



Summer e-Xperience 2020

ZADANIE – DATA ANALYST - Intern

W załączonych do maila plikach CSV znajdziesz dane pewnej wypożyczalni rowerowej z San Francisco. Korzystając z tych danych wykonaj proszę następujące kroki:

1. Zwróć wszystkie rekordy z tabeli TRIPS wraz z nazwami stacji początkowej i końcowej dla podróży rozpoczynających się na Alcatraz Ave at Shattuck Ave.
2. Zwróć liczbę i średnią długość podróży (w sekundach) po nazwach regionów stacji początkowej dla podróży z 2018 r.
3. Zwróć (unikalną) listę 'id' oraz skróconych nazw wszystkich stacji, w których zakończyło się przynajmniej 5 podróży - wyniki zapisz do tabeli STATION_A.
4. Zwróć (unikalną) listę 'id' wszystkich stacji końcowych, do których nie odnotowano podróży, ze stacji, w której obecnie nie ma dostępnych rowerów - wyniki zapisz w tabeli STATION_B.
5. Połącz wyniki z punktu 3 i 4 zwracając tylko te id stacji, które występują w obu listach.
6. Dla każdego z regionów wskaż 'id' oraz długość najdłuższej podróży, która rozpoczęła się w danym regionie (stacja końcowa może być położona w innym regionie). Pod uwagę weź tylko podróże, które rozpoczęły się w 2018. (wskazówka: oszacuj długość trasy wyliczając długość po linii prostej, użyj pola STATION_INFO.STATION_GEOM oraz twierdzenia Pitagorasa).
7. Dla każdej podróży zwróć listę zawierającą datę startu i datę końca podróży oraz jej 'id', nazwę stacji początkowej i końcowej (oraz ich 'id'). W tej samej tabeli podaj także datę startu i końca, wraz z nazwą stacji końcowej poprzedniej podróży, która zaczęła się w tej samej stacji początkowej.



Źródła danych:

1. Do maila załączyliśmy 4 pliki csv z danymi potrzebnymi do wykonania zadania:
 - trips.csv
 - station_info.csv
 - station_status.csv
 - regions.csv
2. Dodatkowo opis danych znajdziesz w pliku:
 - Metadane - opis.xlsx

Narzędzia

- Oracle SQL

Aby ułatwić Tobie pracę nad zadaniem, przygotowaliśmy instrukcję oraz skrypt, które pozwolą w łatwy sposób zbudować potrzebne źródła danych. Znajdziesz je w plikach skrypt_wczytanie_bazy.sql oraz instrukcja_wczytanie_bazy.doc.

Lista plików do przesłania:

1. Kod źródłowy
2. Wyniki zapytań przedstawione w pliku Excel - każde zadanie w osobnej zakładce

Powodzenia!

Zespół CEMS RESEARCH