

## BUAD 5272: Database Management

Term: Fall 2020

### Team Project Instructions

---

#### Project Objective

---

The purpose of the project is to give students the experience of developing a working database for analytics, conducting ETL to load the data, writing SQL queries and visualizing the results.

---

#### Task

---

We are utilizing data provided to William and Mary Mason School of Business. These data are from Medicare Shared Savings Program (Centers for Medicare & Medicaid Services) and is provided to us by Bon Secours.

According to the Centers for Medicare & Medicaid Services (CMS), the Medicare Shared Savings Program (MSSP) aims to encourage coordination and cooperation among providers to improve the quality of care for Medicare Fee-For-Service (FFS) beneficiaries and reduce unnecessary costs. MSSP rewards Accountable Care Organizations (ACO) that lower growth in health care costs while meeting performance standards on the quality of care. Provider participation in the program is purely voluntary, and Medicare patients can seek treatment from any provider they wish.

Bon Secours has provided us with two years of MSSP claims data for the Richmond and Hampton Roads areas. This claims data allows for complete, 360° view of patient encounters for Medicare patients assigned to the Good Help ACO.

The analysis you will conduct will revolve around hospital readmissions, in particular for the following diagnoses: Acute myocardial infarction (AMI), Heart Failure, Pneumonia, Chronic obstructive pulmonary disease (COPD), Coronary artery bypass grafting (CABG), Stroke and Total Hips/Total Knees. Particularly we ultimately seek to determine if there are relationships between provided variables and readmissions that could be used to create additional interventions. Particular attention should be paid to whether: 1) Current Length of Stay > 5 days and 2) They have seen their primary care physician at least one time in last 6 months 3) The number of Admissions in Last Calendar Year 4) If any of the following are true: Married, Living with Significant Other, Assisted Living, Long-Term Acute Care (LTAC), skilled nursing facility (SNF), Rehab.

Though we may not arrive at a complete answer to all of the questions in the previous paragraph by the end of the semester, your task is to use the skills you are learning in this class to organize the data so we can easily formulate these questions with queries and visualizations. Keep in mind, there is more than one correct solution/approach, and I expect most teams will come up with **quite different approaches!**

---

#### The data

---

The data dictionary provided by Bon Secours is in this link (note: *we do not have all the fields listed*). Additionally, there are five files you will be working with you can find them [here](#):

1. **CCLF1\_RI** – High level claims data for the Richmond area
  2. **CCLF1\_HR** – High level claims data for Hampton Roads area
  3. **CCLF2\_HR** - Detailed claims data for Hampton Roads
  4. **CCLF2\_RI\_2016** - Detailed claims data for Richmond, 2016
  5. **CCFL\_RI\_2017** - Detailed claims data for Richmond, 2017
- 

#### Step 1 - Conceptual design

---

Create an ER diagram (using ERDPlus) of your dimensional model. You can add data from other sources (for example you may wish to translate ICD codes).

---

### Step 2 – Logical Design

---

Map the ER diagram you created in Step 2 into a relational schema containing the relations (with their attributes) and the referential integrity constraint lines. This is your blueprint for your final database.

---

### Step 3 – Implementation

---

Examine the relational schema and based on that schema, implement the tables in MySQL. Name the database and its tables appropriately. Make sure to specify primary keys and create all the necessary relationships between the tables.

---

### Step 4 – ETL -Populate the database.

---

Identify data sources (from above and any additional data you may want to add) and create ETL plans for extracting and loading this data into your database.

---

### Step 5 - Queries

---

Create at least 10 queries for this database with varying complexities that help explain the questions posed above. Your queries can also be queries that create views that you can later visualize.

---

### Step 6 - Visualizations

---

Create a basic “Dashboard” in Tableau or Alteryx that displays and incorporates at least 3 different data representations.

---

### Project Deliverables

---

You will be providing four files. Is need to be able to fully replicate your solution on my computer:

1. Report in word/pdf:
  - i. Cover page. List the course, group member’s names and project title.
  - ii. Introduction. Describe your interpretation of the data, the two main aims of your work, a summary of your approach and a description of your conclusions.
  - iii. Modeling approach. Justification of your modeling approach. Include your ER and RS in this section.
  - iv. ETL approach. Describe your database implementation and ETL approach. Use screenshots when necessary.
  - v. Queries and visualizations. Describe your queries, visualizations and results. No need to provide the full code of the queries, you will be submitting these separately
2. SQL files:
  - i. SQL Backup code so I can recreate your database
  - ii. The 10 queries you have created
3. ETL files: all the files needed for me to duplicate your ETL process.
4. Tableau files: All files needed for me to duplicate your tableau solution
5. Conclusion. Reflect on the assignment and describe lessons learned, how you might have approached the project differently.

---

## Presentation

---

Your team will have 20 minutes to present many hours of work. Show me your pearls and don't make me shuck all the oysters! Your classmates will have gone thru very similar process so leave out simple details but rather focus on your most interesting and salient parts of your approach and solution. Some suggestions for components of your presentation:

1. An introduction section similar to what is described above.
2. The ER and RS and a description/justification of your modeling approach.
3. A description and screen pictures of the ETL processes (not all of it, just what was particularly challenging/interesting).
4. Screen shots and descriptions of the analytics based on queries and dashboard application
5. A narrative conclusion section that describes the group's experience with the project (which steps were the most difficult? Which were the easiest? what did you learn that you did not imagine you would have? if you had to do it all over again, what would you have done differently?