The City College of New York
Department of Computer Science
Summer 2020

CSc 33600 Introduction to Database Systems

CitiBike

Refer to the abridged data file "2013-07 - Citi Bike Trip Data.txt" posted on the course page on Blackboard. You are required to design and implement a relational database to help better manage the CitiBike program in New York City.

Deadline to submit your report is 5 pm on 21 July 2020. The deadline is strictly observed. Demonstration of the project is scheduled on Thursday, 23 July 2020 from 5 to 7 pm.

Assignment:

[I]

1- Study the data provided for the use of Citi bikes to gain an understanding of the program utilization.
2- Explore the information on the program available on the Web to verify the basic facts and collect additional information regarding the utilization of Citi Bikes. Write down the additional information and/or requirements you gathered. Give the URLs for the websites contacted.
3- Draw an appropriate E/R diagram that satisfies the basic and additional facts, indicating, weak and subclass entity sets, whenever exist, multiplicity of relationships, and the key, or keys, for each entity set. Distinguish between the parts of the E/R diagram pertaining to the given and additional facts.
4- Translate the E/R diagram in [3] to a relational database schema.
5- Specify a number of essential functional dependencies for each relation. Identify possible keys, whenever exist, and the primary key and foreign keys for each relation.

[II]

Create SQL expressions to break the data provided in the file into the following tables:

   Stations(*Id*, *Name*, *Latitude*, *Longitude*)
   Trips(*StationId*, *MinTripDuration*, *MaxTripDuration*, *AvgTripDuration*, *NumberStartUsers*,
            NumberReturnUsers)
   UsageByDay(*StationId*, *NumberWeekdayStartUsers*, *NumberWeekdayReturnUsers*,
            NumberWeekendStartUsers, *NumberWeekendReturnUsers*)
   UsageByGender(*StationId*, *NumberMaleStartUsers*, *NumberFemaleStartUsers*,
            NumberMaleReturnUsers, NumberFemaleReturnUsers)
   UsageByAge(*StationId*, *NumberMaleUsersUnder18*, *NumberMaleUsers18To40*,
            NumberMaleUsersOver40, NumberFemaleUsersUnder18,
            NumberFemaleUsers18To40, NumberFemaleUsersOver40)

[III]

1- Create appropriate SQL expressions to determine the most frequent trips between any two stations by the day of the week.

2- Create appropriate SQL expressions to find permanently dormant or vacant stations.
3- **Bonus**: Reconstruct the tables and recreate the SQL expressions above to include zip codes, hence allowing for aggregation by area. Note that zip codes are not included in the data provided but may be related to the *StationId* by its *Latitude* and *Longitude*. This could be part of your research and may require some coding!

[IV]

Submit a written report that includes:

1- The complete E/R diagram and schema of the relational database fully specifying the given requirements and any other requirements gathered. Identify all keys and foreign keys of the database relations.
2- SQL code that creates the tables' structure.
3- SQL code that loads the data.
4- SQL code that answer the given queries.
5- Sample outputs for your SQL code.
6- Provide a narrative explaining the outputs produced whenever possible.

Best wishes,

Hesham A Auda
10 July 2020