

Zadanie rekrutacyjne na praktyki 2021 w Daftcode

Wykonawca
Łukasz Jaremek

Temat
**analiza bazy danych
w poszukiwaniu oraz likwidowaniu
przyczyn obciążeń zwrotnych**

Zadanie pierwsze

Polecenie

Napisać skrypt w Pythonie, który będzie dopasowywał zgłoszone chargeback'i (plik excel) do transakcji z bazy danych bazy danych.

Rozwiązanie (Plik: Task1.py)

Napisałem skrypt, który pobiera zgłoszenia z pliku *.xlsx* oraz bazę danych z pliku *.sqlite* a następnie przeszukuję bazę w poszukiwaniu transakcji, które zostały zgłoszone.

Funkcja *match_chargebacks* porównuje oba pliki i zwraca listę numerów id szukanych transakcji.

Natomiast funkcja *reports* zwraca id transakcji podanej marki karty. Uruchomienie pliku wygeneruje plik tekstowy z listą numerów id zgłoszonych transakcji (*reported_id.txt*).

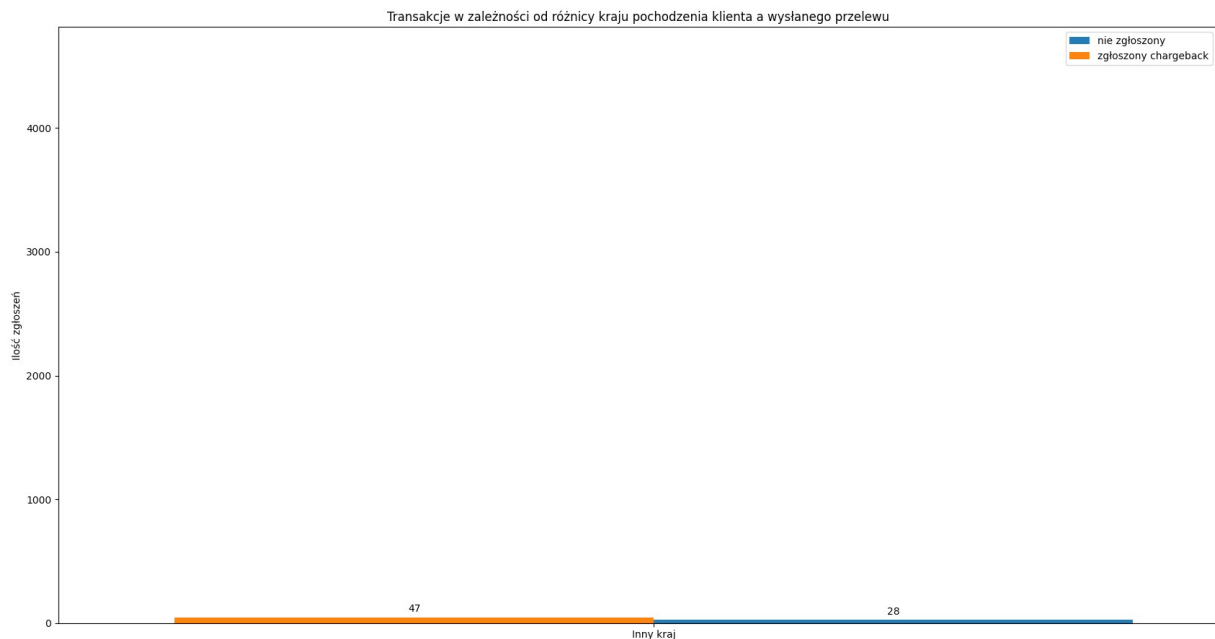
Zadanie drugie

Polecenie

Przeprowadzanie analiz w celu znalezienia źródeł obciążeń zwrotnych.

Rozwiązanie (Pliki: Task2_*.py)

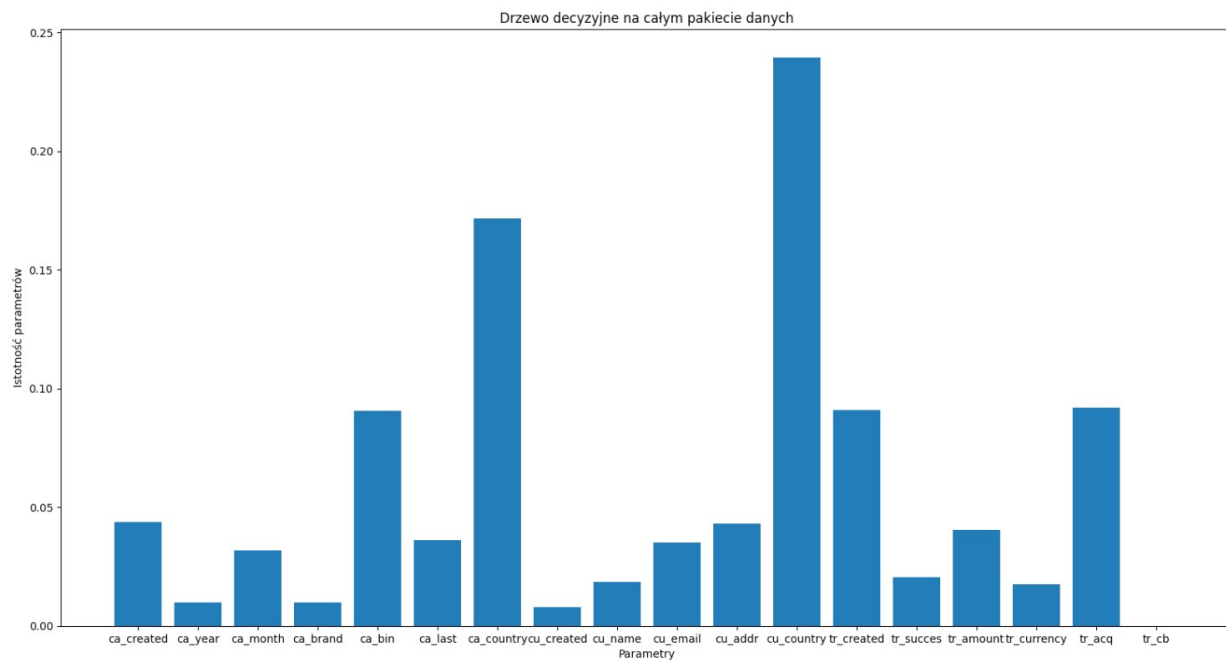
Wykonanie zadania rozpocząłem od próby ręcznego „wyłuskania” danych. Stworzyłem pliki (*Task2_1_*.png*), które generują wykresy różnych zależności w bazie danych. Najciekawszym z nich jest wykres przedstawiający ilość zgłoszeń w zależności od różnicy kraju pochodzenia klienta, a kraju w którym dokonany został przelew



Fragment pliku *Task2_1_chargebacks_another_country.png*

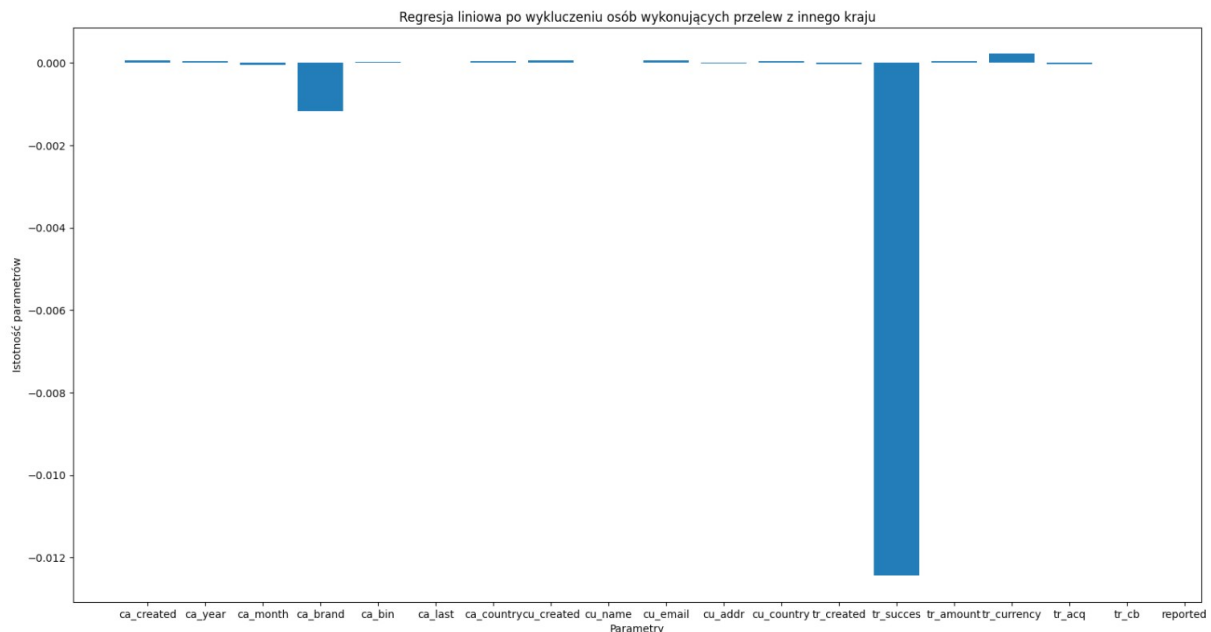
Widać na nim, że osoby które wykonały przelew w innym kraju w ~60% przypadków żądały obciążenia zwrotnego.

Druga wstępna analiza, którą przeprowadziłem polegała na stworzeniu drzewa decyzyjnego (*Task2_decisiontree.py*). Z tej analizy również wynika, że najbardziej istotne parametry to kraj pochodzenia klienta oraz kraj dokonania przelewu.



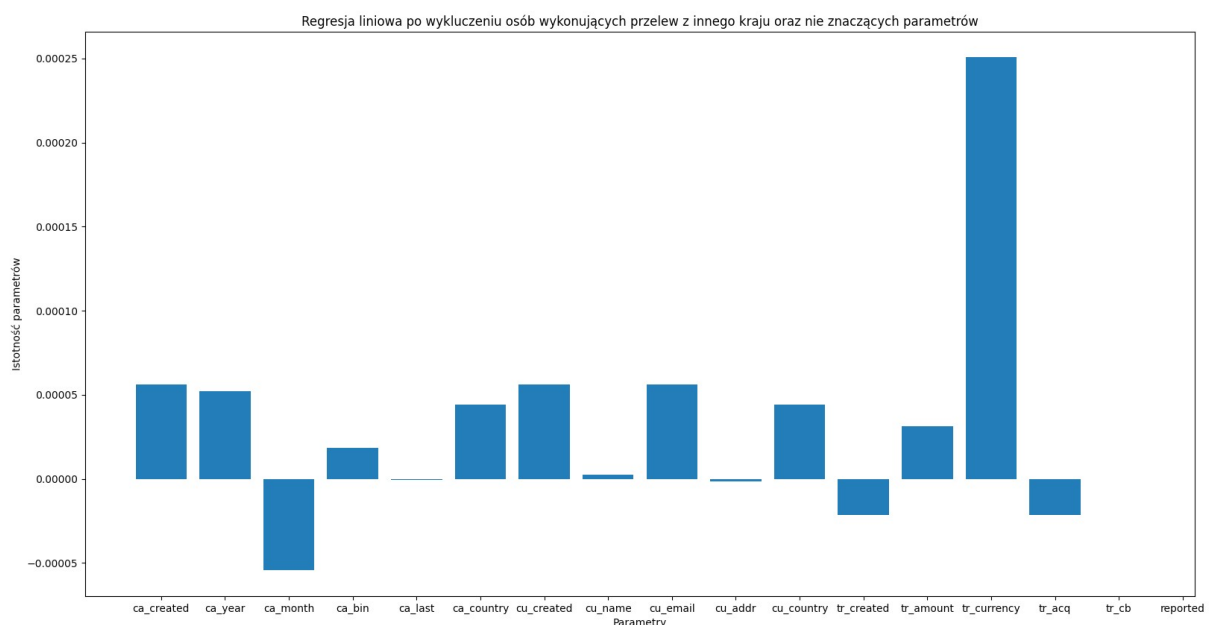
Plik *Task2_2_decisiontree.png*

Następnym krokiem było badanie danych z wykluczeniem osób, które wykonały przelew z innego kraju. Regresja liniowa (Plik *Task2_linearregresion_1.py*) wykazała, że najmniej ważne znaczenie ma sukces wykonania transakcji oraz marka karty, dlatego w kolejnym badaniu zostaną one pominięte.



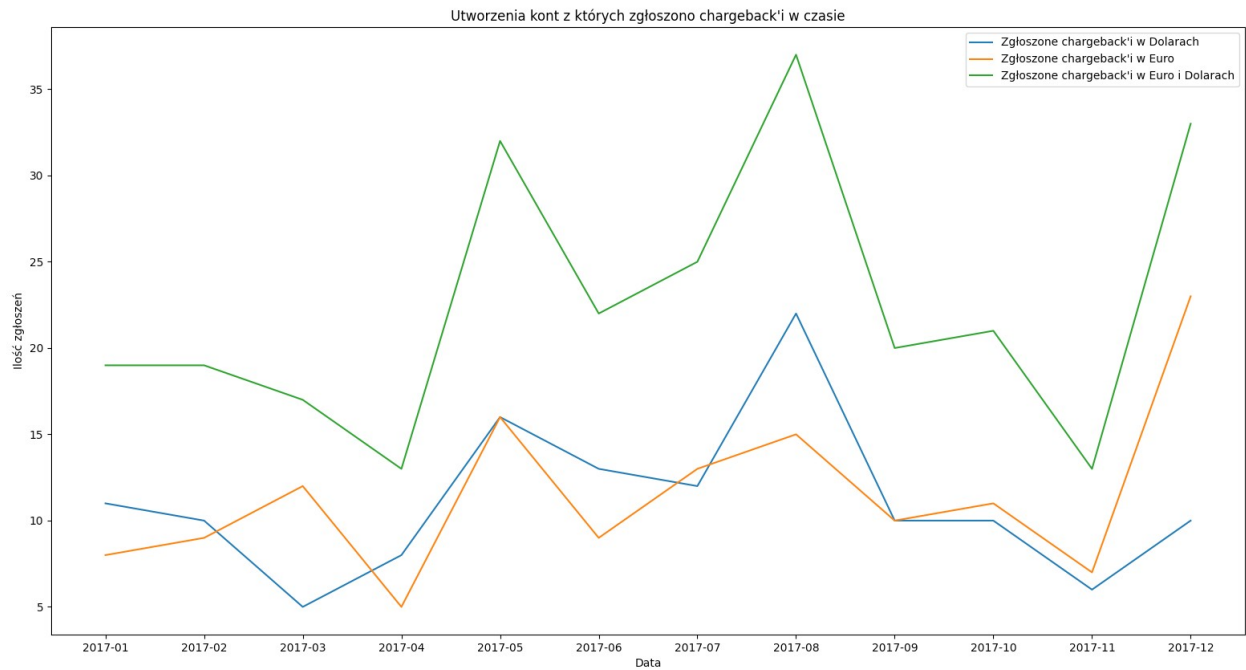
Plik *Task2_2_linearregression_1.png*

Po wykluczeniu dwóch najmniej znaczących parametrów regresja uznała za najbardziej znaczące argumenty: walutę w jakiej została dokonana transakcja oraz dane związane z czasem założenia karty. W kolejnym kroku przyjrzyć się temu bliżej.



Plik *Task2_2_linearregression_2.png*

Ilość zgłoszonych chargeback'ów w zależności od czasu założenia karty (podział na walutę w której został wykonany przelew).



Task2_2_currency_cardcreated.png

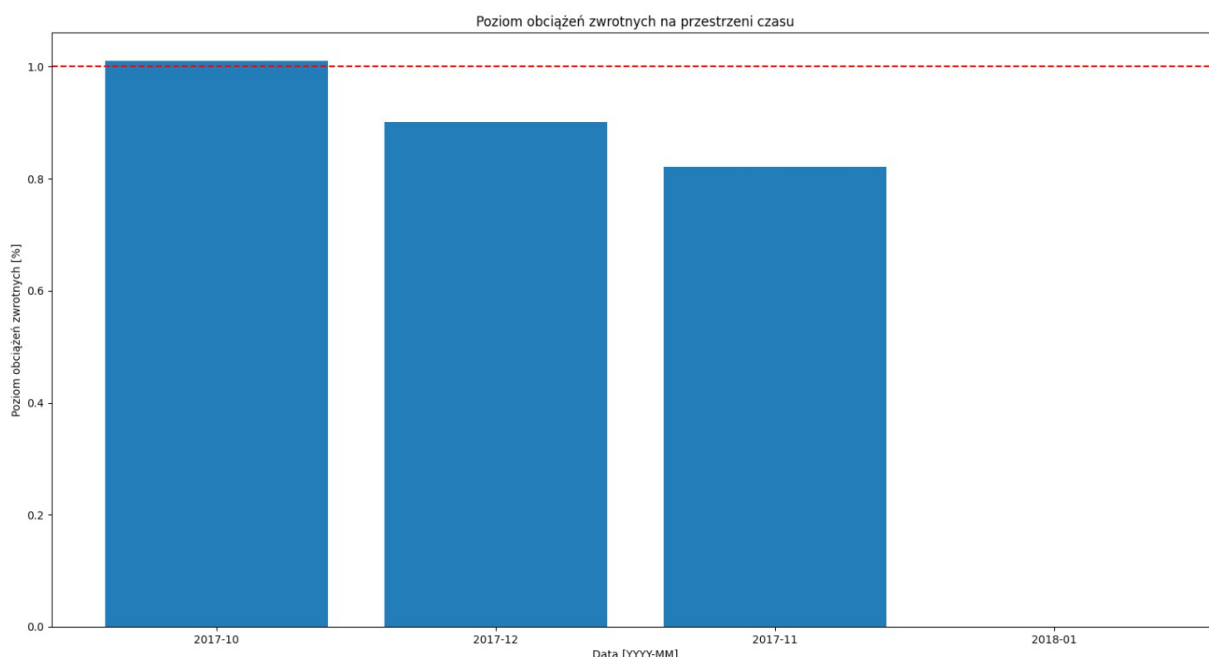
Zadanie trzecie

Polecenie

Krytyczny poziom obciążenia zwrotnego dla handlowców wynosi 1%. W oparciu o przeprowadzoną analizę, proszę zaproponować działania mające na celu obniżenie wskaźnika obciążeń zwrotnych u Akceptanta do poziomu poniżej 1%. Proszę, przedstawić swoje rozwiązanie w formie prezentacji w formacie PDF lub pptx.

Rozwiązanie (Pliki: Task3_*.py)

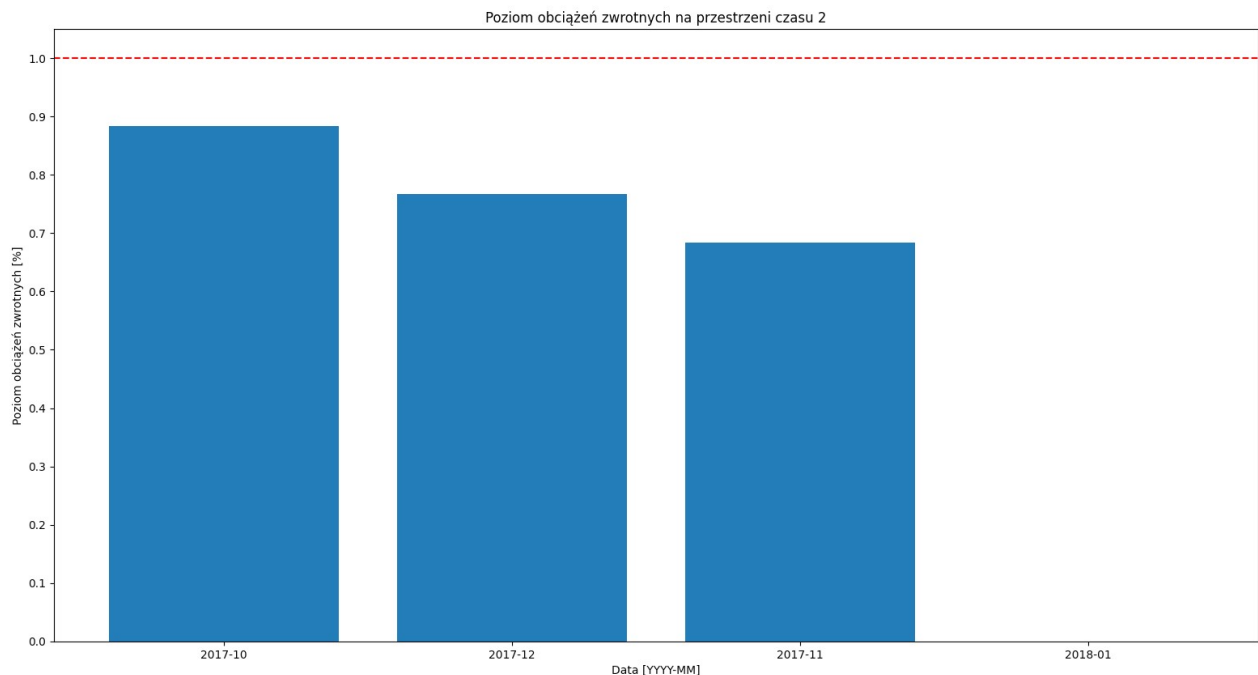
Zacząłem od sprawdzenia, w których miesiącach obciążenie zwrotne przekroczyło krytyczny poziom.



Plik: Task3_chargeback_level.png

Jak widzimy, krytyczny poziom zostaje przekroczony tylko w październiku 2017 roku, spróbuję obniżyć ten poziom dzięki analizom dokonany w wcześniejszym ćwiczeniu.

Zacznę od zbadania poziomu obciążeń zwrotnych w sytuacji, gdy nie będziemy brać pod uwagę przelewów, wykonanych w innym kraju, niż kraj pochodzenia konsumenta.



Task3_chargeback_level_2.png

Możemy zauważyć wyraźną poprawą, poziom obciążeń zwrotnych spadł w każdym miesiącu o $\sim 0.2\%$.

Wniosek:

Głównym problemem są przelewy wykonane w innym kraju, ciężko stwierdzić jaka była przyczyna chargeback'ów, być może były to transakcje robione bez wiedzy właściciela karty (kradzież karty, danych bankowych a następnie wykonywanie zakupów przez złodzieja). Zaleciłbym wprowadzenie dodatkowych zabezpieczeń w takich przypadkach, na przykład w postaci dodatkowej weryfikacji użytkownika.