Individual and Group Project Guidelines

Overview and Rationale

This course is designed to make you a WELL-ROUNDED analytics and informatics professional, with the expectation that you possess these skills, and make you an "agent of change" in whatever business environment you may find yourself in, in the future. This course requires you to execute two projects: one at individual level called *Signature Project* or *Individual Signature Project*; and group level called Group Project. Your projects will be graded not only on the technical accuracy of the solution, but equally on your ability to communicate your thoughts, process, ideas, goals, and challenges.

It is expected that some students will have advance level of technical skills (analytics and informatics) prior to attending this course and some may have basic technical skill due to their prior educational and professional background. Students who are lacking in technical skill should review the lecture videos, consult the TA, and consult the student experts in NU campus available through Wyzant Tutoring Services for students of College for Professional Studies[this platform is accessible through Canvas]. This is very critical, especially for students with lower prior technical skills because they could easily become burden for their group project and may not be able to produce high quality solution for signature project.

Learning Connection

The goals of these projects are to prepare you not just from a technical perspective (i.e. the ability to use database tools, SQL, etc.) but to develop skills in effective writing and communication. You can be the most technical person in the room, but if you cannot describe your thoughts, goals, and challenges effectively then you will not be successful in the "real world." Northeastern University has many resources to help you develop your writing and communication skills.

In addition, the following course outcomes will be addressed:

- Examine the role of databases and database applications in contemporary organizations
- Model data using entity relationships models (i.e., design and implement ERD)
- Enable business use-cases by of information retrieval to improve business processes (i.e., design and implement SQL queries)
- Analyze the impact that database design has on application program design and user interface (e.g., data quality constraints, predefined reporting capabilities, user-defined reporting).

Project Assignment Overview and Mechanism

For your signature project, you will identify a real-world business need, preferably based on some publicly available dataset. Find dataset(s) that reflect different types of data (numeric, qualitative, text, logical) and observations on subjects of different types [if needed think about multiple sources of datasets that can be combined in some way to represent richer context]. The assumption here is that the database application to be developed is expected to solve related business problem, and the existing datasets offers you some indication of underlying business processes and its contextual environment.

Caveat: The business problem context and database solution SHOULD NOT be exactly similar to the example cases covered in the course such as ERDs shared in lecture slides, case studies described in the Harrington (2015) or any other reference materials listed in the syllabus. For example, you can't define your project same as sales databases for Chinook, LG Sales, SmartMart; inventory

Individual and Group Project Guidelines

management databases like MightMiteMotors, East Coast Aquarium; and production management database like academic journal publication {all are used as reference database in this course}. However, you can conceptualize a conceptualize a business problem in the same domain that requires significantly different database solution even though some of the entities will overlap due to commonality of the business context.

Signature Project: The signature project should be executed in four phases spread over six weeks with three incremental milestones and the final phase of project completion. Students are expected to develop the individual project every week and make sufficient progress toward its completion. For each phase of the signature project:

- Preliminary work deliverables are peer reviewed by at least three (randomly assigned) fellow class members, and the teaching assistant (TA).
- Students are expected to incorporate feedbacks from peers and TA in finalizing their project deliverables before submission for evaluation to instructor.
- For each milestone deliverables, the instructor will provide written feedback in the Canvas to improve the quality of database solution.
- Subsequent to instructor's evaluation of project deliverables, students are expected to incorporate changes to address concerns raised by instructor during the next phase of project.
- Students are also encouraged to seek feedback from instructor during development of preliminary solutions during online office hours.

Group Project: This course requires students to develop database solution for a common business context specified in the group project context section of this document. All groups are expected to develop their own version of database solution for the specified problem. The business problem is of adequate complexity that would require a group of 3-4 members to develop high quality database solution. The group project should be executed in three phases spread over three weeks (starting from weekly module-3) in two incremental phases and final phase of project completion. For each phase of the group project:

- Students are expected to leverage the experience gained in the signature project and individual assignments to solve the Group Project problem.
- There is no peer review feedback process for group project.
- Students are encouraged to seek feedback from instructor during development of preliminary solutions during online office hours.
- Instructor will provide written feedback on the deliverables for each milestone. Students are expected to incorporate necessary changes during next phase of group project.

Final Phases of Signature and Group Projects: In the final phase of each project, students are required submit full project report based on the template prescribed in the next section. Additionally, students will prepare presentation slides and video record their presentation for Signature Project and Group Project.

- During last week, reports and presentations will be peer reviewed by randomly assigned peer students outside their groups.
- There will be no in-class presentation of projects. Instead,
- On the final class day, peer reviewers will make solution improvement suggestions after consolidating feedback from members of peer review groups.

Individual and Group Project Guidelines

Project Context: Individual Signature Project

Every student will identify their own project problem. This individual project is considered as signature project of the course and aims to enhance your database design and implementation skills in conjunction with improving your presentation skills and domain knowledge of the chosen project topic.

Potential Data Sources

The choice of Individual Project could be based on an existing publicly available dataset (e.g. Kaggle; UCI machine library; WHO: www.who.int; CDC www.cdc.gov; US Department of Health and Human Services www.hhs.gov; open data initiative of federal and state governments: www.data.gov). This listing is not exhaustive. You can explore on your own and find other datasets.

- Be sure that the chosen dataset(s) has sufficiently large number of data columns such that moderate number of entities can be identified (say 5 -15 entities). However, it does not imply the dataset will have 5-15 columns. Remember that individual columns in the dataset mostly represent attributes of some underlying entity.
- Alternately, you can choose a business context of your liking that has personal appeal to you and create synthetic dataset yourself. However, you are encouraged to find some publicly available dataset for the chosen problem context so that database can be populated with such data.

Individual Signature Project Phases and Milestones

The project consists of four phases with three initial milestones, and the final phase of presentation and report. At each milestone, students are expected to generate cumulative report by documenting the relevant sections as described below.

Milestone-A: Project Proposal & Preliminary Database Design

Note: *Applicable Modules: Module-1*

This phase requires completion of multiple tasks including:

1. Identify the overall business context and justification for the need to develop a database solution.
2. Identify relevant data sources from the web for sample data collection (or generation if not available).
   a. Sample datasets collected from web may have significantly large number of data columns.
   b. When selecting dataset from web you must consider if multiple entities can be identified from dataset. Do not select dataset that can't be converted into a meaningful relational database.
   c. Your database design solution need not capture all the columns of the datasets collected.
   d. It would suffice to design database solution with respect to a smaller set of attributes.
3. Define business rules and use case scenarios (functional requirements) for the database. The list of business rules should be relevant for database design and implementation. Do not list any tangential rules covering business ethics, and corporate governance.

Individual and Group Project Guidelines

    a. The scope of the use case scenarios for the business context should be exhaustive enough so that it leads to an ERD of at least 5, but no more than 10 entities in the design [reduce the scope of the project if it needs more than 15 entities or alternately merge some entities – i.e., compromise on data redundancy].

    b.

4. Develop a preliminary design of the database in the form of entity relationship diagram using the desktop or web version of the Draw.IO app. This is a multi-step process.

    a. The initial draft of ERD along with necessary documentation of project proposal is shared to the class for peer feedback using Discussion board.

    b. Randomly chosen peer students and TA will provide feedback using the Peer-Review functionality of Canvas as part of weekly discussion assignments.

    c. Subsequently, the ERD and project proposal are revised for submission to the instructor.

5. Feedback from Instructor must be addressed by revising the database design during the next phase.

Deliverables: Preliminary ERD image file; sample dataset as CSV files; Report with documentation for Milestones A. Do not zip them.

Milestone -B: Database Design, Implementation, and Basic Queries

Note: *Applicable Modules: Module-1, -2 & -3*

This phase requires finalizing the database by designing the full ERD (in Draw.IO app), implementing the database as SQLite database, and writing several basic SQL queries for decision making purpose. This phase requires completion of multiple tasks including:

1. For finalization of database design, the ERD revision must incorporate feedbacks received from the instructor and your peers and TA (if not already done in the previous phase).

    a. The number of entities in this final version may increase from the initial version of the reported design in previous phase.

    b. All concerns raised by instructor must be addressed by revising the ERD. Any unaddressed concern must be justified in writing.

2. Next, implement the finalized design in the **SQLite** relational database management system using RStudio (preferred) or DB Browser.

    a. For implementation of the database in RStudio the Rmd or sql file must be written in such a way that it's execution as a full file will generate the desired database.

    b. For implementation of the database in DB Browser, all tables should be created by specifying all relevant attributes and their data types, and all applicable constraints.

3. Subsequently, write a SQL script file (as Rmd file for RStudio; or in the SQL tab of DB Browser) to populate the database using the reference sample dataset collected. At this stage, it is possible that not all tables will get populated. However, it is strongly recommended to populate all tables in the database with sufficient number of records to generate meaningful basic queries.

Individual and Group Project Guidelines

      a. In order to avoid integrity alerts/ errors, all relationship can be specified with zero minimum cardinality (i.e., all relationships are optional). This will allow you to populate each table separately without getting referential integrity error alerts.

      b. The SQL statements for populating different tables must be written in such a way that execution of the full file populates the database without any error.

      c. The DB Browser users should save the SQL script as a separate file for submission.

4. Finally, write another SQL data query file (.Rmd file) to demonstrate meaningful data can be extracted from single or multiple tables with and without performing column transformations. These basic queries should satisfy some decision making needs for the organization instead of simply validating your data population process executed previously.

      a. Execute and evaluate the queries to ensure quality and generate the HTML output.

      b. Write your inferences of the query results.

      c. The DB Browser users should save the SQL script as a separate file for submission and paste screen shots of result window to report.

5. Submit all deliverables to instructor for evaluation.

      a. Instructor will give written feedback that must be incorporated during next phase.

Deliverables: revised ERD image file; database file populated with sample data; SQL scripts of data population and data queries; revised Report with documentation for Milestones-A& B including query results and your inference. Do not zip them.

Milestone -C: Database Queries (Intermediate SQL queries)

Note: *Applicable Modules: Module-1, 2, 3 & -4*

In this phase define use-cases for several queries, i.e., specific context of data retrieval for the end users. These queries could be of sufficiently complex in nature requiring the use of SQL features covered in the module-4. This incremental phase requires completion of multiple tasks including:

1. Revise the database design (ERD), and schema file to incorporate changes to address instructor's concerns raised for deliverables of previous phase.

2. Revise and execute the SQL script for data population to insert full data previously collected or at least ensure that sufficiently large number of observations are inserted to implement advance queries.

3. Develop SQL query solutions for the use-cases you identified for this phase.

      a. Write the necessary SQL script using the R/RStudio or DB Browser.

      b. These queries may require joining multiple tables, transformation of data columns, conditional queries, use of regular expression, and union of multiple queries.

      c. Evaluate the queries to ensure it produces desired output and generate the HTML output.

      d. Write your inferences of the query results.

      e. The DB Browser users should save the SQL script as a separate file for submission and paste screen shots of result window to report.

4. Append the project report with the list of query use cases, their description, and SQL code. Also show the result of executed query on the sample dataset.

5. Submit all deliverables to instructor for evaluation.

a. Instructor will give written feedback that must be incorporated during next phase.

Deliverables: revised ERD image file; database file populated with full data; SQL scripts of data population and data queries; revised Report with documentation for Milestones-A, B & C including query results and your inference. Do not zip them.

Individual Signature Project Report and Presentation [includes advanced complex queries]

Note: *Applicable Modules: Module-4 & -5* [+ module-1, -2, & -3 if changes are done in design and implementation]

In this final phase of the signature project, you should define new set of use-cases for several complex queries that require use of advanced/complex SQL query designs covered in the module-5, e.g., view based queries, ranking in a cross-sectional data, and ranking by time frame or within groups. These queries may require joining multiple tables or creation of temporary virtual tables and/ or views. This final phase requires completion of multiple tasks including:

1. Review and revise deliverables of all previous phases where necessary. Specifically, revise the SQL queries and reproduce their result (HTML output), and streamline the report by updating the sections corresponding to previous phases (milestones- A, B, C).
2. **Develop SQL query solutions** for the newly identified use-cases.
   a. Write the necessary SQL script using the R/RStudio or DB Browser.
   b. These queries may require joining multiple tables, transformation of data columns, conditional queries, use of regular expression, union of multiple queries, creation of views or temporary virtual tables (i.e., WITH clause).
   c. Evaluate the queries to ensure it produces desired output and generate the HTML output.
   d. Write your inferences of the query results.
   e. The DB Browser users should save the SQL script as a separate file for submission and paste screen shots of result window to report.
3. **Signature Project Report:** update and complete the project report to include description of deliverables generated in this final phase.
   a. Ensure to review and revise the content of previous phases where required especially if instructor has raised concerns on the quality of project deliverables in the previous phase.
   b. Append the project report with the list of query use cases, their description, and SQL code generated in this final phase. Also paste the snippet of query result and add your inferences of the query results.
   c. Remove any unnecessary text to shorten the final report while maintaining the quality of report.

4. **Signature Project Presentation:** Create PowerPoint slides to present your project to the class and create a video recording of presentation using slide show recording functionality of PowerPoint.
   a. Develop your slides with only necessary text and images.
   b. Presentation slides should include context for explaining the business context, database design (ERD), data preparation (what you did to prepare the data for populating the database), selected

Individual and Group Project Guidelines

        uses-cases of business queries and their SQL implementation, query results and inferences, and concluding slide.

    c. The conclusion slide should emphasize on what future changes can be done to improve the database solution, and what key lessons you have learned from this project.

    d. To record your presentation, see documentation at this web page: https://tinyurl.com/y3v8pyqa

    e. Your presentation video should not be longer than 10 minutes.

5. Final report and presentations of signature project will be peer-reviewed. Reviewers' grades for final report & presentation of signature project will be partially determined based on the quality of anonymous feedbacks they provided to (randomly assigned) students.

Deliverables: Final versions of project Report, ERD image file, database file populated with full data, SQL scripts of data population and data queries; PowerPoint slides; Video recording of project presentation. Do not zip them.

Project Context: Group Project

All groups are expected to solve a commonly defined business problem as described below. The group size (typically 3-4) and membership will be decided by the instructor depending on the class size. Note that group membership may change during first week because students tend to drop/join the course.

Group Project Execution Mechanism

Group project will also follow similar execution mechanism as the signature project, except that the group project will have only two incremental milestone phases, and final phase of report and presentation.

1. All groups should execute their projects by coordinating their efforts.
   a. Each member is responsible for contributing to the project in a professional manner.
   b. If a member lacks technical skill, then he/she must seek out help from other group members. However, they should not become burden on the group by always depending on the help. If necessary, such individuals should seek out tutoring by using the Wyzant Tutoring available to NU students through Canvas.
   c. Group member possessing higher technical skill should provide technical consultation to other group members who lack necessary technical skill.
2. Instructor has randomly assigned one person as group leader who will be responsible
   a. For uploading group level deliverables to the Canvas.
   b. For communicating with TA and instructor for clarifying project related concerns.
3. Every member should contribute to the design and implementation of the database solution.
4. Each group member should obtain feedback on their deliverables from other group members. Each group is responsible to perform peer review within the group to ensure high quality of database solution.
5. There will be no peer-review of the incremental deliverables of group project during the early phases. However, TA will review the deliverables and provide written feedback. Students are allowed to revise the deliverables before instructor evaluates them.

Individual and Group Project Guidelines

6. Instructor will provide written feedback for each incremental milestone phase of the project to improve the database solution quality.

7. Final report and presentations of group project will be peer-reviewed. Reviewers' group project grades for final report & presentation will be partially determined based on the quality of feedbacks they provided to (randomly assigned) groups.

Business Context of Group Project: New Age Research Foundation (NARF)

A leading multi-billionaire has recently launched a non-profit New Age Research Foundation (NARF) with initial outlay of $150 billion to promote basic and applied research to tackle challenging problems facing the society locally and globally. NARF is expected to generate funding capital by engaging other philanthropic multibillionaires.

NARF has subcontracted your team to build a Grant Management App that leverages open-source technologies (e.g. SQLite for database) to manage their grant funding processes. Note that this database will handle only the funded projects. Digitalization of grant submission, review, and approval processes are beyond the scope of this suggested project.

NARF will have different types of grant mechanism of different funding level (e.g., NIH has categorized research (R type) and career development awards (K type) grants using three character codes such as K01, K02, K99/R00, R01, R03, R18, etc. Check out the grant mechanism description page here. You DO NOT need to explore other grant mechanism types such as training grants, resource grants, center grants, and Trans-NIH grants.

NARF will fund research teams that could be affiliated with a single or multiple organizations (academic, private, non-profit) anywhere in the world. The funded teams are expected to disseminate their research findings through publications (journal articles or reports), news articles and/or news media coverage. During the grant period, teams are expected to provide annual updates to NARF on their progress for continued funding support if projects were initially approved for multi-year funding. Additionally, the research could lead to patents.

NARF wants you to study the NIH grants management portal and develop the GMA solution. As a starter, your team should review the NIH Reporter portal, and examine the data fields available on various tabs of the search portal

- To get the default search result, click the "Search" button without entering any value in the search form. The result page will display multiple tabs.
- You should review the tabs: "Projects", "Publications", "Patents", and "News & More".
- You can ignore the "Clinical Studies" tab because implementation of that would require more complex & flexible design beyond the scope of this course.
- Next, click few project number links (first column of search result) and review their details. Note that the main search result gives a button link to view similar projects, and the individual project page lists those similar projects under the respective section.

The group project should be executed in two incremental milestone phases and a final phase of reporting & presentation.

Individual and Group Project Guidelines

Milestone-A: Preliminary Database Design and Implementation

This phase combines the expectations of milestones-A and -B of signature project. The only exception is that student groups are given a common business context to develop a database solution.

1. Review the description given in the section of Signature Project in previous pages to understand what is expected for database design and implementation.
2. Each group member should develop their own version of ERD for the project. Thereafter, group should conduct within-group review of individually prepared ERDs to create and finalize ERD for the group.
3. Coding effort for schema to implement the group's ERD should be shared among group members. Subsequently, the script should be reviewed, revised, and consolidated as a single database schema file.
4. Collect and prepare the necessary sample data by performing appropriate search at NIH Reporter portal. Download the relevant search results from NIH reporter as Excel or CSV file.
   a. To download click the Export button at top right corner of the search result page.
   b. Perform necessary data cleaning in the spreadsheet software.
   c. Arrange the data to match the table structures mapping to the finalized preliminary ERD.
5. Every group member should write data population script for different tables. Subsequently, the script should be reviewed, revised, and consolidated as a single data population script file.
6. Write few basic queries to validate the data inserted in the database.
7. Write the preliminary project report documenting deliverables of this phase. Conduct a within group review of report and revise, if necessary, before submitting for evaluation by Instructor.
6. Feedback from Instructor must be addressed by revising the database design and implementation scripts during the next phase.

Deliverables: Preliminary ERD image file; SQL scripting files (.Rmd or .sql extehsion) for database schema and data population {these files should include comment lines indicating author's name for respective code}; database (.db) file with populated data; Report with documentation of Milestones A. Do not zip them.

Milestone -B: Database Queries Implementation (Basic and Intermediate Queries)

This phase combines the expectations of milestones-B and -C of signature project. During this phase:

1. Group should revise their project deliverables from previous milestone to address concerns raised by the TA and Instructor.
2. Identify use-cases of business queries meaningful for management of NARF grant funding processes. The number of use-cases should be sufficient enough to allow each member to implement few SQL queries.
3. Each member should implement SQL query solutions for few use-cases and execute them.
   a. Review the queries within group and finalize them to generate correct outputs.
   b. Execute the finalized version of queries and generate HTML output files.
4. Append the project report with documentation of deliverables produced in this phase including snippets of query results and your inferences.
5. Submit the interim report along with deliverables for evaluation by Instructor.

Individual and Group Project Guidelines

Deliverables: revised ERD image file; database file populated with full data; SQL scripts of data population and data queries; revised Report with documentation for Milestones-A, & B including query results and your inference. Do not zip them.

Group Project Report and Presentation [includes advanced complex queries]

For this final phase, follow the same processes as described for the Report and Presentation section of Signature Project in previous pages.

Specifically, the group should define new set of use-cases for several complex queries that require use of advanced/complex SQL query designs covered in the module-5, e.g., view based queries, ranking in a cross-sectional data, and ranking by time frame or within groups. These queries may require joining multiple tables or creation of temporary virtual tables and/ or views.

1. Complete the SQL query scripting for newly identified use-cases and generate HTML outputs. Each member should write at least one of these complex queries.
2. Complete the final report for group project.
3. Prepare a PowerPoint slides as a group highlighting the database solution developed in this group project.
4. Video record the PowerPoint slide presentation as a group. A recommended strategy would be to conduct an online meeting using Zoom or MS Teams (that allows screen sharing) to record presentation.
5. The video recording of presentation should be no longer than 20 minutes, and every group member must have presented at least few slides.
6. Submit the final report and group presentation video recording for evaluation by randomly assigned peer groups and the instructor.
7. Each group will peer review 2-3 randomly assigned groups. Reviewers' group project grades for final report & presentation will be partially determined based on the quality of feedbacks they provide to peer groups.

Individual and Group Project Guidelines

Report Format: Individual Signature Project and Group Project

Prepare brief project reports for both individual project and group project. The report MUST use proper English grammar and be CLEAR and CONCISE. Note that this is NOT a typical "more is better" project. Be clear, be precise, and get your points across. Less is more!

Your future supervisors, co-workers, and other business stakeholders do NOT want to read a dozen pages of 'fluff', nor do they need an exhaustive history or business analysis. This report is NOT about designing a new innovative solution or writing long-winded market analysis documents.

These project reports are live documents and should be submitted at the end of each milestone phase representing progress toward project completion [check for latest due dates on Canvas]. At each incremental phase (i.e., module), the report should be updated with the latest progress accomplished.

Note that the final version of the reports upon completion at the end of term is expected to be about 7-10 pages long and the text content should be approximately 1500 words maximum {excluding title page, appendixes, bibliography, images, and tables}.

- The recommended Font Type/Size: Arial 11 or Times New Roman 12 point; and Line-spacing: 1.2 lines.
- All images and tables in the report should be numbered and appropriate title should be written. Place the number and title at the top for tables , and at the bottom for images {this is a common convention in any report writing}.

Report Template

This section describes the recommended template for the full report applicable to both Individual Signature Project and Group Project. For both reports, the main section and sub-section headings of the report MUST have appropriate Heading 1 style (e.g., this document has dark green background and white font) and Heading 2 style (e.g., this document has dark red font without background color). You are free to introduce additional subsections (with Heading 3 style) if your reporting context requires.

The layout of the report template includes the following [in this template, the section headings of Report template have Heading 2 style but in your report they should have Heading 1 style]:

Title Page

This page contains project's title, Project Type (individual or group), Course number and name. Then after few blank lines, write author's name, and the submission date (as last line of the page). DO NOT write the project's phase name on the title page. DO NOT include instructor's name in your reports.

In the case of Group project, write your Group Number, and members' name in alphabetical order.

Individual and Group Project Guidelines

NOTE for NU logo & Seal positioning: if using NU logo (full name and seal) then place it at the page top and align center [height <1"]. If using the NU seal only, then place it at the page bottom align center [max size: 1.5"x 1.5"]

If using this, place it at the top of the title page and align center [height <1"]

If using this, place it at the bottom of the title page and align center [max size: 1.5"x 1.5"]

Introduction { should use Heading 1 style}

Describe the business challenge, use-case scenario, and the application solution.

Individual Project: Provide the reader with a concise yet effective overview of the topic you've selected. Explain the significance of the project, and why you chose it.

Group Project: In the case of group project a common problem statement has been provided. Based on the given description write this section.

Database System Analysis and Design { should use Heading 1 style}

User Profiles (or Personas) { can be Heading 2 style}**:** Briefly describe the profile of end-users who will use the database solution and in what way they are going to use. The goal is to ensure that you, as the database designer, can show that you truly understand WHO uses the application, and WHY they use it, and HOW they use it. A good idea would be to identify these roles from top management, middle management, and lower (operational) management layers of an organization, or some in some context customers.

You should describe two to four different user profiles (personas) with one paragraph each.

Business Rules { can be Heading 2 style}**:** Briefly explain the important business rules or logic that the database will use or enforce using well-written bullet-points (as many as you need). DO NOT write long paragraphs. Each rule should be written separately. Remember, in the real world environment, you (as the database developer) MUST understand the expected business functionality and be able to reflect that understanding back to others. Before you start building tables to store data, you MUST understand all of the ways that data will be created, stored, and retrieved.

Database Design { can be Heading 2 style}**:** Identify and describe a reasonable number of data entities (tables) for your project. There should be at least 5, but no more than 15 entities in the design [reduce the scope of the project if it needs more than 15 entities or alternately merge some entities – i.e., compromise on data redundancy]. Identify all attributes related to each entity and their data type and size constraint for their respective data field.

Individual and Group Project Guidelines

Identify all "association entities" needed to normalize the database to 3rd normal form [splitting the M:M relationship]. Identify all entity relationships, primary and foreign keys for entities.

Create a professional looking ERD using the Draw.IO app. Ensure that:

- Entities and attributes follow standard naming conventions
- Primary keys (PK) and foreign keys (FK) are clearly designated.
- Relationship lines must connect PK and FK fields.
- Minimum and maximum cardinality of relationships are shown using Chen's Crow foot convention.

The ER diagram must be on its OWN page (insert the image version obtained from Draw.IO). A separate image file of ERD should also be uploaded to Canvas.

Database Implementation { should use Heading 1 style}

Database Schema { can be Heading 2 style}**:** Write database schema using R/RStudio [or DB Browser] and SQL data definition languages (e.g., CREATE, ALTER, and DROP statements). Populate the database using SQL data manipulation language (INSERT, UPDATE, and DELETE statements). If sample dataset is not available for your project, then use fictitious data.

Data Query Implementation { can be Heading 2 style}**:** Specify a list of business queries as derived from the use-case scenarios discussed above. If needed, add new use case scenarios to the "Introduction" section to match the newly identified queries which do not correspond to any of the existing use-case scenario. In the real-world settings, many use cases may be reflected in the data queries instead of database design as entities or attributes.

Write the SQL statements required to implement the listed queries. Your queries MUST be of sufficient complexities such that

- It requires joining two or more tables
- It includes complex filters that represent real-world use-cases.
- At least one-two queries require creation of "views" or nested queries
- At least one reporting query with data aggregation over time and/ or group characteristics.

Analytics and Reports { should use Heading 1 style}

Provide examples of what analytics which could be conducted on your database to generate what reports for end-users. They could make use of one or more queries implemented previously.

Execute all the queries on the sample database. Show the query results [no more than 20 rows]. Full query results should be shown in the HTML output generated from .Rmd files.

Conclusions and Future Considerations { should use Heading 1 style}

Describe what you have learned from this project, and what key takeaways for future use.

Bibliography { should use Heading 1 style}

List all references using APA citation style.

Appendixes { should use Heading 1 style}

Add all appendixes as required. Each appendix must be numbered and cited in the report.

Deliverables and File Naming Conventions

All files including this report (.docx), ERD image file, source code files (.Rmd file), database file (.db), and PowerPoint files (.pptx) using following naming conventions:

- **LastNameFirstName_ProjectType_DocumentType.ext** {for individual signature project files}
- **Group_N_ DocumentType.ext** {for group project files common to group}
- **Group_N_ LastNameFirstName_DocumentType.ext** {for individually authored documents such as SQL files of group project}

Replace the "YourLastNameFirstName" in the file name with your last name and first name; the ProjectType by Individual or Group as applicable. Next, replace the phrases DocumentType by Report, ERD, DataSchema, AnalyticsSQL, and Presentation for each of the respective files. Replace "N" with your group number for group project related files.

Academic Integrity in Project Reporting

Academic Integrity is a CRITICAL component of this course. You should NOT be copying-and-pasting anything into your paper and trying to pass it off as your own work. Limited instances of copy-and-paste (2-3 sentences long subparagraphs) followed by an APA citation is acceptable. Copying-and-pasting full (paragraph length or longer) content within double quotes and then putting a generic link in your REFERENCES section is a violation of Academic Integrity, and may result in reduced points, or in cases of egregious plagiarism, a FAIL (F) for the entire course. When in doubt, ask!