READ ME:

**Introduction**

This project uses data from the October 2015 HCAPS survey. The data are available [here](https://data.medicare.gov/data/hospital-compare). Note that this file reflects updated data from last semester’s project, which had data from May 2015.

**Key Files**

1. load\_data\_lake.sh – downloads the data **in zip format**, unzips its and loads the data in into HDFS. Note – please ensure your volume has an unzip program installed.
2. hive\_base\_ddl.sql – loads the data into SQL tables
3. transform\_survey.sql – transforms the relevant patient survey table and outputs a table called survey\_scores with the top hospitals via patient survesy.
4. transform complications.sql and transform\_readmit\_deaths.sql – transforms the complications and readmit\_deaths tables into useable format (probably not needed.)
5. analytical\_output.sql – main SQL script for conducting the analysis. This script results in three key tables:
   1. best\_hospitals: lists the top 10 best hospitals culled from the data
   2. best\_states: lists the best states in the data
   3. variance\_measures: sorts the quality control measures by the measure with the highest standard deviation.

**Analytical Procedure**

To establish the best hospitals and states from all of the various procedures measured, I conducted the following steps.

1. Review the measures to determine which measures showed better performance the higher the number, and which indicated better performance with a lower number.[[1]](#footnote--1)
2. Eliminate all procedures for which a lower number indicated better performance.
3. Attempt to eliminate small hospitals that could skew the data set by eliminating hospitals with low emergency room volume. This procedure was generally unsuccessful (see discussion below.)
4. Generate a list of the most commonly measured indicators of hospital quality – this will allow me to get the widest ranking possible.
5. Take an average of these measures, which are generally numbers out of 100 %.
6. Output the hospitals and states with the highest average score across these top 10 measures.
7. As a sanity check, cross check these top hospitals against the complications and readmissions and deaths tables to make sure that none of my top hospitals appear near the top for complications and readmissions and deaths.

To evaluate variation in these measures, I undertook the following steps;

1. Estimate the standard deviation of each measure.
2. Sort the final table by the highest standard deviation.

To evaluate patient satisfaction I undertook the following steps:

1. Use the table hvbp\_hcahps\_10\_28\_2015.csv, which aggregates patient satisfaction scores.
2. Sort the table by overall achievement score, and output the top hospitals.

**Results, Caveats and Discussion**

The best hospitals from this procedure are:

|  |  |
| --- | --- |
| **Hospital Name** | **State** |
| LANSDALE HOSPITAL | PA |
| NORTH KANSAS CITY HOSPITAL | MO |
| SURGICAL HOSPITAL AT SOUTHWOODS | OH |
| TEXAS HEALTH PRESBYTERIAN HOSPITAL ALLEN | TX |
| SAINT VINCENT HOSPITAL | PA |
| HERRIN HOSPITAL | IL |
| WATERBURY HOSPITAL | CT |
| FRANCISCAN ST ELIZABETH HEALTH - LAFAYETTE EAST | IN |
| CHI HEALTH LAKESIDE | NE |
| VIDANT ROANOKE CHOWAN HOSPITAL | NC |

These results are generally unsatistfactory, and are likely skewed by the inability to eliminate small hospitals from the data set. While it is credible that the Connecticut has one of the top 10 best hospitals in the country, it is unlikely that Indiana and Nebraska do.

|  |  |  |
| --- | --- | --- |
| **Rank** | **State** | **Average Score** |
| 1 | SD | 97.0 |
| 2 | CO | 96.4 |
| 3 | MD | 96.4 |
| 4 | NE | 95.5 |
| 5 | NC | 94.3 |
| 6 | RI | 93.9 |
| 7 | WI | 93.5 |
| 8 | UT | 93.2 |
| 9 | IA | 93.0 |
| 10 | NH | 92.8 |

These results are similarly unsatisfactory, and likely indicate that the top 10 most commonly measured indicators are not a representative sample that actually indicate hospital quality. Those indicators are:

|  |  |
| --- | --- |
| **Measure** | **Count in data set** |
| Healthcare workers given influenza vaccination | 2718 |
| Venous Thromboembolism (VTE) Prophylaxis | 1625 |
| Discharged on Statin Medication | 1591 |
| Stroke Education | 1547 |
| Aspirin at Arrival | 1247 |
| Assessed for Rehabilitation | 1021 |
| Discharged on Antithrombotic Therapy | 1011 |
| Antithrombotic Therapy by End of Hospital Day 2 | 1009 |
| Head CT results | 827 |
| Anticoagulation Therapy for Atrial Fibrillation/Flutter | 671 |

The measures of hospital variation are:

|  |  |
| --- | --- |
| **Measure** | **SD** |
| Head CT results | 21.36 |
| Thrombolytic Therapy | 20.53 |
| Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival | 19.54 |
| Home Management Plan of Care Document | 15.43 |
| Healthcare workers given influenza vaccination | 15.11 |
| Venous thromboembolism prophylaxis | 13.88 |
| Warfarin therapy discharge instructions | 13.48 |
| Stroke Education | 12.32 |
| Immunization for influenza | 12.18 |
| Evaluation of LVS Function | 10.92 |
| Initial antibiotic selection for CAP in immunocompetent patient | 9.05 |
| Discharged on Statin Medication | 8.46 |
| Venous Thromboembolism (VTE) Prophylaxis | 8.11 |
| Anticoagulation overlap therapy | 7.46 |
| Fibrinolytic Therapy Received Within 30 Minutes Of Hospital Arrival | 7.09 |
| Primary PCI Received Within 90 Minutes of Hospital Arrival | 6.64 |
| ICU venous thromboembolism prophylaxis | 6.51 |
| Assessed for Rehabilitation | 5.65 |
| Prophylactic antibiotic received within 1 hour prior to surgical incision | 5.53 |
| Anticoagulation Therapy for Atrial Fibrillation/Flutter | 5.53 |
| Unfractionated heparin with dosages/platelet count monitoring | 5.27 |
| Surgery Patients on a Beta Blocker Prior to Arrival Who Received a Beta Blocker During the Perioperative Period | 5.13 |
| Antithrombotic Therapy by End of Hospital Day 2 | 5.12 |
| Aspirin at Arrival | 5.12 |
| Postoperative Urinary Catheter Removal | 4.69 |
| Prophylactic antibiotics discontinued within 24 hours after surgery end time | 4.57 |
| Prophylactic Antibiotic Selection for Surgical Patients | 4.42 |
| Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis within time frame | 4.09 |
| Discharged on Antithrombotic Therapy | 3.13 |
| Emergency department volume | NULL |
|  |  |

These results also must be considered carefully, as some of the units are not bounded bewteen 0-100% and are therefore not comparable. A better methodology would have been to transform everything into z scores and take the range, but I did not have sufficient time to solve the sql programming to do so.

The top ranked hospitals by patients are:

|  |  |  |
| --- | --- | --- |
| **Hospital** | **State** | **Score** |
| FAYETTE MEDICAL CENTER | AL | 10 |
| SACRED HEART HOSPITAL ON THE EMERALD COAST | FL | 10 |
| MAYO CLINIC HOSPITAL | AZ | 10 |
| ARIZONA ORTHOPEDIC AND SURGICAL SPECIALITY HOSP | AZ | 10 |
| ARKANSAS HEART HOSPITAL, LLC | AR | 10 |
| ARKANSAS SURGICAL HOSPITAL | AR | 10 |
| SHARP MEMORIAL HOSPITAL | CA | 10 |
| RONALD REAGAN U C L A MEDICAL CENTER | CA | 10 |
| SUTTER DAVIS HOSPITAL | CA | 10 |
| KAISER FOUNDATION HOSPITAL - ORANGE COUNTY - ANAHE | CA | 10 |
| PATIENTS' HOSPITAL OF REDDING | CA | 10 |
| FRESNO SURGICAL HOSPITAL | CA | 10 |
| SUTTER MATERNITY & SURGERY CENTER OF SANTA CRUZ | CA | 10 |
| STANISLAUS SURGICAL HOSPITAL | CA | 10 |
| FRESNO HEART AND SURGICAL HOSPITAL | CA | 10 |
| SUTTER SURGICAL HOSPITAL - NORTH VALLEY | CA | 10 |
| HOAG ORTHOPEDIC INSTITUTE | CA | 10 |
| YAMPA VALLEY MEDICAL CENTER | CO | 10 |
| ORTHOCOLORADO HOSPITAL AT ST ANTHONY MED CAMPUS | CO | 10 |
| MAYO CLINIC | FL | 10 |

This list is more credible. While there are small hospitals on here, it is expected that small hospitals would have high patient scores. Similarly, there are nationally renowned top hospitals such as the Mayo clinic appearing on the list.

1. With thanks to fellow classmate Mary Lewis for the research on health care quality indicators. [↑](#footnote-ref--1)