Wyobraź sobie, że jesteś Data Consultantem / Big Data Analitykiem w zespole Data&AI w InPost. Trafia do Ciebie zadanie od zespołu marketingu związane z analizą klientów nadających paczki. Zespół marketingu zastanawia się, czy użytkownicy nadający paczki przez naszą stronę www lub aplikację mobilną mogą zostać podzieleni na segmenty. Twoim zadaniem jest przeprowadzenie analizy użytkowników, która pomoże lepiej zrozumieć klientów nadających paczki oraz przygotowanie rekomendacji dotyczącej potencjalnej segmentacji, która mogłaby zostać wykorzystana do realizacji kampanii targetowanych.

Po pierwszym spotkaniu dotyczącym wymagań, zespół marketingu wrócił z pomysłami w jakich ujęciach chciałby spojrzeć na użytkowników nadających paczki. Poproszono Cię, żeby przeprowadzić analizę m.in. liczby paczek na użytkownika, daty ostatniego nadania paczki, częstotliwości nadawania paczek, sposobu nadawania paczki, sposobu odbierania paczki, miast gdzie nadawane są paczki, kategorii z których najczęściej Ci klienci zamawiają paczki. Możesz też zaproponować inne spojrzenia i analizy klientów. Wszystkie analizy powinny zostać przeprowadzone uwzględniając L12M.

Do Twojej dyspozycji jest zestaw danych składający się z 3 plików csv. Poniżej znajduje się krótki opis zbiorów danych.

1. parcels

id\_paczka - Id paczki | id\_klient = id klienta | data\_nadania - data nadania paczki| kanał – strona internetowa lub aplikacja| sposób dostawy – paczkomat lub kurier| id\_paczkomat – id paczkomatu

Obraz zawierający stół

Opis wygenerowany automatycznie

1. apms

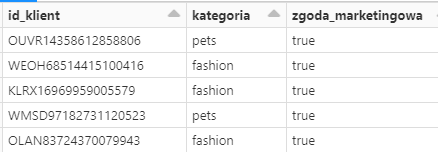
id\_paczkomat – id paczkomatu | miejscowosc – miejscowość, w której znajduje się paczkomat| klasa\_miejscowosci - kategoria miejscowości, w której znajduje się paczkomat

Obraz zawierający stół

Opis wygenerowany automatycznie

1. customers

id\_klient - Id klienta | kategoria - najczęściej wybierana kategoria zakupowa | zgoda\_marketingowa – flaga wskazująca czy użytkownik ma zgodę marketingową



Analizę możesz przeprowadzić w SQL, python, pySpark. Wyniki analiz warto zaprezentować w formie raportu w narzędziu do wizualizacji danych (np. Power BI), przy czym w raporcie warto skupić się na tym, żeby odpowiednio pokazać wyniki, a nie na warstwie wizualnej.