

Zadanie będzie polegało na policzeniu wskaźnika liquidity

Liquidity rozumiane będzie jako % ogłoszeń, które otrzymały co najmniej 1 odpowiedź (telefoniczną lub mailową) w okresie 7 dni

Przykład:

Użytkownik dodał 1 kwietnia do serwisu 10 ogłoszeń

W dniach od 1 do 7 kwietnia otrzymał odpowiedzi na 6 ogłoszeń.

W dniu 2 kwietnia dodał kolejne 5 ogłoszeń i otrzymał odpowiedzi na wszystkie z nich w czasie 7 dni od pojawienia się tych ofert w serwisie

Liquidity wynosi $(6+5)/(10+5) = 73\%$

Do dyspozycji mamy 4 bazy danych:

1. Data_ads - tutaj znajdują się informacje o ogłoszeniach
2. Data_replies - informacje o odpowiedzi per ogłoszenie per dany dzień
3. Data_categories - mapowanie do drzewa kategorii
4. Data_segments - mapowanie do segmentacji dla każdego użytkownika

Nazwy kolumn dla data_ads

date
user_id
ad_id
category_id
params

Nazwy kolumn dla data_replies

date
user_id
ad_id
mails
phones

Nazwy kolumn dla data_segmentation

user_id
segment

Nazwy kolumn dla data_categories

category_id
category_name

Pytania do zadania

- 1 Ile wynosi liquidity dla każdego użytkownika, który posiadał ogłoszenie na serwisie
- 2 Jakie dostrzegasz różnice pomiędzy segmentami w zakresie danych, które masz dostępne
- 3 Co wg Ciebie może mieć wpływ na wyższy lub niższy poziom liquidity

Do zrobienia

- 1 Przygotowanie skryptu w pythonie/R, który dane zaimportuje do bazy danych (może być dowolna baza: mysql, postgresql, etc.)
- 2 Treść zapytań SQL, które pozwoliły na obliczenie liquidity
- 3 Przygotowanie skryptu w pythonie/R, który liczy liquidity dla wszystkich użytkowników
- 4 Analiza wyników - proszę o przedstawienie pełnej analizy w notebooku (preferowany Jupyter/R Markdown)
- 5 Wyniki proszę przedstawić za pomocą narzędzia google/slides