# Visualization of dynamic reporting data of an information and analytical system for making managerial decisions

Anastasia Kiryushina HSE university Department of Computer Engineering Moscow, Russia akiryushina@hse.ru

Anton Drozhzhin HSE university Department of Electronic Engineering Moscow, Russia aadrozhzhin@edu.hse.ru Olga Ivanenko HSE university Department of Applied Mathematics Moscow, Russia ovivanenko@edu.hse.ru Diana Shayakhmetova HSE university Department of Applied Mathematics Moscow, Russia dtshayakhmetova@edu.hse.ru

Anastasia Koroleva HSE university Department of Computer Engineering Moscow, Russia aakoroleva\_4@edu.hse.ru

Abstract—Nowadays, in the age of digital transformation in the field of analytical systems for managerial decision-making, there are a number of main problems associated with changing business processes of the organization. The solution of modern problems in this field is based on the analysis of the subject area with the conjugation with the hierarchical structure of the organization's management and the way of its visual representation.

This article proposes an approach for information and analytical system developers, including recommendations on ways and forms of visual representation of indicators in the formation of analytical reports, as well as an analysis of errors that occur as you work with multidimensional data and propose ways to solve them, which makes working with data more efficient and scientifically sound from the standpoint of providing effective support for management decisions.

#### I. INTRODUCTION

The purpose of the work is to develop an approach to visualization in the information and analytical system (hereinafter IAS) from the standpoint of providing effective support for management decisions (hereinafter the Approach).

When developing the approach there were used the following management models:

- Anthony Triangle Model;
- General control scheme of the control object.



Fig. 1. Anthony 's Triangle model

According to the hierarchical view of the management structure (Anthony's triangle) proposed by Robert Anthony, managerial decisions are made at the strategic, management control and operational levels (A framework for management information systems, HBR's 10 Must Reads on Strategy)

In order to visualize on dashboards the indicators corresponding to each of the decision-making levels, firstly, it is necessary to determine what to visualize:

- List of strategic management level indicators;
- List of indicators of the management control level of management;
- List of indicators of the operational level of management.

This point is developed in paragraph (IV) of the work.

Secondly, the quality of data mining should meet such criteria for the success of managerial decision-making as speed, awareness and intelligence. This is perhaps the most difficult and important aspect in this work and it is given special attention in paragraphs (V) and (VI) on the stages of data mining.

Finally, it is necessary to ensure the interpretability of data mining. This issue is discussed in paragraph (VII) of the work.

### II. SUBJECT AREA OF RESEARCH

This study was conducted on the example of data from one of a Corporation in Moscow with state participation in the authorized capital of the organization (hereinafter the Corporation), on the territory of which there is an exposition, museum and recreational complex, where museum and exhibition projects, international business exhibitions and congresses, festivals and holidays are held.

The problem domain of research of the subject of management of one Corporation in Moscow was investigated on the

Flements of the Designations Values of controls used in the development of a visualization approach using the example					
control model	Designations	of a Concoration			
control model		Strategic management level	Management control level	Operational management level	
Subjects of manage-	$U = \phi(W, X, Z, t)$	CEO of the Corporation	Heads of departments	Heads of departments and group	
ment		1		heads of the relevant structural	
				divisions of the Corporation	
Management	X = f(X, U, Z, t)	The Corporation is an exhibition,	Elements of a Corporation, on	Territory, real estate, residents,	
objects		museum and recreational com-	the territory of which real estate	events, staff, equipment	
		plex, a unique space for mu-	objects (including cultural her-		
		seum and exhibition projects, in-	itage objects) are located, leased		
		ternational business exhibitions	to residents doing business on		
		and congresses, festivals and nol-	the territory of the Corporation		
Variables that reflect	V	Indicators characterizing the	Indicators characterizing the	1) Sorviges provided to residents:	
the state of an ob	A-	state of the exposition museum	state of the territory including	2) Processes related to the hold	
iect at a time		and recreational complex	the state of real estate objects	ing of events on the territory of	
Jeet at a time		and recreational complex	their attractiveness to residents	the Corporation:	
			and visitors	3) Activity of residents on the	
				territory of the Corporation;	
				4) Processes related to ensuring	
				the quality of real estate objects	
				to the specified requirements	
Control actions de-	U-	Moscow development programs	Orders, instructions and memos	Actions on the part of the per-	
signed to optimize		and other regulatory acts to	at the level of heads of depart-	sonnel of the relevant structural	
the performance in-		which the Corporation is subject.	ments	units	
dicator		Orders and instructions at the			
External influences	7	level of the CEO	Natural and other disasters day	Staff me ductivity activity of rea	
External influences	2-	seasonal nature, calendar of	watural and other disasters, doc-	idents, etc.	
		tion factors affecting the state of	ket conditions etc	idents, etc.	
		tourism in Moscow	ket conditions, etc.		
Management	W	Attendance of the Corporation's	Performance indicators of heads	1) Performance indicators of	
performance		territory	of departments of the Corpo-	heads of departments and heads	
indicator			ration: the number of residents	of groups of structural divisions	
			doing business on the territory;	of the Corporation;	
			balance of financial results; the	2)The number of successfully	
			number of restored objects of	implemented projects, the num-	
			cultural heritage of Moscow on	ber of agreements concluded	
			the territory of the corporation,	with residents, the number of	
			etc.	events held, etc.	

 TABLE I

 Subject area of research of the subject of management of the Corporation of Moscow

basis of the methodological foundations of management. The results of analysis are presented in the table (I).

Thus, within the framework of the development of the approach, the main elements of the management model are identified in relation to each other, which underlie the architecture created by the IAS to ensure effective support of management decisions on the example of a Corporation for all levels of management.

# III. DEVELOPMENT OF AN APPROACH TO THE STAGES OF DATA MINING

The main stages of data mining are (Data and process analysis):

- 1) Understanding and formulation of the analysis task;
- Preparation of data for automated analysis (preprocessing);
- Application of Data Mining methods and construction of models;
- 4) Checking the constructed models;
- 5) Human interpretation of models.



Fig. 3. Stages of data mining



Fig. 2. General scheme of the problem statement

# IV. PROBLEM STATEMENT

To develop the approach, tasks from the management activities of the Corporation were used according to which it is necessary to visualize indicators characterizing the state of management facilities at the strategic, management control and operational levels.

Internal corporate data sources are used to obtain indicators. When building models for the formation of indicators, the main standards of the company were used, the documents in which the Corporation's strategy is reflected.

General scheme of the current problem is represented in the figure (2).

# V. DESCRIPTION OF THE PROCESS OF WORKING WITH DATA, IDENTIFICATION OF PROBLEMS AT THE STAGE OF DATA PREPROCESSING AND DETERMINATION OF WAYS TO SOLVE THEM

Nowadays, there is a need to deal with huge amounts of data located in various sources and with variety of representations, formats and encodings, as well as violations of the structure, completeness and integrity, data preprocessing is required. In the process of data preprocessing, there is a preporation for analysis, as a result of which they are brought into compliance with the requirements determined by the specifics of the task being solved. Without preprocessing during the analysis in the future either the algorithms will not be able to work, or the result of their work will be incorrect. There are two directions of data preprocessing:

 Cleaning — processing duplicates, contradictions and fictitious values, restoring and filling in gaps. Also smoothing, noise suppression and editing abnormal values, as well as restoring the structure, completeness and integrity of data, converting incorrect formats;  Optimization — reducing dimensionality, identifying and eliminating insignificant features, i.e. adapting data to a specific task and improving the analysis efficiency.

Based on the results of the analysis of data from several internal sources (corporate information systems and files) of the Corporation, the following key data quality problems were identified:

- Naming conflicts;
- Missing values;
- Difference in the representation of values;
- Operator input errors.

One of the problem in the data of the source Nº1 personal account of the resident is the differences in the names of columns in various departments, and there is also a space at the end of the line.

For both sources, the problem is the discrepancy between the name of the resident in the personal account and in the source  $N^{\circ}2$  "Register":

1) Different name;

ЛКР	Реестр
АО "Союзмультфильм"	АО "Киностудия Союзмультфильм"
АО "БРПИ" (417, 575)	АО "БРПИ
ИП Ахембе А.В.	ИП Ахембе Анастасия Владимировна
ООО «Ласточка» (Осипова С.С.)	ООО "Ласточка"
ООО "Макдональдс"	ООО "Макдоналдс"
Управление ЗАГС	Управление ЗАГС Москвы
И др.	

2) Uppercase/lowercase letters;

ЛКР	Реестр
ООО "МОСКОВСКОЕ НЕБО"	ООО "Московское небо"
ООО "Барком"	ООО "БАРКОМ"
ООО "Журнал мод"	ООО "ЖУРНАЛ МОД"
Фонд гуманитарных проектов	Фонд Гуманитарных Проектов

3) Abbreviations;

ЛКР	Реестр	
ООО "НПП "ЧИП"	000	"Научно-производственное
	предприятие "Ч	ИП"

#### 4) Extra spaces;

1		
JKP	Реестр	
B	Значале¶	
АО "Риччи и Каприччи"		
B	середине¶	
ООО·"ЛУКОЙЛ-Центрнефтепродукт" ООО·"ЛУКОЙЛ-Центрнефтепродукт"		
1	В конце¶	
ИП Романчук Р.В. 9		

### 5) Special symbols;

ЛКР	Реестр
Православный приход Тихвинского храма в	Православный приход Тихвинского храма
Алексеевском	в Алексеевском

#### 6) Quotation marks;

-	
ЛКР	Реестр
Использование разных кавычек	
АО «Электрификация»	АО "Электрификация"
ООО «Международная компания связи»	ООО "Международная компания связи"
ООО «Возрождение ВВЦ»	ООО "Возрождение ВВЦ"
Незакрытые кавычки или их отсутс	гвие
ПАО ВымпелКом	ПАО "ВымпелКом"

7) The absence of a resident in the personal account or in the Register;

ЛКР			Реестр	
OCOO	«Олимпийский	комитет	-	
России»				

- 8) Extra characters in the direction name or "Null";
  - ["ВДНХ Гастрономическая"]
  - ["ВДНХ Сервисная"]
  - ["ВДНХ Музейная"]
  - [null]
- 9) Not all residents have directions specified.

# VI. COMPARATIVE ANALYSIS OF THE USE OF DATA TOOLS AND DATA MINING METHODS IN THE CONSTRUCTION OF MODELS AND VERIFICATION OF THEIR RESULTS

According to the statement of the task, it is necessary to visualize indicators characterizing the state of management facilities "residents" and "capital construction facilities".

At the strategic level:

- The main characteristics of the "capital construction facilities" total area, leasable, leased area;
- Rating of residents by a set of indicators.

At the management control level:

- The ratio of the number of residents who signed the contract to all residents with reference to the responsible departments of the Corporation;
- Residents' activity in the Corporation's digital platform.

In the process of developing the approach, a comparative analysis of the use of the following tools for working with data was carried out:

- Microsoft Excel;
- Microsoft Power BI;
- Python libs.

The results of the comparative analysis are presented in the figures (4) and (5) and in the table (II).



Fig. 4. An example of using Microsoft Power BI to visualize indicators at the strategic management level: The main characteristics of the "capital construction facilities" total area, leasable, leased area

The result of the visualization with the assistance of Python libraries:



Fig. 5. An example of using Python to visualize indicators at the strategic management level: The main characteristics of the "capital construction facilities" total area, leasable, leased area

	XX7 4 1 1 1 C 1 4				
Name of the tool	what kind of data preprocessing was re- quired	The need for an In- ternet connection	Time spent on data preprocessing	Time to search for a visualization form	The ability to cus- tomize the result ob- tained
Microsoft Excel	In order to start work- ing with pivot tables and build any visual- izations, you need to remove all unneces- sary rows and columns in the source data. This will save you from unnecessary in- terfering filters in the slices and in the rest of the filtering and sort- ing windows in the fu- ture	Not required	When executing Case 1, it took $\sim 10 - 20$ minutes to pre-process the data and create pivot tables (when understanding which slices and samples need to be visualized)	It took $\sim 10-15$ min- utes to choose all the necessary visualization forms	It is possible to cre- ate a simple interactive dashboard completely for yourself. The set of diagrams and the func- tionality of their add-in is very limited
Microsoft Power BI	When executing Case 1, data preprocessing was not needed. All extra rows and columns are easily disabled and do not interfere with further work	Not required (when working in the free Desktop version). It is possible to use the paid Power BI cloud service to view data and dashboards from different devices, as well as to work with the Power BI Rest API (to output dashboards and work with them on your own web service)	Thanks to the built- in PowerQuery pivot table editor, the data preprocessing process takes much less time. When executing Case 1, no time was spent on preprocessing	It also took $\sim 10$ min- utes to decide on all the necessary forms	Almost limitless func- tionality for the imple- mentation of a conve- nient dashboard com- pletely for yourself, with the ability to cus- tomize each element. A huge selection of different charts. Full interactivity
Python libs	It is necessary to remove unnecessary lines when working directly with data in Python. This is done by simple actions with line slices	Not required (when working with Anaconda)	On average $\sim 2-5$ minutes. When execut- ing Case 1, no time was spent on prepro- cessing	$\sim 10$ minutes to an- alyze and select the necessary form of vi- sualization	It is possible to add new data that will automatically be re- flected on the chart. To get a different form of visualization, you need to write the appropri- ate part of the code

 TABLE II

 Results of a comparative analysis of the use of data tools

Thus, after visualizing the test case in three different environments, we came to the conclusion that for large volumes of disparate data and more complex tasks, it is much more convenient and functional for Power BI with its capabilities to create complex data models, interactive dynamic dashboards, as well as a huge library of various visualizations available. For full use, it is recommended to master the DAX (Data Analysis Expressions) language for writing tabular queries and formulas. Excel may be required only at the phase of studying data and working out general technical solutions for visualization and data preparation. Python libraries are designed for more experienced users and are convenient for various kinds of tasks from simple and urgent to the use of data mining methods in solving corporate tasks at various levels of management.

# VII. RESULTS OF INTERPRETATION OF MODELS AT DIFFERENT LEVELS OF MANAGEMENT

It is a well-known fact that a person perceives the main amount of information through the visual channel, interpreting the picture of the environment, and in the case of presenting information, trying to convey specific idea to an audience, data visualization is the most preferred way of transmitting it. Visualization of multidimensional data is an integral part of the workflow of every company. The formation of reports and current summaries, time and frequency correlations between producers and consumers, income and expense data, and much more is part of a large system that uses charts to work with analytics at every level of management.

A lot of companies neglect the time and energy to create charts, considering the tool for creating them primitive. In majority of cases, unfortunately, at the output a large number of illogical, not visual graphs with an overabundance of information, details and colors are obtained, which eliminates the opportunity for an ordinary viewer to understand the basic idea of the graph. For example, a pie chart is the most used in data visualization, which is absolutely unjustified, given the low practicality of this type.

Based on the results of working out the requirements for data visualization for decision makers, in order to increase the efficiency and speed of decision-making at each level of management, several techniques have been selected to create high-quality graphs:

# 1) Immersion in the topic

Firstly, you need to understand what the data presented in the table displays. All quantitative values have a certain meaning. They are arranged according to some principle, following a certain methodology. It is crucial to understand which subject area these values belong to. Also, it is important to understand exactly which data located in the table needs to be visualized.

2) Fundamentals of statistics

We work with an array of data on which, if necessary, we can perform actions using various statistical methods, but there is a need to be guided only by technically sound solutions in order to avoid meaningless and incorrect calculations.

3) Modern business analytics data analysis

Undoubtedly, there is a need to be able to analyze data. This is the main point when visualizing data.

There are two types of analysis:

Statistical analysis is the ability to work with certain values.

Visual analysis is finding the nature and patterns of data changes in the process of their graphical representation. It is necessary to immediately represent a series of numbers in the form of graphs.

4) Creating a pivot data table

In order to highlight the main information and/or further work only with the data that can be depicted on the graph, you need to create a pivot table. This greatly facilitates the work process and avoids errors caused by mixing different types of data.

5) Formulation of the idea

Data visualization is a communication tool, so there is a need to formulate a clear message that we want to convey through graphs. Depending on this, it will be possible to start from the subject area of the data and select the optimal way of visualization.

6) Selection of the optimal visualization method Based on all the previous paragraphs, a complete picture of the data presented in the table will be created.

The effectiveness of the type of visualization as a tool for delivering a message is determined by:

- Clarity. The case when a person immediately understands the essence of the graph, how numeric values are encoded, etc;
- Convenience. It is important that the human eye accurately decode the graphic images back into numerical values;
- The correctness of the chosen visualization method.

Any aspect of the data that needs to be paid attention to can be expressed through one of the five main types of comparison:

- 1) Component-by-component comparison shows the size of each component as a percentage of the whole;
- Positional comparison shows the location of objects relative to each other;
- 3) Time comparison shows the change over time;

- Frequency comparison shows how many objects fall into a series of consecutive intervals;
- 5) Correlation comparison shows the presence or absence of dependence between variables.

There are basic types of relationships between data and the most effective graphs.

Component-by-component comparison: pie chart.

Positional comparison: bar chart.

Time and frequency comparisons: histogram (bar chart), line chart ( slope).

Correlation comparison: bar chart, dot chart.

The main purpose of graphs is to display the main idea and connections faster and more clearly than tables and simple data summaries, so it is very important to ensure that the form does not overwhelm the content.

It is always better to use the idea that needs to be expressed using a diagram as the title of the graph. It should express the main idea, be clear and concise, so it will immediately focus on a certain, most important aspect of the data presented. Sometimes it is necessary to make diagrams without headings, but this does not affect the stage of formulating an idea and the correct placement of accents, it still plays a key role.

A. Application of the approach on the example of cases of visualization of indicators at the strategic and management control levels

# The first example

Let's use an example to illustrate the graph obtained by following all the steps described above.

The task was to display indicators in a summary table and chart: The number (absolute value and percentage) of signed agreements of residents (the name of the resident in the "Name" field of the RezidentReestr file in the "Direction" fields of the RezidentReestr file, the name of the Department of the "personal account of the resident" file in the names of the sheets.

An Excel file with a large number of sheets and a large amount of heterogeneous data was received at the input.

The first step was to understand what subject area we are working with. During the review of the file, it becomes clear that we are working with residents of the Corporation, distributed by departments. Then, after looking at the task, we understand that we need to display the number of signed agreements. Therefore, for visualization, the following distribution is needed for 3 sets of groups: department  $\rightarrow$  resident  $\rightarrow$  signed/not signed contract. For the convenience of visualization on the platform (Jupiter Notebook/Google Collab), you need to make a pivot table, which is the next step that brings us closer to the visualization itself.

Part of the pivot table looks as follows:

Попортомонт		Desugar	Договор ЛКР (текущий	
департамент	٣	Резидент	статус)	*
ДМВОиЭД		ООО "Галерея Анастасии Чиковой"	Зарепистрирован	
ДМВОиЭД		ООО " ЛУКОЙЛ- Центрнефтепродукт"	Зарегистрирован	
ДМВОиЭД		ФГБУК "Государственный центральный музей кино"	Подготовлен	7
ДМВОиЭД		Государственный музей Востока	Подготовлен	
ДМВОиЭД		АНО "Цифровая трансформация"	Подготовлен	
ДМВОиЭД		"CMK"	Подготовлен	
ДМВОиЭД		000 "POEOMAPKET"		
ДМВОиЭД		АНО "Московская дирекция транспортного обслуживания"	Зарегистрирован	
ДМВОиЭД		АНО "МВТЦ АМК ФСО России"	Зарепистрирован	
ДМВОиЭД		Фонд гуманитарных проектов	Зарепистрирован	
ДМВОиЭД		АНО "Развитие человеческого калитала"	Подготовлен	
ДМВОиЭД		Департамент градостроительной попитики города Москвы	Подготовлен	
ДМВОиЭД		OOO MCE "ЭКСПОХЛЕБ"	Подготовлен	
ДМВОиЭД		ГПБУ "Мосприрода"	Зарепистрирован	
ДМВОиЭД		ООО "Детское Посопьство"	Зарепистрирован	
ДМВОиЭД		ООО "ТОПФРЕНД"	Подготовлен	
ДМВОиЭД		ФРиПИ "АЙРИС"	Подготовлен	
ДМВОиЭД		ФГБУК ГМВЦ "РОСИЗО"	Подготовлен	-

Fig. 6. Example of forming a summary table to obtain indicators for the management control level

You can now work with this table to visualize certain indicators.

Now that everything is ready, you need to start choosing the optimal visualization method. Since the purpose of visualization is to display the number of signed agreements and the difference in the number between departments, the bar chart will be the most suitable type of chart. In each department, you can easily display both the percentage and the quantitative ratio, in which case the information will be clear to the reader.

It will also be good to display the ratio of the number of residents who signed the contract to all residents.

The final result of visualization:



Fig. 7. Example of visualization of indicators for the management control level "The ratio of the number of residents who signed the contract to all residents with reference to the responsible departments of the Corporation"

#### The second example

A good example for a visual representation of the work done is also the analysis of the following case.

The subject area, as in the previous case, is numerical characteristics reflecting the number of points scored by each resident of the Corporation, distributed by departments.

Our task is to display the number of points scored by each of the five residents according to various criteria, taking into account also the description, representing a more complete and accurate formulation for each area.

Below is a fragment of the table obtained for visualization:

Критерии	Описание	Оценка	Балл
<ol> <li>Своевременность платежей</li> </ol>	Сколько раз Резидент задержал платежи по договорам с ВДНХ?	2	-40,0
<ol> <li>Своевременность предоставления отчётности</li> </ol>	Сколько раз Резидент несвоевременно предоставил данные для суточного отчёта?	2	-5,0
	Количество мероприятий, в которых участвовал Резидент	4	30,0
<ol> <li>участие в мероприятиях</li> </ol>	Количество обязательных мероприятий в которых Резидент не участвовал	1	-10,0
<ol> <li>Объём скидок для сотрудников ВДНХ</li> </ol>	Какую скидку (в процентах) Резидент предоставляет сотрудникам ВДНХ?	10	2,5
5. Чистота объекта Сколько поступило жалоб на чистоту объекта Резилента?		0	5,0
6. Отзывы и жалобы Сколько мы получили негативных отзывов о посетителей Резиденте?		0	7,5
<ol> <li>Интеграция в билетную систему</li> </ol>	Резидент интегрирован в билетную систему ВДНХ?	1	2,5
Downey you guidentia			
<b>Результат оценки:</b>	-7,5		
Итоговый рейтинг	2,5		

Fig. 8. A fragment of the working table for obtaining indicators for the strategic management level "Rating of residents by a set of indicators"

As we can note, there are five residents, 8 criteria and a complete formulation of each of them is required, as well as the number of points for each criterion for each resident. When analyzing the source data, we came to the decision that the most visual way to visualize this case — a horizontal lollipop chart.

We have chosen an own color for each resident, the palette of all residents is contrasting, which allows us not to confuse the data of objects.

•	Резилент 1
Ξ.	Desugeur 2
•	Резидент 2
•	Резидент 3
٠	Резидент 4
٠	Резидент 5

Fig. 9. Selection of color visualization elements on the example of indicators for the strategic management level "Rating of residents by a set of indicators"

The left side of the diagram fully contains the criteria for evaluating residents.



Fig. 10. Selection of visualization elements in terms of the names of one of their axes of the diagram on the example of indicators for the strategic management level "Rating of residents by a set of indicators"

The lower panel reflects the number of points of each object under study, depending on the criterion under consideration.

#### As a result, we get a graph:



Fig. 11. An example of the visualization result on the example of indicators for the strategic management level "Rating of residents by a set of indicators"

So, using the examples discussed above, you can clearly see the whole underside of the work, the value of which lies not only in selecting the most appropriate form of visualization, but also working with colors, formulating headings, numerical characteristics and their placement on the user's field of view, data readability and fulfilling the basic requirement — faster user interpretability of the main statements, trends or the general structure of this array of information.

#### VIII. CONCLUSION

Thus, we have developed an approach to visualization in an Information and Analytical System from the standpoint of providing effective support for management decisions. Analyzing different levels of management in a hierarchical view of the management structure, we came to the direct definition of visualization objects and characteristics of the states of their indicators. We also conducted a comparative analysis of tools for creating a visual form of data representation. We have also developed a new approach to data visualization that takes into account specific recommendations for developers, which includes recommendations for presenting data, improving their perception.

#### REFERENCES

- G. Anthony Gorry, Michael S. S. Morton, "A FRAMEWORK FOR MANAGEMENT INFORMATION SYSTEMS\*", February 1971, 510-71
- [2] RYSZARD BARNAT, LLM., DBA, PHD (STRAT. MGMT), "Strategic Management: Formulation and Implementation"
- [3] Harvard Business Review, April 7, 2020, "HBR's 10 Must Reads on Strategy", 2-Volume Collection Paperback
- [4] Robert N. Anthony, "Planning and Control Systems: A Framework for Analysis", 1965