Oczyszczanie danych sprzedaży kawiarni (Python)

1. Cel projektu

Celem projektu było przygotowanie surowego pliku CSV z transakcjami kawiarni do dalszej analizy. Projekt miał pokazać umiejętność oczyszczania danych, walidacji i przygotowania ich do analizy biznesowej.

1. Kroki procesu oczyszczania danych

-Importowanie potrzebnych bibliotek

-Użycie biblioteki pandas do wczytania pliku CSV

-Utworzenie kopii zapasowej pliku

-Zapoznanie sie z podstawowymi informacjami o danych

-Zamienianie błędów w kolumnach jako pustych wartości

-Konwersja kolumn na odpowiedni typ

-Sprawdzanie ile wartości jest pustych

-Konwersja błędnych wartości na NaN

-Tworzenie tymczasowych kolumn, by później podmienić je w zwykłych kolumnach

-Uzupełnianie wartości, które zostały obliczone w tymczasowych kolumnach

-Usuwanie tymczasowych kolumn

-Sprawdzenie, czy istnieją duplikaty

-Wykrywanie potencjalnych outlierów

-Uzupełnianie brakujących wartości

-Zapisanie pliku .csv

3. Umiejętności pokazane w projekcie

* **Python & pandas:** wczytywanie danych, manipulacja DataFrame, konwersja typów.
* **Data Cleaning:** obsługa braków, duplikatów, outlierów, normalizacja danych tekstowych.
* **Data Validation:** walidacja kolumn liczbowych, sprawdzanie spójności wartości (Total Spent).
* **Problem Solving:** analiza i rozwiązywanie typowych problemów w danych sprzedażowych.
* **Portfolio-ready Skills:** przygotowanie danych do dalszych analiz i wizualizacji w BI.

4. Wynik

Oczyszczony i spójny zestaw danych gotowy do analizy: np. raportów sprzedaży według produktów, metod płatności i lokalizacji.