

Python w Analizie Danych – Wprowadzenie

Python zdobywa ogromną popularność w świecie Data Science dzięki swojej wszechstronności i prostocie. Jest niezastąpiony w analizie danych, automatyzacji procesów oraz wizualizacji wyników. Rynek pracy dynamicznie rośnie, zapotrzebowanie na analityków danych jest coraz większe.

W tej prezentacji omówimy kluczowe etapy analizy danych w Pythonie, najważniejsze używane biblioteki, podstawowe komendy oraz praktyczne zastosowania analizy danych w różnych branżach.

 **by r janiak**



Etapy Analizy Danych w Pythonie

1

Cel i pytanie

Definiowanie celu biznesowego i sformułowanie pytania badawczego.

2

Eksploracja danych

Zbieranie danych i wstępna analiza ich struktury oraz jakości.

3

Czyszczenie

Usuwanie błędów i braków, formatowanie danych do analizy.

4

Analiza

Wykorzystanie statystyki, modeli matematycznych i algorytmów.

5

Wizualizacja

Prezentowanie wyników za pomocą wykresów i raportów.



Kluczowe Biblioteki w Pythonie – Przegląd

Pandas

Manipulacja i analiza danych tabelarycznych, szerokie możliwości importu i agregacji.

NumPy

Szybkie obliczenia na tablicach wielowymiarowych, podstawa dla wielu innych bibliotek.

Matplotlib & Seaborn

Tworzenie estetycznych wykresów i wizualizacji, od prostych po złożone analizy graficzne.

Pandas – Najważniejsze Informacje

Pandas operuje na strukturze DataFrame, która umożliwia efektywną pracę z tabelarycznymi danymi. Biblioteka umożliwia łatwe wczytywanie i eksport danych z różnych formatów, takich jak CSV, Excel czy SQL. Obsługuje również brakujące wartości i umożliwia ich wygodne przetwarzanie.

Biblioteka oferuje potężne narzędzia do filtrowania, grupowania i agregowania danych. Jest uznawana za najczęściej stosowany pakiet w analizie danych w Pythonie, dzięki czytelności i szerokim możliwościom manipulacji danymi.

Pandas – Przykładowe Podstawowe Komendy

Import biblioteki

import pandas as pd – umożliwia korzystanie z funkcji Pandas.

Podgląd i inspekcja

df.head() – pokazuje pierwsze wiersze danych, **df.info()** – wyświetla metadane tabeli, **df.describe()** – zwraca statystyki opisowe.

Wczytywanie danych

df = pd.read_csv("plik.csv") – import danych z pliku CSV do DataFrame.

Agregacja

df.groupby("kolumna").mean() – grupowanie po wybranej kolumnie i obliczanie średniej.

NumPy – Najważniejsze Informacje

NumPy to fundament obliczeń numerycznych w Pythonie. Umożliwia szybkie operacje na tablicach wielowymiarowych, które są znacznie bardziej efektywne niż tradycyjne listy Pythona. Obsługuje operacje wektorowe i macierzowe, co jest niezbędne do analizy danych i uczenia maszynowego.

Biblioteka ta stanowi techniczną podstawę dla Pandas, zapewniając szybkie przeliczanie danych numerycznych w tabelach oraz macierzach.



NumPy – Przykładowe Podstawowe Komendy

Import biblioteki

import numpy as np – dostęp do funkcji NumPy.

Statystyki

np.mean(a) i **np.std(a)** – obliczanie średniej arytmetycznej i odchylenia standardowego.

Tworzenie tablicy

a = np.array([1, 2, 3]) – definiowanie jednowymiarowej tablicy.

Zmiana wymiarów

np.reshape(a, (3, 1)) – transformacja tablicy na inną strukturę wymiarową.

Matplotlib i Seaborn – Wizualizacja Danych

Matplotlib pozwala na tworzenie różnorodnych wykresów, od prostych liniowych po złożone wykresy słupkowe i rozrzutu. Komenda **plt.plot(x, y)** umożliwia szybkie wyświetlenie danych.

Seaborn rozszerza możliwości graficzne, oferując estetyczne wykresy korelacji, heatmapy oraz zaawansowane style wizualne, ułatwiające interpretację danych.

Obie biblioteki współpracują bezproblemowo z Pandas i NumPy, co pozwala efektywnie eksplorować i prezentować dane w atrakcyjnej formie wizualnej.

Przykładowe Zastosowania Analizy Danych Pythonem



Analiza biznesowa

Wsparcie
decyzji
strategicznych
i operacyjnych
poprzez
raportowanie i
monitoring.



Wykrywan ie anomalii

Identyfikacja
niestandardow
ych zachowań i
błędów w
danych.



Prognozo wanie

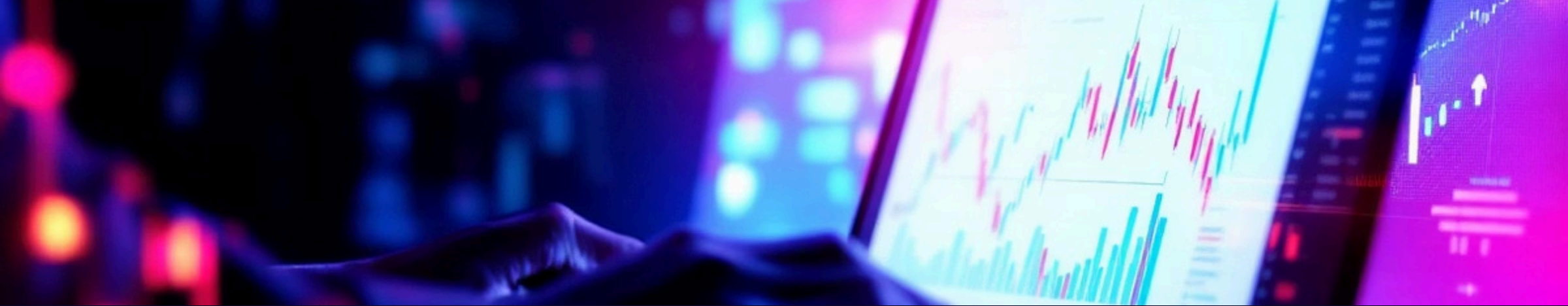
Estymacje
sprzedaży,
popytu i
trendów
rynkowych.



Badania naukowe

Obróbka i
analiza dużych
zbiorów
danych
eksperymental
nych i
obserwacyjnyc
h.





Podsumowanie i Gdzie Wykorzystać Analizę Danych

Python to podstawowe narzędzie dla współczesnych analityków danych. Kluczowe biblioteki takie jak Pandas, NumPy oraz Matplotlib i Seaborn, tworzą kompletny ekosystem do zbierania, opracowywania i prezentowania danych.

Znajomość tych narzędzi jest ceniona w finansach, marketingu, nauce i przemyśle, a rozwijanie umiejętności analitycznych podnosi konkurencyjność na rynku pracy i otwiera szerokie możliwości zawodowe.