Analyst Exercise

July 01, 2019

Overview:

The ability for a company to retain its customers for a period of time is called customer **retention**.**Customer retention metrics** are a set of measures that tells how well a company is doing in retaining its customer base. There are various measures of customer retention and every business/function chooses one that's most suitable for their business needs. The rider CRM team in Uber would want to find the retention of a customer cohort on a weekly-rolling basis, so that they can take necessary intervention if a rider has not taken a ride for 28 days.

Exercise 1:

Write a Query in SQL which will give me the following Columns city wise and date-wise. The following data needs to be found '**2019-06-01**'. You could choose any sql language you are comfortable with. Please use the sample data to create a table and write the query.

Definitions of Columns in the Output Query (Single Query):

1. **date:** the date in respect to which the below metric will be computed.
2. **city_id:** the id of the city
3. **mau_28:** Count of distinct riders who have completed min one trip in the last 28 days with respect to the date in column 'date'
4. **previous_mau:** count of distinct riders who have taken min one trip between the last 56 to last 29 days with respect to the date in column 'date'
5. **retained:** intersection of **distinct** riders in the **previous_mau** and **mau_28** phase
6. **resurrect:** count of distinct riders who were inactive in the **previous_mau** phase but were active in the **mau_28** phase.

   ***Active**: If a rider has completed at least one trip in the respective period*
   ***Inactive**: If a rider has not taken a single trip in the respective period*

Tables Schema:

Table 1: Trips_Data which stores the data for all the trips

Columns:

1. *date : The date on which the trip was requested*
2. *rider_id: Id of the rider requesting the trip*
3. *trip_id: id of the trip request*

4. *city_id: Id of the city where the trip was taken*
5. *status: Status can be 'completed' , 'cancelled' or 'unfulfilled'*

Special Considerations:

1. A rider may take a trip from multiple cities, which might lead to counting the riders active or inactive in multiple cities. Hence to solve this problem a rider needs to be mapped to one city only. A rider should be mapped to a city from which they have taken the maximum trips considering only their recent 20 trips.
2. For all calculations related to a city it is important to consider the city mapped to a rider instead of the city where the trip occurred.
3. Our database system does not have a standard mode function so the rider city mapping needs to be derived.
4. Ensure that each city has only one row for a respective date.
5. Please give enough comments/explanations of the logic in the query so we could assess how/what you are trying to achieve while writing the query. Please put a comment for all your filters, joins and join conditions, aggregate statement, case statements, etc.
6. Use "with" clauses or create temporary tables to write the query and avoid using subqueries as it impact readability of your code.

Exercise 2:

The Marketing team uses this definition measuring Retention and Reactivations from the above query result:

- **Retention: retained / mau_28**
- **Reactivation: resurrect / previous_mau**

Could you assess if this is a correct way to measure retention and reactivation? Else could you improve on these definitions with explanations.