Logo

Description automatically generated

**College of Professional Studies**

**Northeastern University San Jose**

**MPS Analytics**

**Course: ALY6030: Data Warehousing & SQL**

**Assignment:**

Module 2 Assignment

**Submitted on:**

Mar 16, 2024

**Submitted to:**  **Submitted by:**

Prof: Kayal Chandrasekaran Nikshita Ranganathan

**Introduction**

This project focuses on a comprehensive analysis in support of the initiative toward optimizing the level of nursing staff within the ICU and SICU throughout the hospital network of ACME Integrated Delivery System. The primary goal is to make data driven recommendations for pilot sites that could significantly benefit from enhanced nursing support.

**Data Overview**

The analysis will be based on three key datasets:

* business.csv: Contains information about hospitals within the ACME network.
* bed\_type.csv: Details the types of beds available, focusing on ICU and SICU beds.
* bed\_fact.csv: A fact table aggregating bed data across different dimensions, including license, census, and staffed beds.

A screenshot of a data

Description automatically generated

A screenshot of a computer

Description automatically generatedA table with numbers and text

Description automatically generated

**Introduction**

1. **Dimensions:**

**bed\_type Table:**

**Dimensions:** bed\_id, bed\_code, bed\_desc

bed\_id is the primary key (PK) and thus qualifies as a dimension. bed\_code and bed\_desc are descriptive attributes of the bed, making them dimensions as well.

**Fact:** None of the variables, as all of them describe the type of bed rather than quantifiable measures.

**business Table:**

**Dimensions:** ims\_org\_id (Primary Key), business\_name, bed\_cluster\_id.

ims\_org\_id is the primary key and a dimension. business\_name and bed\_cluster\_id are attributes that describe the business, thus they are dimensions.

**Fact:** None, as this table, would primarily contain descriptive information about businesses (hospitals).

1. **Fact:**

**bed\_fact Table:**

**Dimensions:** Primary key - Composite key made of (ims\_org\_id, bed\_id)

**Fact:** license\_beds, census\_beds, staffed\_beds - These variables represent quantifiable metrics about the beds in each hospital and are considered facts.

1. **Schema**

A diagram of a data flow

Description automatically generated with medium confidence

The star schema is a fundamental concept in data warehousing and business intelligence that organizes data into a format optimized for querying and analysis. A star schema forms a structure that resembles a star by grouping data with two major types of tables: fact tables and dimension tables.

Here at the center of the star schema is the fact table (bed\_fact), which holds transactional data with numerical metrics that are the focus of analysis—like the number of licensed, census, and staffed beds. We have the dimension tables surrounding the fact tables, each connected to the fact table by a foreign key relationship.

**4a.**

1. **Licensed beds by total ICU/SICU (Top 10)**

**Code –**

**A screenshot of a computer program

Description automatically generated**

**Result –**

A screenshot of a medical report

Description automatically generated

1. **Census beds by total ICU/SICU (Top 10)**

**Code –**

**A screenshot of a computer

Description automatically generated**

**Result –**

**A screenshot of a computer

Description automatically generated**

1. **Staffed beds by total ICU/SICU (Top 10)**

**Code –**

**A computer screen shot of a computer code

Description automatically generated**

**Result –**

**A screenshot of a data

Description automatically generated**

**4b. Top one/two hospitals on the basis of bed volume**

**Result –**

A screenshot of a data

Description automatically generated

**4b. Hospitals that appear on multiple lists**

**Result –**

**A screenshot of a medical center

Description automatically generated**

**5a.**

1. **Hospitals with both ICU and SICU (Top 10 – Licensed)**

**Code –**

**A computer screen shot of a code

Description automatically generated**

**Result –**

A screenshot of a medical center

Description automatically generated

1. **Hospitals with both ICU and SICU (Top 10 – census)**

**Code –**

**A screenshot of a computer code

Description automatically generated**

**Result –**

A screenshot of a medical center

Description automatically generated

1. **Hospitals with both ICU and SICU (Top 10 – staffed)**

**Code –**

A computer screen shot of a code

Description automatically generated

**Result –**

A screenshot of a medical center

Description automatically generated

**5b. Final recommendation**

Through the analysis conducted, the hospitals most appropriate for the pilot program are - University of Maryland Medical Center and Vidant Medical Center.

University of Maryland Medical Center and Vidant Medical Center appear three times in the top 10 listings across different categories, indicating a strong infrastructure and the capacity to handle intensive care at multiple levels.

**Conclusion**

The analysis shows us the importance of differentiating between dimensions (contextual data) and facts (measurable data), and the effectiveness of the Star Schema for data analysis. It also demonstrates how data driven decision making can identify optimal candidates for healthcare pilot programs. Lastly, these insights underline the value of a data-driven approach in strategic healthcare decision-making.

**References**

Java Guides. (2020, May 30). *Create ER diagram of a database in MySQL workbench* [Video]. YouTube. <https://www.youtube.com/watch?v=BjPsm7Ny1MA>