GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

| | | | |
|---|---|---|---|
| Subject name and code | Descriptive statistics, PG_00045293 | | |
| Field of study | Data Engineering, Data Engineering | | |
| Date of commencement of studies | October 2026 | Academic year of realisation of subject | 2026/2027 |
| Education level | first-cycle studies | Subject group | Obligatory subject group in the field of study<br>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 -> Faculties of Gdańsk University of Technology | | |
| Name and surname of lecturer (lecturers) | Subject supervisor | dr inż. Karol Flisikowski | |
| | Teachers | dr inż. Karol Flisikowski | |

| Lesson types | Lesson type | Lecture | Tutorial | Laboratory | Project | Seminar | SUM |
|---|---|---|---|---|---|---|---|
| | Number of study hours | 15.0 | 0.0 | 15.0 | 0.0 | 0.0 | 30 |
| | E-learning hours included: 18.0 | | | | | | |

| 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 | 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_U07] uses information technologies to improve the acquisition, analysis and processing of data in business applications | The student uses Python to develop a professional EDA (exploratory data analysis) report for a given dataset in a specific context, such as business. | [SU4] Assessment of ability to use methods and tools<br>[SU2] Assessment of ability to analyse information<br>[SU1] Assessment of task fulfilment |
| | [K6_U05] develops innovative solutions for data analysis and processing, using appropriate methods and tools | 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. | [SU4] Assessment of ability to use methods and tools<br>[SU2] Assessment of ability to analyse information<br>[SU1] Assessment of task fulfilment |
| | [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. | [SW3] Assessment of knowledge contained in written work and projects<br>[SW1] Assessment of factual knowledge |

| | |
|---|---|
| Subject contents | Course content – lecture<br>**0. Introduction to Python.** Basic operations, data types, data structures, and the use of Python in statistical analysis.<br><br>PART 1 STRUCTURE ANALYSIS<br><br>Basic statistical concepts (population, variable, statistical study), tabular and graphical presentation of data, types of statistical tables and charts.<br><br>Empirical distribution of a variable. Statistical series (discrete and grouped), data grouping, determination of the number of classes.<br><br>Measures of central tendency: mean, mode, median, quantiles. Methods of calculation and interpretation.<br><br>Measures of variability: variance, standard deviation, coefficient of variation, range, interquartile range. Calculation and interpretation.<br><br>Measures of asymmetry and concentration: skewness coefficient, Gini coefficient. Comparison of distributions.<br><br>PART 2 CORRELATION AND REGRESSION ANALYSIS<br><br>Correlation of quantitative variables: concept of correlation, graphical presentation, Pearson correlation coefficient, interpretation.<br><br>Introduction to statistical inference: hypotheses, significance level, p-value. Correlation of qualitative variables, chi-square test, measures of association (e.g., Cramer's V).<br><br>Rank, partial, and multiple correlation. Spearman and Kendall correlation coefficients.<br><br>Linear regression: least squares method, coefficient of determination, interpretation of the model.<br><br>Multiple and nonlinear regression: dependent and independent variables, functional transformations, interpretation of results.<br><br>PART 3 DYNAMICS ANALYSIS<br><br>Time series and dynamics indices (chain and fixed-base indices), rate of change, interpretation.<br><br>Analysis of changes over time, application of indices, short-term forecasting.<br><br>Trend models, time series decomposition, seasonal analysis, linear and nonlinear models, interpretation.<br><br>Course content – laboratory<br>**0. Introduction to Python and data analysis environment**<br>Working environment setup (Anaconda, VS Code, Jupyter Notebook). Basic data types, structures, and operations. Importing and exploring data (CSV, Excel). Introduction to pandas, numpy, matplotlib, seaborn.<br><br>**1. Analytical reporting (Markdown, Jupyter Notebook)**<br>Creating reports with code, results, and interpretation. Document formatting, visualization, and export to HTML and PDF.<br><br>**2. Teamwork and version control (Git, GitHub)**<br>Fundamentals of Git, local and remote repositories, version history. Team collaboration and cloud-based workflows. |

**PART 1 STRUCTURE ANALYSIS**
Data preparation, cleaning, and visualization. Statistical distributions and grouped data. Measures of central tendency and variability. Skewness, kurtosis, and Gini coefficient.

**PART 2 CORRELATION AND REGRESSION ANALYSIS**
Correlation analysis (Pearson, Spearman, Kendall). Association measures for qualitative variables. Linear, multiple, and nonlinear regression. Model interpretation.

**PART 3 DYNAMICS ANALYSIS**
Time series visualization and trend analysis. Dynamics indices. Trend models and short-term forecasting, including ARIMA models.

**Prerequisites and co-requisites**

Mathematics, English (intermediate level), basic programming skills - Python.

**Assessment methods and criteria**

| Subject passing criteria | Passing threshold | Percentage of the final grade |
|---|---|---|
| Laboratory - final test, quizzes, self-tests | 60.0% | 50.0% |
| Lecture (final exam) | 60.0% | 50.0% |

**Recommended reading**

| | |
|---|---|
| Basic literature | 1. Grus, J. (2019). Data science from scratch: First principles with Python (2nd ed.). O'Reilly Media.<br>2. VanderPlas, J. (2022). Python data science handbook: Essential tools for working with data (2nd ed.). O'Reilly Media.<br>3. McKinney, W. (2017). Python for data analysis: Data wrangling with pandas, numpy, and ipython (2nd ed.). O'Reilly Media. |
| Supplementary literature | 1. O'Reilly, T. (2019). Learning pandas: Getting started with data analysis and visualization in Python. O'Reilly Media.<br>2. Beyer, H., & Laubacher, R. (2021). Data science for business and decision making: A hands-on guide to data science and machine learning with Python. Packt Publishing.<br>3. Downey, A. B. (2017). Think stats: Exploratory data analysis in Python (2nd ed.). O'Reilly Media. |
| eResources addresses | |

**Example issues/ example questions/ tasks being completed**

1. **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).
2. **Explain measures of distribution, such as skewness and kurtosis.** What information do these measures provide about the data?
3. **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).
4. **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 using Python.
5. **Perform a correlation analysis between two variables 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.

**Practical activites within the subject**

Not applicable

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