**CSI 300: Project 1 – Option 1**

**Introduction**

In this project, you will be analyzing Burlington housing data using MySQL. A growing

movement in government has been open data as a way to further transparency. The City of

Burlington Open Data website (https://data.burlingtonvt.gov) includes several interesting data

sets about the city. We will be utilizing the data set known as “City of Burlington Property

Details”. Unfortunately the City of Burlington no longer supports the download of this dataset. Instead of an Excel spreadsheet, you are provided with a database that includes all of the listings from several years ago. The data contained in it represents the last update to the file before it was no longer made publically available.

This data set includes a lot of information about every taxable property in Burlington. You will

find things like year built, number of bathrooms, heating type, last sale price, and more. The fact that the data may not be current, it in no way negatively impacts your ability to complete this project. Yu can assume that the data is current for any question you want to answer

You may work alone or with one partner. If you choose to work with a partner, you will both

receive the same grade for the project. You have a month to work on this project, but I don’t

suggest waiting until the last minute! Since there is some flexibility in your topic, it is good to

have time to run your plan by me.

**Decisions and Design**

You need to decide what subset of the data set you want to analyze with regards to what type of data is of interest to you. For example, you may be interested in the correlations between house age and house size/value. Explore the data set and consider what interests you most.

Your goal is to develop at least 5 questions you have about the properties listed in the database. The database is saved in a .zip file on Canvas. The name of the .zip file is **all\_burl\_properties.zip** and should be download from Canvas. Your questions should be specific and have a clear result sought. Here are some examples:

- Analysis of properties by age: In what decade were the most valuable (at present value)

properties built? Is there a correlation between property age and lot size? Is there a

correlation between property age and property value? What is the most common condition of properties from each decade?

- Analysis of properties by street: What is the wealthiest and poorest street in Burlington by

property value? What street in Burlington has the most Tudor style homes? What are the five

oldest streets in Burlington by property age? What street has the largest homes? Is this the same as the street with the highest property values?

You should have at least five interesting questions that you want to answer. You should then consider if you want to redesign the database and move the data from the current design into your newly designed database. If you choose this option your project should be of a complexity such that it utilizes more than one table. The tables should have a relationship with one another that requires the utilization of JOIN statements. For many, keeping the current structure of the database will be the easiest way to proceed which is fine.

**PLEASE NOTE: The database was already designed. You will need to reverse engineer it to create the ERD. From thee you can re-design the database if you wish. You are not required to re-design it. You are required to reverse engineer it.**

**Data Input and Analysis**

How are you going to get the data from the data set into MySQL? The first thing you must do is to take the file provided to you and using your knowledge of MySQL and database systems, create the database with all of its tables and data on your database server. Once it is on your server, using your knowledge of MySQL, create an ERD (Entity Relationship Diagram) of the database.

Once you actually have the database with its tables and data in MySQL, along with the ERD, it’s time to utilize MySQL statements to do your analysis. Keep track of these statements (save them in .SQL query files) because you will need to include them in your final report. You should not just be using basic SELECT statements for your analysis. For example, you must use WHERE clauses, JOIN clauses, and ORDER clauses. Your project should be more complex than that. *Your final results for each question you are trying to answer should be inserted into a separate results table*.

**Final Report**

Your final report should include three files. One must your ERD of the database. The ERD must include a view of the tables, their respective attributes, keys and datatypes. The suggested file you submit is the Workbench file that holds your ERD. Another file should be a .zip or type of compression file that contains all of your SQL code. The other should be a short report (2-4 pages) explaining your project’s goals, its results, and how you used MySQL. Be sure to include your results tables in the report, nicely formatted for human consumption.

**Grading**

Grading is based on the following criteria:

**Database Design & Analysis (12 points)**

* Did you select questions for analysis that meet the criteria of the questions asked for in this assignment?
* Does your code support providing information to help answer your questions?

**SQL (6 points)**

* Did you properly use SQL syntax?
* Was your SQL of a clean style?
* Was your SQL of the sophistication necessary to complete the analysis?
* Did it meet the criteria laid out in this document?

**Report (2 points)**

* Did your report coherently explain your project?
* Does your report show an understanding of database design principles and MySQL?
* Does your report meet the requirements outlined in this document?