

Assignment 3 — Hospital Nursing Intervention Pilot Program

Heejae Roh (ID 002795339)

College of Professional Studies, Northeastern University

ALY 6030: Data Warehousing & SQL

Professor Venkata Duvvuri

December 4, 2023

Table of Contents

Introduction.....	3
Analysis.....	3
Questions to investigate	3
Data and Rationale	4
Step 1: Identify the dimensions from each dimension table	5
Step 2: Identify the Facts variables from the single Fact Table.....	5
Step 3: Sketch out a Star Schema using MySQL Workbench	6
Answer for the questions	7
Step 4a & Step 5a : Question 1 - 3.....	7
1) Top 10 Licensed beds by the total ICU or SICU license.....	7
2) Top 10 census beds by total ICU or SICU license	7
Step 4b: Interpretation of findings Question 4 - 5	8
Step 5b: Final recommendation	8
Conclusion	9
References.....	10
SQL query	11

Introduction

In this analysis, I will understand the structure of Star Schema and apply it to a real dataset. I will analyze data and answer practical business questions based on Star Schema. I will complete the analysis by looking for primary questions for business recommendations and answering them.

Analysis

Questions to investigate

- 1) Number of Licensed beds (total beds allowed by state license)
- 2) Number of Census beds (total beds at the hospital)
- 3) Number of Staffed beds (total beds for which staffing)
- 4) What are the top one or two hospitals per list based on bed volume?
- 5) Are there any hospitals that appear on multiple lists? They might make good candidates for the intervention pilot program.

Data and Rationale

1. Dataset consist of 3 tables which are bed_fact, bed_type, and business

fact_id	ims_org_id	bed_id	license_beds	census_beds	staffed_beds
1	INS00000684	5	68	51	68
2	INS00000519	2	10	7	10
3	INS00000686	5	170	25	60
4	INS00000519	5	566	394	566
5	INS00000784	16	125	54	125

2. 'bed_fact' is the table with fact (or information). 'bed_fact' table has 2,000 entries and 5 variables. 'fact_id' column is created as a Primary key

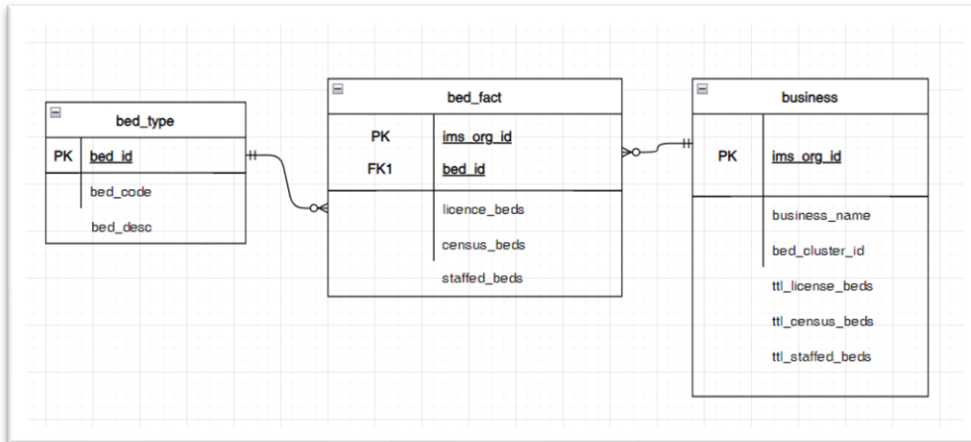
bed_id	bed_code	bed_desc
1	BU	Burn
2	CC	CCU
3	DE	Detox ICU
4	IC	ICU
5	MS	Med/Surg

3. 'bed_type' table has total 20 entries and 3 variables. 'bed_id' is added as a Primary key.

business_id	ims_org_id	business_name	ttl_license_beds	ttl_census_beds	ttl_staffed_beds	bed_cluster_id
1	INS00040779	Adams Street Place	61	61	61	1
2	INS00077200	140 Prescott Street Corporation	126	122	126	2
3	INS00000594	366th Medical Group	10	4	10	1
4	INS00074158	AdCare Hospital of Worcester, Inc	114	100	114	1
5	INS00637122	Alfredo Gonzales VA Nursing Home	160	150	160	2

4. 'business' table has total 2,000 entries and 6 variables. 'business_id' is created as a Primary key.

Step 1: Identify the dimensions from each dimension table



This graph explains the relationship between **bed_fact** table and others. The **bed_fact** table has the hospital ID called **ims_org_id** as its primary key. The primary key of the **bed_type** table called **bed_id** is used as a foreign key of the **bed_fact** table which is primary key for **bed_type** table.

The **bed_type** table consists of a term indicating the actual bed type called **bed_desc** and an abbreviation for **bed_desc** called **bed_code**. The **business** table has **ims_org_id** as PK, which is the same as the primary key of **bed_fact** table. The **business** table contains information called **business_name** and **bed_cluster_id**. The **business** table has a total number as well, but I will exclude it because the fact table is used as the **bed_fact** in this analysis.

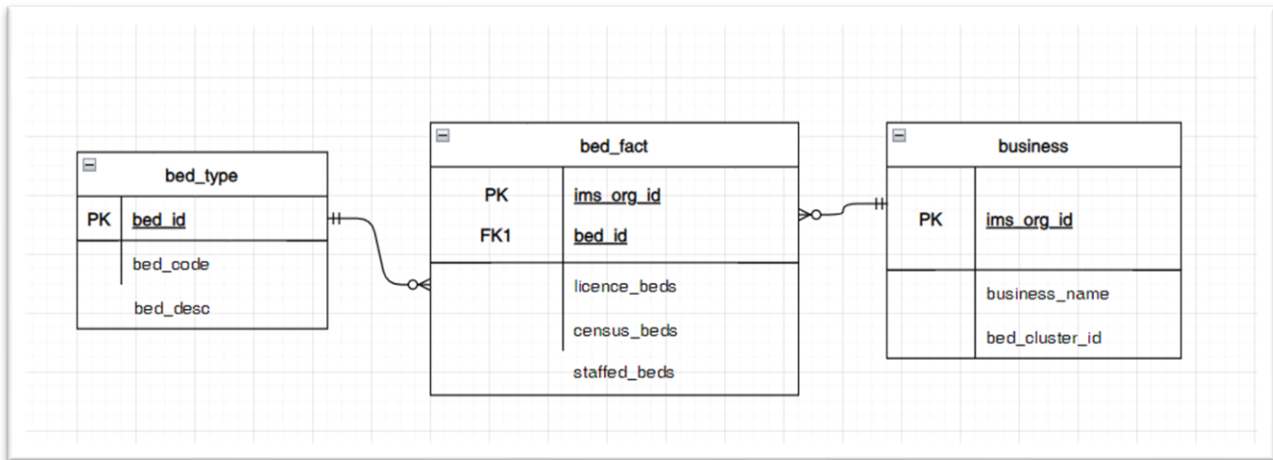
Step 2: Identify the Facts variables from the single Fact Table

SELECT * FROM bed_fact;

ims_org_id	bed_id	license_beds	census_beds	staffed_beds
INS00000784	16	125	54	125
INS00000519	2	10	7	10
INS00000784	18	429	219	429
INS00000519	5	566	394	566
INS00000784	4	33	14	33
INS00000785	18	17	9	17
INS00000785	5	17	9	17
INS00000786	5	55	18	35
INS00000786	16	59	20	37
INS00000786	4	6	2	3

The bed_fact table has bed_id as a foreign key, and license_beds, census_beds, and staffed_beds each contain the number of beds.

Step 3: Sketch out a Star Schema using MySQL Workbench



The star schema architecture is the simplest data warehouse schema. It is called a star schema because the diagram resembles a star, with points radiating from a center. The center of the star consists of fact table and the points of the star are the dimension tables. Usually the fact tables in a star schema are in third normal form(3NF) whereas dimensional tables are de-normalized (Northeastern, 2023).

Accordingly, the star shape is completed by placing only the fact table in the center and forming a bed_type table and business table on both sides that explain the dimensions of the fact table. In the business table, information other than that describing the fact table is excluded and analyzed.

Answer for the questions

Step 4a & Step 5a : Question 1 - 3

1) Top 10 Licensed beds by the total ICU or SICU license

business_name	ICU or SICU licensed beds
Phoenix Childrens Hospital	247
University of Maryland Medical Center	220
UC Health University Hospital	218
Wesley Medical Center, LLC	214
Vidant Medical Center	204
Rady Childrens Hospital and Health Center	200
Dallas County Hospital Association	195
Saint Lukes Episcopal Hospital Texas Medical Ce...	178
The Methodist Hospital	170
Emory University Hospital	169

2) Top 10 census beds by total ICU or SICU license

business_name	ICU or SICU census beds
Shands Hospital at the University of Florida	167
Dallas County Hospital Association	145
Mercy Medical Center Saint Louis	142
Los Angeles County University of Southern Calif...	139
The Methodist Hospital	138
University of Minnesota Medical Center Fairview	129
University of Maryland Medical Center	127
Brigham and Womens Hospital	124
Vidant Medical Center	123
Ronald Reagan University of California Los Ang...	122

3) Top 10 staffed beds by total ICU or SICU license

business_name	ICU or SICU staffed beds
Vidant Medical Center	203
Rady Childrens Hospital and Health Center	200
University of Maryland Medical Center	171
Emory University Hospital	169
Shands Hospital at the University of Florida	167
Mercy Medical Center Saint Louis	163
Wesley Medical Center, LLC	162
Phoenix Childrens Hospital	159
Grady Memorial Hospital	154
Los Angeles County University of Southern Calif...	151

Step 4b: Interpretation of findings Question 4 - 5

- 4) What are the top one or two hospitals per list based on bed volume?

business_name	licensed_beds	rnk_licensed	bed_type
Phoenix Childrens Hospital	247	1	licensed
University of Maryland Medical Center	220	2	licensed
Shands Hospital at the University of Florida	167	1	census
Dallas County Hospital Association	145	2	census
Vidant Medical Center	203	1	staffed
Rady Childrens Hospital and Health Center	200	2	staffed

- 5) Are there any hospitals that appear on multiple lists? They might make good candidates for the intervention pilot program.

business_name	count
University of Maryland Medical Center	3
Vidant Medical Center	3
Phoenix Childrens Hospital	2
Wesley Medical Center, LLC	2
Rady Childrens Hospital and Health Center	2
Dallas County Hospital Association	2
The Methodist Hospital	2
Emory University Hospital	2
Shands Hospital at the University of Florida	2
Mercy Medical Center Saint Louis	2
Los Angeles County University of Souther...	2

Step 5b: Final recommendation

Finally, we will look at questions 4 and 5, and recommend candidates for the program among them. First, through question 5, you can find out the names of hospitals that appear more than once in license, census, and staffed top 10. Among these hospitals, all hospitals in questions number 5 (multiple appearance in Top 10) that are in the top 2 in each license, census, and staffing that appear in the answer to question 4, and in fact are in the top 2, all appear in the top 10 more than once, so in the answer to question 4, I would recommend those 6 hospitals.

Among them, University of Maryland Medical Center and Vidant Medical Center all appear (count 3) in license, census, and staffed, so I would recommend these two hospitals as the best.

Conclusion

In this analysis, based on the fact table about hospital beds, the two dimensions, type and business (name), were organized into a star schema. Ultimately, the University of Maryland Medical Center and Vidant Medical Center, which appeared in all of the TOP 10, were included as the top recommendations. Additionally, because the hospitals that ranked top 2 in licensing, staffing, and census all appeared in the top 10 list more than once, I recommended four more hospitals as second-choice candidates. Through this analysis, I learned how to understand the structure of data through Star Schema and utilize fact tables.

References

- Zanini, A. (n.d.). Deleting a column in MySQL. DBVis Software. Retrieved from <https://www.dbvis.com/thetable/deleting-a-column-in-sql/#:~:text=Deleting%20a%20column%20in%20MySQL,the%20COLUMN%20keyword%20is%20optional.>
- Peterson, R. (2023, October 28). SQL Server Primary Key: How to Create & Add to Existing Table. Guru99. Retrieved from <https://www.guru99.com/sql-server-primary-key.html#:~:text=You%20can%20use%20the%20ALTER,drop%20and%20recreate%20the%20table.>
- Northeastern. (2023). Lesson 3.3: Data Warehouse Schema. Retrieved from https://northeastern.instructure.com/courses/160692/pages/lesson-3-3-data-warehouse-schema?module_item_id=9163762

SQL query

```

-- Setting the schemas used
USE hospital;

-- checking the dataset
SELECT * FROM bed_type;
SELECT * FROM bed_fact;
SELECT * FROM business;

-- 1) Top 10 Licensed beds by the total ICU or SICU license beds

SELECT
    b.business_name,
    SUM(f.license_beds) AS 'ICU or SICU licensed beds'
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10;

-- 2) Top 10 census beds by total ICU or SICU license

SELECT
    b.business_name,
    SUM(f.census_beds) AS 'ICU or SICU census beds'
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10;

-- 3) Top 10 staffed beds by total ICU or SICU license

SELECT
    b.business_name,
    SUM(f.staffed_beds) AS 'ICU or SICU staffed beds'
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10;

-- 4) Top one or two hospitals per list based on bed volume
SELECT
    business_name,
    licensed_beds,
    rnk_licenced,
    bed_type
FROM
(
SELECT
    b.business_name,
    SUM(f.license_beds) AS 'licensed_beds',
    RANK() OVER(ORDER BY SUM(f.license_beds) DESC) AS rnk_licenced,
    'licensed' AS bed_type
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'

```

```

GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10) a
WHERE rnk_licenced = 1 OR rnk_licenced = 2
UNION ALL
SELECT
    business_name,
    census_beds,
    rnk_census,
    bed_type
FROM
(
SELECT
    b.business_name,
    SUM(f.census_beds) AS 'census_beds',
    RANK() OVER(ORDER BY SUM(f.census_beds) DESC) AS rnk_census,
    'census' AS bed_type
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10) a
WHERE rnk_census = 1 OR rnk_census = 2
UNION ALL
SELECT
    business_name,
    staffed_beds,
    rnk_staffed,
    bed_type
FROM
(
SELECT
    b.business_name,
    SUM(f.staffed_beds) AS 'staffed_beds',
    RANK() OVER(ORDER BY SUM(f.staffed_beds) DESC) AS rnk_staffed,
    'staffed' AS bed_type
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10) a
WHERE rnk_staffed = 1 OR rnk_staffed = 2;

-- 4) Multiple appearance
WITH a AS
(
SELECT
    b.business_name,
    SUM(f.license_beds) AS 'ICU or SICU licensed beds'
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10)
,b AS
(
SELECT
    b.business_name,
    SUM(f.census_beds) AS 'ICU or SICU census beds'
FROM business b

```

```

INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10)

,c AS
(
SELECT
    b.business_name,
    SUM(f.staffed_beds) AS 'ICU or SICU staffed beds'
FROM business b
INNER JOIN bed_fact f USING (ims_org_id)
INNER JOIN bed_type t USING (bed_id)
WHERE t.bed_desc = 'ICU' or t.bed_desc = 'SICU'
GROUP BY 1, b.ims_org_id
ORDER BY 2 DESC
LIMIT 10)

SELECT
    business_name,
    cnt AS count

FROM (
SELECT
    business_name,
    COUNT(business_name) AS cnt
FROM
    (SELECT *
    FROM a
    UNION ALL
    SELECT *
    FROM b
    UNION ALL
    SELECT *
    FROM c) a
GROUP BY business_name
) b
WHERE cnt > 1
ORDER BY cnt DESC;

```