**Homework – Spatial and Temporal Data**

The following practice may need to be completed using the databases hosted on wmc3317-2 using WB. Use WB whether you are running queries or creating a relational schema (a data model, essentially). Submit homework as a SQL script file. To mark the answers I will copy+paste the SQL script into my WB and execute the code. You could even write essay answers in an MS Word file or the SQL script itself – or in a separate MS Word/PDF file. If there is any dirty data, clean it by making reasonable assumptions.

1. Spatial data:

You will extend the Sakila database with your own randomly created geospatial data as follows. Find the 5-10 customers who have the most number of transactions[[1]](#footnote-1). Extend or add the Sakila database so as to accommodate the point coordinates of the cities/zip codes that the five customers live in (based on the zip code). Similarly, alter the Store table so that it can capture the coordinates of each store. Then write a query that calculates the distance of each of the top customers from each of the stores. Make any other assumptions you find necessary. (Your queries should include those for copying the necessary tables from Sakila to your space on the SQL server).

1. A bit of practice with manipulating datetime values. Consider the database expia (table ‘train). This table contains data about the internet searches for hotels by Expedia users. Copy 10000 records to your database space as follows:

###take special care to limit to 10000 records, not more since train is a large table. If you ignore this limit, the server might hang up.

create table db\_<your sfuemailid>.BUS464Assign as select \* from train limit 10000;

Next, work solely on the new table. Change the format of the dttime field to ‘datetime’ and answer the following questions by referring to the datetime functions at http://dev.mysql.com/doc/refman/5.6/en/date-and-time-functions.html . Copy the answer here.

1. Which year(s) does the data pertain to?
2. Extract the hour of the day for each search result and find which hour has the most number of searches?
3. Which FIVE weeks of the year saw the highest number of searches?
4. Which weekdays saw the lowest number of searches? Any guesses why?
5. The following questions pertain to the data model discussed by the Industry Speaker and provided with this assignment. (**Eric Leung, Beedie School of Business**)
6. Find the all students and their email address (if available) who are currently on the waitlist for BUS464 in Fall 2015 (1157). Note: You should notice some students do not have an email address -- Use LEFT, RIGHT JOINS.
7. Find the name of students who are enrolled in a course taught by a professor who wrote an article called "Market positioning by IT Service Vendors through imitation".
8. Find a list of stakeholders without an email address
9. Find a list of stakeholders with more than one email addresses

\_\_\_\_\_\_\_

1. The Sakila data is fake, so all customers have only one transaction. However, write the SQL code for the top 10 customers, nevertheless. [↑](#footnote-ref-1)