00	$72,00	$204,45	$425,45
Berries			$72,00		$72,00
Onion				$204,45	$204,45
Potatoes		$149,00			$149,00
Morisson	$251,60				$251,60
Strawberries	$251,60				$251,60
Murphy				$36,00	$36,00
Peaches				$36,00	$36,00
Savage		$110,20			$110,20
Parsley		$110,20			$110,20
Savage	$110,20	$24,00	$53,40	$178,80	$366,40
Cherries			$53,40		$53,40
Parsley	$110,20				$110,20
Peaches		$24,00			$24,00
Potatoes				$178,80	$178,80
Smith	$270,00	$69,50	$135,80		$475,30
Apricots			$38,90		$38,90
Cucumbers			$96,90		$96,90
Raspberries	$270,00				$270,00
Tomatoes		$69,50			$69,50
Smith				$119,20	$119,20
Potatoes				$119,20	$119,20
Grand Total	$1 556,90	$832,55	$883,05	$829,04	$4 101,54

Figure 1.4. Example of a pivot table report in Excel

Source: own study

Because the fields in the pivot table can be positioned in any configuration, the result is a different report layout effect each time. Figure 1.5 shows a pivot table created from the same source list as the pivot table in Figure 1.4, but this time it presents a report to analyse the average quantity (Values area) of products sold (Rows area) and delivered to a given customer (Filters area) using a given Transport type (Columns area).



	A	B	C	D	E
1	Customer	(All)			
2					
3	**Average of Quantity**	**Column Labels**			
4	**Row Labels**	**air**	**land**	**maritime**	**Grand Total**
5	Apples	15			15
6	Apricots	12,5	12,5		12,5
7	Berries	15			15
8	Carrot		10		10
9	Cherries	10	20		15
10	Cucumbers	12,5			12,5
11	Onion		13	15	14
12	Parsley	10	10	20	13,33333333
13	Peaches		12,5		12,5
14	Pears			15	15
15	Plums		30	30	30
16	Potatoes	20	25		21,25
17	Raspberries	10	15		12,5
18	Strawberries	20			20
19	Tomatoes			10	10
20	**Grand Total**	**14,64285714**	**15,72727273**	**18**	**15,6**

Figure 1.5. Pivot table showing the Average number of sold products delivered to a given Customer using individual types of transport

Source: own study

Currently, there is an increased interest among entrepreneurs in 'what if?' spreadsheets with optimisation capabilities, such as **EXCEL Solver** (Microsoft Co.). Excel Solver is primarily used to solve and optimise process design and integration. Practicing engineers also use spreadsheets for many tasks, as process optimisation is becoming an increasingly common task in process synthesis, design and integration.

Due to its usefulness, the solver is very often used in the decision-making process for optimising issues such as: efficient use of existing materials, reducing delivery and transport costs, determining production volumes or determining the best multi-shift work schedule.

Solver is a free add-in for the Microsoft Excel spreadsheet. Excel Solver has two non-linear unconstrained optimisers, the quasi-Newton method and the reduced gradient method. These are used within the Generalized Reduced Gradient algorithm for solving

constrained optimisation problems. The linear simplex method with constraints on variables and the branch-and-bound method can be used to solve linear and integer problems. The approach used to obtain better initial estimates of the underlying variables in any one dimensional search can be specified in the Solver options. Linear extrapolation from the tangent vector or quadratic extrapolation can be used, which can improve results for highly non-linear problems. It is also possible to specify a differential method to estimate the derivatives of the objective and constraint functions: Forward, when the values of the constraints change relatively slowly, or Central, used for problems when the constraints change rapidly, especially near the limits of the active constraints (Ferreira et al., 2004).

In order to perform the calculation, you must first start by coding the contents of the Excel worksheet and place the formula that calculates the function in the selected cell. The parameter values of the function, as well as the arguments to be sought, must be coded in the cells of the selected worksheet range. In addition, the formulas needed to include variable constraints in the calculation should be saved. Next, the Solver dialog box should be displayed, designed to determine the relationships necessary to reach a solution. In the window, by referring to the cell addresses, one should indicate (Bomba & April, 2012):

- objective cell, e.g. $A$2,
- the sought values of the objective function (Max, Min or Value),
- the range of variables sought, e.g. $H$8:$H$13,
- constraint relations,
- solution method (non-linear GRG, LP simplex or Evolutionary).

The Solver will perform the optimisation calculation when the Solver button is pressed, resulting in a displayed report with the results.

Task description: The company produces two products: A and B. Each of them generates profit, but requires different amounts of work time and materials (input data can be found in Table 1.1).

Table 1.1. Input data for an optimisation task solved by the Solver

Maximum	Product	Profit per unit	Work time (hours)	Material (kg)
	A	50	2	1
	B	40	1	2
Work time			100	
Quantity of materials				80

Source: own study

Determine how many units of each product to produce in order to maximise profit, with constraints on work time and material availability?

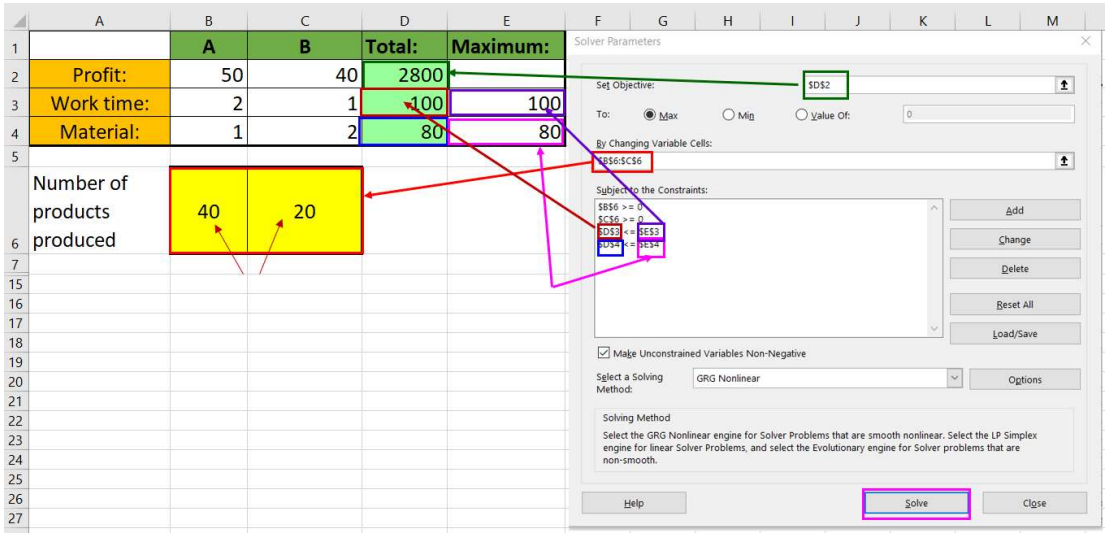


Figure 1.6. Applying Solver to an Example Optimisation Task

Source: own study

Macros in Excel are command sequences written in the Visual Basic for Applications (VBA) programming language, allowing the automation of frequently performed actions in a spreadsheet. Thanks to macros, a series of tedious and repetitive operations can be reduced to a single click of a button or key combination. If a number of the same operations are frequently performed on some data in a company, it is likely that they can be automated using a macro. Obviously, it is worth creating a universal macro, capable of performing a sequence of specific actions not only on one piece of data, but also if the layout of the data changes slightly or if the number of data increases or decreases. For this reason, when

designing a macro, it is important to use a command that can be applied to multiple sheets, not just one, in a particular situation.

Examples of work automation using macros: (1) automatic data sorting, (2) automatic data input, (3) automatic filling of forms, (4) automatic report generation, (5) automatic creation and processing of forms and surveys, (6) automatic integration with other systems.

Microsoft Power Query (PQ) is a Microsoft Excel spreadsheet add-in for versions prior to Excel 2016, designed by Microsoft to support Self-Service Business Intelligence solutions. It is also worth emphasizing its usefulness when working with data, collecting it or analysing it. PQ allows you to download data from many different areas; starting from relational databases, through data from SharePoint and the operating system, and ending with data downloaded from any website. An additional advantage is that it allows for preliminary data processing, as well as preparing it for further analysis or visualization. The implementation of the above functions is possible thanks to the special language "M", used in Power Query to create formulas and giving great possibilities of using advanced functions to work on data using selected operators.

Power Map is an add-in for Microsoft Excel Professional and Office 365 Professional that lets you create clear geospatial visualizations of your data directly from Excel. In many data visualization cases, Power Map completely eliminates the need for programming, allowing you to work directly with data in a spreadsheet and present relationships directly on a map. As a helpful debugging tool, it lets you retrieve information from a database, import it into Excel, visualize it, and draw conclusions without having to write code to plot the data on a map (Au & Rischpater, 2015).

Power Map can help create 3D visualisations by plotting up to a million data points as column, heat and bubble maps on a Bing map. If the data is time-stamped, it can also create interactive views that show the change in data over time and space (Clark, 2014).

1.5. Spreadsheet applications in various sectors

The Excel spreadsheet is most often used as a tool in manufacturing companies that do not have integrated information systems.

The first association with the use of spreadsheets is, of course, various types of office work. It is used to create employee lists, sales reports and calculate employee salaries. Programmes of this type are also widely used in accounting, thanks to advanced financial calculation functions. The charts and pivot tables available in the spreadsheets also help you visualize interest accumulating when selling financial products, both in banking, insurance companies and the investment sector. In addition, spreadsheets are also used to collect and process information needed to optimize processes and machines used in industry. The inherent use of sheets also includes creating various types of patterns and templates for commercial offers (www_1.2).

Excel is also used in the field of logistics. In their work, logisticians very often use spreadsheets to support decisions in the area of logistics. These include:

1) Stock control analysis:
    - analysis of the structure, dynamics and volume of purchases,
    - analysis of completeness and suitability of stocks,
    - stocks productivity indicators,
    - analysis of stock-related costs.
2) Dynamic analysis of the development of logistic measures over time using a pivot table:
    - analysing and comparing purchase costs,
    - comparison of indicators over time,
    - multi-currency analysis and calculations,
    - analysis of extensive databases using a pivot table,
    - analysis in the field of transport.
3) Graphical presentation of data in reports in the form of various types of charts.
4) Budget and operational logistics reports.

The Excel spreadsheet is used as a reliable tool for preparing complex reports when there are problems with reading them after they have been generated by logistics systems such as ERP or WMS, which have their own reporting systems. Pivot tables are very useful, as they cannot be created in other IT systems supporting logistics processes and those created in Excel are clear, understandable and transparent (www_1.3).

Logistics is an area of management where we often deal with quantifiable values that can be described with a mathematical model, which makes Excel ideal for analysing specific decision-making issues in the field of logistics. Here are examples of some of them:

- coordination of the flow of resources in the supply chain,
- implementation of transport tasks,
- minimization of empty runs in transport,
- intermediary issue,
- the travelling salesman issue,
- economic order quantity model,
- optimisation of the production range to maximize profit,
- capacity and material demand planning in a production company,
- optimisation of production and distribution of products,
- optimisation of order fulfilment of orders,
- XYZ analysis in stocks management,
- solving the problem of calculating logistics service prices,
- multi-criteria analysis and evaluation.

Chapter Questions

1. What are the main challenges related to information and data management in the context of a dynamically changing organizational environment and logistics system?
2. What actions can organizations take to ensure high quality information and meet user expectations in the context of supply chain management?

3. What are the benefits of using spreadsheets in data analysis and presentation of results in various projects?

REFERENCES

Au, C. & Rischpater, R., (2015). Power Map for Excel. Microsoft Mapping, Geospatial Development in Windows 10 with Bing Maps and C#. Second Edition, Apress, 159-165.

Broman, K.W. & Woo, K.H. (2018). Data Organization in Spreadsheets, The American Statistican, Taylor & Francis Group, Vol. 72, No.1, 2–10.

Clark, D., (2014), Beginning Power BI with Excel 2013. Self-Service Business Intelligence Using Power Pivot, Power View, Power Query, and Power Map, Apress.

Ferreira, E., Lima, R. & Salcedo, R., (2004). Spreadsheets in Chemical Engineering Education ĐA Tool in Process Design and Process Integration, Int. J. Engng Ed. Vol. 20, No. 6, 928-938.

Shinsato Jr, Ch., de Mattos Veroneze G., da Costa Craveiro, J M., Neto, T M. (2023). Proposal of an inventory control system based on the flow of materials in a warehouse using Excel/VBA, E-tech-Vol.16, Iss:1

Szymczak, M. (ed.) (2011). Decyzje logistyczne z Excelem, Difin S.A., Warszawa.

Szymonik, A. (2010). Technologie informatyczne w logistyce, Placet, Warszawa.

Szymonik, A. (2015). Informatyka dla potrzeb logistyka (I), Difin S.A., Warszawa.

Zhang, C. & Shao, X. (2020). Research on intelligent analysis of port logistics information based on dynamic data mining, Journal of Coastal Research, vol. 115.

(www_1.1) https://blog.strefakursow.pl/wprowadzenie-do-analizy-danych-kluczowe-pojecia-i-narzedzia-dla-poczatkujacych/, (access 2024.01.04)

(www_1.2) https://www.xblue.pl/co-to-jest-i-do-czego-sluzy-arkusz-kalkulacyjny/, (access 2024.01.04)

(www_1.3) https://www.projektgamma.pl/strefa-wiedzy/wiki-eksperckie/microsoft-office-excel-w-logistyce/, (access 2024.01.04)