

Tytuł Projektu: Analiza Danych z E-commerce Kawiarni z Web Scrapingiem i EDA

Cel:

Projekt ma na celu naukę technik web scrapingu oraz eksploracyjnej analizy danych (EDA) poprzez zebranie danych ze stron internetowych kawiarni online, ich przetworzenie, wizualizację oraz porównanie kluczowych cech różnych kawiarni. Realizacja projektu umożliwi rozwinięcie umiejętności w zakresie web scrapingu, czyszczenia danych, wizualizacji danych oraz EDA.

Wymagania Projektowe:

1. Zbieranie Danych przez Web Scraping (10 punktów)

- **Wybierz 2-3 sklepy internetowe z kawą** (np. Starbucks, Dunkin', lokalne kawiarnie online).
- Zidentyfikuj kluczowe dane do zebrania, takie jak:
 - Produkty (nazwa, rodzaj napoju, rozmiar, cena).
 - Oceny lub recenzje (jeśli dostępne).
 - Informacje o wartościach odżywczych (kalorie, zawartość cukru, poziom kofeiny).
- Użyj bibliotek Pythona (np. BeautifulSoup lub Scrapy) do pozyskania i ekstrakcji danych.
- Przechowaj zebrane dane w uporządkowanej formie (CSV lub DataFrame).

2. Czyszczenie i Przygotowanie Danych (6 punktów)

- Wykonaj podstawowe czyszczenie danych, aby były strukturalne, spójne i kompletne. Zadania mogą obejmować:
 - Obsługę brakujących lub niespójnych wartości.
 - Standaryzację jednostek miary (np. uncje na mililitry, kalorie).
 - Konwersję cen na wspólną walutę, jeśli to konieczne.

3. Eksploracyjna Analiza Danych (EDA) (6 punktów)

- Przeprowadź analizę danych, aby odkryć wzorce i uzyskać wgląd.
- Sugerowane analizy obejmują:
 - Porównanie cen w różnych typach i rozmiarach napojów.
 - Rozkład typów napojów i popularność (jeśli dostępne są oceny).
 - Analiza wartości odżywczych napojów (kalorie, cukier, poziom kofeiny).

4. Wizualizacja Danych (5 punktów)

- Użyj bibliotek Pythona, takich jak Matplotlib lub Seaborn, do wizualizacji danych i wyników.
- Stwórz co najmniej **3 wizualizacje**, które zilustrują różnice między kawiarniami. Sugerowane wizualizacje:
 - Wykres słupkowy porównujący średnie ceny popularnych typów napojów (np. latte, espresso).
 - Wykres kołowy lub histogram przedstawiający rozkład typów napojów w każdej kawiarni.
 - Heatmapa porównująca wartości odżywcze (np. kalorie vs. cukier) dla napojów z różnych kawiarni.

5. Porównanie i Wnioski (3 punktów)

- Podsumuj swoje odkrycia, podkreślając kluczowe różnice między kawiarniami.
 - Porównaj takie metryki jak cena, wartości odżywcze oraz różnorodność oferty.
 - Zakończ analizę wnioskami lub rekomendacjami, które wynikają z analizy (np. która kawiarnia oferuje najlepszy stosunek jakości do ceny lub najzdrowsze opcje).
-

Wymagane Materiały:

1. **Pliki Kodów** – Skrypty Pythona lub Jupyter Notebook zawierające kod do scrapingu, czyszczenia oraz analizy.
2. **Pliki Danych** – Pliki CSV lub DataFrame z oczyszczonymi, uporządkowanymi danymi.
3. **Wizualizacje** – Co najmniej 3 wykresy zapisane jako obrazy lub osadzone w Notebooku.
4. **Raport** – Krótkie podsumowanie (100-200 słów) opisujące proces, analizę oraz kluczowe spostrzeżenia. (Można skomentować w Jupyter notebook)

Opcjonalne: Za dodatkowe punkty, można przeprowadzić analizę sentymentu na podstawie opinii klientów, jeśli są dostępne.

Praktykowane Umiejętności:

- Web Scraping w Pythonie
- Czyszczenie i Strukturyzacja Danych
- Eksploracyjna Analiza Danych (EDA)
- Wizualizacja Danych ma menu kontekstowe