MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

BUSINESS INTELLIGENCE IN FINANCE

Guidelines to laboratory and independent work

of Master's (second) degree higher education students
of speciality D2 "Finance, Banking,
Insurance and Stock Market"
of the study program "Finance and Credit"

Kharkiv
S. Kuznets KhNUE
2025

UDC 336:004(072.034)
B97

Compiled by: M. Berest
S. Leliuk

Затверджено на засіданні кафедри фінансів і кредиту.
Протокол № 8 від 14.01.2025 р.

Самостійне електронне текстове мережеве видання

B97 **Business** Intelligence in Finance [Electronic resource] : guidelines to laboratory and independent work of Master's (second) degree higher education students of speciality D2 "Finance, Banking, Insurance and Stock Market" of the study program "Finance and Credit" / compiled by M. Berest, S. Leliuk. – Kharkiv : S. Kuznets KhNUE, 2025. – 47 p. (English)

Tasks for laboratory classes and independent work of students on the academic discipline, guidelines and the order of doing the tasks are offered.

For Master's (second) degree higher education students of speciality D2 "Finance, Banking, Insurance and Stock Market" of the study program "Finance and Credit".

UDC 336:004(072.034)

Introduction

The academic discipline "Business Intelligence in Finance" belongs to the cycle of professional and practical training, to the group of compulsory educational components of master's training. The knowledge gained as a result of studying the discipline will help students to solve a wide range of tasks – from analyzing, processing and systematizing arrays of financial information to developing relational databases in MS Access, implementing work in environments for creating dashboards for corporate information systems of business entities.

The purpose of the academic discipline is to develop in-depth knowledge and skills of existing information systems and digital technologies for processing and accounting, rational use of financial information, as well as practical skills in the effective use of modern digital technologies in the process of financial and economic activities of enterprises and organizations.

The objectives of the academic discipline are: to study the theoretical and practical aspects of the selection and application of specific tools and instruments for automated analysis of financial and economic information, the peculiarities of the functioning of information in automated information systems, the formation of theoretical and practical knowledge of financial calculations using tabular databases, the formation of knowledge of the use of modern tools for the formation of financial data arrays and dashboards for tracking them.

The subject of the academic discipline is business intelligence tools that provide processing of financial and economic information.

The learning outcomes and competencies that the discipline forms are given in Table 1.

Table 1

Learning outcomes and competencies formed by the academic discipline

Learning outcomes	Competencies
PR04	SC7, SC9
PR05	GC2, GC8
PR08	SC8, SC9
PR09	SC9
PR10	SC6, SC7
PR15	GC2, SC11

PR04. Ability to search for, process, systematize and analyze information necessary for solving professional and scientific problems in the field of finance, banking, insurance and the stock market.

PR05. Ability to communicate fluently in a foreign language orally and in writing on professional and scientific issues, present and discuss research results.

PR08. Ability to apply and manage innovative approaches in the field of finance, banking, insurance and the stock market.

PR09. Ability to apply management skills in the field of finance, banking, insurance and stock market.

PR10. Ability to carry out diagnostics and modeling of financial activities of economic entities.

PR15. Ability to apply BI (Business Intelligence) tools to organize quick access to heterogeneous data from various sources, process it, and visualize it to support financial decision-making.

SC6. Ability to apply interdisciplinary approaches in solving complex problems and issues in the field of finance, banking, insurance and stock market.

SC7. Ability to search, use and interpret information necessary for solving professional and scientific problems in the field of finance, banking, insurance and stock market.

SC8. Ability to apply innovative approaches in the field of finance, banking, insurance and stock market.

SC9. Ability to develop technical specifications for the design of information systems in the field of finance, banking, insurance and stock market.

SC11. Ability to provide quick access to heterogeneous data from various sources, process and visualize them to support financial decision-making using a BI (Business Intelligence) tool.

GC2. Ability to communicate in a foreign language.

GC8. Ability to work in an international context.

Laboratory work 1
Formation of a database in MS Excel on the topic of analytical research

The purpose of the laboratory work is to determine the key characteristics of analytical research in the field of financial activities of business entities or financial institutions; to gain practical skills in collecting analytical information for research; to form a data set for further analytical research.

General guidelines

Data analysis is a highly iterative and non-linear process, better reflected by a series of epicycles, in which information is learned at each step, which then informs whether (and how) to refine, and redo, the step that was just performed, or whether (and how) to proceed to the next step. In data analysis, the iterative process that is applied to all steps of the data analysis can be conceived of as an epicycle that is repeated for each step along the circumference of the entire data analysis process. Some data analyses appear to be fixed and linear, such as algorithms embedded into various software platforms, including apps. However, these algorithms are final data analysis products that have emerged from the very non-linear work of developing and refining a data analysis so that it can be "algorithmized".

While a study includes developing and executing a plan for collecting data, a data analysis presumes the data have already been collected. More specifically, *a study* includes the development of a hypothesis or question, the designing of the data collection process (or study protocol), the collection of the data, and the analysis and interpretation of the data.

Because a data analysis presumes that the data have already been collected, it includes development and refinement of a question and the process of analyzing and interpreting the data. It is important to note that although a data analysis is often performed without conducting a study, it may also be performed as a component of a study.

For each of the core activities, it is critical to realize the following steps:

setting expectations;

collecting information (data), comparing the data to your expectations, and if the expectations don't match;

revising your expectations or fixing the data so your data and your expectations match.

Iterating through this 3-step process is the *"epicycle of data analysis"*.

There are 5 core activities of data analysis:

stating and refining the question;

exploring the data;

building formal statistical models;

interpreting the results;

communicating the results.

The repeated cycling through each of these five core activities that is done to complete a data analysis forms the larger circle of data analysis.

There are six basic types of questions. Understanding the type of question may be the most fundamental step to ensure that, in the end, interpretation of the results is correct. The six types of questions are:

a *descriptive question* is one that seeks to summarize a characteristic of a set of data. There is no interpretation of the result itself as the result is a fact, an attribute of the set of data that you are working with;

an *exploratory question* is one in which you analyze the data to see if there are patterns, trends, or relationships between variables. These types of analyses are also called "hypothesis-generating" analyses because rather than testing a hypothesis as would be done with an inferential, causal, or mechanistic question, you are looking for patterns that would support proposing a hypothesis;

an *inferential question* would be a restatement of this proposed hypothesis as a question and would be answered by analyzing a different set of data;

a *predictive question* makes it possible to determine the predictive behavior of the studied indicator, phenomenon or object;

a *causal question* asks about whether changing one factor will change another factor, on average, in a population. Sometimes the underlying design of the data collection, by default, allows for the question that you ask to be causal;

finally, a *mechanistic question* allows us to find the answer to how a change in one indicator affects the change in another indicator or the phenomenon under study.

There are five key characteristics of a good question for a data analysis, which range from the very basic characteristic that the question should not have already been answered to the more abstract characteristic that each of the possible answers to the question should have a single interpretation and be meaningful.

The question should be of interest to your audience, the identity of which will depend on the context and environment in which you are working with data. If you are in academia, the audience may be your collaborators, the scientific community, government regulators, your funders, and/or the public. If you are working at a startup, your audience is your boss, the company leadership, and the investors. You should also check that the question has not already been answered. With the recent explosion of data, the growing amount of publicly available data, and the seemingly endless scientific literature and other resources, it is not uncommon to discover that your question of interest has already been answered. The question, should also, of course, be answerable. Specificity is also an important characteristic of a good question.

A database, as a means of collecting and organizing information, contains information about indicators that characterize a particular process, phenomenon, object, etc.

Work in progress

Task 1. Determine the purpose of the analytical research in accordance with the topic of the student's research or in accordance with the topic of the master's thesis.

Task 2. Formulate a hypothesis and expectations for the research results. Determine what questions (by type) will be raised and solved in the course of the research, characterize them.

Task 3. Identify potential audience (stakeholders) of the research results.

Task 4. Using open sources of information on the Internet, form a set of relevant data for analytical research.

The data set should be presented in the form of tables and (if necessary and appropriate) graphs or diagrams in the form of a file in the Microsoft Excel spreadsheet editor. The data should meet the purpose and expectations regarding the results of the analytical study.

The data may relate to the financial activities of individual enterprises, banks, financial institutions, the banking system as a whole, financial, economic and budgetary indicators of the macro level, depending on the focus and purpose of the analytical study. The time period for data collection is at least 3 years, and if the research involves building models or contains forecasting questions, at least 10 time points. References to data sources are required.

Requirements for a report on laboratory work

The report on the laboratory work should contain:

1) a title page;

2) the purpose of the laboratory work;

3) a file with data in MS Excel;

4) a file in MS Word containing answers to the questions raised and conclusions reached in the progress of the laboratory work.

Checklist questions

1. Name and describe the iterations of data mining.
2. Identify and describe how expectations for data analysis are set.
3. What sources of information should be used in the process of generating data for analysis?
4. What types of analytical questions do you know? Describe them.
5. What is the need to revise the expectations of the analyst depending on the data collected?
6. Who can be potential users (stakeholders) of the results of analytical research in the financial sector?
7. Justify the need for correct interpretation of the results of analytical research.
8. What is the need for communication skills and presentation of research results?
9. Describe the characteristics of a qualitative research question.

Task for self-study

Based on open sources of Internet data, formulate the topic and purpose of the analytical research and create a data set in Microsoft Excel, taking into account that the objects of research can be any financial phenomena, processes, aspects of financial activities of business entities, banking institutions, budgetary institutions at both the micro and macro levels.

Please indicate your expectations regarding the results of the future study. Formulate analytical questions and name their types. Identify the potential audience interested in the research results.

Laboratory work 2
Analytical study of the data set

The purpose of the laboratory work is to perform analytical calculations in accordance with the formed expectations and questions for the generated data set, to form analytical conclusions.

General guidelines

There are *six basic types of questions*, as we could see in the general guidelines to laboratory work 1:

descriptive;
exploratory;
inferential;
predictive;
causal;
mechanistic.

And the type of question directly affects which mathematical methods and models you should choose to process the generated data set and how you interpret your results.

Exploratory data analysis is the process of exploring your data, and it typically includes examining the structure and components of your dataset, the distributions of individual variables, and the relationships between two or more variables. The most heavily relied upon tool for exploratory data analysis is visualizing data using a graphical representation of the data. *Data visualization* is arguably the most important tool for exploratory data analysis because the information conveyed by graphical display can be very quickly absorbed and because it is generally easy to recognize patterns in a graphical display.

Making a *plot* to visualize data is a good way to further understanding of question and data. Plotting can occur at different stages of a data analysis. For example, plotting may occur at the exploratory phase or later on in the presentation/communication phase.

There are two key reasons for making a plot of data. They are creating expectations and checking deviations from expectations. At the early stages of analysis, you may be equipped with a question/hypothesis, but you may have little sense of what is going on in the data. So, making some sort of plot, which serves as a summary, will be a useful tool for setting expectations for what the data should look like.

Plots are particularly good at letting you see deviations from what you might expect. Tables typically are good at summarizing data by presenting things like means, medians, or other statistics. Plots, however, can show you those things, as well as show you things that are far from the mean or median, so you can check to see if something is supposed to be that far away. Often, what is obvious in a plot can be hidden away in a table.

In addition to visualization and graphing, depending on the questions posed when forming the data set, it is advisable to use various methods of analytical research.

For example, if the research poses *descriptive* questions, it is advisable to use vertical and horizontal analysis methods to study the data set, accompanied by the construction of structural diagrams or bar charts.

A *horizontal analysis* is used to assess the development of financial phenomena and processes over time (dynamics), and is based on the study of the dynamics of individual financial indicators, items included in the financial statements or their groups. The most typical forms of horizontal analysis are:

assessment of the dynamics of financial indicators of the reporting and previous periods;

comparison of financial indicators of the reporting period with the planned ones;

assessment of the dynamics of financial indicators over a number of periods.

The purpose of this method is to identify trends in the indicators characterizing the financial activities of the enterprise. In the process of analyzing the indicators by series of dynamics, the following characteristics of values are calculated and used: absolute growth, growth rate, growth rate and the absolute value of one percent of growth.

A *vertical (structural) analysis* is aimed at determining the structure of financial indicators and identifying the impact of each reporting item on the overall result. It is used for indicators that have additive calculation models.

If the study contains *prognostic, exploratory and causal* questions, it is advisable to use a time-series forecasting method or a regression model to build predictive values of indicators.

The *trend-based approach to forecasting* economic indicators is based on extrapolating the leveled values of the dynamic time series of the forecasted indicator. Extrapolation means extending the patterns, relationships, and correlations identified in the period under study to the future. This means that the identified trends observed in the past are used to estimate future changes. Extrapolation is useful for prospective analysis of both the factors of influence and the final indicators.

Trend models allow for short- and medium-term forecasting under certain conditions:

the observation period should be sufficient to identify patterns;

the trend model should develop gradually;

the process must demonstrate inertia, i. e. significant changes take a long time;

data autocorrelation should decay rapidly, which means that newer data has a greater impact on the forecast.

The trend line is widely used to solve forecasting problems using regression analysis methods. The trend function is selected using the least squares method. To evaluate the accuracy of the model, the coefficient of determination is used, based on estimates of the variance of empirical data and the values of the trend model. The trend model shows a trend in the development of the process if the coefficient of determination tends to 1.

Correlation and regression analysis is the construction and analysis of an economic and mathematical model in the form of a regression equation (correlation equation) that reflects the dependence of the resultant feature on one or several attributes of the factors and provides an assessment of the degree of closeness of the relationship.

Correlation and regression analysis is a classic method of probabilistic modeling that examines the relationship between performance indicators when the dependence between them is not strictly functional or distorted by the influence of external (random) factors. As a result, the closeness of the relationship between two random attributes or factors is searched for and evaluated *(correlation analysis)*, and a specific type of dependence between the studied parameters is subsequently established *(regression analysis)*.

The process of building a model involves imposing a specific structure on the data and creating a summary of the data. The model is a mathematical equation that reflects the shape or pattern of the data, and the equation allows you to summarize the thousands of observations with one number.

A *statistical model* serves two key purposes in a data analysis, which are to provide a quantitative summary of your data and to impose a specific structure on the population from which the data were sampled. A statistical model must also impose some structure on the data. At its core, a statistical model provides a description of how the world works and how the data were generated. The model is essentially an expectation of the relationships between various factors in the real world and in your dataset. What makes a model a statistical model is that it allows for some randomness in generating the data.

Building statistical models is also often used to find answers to *inferential* and *mechanistic* questions.

It is also advisable to use the method of cluster analysis to find answers to *exploratory* and *inferential* questions.

Cluster analysis is a multivariate statistical research method that involves collecting data containing information about sample objects and organizing them into relatively homogeneous, similar groups. The purpose of cluster analysis is to search for existing structures, which is expressed in the formation of groups of similar objects – clusters.

Unlike other methods, this type of analysis makes it possible to classify objects not by one feature, but by several simultaneously. For this purpose, appropriate indicators are introduced that characterize a certain degree of proximity in all classification parameters.

A *cluster* is a term that refers to a group of objects with homogeneous characteristics. Cluster analysis solves the problem of classifying objects in the absence of a priori information about observations within the classes.

There are such clustering methods as:

joining (tree clustering);

k-means clustering;

two-way joining.

Cluster analysis methods can be divided into two main groups: hierarchical (natural) and non-hierarchical (artificial). It is recommended that both methods be used alternately.

The factor analysis is also appropriate for use in the process of exploratory data analysis. *Factor analysis* is a set of methods that, based on the actual relationships of features (or objects), allow identifying latent (hidden) generalized characteristics of the organizational structure and the mechanism of development of the phenomena and processes under study.

Factor analysis helps to solve two important problems of the researcher: to describe the object of measurement comprehensively and at the same time compactly. Factor analysis helps to identify hidden variables of factors responsible for the presence of linear statistical relationships between the observed variables.

The two main goals of factor analysis are:

determination of the relationships between variables, (classification of variables), i. e. objective R-classification;

reducing the number of variables required to describe the data.

In this context, factor analysis is used as a method of incomplete feature reduction. The essence of the task of reducing the dimensionality of the feature space is to present the initial information, which is given in the form of a sufficiently large number of feature descriptions, in a space of a smaller dimensionality, if possible, with minimizing the loss of information.

At the same time, it is possible to apply the method of full feature reduction – *taxonomic analysis*.

The main purpose of using the taxonomy method is to build a generalized assessment of a complex object or process. The taxonomic indicator is calculated according to the classical algorithm of taxonomic analysis, which includes the following steps:

formation of the observation matrix;

standardization of the values of the elements of the observation matrix;

identification of the reference vector;

determination of the distance between individual observations and the reference vector;

calculation of the taxonomic development coefficient.

The taxonomic indicator of the level of development is a synthetic value that is equally effective for all the features that characterize objects. This allows it to linearly organize the elements of the studied population.

Work in progress

Task 1. Justify the choice of analytical research methods based on the hypothesis, expectations from the study and the questions posed, taking into account their type.

Task 2. Using the data set generated in Laboratory work 1, perform analytical calculations using the methods selected above.

Task 2.1. Application of the time series forecasting method for forecasting financial indicators in the Microsoft Excel. (Requirements – the time series of the indicator is at least 10 points).

Step 1. Drawing a graph with markers for the indicator (Fig. 1).

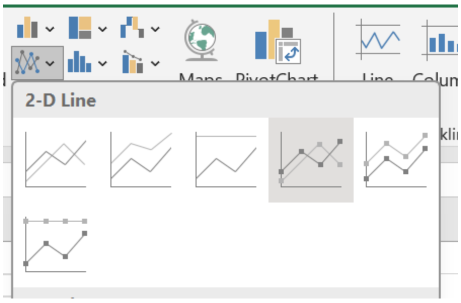


Fig. 1. **Inserting a graph with markers**

Step 2. Right-click on the chart line and select the "Add trend line" option. The "Trend Lines" window offers different types of trend lines. Select each of the available lines in sequence, setting the forecast period for 1 period ahead in the parameters tab, as well as specifying the options: "show equations on the chart" and "put the approximation probability value R^2 on the chart".

Step 3. Create trend lines for five types of approximations on different charts (linear, logarithmic, polynomial, power, exponential). Then, based on the value of the approximation coefficient, select the optimal function (Fig. 2).

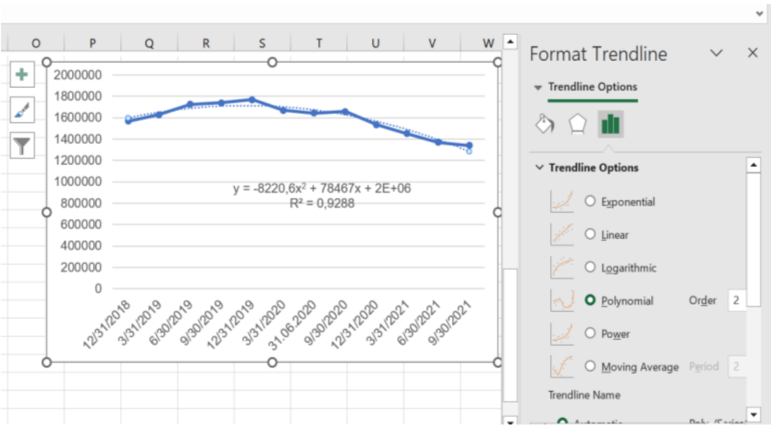


Fig. 2. **Plotting a trend line**

Justify your decision in your report, making sure to include the function and the value of the approximation coefficient.

Step 4. To determine the forecast value of the indicator in the dialog box, select the "Forward" option and specify +1 or another number of periods. The forecast value of the indicator will be displayed on the chart (Fig. 3).

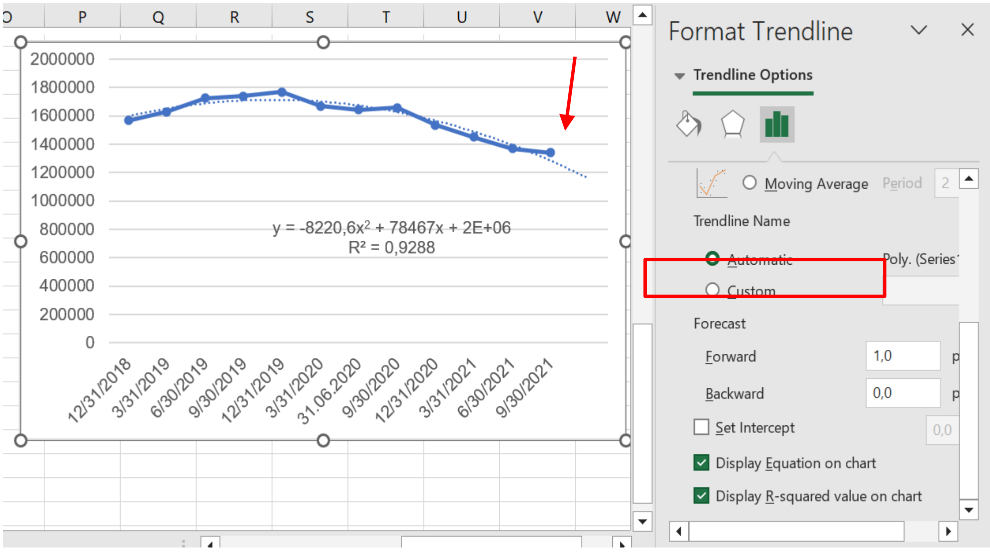


Fig. 3. **Building the forecast value of the indicator**

Thus, the result of applying the time-series forecasting method will be the forecast values of financial indicators.

Task 2.2. Application of the cluster analysis method for diagnosing and grouping objects with similar characteristics. (Requirements – availability of a data sample with at least 10 objects, each with at least 3 characteristics).

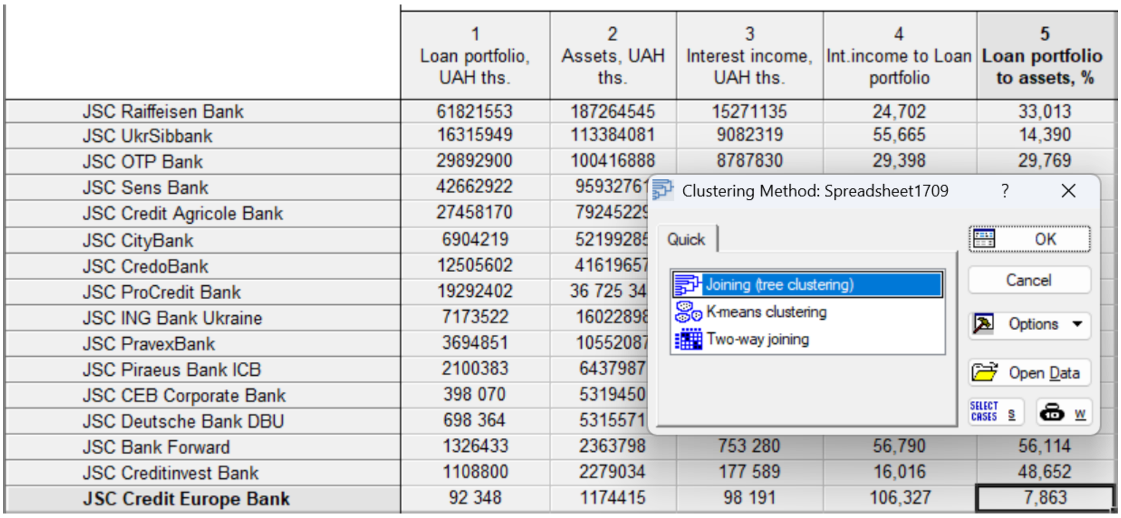
Step 1. Generate initial data in the Statistica (Fig. 4).

	1 Loan portfolio, UAH ths.	2 Assets, UAH ths.	3 Interest income, UAH ths.	4 Int.income to Loan portfolio	5 **Loan portfolio to assets, %**
JSC Raiffeisen Bank	61821553	187264545	15271135	24,702	33,013
JSC UkrSibbank	16315949	113384081	9082319	55,665	14,390
JSC OTP Bank	29892900	100416888	8787830	29,398	29,769
JSC Sens Bank	42662922	95932761	11659710	27,330	44,472
JSC Credit Agricole Bank	27458170	79245229	7047687	25,667	34,650
JSC CityBank	6904219	52199285	4797352	69,484	13,227
JSC CredoBank	12505602	41619657	3062659	24,490	30,047
JSC ProCredit Bank	19292402	36 725 345	3002210	15,562	52,532
JSC ING Bank Ukraine	7173522	16022898	1762616	24,571	44,770
JSC PravexBank	3694851	10552087	902 462	24,425	35,015
JSC Piraeus Bank ICB	2100383	6437987	560 003	26,662	32,625
JSC CEB Corporate Bank	398 070	5319450	363 120	91,220	7,483
JSC Deutsche Bank DBU	698 364	5315571	595 853	85,321	13,138
JSC Bank Forward	1326433	2363798	753 280	56,790	56,114
JSC Creditinvest Bank	1108800	2279034	177 589	16,016	48,652
JSC Credit Europe Bank	92 348	1174415	98 191	106,327	7,863

Fig. 4. **Input data for cluster analysis**

Before starting a cluster analysis, it is necessary to standardize the initial data.

Step 2. First, the initial data should be clustered using hierarchical classification (Fig. 5).



	1 Loan portfolio, UAH ths.	2 Assets, UAH ths.	3 Interest income, UAH ths.	4 Int.income to Loan portfolio	5 Loan portfolio to assets, %
JSC Raiffeisen Bank	61821553	187264545	15271135	24,702	33,013
JSC UkrSibbank	16315949	113384081	9082319	55,665	14,390
JSC OTP Bank	29892900	100416888	8787830	29,398	29,769
JSC Sens Bank	42662922	9593276			
JSC Credit Agricole Bank	27458170	7924522			
JSC CityBank	6904219	5219928			
JSC CredoBank	12505602	4161965			
JSC ProCredit Bank	19292402	36 725 34			
JSC ING Bank Ukraine	7173522	1602289			
JSC PravexBank	3694851	1055208			
JSC Piraeus Bank ICB	2100383	6437987			
JSC CEB Corporate Bank	398 070	5319450			
JSC Deutsche Bank DBU	698 364	5315571			
JSC Bank Forward	1326433	2363798	753 280	56,790	56,114
JSC Creditinvest Bank	1108800	2279034	177 589	16,016	48,652
JSC Credit Europe Bank	92 348	1174415	98 191	106,327	7,863

Fig. 5. **Selecting a data clustering method**

The result of clustering is a dendrogram (Fig. 6).

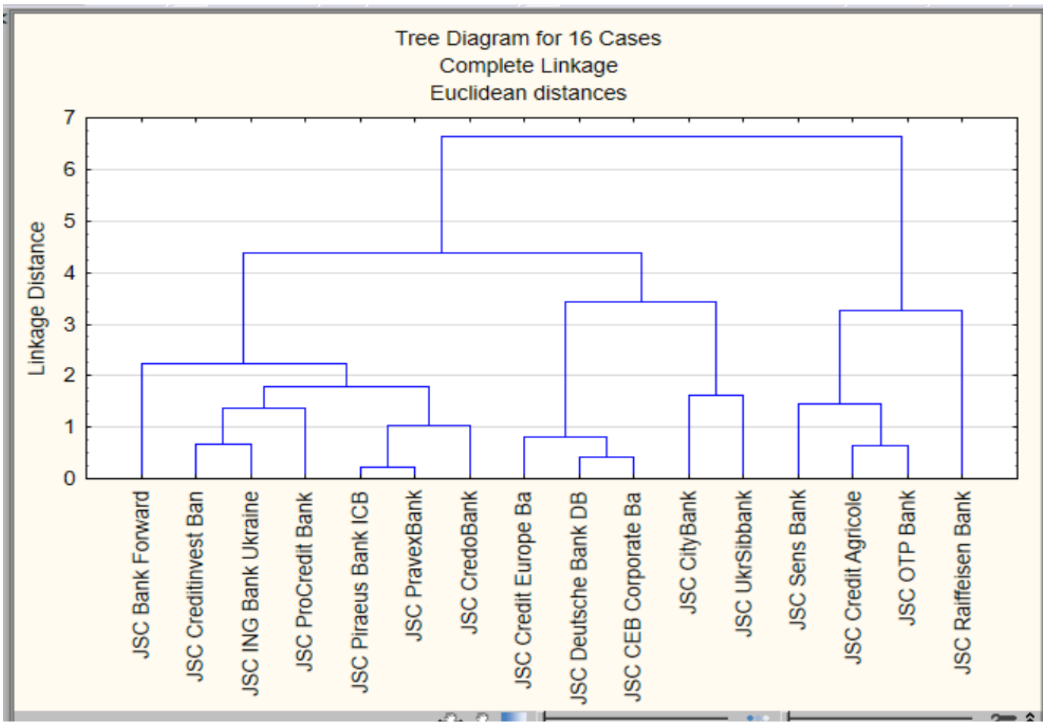


Fig. 6. **A vertical dendrogram constructed by the complete linkage method**

When performing clustering by hierarchical classification, a student can choose any method presented in the Statistica program (single linkage, complete linkage, Ward's method, etc.). The final choice of method is at the discretion of the student and must be justified in the laboratory report.

The student must provide an interpretation of the results – indicate and justify the number of clusters. Additionally, the Euclidean distances between the objects of observation and the scheme of clustering should be provided (Fig. 7 – 8).

	Euclidean distances (Spreadsheet1709_1)							
Case No.	**JSC Raiffeisen Bank**	JSC UkrSibbank	JSC OTP Bank	JSC Sens Bank	JSC Credit Agricole Bank	JSC CityBank	JSC CredoBank	JSC Pro
JSC Raiffeisen Bank	0,00	3,54	2,77	2,26	3,27	4,94	4,64	
JSC UkrSibbank	3,54	0,00	1,54	2,66	1,90	1,61	2,34	
JSC OTP Bank	2,77	1,54	0,00	1,32	0,64	2,47	1,90	
JSC Sens Bank	2,26	2,66	1,32	0,00	1,47	3,55	2,82	
JSC Credit Agricole	3,27	1,90	0,64	1,47	0,00	2,41	1,40	
JSC CityBank	4,94	1,61	2,47	3,55	2,41	0,00	1,93	
JSC CredoBank	4,64	2,34	1,90	2,82	1,40	1,93	0,00	
JSC ProCredit Bank	4,66	3,35	2,34	2,58	1,71	3,18	1,49	
JSC ING Bank Ukraine	5,29	3,26	2,67	3,24	2,07	2,66	1,11	
JSC PravexBank	5,51	3,14	2,79	3,57	2,25	2,35	0,94	
JSC Piraeus Bank ICB	5,65	3,17	2,91	3,73	2,39	2,28	1,03	
JSC CEB Corporate Ba	6,37	3,13	3,91	4,91	3,74	1,56	2,89	
JSC Deutsche Bank DB	6,19	2,99	3,66	4,64	3,45	1,39	2,56	
JSC Bank Forward	5,98	3,85	3,51	3,89	2,97	3,01	2,24	
JSC Creditinvest Ban	5,86	3,86	3,28	3,80	2,68	3,18	1,66	
JSC Credit Europe Ba	6,64	3,44	4,27	5,21	4,11	1,92	3,33	

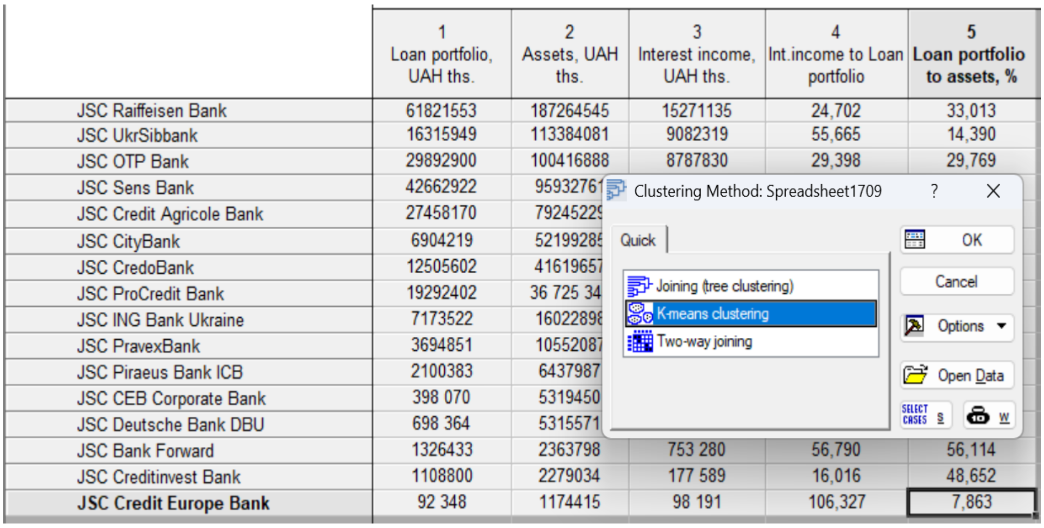
Fig. 7. **The Euclidean distances (part of a chart)**

	Amalgamation Schedule (Spreadsheet1709_1) Complete Linkage Euclidean distances						
linkage distance	**Obj. No. 1**	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. 7
,2172041	JSC PravexBank	JSC Piraeus Bank ICB					
,4105734	JSC CEB Corporate Bank	JSC Deutsche Bank DBU					
,6448078	JSC OTP Bank	JSC Credit Agricole Bank					
,6594165	JSC ING Bank Ukraine	JSC Creditinvest Bank					
,7977383	JSC CEB Corporate Bank	JSC Deutsche Bank DBU	JSC Credit Europe Bank				
1,034643	JSC CredoBank	JSC PravexBank	JSC Piraeus Bank ICB				
1,360995	JSC ProCredit Bank	JSC ING Bank Ukraine	JSC Creditinvest Bank				
1,465259	JSC OTP Bank	JSC Credit Agricole Bank	JSC Sens Bank				
1,610982	JSC UkrSibbank	JSC CityBank					
1,789486	JSC CredoBank	JSC PravexBank	JSC Piraeus Bank ICB	JSC ProCredit Bank	JSC ING Bank Ukraine	JSC Creditinvest Bank	
2,242403	JSC CredoBank	JSC PravexBank	JSC Piraeus Bank ICB	JSC ProCredit Bank	JSC ING Bank Ukraine	JSC Creditinvest Bank	JSC
3,269287	JSC Raiffeisen Bank	JSC OTP Bank	JSC Credit Agricole Bank	JSC Sens Bank			
3,440005	JSC UkrSibbank	JSC CityBank	JSC CEB Corporate Bank	JSC Deutsche Bank DBU	JSC Credit Europe Bank		
4,393947	JSC UkrSibbank	JSC CityBank	JSC CEB Corporate Bank	JSC Deutsche Bank DBU	JSC Credit Europe Bank	JSC CredoBank	JS
6,642991	JSC Raiffeisen Bank	JSC OTP Bank	JSC Credit Agricole Bank	JSC Sens Bank	JSC UkrSibbank	JSC CityBank	JSC CEB C

Fig. 8. **The amalgamation schedule (part of a chart)**

As a result, in Fig. 6 we can distinguish 3 clusters, where the members of the first cluster are JSC Bank Forward, JSC Creditinvest Bank, JSC ING Bank Ukraine, JSC ProCredit Bank, JSC Piraeus Bank ICB, JSC CredoBank, JSC PravexBank. The second cluster includes JSC Credit Europe Bank, JSC Deutsche Bank DBU, JSC CEB Corporate Bank, JSC CityBank, and JSC UkrSibbank. The third cluster included such banks as Sens Bank, Credit Agricole Bank, OTP Bank, Raiffeisen Bank.

Step 3. The next step is to perform a cluster analysis using the k-means method (Fig. 9), in which the number of clusters is initially set. In this case, the number of clusters is 3.



	1 Loan portfolio, UAH ths.	2 Assets, UAH ths.	3 Interest income, UAH ths.	4 Int.income to Loan portfolio	5 **Loan portfolio to assets, %**
JSC Raiffeisen Bank	61821553	187264545	15271135	24,702	33,013
JSC UkrSibbank	16315949	113384081	9082319	55,665	14,390
JSC OTP Bank	29892900	100416888	8787830	29,398	29,769
JSC Sens Bank	42662922	95932761			
JSC Credit Agricole Bank	27458170	79245229			
JSC CityBank	6904219	52199285			
JSC CredoBank	12505602	41619657			
JSC ProCredit Bank	19292402	36 725 34			
JSC ING Bank Ukraine	7173522	16022898			
JSC PravexBank	3694851	10552087			
JSC Piraeus Bank ICB	2100383	6437987			
JSC CEB Corporate Bank	398 070	5319450			
JSC Deutsche Bank DBU	698 364	5315571			
JSC Bank Forward	1326433	2363798	753 280	56,790	56,114
JSC Creditinvest Bank	1108800	2279034	177 589	16,016	48,652
JSC Credit Europe Bank	92 348	1174415	98 191	106,327	7,863

Fig. 9. **A window for selecting the k-means method**

The graph of means for the clusters is shown in Fig. 10.

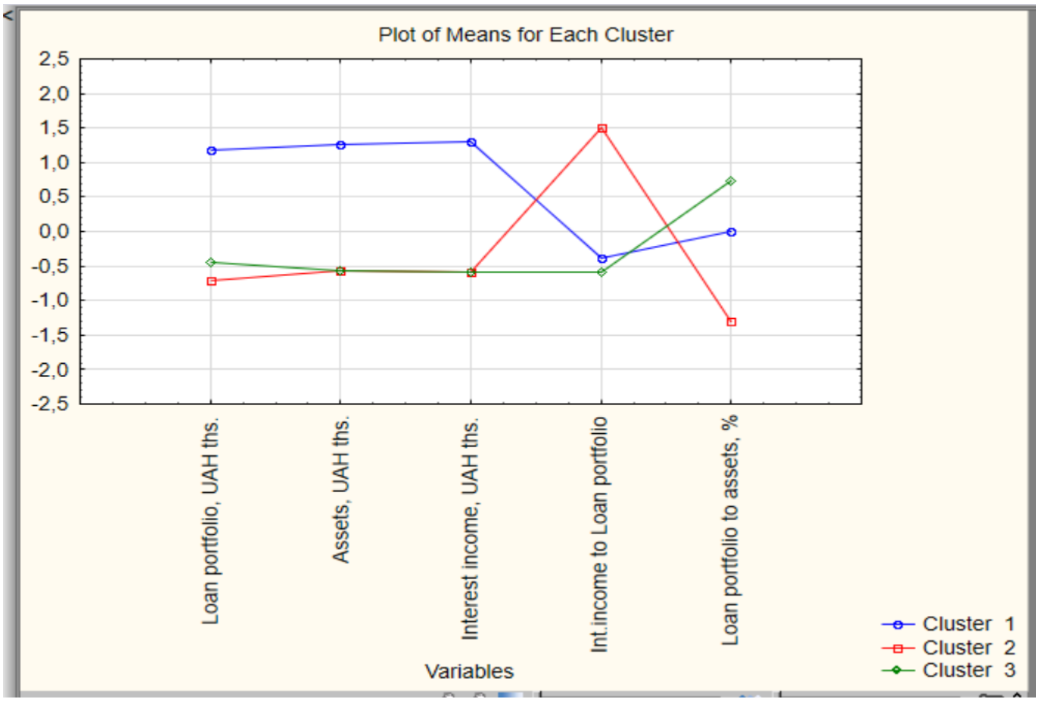


Fig. 10. **The graph of means for clusters**

The quality of the cluster analysis should be determined by analysis of variance (Fig. 11).

Variable	Analysis of Variance (Spreadsheet134) Between SS	df	Within SS	df	F	signif. p
Loan portfolio, UAH ths.	10,23993	2	4,760071	13	13,98289	0,000575
Assets, UAH ths.	11,45131	2	3,548690	13	20,97493	0,000085
Interest income, UAH ths.	12,14494	2	2,855061	13	27,64989	0,000021
Int.income to Loan portfolio	12,07652	2	2,923483	13	26,85063	0,000024
Loan portfolio to assets, %	10,50452	2	4,495478	13	15,18846	0,000397

Fig. 11. **Results of the analysis of variance**

Since the p-value for each of the factors does not exceed 0.05, it is inappropriate to exclude any indicator from clustering.

Descriptive characteristics for each cluster are shown in Fig. 12.

Variable	Descriptive Statistics for Cluster 1 (Sprea Cluster contains 5 cases Mean	Standard Deviation	Variance
Loan portfolio, UAH ths.	1,176274	0,971632	0,944068
Assets, UAH ths.	1,254797	0,776485	0,602929
Interest income, UAH ths.	1,292257	0,674566	0,455039
Int.income to Loan portfolio	-0,388009	0,442971	0,196224
Loan portfolio to assets, %	0,009317	0,684012	0,467873

Variable	Descriptive Statistics for Cluster 2 (Spre Cluster contains 4 cases Mean	Standard Deviation	Variance
Loan portfolio, UAH ths.	-0,70259	0,182445	0,033286
Assets, UAH ths.	-0,57705	0,446863	0,199687
Interest income, UAH ths.	-0,58692	0,470899	0,221745
Int.income to Loan portfolio	1,49814	0,517393	0,267696
Loan portfolio to assets, %	-1,29694	0,199702	0,039881

Variable	Descriptive Statistics for Cluster 3 (Sprea Cluster contains 7 cases Mean	Standard Deviation	Variance
Loan portfolio, UAH ths.	-0,438716	0,383827	0,147323
Assets, UAH ths.	-0,566541	0,299421	0,089653
Interest income, UAH ths.	-0,587659	0,248217	0,061612
Int.income to Loan portfolio	-0,578929	0,471788	0,222583
Loan portfolio to assets, %	0,734451	0,646058	0,417391

Fig. 12. **Descriptive characteristics of clusters**

The composition and characteristics of the clusters are shown in Fig. 13.

	Members of Cluster Number 1 (Spreadsheet1709_1) and Distances from Respective Cluster Center Cluster contains 5 cases
	Distance
JSC Raiffeisen Bank	1,006349
JSC UkrSibbank	0,771501
JSC OTP Bank	0,248776
JSC Sens Bank	0,463412
JSC Credit Agricole Bank	0,498312

	Members of Cluster Number 2 (Spreadsheet1709_1) and Distances from Respective Cluster Center Cluster contains 4 cases
	Distance
JSC CityBank	0,537707
JSC CEB Corporate Bank	0,171150
JSC Deutsche Bank DBU	0,152075
JSC Credit Europe Bank	0,340333

	Members of Cluster Number 3 (Spreadsheet1709_1) and Distances from Respective Cluster Center Cluster contains 7 cases
	Distance
JSC CredoBank	0,464862
JSC ProCredit Bank	0,501252
JSC ING Bank Ukraine	0,072252
JSC PravexBank	0,245822
JSC Piraeus Bank ICB	0,330884
JSC Bank Forward	0,617392
JSC Creditinvest Bank	0,320345

Fig. 13. **The composition and characteristics of clusters**

Thus, as a result of the cluster analysis, the number of clusters should be determined and their economic characteristics should be formed. In this case, we can see that according to the results of using the k-means method, the first cluster includes Sens Bank, Credit Agricole Bank, OTP Bank, Raiffeisen Bank, JSC PravexBank. The second cluster united JSC Credit Europe Bank, JSC Deutsche Bank DBU, JSC CEB Corporate Bank, JSC CityBank, and the third one included JSC Bank Forward, JSC Creditinvest Bank, JSC ING Bank Ukraine, JSC ProCredit Bank, JSC Piraeus Bank ICB, JSC CredoBank, JSC PravexBank. The first cluster was formed by banks that are characterized by a large scale of operations, but have an average share of the loan portfolio in assets and a rather low level of lending efficiency (the ratio of interest income to the size of the loan portfolio). The second and third clusters include banks with a significantly smaller scale of operations, but the banks in the second cluster have the lowest share of loans in assets, while the efficiency of lending operations is the highest among the banks studied. Banks in the third cluster, on the contrary, have the highest share of lending operations, but the lowest level of their profitability.

Thus, the cluster analysis made it possible to diagnose a group of Ukrainian banks in terms of the efficiency of their lending activities, given their scale.

Requirements for a report on laboratory work

The report on the laboratory work should contain:

1) a title page;

2) the purpose of the laboratory work;

3) a file with data in MS Excel, Statistica or another program student used;

4) a file in MS Word containing answers to the questions raised and conclusions reached in the progress of the laboratory work.

Checklist questions

1. Describe the characteristics and scope of use of a descriptive question.
2. Describe the characteristics and scope of use of an exploratory question.
3. Describe the characteristics and scope of use of an inferential question.
4. Describe the characteristics and scope of use of a predictive question.
5. Describe the characteristics and scope of use of a causal question.
6. Describe the characteristics and scope of use of a mechanistic question.
7. What is data visualization and what tools should be used for data visualization?
8. What is the feasibility and necessity of building graphs and charts when analyzing data?
9. Describe the methods of horizontal and vertical analysis. What is the purpose of using them?
10. Describe the method of trend forecasting. In your opinion, what type of questions should be used to answer in this case?
11. Describe the method of cluster analysis. In your opinion, what type of questions should be used to answer?
12. Describe the method of factor analysis. In your opinion, what type of questions should be used to answer?

13. Describe the method of taxonomic analysis. In your opinion, what type of questions should be used to answer?

Task for self-study

Using the data set generated in Laboratory work 1, formulate research questions, indicating their type (at least 2 types of questions should be selected). Justify the choice of the type of questions based on the generated data set.

Using the proposed methods (at the student's choice), conduct analytical calculations using MS Excel, Statistica or another program depending on the chosen research methods. Prepare a report, formulate analytical conclusions.

Laboratory work 3
Building statistical models for data analysis and forecasting

The purpose of the laboratory work is building statistical models for data analysis and forecasting in accordance with the formed set of financial and economic indicators.

General guidelines

In a very general sense, a model is something we construct to help us understand the real world. For example, a financier has retrospective data on a company's revenues, expenses, and profits and wants to use it to predict the company's financial position in the future. The data analyst uses the data to construct a model to predict what will happen with the company's financial position in the future. The process of building a model involves imposing a specific structure on the data and creating a summary of the data. In the financial data example, you may have time series of observations, so the model is a mathematical equation that reflects the shape or pattern of the data, and the equation allows you to summarize the observations with, for example, one number, which might be an indicator of the probability of a company's bankruptcy.

A statistical model serves two key purposes in a data analysis, which are to provide a quantitative summary of your data and to impose a specific structure on the population from which the data were sampled. It's sometimes helpful to understand what a model is and why it can be useful through the illustration of extreme examples.

The first key element of a statistical model is data reduction. The basic idea is you want to take the original set of numbers consisting of your dataset and transform them into a smaller set of numbers. If you originally started with 20 numbers, your model should produce a summary that is fewer than 20 numbers. The process of data reduction typically ends up with a statistic. Generally speaking, a statistic is any summary of the data. The sample mean, or average, is a statistic. So is the median, the standard deviation, the maximum, the minimum, and the range. Some statistics are more or less useful than others but they are all summaries of the data.

But a simple summary statistic, such as the mean of a set of numbers, is not enough to formulate a model. A statistical model must also impose some structure on the data. At its core, a statistical model provides a description of how the world works and how the data were generated. The model is essentially an expectation of the relationships between various factors in the real world and in your dataset. What makes a model a statistical model is that it allows for some randomness in generating the data.

The usefulness of a statistical model will depend on how accurately it reflects the data we collect in the real world. The following steps should be taken to check whether the data collected reflects reality:

1. Develop expectations: draw a fake picture – what do we expect to see before looking at the data?
2. Compare our expectations to the data.
3. Refine our expectations, given what the data show.

It's common to look at data and try to understand linear relationships between variables of interest. The most common statistical technique to help with this task is linear regression. The simplest model that we might formulate for characterizing the relationship between financial performance, as we are investigating, is a linear model. This model says that, for example, if the company's receivables increase, its profit decreases. Statistical theory suggests a number of different approaches to determining when a statistical model is "good enough" and fits the data well.

Model building, like the entire process of data analysis itself, is an iterative process. Models are used to provide data reduction and to give you some insight into the population about which you are trying to make inference. It's important to first set your expectations for a how a model should characterize a dataset before you actually apply a model to data. Then you can check to see how your model conforms to your expectation. Often, there will be features of the dataset that do not conform to your model and you will have to either refine your model or examine the data collection process.

Associational analyses are ones where we are looking at an association between two or more features in the presence of other potentially confounding factors. There are three classes of variables that are important to think about in an associational analysis.

1. The *outcome* is the feature of your dataset that is thought to change along with your key predictor.

2. *Key predictor.* Often for associational analyses there is one key predictor of interest (there may be a few of them). We want to know how the outcome changes with this key predictor. However, our understanding of that relationship may be challenged by the presence of potential confounders.

3. *Potential confounders*. This is a large class of predictors that are both related to the key predictor and the outcome. It's important to have a good understanding what these are and whether they are available in your dataset.

Once you have identified these three classes of variables in your dataset, you can start to think about formal modeling in an associational setting.

The basic form of a model in an associational analysis is:

$$y = a + bx + cz + d,$$

where y is the outcome;

a is the intercept, i. e. the value y when x = 0 and z = 0;

b is the change in y associated with a 1-unit increase x, adjusting for z;

x is the key predictor;

c is the change in y associated with a 1-unit increase in z, adjusting for x;

z is a potential confounder;

d is independent random error.

This is a linear model, and our primary interest is in estimating the coefficient b, which quantifies the relationship between the key predictor x and the outcome y.

The model shown above could be thought of as the primary model. There is a key predictor and one confounder in the model where it is perhaps well known that you should adjust for that confounder. This model may produce sensible results and follows what is generally known in the area. But sometimes the goal is to use all of the information available to you to predict *y*.

With prediction models, we have outcome variables – features about which we would like to make predictions – but we typically do not make a distinction between "key predictors" and other predictors. In most cases, any predictor that might be of use in predicting the outcome would be considered in an analysis and might, *a priori*, be given equal weight in terms of its importance in predicting the outcome.

Prediction analyses will often leave it to the prediction algorithm to determine the importance of each predictor and to determine the functional form of the model. For many prediction analyses it is not possible to literally write down the model that is being used to predict because it cannot be represented using standard mathematical notation. Many modern prediction routines are structured as algorithms or procedures that take inputs and transform them into outputs. The path that the inputs take to be transformed into outputs may be highly nonlinear and predictors may interact with other predictors on the way.

Typically, there are no parameters of interest that we try to estimate – in fact many algorithmic procedures do not have any estimable parameters at all.

The key thing to remember with prediction analyses is that we usually do not care about the specific details of the model. In most cases, as long as the method "works", is reproducible, and produces good predictions with minimal error, then we have achieved our goals.

With prediction analyses, the precise type of analysis you do depends on the nature of the outcome (as it does with all analyses). Prediction problems typically come in the form of a classification problem where the outcome is binary. In some cases the outcome can take more than two levels, but the binary case is by far the most common.

Work in progress

Task. Use the regression analysis method to build a model (requirements: the length of the time series for the dependent and independent variables is at least 10 years).

Step 1. Define the dependent and independent variables in the dataset.

The dependent variable (y) is a variable that describes the process that needs to be predicted or understood. Independent variables (X) are variables used to model or predict the values of dependent variables.

Using the *Statistica*, create a file with the initial data for building the model (Fig. 14).

1 R1	2 R2	3 R3	4 R4	5 R5	6 R6	7 R7	8 R8	9 R9	10 Int_fr
0,0004179	1,06834043	12,6074466	0,07400564	0,85807432	2,0939834	0,91704918	0,07513444	0,69671813	0,480
0,07980632	3,2636022	0,92860708	0,83650475	2,90763482	25,3656573	1,05690009	0,41867873	0,49084113	0,117
0,00121212	0,70757727	1,81519329	0,48794051	17,6218014	4,92661062	0,9956803	0,03168899	0,09944122	0,349
0,10169689	1,05429024	11,1291732	0,08611009	14,3727279	5,19915531	0,99204158	0,02940365	1,09263838	0,355
0,67840232	2,99763247	1,53480084	0,6906836	4,16353042	8,84936708	1,48006042	0,33127466	1,16498651	0,022
0,00040803	1,07830402	50,9391418	0,07274038	2,07022176	2,89403561	0,95128577	0,00252968	0,11976984	0,642
0,00019974	1,00147442	0,07801331	0,92763233	0,02653773	0	0,96551493	-0,0326057	-0,0000314	0,392
0,00702795	0,23821622	1,1674212	0,58769157	1,10375168	3,73458091	0,93995137	-1,086851	-0,2691936	0,565
0,01963322	1,12083348	1,92133029	0,41644703	7,35963824	15,3068769	1,06862889	0,03874083	0,12162104	0,379
0,09839637	1,15510259	3,14875817	0,24110174	6,05058477	9,36485412	0,90249799	-0,0200404	-0,1538286	0,459
0,02338782	1,05897952	14,989489	0,06273807	0,64726559	3,18265457	1,04148063	-0,0065224	-0,0502143	0,554
0,02107291	2,66442345	0,91009341	0,77965624	3,40304526	15,8753968	1,09533724	0,26113826	0,31512418	0,141
0,01065285	0,30248149	7,71856546	0,19367799	7,1943684	2,01101925	0,838099	-0,6157783	-1,1111186	0,646
0,13763604	1,51862394	1,62360838	0,38496697	8,73772457	3,66831644	1,12682922	0,07657831	0,94717556	0,126
0,00134931	1,93347208	5,20035071	0,53510477	3,66410731	4,17302342	0,73222942	-0,0220564	-0,0636754	0,426
0,00034939	1,16499882	18,6418632	0,14174386	1,45882157	3,96509455	1,02539518	0,00998765	0,30999153	0,505
0,00002312	1,00015172	1,11403585	0,47302888	0,00444392	36,4817927	0,96342573	-0,0703317	-0,0000996	0,334
0,04188714	0,40361961	0,72135128	0,73444382	1,77219165	4,33656906	1,27764857	-0,4948095	-0,1436357	0,351
0,10845395	1,78318599	1,30004082	0,60956814	7,88063007	17,455336	1,10823724	0,05035287	0,1671174	0,192
0,08549374	1,27334821	2,39510886	0,29500623	3,12175865	5,99228121	1,257479	0,07564622	0,39891305	0,274
0,0063694	1,03831516	19,7476182	0,0481983	0,76502937	12,1578963	1,00652984	0,00121699	0,01564671	0,620
0,04862619	1,76241736	1,21847255	0,72687622	3,6676179	8,48044973	1,09533724	0,00735244	0,00959949	0,399
0,05711122	0,32171686	14,86665	0,13439702	26,8582803	4,7921669	0,838099	-0,0530007	-0,5011611	0,535

Fig. 14. **Initial data for model building**

The input data for the analysis in this example are:

Int_fr is integral indicator of the level of company's financial risk;

R1 is cash ratio;

R2 is current ratio;

R3 is financial leverage ratio;

R4 is financial stability ratio;

R5 is receivable turnover ratio;

R6 is payable turnover ratio;

R7 is revenue growth rate to cost price rate;

R8 is net profit margin;

R9 is return on equity (ROE).

Step 2. Build a regression model using the standard method in the Statistica program (Fig. 15).

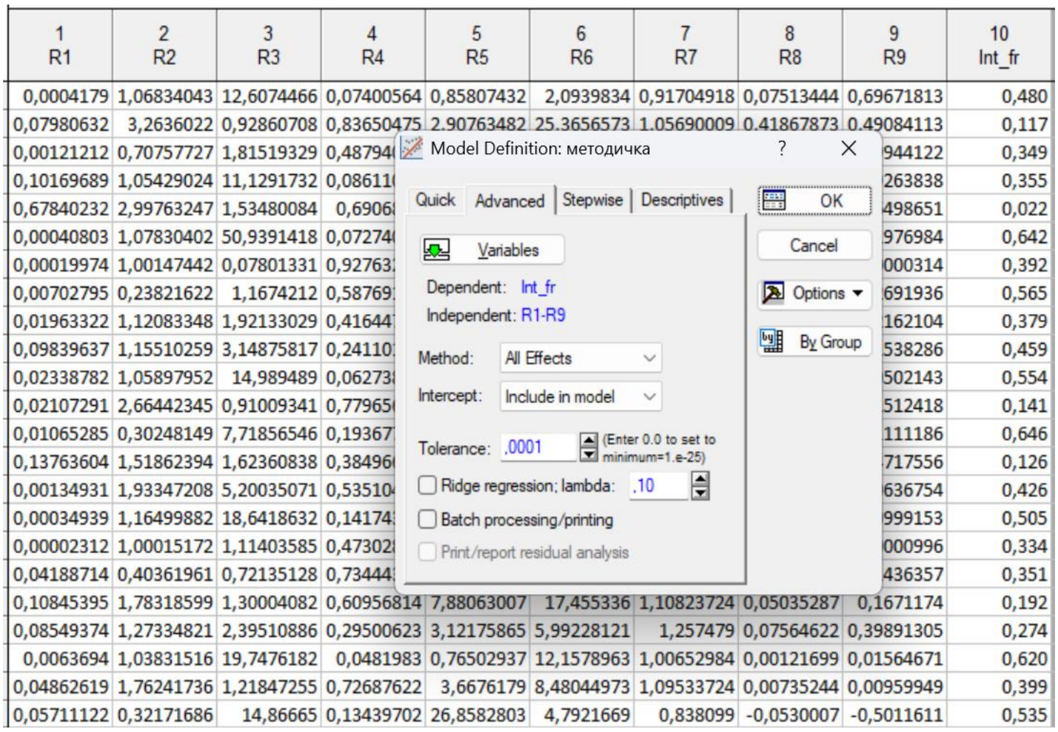


Fig. 15. **Choosing a standard method for building a model**

The result of the model is shown in Fig. 16. Thus, we can see that the model built by the standard method is:

$$y = 0.827944 + 0.003579R3 - 0.22688R4 - 0.00587R5 - 0.002804R6 - \\ - 0.279864R7 - 0.035906R9.$$

R^2 (determination coefficient) indicates how much of the variation in the dependent variable is explained by the independent variables included in the model. It's equal to 0.92104092, which indicates a high quality of the model.

Regression Summary for Dependent Variable: Int_fr (model)
R= ,95970877 R?= ,92104092 Adjusted R?= ,88550934
F(9,20)=25,922 p<,00000 Std.Error of estimate: ,06181

N=30	b*	Std.Err. of b*	b	Std.Err. of b	t(20)	p-value
Intercept			0,827944	0,115169	7,18897	0,000001
R1	-0,039654	0,094217	-0,047727	0,113399	-0,42088	0,678331
R2	-0,026242	0,097511	-0,004046	0,015033	-0,26912	0,790593
R3	0,203759	0,092314	0,003579	0,001621	2,20724	0,039146
R4	-0,347363	0,107834	-0,226888	0,070434	-3,22128	0,004283
R5	-0,188804	0,070011	-0,005870	0,002177	-2,69676	0,013876
R6	-0,159607	0,068934	-0,002804	0,001211	-2,31536	0,031333
R7	-0,251060	0,094217	-0,279864	0,105027	-2,66468	0,014883
R8	-0,152299	0,085978	-0,096616	0,054543	-1,77137	0,091738
R9	-0,342329	0,089763	-0,136935	0,035906	-3,81367	0,001087

Fig. 16. **Results of model building**

Step 3. Build a regression model using the Forward stepwise method (Fig. 17).

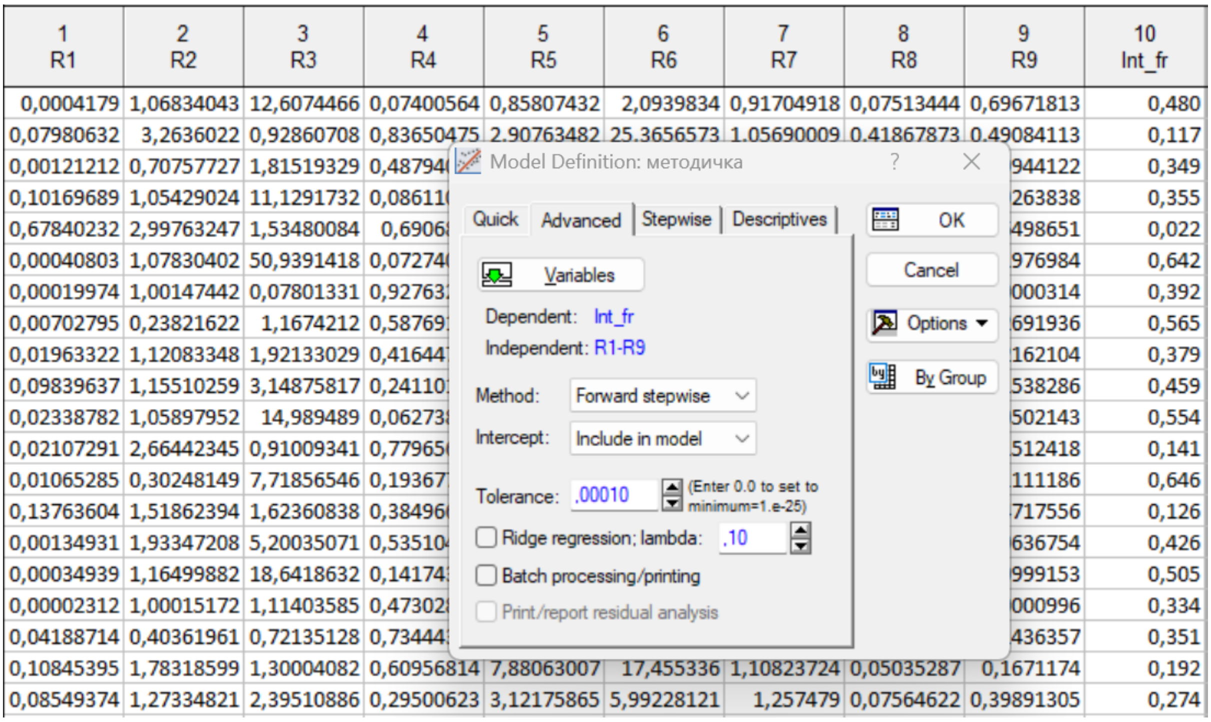


Fig. 17. **Choosing the Forward stepwise method for building a model**

The result of the model is shown in Fig. 18. As we can see, the coefficient of determination of the model R^2 is 0.91938722, which is lower than in the previous version.

Regression Summary for Dependent Variable: Int_fr (model)
R= ,95884682 R?= ,91938722 Adjusted R?= ,89373770
F(7,22)=35,844 p<,00000 Std.Error of estimate: ,05955

N=30	b*	Std.Err. of b*	b	Std.Err. of b	t(22)	p-value
Intercept			0,844612	0,097033	8,70441	0,000000
R7	-0,265879	0,077067	-0,296383	0,085909	-3,44997	0,002282
R4	-0,369604	0,086276	-0,241415	0,056353	-4,28398	0,000302
R9	-0,359359	0,082658	-0,143747	0,033064	-4,34756	0,000258
R5	-0,197661	0,064992	-0,006145	0,002021	-3,04132	0,005991
R6	-0,160022	0,065415	-0,002812	0,001149	-2,44627	0,022892
R3	0,198729	0,086901	0,003490	0,001526	2,28685	0,032191
R8	-0,156736	0,077599	-0,099430	0,049228	-2,01980	0,055747

Fig. 18. **Results of model building**

Step 4. Build a regression model using the Backward stepwise method (Fig. 19).

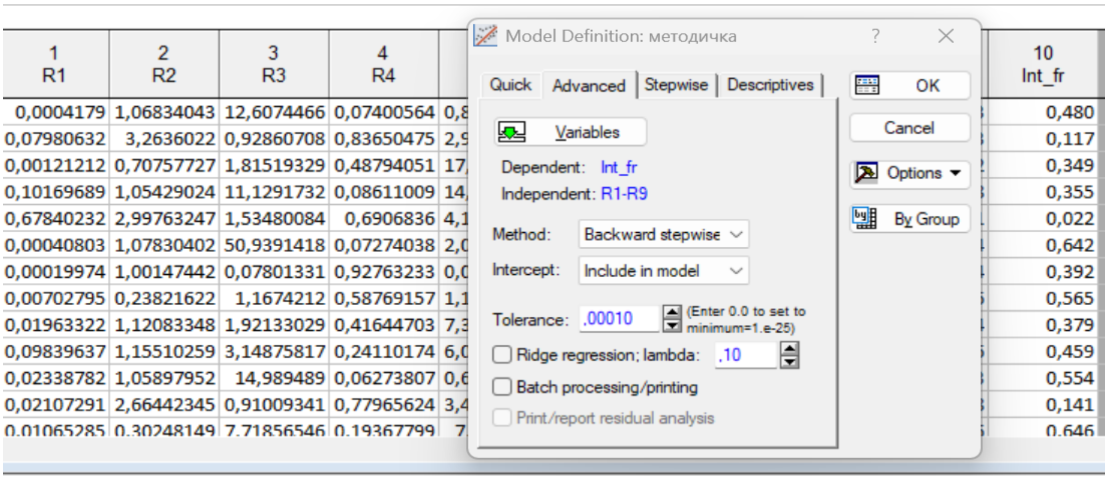


Fig. 19. **Choosing the Backward stepwise method for building a model**

The result of the model is shown in Fig. 20. We can see that the coefficient of determination is 0.76924869, which is significantly lower than both previous versions. In addition, only the coefficients R4 and R9 are included in the model parameters. Thus, we can conclude that the regression model obtained by the standard method is the best in terms of R2 (closest to 1). From the list of indicators that affect the level of financial risk of the studied enterprises, we should exclude *R1* – cash ratio; *R2* – current ratio and *R8* – net profit margin as insignificant.

	Regression Summary for Dependent Variable: Int_fr R= ,87706824 R?= ,76924869 Adjusted R?= ,75215600 F(2,27)=45,005 p<,00000 Std.Error of estimate: ,09094					
N=30	b*	Std.Err. of b*	b	Std.Err. of b	t(27)	p-value
Intercept			0,581032	0,031759	18,29487	0,000000
R4	-0,590817	0,092992	-0,385905	0,060740	-6,35343	0,000001
R9	-0,587469	0,092992	-0,234994	0,037198	-6,31744	0,000001

Fig. 20. **Results of model building**

The construction of a statistical model can also be done in MS Excel or other programs that allow for this type of calculation.

Requirements for a report on laboratory work

The report on the laboratory work should contain:

1) a title page;

2) the purpose of the laboratory work;

3) a file with data in MS Excel or Statistica;

4) a file in MS Word containing answers to the questions raised and conclusions reached in the progress of the laboratory work.

Checklist questions

1. Describe what a statistical model is.
2. Specify the purpose of building and assigning statistical models.
3. Specify the requirements for the data used to build statistical models.
4. Describe the steps that should be taken to check whether the data collected reflects financial processes under study.
5. Describe what linear regression is.

Task for self-study

Using the data set generated in Laboratory work 1, formulate the purpose of building a statistical model and the hypothesis regarding the expected results.

Using the proposed methods of building statistical models, conduct analytical calculations using MS Excel, Statistica or another program depending on the chosen research methods. Prepare a report, formulate analytical conclusions.

Laboratory work 4
Formation of data arrays using SQL

The purpose of the laboratory work is to gain skills in building SQL queries to retrieve financial data.

General guidelines

SQL (Structured Query Language) is a structured query language used to work with databases. The elements of the SQL language syntax are: Literals, Identifiers, Commands, Keywords, Functions and Comments.

Literals are fixed values, such as numbers, strings, dates, and others, expressed directly in a query, that are used to specify specific data or conditions. They are explicitly specified values, rather than referring to data stored in tables or calculated during query execution.

Identifiers are names for various database objects (tables, columns, indexes, procedures, etc.).

SQL commands can be divided into: DDL (Data Definition Language) – create, alter, drop; DML (Data Manipulation Language) – select, update, insert, delete; DCL (Data Control Language) – grant, revoke; DQL (Data Query Language) – select.

Keywords are reserved words that have a special meaning and are used to define operations and database structure. Keywords cannot be used for names of tables, columns, or other database objects:

where – filters a result set to include only records that fulfill a specified condition;

between – selects values within a given range;

and/or – includes rows where both/either condition(s) is (are) true;

like – looking for a specified pattern in a column;

in – allows you to specify multiple values in a *where* clause;

order by – sorts the result set in ascending order by default (ASC) or in descending order, if the DESC keyword is used.

Comments allow you to add code explanations, do not affect the execution of it, but help developers better understand the logic of requests and data structures.

To perform various mathematical operations with numerical values, numerical functions are used in SQL (Table 2).

Table 2

SQL numeric functions in MS Access

Function	Result of execution
Abs	The absolute value of a number
Avg	The average value of an expression
Count	The number of records returned by a select query
Max/Min	The maximum/minimum value in a set of values
Round	Rounds a number to a specified number of decimal places
Sum	The sum of a set of values

Work in progress

Task. Load primary financial data into the MS Access DBMS. Create a query for the imported data table using the Wizard. Configure queries in SQL mode using commands and numeric functions.

Step 1. Create data array in MS Excel (Fig. 21). The data array must consist of at least 10 columns – financial indicators, and 10 rows displaying the values of the indicators for 10 years.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Period	Income	Expenses	Net profit/loss	Total deposits by maturity	Deposits On demand	Deposits Short-term	Deposits Medium-term	Deposits Long-term	Deposits in national currency	Deposits in foreign currency	Total credits by terms of validity	Short-term loans	Medium-term loans	Long-term loans	Loans In national currency	Loans in foreign currency	Interest rate on loans
2	2009	142 996	181 445	-38 450	334 953	119 793	135 357	60 425	19 378	173 091	161 862	723 295	230 974	266 699	225 623	355 521	367 774	18,26
3	2010	136 848	149 875	-13 027	416 650	152 477	133 499	106 718	23 956	239 302	177 348	732 823	242 772	275 605	214 446	395 504	337 319	14,61
4	2011	142 778	150 486	-7 708	491 756	174 959	153 983	132 907	29 907	280 440	211 316	801 809	297 422	301 002	203 385	478 596	323 213	14,29
5	2012	150 449	145 550	4 899	572 342	185 314	187 948	163 861	35 219	320 268	252 074	815 142	344 237	294 511	176 394	515 580	299 562	15,50
6	2013	168 888	167 452	1 436	669 974	205 565	188 831	249 362	26 216	421 754	248 220	910 782	426 082	310 739	173 961	602 776	308 006	14,41
7	2014	210 201	263 167	-52 966	675 093	250 153	210 442	186 535	27 963	365 454	309 638	1 020 667	413 045	390 343	217 280	547 987	472 680	15,03
8	2015	199 193	265 793	-66 600	716 728	305 383	271 454	120 055	19 836	391 911	324 817	981 627	437 977	329 593	214 057	433 826	547 802	17,54
9	2016	190 691	350 078	-159 387	793 475	364 547	271 849	148 497	8 581	426 418	367 056	998 682	350 475	389 798	258 409	504 999	493 683	15,90
10	2017	178 054	204 545	-26 491	898 844	419 126	293 337	151 485	34 896	490 971	407 872	1 016 657	405 718	342 147	268 792	570 627	446 031	14,58
11	2018	204 554	182 215	22 339	932 967	456 467	322 065	125 571	28 863	540 683	392 284	1 073 131	491 942	336 898	244 290	613 994	459 137	17,18

Fig. 21. **Initial data**

Step 2. Create an empty database in MS Access (Fig. 22).

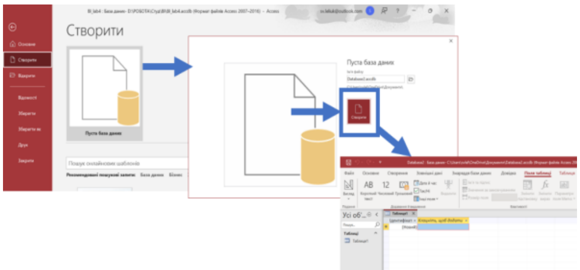


Fig. 22. **Database creating in MS Access**

Step 3. Import data in MS Access from MS Excel (Fig. 23 – 26).

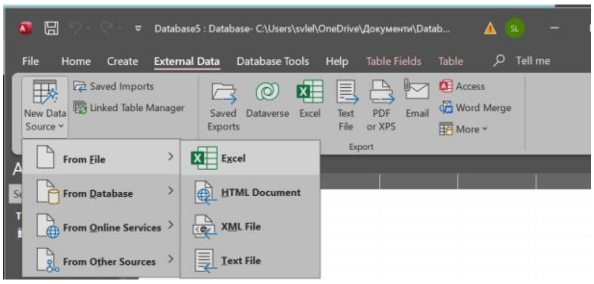


Fig. 23. **Connecting to the external data**

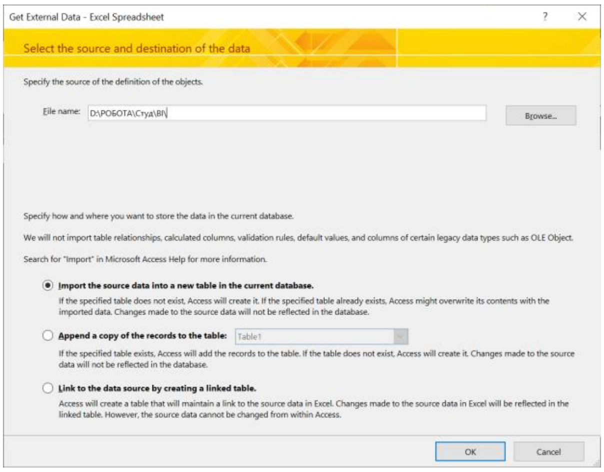


Fig. 24. **The import results saving options**

Select a sheet with the database in the MS Excel (Fig. 25).

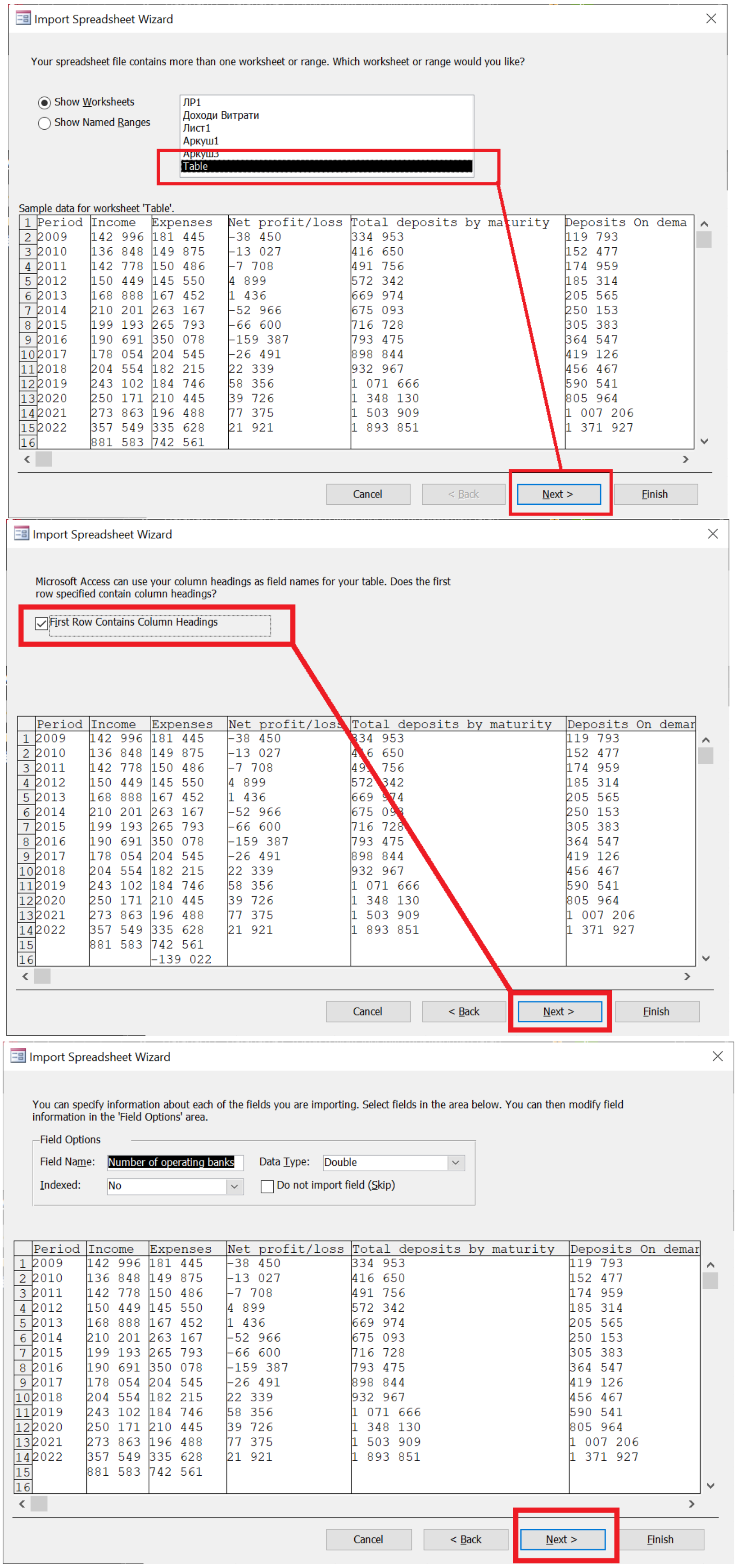


Fig. 25. **Using the Spreadsheet Import Wizard**

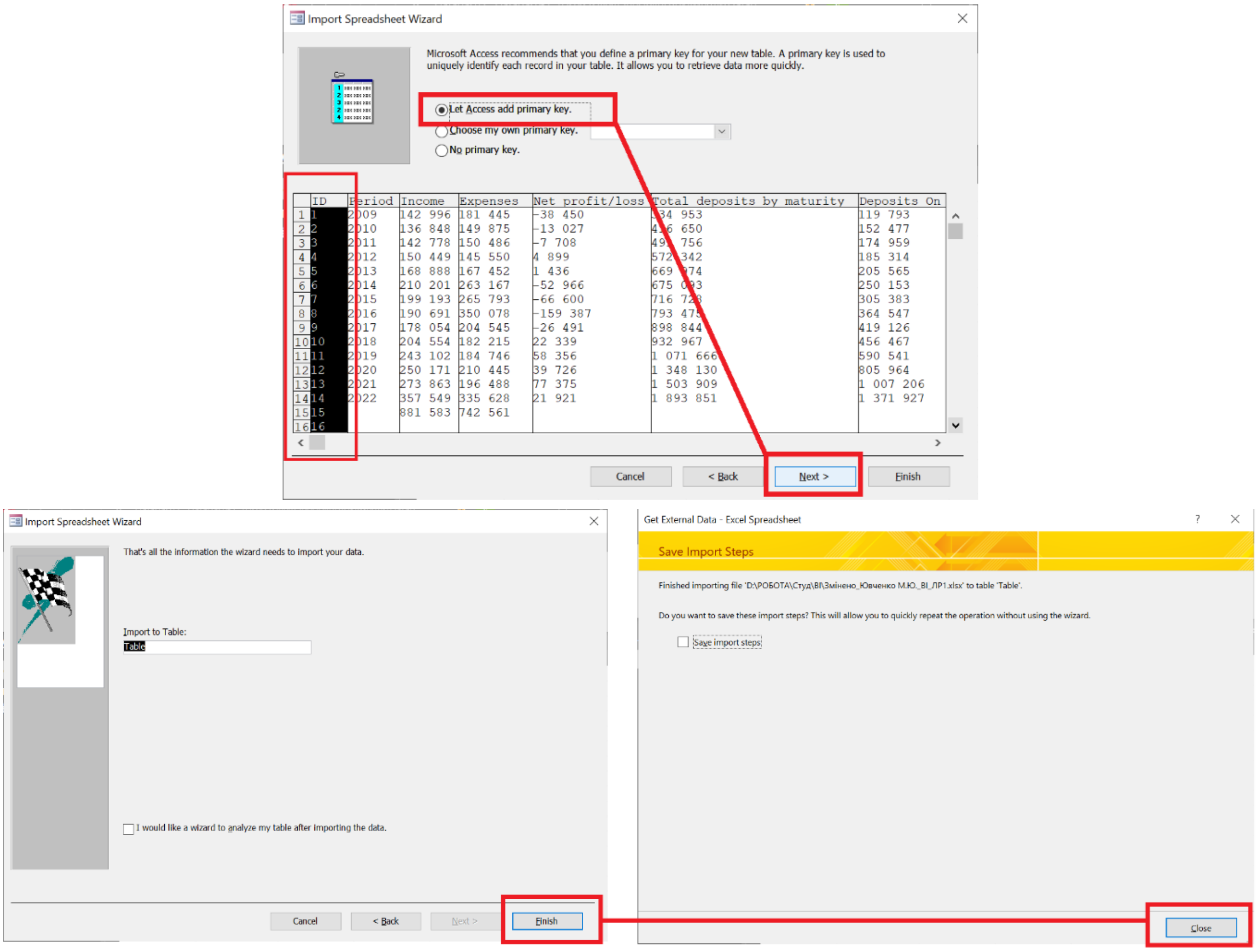


Fig. 26. **Setting up a key field and saving results**

Open the created Table in MS Access (Fig. 27).

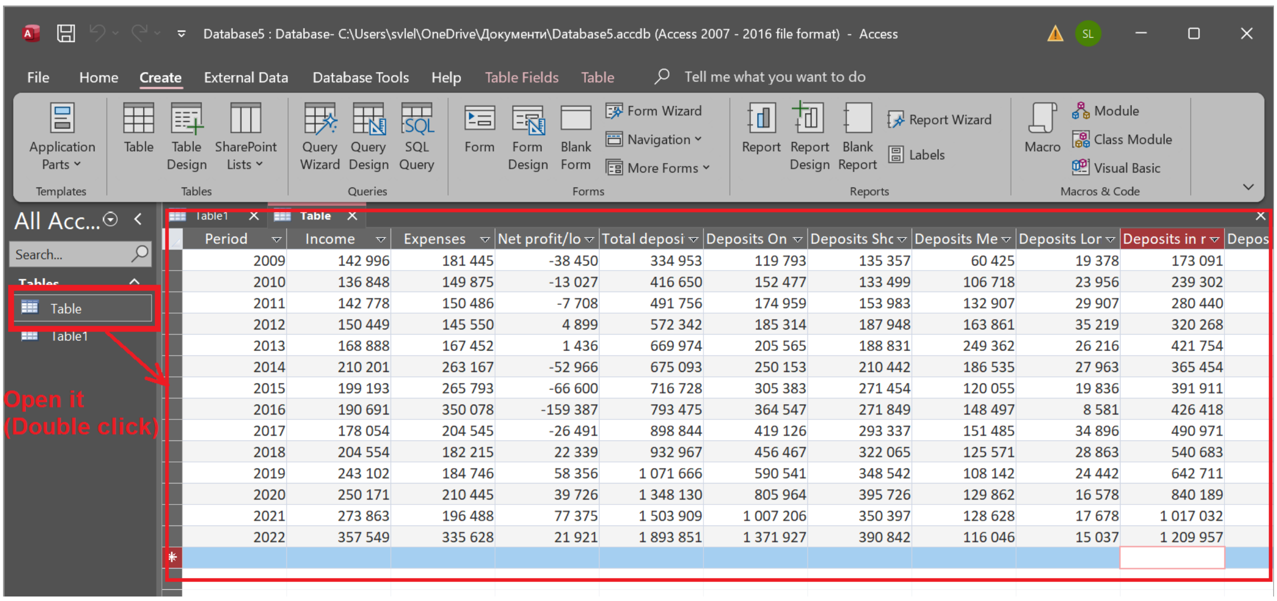


Fig. 27. **Imported financial data table**

Step 4. Create a simple query using the Query Wizard (Fig. 28).

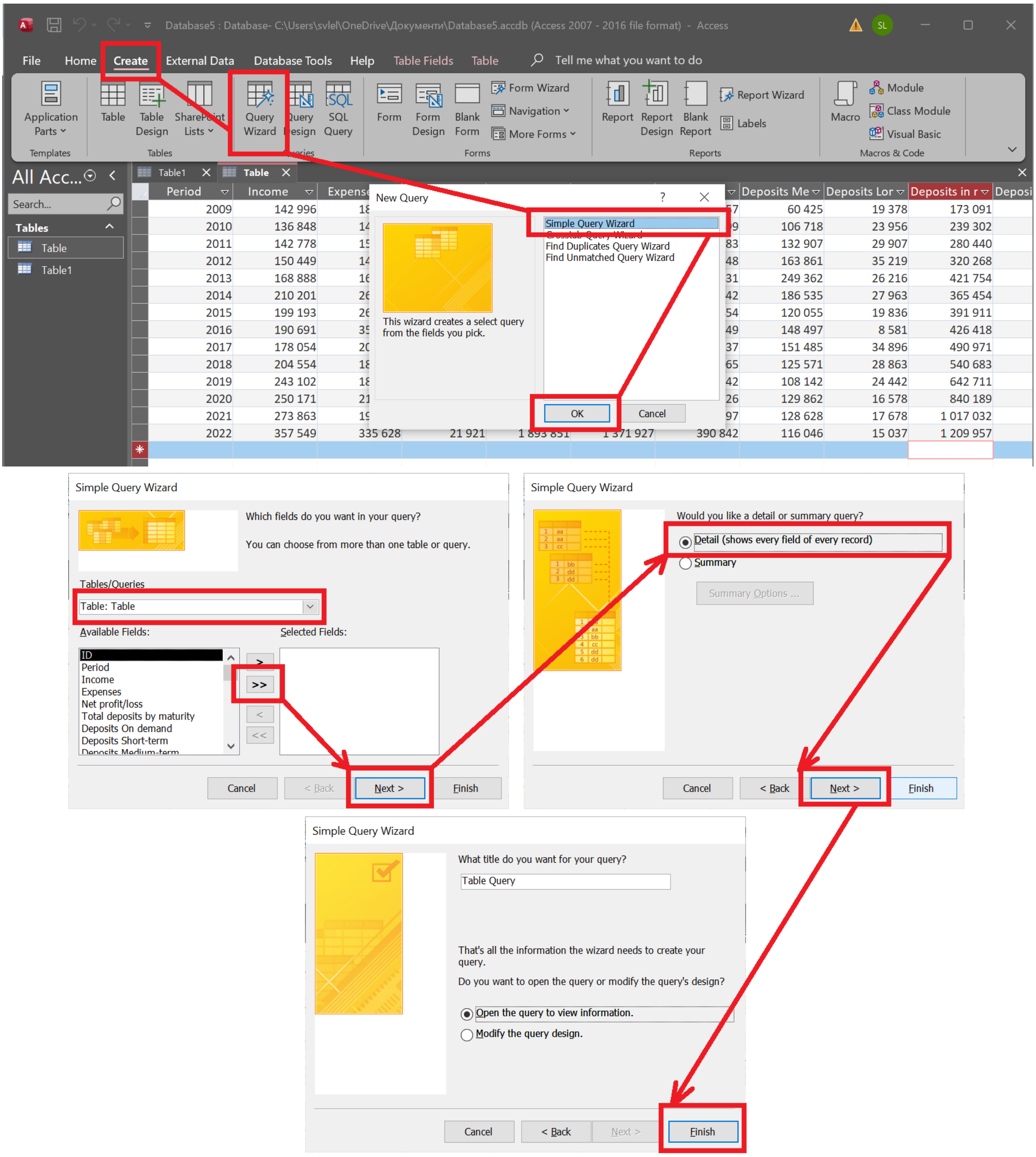


Fig. 28. **Query creation using the Wizard**

Step 5. Save and copy created Query (Fig. 29).

Step 6. Open the Query in the SQL mode (Fig. 30).

Change the created Query in the SQL mode and run it (Fig. 31). The result of query implementation is shown on the Fig. 32.

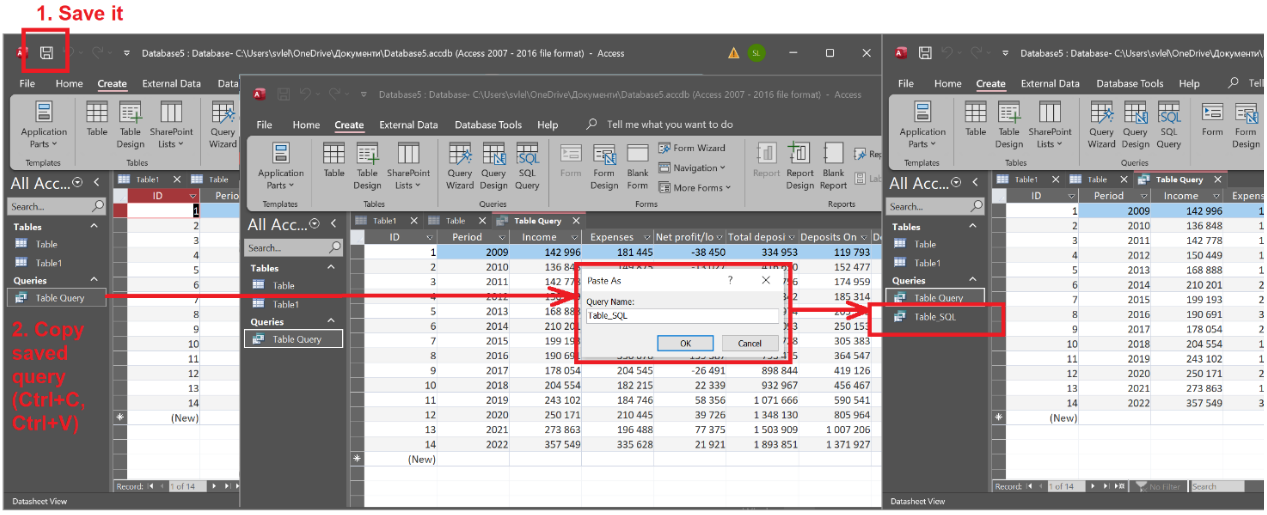


Fig. 29. **Saving the created query**

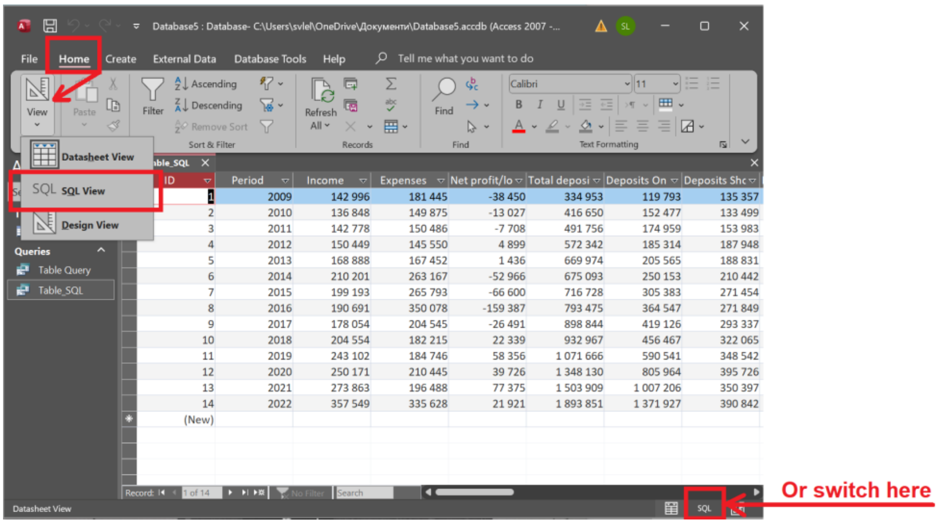


Fig. 30. **Switching to SQL query writing mode**

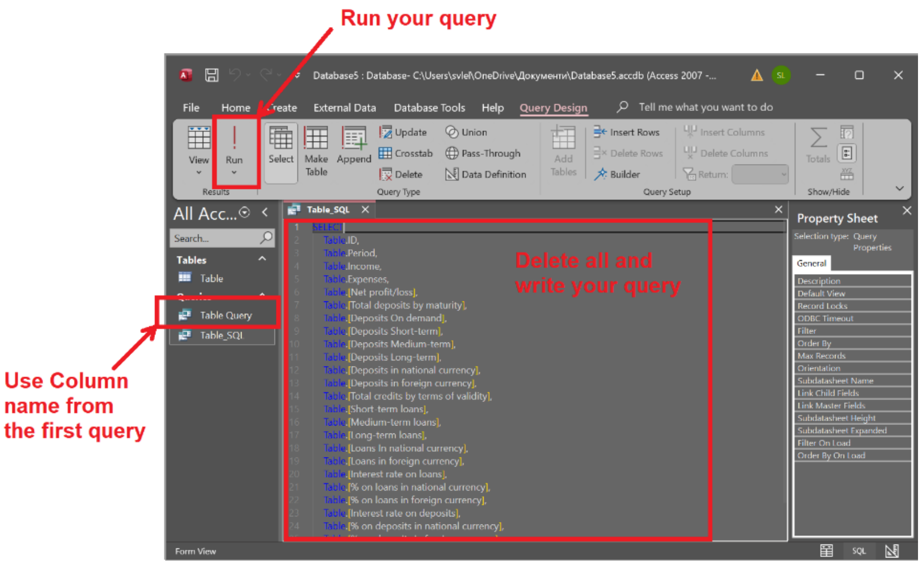


Fig. 31. **Initiating a request**

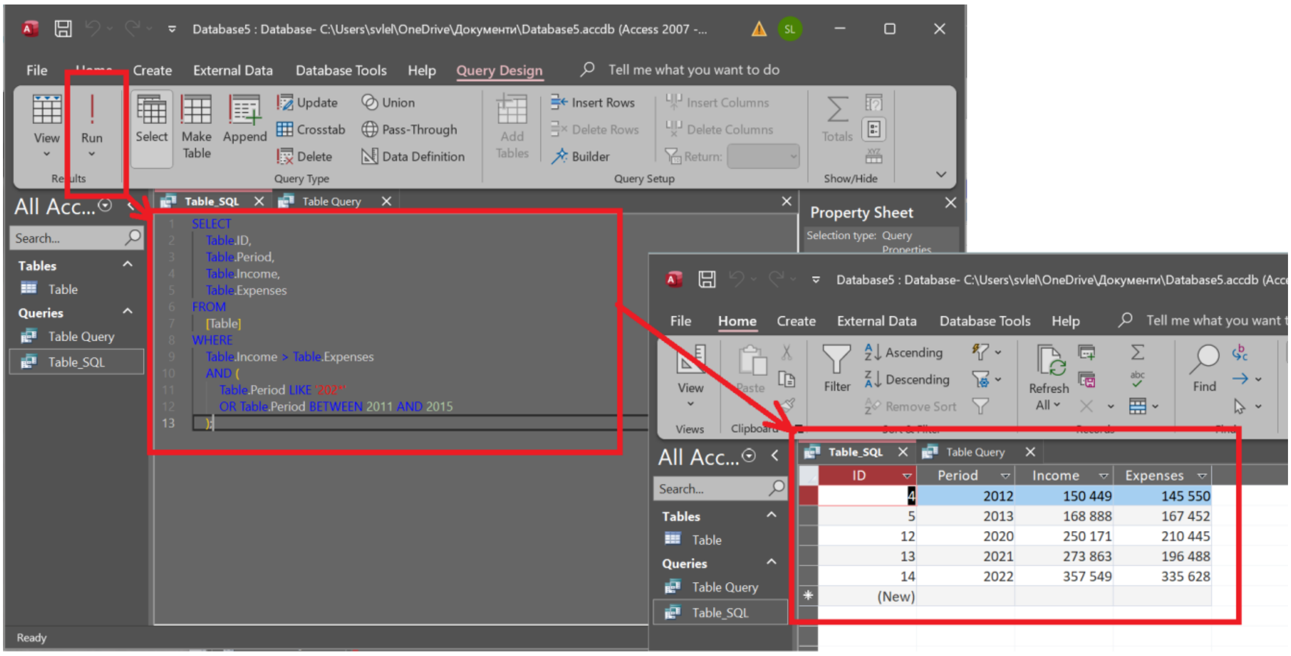


Fig. 32. **Sampling based on query results**

Requirements for a report on laboratory work

The report on the laboratory work should contain:

1) a title page;

2) the purpose of the laboratory work;

3) a file with data in MS Excel and MS Access;

4) a file in MS Word containing answers to the questions raised and conclusions reached in the progress of the laboratory work.

Checklist questions

1. What functions allow you to clean up the minimum/maximum value in the data sample under study?

2. How to write a SQL query to search for values in table cells based on a specific sequence of characters?

3. What are the basic elements used to write a SQL query?

Tasks for self-study

Task 1. Create the database in MS Access using financial data import from MS Excel.

Task 2. Make your query in MS Access database using:

functions (AVG(), MIN() and MAX(), COUNT(), SUM());

keywords (AND, OR, LIKE, BETWEEN, IN, ORDER BY (ASC/DESC)).

Laboratory work 5
Data preparation and modeling in the MS Power BI environment

The purpose of the laboratory work is to gain skills in building a data model and a financial performance monitoring dashboard in MS Power BI.

General guidelines

MS Power BI is designed to create reports that can be used to analyze data. The data in such reports is presented in a clear form. Report generation is carried out in several stages:

1) data acquisition;

2) data processing;

3) modeling (establishing relationships, creating calculated columns and dimensions);

4) creating visualizations (reports).

There are three main areas of work in MS Power BI: query editor; model editor; report editor. Each of the sections is focused on the implementation of the corresponding tasks.

The Power Query query editor allows you to connect to data, combine data from several tables and process them for further use in modeling and visualization.

Work in progress

Task. Prepare a financial data set based on the reporting of 5 enterprises for the last 5 years in MS Excel. Load information into MS Power BI. Edit the data model. Create a financial indicator monitoring dashboard.

Step 1. Prepare financial data in MS Excel using the "Enterprise", the "Indicators" and the "Data" tables (Fig. 33 – 35).

The structure of the input data is as follows:

the "Enterprise" table consists of the "Enterprises code" and "Enterprise" fields;

the "Indicators" table includes the "Code" and the "Indicator" fields;

the "Data" table consists of the "EnterpriseCODE" field, the "IndCODE" field, the "Period" field and the "Value" field.

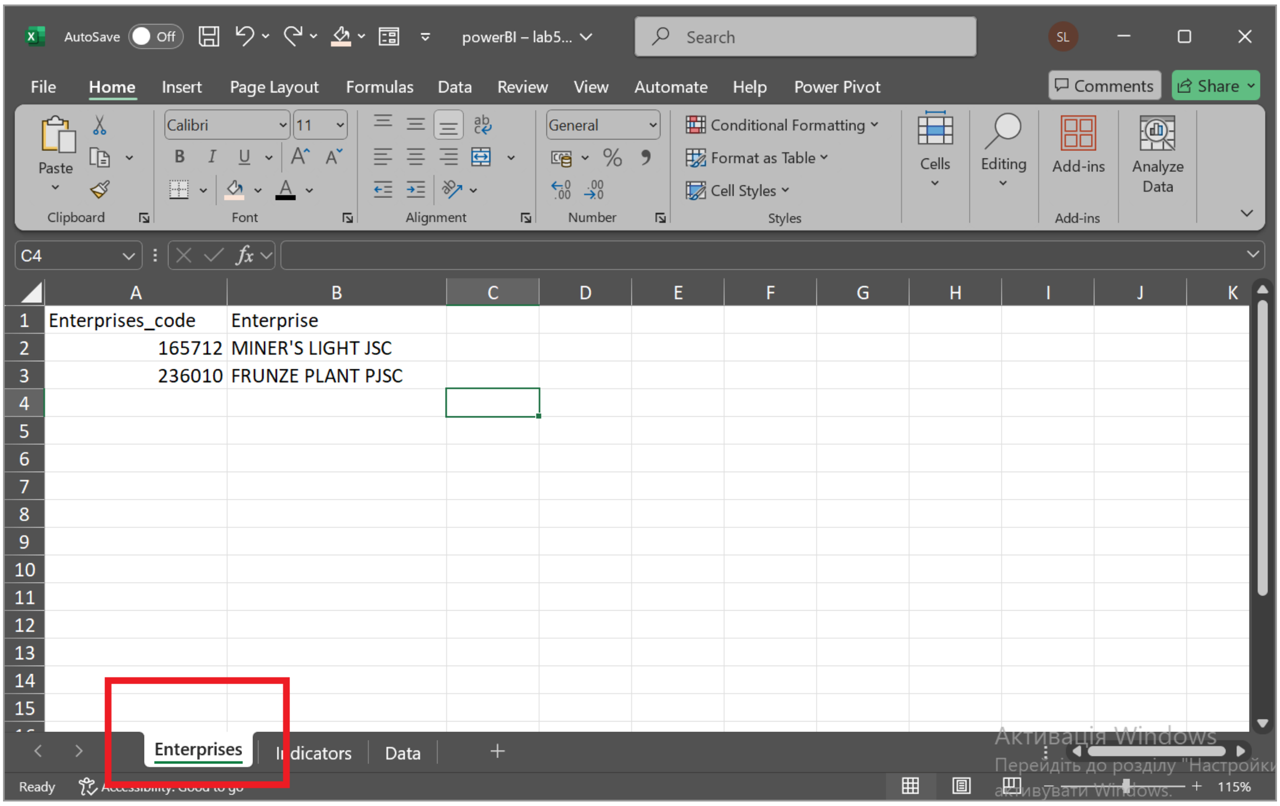


Fig. 33. **The "Enterprise" table structure**

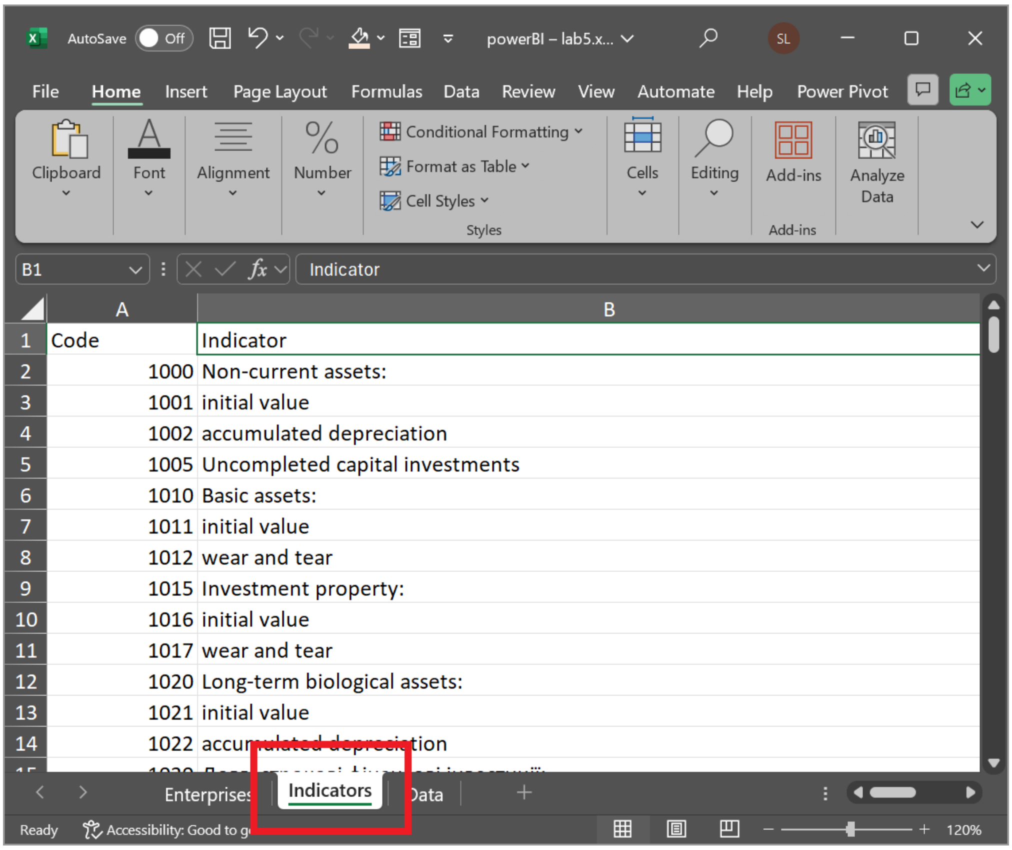
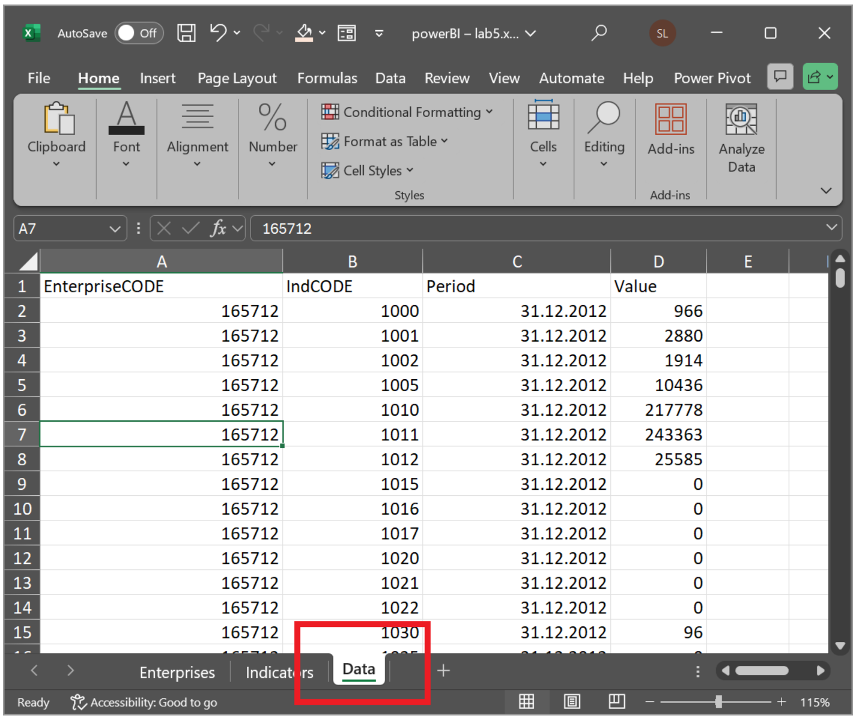


Fig. 34. **The "Indicators" table structure**



	A	B	C	D
1	EnterpriseCODE	IndCODE	Period	Value
2	165712	1000	31.12.2012	966
3	165712	1001	31.12.2012	2880
4	165712	1002	31.12.2012	1914
5	165712	1005	31.12.2012	10436
6	165712	1010	31.12.2012	217778
7	165712	1011	31.12.2012	243363
8	165712	1012	31.12.2012	25585
9	165712	1015	31.12.2012	0
10	165712	1016	31.12.2012	0
11	165712	1017	31.12.2012	0
12	165712	1020	31.12.2012	0
13	165712	1021	31.12.2012	0
14	165712	1022	31.12.2012	0
15	165712	1030	31.12.2012	96

Fig. 35. **The "Data" table structure**

Step 2. Import financial data from MS Excel to MS Power BI (Fig. 36 – 37).

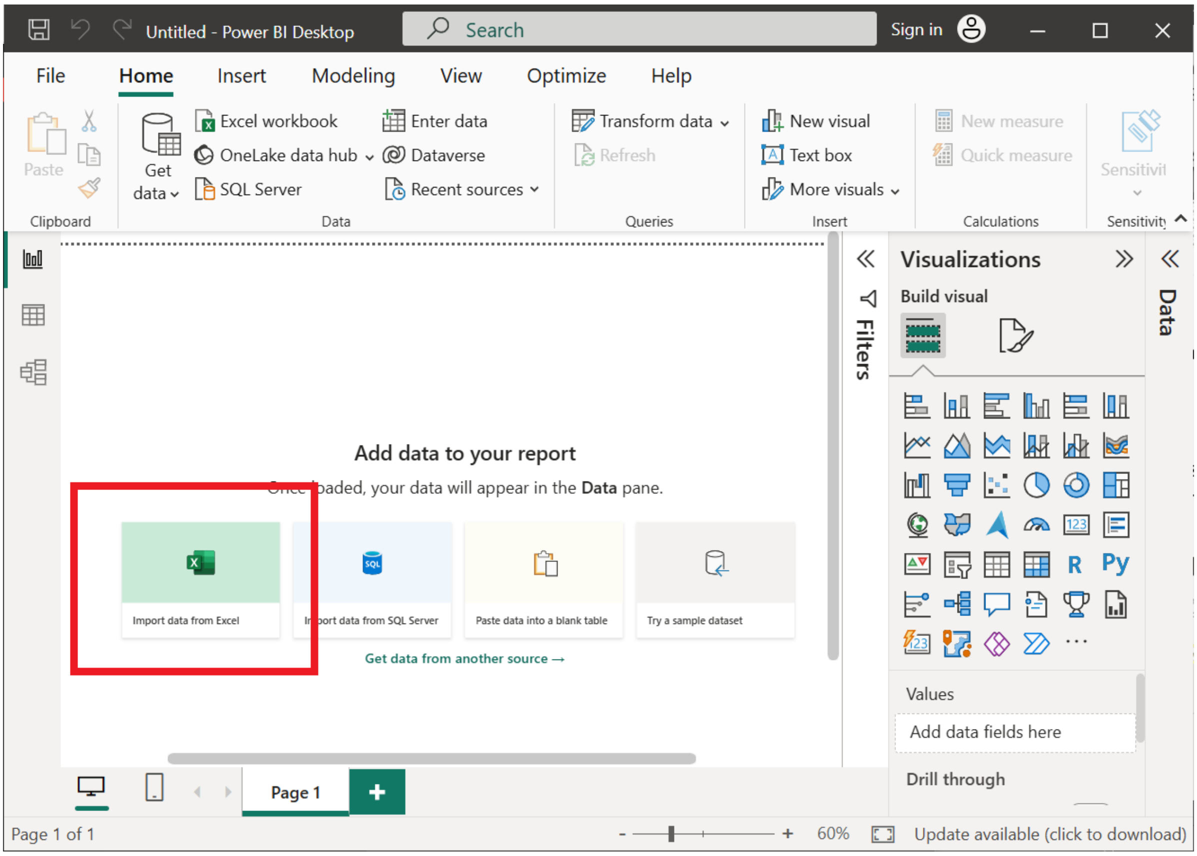


Fig. 36. **Connecting to the data source in MS Power BI**

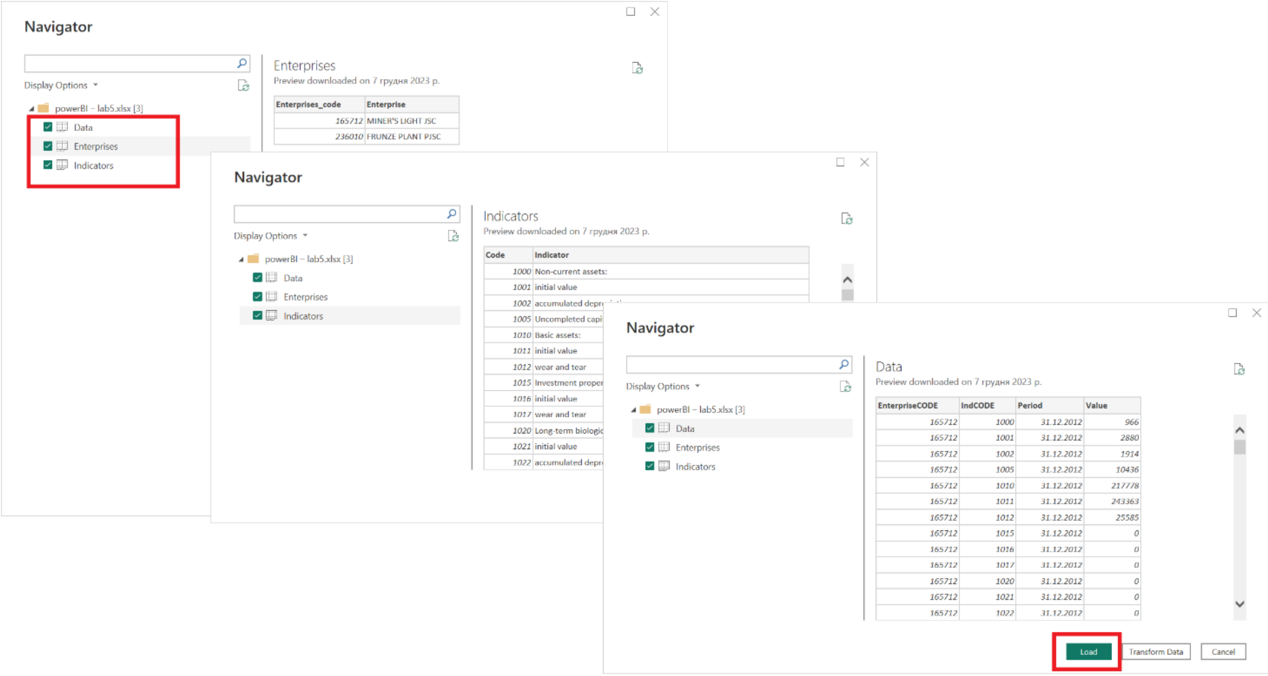


Fig. 37. **Checking the imported data**

Step 3. Create the data model (set the relationships between tables) MS Power BI (Fig. 38).

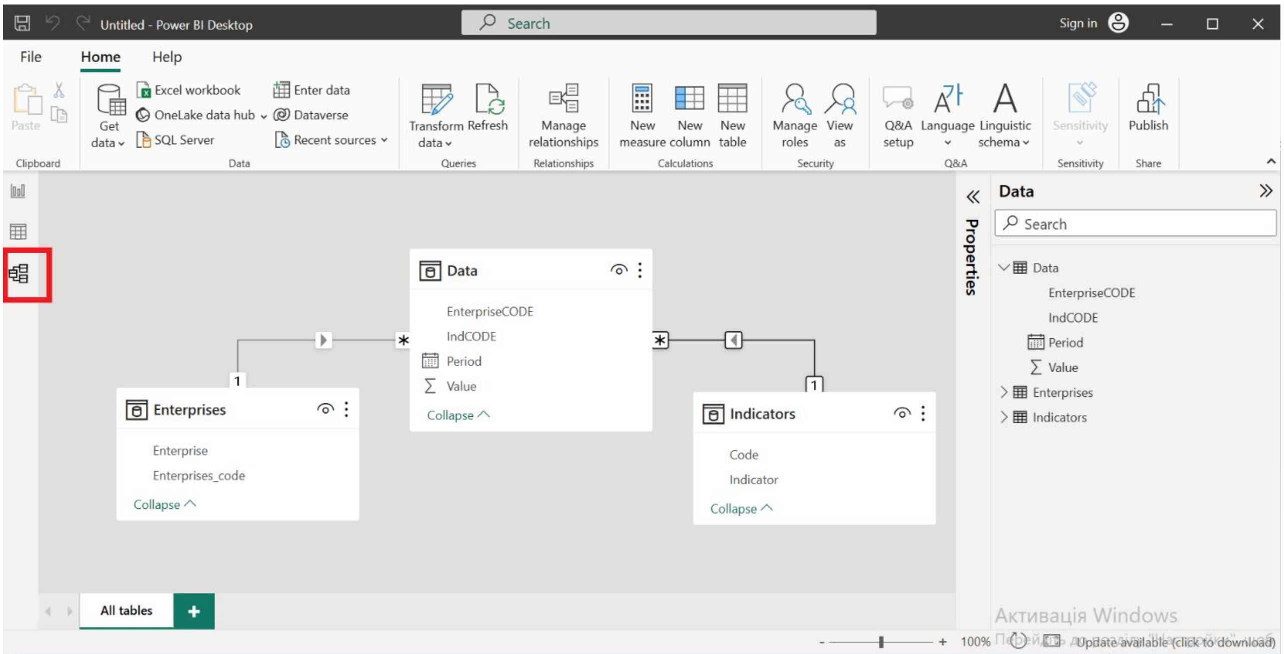


Fig. 38. **Checking (creating) the data model**

Step 4. Create the visualizations of imported financial data (Fig. 39). Edit and customize the rendering structure (Fig. 40).

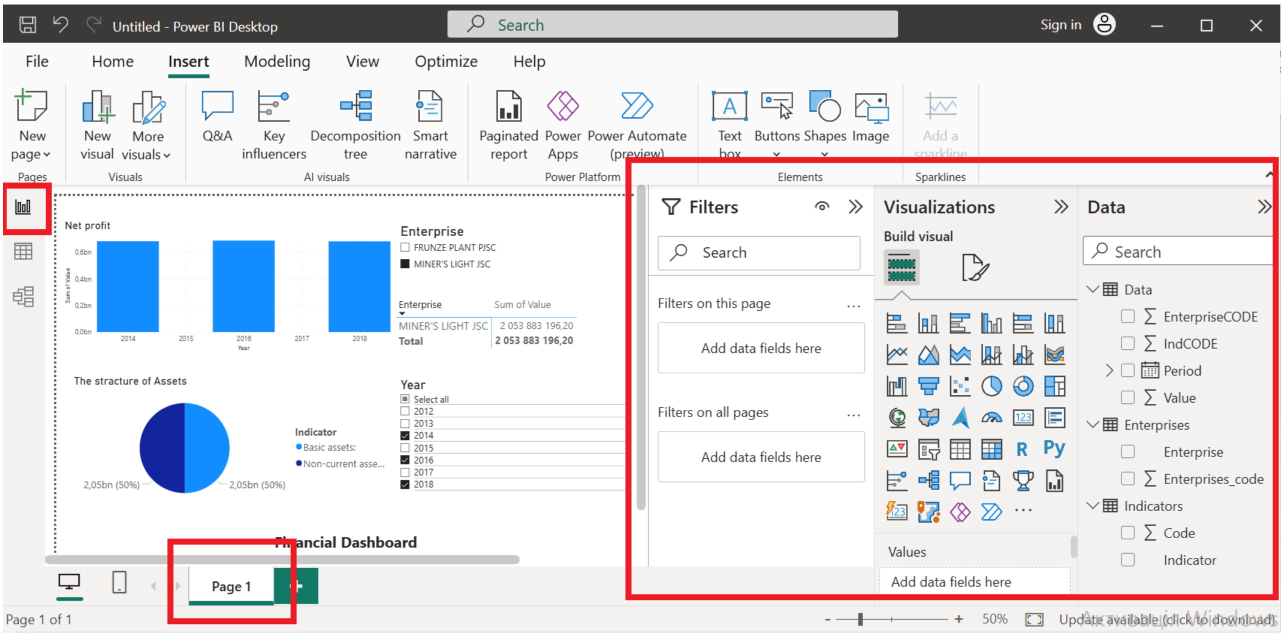


Fig. 39. **Creating the financial data dashboard**

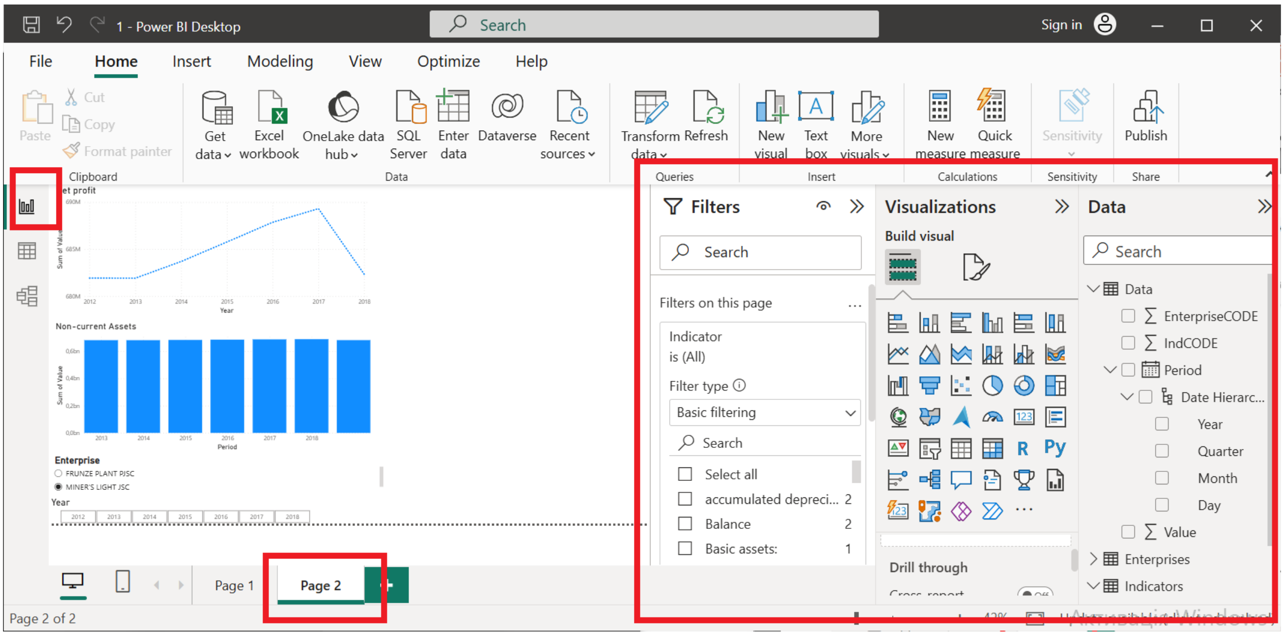


Fig. 40. **Editing the financial data dashboard**

Requirements for a report on laboratory work

The report on the laboratory work should contain:

1) a title page;

2) the purpose of the laboratory work;

3) a file with data in MS Excel and MS Power BI;

4) a file in MS Word containing answers to the questions raised and conclusions reached in the progress of the laboratory work.

Checklist questions

1. Describe the procedure for creating a financial data dashboard (report creation) in the PowerBI.

2. Describe the main modes of work in MS Power BI.

3. What is the main purpose of the Power Query editor in MS Power BI?

Tasks for self-study

Task 1. Create a financial data source to build a financial data monitoring panel on the topic of the diploma project.

Task 2. Import the created financial data into the PowerBI environment. Form a financial data structure model according to the example given in Fig. 27.

Task 3. Build a financial data monitoring panel on the topic of the diploma project (report) in the MS PowerBI environment.

Recommended reading

Main

1. Проноза П. В., Лелюк С. В. Системи підтримки прийняття антикризових фінансових рішень : навчальний посібник. Харків : ХНЕУ ім. С. Кузнеця, 2019. 117 с. URL: http://repository.hneu.edu.ua/bitstream/123456789/21442/1/2019-5-%D0%95%D0%9D%D0%9F%20%D0%9F%D1%80%D0%BE%D0%BD%D0%BE%D0%B7%D0%B0%20%D0%9F%20%D0%92%2C%20%D0%9B%D0%B5%D0%BB%D1%8E%D0%BA%20%D0%A1%20%D0%92.pdf.

2. Loshin D. Business Intelligence. The Savvy Manager's Guide Getting Onboard with Emerging IT. S. l. : Morgan Kaufmann Publishers, 2003. 292 p.

Additional

3. Алексєєнко І. І., Лелюк С. В., Полтініна О. П. Управління проєктами та вартістю підприємства в умовах цифровізації: аналіз бізнес-процесів і візуалізація фінансових даних. *Цифрова економіка як фактор економічного зростання держави* : колект. монографія. Херсон : Гельветика, 2021. С. 80–98. URL: http://repository.hneu.edu.ua/handle/123456789/27824.

4. Аксак Н. Г., Гризун Л. Е., Щербаков О. В. [та ін.]. Сучасні інформаційні технології та системи: монографія / за заг. ред. Пономаренка В. С. Харків : ХНЕУ ім. С. Кузнеця, 2022. 270 с. URL: http://repository.hneu.edu.ua/handle/123456789/29233.

5. Лелюк С., Алексєєнко І., Полтініна О. Візуалізація даних в управлінні проектами фінансової сфери. *Економіка та суспільство*. 2021. Випуск № 26. URL: http://repository.hneu.edu.ua/bitstream/123456789/26119/1/2.pdf.

6. Abduldaem A., Gravell A. Success Factors of Business Intelligence and Performance Dashboards to Improve Performance in Higher Education. *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS 2021)*. Vol. 2, P. 392–402. URL: https://www.scitepress.org/PublishedPapers/2021/104995/104995.pdf.

7. Advanced Analytics with Power BI. URL: https://www.marquam.com/Documents/Advanced%20Analytics%20with%20Power%20BI%20White%20Paper.pdf.

8. Agiu D., Mateescu V., Muntean Iu. Business Intelligence overview. *Database Systems Journal.* 2014. Vol. V, No. 3/2014. P. 23–36.

9. Aleksieienko I., Leliuk S. Modeling business processes of making financial decisions using digital technologies. *Transformational economy: theoretical and practical aspects* : *Collective monograph.* Riga, Latvia : Baltija Publishing, 2024. P. 383–402. URL: http://www.repository.hneu.edu.ua/handle/123456789/35238.

10. Bataweel D. S. Business Intelligence: Evolution and future trends. Theses. 2015. 283 p. URL: https://core.ac.uk/download/pdf/327255786.pdf.

11. Berest M., Sablina N. Evaluation of the effectiveness of strategic and tactical controlling based on the analysis of the company's financial reports. *Development Management*. 2024. Vol. 23 (1). P. 8–18. URL: http://repository.hneu.edu.ua/handle/123456789/35284.

12. Creating Visualizations using Microsoft Power BI. URL: https://home.ictp.it/Members/oayub/Lecture-2.pdf.

13. Gad-Elrab A. A. Modern Business Intelligence: Big Data Analytics and Artificial Intelligence for Creating the Data-Driven Value. *E-Business – Higher Education and Intelligence Applications*. 2021. URL: http://dx.doi.org/10.5772/intechopen.97374.

14. Roger D. Peng and Elizabeth Matsui. The Art of Data Science: A Guide for Anyone Who Works with Data. 2017. URL: https://bookdown.org/rdpeng/artofdatascience/.

Information resources

15. Educational materials on the academic discipline "Business Intelligence in Finance" on the website of personal learning systems of S. Kuznets KhNUE. URL: https://pns.hneu.edu.ua/enrol/index.php?id=10252.

16. Get started with Power BI Desktop. URL: https://learn.microsoft.com/en-gb/power-bi/fundamentals/desktop-getting-started.

17. Ministry of Finance of Ukraine. URL: https://www.mof.gov.ua/uk/.

18. National Bank of Ukraine. URL: https://bank.gov.ua.

19. State Statistics Service of Ukraine. URL: http://www.ukrstat.gov.ua.

Contents

НАВЧАЛЬНЕ ВИДАННЯ

BUSINESS INTELLIGENCE У ФІНАНСАХ

Методичні рекомендації до виконання лабораторних і самостійних робіт здобувачів вищої освіти спеціальності D2 "Фінанси, банківська справа, страхування та фондовий ринок" освітньої програми "Фінанси і кредит" другого (магістерського) рівня

(англ. мовою)

Самостійне електронне текстове мережеве видання

Укладачі: **Берест** Марина Миколаївна
Лелюк Світлана Валеріївна

Відповідальний за видання *І. В. Журавльова*

Редактор *З. В. Зобова*

Коректор *З. В. Зобова*

Запропоновано завдання для лабораторних занять та самостійної роботи здобувачів вищої освіти з навчальної дисципліни, методичні рекомендації та порядок виконання завдань.

Рекомендовано для здобувачів вищої освіти магістратури (другого) ступеня спеціальності D2 "Фінанси, банківська справа, страхування та фондовий ринок" освітньої програми "Фінанси і кредит".

План 2025 р. Поз. № 39 ЕВ. Обсяг 47 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61165, м. Харків, просп. Науки, 9-А

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.