

Raport Przygotowania i Działania Programu Obliczania Różnic Kursowych

Wprowadzenie

W tym dokumencie przedstawiam proces przygotowania i sposób działania programu służącego do wyliczania aktualnej, pozostającej do zapłaty należności z faktur zagranicznych kontrahentów. Dokument ten zawiera opis poszczególnych etapów działania programu, napotkane trudności w implementacji oraz rozważane alternatywne podejścia.

Zarys problemu

Dany jest zbiór danych opisujących faktury zagranicznych kontrahentów. Każdy rekord opisuje jedną fakturę i zawiera jej numer identyfikacyjny, kwotę należności, walutę w której ta kwota jest denominowana oraz datę wystawienia tejże faktury. Dodatkowo dany jest drugi zbiór danych opisujący zapłacone kwoty dla każdej faktury denominowane w złotych wraz z datą zapłaty. Naszym celem jest określenie kwoty pozostającej do zapłaty dla każdej z faktur.

Opis danych wejściowych i wyjściowych

Dane wejściowe mają formę dwóch plików csv. Pierwszy opisuje dane faktur i ma następujące kolumny:

- id - unikatowy numer identyfikacyjny
- liability date - data wystawienia faktury w formacie rok-miesiąc-dzień (np. 2022-01-30)
- liability value - kwota należności
- currency - 3 literowy kod waluty, w której denominowana jest kwota należności (np. gbp)

Drugi plik opisuje dane płatności i ma następujące kolumny:

- id - unikatowy numer identyfikacyjny zgodny z id z poprzedniego zbioru danych
- payment date - data zapłaty w formacie rok-miesiąc-dzień
- payment value - kwota zapłaty denominowana w złotych

Zbiór wyjściowy zawiera dane dla każdej faktury z obu danych wejściowych oraz dodatkowo kolumny:

- original exchange rate - kurs waluty na dzień wystawienia faktury
- payment exchange rate - kurs waluty na dzień zapłaty
- today exchange rate - kurs waluty na dzień uruchomienia programu
- original value in pln - kwota należności denominowana w złotych na dzień wystawienia faktury
- paid value in foreign - kwota zapłaty denominowana w walucie należności faktury na dzień zapłaty
- remaining liability in foreign - pozostająca kwota należności w walucie należności faktury po zapłacie
- remaining liability in pln - kwota pozostającej należności denominowana w złotych na dzień uruchomienia kodu

Źródło danych o kursie walut

Kursy walut na dany dzień są pobierane poprzez publicznie dostępne API Narodowego Banku Polskiego dostępnego pod adresem: <https://api.nbp.pl/api/exchangerates/>.

Działanie programu

Program składa się z trzech funkcji:

- `fetch_exchange_rate`: przyjmuje ona zmienne `currency` oraz `date` oznaczające odpowiednio trzy literowy kod waluty oraz datę. Zwraca on kurs danej waluty na daną datę pobierając ją z API NBP.
- `read_and_join_data`: przyjmuje dwie ścieżki do danych wejściowych. Łączy je za je i zwraca jako obiekt typu `pandas.DataFrame`
- `calculate_balance`: przyjmuje obiekt typu `pandas.DataFrame` opisujący faktury i dokonane zapłaty oraz ścieżkę do pliku, gdzie ma zostać zapisany wynik działania programu. Dla każdej faktury pobiera kursy wymiany waluty należności faktury na dzień wystawienia, zapłaty oraz obecny. Następnie poprzez proste kalkulek wyznacza wartości kolumn opisane w sekcji Opis danych wejściowych i wyjściowych. Wyliczony zbiór danych jest zapisywany pod określoną na wejściu ścieżką oraz zwracany przez funkcję jako obiekt typu `pandas.DataFrame`

Celem poprawnego działania programu użytkownik musi zmienić wartości zmiennych `liability_path`, `payment_path`, `result_path`. Dwie pierwsze są następnie przyjmowane przez funkcję `read_and_join_data`, której rezultat jest przekazywany do funkcji `calculate_balance` wraz z `result_path`.

Limitacje metody

Aktualna implementacja programu zakłada poprawny format danych wejściowych, w szczególności daty. Ponadto zakłada, że kurs zadanej waluty jest dostępny na zadany dzień. Dodatkowo przyjmuję, że każda faktura została opłacona co najwyżej raz. W przypadku, gdy nie dokonano żadnej płatności, wartość w kolumnie `payment value` jest równa 0, a w kolumnie `payment date` jest równa wartości `liability date`.

Program jest dość wolny. Jego czas wykonania dla zbioru 1000 rekordów wygenerowanych losowo za pomocą skryptu `generate_data.py` wyniósł ponad 17 minut.

Alternatywne podejście do problemu

Moim początkowym podejściem do problemu było wstępne skonstruowanie tablicy kursów wymiany dla przedziału czasu pomiędzy najstarszą oraz najnowszą datą występującą w zbiorach wejściowych. Znacząco przyspieszyłoby to działanie programu dla bardzo dużych danych wejściowych, bo nie wymagałoby każdorazowego pobierania kursu waluty przez API NBP. Dodatkowo ułatwiłoby radzenie sobie z brakiem kursu waluty na dany dzień poprzez branie ostatniego dostępnego kursu. API NBP umożliwia pobranie kursów wszystkich dostępnych walut z przedziału o maksymalnej długości 90 dni (pomijając wszystkie dni na które kurs jest niedostępny, np. weekendy). Z tego względu możliwym jest pobieranie danych na każdy kwartał roku. Pobranie w ten sposób danych z okresu 2002-2023 trwało mniej niż minutę. Jednakże, dużym problemem okazało się wydajne przetwarzanie i łączenie pozyskanych danych. API NBP umożliwia pobranie danych w formacie JSON oraz XML o dość złożonej strukturze, której przetwarzanie do prostego formatu danych jest czasochłonne. Ponadto pula dostępnych walut w zależności od okresu waha się od 20 do 30 co stwarza dodatkowe trudności. Z tych względów podejście to okazało się bardzo problematyczne i nie dawało gwarancji optymalizacji czasowej.

Raport przygotowany przez: Julia Flaszyńska

Data: 15.02.2024