



APPROVED

Vice-rector for educational activities

Veronika.G. Shubaeva

«26» мая 2023.

## Анализ данных / Data Analysis

### Syllabus of the course

Specialty	38.03.02 Management
Specialization	Business management and digital innovations
Level of higher education	Bachelor
Form of training	Full-time
Year of enrolment	2023

Authored by:

PhD, Fridman Grigory Moritsovich

Total number of hours	144	<b>Form of final attestation:</b>  Exam: semester 5
incl:		
contact work	48	
self-study	60	
practical training	0	
control hours	36	

#### Hours distribution:

Semester:	5
Type of classes	Hours
Contact hours	20
Practical training	28
Laboratory work	
<b>Total contact hours</b>	<b>48</b>
Self-study	60
Control hours	36
<b>Total academic hours</b>	<b>144</b>
<b>Total credits</b>	<b>4</b>

# CONTENTS

<b>1. LEARNING OBJECTIVES .....</b>	<b>3</b>
<b>2. COURSE PLACE IN THE PROGRAMME STRUCTURE.....</b>	<b>3</b>
<b>3. EXPECTED LEARNING OUTCOMES .....</b>	<b>3</b>
<b>4. COURSE STRUCTURE AND CONTENT .....</b>	<b>3</b>
<b>5. TEACHING AND LEARNING TOOLS OF THE COURSE.....</b>	<b>4</b>
<b>5.1 Recommended literature .....</b>	<b>4</b>
<b>5.2 List of software (including national production).....</b>	<b>5</b>
<b>5.3 List of reference systems and modern professional databases .....</b>	<b>5</b>
<b>6. TECHNICAL FACILITIES.....</b>	<b>5</b>
<b>7. METHODOLOGICAL GUIDELINES FOR STUDENTS .....</b>	<b>6</b>
<b>8. SPECIFICATIONS FOR TEACHING DISABLED PERSONS .....</b>	<b>7</b>
<b>ASSESSMENT RESOURCES .....</b>	<b>8</b>
<b>1.1 Control tasks and assignments for interim attestation .....</b>	<b>8</b>
<b>1.2 Topics for written task .....</b>	<b>8</b>
<b>1.3 Interim checkpoints.....</b>	<b>9</b>
<b>1.4 Other assessment objects .....</b>	<b>9</b>
<b>1.5 Self-study.....</b>	<b>9</b>
<b>1.6 Grading scale .....</b>	<b>9</b>

## 1. LEARNING OBJECTIVES

<b>Objective:</b>	Gaining skills in using methods and models of data analysis and appropriate computer tools for solving economic problems.
-------------------	---

## 2. COURSE PLACE IN THE PROGRAMME STRUCTURE

The discipline B1.O. Data Analysis is a part of Block 1.

## 3. EXPECTED LEARNING OUTCOMES

Code and name of graduate competence	Code and name of the competence achievement indicator	Expected learning outcomes
GPC-6 – Able to understand the principles of operation of modern information technologies and use them to solve the problems of professional activity	GPC-6.2 – Uses methods and software for data collection, processing and analysis	<p>To know: the main methods of collecting, processing and analyzing data to solve the problems of professional activity</p> <p>To be able to: apply the methods of collecting, processing and analyzing data to solve the problems of professional activity.</p> <p>To possess: the skills of using modern software tools for collecting, processing and analyzing data.</p>
UC-1 – Able to search, critical analysis and synthesis of information, apply a systematic approach to solve tasks	UC-1.2 – Develops options for solving a problem situation based on a critical analysis of available sources of information	<p>To know: the main methods and means of analyzing available data sources and approaches to making decisions based on the information collected</p> <p>To be able to: apply methods and methods of analyzing available data sources to make decisions based on the information collected</p> <p>To possess: the skills of using modern software tools to analyze available data sources and make decisions based on the information collected</p>

## 4. COURSE STRUCTURE AND CONTENT

Code and name of the topics	Course content	Academic hours			
		Contact work			Self-study
		Lectures	Practices	Workshops	
Topic 1. Exploratory data analysis	Types and methods of representing data of various nature. Data visualization. Analysis of the main properties of data, finding common patterns and distributions in them. Univariate and multivariate data analysis.	4	4		8

Topic 2. Basic statistical methods for data analysis.	Descriptive statistics. Testing statistical hypotheses. Correlation analysis. Analysis of contingency tables.	4	6		12
Topic 3. Data processing of various nature.	Work with passes and emissions. Normalization and standardization of data. Processing of categorical features.	4	6		12
Topic 4. Methods for reducing the dimension of source data	Principal component method. T-SNE method. UMAP method.	4	6		14
Topic 5. Cluster analysis.	Clusters. proximity metrics. Methods for combining clusters. Hierarchical cluster analysis. Dendrograms. K-means method. DBSCAN method. Methods for assessing the quality of clustering. silhouette	4	6		14
<b>Control hours:</b>					<b>36</b>
<b>Total hours:</b>		<b>20</b>	<b>28</b>	<b>0</b>	<b>60</b>

## 5. TEACHING AND LEARNING TOOLS OF THE COURSE

### 5.1 Recommended literature

Bibliographic description of the publication (author, title, type, place and year of publication, number of pages)	Digital resources
Nivorozhkina L.I. Statistical methods of data analysis: textbook / L.I. Nivorozhkina, S.V. Arzhenovsky, A.A. Rudyaga [and others]; under total ed. L.I. Nivorozhkina. — Electron. Dan. - M.; RIOR Publishing Center: LLC "Scientific Publishing Center INFRA-M", 2016. - 333 p.	<a href="http://znanium.com/go.php?id=556760">http://znanium.com/go.php?id=556760</a>
Data analysis: a textbook for universities / V. S. Mkhitarian [and others]; edited by V. S. Mkhitarian. - Moscow: Yurayt Publishing House, 2021. - 490 p.	<a href="https://urait.ru/bcode/469022">https://urait.ru/bcode/469022</a>
Mirkin, B. G. Introduction to data analysis: textbook and practical work / B. G. Mirkin. - Moscow: Yurayt Publishing House, 2020. - 174 p.	<a href="https://urait.ru/bcode/450262">https://urait.ru/bcode/450262</a>
Gurikov S.R. Fundamentals of algorithmization and programming in Python: a tutorial. — Electron. Dan. – M.: FORUM: INFRA-M, 2019. – 343 p.	<a href="http://znanium.com/go.php?id=970143">http://znanium.com/go.php?id=970143</a>
Kazansky, A. A. Applied programming on Excel 2019: a textbook for secondary vocational education / A. A. Kazansky. - 2nd ed., revised. and additional - Moscow: Yurayt Publishing House, 2021. - 171 p.	<a href="https://urait.ru/bcode/470405">https://urait.ru/bcode/470405</a>

## 5.2 List of software (including national production)

- 7-Zip
- LibreOffice
- ОС АЛЪТ образование 10
- Jupyter Notebook
- Python

## 5.3 List of reference systems and modern professional databases

№	Name of reference systems and professional databases
1.	Digital library Grebennikon.ru – <a href="http://www.grebennikon.ru">www.grebennikon.ru</a>
2.	Science Digital Library eLIBRARY – <a href="http://www.elibrary.ru">www.elibrary.ru</a>
3.	Science Digital Library КиберЛеника – <a href="http://www.cyberleninka.ru">www.cyberleninka.ru</a>
4.	Database ПОЛПРЕД Справочники – <a href="http://www.polpred.com">www.polpred.com</a>
5.	Database OECD Books, Papers & Statistics on the platform OECD iLibrary <a href="http://www.oecd-ilibrary.org">www.oecd-ilibrary.org</a>
6.	Legal reference system КонсультантПлюс (installed resource UNECON or <a href="http://www.consultant.ru">www.consultant.ru</a> )
7.	Legal reference system «ГАРАНТ» (installed resource UNECON or <a href="http://www.garant.ru">www.garant.ru</a> )
8.	Information and referral system «Кодекс» (installed resource UNECON or <a href="http://www.kodeks.ru">www.kodeks.ru</a> )
9.	Digital library system BOOK.ru - <a href="http://www.book.ru">www.book.ru</a>
10.	Digital library system ЭБС ЮРАЙТ – <a href="http://www.urait.ru">www.urait.ru</a>
11.	Digital library system ЗНАНИУМ (ZNANIUM) – <a href="http://www.znanium.com">www.znanium.com</a>
12.	Digital library UNECON – <a href="http://opac.unecon.ru">opac.unecon.ru</a>

## 6. TECHNICAL FACILITIES

There are special rooms for lectures, seminars, coursework, group and individual consultations, current and interim assessments, as well as rooms for self-study.

The premises are equipped with equipment and teaching aids.

The rooms for students' independent work are equipped with computers with Internet connection and access to the university's electronic learning environment.

Name of classroom	Classroom location
Classroom 2004 Training classroom (for conducting lecture-type classes and seminar-type classes, course design (term papers), group and individual consultations, current control and intermediate certification), is equipped with a multimedia complex. Specialized furniture and equipment: Educational furniture for 54 seats, work teacher's place, department - 1 pc., chalk board (3-section) - 1 pc., chair - 1 pc., blinds - 2 pcs., Computer Intel i3-2100 2.4 Ghz/4Gb/500Gb/Acer V193 19" - 1 pc., Multimedia projector Type 1 Optoma x 400 - 1 pc. Sets of demonstration equipment and teaching and visual aids: multimedia applications for lecture courses and practical exercises, interactive teaching and visual aids.	191023, St. Petersburg, Griboedova canal, 30-32, lit. A, Б, Р
Classroom 2022 Laboratory "Laboratory complex" Specialized furniture and equipment: Educational furniture for 19 seats (19 computer tables, 19 black chairs) teacher's workplace (computer table 1 pc., chair 1 pc.), single-section chalk board 1	191023, St. Petersburg, Griboedova canal, 30-

pc., marker board on wheels 1 pcs., table 1 pc., chair 1 pc., blinds 1 pc., rack hanger 1 pc. Computer Intel i5 4460/1TB/8GB/monitor Samsung 23" - 1 pc., Computer Intel i5 4460/1TB/8GB/monitor Samsung 23" - 18 pcs. Sets of demonstration equipment and teaching aids: multimedia applications for lecture courses and practical exercises, interactive teaching and visual aids	32, lit. A, B, P
---	------------------

## 7. METHODOLOGICAL GUIDELINES FOR STUDENTS

The following documents should be made available to the trainee before the start of the course:

- training and methodological documentation;
- local normative acts regulating the main issues of the organisation and implementation of educational activities, including those regulating the procedure for current monitoring and interim assessment of students;
- the schedule of consultations of the teaching staff.

The level and depth of mastering the discipline is determined by the active and systematic work of students in lectures, seminars, independent work, including in terms of identifying the most significant and relevant problems for further study. A special condition for qualitative mastering of the discipline is an effective organisation of work, which allows distributing the academic workload evenly in accordance with the schedule of the educational process.

When preparing for classes, students have the opportunity to attend consultations with the staff of UNECON according to the timetable set out in the schedule of consultations.

The students' in- and out-of-classroom work should aim to form:

- the fundamentals of the learner's world view and scientific understanding;
- basic knowledge relevant to the training area and the declared professional field, forming the target and professional basis for training;
- professional competences oriented towards the needs of the labour market;
- an individual trajectory by mastering a unique set of professional competences that complement the learner's competence model, through a focus on specific professional specialised areas of knowledge defined by labour market representatives;
- meta-skills for learners, such as teamwork and leadership, data analysis, digital skills, project design and implementation, intercultural interaction.

## **8. SPECIFICATIONS FOR TEACHING DISABLED PERSONS**

Students with disabilities, if necessary, are taught on the basis of an adapted work programme using special teaching methods and didactic materials that take into account the particularities of their psychophysical development, individual capacities and health status.

In order for disabled persons and persons with disabilities to master the curriculum, the University shall ensure that:

- for the visually impaired and visually impaired: availability of information on the timetable in accessible places and adapted forms for learners who are blind or visually impaired; presence of an assistant to assist the learner as needed; production of alternative formats of teaching materials (large print or audio files);

- for the hearing-impaired and hearing-impaired: adequate sound reproduction of information;

- for persons with disabilities and persons with mobility impairments: the possibility of unimpeded access for students to classrooms, restrooms and other areas of the department, as well as their stay in these areas.

Learners with disabilities and persons with disabilities are provided with printed and/or electronic educational resources in forms adapted to their disabilities. The education of students with disabilities may be organised with other students or in separate groups or organisations.

## **ASSESSMENT RESOURCES**

### **1.1 Control tasks and assignments for interim attestation**

1. Statement of the problem of data analysis. Basic concepts. Task types. Examples of applied tasks.
2. Name the types of measuring scales known to you.
3. Estimation of parameters and construction of statistical conclusions.
4. Correlation analysis of a multidimensional general population, its purpose and conditions of application.
5. Application of the contingency table and the chi-square test.
6. Methods for processing categorical variables. Coding of categorical variables.
7. Algorithms for constructing a decision tree. Handling gaps in data.
8. Methods for searching for local anomalies in a data set.
9. Outlier detection methods. Emission processing. Tukey borders.
10. Methods for reducing the dimension of initial data. Main characteristics.
11. t-SNE and UMAP methods for data dimensionality reduction.
12. Principal component method. Formulation of the problem. Covariance matrix of initial features.
13. Selection of factors in the method of principal components. factor loads. Model quality estimates. Interpretation of the obtained results.
14. Basic Numerical Characteristics and Optimality Properties of Principal Components.
15. Conditions for applicability of the principal component method. Graphic approach.
16. For what research purposes is cluster analysis used? Cluster definition.
17. Proximity metrics in clustering problems. Methods for combining clusters.
18. Methods of hierarchical cluster analysis.
19. Nearest neighbor method. Method  $k$  nearest neighbors. The method of weighted nearest neighbors. Choice of weight function.
20. Methods for assessing the quality of clustering.
21. Application of dendrograms in hierarchical cluster analysis.

### **1.2 Topics for written task**

Is not provided by the work programme of the discipline.



### 1.3 Interim checkpoints

Number	Type	Method of conduct	Topic number
1	Analytical work	with the help of technical means and information systems	1-3
2	Analytical work	with the help of technical means and information systems	4-5
3	Monitoring	with the help of technical means and information systems	1-5

### 1.4 Other assessment objects

Is not provided by the work programme of the discipline.

### 1.5 Self-study

Name of self-study	Topic number
Doing homework	1-5
Preparation for lectures and practical classes	1-5
Exam preparation	1-5

### 1.6 Grading scale

Scales of assessment and procedures for assessing learning outcomes of the discipline are regulated by the Regulations on the current control of progress and interim attestation of students in higher education programmes and the Regulations on the scoring and rating system.

**A grading and rating system** is used to assess the learning outcomes of the discipline:

The final control of the discipline is an examination (or a differentiated test), the final grade being formed in accordance with the scale given in the table below:

Points	Grade
$\leq 54$	fail
55-69	satisfactory
70-84	good
$\geq 85$	excellent

### Grading scale

2 (points to 54)	Demonstrates a lack of understanding of the problem. Many of the requirements of the assignment are not met. An initial perception of the material is demonstrated. The work is incomplete and/or plagiarised.
------------------	---

3 (points 55-69)	Demonstrates a partial understanding of the problem. Most of the requirements of the task have been met. Mastery of the elements of the assigned material. The material is mostly clear and coherent.
4 (points 70-84)	Demonstrates considerable understanding of the issue by the discipline. All requirements of the assignment are fulfilled. The content of the completed tasks is disclosed and examined from different perspectives.
5 (points 85-100)	Demonstrates full understanding of the problem. All requirements of the assignment are fulfilled. Demonstrates proficiency in the discipline. The completed assignments are holistic, complete, structured, present different points of view and demonstrate creativity.