# DA5020.P2

Josh Okon, Hsiao-Yu Peng, and Chandani Shrestha

2023-11-01

```
# Loading all libraries
library(dplyr)
library(tidyverse)
library(sqldf)
```

## load csv files and inspect the data

We loaded all the fact and dimension tables as: bed_fact.csv as bed_fact_df, business.csv as business_df and bed_type.csv as bed_df. To check if we had duplicates values for "ims_org_id" in business_df, we used `distinct()` and checked the dimension of the business_df before and after perfroming that command. The evaluation suggested that there were no duplicates as there was no difference in the dimension.

```
# Reading in CSV files
bed_fact_df <- read.csv("bed_fact.csv", header = TRUE)
bed_type_df <- read.csv("bed_type.csv", header = TRUE)
business_df <- read.csv("business.csv", header = TRUE)
# check dimension before using distinct()
dim(business_df)
```

```
## [1] 22202     6
```

```
# check if there are duplicates ims_org_id in the business_df
business_df %>% distinct(ims_org_id, .keep_all=TRUE)
```

```
#check after using distinct()
dim(business_df)
```

```
## [1] 22202     6
```

## 1. Identify the dimensions from each dimension table

**You'll notice in bed_type Download bed_type there are only three variables, bed_id, bed_code, and bed_desc. Consider which of these is a fact if any and which is a dimension. Note that the PK qualifies as a dimension.For the business Download business table, pay attention to the differences between a fact and a dimension variable as we discussed in class. In this table there are only three dimensions to select. See if you can correctly identify which ones they are (again the PK can be counted as one dimension).**

After loading the data, we did `head()` command on dimension tables to get the preview of the data.

```
head(bed_type_df)
```

```
##   bed_id bed_code     bed_desc
## 1      1       BU         Burn
## 2      2       CC          CCU
## 3      3       DE    Detox ICU
## 4      4       IC          ICU
## 5      5       MS     Med/Surg
## 6      6       NE NeoNatal ICU
```

```
head(business_df)
```

```
##    ims_org_id                       business_name ttl_license_beds
## 1 INS00077200     140 Prescott Street Corporation              126
## 2 INS00000594                 366th Medical Group               10
## 3 INS00039181            7 Hills Pediatric Center               83
## 4 INS00011388 92 Brickroad Operating Company, LLC               49
## 5 INS00000593                   96th Medical Group               53
## 6 INS00036273    A and C Healthcare Services, Inc              130
##   ttl_census_beds ttl_staffed_beds bed_cluster_id
## 1             122              126              2
## 2               4               10              1
## 3              75               83              1
## 4              46               49              1
## 5              22               53              1
## 6             122              130              2
```

Fact variables are those, which contain numerical values and at the same time are also measurable i.e., which are quantitative. On the otherhand, dimension variables are those, which contain descriptive information that are not quantitative, primary keys are also considered as dimension.

In the business_df(dimension table), ims_org_id, business_name and bed_cluster_id are dimension variables.

In the bed_type_df(dimension table), bed_id, bed_code and bed_desc are dimensions. There are no fact variables here.

## 2. Identify the Facts variables from the single Fact Table

### Consider what type of variable would be a fact vs. a dimension and select three of those from the bed_fact Download bed_facttable. Be aware, it is likely that some variables in a fact table are dimensions, for example a foreign key or anything of that sort is a dimension even if it's listed in the fact table.

After loading the data, we did `head()` command on fact table to get the preview of the data.

```
head(bed_fact_df)
```

```
##    ims_org_id bed_id license_beds census_beds staffed_beds
## 1 INS00000519      2           10           7           10
## 2 INS00000519      5          566         394          566
## 3 INS00000519     15           25          17           25
## 4 INS00000519      4           38          26           38
## 5 INS00000519      6           32          22           32
## 6 INS00000519     18          671         466          671
```

Fact variables are those, which contain numerical values and at the same time are also measurable i.e., which are quantitative. On the otherhand, dimension variables are those, which contain descriptive information that are not quantitative, primary keys are also considered as dimension.

In the bed_fact_df(fact table), there are two dimensions variables and three fact variables. ims_ord_id, bed_id are dimensions. license_beds, census_beds and staffed_beds are fact variables.

## 3a. Analysis for Leadership

**Identify which hospitals have an Intensive Care Unit (ICU bed_id = 4) bed or a Surgical Intensive Care Unit (SICU bed_id = 15) bed or both.**

**Create three summary reports that show the following:**

**1. License beds: List of Top 10 Hospitals ordered descending by the total ICU or SICU license beds.**

**Include just two variables, hospital_name (business_name) and the total license beds from above as one summary fact. But include only 10 rows in your output table.**

```
#join tables and query for hospitals that have both ICU or SICU license beds or both
result1 <- sqldf("SELECT business_df.business_name AS hospital_name,
SUM(bed_fact_df.license_beds) AS total_license_beds
FROM bed_fact_df
JOIN business_df
ON bed_fact_df.ims_org_id = business_df.ims_org_id
WHERE bed_fact_df.bed_id = 4 OR bed_fact_df.bed_id = 15
GROUP BY hospital_name
ORDER BY total_license_beds DESC
LIMIT 10")

result1
```

```
##                                       hospital_name total_license_beds
## 1                           Phoenix Childrens Hospital                247
## 2                 University of Maryland Medical Center                220
## 3                         UC Health University Hospital                218
## 4                            Wesley Medical Center, LLC                214
## 5                                 Vidant Medical Center                204
## 6            Rady Childrens Hospital and Health Center                200
## 7                   Dallas County Hospital Association                195
## 8  Saint Lukes Episcopal Hospital Texas Medical Center                178
## 9                               The Methodist Hospital                170
## 10                           Emory University Hospital                169
```

**2. Do the same thing for Census beds. List of Top 10 Hospitals ordered by total icu or sicu census beds. Include just two variables, hospital_name (business_name) and the total census beds from above as one summary fact. Include only 10 rows again.**

```
#join tables and query for hospitals that have both ICU or SICU census beds or both
result2 <- sqldf("SELECT b.business_name AS hospital_name,
          SUM(f.census_beds) AS total_census_beds
          FROM bed_fact_df AS f
          JOIN business_df AS b ON f.ims_org_id = b.ims_org_id
          WHERE f.bed_id = 4 OR f.bed_id = 15
```

```
          GROUP BY business_name
          ORDER BY total_census_beds DESC
          LIMIT 10;")

result2
```

```
##                                                                hospital_name
## 1                            Shands Hospital at the University of Florida
## 2                                      Dallas County Hospital Association
## 3                                        Mercy Medical Center Saint Louis
## 4  Los Angeles County University of Southern California Healthcare Network
## 5                                                   The Methodist Hospital
## 6                          University of Minnesota Medical Center Fairview
## 7                                    University of Maryland Medical Center
## 8                                              Brigham and Womens Hospital
## 9                                                   Vidant Medical Center
## 10       Ronald Reagan University of California Los Angeles Medical Center
##    total_census_beds
## 1                167
## 2                145
## 3                142
## 4                139
## 5                138
## 6                129
## 7                127
## 8                124
## 9                123
## 10               122
```

**3. Do the same thing for Staffed beds. List of Top 10 Hospitals ordered by the total icu or sicu staffed beds. Include just two variables, hospital_name (business_name) and the sum of staffed beds from above as one summary fact. Include only 10 rows again.**

```
#join tables and query for hospitals that have both ICU or SICU staffed beds or both
result3 <- sqldf("SELECT b.business_name AS hospital_name,
          SUM(f.staffed_beds) AS total_staffed_beds
          FROM bed_fact_df AS f
          JOIN business_df AS b ON f.ims_org_id = b.ims_org_id
          WHERE f.bed_id = 4 OR f.bed_id = 15
          GROUP BY business_name
          ORDER BY total_staffed_beds DESC
          LIMIT 10;")
result3
```

```
##                                hospital_name total_staffed_beds
## 1                      Vidant Medical Center                203
## 2     Rady Childrens Hospital and Health Center                200
## 3         University of Maryland Medical Center                171
## 4                       Emory University Hospital                169
## 5  Shands Hospital at the University of Florida                167
## 6               Mercy Medical Center Saint Louis                163
## 7                    Wesley Medical Center, LLC                162
## 8                    Phoenix Childrens Hospital                159
```

```
## 9                       Grady Memorial Hospital               154
## 10                UC Health University Hospital               151
```

## 3b. Interpretation of Findings

**Based on your results from step 3a, discuss your insights from the data summary that you want to bring to the attention of Leadership.**

**For example, what are the top one or two hospitals per list based on bed volume? Are there any hospitals that appear on multiple lists? They might make good candidates for the intervention pilot program.**

```
# Merge three lists to identify which hospitals are in multiple lists
merged_data <- merge(merge(result1, result2, by = "hospital_name"), result3, by = "hospital_name")
merged_data
```

```
##                           hospital_name total_license_beds total_census_beds
## 1 University of Maryland Medical Center               220               127
## 2                 Vidant Medical Center               204               123
##   total_staffed_beds
## 1                171
## 2                203
```

Among the hospitals with ICU or SICU in the list, Phoenix Children's Hospital has the highest number of licensed beds at 247, followed by the University of Maryland Medical Center with 220. Shands Hospital at the University of Florida leads in total census beds with 167, closely followed by Dallas County Hospital Association with 145. Vidant Medical Center and Rady Children's Hospital and Health Center are the top two hospitals with 203 and 200 total staffed beds, respectively.

We merged three lists and identified that both University of Maryland Medical Center and Vidant Medical Center are included in these three lists. Both hospitals have significant numbers of total license beds, which are allowed by the state license, indicating that they have considerable capacity for patients. The total census beds are lower than the total licensed beds, suggesting that not all beds are being utilized to their full capacity. This provides room for additional patients or fluctuations in occupancy. The total staffed beds are also noteworthy. Both hospitals are adequately staffed, which is crucial for patient care, especially in ICU and SICU settings.

These insights provide valuable information about the capacity and staffing of these hospitals in the context of ICU and SICU beds.

## 4a. Drill down investigation

**Leadership is also interested in hospitals that have sufficient volume of both ICU and SICU beds, as opposed to either type of bed that you developed in step 3a.**

**Conduct the same investigation as you did for 3a and list the same output of top 10 hospitals by descending bed volume, only this time select only those top 10 hospitals that have both kinds of ICU and SICU beds, i.e. only hospitals that have at least 1 ICU bed and at least 1 SICU bed can be included in this part of the analysis.**

Conduct separate data investigations for Census beds, License beds, and staffed beds, like step 3a.

### 1. License beds: List of Top 10 Hospitals ordered descending by the total ICU and SICU license beds.

```
#join tables and query for hospitals that have both ICU and SICU license beds
result4 <- sqldf("SELECT business_df.business_name AS hospital_name,
          SUM(bed_fact_df.license_beds) AS total_license_beds
          FROM bed_fact_df
          JOIN business_df
          ON bed_fact_df.ims_org_id = business_df.ims_org_id
          WHERE bed_id IN (4,15)
          GROUP BY hospital_name
          HAVING COUNT(DISTINCT bed_id) = 2
          ORDER BY total_license_beds DESC
          LIMIT 10")

result4
```

```
##                                     hospital_name total_license_beds
## 1            University of Maryland Medical Center                220
## 2                    UC Health University Hospital                218
## 3     Shands Hospital at the University of Florida                167
## 4                                    MCGHealth, Inc                155
## 5                          Grady Memorial Hospital                154
## 6                        Jackson Memorial Hospital                151
## 7  University of Minnesota Medical Center Fairview                144
## 8              University Hospital in Bexar County                144
## 9                          Carolinas Medical Center                137
## 10                          Yale New Haven Hospital                136
```

### 2. Census beds: List of Top 10 Hospitals ordered descending by the total ICU and SICU census beds.

```
# join tables and query for hospitals that have both ICU and SICU census beds
result5 <- sqldf(
          "SELECT business_name AS hospital_name, SUM(census_beds) AS census_bed_count
          FROM business_df
          JOIN bed_fact_df ON business_df.ims_org_id = bed_fact_df.ims_org_id
          WHERE bed_id IN (4,15)
          GROUP BY business_name HAVING COUNT(DISTINCT bed_id) = 2
          ORDER BY census_bed_count DESC
          LIMIT 10;")

result5
```

```
##                                     hospital_name census_bed_count
## 1     Shands Hospital at the University of Florida              167
## 2  University of Minnesota Medical Center Fairview              129
## 3            University of Maryland Medical Center              127
## 4                          Jackson Memorial Hospital              117
## 5                    UC Health University Hospital              110
## 6                         Carolinas Medical Center              106
## 7                       Cedars Sinai Health System               92
## 8               University Hospital in Bexar County               91
## 9                     Duke University Health System               91
```

```
## 10                The Cleveland Clinic Foundation              88
```

**3. Staffed beds: List of Top 10 Hospitals ordered descending by the total ICU and SICU staffed beds.**

```
# join tables and query for hospitals that have both ICU and SICU staffed beds
result6 <- sqldf(
          "SELECT business_name AS hospital_name, SUM(staffed_beds) AS staffed_bed_count
          FROM business_df
          INNER JOIN bed_fact_df ON business_df.ims_org_id = bed_fact_df.ims_org_id
          WHERE bed_id IN (4,15)
          GROUP BY business_name HAVING COUNT(DISTINCT bed_id) = 2
          ORDER BY staffed_bed_count DESC
          LIMIT 10;")

result6
```

```
##                                     hospital_name staffed_bed_count
## 1             University of Maryland Medical Center               171
## 2     Shands Hospital at the University of Florida               167
## 3                           Grady Memorial Hospital               154
## 4                     UC Health University Hospital               151
## 5  University of Minnesota Medical Center Fairview               144
## 6                          Carolinas Medical Center               137
## 7        Saint Josephs Hospital and Medical Center               134
## 8   Chattanooga Hamilton County Hospital Authority               134
## 9                          Jackson Memorial Hospital               128
## 10        Sunrise Hospital and Medical Center, LLC               125
```

## 4b. Final recommendation

**Based on your analyses in step 3a and 4a, state your final recommendation here for Leadership as to which hospitals are the best candidates for their pilot intervention program. Remember, Leadership stated they are only interested in one or two hospitals for their pilot sites so it's best to tailor your recommendation to their business need and avoid unnecessary details that might confuse them. Identify your hospitals and briefly explain why you chose them.**

```
# Merge three lists to identify which hospitals are in multiple lists
merged_data2 <- merge(merge(result4, result5, by = "hospital_name"), result6, by = "hospital_name")
merged_data2
```

```
##                                     hospital_name total_license_beds
## 1                          Carolinas Medical Center               137
## 2                         Jackson Memorial Hospital               151
## 3    Shands Hospital at the University of Florida               167
## 4                     UC Health University Hospital               218
## 5             University of Maryland Medical Center               220
## 6 University of Minnesota Medical Center Fairview               144
##   census_bed_count staffed_bed_count
## 1              106               137
## 2              117               128
## 3              167               167
```

```
## 4              110              151
## 5              127              171
## 6              129              144
```

Based on the analyses in step 3a and 4a, our final recommendation for Leadership as to which hospital is the best candidate for the pilot intervention program is the University of Maryland Medical Center. Leadership is aiming to launch an intervention to hire more nurses in ACME's hosptial network for hospitals with Intensive Care Units (ICU) and Surgical Intensive Care Units (SICU) to better care for critical patients that are admitted to the facilities. Among hospitals that contain both an ICU and a SICU, the hosptial with the greatest sum of licensed beds in the ICU and SICU combined is the University of Maryland Medical Center with 220 beds. Among hospitals that contain both an ICU and a SICU, the hosptial with the third greatest sum of census beds in the ICU and SICU combined is the University of Maryland Medical Center with 127 beds. The University of Maryland Medical Center is only 40 beds behind the hospital with the greatest sum of census beds in the ICU and SICU combined. Among hospitals that contain both an ICU and a SICU, the hosptial with the greatest sum of staffed beds in the ICU and SICU combined is the University of Maryland Medical Center with 171 beds.

Upon merging the data from each bed type, we identified that 6 hospitals are included in the three lists. Among the 6 hospitals, the Universtiy of Maryland Medical Center has the greatest total licensed beds, which means the hopital has the greatest capacity for patients. The total census beds for the Universtiy of Maryland Medical Center is less than the total licensed beds, which means that the hospital is not taking advantage of the maximum number of beds allowed by state license (93 bed difference between licensed beds and census beds). Therefore there is potential to increase the number of patients admitted to the hospital. As more patietns are admitted to the hospital, more nurses will need to be hired to maintain a higher nurse to patient ratio that will lead to better outcomes in these intensive care settings. If the University of Maryland Medical Center plans to take in more patients to fill the 220 licensed beds, then the hospital will absolutely require more nurses becasue the facility is only staffed for 171 beds right now (49 bed difference between licensed beds and staffed beds). Therefore, the University of Maryland Medical Center serves as a prime candidate to launch an intervention to hire more nurses in ACME's hospital network, for hospitals with Intensive Care Units (ICU) and Surgical Intensive Care Units (SICU), to better care for critical patients who are admitted to these facilities.