

# BUAN 6320

## Database Foundations for Business Analytics

### Project 2

Project 2 includes 2 steps as follows. For each step follow the description and requirements of the step and provide your answers/results in the report.

- The report should not be longer than 15 pages.
- Make sure to document every step you take and explain what you have done.
- Document all your SQL queries in the appendix of the report (wherever applicable)

#### Step 1: Design a Database [20]

In this step you should design a database given the business requirements (discussed in step 2 of project 1) and your findings (from step 3 of project 1). Follow the same steps discussed in the class for the database design process.

Please make sure to follow all the following steps one by one and perform the tasks as described in the slides in the class. Document everything that is asked for and explained in the class. Each bullet point requires its own section in your report.

1. Requirement Analysis (this is already done in step 2 of project 1)
2. Data Understanding (this is already done in step 3 of project 1)
3. Schema Design
  - a. Find entities, their attributes, their primary keys, and relationships between them.
  - b. Model all the constraints you believe should be there in your schema.
  - c. Draw an ER diagram of your dataset.
  - d. Translate your ER diagram into relations .
4. Schema Normalization
  - a. Find all the functional dependencies you can from your schema and list them.
  - b. Check if the keys you have chosen for your relations are minimal (prove)
  - c. Check if your schema is in BCNF (Boyce-Codd Normal Form) (prove)
  - d. If your schema violates BCNF, bring it to BCNF by decomposing it (show)
  - e. Update your ER diagram with the latest schema.
5. Create your database in MySQL using the latest version of your schema.
6. Import the data into your database.
  - a. If there are errors while importing, document these errors in your report and mention how you dealt with them.

#### Step 2: Data Cleaning and Database Testing [10]

Now that you have your database ready and the data in it, you can start working with the data and querying it.

In this step you will investigate the data quality and deal with what you find. Your data might need some data cleaning which you will take care of in this step.

For every task you perform in this step please provide the SQL query in the appendix of the project. Note that all the tasks in this step should be done using SQL only.

- For each table in your database, check all the columns and the values they contain.
- For numeric columns, provide all the statistics, check them, and see what you find.
  - o Check them against the information you found in step 3 of project 1.

- You should be looking for missing values, values that seem to be outliers (typically far away from the mean), or data errors or any values that does not seem to be valid (like a typo)
- Make sure all the values of these columns are from the same type (all numeric)
- Document the problems you find; fix them and explain how you dealt with them.
- For character columns, provide frequency table of some of the values they contain, then do a quick check on these values.
  - Check them against the information you found in step 3 of project 1.
  - You should be looking for missing values or data errors or values that does not seem to be valid (e.g., sometimes there are white spaces in some of the cells either before or after the value)
  - Make sure all the values are from the same type and domain.
  - Document the problems you find, fix them, and explain how you dealt with them.
- Try to query your database especially using more than one table (by joining them) to see if the results make sense or not.
  - Try at least 3 join queries to see if the joins are working properly. Explain how you believe the joins are working. Check if the results of these queries match what you expect.
  - Try at least 3 queries with regard to the constraints you have in your database to see if those constraints are working properly. Explain how you believe they are working.