BAN545 Capstone Assignment

# Purpose:

During this capstone assignment you will use the knowledge you acquired over previous labs in BAN545 to build a new data warehouse using dimensional modeling

# Instructions:

The capstone assignment consists of three parts.

First, working by yourself or with a partner, you will locate a suitable dataset to use and complete a written plan describing how you plan to create a dimensional model from your dataset in Microsoft Access.

Second, you will create the dimensional model in Access, including some queries of the data.

Third, you will create a written lessons learned document which will describe any issues and potential solutions you discovered during the first and second part of the capstone assignment. Specific requirements of each part are described below.

# Part 1: Data Selection and Plan:

To begin your capstone project, you will need to locate a dataset which will allow you to create a fact table as well as three connected dimension tables. Each dimension table must have at least 3 attribute fields (not counting the primary key fields). While there are a number of places to locate datasets online, I recommend Kaggle ([Find Open Datasets and Machine Learning Projects | Kaggle](https://www.kaggle.com/datasets)). When looking at datasets, keep in mind that you will need enough data to eventually create a fact table and three connected dimension tables.

Keep in mind that you do not need to use all the data in the dataset you choose (if your preferred dataset is large, please limit the data you use to no more than 500 records). Also keep in mind that it will likely be possible that you can create additional attributes in your dimension tables not found in your dataset. The date dimension is an excellent example.

After you have selected an appropriate dataset, you will create a written plan that will explain the following:

1. A description of the dataset your will be using
2. A description of the fact table and your three dimension tables including a visual illustration of the tables listing the fields that will be used. Your visual illustrations should include the names of fields, and example data to be held in the fields (similar to the technical design process part of Lab 4)..

**You must submit your plan as an email attachment before moving on to the second part of the capstone.** You written plan must have the names of both students if you are working with a partner. **Please keep in mind that your dataset must be different from any other capstone projects in the class. In other words, if you submit a plan using the same data as a plan already submitted by other students, you will be asked to find a different dataset.**

HealthCare Record Fact Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RecordID | AdmisionDateID | PatientID | DoctorID | BillingAmount |
| 1 | 11172022 | 1 | 1 | 37490.98336352819 |

Date Dimension Table:

|  |  |  |  |
| --- | --- | --- | --- |
| DateID | Month | Day | Year |
| 11172022 | 11 | 17 | 2022 |

Patient Dimension Table:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PatientID | PatientName | Age | Gender | BloodType | MedicalCondition | InsuranceProvider | AdmissionType | RoomNo | DischargeDateID | Medication | TestResult |
| 1 | Tiffany Ramirez | 81 | Female | O- | Diabetes | Medicare | Elective | 146 | 12012022 | Aspirin | Inconclusive |

Doctor Dimension Table:

|  |  |  |  |
| --- | --- | --- | --- |
| DoctorID | FirstName | LastName | Hospital |
| 1 | Patrick | Parker | Wallace-Hamilton |

# Part 2: Creation of the Data Warehouse:

Once your plan has been approved by your professor, you will build your data warehouse in Microsoft Access. Your data warehouse will contain the fact table and three connected dimension tables as shown in your plan. You will need to use several of the dimension table creation techniques demonstrated in previous labs to build your dimension tables. You will need to create surrogate keys to connect your fact and dimension tables. After creating your tables in Access, you will load your tables using the same process you used during the labs earlier in the semester. After successfully loading your data, you will create at least 5 meaningful queries in your data warehouse. A meaningful query means you have looked at the data, and created a query which shows something interesting about the data. In other words, don’t just pick a bunch of fields and create queries without a reason. Once your have created and saved your queries, you will move on to Part 3.

# Part 3: Lessons Learned Report:

For the final part of the capstone, you will create a written lessons learned report. This report will describe any difficulties you experienced during parts 1 and 2 of the capstone as well as your solutions to the difficulties. Your lessons learned report should also address any ways in which you feel your capstone project could have been better if you could repeat the assignment.

Once you complete parts 2 and 3, you will upload your Access database file and Lessons Learned report in Canvas. If you are working with a partner, please make sure that both names are on the lessons learned report, and that only one student submits the part 2 and 3 files in Canvas.

**Lessons Learned Report:**

While I am trying to split the Doctor names into first and last names there are Dr., Mr. which are being considered as first name and the rest was split into last name so where I cleansed the raw data first then I used the cleaned data for splitting using substitute method.

While removing duplicates from the doctor list I forgot to copy as text data where it got copied as formula data which caused weird result when I tried to remove duplicates and then I got to know and copied as text and then removed duplicates and utilized the non-duplicate data for next steps.

When I am trying to VLOOKUP to look for the column 2 matching and return column 1 data which couldn’t let me do using that where I used XLOOKUP