

Zadanie

Dokonaj następującej eksploracji danych z pliku zakupy-online.xlsx:

1. Zapoznaj się z danymi, zwróć uwagę na powtarzające się rekordy w dwóch arkuszach
2. Wyczyść dane z transakcji niezrealizowanych (nieopłaconych) i zwrotów
3. Dokonaj opisu danych (opowiedz je) poprzez wyznaczenie podstawowych statystyk, na przykład: z jakiego okresu pochodzą dane (łącznie), z ilu (jakich) krajów pochodzą klienci, jaka była zmienność w czasie liczby transakcji, wysokości transakcji, jakie artykuły były najpopularniejsze (sprzedano najwięcej sztuk), na jakich artykułach najwięcej zarobiono (najwyższa łączna cena sprzedaży)
4. Dokonaj analizy koszykowej wyznaczając:
 - a. Zbiory częste (min_sup = częstotliwość 20% najczęstszych artykułów w zbiorze wszystkich transakcji)
 - b. Reguły asocjacyjne (min_sup = jak wyżej, $\text{min_conf} \geq 0.7$)

Jeśli trzeba, dokonaj transformacji danych do wymaganego formatu.

- W tym celu wykorzystaj elementy bibliotek Pythona, np. biblioteki Pythona przydatne w analizie koszykowej https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/
 - openpyxl do wczytania danych z pliku xlsx,
 - mlxtend.frequent_patterns, algorytm apriori,
 - mlxtend.frequent_patterns, algorytm association_rules
-
- apyori: To biblioteka do implementacji algorytmu Apriori. Udostępnia funkcje do odczytu i manipulowania danymi transakcyjnymi, a także do generowania reguł asocjacyjnych i oceny ich jakości.
 - mlxtend: To biblioteka do implementacji różnych algorytmów uczenia maszynowego i narzędzi, w tym indukcji reguł asocjacyjnych. Udostępnia funkcje do odczytu i manipulowania danymi transakcyjnymi, a także do generowania reguł asocjacyjnych i oceny ich jakości.
 - PyCaret: PyCaret to niskokodowa (low-code) biblioteka uczenia maszynowego typu open source do automatyzacji procesów uczenia maszynowego. Udostępnia opakowanie na mlxtend w celu łatwej implementacji algorytmu Apriori.