Northeastern University
College of Professional Studies

ALY 6030 Data Warehousing and SQL Module 2 Assignment

Class ALY6030 – Data Warehousing and SQL

Module 2 Assignment: Data Modeling and Basic SQL

Overview and Rationale

In this assignment you will design ERD and normalize database from the flat file. You will need to identify at least 3 business questions and provide answers using SQL by joining the data from multiple tables.

Assignment Instructions:

Part 1: Database Normalization

- Select one of the following datasets for this assignment:
  - airport.csv
  - cars.csv
  - ebay.csv
  - movies.csv
  - sacramento.csv
  - techcrunch.csv
- Design a normalized database (up to 3NF) using the selected dataset as a base.
- Create at least 5 tables in total holding the data based on the data set selected.
- Include SQL code, screen shots, results and explanations.
- Create Database ERD using MySQL Workbench.
- Include screen shot of the complete schema into your paperwork.
- Create relationships between tables using MySQL Workbench.
- Define primary keys, foreign keys, constrains, etc. using SQL.
- Include screens shots and explanations for each entity, table, attributes, relationships

Part 2: SQL Query Design

- Design at least 3 SQL SELECT queries that would return data from multiple tables (at least 3 tables in each query) of your designed database.
- Explain what each query is for and what results it returns
- Include screen shots and explanations for each screen shot.

Northeastern University
College of Professional Studies

ALY 6030 Data Warehousing and SQL Module 2 Assignment

Submission requirements:

- Submit paperwork using .docx file in Microsoft Word format named as ALY6030_80528_Assignment_2_[your name] .docx
- Submit SQL script .sql file in text format names as ALY_6030_80528_Assignment_2_[your name].sql

Paperwork file must meet the following requirements:

- Title Page must have:
  - Class name
  - Class number
  - Class CRN number
  - Assignment name
  - Your full name as registered in Canvas
- Introduction section
  - explain what this assignment is about
  - set goals and expectations
  - explain the selected dataset
  - introduce tools and languages you plan to use
- Final conclusions section
  - Explain if goals and expectations were met or not
  - Explain cons and pros of using the tools, methods and techniques
  - Explain what would you do differently
  - References (optional)

SQL Script file must meet the following requirements:

- Only .sql text format will be accepted.
- Each line of the SQL code must have a meaningful comment.
- Code should be executed on other computers without any modifications (add extra steps in the comments if necessary).