

Karta przedmiotu

Nazwa i kod przedmiotu	Descriptive statistics, PG_00045293						
Kierunek studiów	Inżynieria danych						
Data rozpoczęcia studiów	październik 2024 r.		Rok akademicki realizacji przedmiotu		2024/2025		
Poziom kształcenia	I stopnia - inżynierskie		Grupa zajęć		Grupa zajęć obowiązkowych z zakresu kierunku studiów Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	stacjonarne		Sposób realizacji		mieszane (blended-learning)		
Rok studiów	1		Język wykładowy		angielski		
Semestr studiów	2		Liczba punktów ECTS		3.0		
Profil kształcenia	ogólnoakademicki		Forma zaliczenia		zaliczenie		
Jednostka prowadząca	Wydział Zarządzania i Ekonomii -> Katedra Statystyki i Ekonometrii						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr inż. Karol Flisikowski				
	Prowadzący zajęcia z przedmiotu		dr inż. Karol Flisikowski				
Formy zajęć i metody nauczania	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	15.0	0.0	0.0	30
	W tym liczba godzin zajęć na odległość: 18.0						
Aktywność studenta i liczba godzin pracy	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów		Udział w konsultacjach		Praca własna studenta	RAZEM
	Liczba godzin pracy studenta	30		4.0		41.0	75
Cel przedmiotu	Zdobycie umiejętności statystycznej analizy danych w próbie przy wykorzystaniu języka programowania Python.						

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[K6_U07] wykorzystuje technologie informatyczne w celu usprawnienia pozyskiwania, analizy i przetwarzania danych w zastosowaniach biznesowych	Student wykorzystuje język Python w celu opracowania profesjonalnego raportu EDA (exploratory data analysis) dla zadanego zbioru danych w kontekście np. biznesowym.	[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania
	[K6_U05] projektuje innowacyjne rozwiązania analizy i przetwarzania danych, wykorzystując odpowiednie metody i narzędzia	Student potrafi przeprowadzać zaawansowaną analizę danych, wykorzystując odpowiednie miary statystyczne (np. średnia, mediana, wariancja, współczynnik skośności, kurtoza) oraz techniki eksploracji danych, takie jak analiza rozkładu i korelacji, przy użyciu Pythona. Student potrafi efektywnie prezentować wyniki analizy za pomocą odpowiednich wykresów (np. histogramów, wykresów pudełkowych) i tabel w Pythonie, wykorzystując biblioteki takie jak Matplotlib i Seaborn, umożliwiając lepszą interpretację wyników.	[SU4] Ocena umiejętności korzystania z metod i narzędzi [SU2] Ocena umiejętności analizy informacji [SU1] Ocena realizacji zadania
	[K6_W02] demonstruje zaawansowane przygotowanie w zakresie metod oraz technik formułowania i rozwiązywania problemów	Student posiada zaawansowaną wiedzę na temat metod statystyki opisowej, takich jak miary tendencji centralnej (średnia, mediana, tryb), miary zmienności (wariancja, odchylenie standardowe), miary kształtu rozkładu (skośność, kurtoza), a także technik wykrywania wartości odstających. Student zna i rozumie zaawansowane narzędzia oraz biblioteki Pythona (takie jak NumPy, Pandas, Matplotlib, Seaborn, SciPy) wykorzystywane w analizie statystycznej, rozwiązywaniu problemów oraz wizualizacji danych. Student ma zaawansowaną wiedzę na temat technik eksploracji danych, w tym analizy rozkładów, analizy korelacji między zmiennymi oraz identyfikacji i interpretacji wartości odstających. Student rozumie pełen cykl analizy danych – od wstępnego oczyszczania danych, przez eksplorację, po interpretację wyników i formułowanie wniosków, oraz zna metody weryfikacji poprawności przeprowadzonych analiz.	[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym [SW1] Ocena wiedzy faktograficznej

Treści przedmiotu	<p>0. Wprowadzenie do języka Python. Podstawowe operacje w Pythonie.</p> <p>CZĘŚĆ 1 - ANALIZA STRUKTURY</p> <p>1.Podstawowe pojęcia (statystyka, zbiorowość statystyczna, cecha), badanie statystyczne etapy, prezentacja graficzna i tabelaryczna rozkładu cechy rodzaje tabel i wykresów statystycznych, przykłady.</p> <p>2.Empiryczny rozkład cechy. Szeregi strukturalne i rozdzielcze (punktowe i przedziałowe), ustalanie liczby klas.</p> <p>3.Miary położenia klasyczne i pozycyjne. Średnia, dominanta, mediana, kwantyle. Sposoby liczenia, przykłady.</p> <p>4.Miary zmienności (dyspersji) klasyczne i pozycyjne. Wariancja, odchylenie standardowe, współczynnik zmienności, rozstęp, odchylenie ćwiartkowe. Sposoby liczenia, przykłady.</p> <p>5.Miary asymetrii i koncentracji. Rodzaje rozkładów, krzywa Lorenza, współczynnik Giniego. Sposoby liczenia, przykłady. Porównywanie rozkładów względny wskaźnik podobieństwa struktur.</p> <p>CZĘŚĆ 2 ANALIZA KORELACJI I REGRESJI</p> <p>1.Analiza korelacji dla cech ilościowych. Pojęcie korelacji, statystyczna prezentacja korelacji, związki korelacyjne, miary korelacji, współczynnik korelacji liniowej Pearsona.</p> <p>2.Wprowadzenie do wnioskowania statystycznego (rodzaje hipotez, poziom istotności, prawdopodobieństwo testowe) . Analiza korelacji dla cech jakościowych. Test zgodności Pearsona, poprawka Yatesa, współczynnik Yula, współczynniki zbieżności korelacyjnej T Czuprowa, V Cramera, C Pearsona.</p> <p>3.Korelacja cząstkowa i wieloraka. Współczynniki korelacji rang Spearmana, Kendalla</p> <p>4.Podstawy analizy regresji. Regresja liniowa, klasyczna metoda najmniejszych kwadratów, współczynnik determinacji i zbieżności, indeks korelacji, interpretacja siły i kierunku oddziaływania zmiennych.</p> <p>5.Regresja wieloraka i nieliniowa. Zmienna zależna i zmienne niezależne, transformacje nieliniowych funkcji regresji potęgowa, wykładnicza, hiperboliczna, wielomiany, interpretacje.</p> <p>CZĘŚĆ 3 ANALIZA DYNAMIKI</p> <p>1.Szereg czasowy. Indeksy dynamiki - łańcuchowe i jednopodstawowe, transformacje, średnie tempo zmian, interpretacja, sposoby liczenia, przykłady.</p> <p>2.Zastosowanie indeksów dynamiki analiza zmian, urealnianie, prognozowanie krótkookresowe.</p> <p>3.Modele trendu dekompozycja szeregu czasowego, analiza wahań sezonowych, liniowe i nieliniowe modele trendu, interpretacja, przykłady.</p>		
Wymagania wstępne i dodatkowe	Matematyka, język angielski (poziom średniozaawansowany), podstawowe umiejętności programistyczne - Python.		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	Laboratorium - kolokwium, quizy, self-testy	60.0%	50.0%
	Wykład - kolokwium końcowe (test)	60.0%	50.0%

Zalecana lista lektur	Podstawowa lista lektur	<ol style="list-style-type: none"> 1. Grus, J. (2019). Data science from scratch: First principles with Python (2nd ed.). O'Reilly Media. 2. VanderPlas, J. (2022). Python data science handbook: Essential tools for working with data (2nd ed.). O'Reilly Media. 3. McKinney, W. (2017). Python for data analysis: Data wrangling with pandas, numpy, and ipython (2nd ed.). O'Reilly Media.
	Uzupełniająca lista lektur	<ol style="list-style-type: none"> 1. O'Reilly, T. (2019). Learning pandas: Getting started with data analysis and visualization in Python. O'Reilly Media. 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. 3. Downey, A. B. (2017). Think stats: Exploratory data analysis in Python (2nd ed.). O'Reilly Media.
	Adresy eZasobów	Uzupełniające Adresy na platformie eNauczanie: Descriptive Statistics 2025 - Moodle ID: 42784 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=42784
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<ol style="list-style-type: none"> 1. Przeprowadź analizę EDA na zadanym zbiorze danych. Oblicz miary tendencji centralnej (średnia, mediana, tryb), miary zmienności (wariancja, odchylenie standardowe), oraz przedstaw rozkład danych za pomocą odpowiednich wykresów (np. histogram, wykres pudełkowy). 2. Wyjaśnij miary rozkładu, takie jak skośność i kurtoza. Jakich informacji dostarczają te miary o danych? 3. Z użyciem Pythona i bibliotek Pandas oraz Matplotlib, wykonaj analizę danych z pliku CSV zawierającego dane o sprzedaży w sklepie internetowym. Stwórz wykresy, które pomogą zwizualizować zależności między poszczególnymi zmiennymi (np. sprzedaż a dzień tygodnia). 4. Jakie techniki wykrywania wartości odstających w danych opisujesz w kontekście statystyki opisowej? Zastosuj jedną z metod do wykrycia wartości odstających w zadanym zbiorze danych przy użyciu Pythona. 5. Przeprowadź analizę korelacji między dwoma zmiennymi w zadanym zbiorze danych. Wyjaśnij, co oznacza wynik korelacji, i omów, jak może to wpłynąć na interpretację danych oraz podejmowanie decyzji w kontekście biznesowym. 	
Praktyki zawodowe w ramach przedmiotu	Nie dotyczy	

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.