GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	BIG DATA, PG_00070564		
Field of study	Economic Analytics		
Date of commencement of studies	October 2026	Academic year of realisation of subject	2027/2028
Education level	second-cycle studies	Subject group	Specialty subject group Subject group related to scientific research in the field of study
Mode of study	Part-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	English
Semester of study	3	ECTS credits	4.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Statistics and Econometrics -> Faculty of Management and Economics -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr hab. Michał Pietrzak	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	9.0	0.0	27.0	0.0	0.0	36

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	36	3.0	61.0	100

Subject objectives

preparing students to formulate innovative solutions to analytical problems under uncertainty, based on knowledge of data reliability assessment and economic analytics, as well as to shape attitudes related to the responsible use of data in the context of designing and applying advanced Big Data analysis methods.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_W06] knows and understands the principles of evaluating the reliability of utilized data, applying in-depth specialized knowledge in the field of economic analysis.	knows and understands the principles of data reliability assessment and their importance in Big Data analysis, using specialised knowledge in economic analytics.	[SW3] Assessment of knowledge contained in written work and projects
[K7_U01] creates innovative solutions for complex and unstructured processes, considering unpredictable environmental conditions through the synthesis of information from various sources.	is able to develop analytical solutions in the field of Big Data for complex and dynamic problems by integrating data from multiple sources.	[SU2] Assessment of ability to analyse information [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools [SU5] Assessment of ability to present the results of task
[K7_K03] responsibly fulfills professional roles, demonstrating the ability to identify ethical dilemmas and recognize and evaluate alternative courses of action.	is ready to use data and analytical results responsibly, taking into account ethical dilemmas and the consequences of decisions made.	[SK2] Assessment of progress of work [SK5] Assessment of ability to solve problems that arise in practice

Subject contents

Course content – lecture

1. Introduction to Big Data and the ETL process (ExtractTransformLoad) data characteristics, data sources, data processing, data analytics and solving economic problems
2. Excel as a data source and programming in VBA (Visual Basic for Applications) tabular structures, VBA programs for data cleaning, transformation and processing, ETL automation
3. DataFrame in Python and R and data analytics on DataFrame operations (filtering, aggregation, join), data preparation, preliminary data analysis
4. Data acquisition from API and JSON API queries, analysis based on JSON, DataFrame preparation (flattening JSON structures)
5. Data transformation cleaning (missing values, duplicates, errors), type conversion, standardization, validation and data preparation for analysis
6. Databases: relational (PostgreSQL) and non-relational (NoSQL) creating databases and tables, SQL queries, data modeling and storage
7. Integration of databases with Python and R database connections, data reading and writing, use of SQL in analysis
8. Introduction to Apache Spark distributed processing, Spark DataFrame, working with large datasets
9. Data processing in Spark (DataFrame, API, JSON) data loading, transformations, large-scale operations
10. Data analysis using open source software (Python, R) and databases exploratory data analysis, visualization, working with data from PostgreSQL and Spark

Course content – laboratory

1. Designing the ETL process and analysis of example economic problems on real data (Introduction to Big Data and ETL)
2. Creating and automating data processing in Excel using VBA macros for data cleaning and transformation (Excel and VBA)
3. Working with data in DataFrame structure in Python and R filtering, aggregation, joining data and preliminary data analysis (DataFrame and data analytics)
4. Data acquisition from API and processing JSON format into DataFrame in Python and R HTTP queries and flattening data structures
5. Data cleaning and transformation handling missing data, removing duplicates, type conversion, preparing data for analysis (Data transformation)
6. Creating and managing relational and non-relational databases table design, SQL queries, data organization (Relational and non-relational databases)
7. Integration of Python and R with a database reading, writing and analyzing data directly from programming languages (Database integration)
8. Data processing in the Apache Spark environment operations on large datasets using Spark DataFrame (Introduction to Apache Spark)
9. Advanced data transformations in Spark loading data from various sources (JSON, API, databases) and large-scale processing (Data processing in Spark)
10. Exploratory data analysis and visualization using open source tools and data from databases building analytical insights (Data analysis in Python, R, Spark and PostgreSQL, NoSQL)

Prerequisites and co-requisites

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Computational and decision-making tasks	60.0%	40.0%
Written exam with open questions	60.0%	60.0%

Recommended reading

Basic literature	1. VanderPlas, J. (2016). Python Data Science Handbook. OReilly Media. 2. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of Massive Datasets (3rd ed.). Cambridge University Press.
Supplementary literature	1. Date, C. J. (2019). An Introduction to Database Systems. Addison-Wesley. 2. Elmasri, R., & Navathe, S. (2016). Fundamentals of Database Systems (7th ed.). Pearson.
eResources addresses	

Example issues/ example questions/ tasks being completed

1. Discussion of the concept of Big Data and the stages of the ETL process along with examples of applications in analyzing economic problems
2. Design and implementation of a simple ETL process using data from various sources (e.g. Excel, API)
3. Creating and using VBA macros in Excel to automate data cleaning and transformation
4. Operations on DataFrame structures in Python and R filtering, aggregation, joining data and interpretation of results
5. Data acquisition from API and transformation of JSON data into tabular form (DataFrame) by flattening structures
6. Identification and solving data quality problems missing data, duplicates, errors and their impact on analysis
7. Designing and executing SQL queries in relational databases and comparison with the NoSQL approach
8. Integration of Python and R with databases performing data writing, reading and analysis
9. Processing large datasets in the Apache Spark environment DataFrame operations and optimization of processing
10. Conducting exploratory data analysis and visualization using open source tools on data from databases and Big Data systems

Practical activites within the subject

Not applicable