



Karta przedmiotu

Nazwa i kod przedmiotu	Descriptive statistics, PG_00045293						
Kierunek studiów	Inżynieria danych, Inżynieria danych						
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu			2026/2027		
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	Wydziały Politechniki Gdańskiej -> Wydział Zarządzania i Ekonomii -> Katedra Statystyki i Ekonometrii						
Imię i nazwisko wykładowcy (wykładowców)	Od odpowiedzialny za przedmiot	dr inż. Karol Flisikowski					
	Prowadzący zajęcia z przedmiotu	dr inż. Karol Flisikowski					
Formy zajęć	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

0. Wprowadzenie do języka Python. Podstawowe operacje, typy danych, struktury danych oraz wykorzystanie Python w analizie statystycznej.

CZĘŚĆ 1 ANALIZA STRUKTURY

1. Podstawowe pojęcia statystyczne (zbiorowość, cecha, badanie statystyczne), prezentacja tabelaryczna i graficzna danych, rodzaje tabel i wykresów.
2. Empiryczny rozkład cechy. Szeregi statystyczne (punktowe i przedziałowe), grupowanie danych, ustalanie liczby klas.
3. Miary położenia: średnia, dominanta, mediana, kwantyle. Metody obliczania i interpretacja.
4. Miary zmienności: wariancja, odchylenie standardowe, współczynnik zmienności, rozstęp, odchylenie ćwiartkowe. Obliczanie i interpretacja.
5. Miary asymetrii i koncentracji: współczynnik asymetrii, współczynnik Giniego. Porównywanie rozkładów.

CZĘŚĆ 2 ANALIZA KORELACJI I REGRESJI

1. Korelacja zmiennych ilościowych: pojęcie korelacji, prezentacja graficzna, współczynnik korelacji Pearsona, interpretacja.
2. Wprowadzenie do wnioskowania statystycznego: hipotezy, poziom istotności, wartość p. Korelacja zmiennych jakościowych, test chi-kwadrat, współczynniki zależności (np. V Cramera).
3. Korelacja rangowa, cząstkowa i wieloraka. Współczynniki Spearmana i Kendalla.
4. Regresja liniowa: metoda najmniejszych kwadratów, współczynnik determinacji, interpretacja modelu.
5. Regresja wieloraka i nieliniowa: zmienne zależne i niezależne, transformacje funkcji, interpretacja wyników.

CZĘŚĆ 3 ANALIZA DYNAMIKI

1. Szeregi czasowe i indeksy dynamiki (łańcuchowe i jednopodstawowe), tempo zmian, interpretacja.
2. Analiza zmian w czasie, zastosowanie indeksów, prognozowanie krótkookresowe.
3. Modele trendu, dekompozycja szeregu czasowego, analiza sezonowości, modele liniowe i nieliniowe, interpretacja.

Treści przedmiotu - laboratoria

0. Wprowadzenie do języka Python i środowiska analizy danych

Konfiguracja środowiska pracy (Anaconda, VS Code, Jupyter Notebook). Podstawowe typy i struktury danych, operacje oraz wczytywanie i eksploracja danych (CSV, Excel). Wprowadzenie do bibliotek pandas, numpy, matplotlib, seaborn.

1. Tworzenie raportów analitycznych (Markdown, Jupyter Notebook)

Tworzenie raportów zawierających kod, wyniki i interpretację. Formatowanie dokumentów, wizualizacja wyników, dokumentowanie analizy. Eksport raportów do HTML i PDF.

2. Praca zespołowa i zarządzanie wersjami Git i GitHub

Podstawy Git, repozytoria lokalne i zdalne, historia zmian. Współpraca zespołowa, udostępnianie kodu i raportów, wprowadzenie do pracy w chmurze.

CZĘŚĆ 1 ANALIZA STRUKTURY

3. Przygotowanie i prezentacja danych

Wczytywanie, czyszczenie i przygotowanie danych, tworzenie tabel liczebności, wizualizacja i interpretacja danych.

4. Rozkład cechy i szeregi statystyczne

Grupowanie danych, szeregi punktowe i przedziałowe, wizualizacja i interpretacja rozkładu.

5. Miary położenia i zmienności

Średnia, mediana, dominanta, kwantyle, wariancja, odchylenie standardowe, współczynnik zmienności, interpretacja wyników.

6. Miary asymetrii i koncentracji

Asymetria, kurtoza, współczynnik Giniego, interpretacja rozkładu.

CZĘŚĆ 2 ANALIZA KORELACJI I REGRESJI

7. Analiza korelacji

Wykresy rozrzutu, współczynniki Pearsona, Spearmana i Kendalla, interpretacja zależności.

8. Korelacja zmiennych jakościowych

Tabele kontyngencji, miary zależności (np. V Cramera), interpretacja.

9. Regresja liniowa i wieloraka

Budowa modeli, współczynnik determinacji, interpretacja parametrów.

10. Modele nieliniowe

Transformacje zmiennych i interpretacja modeli.

CZĘŚĆ 3 ANALIZA DYNAMIKI

11. Szeregi czasowe i indeksy dynamiki

Wizualizacja danych czasowych, analiza trendów i zmian.

12. Modele trendu i prognozowanie

Modele liniowe i nieliniowe, prognozowanie krótkookresowe, modele ARIMA.

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

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"> O'Reilly, T. (2019). Learning pandas: Getting started with data analysis and visualization in Python. O'Reilly Media. 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. Downey, A. B. (2017). Think stats: Exploratory data analysis in Python (2nd ed.). O'Reilly Media.
	Adresy eZasobów	
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania	<ol style="list-style-type: none"> 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). Wyjaśnij miary rozkładu, takie jak skośność i kurtoza. Jakich informacji dostarczają te miary o danych? 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). 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. 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. 	
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy	

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.