GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	DATA ANALYSIS WITH R AND PYTHON, PG_00061091		
Field of study	Economic Analytics		
Date of commencement of studies	October 2023	Academic year of realisation of subject	2024/2025
Education level	second-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-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	Katedra Statystyki i Ekonometrii -> Faculty of Management and Economics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr inż. Karol Flisikowski	
	Teachers	dr inż. Karol Flisikowski	

Lesson types and methods of instruction	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	0.0	0.0	45.0	0.0	0.0	45
	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	45	8.0	47.0	100

Subject objectives	Proposes innovative solutions to complex and unstructured problems using modern analytical methods, supporting its activities the use of information technology

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K7_W03] demonstrates in-depth preparation in the application of analytical methods and techniques for formulating and solving problems	uses the possibilities of support by modern information technologies in in-depth analysis of economic and social data	[SW1] Assessment of factual knowledge
	[K7_U01] creates innovative solutions to complex and unstructured problems, taking into account the variability of the environment by synthesising information from many sources	formulates innovative solutions to economic and social problems, synthesizing information from many sources	[SU3] Assessment of ability to use knowledge gained from the subject

Subject contents	Introduction to R and Python languages. Basic operations. Data sources. Importing data from different formats in R vs Python Variables and data types in R vs Python (vector, data frame, array, list, arrays, sets, dictionaries) Basic functions - descriptive and mathematical statistics in R vs Python Basic data processing (new variables, filters, combining frames: transform, split, concatenate) in R vs Python Dirty data - missing observations; duplicates; outliers; formatting errors. Naniar package. Python libraries Data processing using Dplyr and Tidyr. data transformation in Python Data cleaning - outliers and missing values. Imputations. Transformations and discretization of variables in Python. Optimal binning Graphics in R - basic and advanced graphical presentation of data (packages: ggplot2; Lattice; Grid) vs Graphics in Python (Matplotlib; Plotly, etc.) Analysis reporting with R/Python - introduction to R-Markdown and Quarto (notepad; presentations - R and Powerpoint; HTML slides; PDF beamer, etc.). Shiny Apps for Python Machine learning - linear models, generalized linear models. Iterative model fitting. Reporting Application of k-nearest neighbor (KNN) method Classification and clustering. Linear discriminant analysis; graphical trees; logistic regression Bagging and random forests Boosting method. XGBoost
Prerequisites and co-requisites	

Assessment methods and criteria	Subject passing criteria	Passing threshold	Percentage of the final grade
	Project - reports	60.0%	50.0%
	Test	60.0%	50.0%

Recommended reading	Basic literature	Bruce Peter, Bruce Andrew, Gedeck Peter, Statystyka praktyczna w data science. 50 kluczowych zagadnień w językach R i Python, Helion, 2021 Chantal D. Larose, Daniel T. Larose, Data Science Using Python and R, Wiley, 2019 Rick J. Scavetta, Boyan Angelov, Python and R for the Modern Data Scientist, O'Reilly Media, 2021
	Supplementary literature	Wes McKinney, Python w analizie danych. Przetwarzanie danych za pomocą pakietów Pandas i NumPy oraz środowiska IPython. Wydanie II. Helion, 2018 Marek Gągolewski, Maciej Bartoszuk, Anna Cena, Przetwarzanie i analiza danych w języku Python, PWN, 2017 Ajay Ohri, Python for R Users: A Data Science Approach, Wiley, 2017 Hadley Wickham, Garrett Grolemund. R for Data Science, https://r4ds.had.co.nz J. Hathaway, Katie Larson, Python for Data Science, https://byuidatascience.github.io/python4ds/
	eResources addresses	Adresy na platformie eNauczanie:

Example issues/ example questions/ tasks being completed	Data preprocessing in R and Python languages Exploratory Descriptive Data Analysis (DEA) report in R and Python Statistical inference from a sample in R and Python
Work placement	Not applicable