

## 。 GDAŃSK UNIVERSITY OF TECHNOLOGY

## Subject card

Subject name and code	Descriptive statistics, PG_00045293							
Field of study	Data Engineering							
Date of commencement of studies	October 2025		Academic year of realisation of subject		2025/2026			
Education level	first-cycle studies		Subject group			Obligatory subject group in the field of study		
						Subject group related to scientific research in the field of study		
Mode of study	Full-time studies		Mode of delivery			blended-learning		
Year of study	1		Language of instruction		English			
Semester of study	2		ECTS credits		3.0			
Learning profile	general academic profile		Assessment form		assessment			
Conducting unit	Department Of Statistics And Econometrics -> Faculty Of Management And Economics -> Wydziały Politechniki Gdańskiej							
Name and surname	Subject supervisor	dr inż. Karol Flisikowski						
of lecturer (lecturers)	Teachers		dr inż. Karol Flisikowski					
Lesson types and methods	Lesson type	Lecture	Tutorial	Laboratory	Project	t	Seminar	SUM
of instruction	Number of study hours	15.0	0.0	15.0	0.0		0.0	30
	E-learning hours inclu	uded: 18.0						
Learning activity and number of study hours	Learning activity	y Participation in dic classes included in plan		Participation in consultation hours		Self-study		SUM
	Number of study hours	30		4.0		41.0		75
Subject objectives	The main aim of the course is to acknowledge students with the exploratory statistical analysis based on the sample data using Python programming language.							

Learning outcomes	Course outcome	Subject outcome	Method of verification
	[K6_W02] demonstrates advanced preparation in methods and techniques for formulating and solving problems	The student has advanced knowledge of descriptive statistics methods, such as measures of central tendency (mean, median, mode), measures of variability (variance, standard deviation), measures of distribution shape (skewness, kurtosis), as well as techniques for detecting outliers. The student is familiar with and understands advanced Python tools and libraries (such as NumPy, Pandas, Matplotlib, Seaborn, SciPy) used in statistical analysis, problem-solving, and data visualization. The student has advanced knowledge of data exploration techniques, including distribution analysis, correlation analysis between variables, and identification and interpretation of outliers. The student understands the full data analysis cycle – from initial data cleaning, through exploration, to result interpretation and formulation of conclusions, and is familiar with methods for verifying the correctness of conducted analyses.	[SW1] Assessment of factual knowledge [SW3] Assessment of knowledge contained in written work and projects
	[K6_U05] develops innovative solutions for data analysis and processing, using appropriate methods and tools [K6_U07] uses information technologies to improve the	The student is able to perform advanced data analysis using appropriate statistical measures (e.g., mean, median, variance, skewness, kurtosis) and data exploration techniques, such as distribution and correlation analysis, using Python. The student is able to effectively present the results of the analysis using appropriate plots (e.g., histograms, box plots) and tables in Python, utilizing libraries such as Matplotlib and Seaborn, enabling better interpretation of the results. The student uses Python to develop a professional EDA	[SU1] Assessment of task fulfilment [SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools [SU1] Assessment of task fulfilment
	technologies to improve the acquisition, analysis and processing of data in business applications	develop a protessional EDA (exploratory data analysis) report for a given dataset in a specific context, such as business.	tultilment [SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools

Subject contents	Part 1. Introduction to Python & Visual Studio Code.				
	1. Types of data, basic programming skills (Python)				
	2. Types of statistical variables, basic operations, graphs				
	3. Methods of raw data aggregating				
	Part 2. Describing univariate data				
	4. Measures of central tendency, dispersion, assymetry (skewness) and kurtosis, inequality & concentration				
	+ aggregating data (in respect to the type of variable) + graphs				
	Part 3. Describing bivariate data				
	5. Correlation analysis for pair of quantitative variables + graphs				
	6. Rank correlation				
	7. Two-way tables, correlation analysis of qualitative variables + graphs				
	Part 4. Regression analysis				
	8. OLS. Simple regression, multiple linear regression.				
	9. Nonlinear regression.				
	10. Regression analysis for qualitative dependend variable (optional)*.				
	Part 5. Time series				
	11. Describing time series graphs, tables, dynamics & indices.				
	12. Decompositions. Time series models (MA, AR, ARMA).				
Prerequisites and co-requisites	Mathematics, English (intermediate level), basic programming skills - Python.				
Assessment methods	Subject passing criteria	Passing threshold	Percentage of the final grade		
and criteria	Lecture (final exam)	60.0%	50.0%		
	Laboratory - final test, quizzes, self-tests	60.0%	50.0%		
Recommended reading	Basic literature	<ol> <li>Grus, J. (2019). Data science from scratch: First principles with Python (2nd ed.). O'Reilly Media.</li> <li>VanderPlas, J. (2022). Python data science handbook: Essential tools for working with data (2nd ed.). O'Reilly Media.</li> <li>McKinney, W. (2017). Python for data analysis: Data wrangling with pandas, numpy, and ipython (2nd ed.). O'Reilly Media.</li> </ol>			
	Supplementary literature	<ol> <li>O'Reilly, T. (2019). Learning pandas: Getting started with data analysis and visualization in Python. O'Reilly Media.</li> <li>Beyer, H., &amp; Laubacher, R. (2021). Data science for business and decision making: A hands-on guide to data science and machine learning with Python. Packt Publishing.</li> <li>Downey, A. B. (2017). Think stats: Exploratory data analysis in Python (2nd ed.). O'Reilly Media.</li> </ol>			
	eResources addresses	Adresy na platformie eNauczanie:			

Example issues/ example questions/ tasks being completed	<ol> <li>Perform an EDA (Exploratory Data Analysis) on the given dataset. Calculate measures of central tendency (mean, median, mode), measures of variability (variance, standard deviation), and present the distribution of the data using appropriate plots (e.g., histogram, box plot).</li> <li>Explain measures of distribution, such as skewness and kurtosis. What information do these measures provide about the data?</li> <li>Using Python and the Pandas and Matplotlib libraries, perform an analysis of the data from a CSV file containing sales data from an online store. Create plots that help visualize the relationships between variables (e.g., sales vs. day of the week).</li> <li>What techniques for detecting outliers in data are discussed in the context of descriptive statistics? Apply one of the methods to detect outliers in the given dataset. Explain what the correlation result means and discuss how it can influence data interpretation and decision-making in a business context.</li> </ol>
Work placement	Not applicable

Document generated electronically. Does not require a seal or signature.