



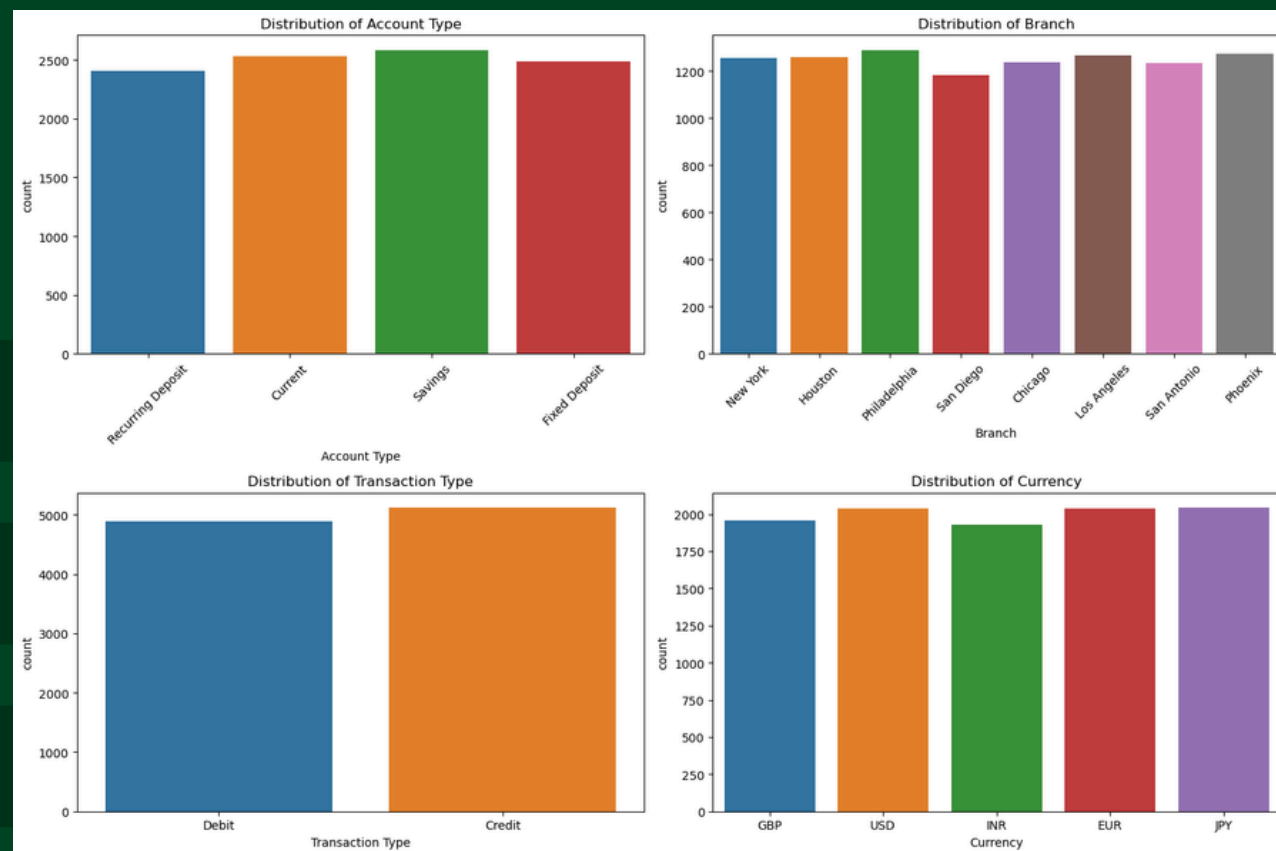
BANK CLIENTS CLUSTERING

WIKTORIA KAWA, WOJCIECH ŁUKASZYK

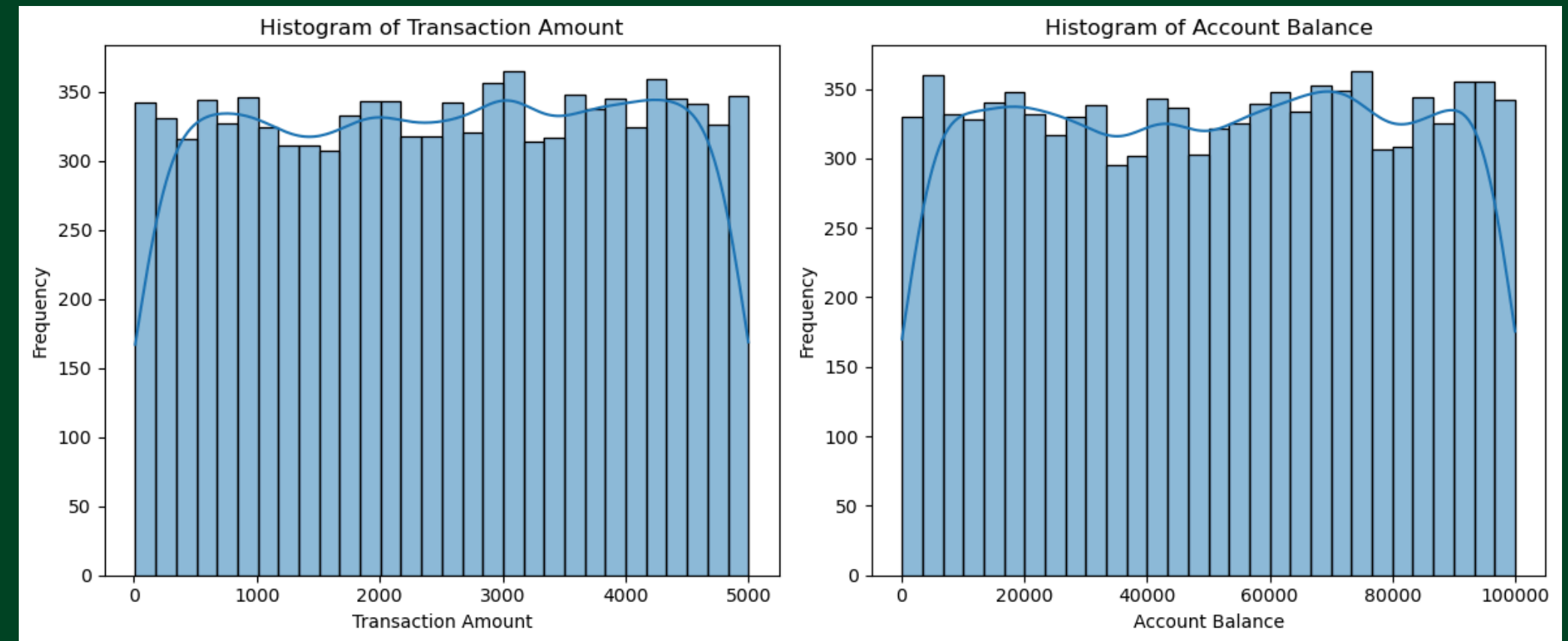
PRZYPOMNIENIE PROBLEMU:

Segmentacja klientów banku na podstawie ich zachowań transakcyjnych i typów kont w celu personalizacji ofert marketingowych i zwiększenia retencji klientów

PROBLEMY Z RAMKĄ DANYCH



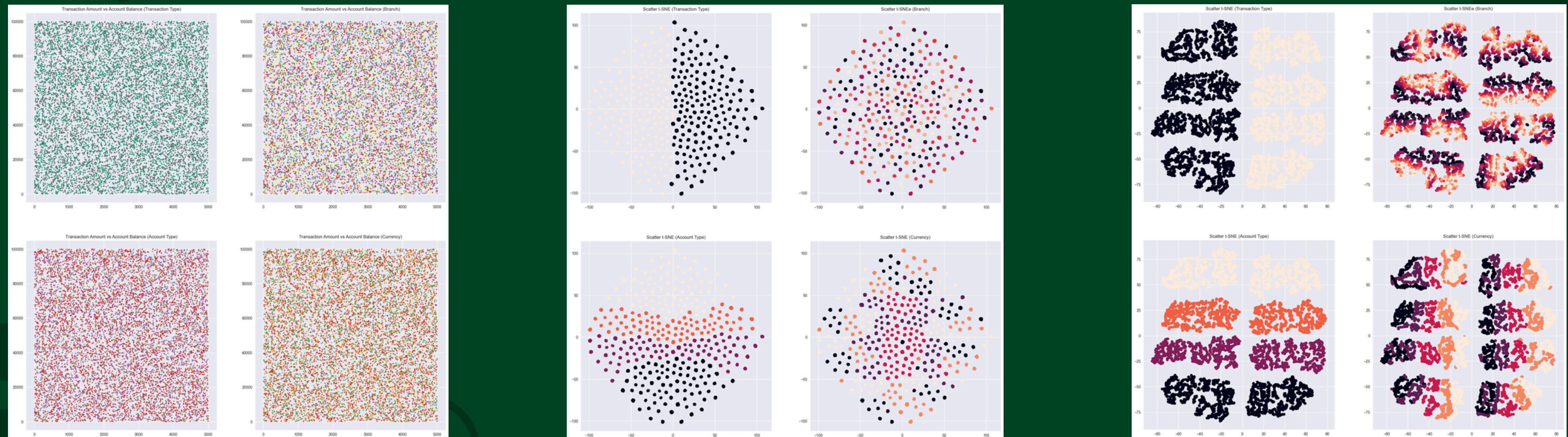
count zmiennych kategorycznych



rozkład zmiennych ciągłych (numerycznych)

PRZYPOMNIENIE PROBLEMU:

Dane w ramce są wygenerowane, nie są też zgodne z opisem autora. Wszystkie dane są wygenerowane i równomiernie rozłożone. Nie ma też między nimi żadnej zależności.

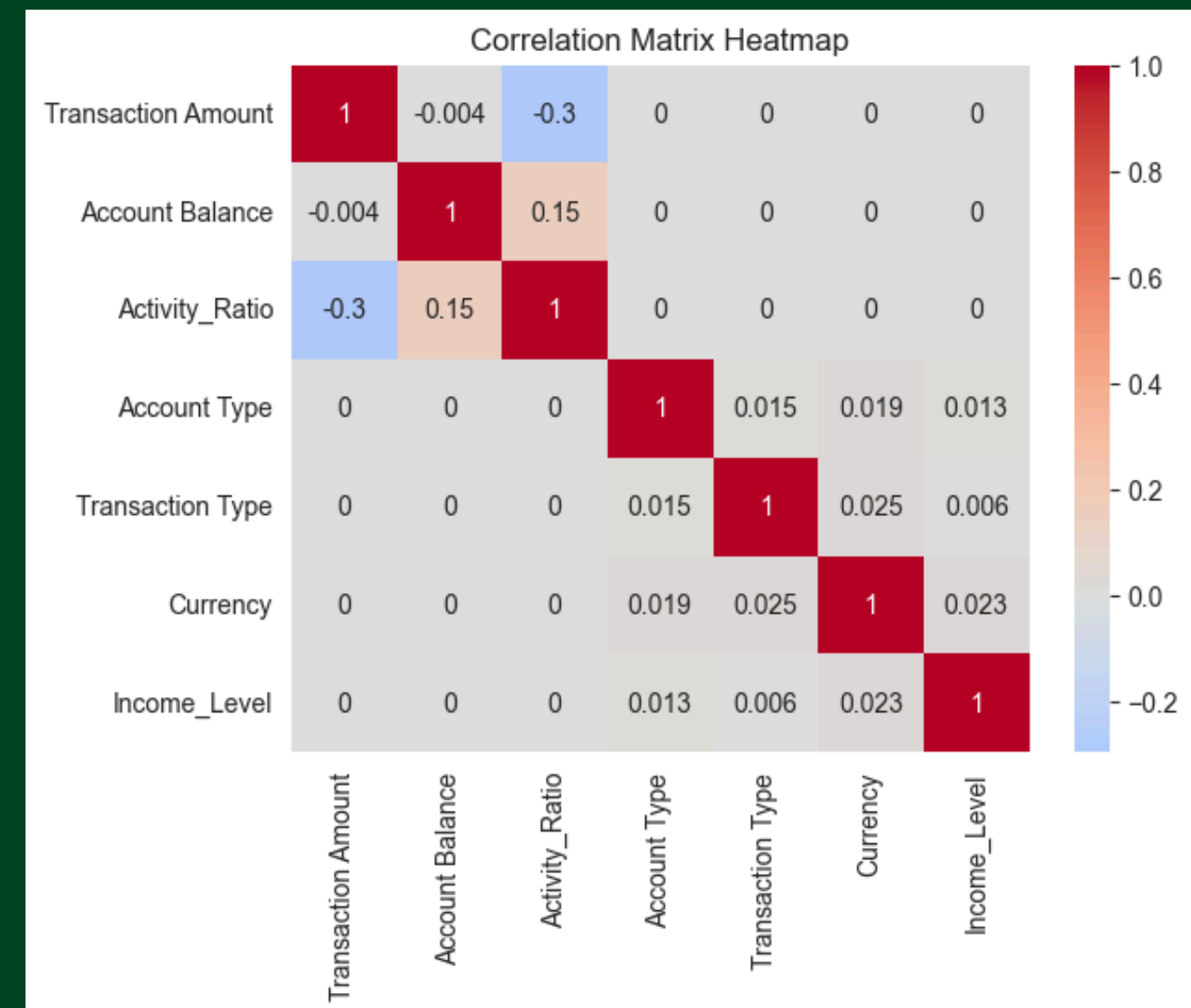


Próbowaliśmy utworzyć nowe zmienne, zakodować zmienne kategoryczne, mieszać biorąc różne zmienne. Niestety nie dało to zbyt wiele

PO KONSULTACJACH:

- Utworzona została nowa zmienna kategoryczna, która przypisywała na podstawie 'Branch' poziom przychodu w danym rejonie (Medium, High, Low) - jako Income_Level
- Zostawiona została utworzona już wcześniej zmienna $\text{ActivityRatio} = \text{AccountBalance} / \text{TransactionAmount}$

Dla zmiennych numerycznych liczymy korelację, dla zmiennych kategorycznych współczynnik V Cramera





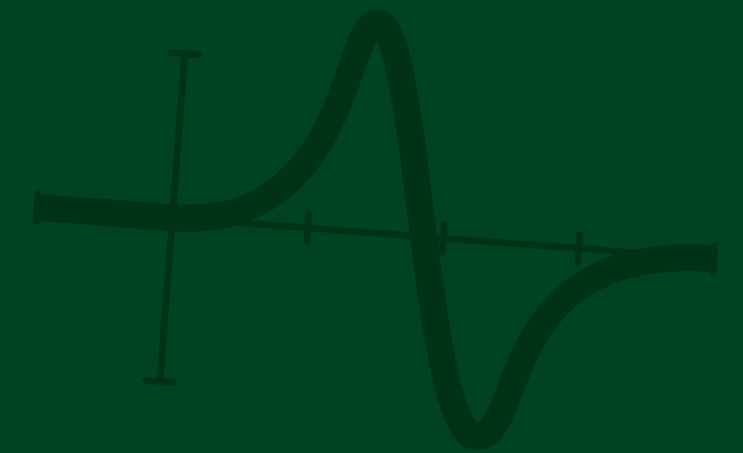
Shillouette jest niejednoznaczne, jako miarę klastrów
przyjmujemy dunn_index

Tworzymy i dostrajamy modele:

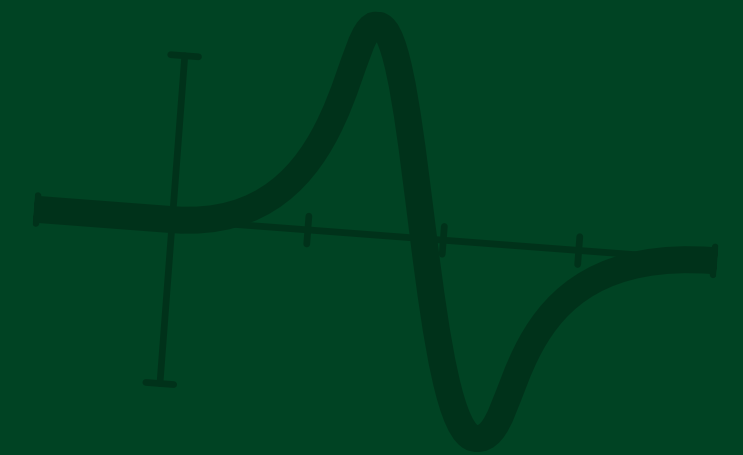
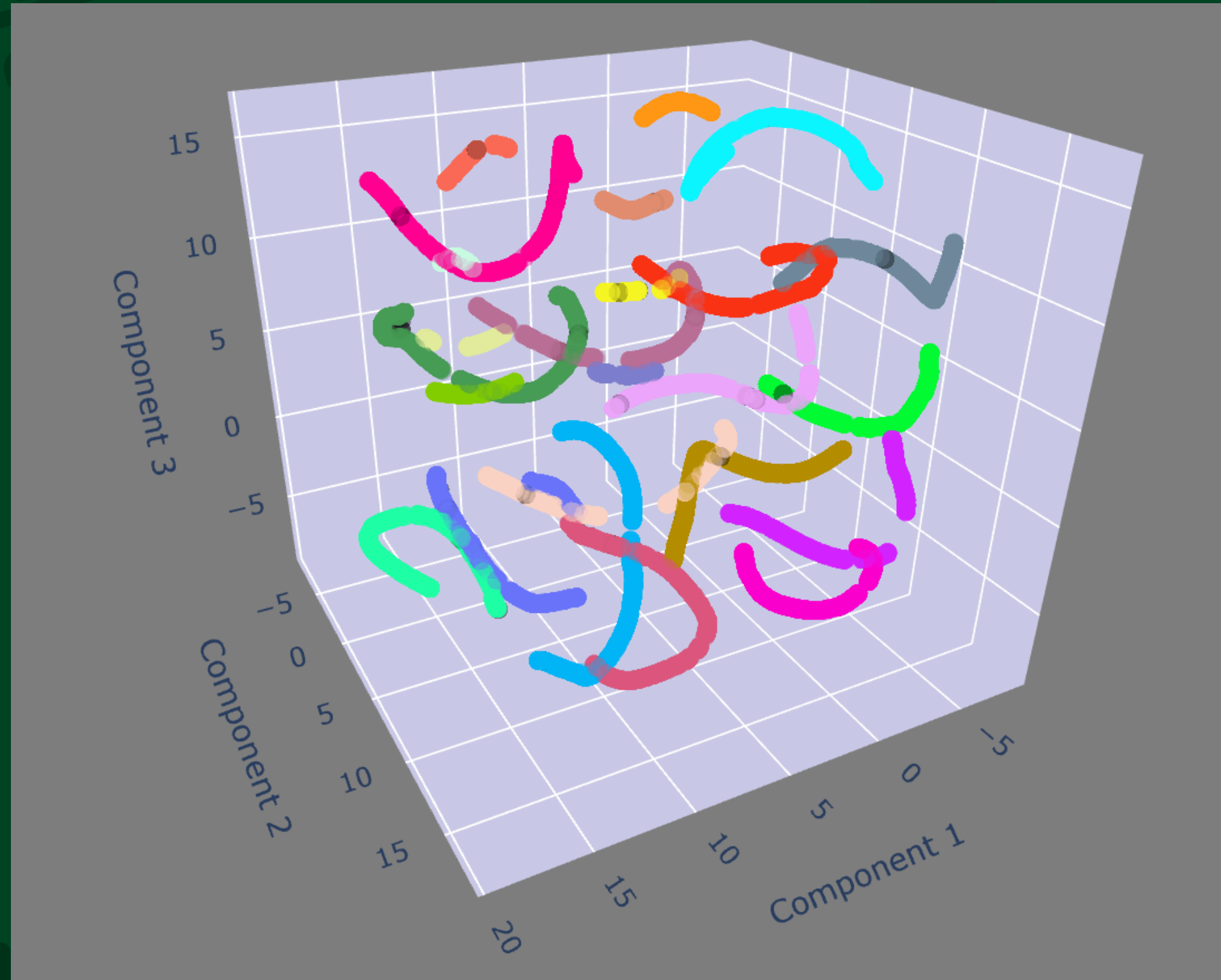
- K-means
- DBSCAN
- HDBSCAN
- AgglomerativeClustering



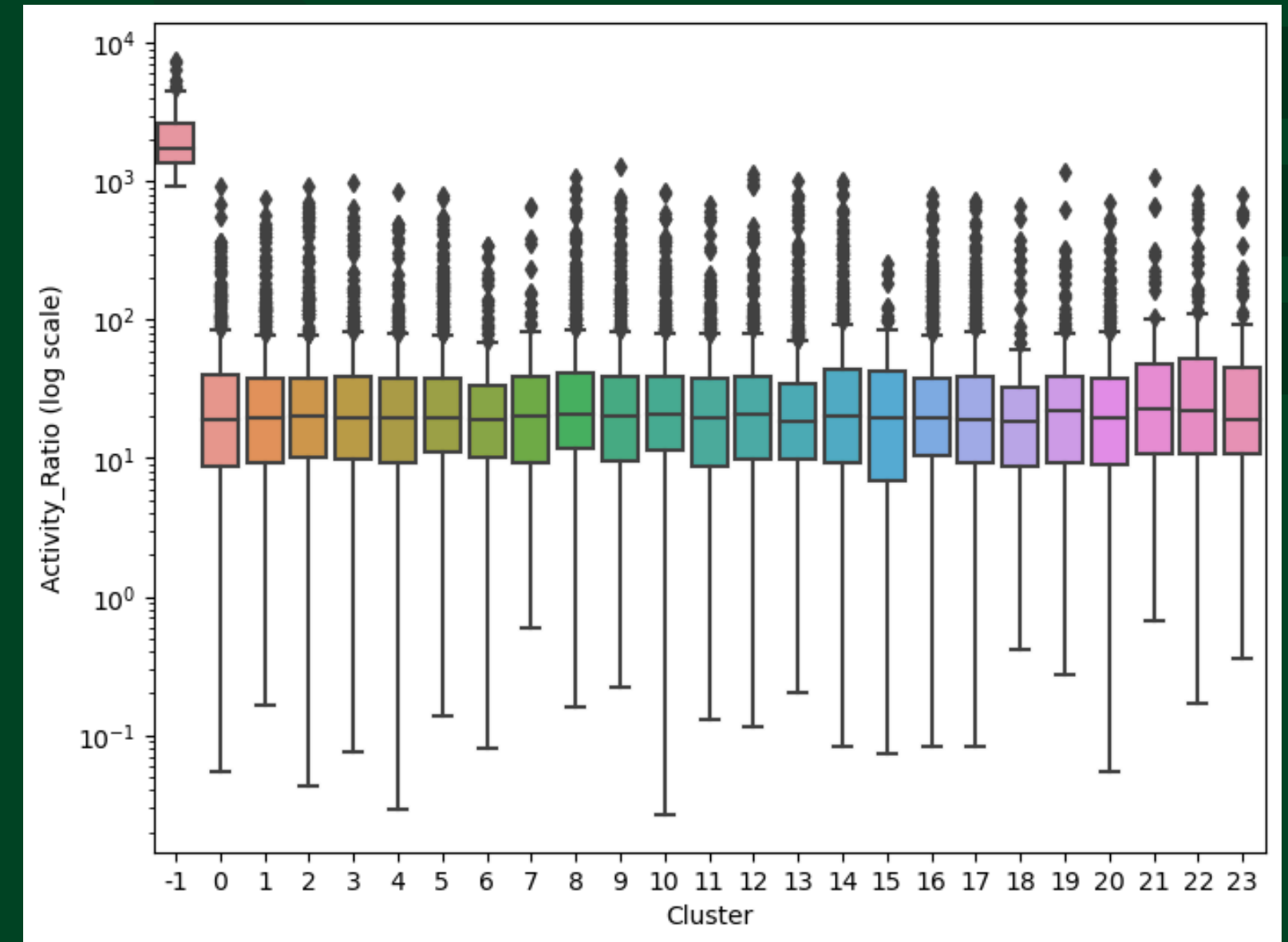
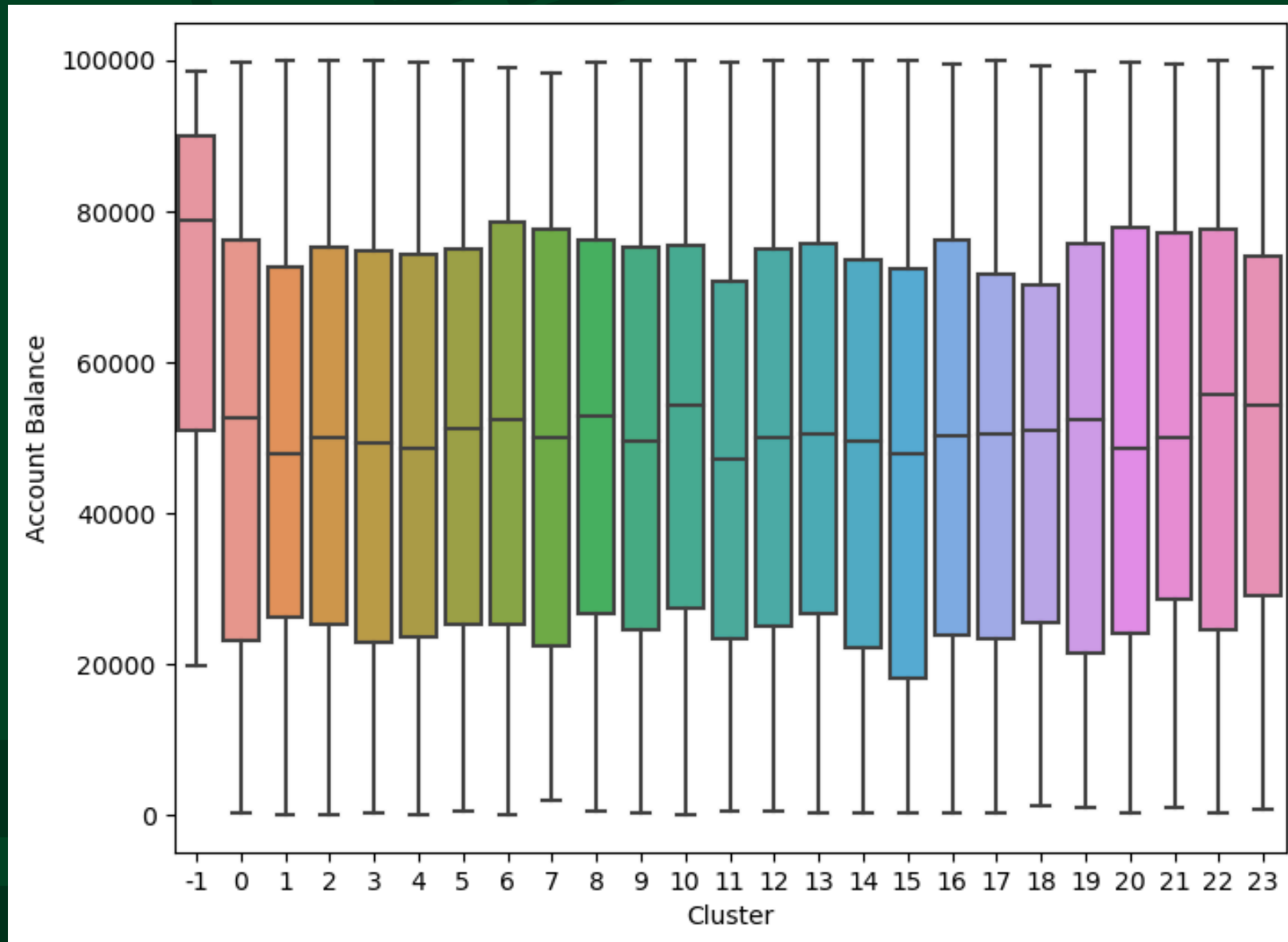
Najlepiej wychodzi DBSCAN



Finalne klastry:



Interpretacja klastrów

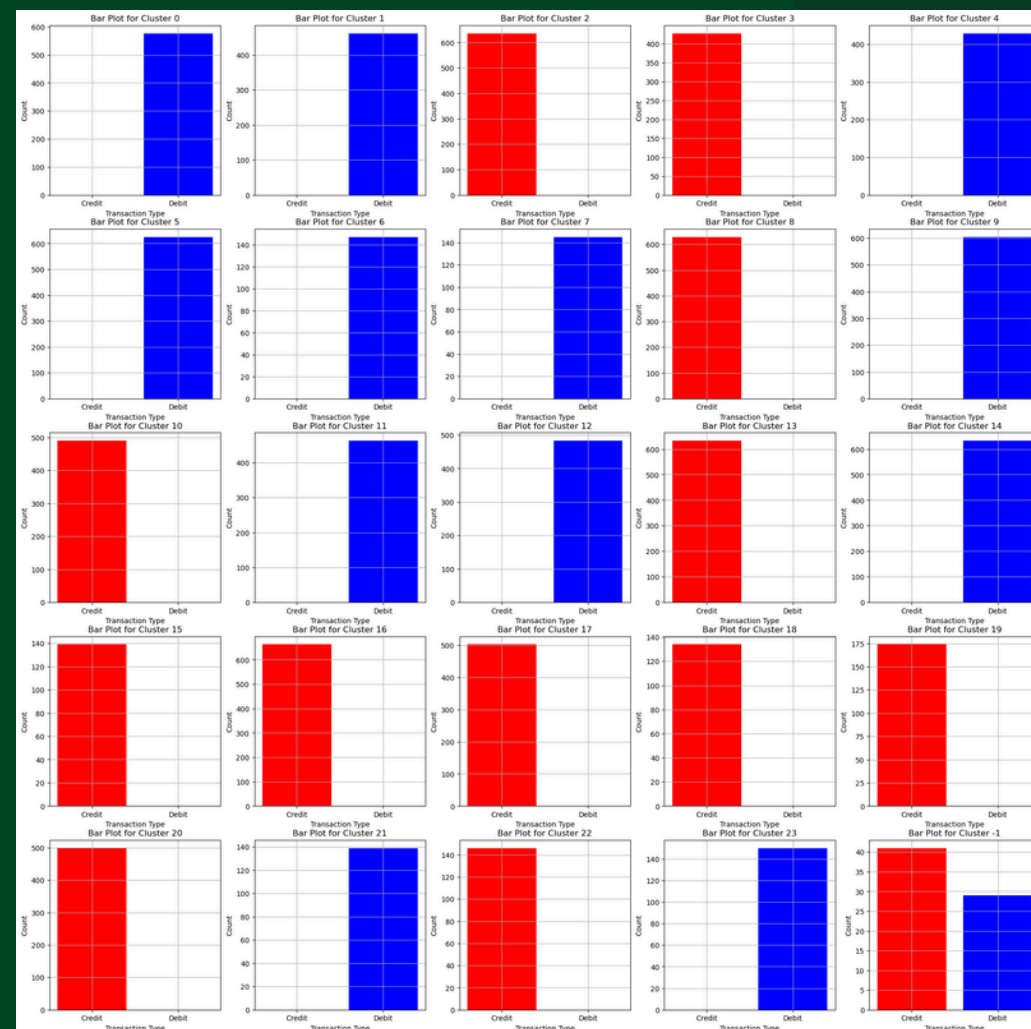


zmienne numeryczne

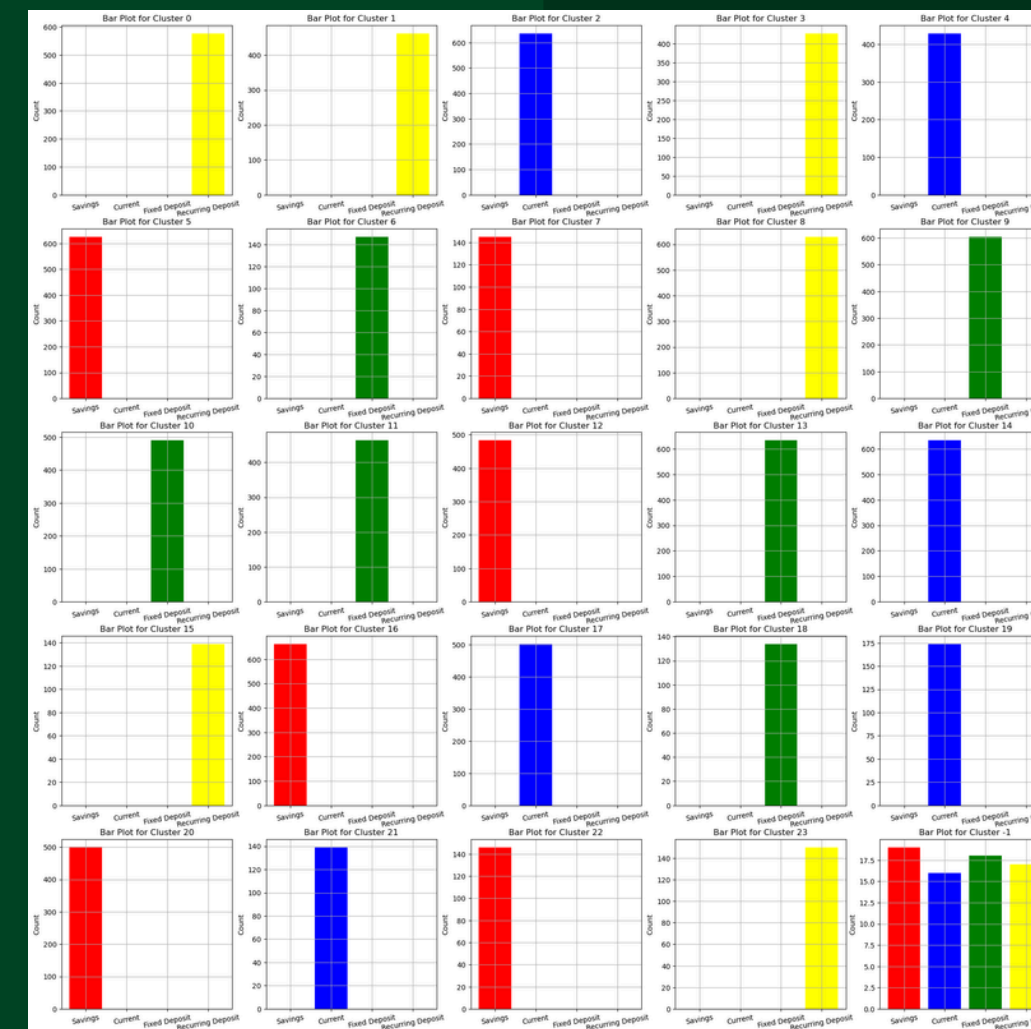
Interpretacja klastrów



Income_Level



Transaction_Type



Account_Type

zmienne kategoryczne

WNIOSKI:

Tak naprawdę, zmienne kategoryczne tworzą nam podział na klastry, zmienne numeryczne (kolumna Activity_Ratio) są nam w stanie wskazać pewne outliery w zachowaniach klientów.

Zmienna Activity_Ratio została stworzona, aby móc wskazać klientów o 'niebezpiecznym' zachowaniu biznesowym, zatem klienci będący outlierami w danej zmiennej mogą być klasyfikowani jako klienci podejrzani.

Ramka danych jednak ze względu na swoją strukturę może być traktowana tylko jako trening, dane są wygenerowane, ciężkie do utworzenia jakichkolwiek wiarygodnych i prawdziwych wniosków.



DZIĘKUJEMY

