

Nazwa przedmiotu	Statystyka i eksploracyjna analiza danych
Kod przedmiotu	2400-M1ABSEA
Profil studiów	ogólnoakademicki
Stopień studiów	drugi
Kierunek studiów	Ekonomia
Specjalność (jeśli dotyczy)	Analityka Biznesowa
Forma studiów	studia stacjonarne
Rok studiów	I
Semestr	zimowy
Liczba godzin	30
Forma/typ zajęć	konwersatorium
Punkty ECTS	4
Rodzaj przedmiotu (przedmiot obowiązkowy / kierunkowy do wyboru / fakultatywny / seminarium)	obowiązkowy
Sposób realizacji przedmiotu (stacjonarny / zdalny / kurs internetowy / kurs realizowany metodą blended learning)	stacjonarny
Język wykładowy	polski
Wymagania formalne	brak
Założenia wstępne	brak
Skrócony opis przedmiotu	Celem zajęć jest wprowadzenie studentów w podstawy analizy statystycznej, obejmujące zarówno teorię, jak i praktyczne narzędzia eksploracji danych. Kurs obejmuje zagadnienia z zakresu podstaw programowania, statystyk opisowych, analizy graficznej, miar zależności i asocjacji, wnioskowania statystycznego oraz wybranych testów statystycznych. Studenci poznają w szczególności miary położenia, rozproszenia oraz kształtu rozkładu, a także metody wykrywania obserwacji odstających. Omówione zostaną także testy statystyczne dla różnych rodzajów danych oraz metody wizualizacji, takie jak histogramy, wykresy pudełkowe i wykresy QQ. Wiedza ta pozwoli na efektywne analizowanie i interpretowanie zbiorów danych.
Treści kształcenia dla przedmiotu	Celem zajęć jest zapoznanie studentów z podstawami analizy statystycznej w stopniu wystarczającym do samodzielnego i efektywnego analizowania danych ustrukturyzowanych, interpretowania i wyciągania wniosków z analiz tychże. Na kurs składają się następujące bloki tematyczne: 1. Wprowadzenie do statystyki: <ul style="list-style-type: none"> • znaczenie statystyki i analizy danych w biznesie, • typy struktur danych, • pojęcia populacji i próby, pojęcie zmiennej losowej, rozkład zmiennej losowej, funkcja gęstości, dystrybuanta, • wybrane rozkłady statystyczne: rozkład normalny, t-Studenta, F-Snedecora, Poissona, Gamma, Beta. 2. Wprowadzenie do języka programowania Python: <ul style="list-style-type: none"> • wczytywanie danych z różnych źródeł,

	<ul style="list-style-type: none"> • wybieranie obserwacji i zmiennych, filtrowanie, sortowanie, • mówienie podstawowych bibliotek języka programowania Python: numpy, pandas, statsmodels, • czyszczenie danych, wykrywanie i usuwanie duplikatów, wykrywanie, usuwanie i uzupełnianie braków danych, • przekształcanie struktur danych, łączenie zbiorów danych. <p>3. Miary położenia, rozproszenia i kształtu rozkładu:</p> <ul style="list-style-type: none"> • średnia, mediana, moda, kwartyle, decyle, kwantyle, percentyle, • rozstęp, rozstęp międzykwartyłowy, wariancja, odchylenie standardowe, współczynnik zmienności, • skośność, kurtoza. <p>4. Miary zależności i asocjacji:</p> <ul style="list-style-type: none"> • korelacja Pearsona, Spearmana, Kendalla, współczynnik V Cramera. <p>5. Wizualizacja danych:</p> <ul style="list-style-type: none"> • obserwacje odstające Wykres pudełkowy, obserwacje odstające, IQR, Z-score, zmodyfikowany Z-score, • przekształcenia zmiennych: kategoryzacja zmiennych ciągłych, one-hot-encoding, normalizacja, standaryzacja, winsoryzacja, • znaczenie i cele wizualizacji danych, • złe i dobre praktyki w wizualizacji, • rodzaje wykresów a typ danych (ciągłe, kategoryczne, szeregi czasowe), • przegląd bibliotek języka programowania Python: matplotlib, seaborn, plotly, pandas plotting. <p>6. Testowanie hipotez statystycznych:</p> <ul style="list-style-type: none"> • hipotezy, statystyki testowe, wartości krytyczne, p-value, poziom istotności, błędy I i II rodzaju, • moc testu, moc a wielkość próby. <p>7. Wybrane testy statystyczne:</p> <ul style="list-style-type: none"> • testy dla jednej grupy: testy na normalność rozkładu, test t, test Z, test dwumianowy, test χ^2 dla jednej grupy, • testy dla dwóch i więcej grup niezależnych: test t, test Wilcoxona, test χ^2, test Fishera. ANOVA, test Kruskala-Wallisa, test χ^2. Testy dla dwóch i więcej grup zależnych: test t, test Wilcoxona, test McNemara. ANOVA, test Q-Cochrana.
Przedmiotowe efekty uczenia się	<p>Po ukończeniu przedmiotu, student:</p> <p>W ZAKRESIE WIEDZY:</p> <ul style="list-style-type: none"> • zna i rozumie podstawowe pojęcia statystyczne, • zna i rozumie specyfikę wybranych rozkładów statystycznych, • zna i rozumie narzędzia pozwalające na przeprowadzanie wnioskowania statystycznego. <p>W ZAKRESIE UMIEJĘTNOŚCI:</p> <ul style="list-style-type: none"> • potrafi wczytywać, oczyszczać i przygotowywać dane do analiz, • potrafi wyciągać wnioski w oparciu o miary położenia, rozproszenia, kształtu rozkładu, miary zależności i asocjacji, • potrafi wizualizować dane w sposób jasny i zrozumiały, jak również wyciągać wnioski na podstawie takich wizualizacji, • potrafi przeprowadzać i interpretować testy statystyczne. <p>W ZAKRESIE KOMPETENCJI:</p> <ul style="list-style-type: none"> • jest gotów interpretować wyniki uzyskane w toku analiz statystycznych, co wynika ze zaznajomienia z najważniejszymi metodami statystycznymi i dobrymi praktykami w ich wykorzystywaniu,

	<ul style="list-style-type: none"> wykazuje się zdolnością do krytycznej analizy uzyskanych wyników, co wynika z faktu omówienia przykładów zastosowań wspomnianych metod w trakcie zajęć.
Powiązanie efektów przedmiotowych z efektami kierunkowymi/specjalnościowym i (oznaczonymi kodami z programu nauczania)	specjalność Analityka Biznesowa: S_W02 S_W04 S_U01 S_U02 S_K01
Nakład pracy studenta	<p>Szacunkowy nakład pracy studenta: 4ECTS x 25h = 100h</p> <p>(K) - godziny kontaktowe (S) - godziny pracy samodzielnej</p> <p>zajęcia: 30h (K) 0h (S)</p> <p>konsultacje: 8h (K) 0h (S)</p> <p>egzamin: 2h (K) 0h (S)</p> <p>przygotowanie do zajęć: 0h (K) 30h (S)</p> <p>przygotowanie zaliczenia: 0h (K) 30h (S)</p> <p>Razem: 40h (K) + 60h (S) = 100h</p>
Metody i kryteria oceniania	<p>Uzyskanie zaliczenia przedmiotu wymaga:</p> <ol style="list-style-type: none"> uzyskania co najmniej połowy punktów z egzaminu pisemnego; uzyskania co najmniej połowy punktów z projektu przygotowanego w domu; uzyskania co najmniej połowy punktów z prezentacji wyżej wspomnianego projektu; uzyskania co najmniej połowy z łącznej sumy punktów uzyskanych ze wszystkich wymienionych powyżej elementów. <p>5. Skala ocen:</p> <p>[0%-50%) - ndst</p> <p>[50%-60%) - dst</p> <p>[60%-70%) - dst+</p> <p>[70%-80%) - db</p> <p>[80%-90%) - db+</p> <p>[90%-100%) - bdb</p>
Literatura	<ul style="list-style-type: none"> Downey, A. B. (2014). Think stats: Exploratory data analysis in Python (Version 2.0.27). Knafllic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons, Inc. Mangiafico, S. S. (2015). An R companion for the Handbook of Biological Statistics (Version 1.3.9, revised 2023) [Online handbook]. Rutgers Cooperative Extension. Retrieved from https://rcompanion.org/handbook. Molin, S. (2021). Hands-on data analysis with pandas: Efficiently perform data collection, wrangling, analysis, and visualization using Python (2nd ed.). Packt Publishing. Pant, D., & Mukhiya, S. K. (2025). Statistics for Data Scientists and Analysts: Statistical approach to data-driven decision making using Python. BPB Publications.
Metody dydaktyczne	<p>W ramach konwersatorium wykorzystywane są następujące metody dydaktyczne:</p> <ul style="list-style-type: none"> -prezentacja wykładowcy, -prezentacja studencka, -analiza źródeł danych,

	-analiza przypadku, -metoda projektu, -ćwiczenia rachunkowe i analityczne, -metoda projektów zespołowych, -prezentacja multimedialna, -analiza danych w programach komputerowych.
Oprogramowanie	Visual Studio Code, wtyczki: Jupyter, Python