HEALTHCARE CAPSTONE PROJECT

Business understanding

- CMS rates hospitals in the US on a scale of 1-5, with the objective of making it easier for patients and consumers to compare the quality of services offered by hospitals.
- The ratings directly influence the choice of hospitals made by consumers and may significantly impact hospitals' revenues.
- Thus, it is extremely important for hospitals to understand the methodology that CMS uses for calculating the ratings so that they can work on improving the factors that affect their ratings.
- Hospital Compare allows consumers to select multiple hospitals and directly compare performance measure information related to heart attack, heart failure, pneumonia, surgery and other conditions. These results are organized by:
    - General information
    - Survey of patients' experiences
    - Timely & effective care
    - Complications
    - Readmissions & deaths
    - Use of medical imaging
    - Payment & value of care

Data Understanding

- The data from Hospital Compare has 60 files containing information about of 4818 hospitals across the United States.
- The files contain demographic information of the hospitals, the scores of hospitals based on various factors categorized into different groups, the ratings of hospitals based on various factors categorized into different groups, the overall performance of the hospitals, position of hospitals when compared with the national averages etc.
- CMS has a long list of measures based on which hospitals are rated. These measures are from the Hospital Inpatient Quality Reporting (IQR) Program and the Hospital Outpatient Quality Reporting (OQR) Program. CMS keeps updating the list with new measures when required.
- To calculate the Overall Hospital Quality Star Rating, the measures are categorized into seven mutually exclusive groups:
    1. Outcomes – Mortality (7 measures)
    2. Outcomes – Safety of Care (8 measures)
    3. Outcomes – Readmissions (8 measures)

4. Patient Experience (11 measures)
5. Process – Effectiveness of Care (18 measures)
6. Process – Timeliness of Care (7 measures)
7. Efficiency – Outpatient Imaging Use (5 measures)

- CMS uses the following criteria to exclude measures from the star rating calculation:

    1. Measures suspended, retired, or delayed in the programs and/or from public reporting on Hospital Compare
    2. Measures with no more than 100 hospitals reporting performance publicly
    3. Structural measures
    4. Measures for which it is unclear whether a higher or lower score is better (nondirectional)
    5. Measures no longer required for Hospital IQR Program or Hospital OQR Program
    6. Duplicative measures (e.g., identical measures or a composite measure consisting of individual measures) CMS will continue to consider feedback when evaluating future measures along these criteria.

Methodology:

Business Approach:

The Overall Hospital Quality Star Rating is calculated using a five-step process:

1. Selection and standardization of measures for inclusion in Star Ratings
2. Assignment of measures to groups
3. Calculation of group scores using latent variable model
4. Calculation of hospital summary scores as a weighted average of group scores
5. Application of clustering algorithm to categorize summary scores into Star Ratings

<u>Step 1: Measure selection</u>

To calculate the Overall Hospital Quality Star Rating, 64 measures have been selected based on the criteria mentioned above.

<u>Step 2: Assignment of measures to groups:</u>

The selected 64 measures are categorized into the below seven mutually exclusive groups

1. Mortality (7 measures)
2. Safety of Care (8 measures)
3. Readmissions (8 measures)
4. Patient Experience (11 measures)

5. Effectiveness of Care (18 measures)
6. Timeliness of Care (7 measures)
7. Outpatient Imaging Use (5 measures)

Step 3: Calculation of group scores using latent variable model :

The loadings of each measure within a group are calculated using latent variable method. These values are used for calculating the group score.

Step 4: Calculation of hospital summary scores:

The hospital summary scores are calculated as the weighted average of the group scores. CMS evaluated potential weighting options considering the following three criteria:

- Group Importance
- Consistency with Existing CMS Policies and Priorities
- Stakeholder Input

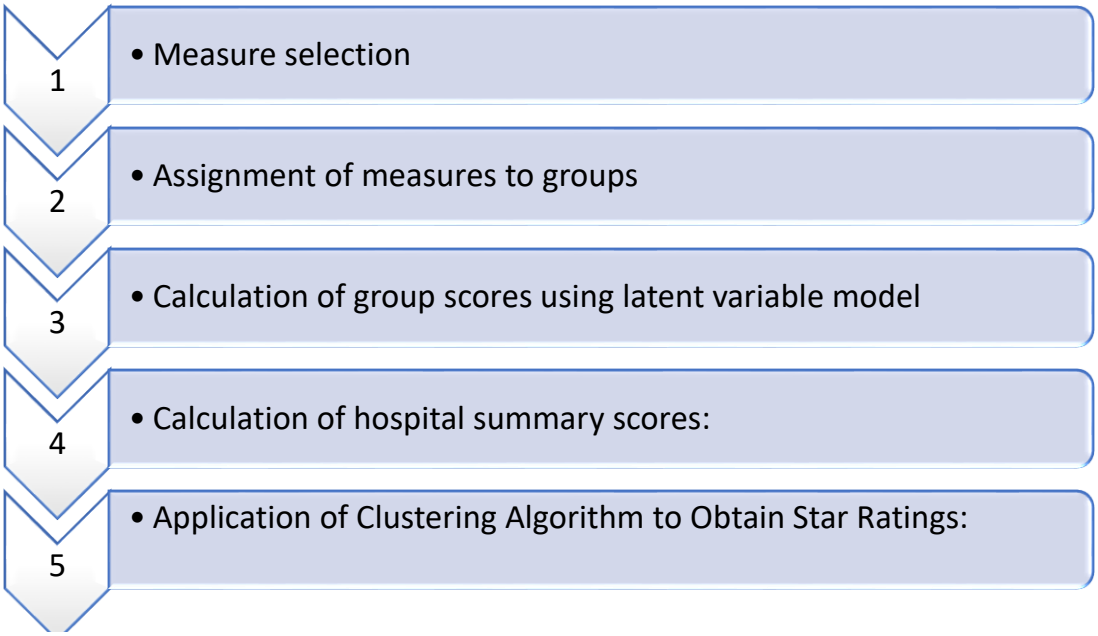
The weighting scheme evaluated based on the above criteria is as below:

Mortality (N=7)	22%
Safety of Care (N=8)	22%
Readmission (N=8)	22%
Patient Experience (N=11)	22%
Effectiveness of Care (N=16)	4%
Timeliness of Care (N=7)	4%
Efficient Use of Medical Imaging (N=5)	4%

Step 5: Application of Clustering Algorithm to Obtain Star Ratings:

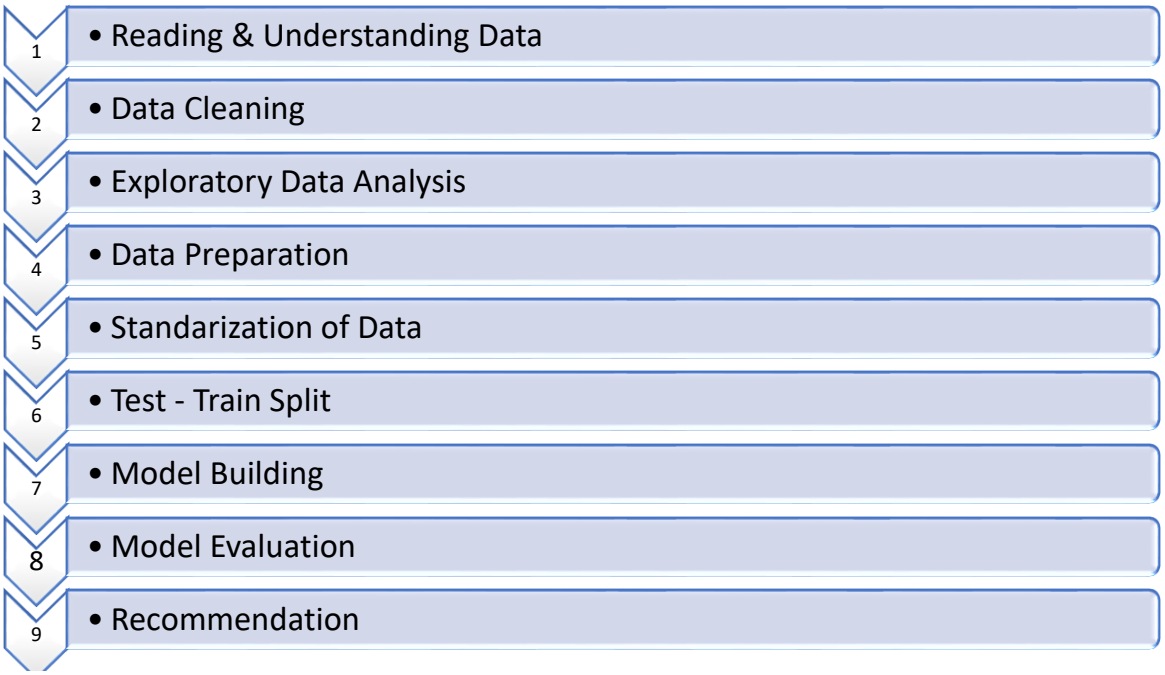
Finally the hospital summary scores are translated to star rating by using Clustering algorithm.

BUSINESS APPROACH:



Technical Approach:

TECHNICAL APPROACH:



The solution for this problem is developed using Python. Below are the steps followed:

Reading & Understanding Data:

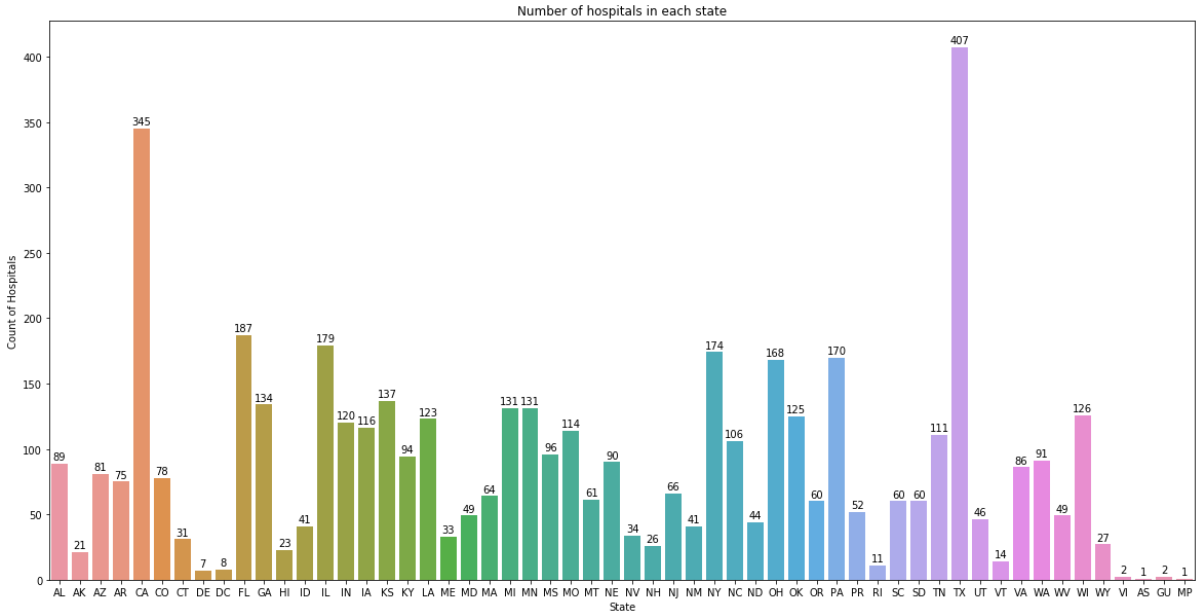
- The required files are read and stored. Only the required measures are retained in each group and the data frames are pivoted to see the score of each hospital against each measure within the group.

Data Cleaning and Data Preparation:

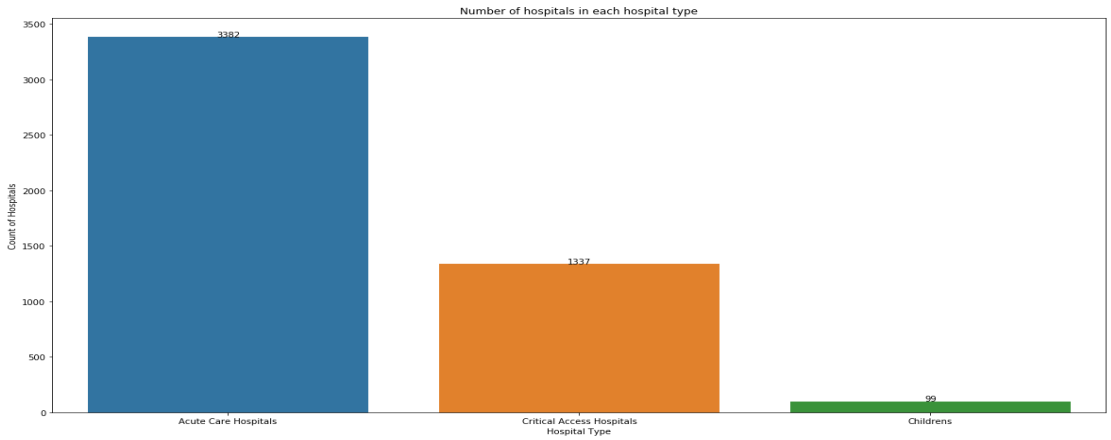
- The data is checked for missing values.The 'Not Available' values are replaced with null.
- The distributions of all the attributes are converted to normal using Quantile Transformation.
- The missing values are imputed with the mean.
- The values are standardized using Standard Scaler.

EDA & Visualization:

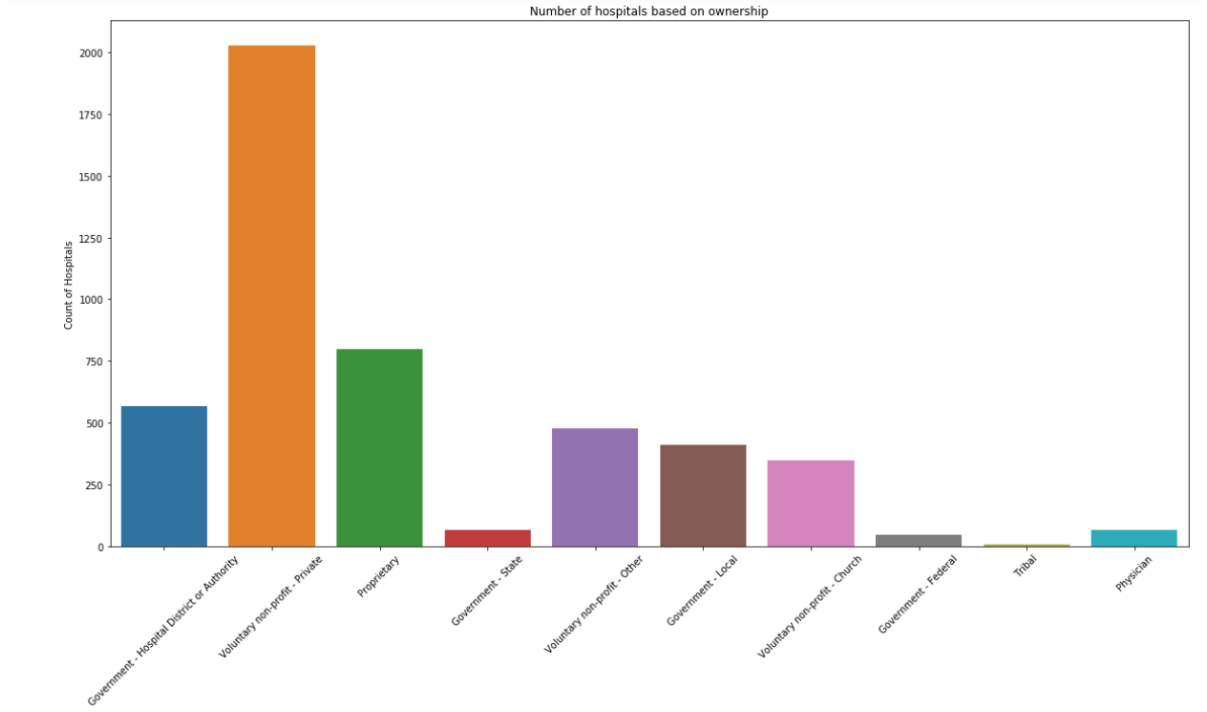
- To perform the exploratory data analysis the Num_of_Measures column is created which stores the number of measures reported by each hospital in that group.
- Various countplots, boxplots, distribution plots, heatmaps are made as part of EDA.



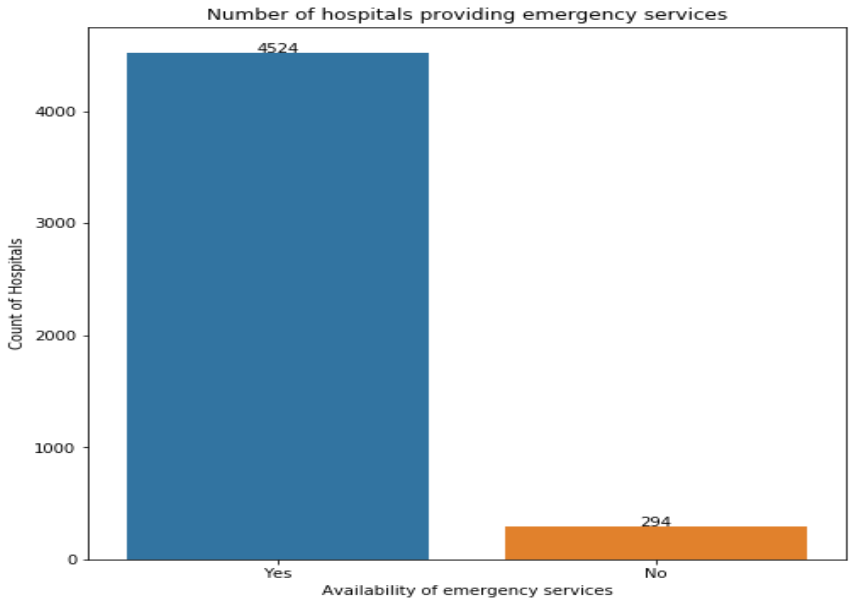
- The number of hospitals are highest in Tx (407) and CA (345) and the least number of hospitals are found in MP(1) ,AS (1



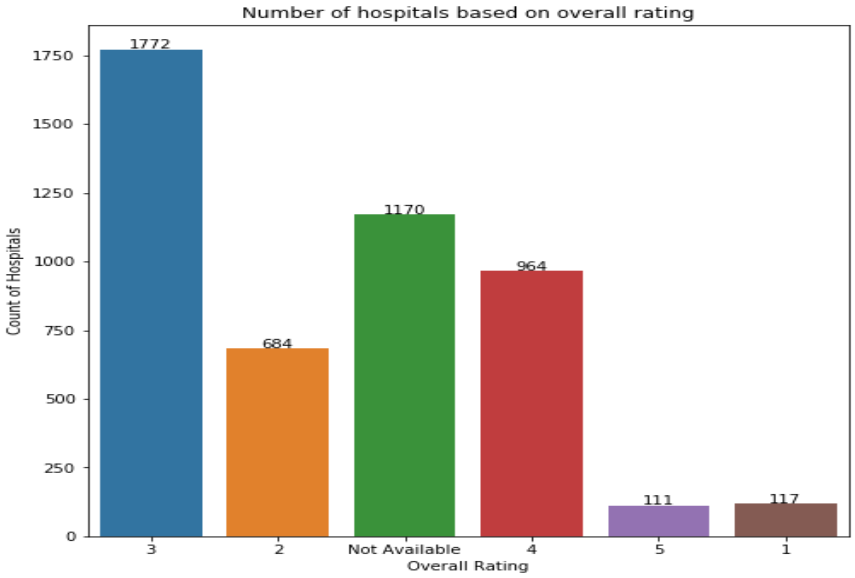
- Acute care hospitals are more when compared to critical Access hospitals or Children hospitals.
- The count of Acute care hospitals is 3382, Critical Access Hospitals is 1337 and Children Hospitals is 99.



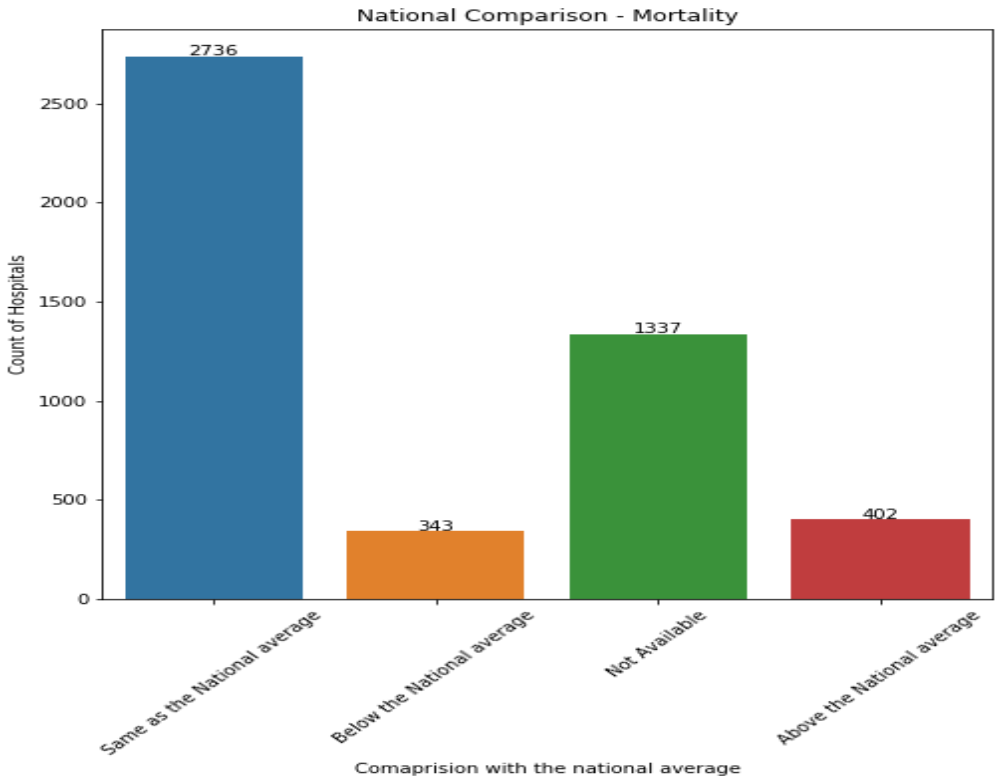
- Voluntary non-profit –private ownership hospitals are more compared to other hospitals
- The count of Voluntary non-profit-private owner ship is 2027
- The least type of ownership hospitals are Tribal i.e 8 no of hospitals



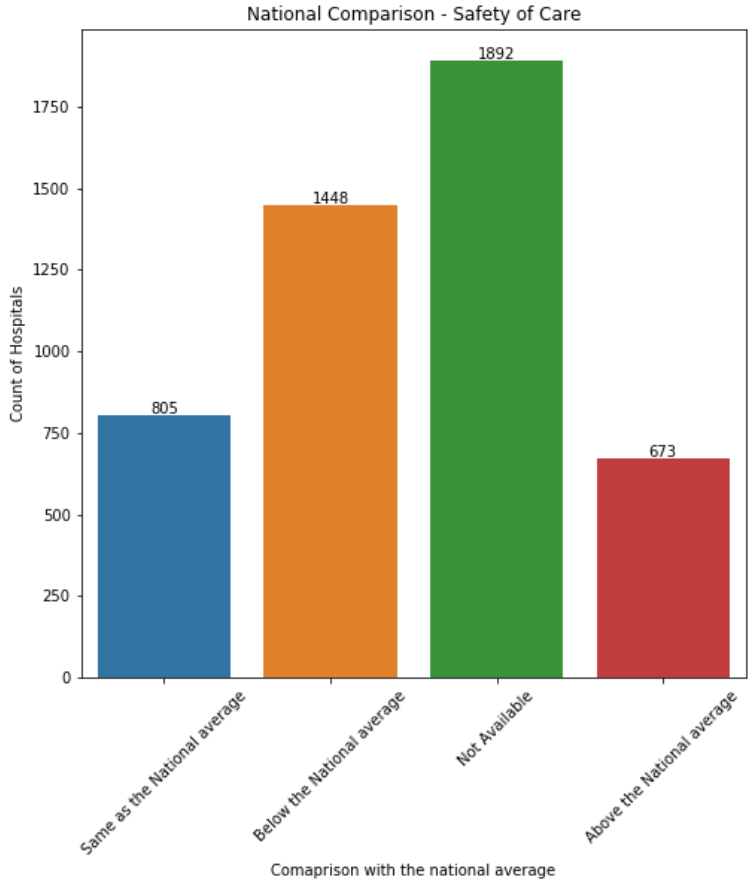
- Hospitals providing emergency services are more.
- The number of hospitals which are providing emergency services is 4524
- The number of hospitals which are not providing the emergency services is 294.



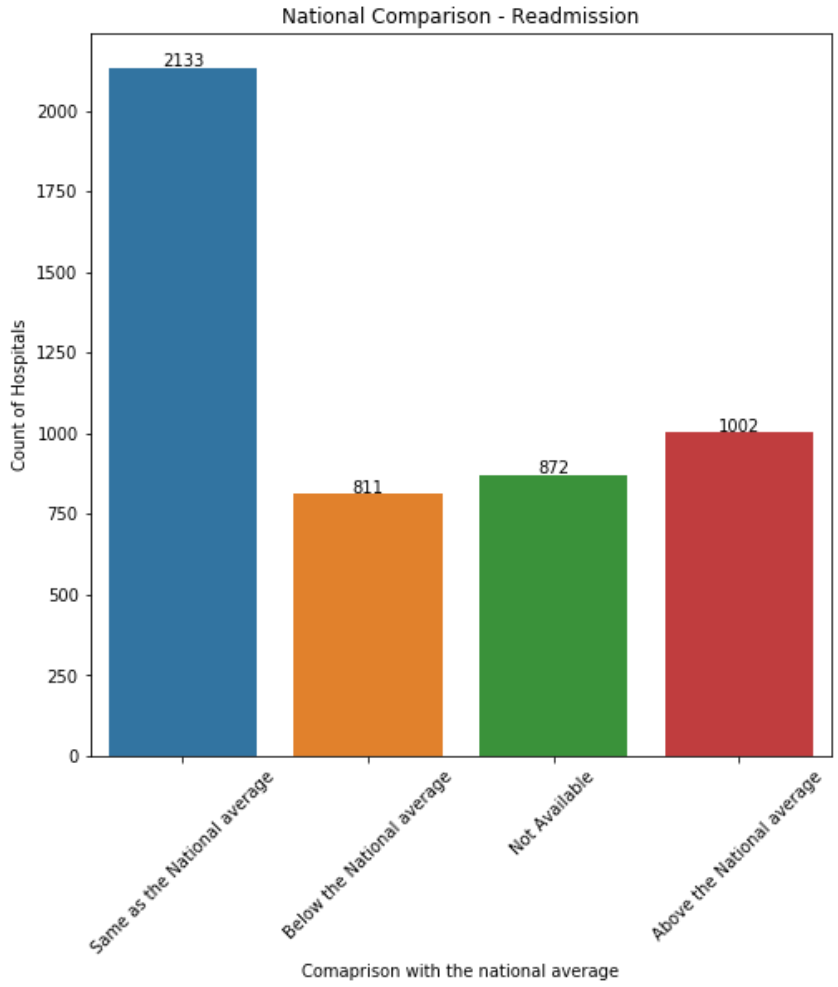
- From the above plot,3 star rating hospitals are more than compared to other hospitals .In the above plot the hospitals which are rated 3 star are 1772.
- The number of hospitals with 5 and 1 star ratings are the least.



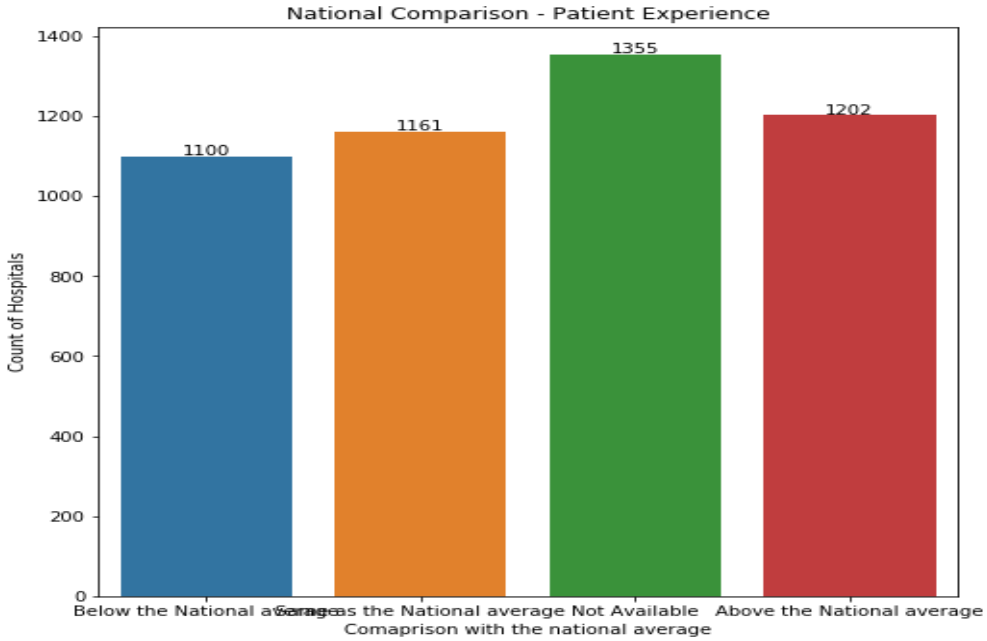
- The hospitals are compared with national average.
- It is divided into 3 levels
- If the score meets the expectation given in national file its is remarked as “Same as the national average” and the rating falls under this are 3 star rated hospitals
- If it does not meet the expectations its marked as “Below the national average”, the hospitals which fall under this are 1 and 2 star rated hospitals
- If the score is more than the expected it is remarked as “Above the National Average” the hospitals fall under this category are 4 and 5 star rated hospitals.
- In the above plot the hospitals which are “same as the national average” are 2736
- The hospitals which are “Below the national average” are 343
- The hospitals which are “Above the national average” are 402
- The “not available” data in this case is 1337



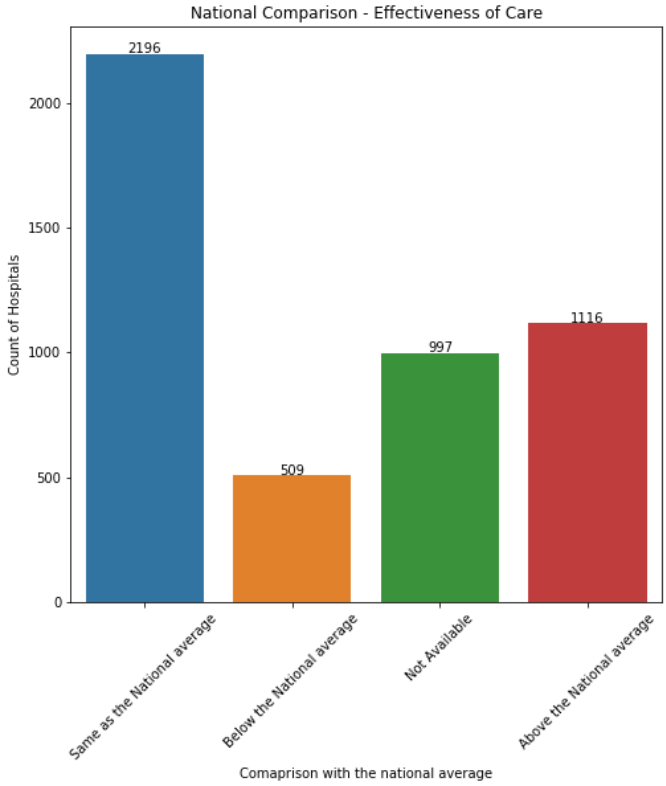
- For the group of Safety of care the “not available” is more i.e the number of hospitals which are not providing the data is 1892
- The number of hospitals which are “below the national average” are 1448
- The number of hospitals which are “Same as the national average” are 805
- The number of hospitals which are “Above the national average” are 673



- The number of hospitals which provide readmission and meets the national comparision are 2133.
- The number of hospitals which are “below the National average” are 811
- The number of hospitals which are “Above the national average” are 1002
- The “not available” data for this measure are 872
- So from the above we got insights that providers are providing readmissions and the score for this is good

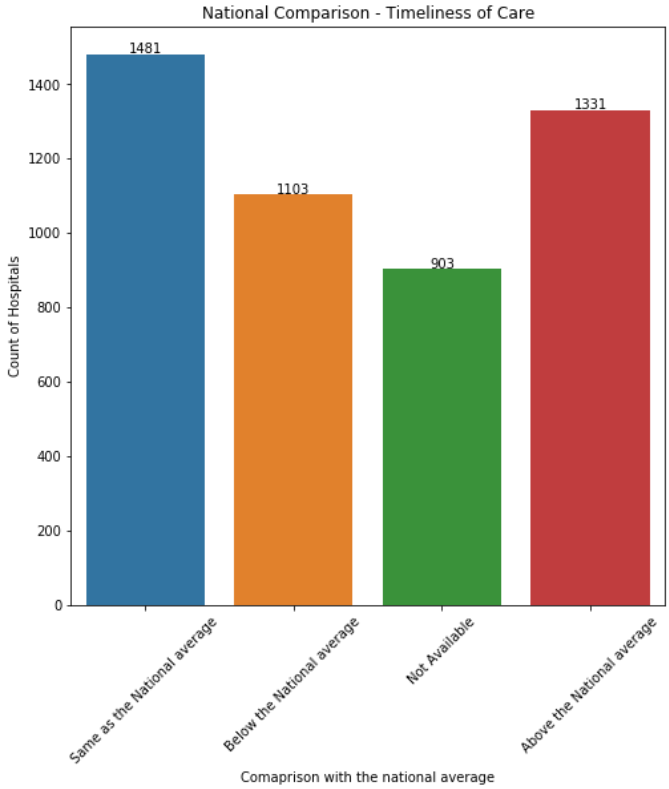


- The count of hospitals which contribute “Not available” data is more 1355
- The number of hospitals which meet national compare is 1161
- The number of hospitals which are “below the national average” is 1100
- The number of the hospitals which are “Above the national average” are 1202

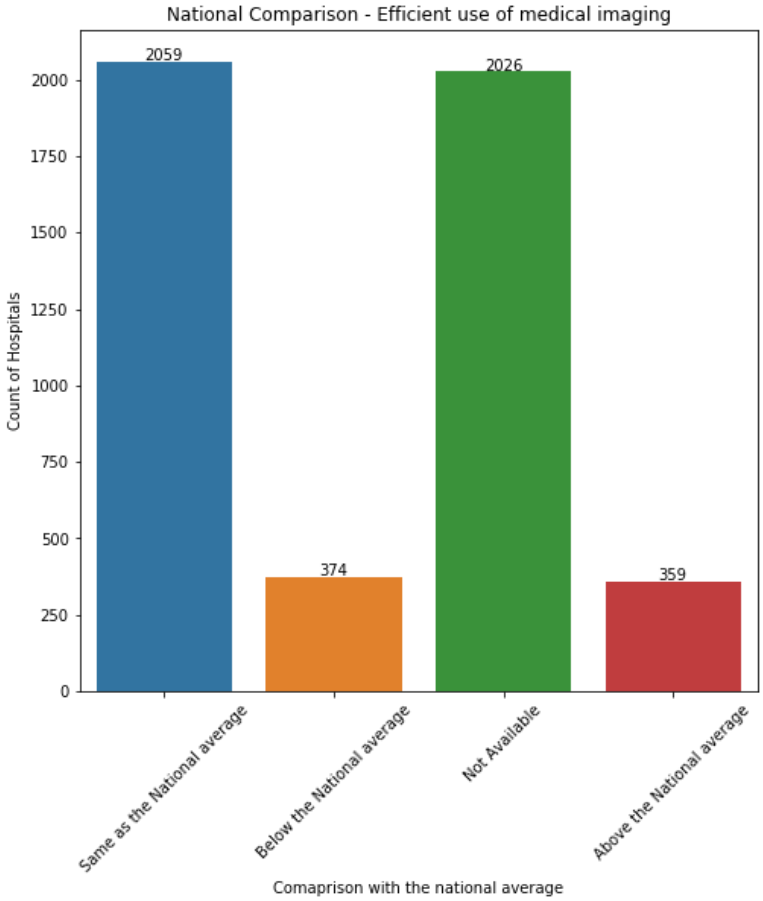


- The hospitals which perform as “same as national average” are 2196
- The hospitals which perform “Above the national average” are 1116
- The hospitals which perform “below the national average” are 509

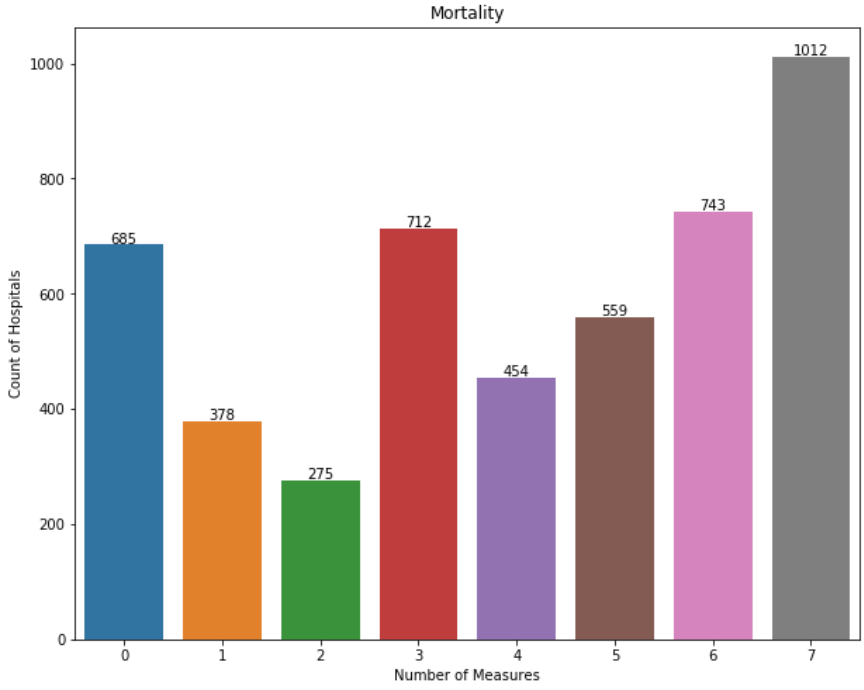
- The “Not Available” data is 997



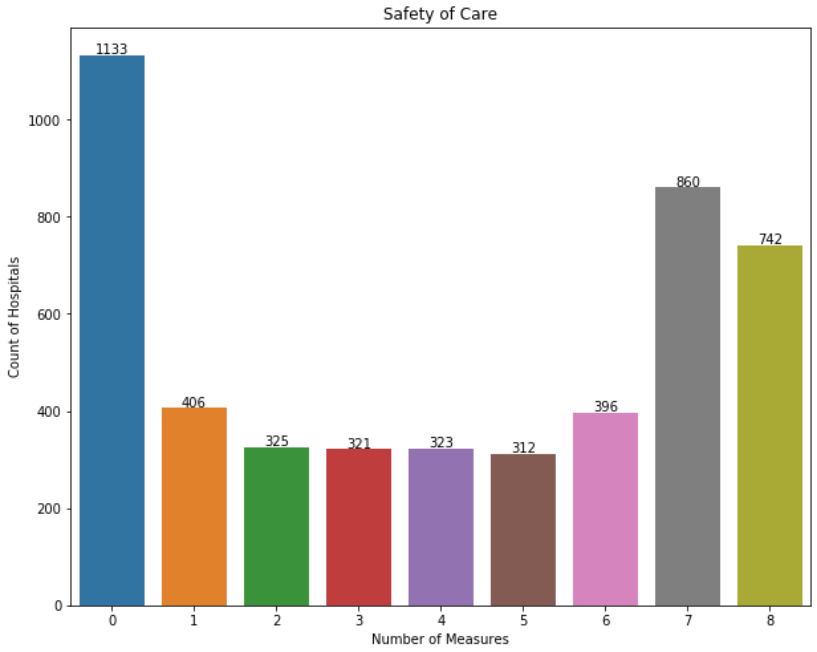
- The Hospitals are performing very good on this measure
- The number of hospitals which are “above the national average” are 1331
- The number of hospitals which are “same as the national average” are 1481
- The number of hospitals which are “below the national average “ are 1103
- The number of hospitals which has “not available “ data are 903



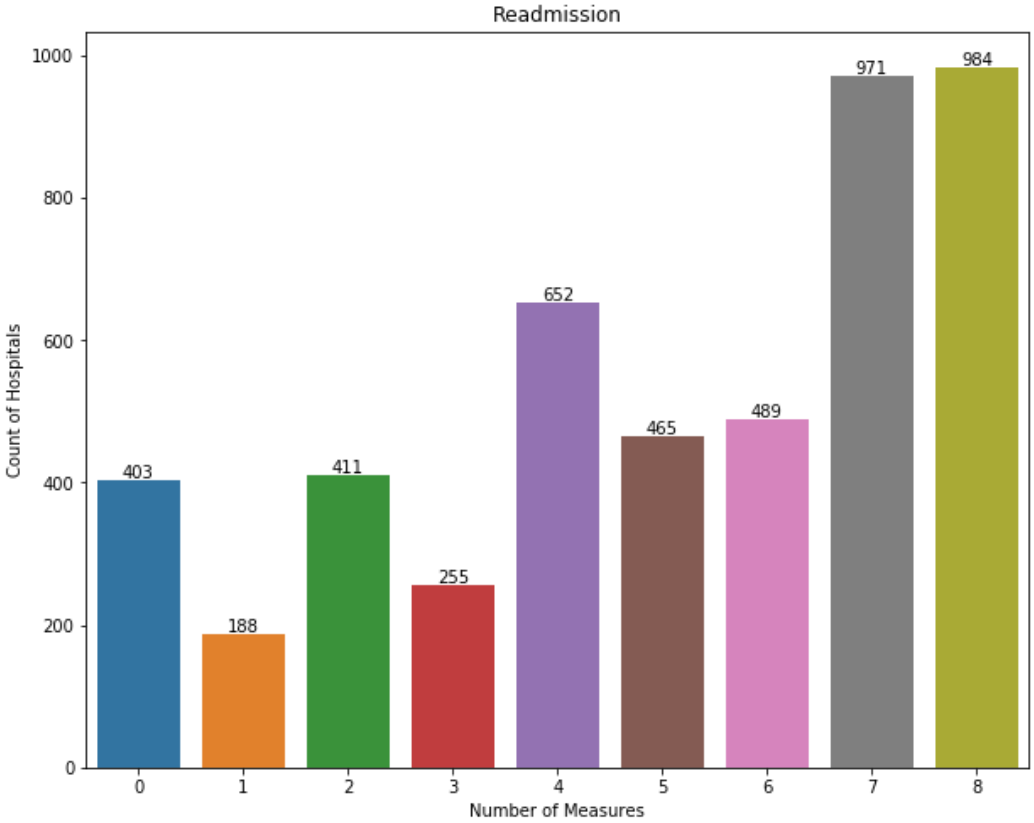
- The hospitals which have “not available” data is 2026
- The providers providing the services for this measure for “same as National average” is 2059
- The hospitals which are “below the national average” are 374
- The hospitals which are “above the national average” are 359



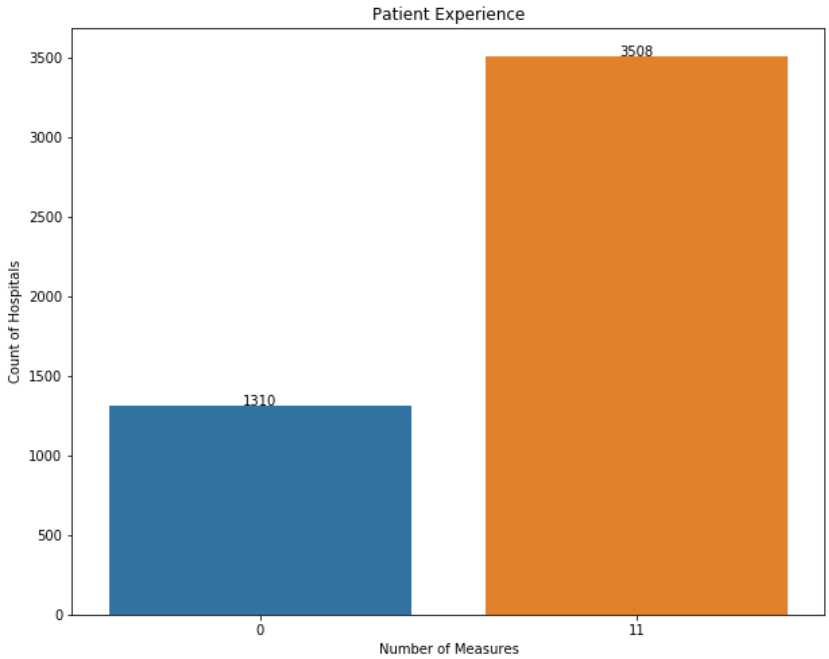
- 1012 hospitals report all the 7 measures in the mortality
- 743 hospitals reported 6 measures
- 685 hospitals does not report the any measures



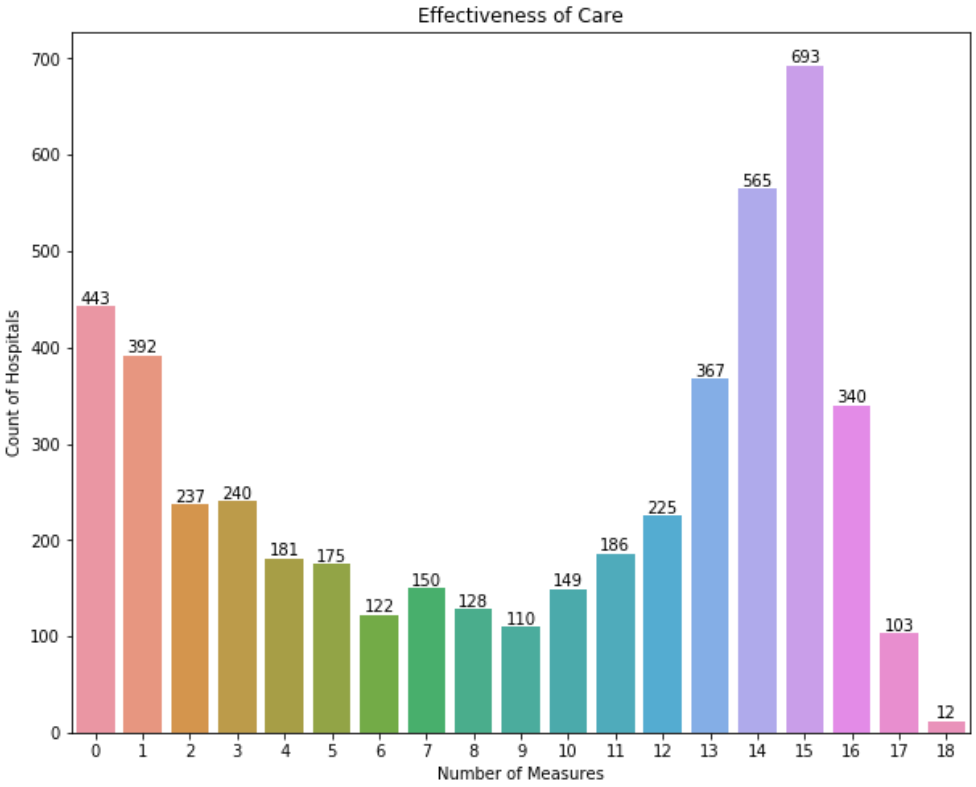
- For safety of care there are 1133 hospitals which does not report any measure
- Out of 8 measures for safety of care 860 hospitals report 7 measures
- 742 hospitals report all the 8 measures



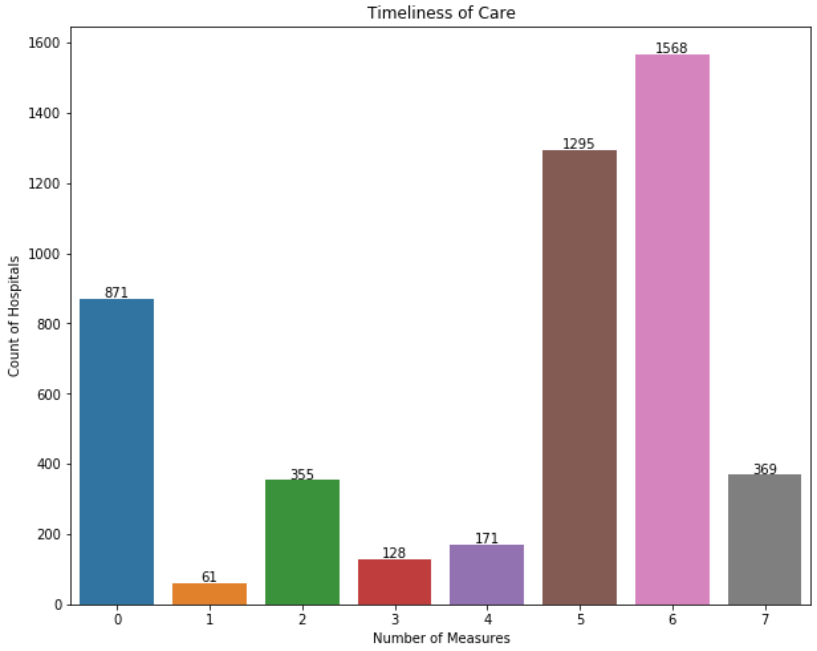
- From the above plot there are 984 Hospitals which report all the 8 measure s
- 971 hospitals report 7 measures each and 403 hospitals doesn't report any data.



- The reports of hospitals in patient experience category are interesting either hospitals report all the measures or else 0.The number of hospitals which report all the measures is 3508 the number of hospitals where the count of measure is 0 is 1310.

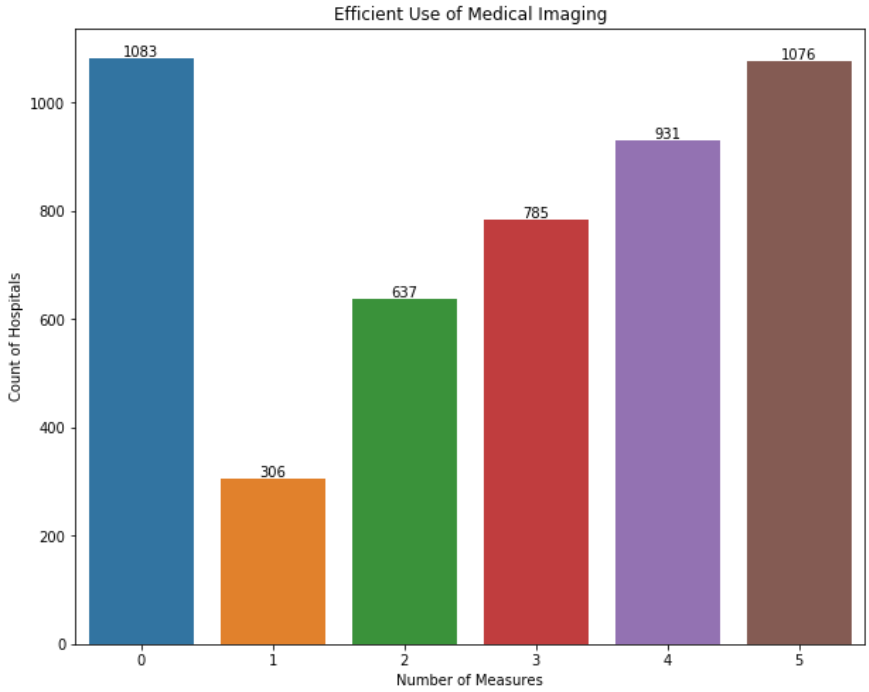


- Out of 18 measures for effective of care, 15 measure data is reported by 693, 443 hospitals does not report any information.



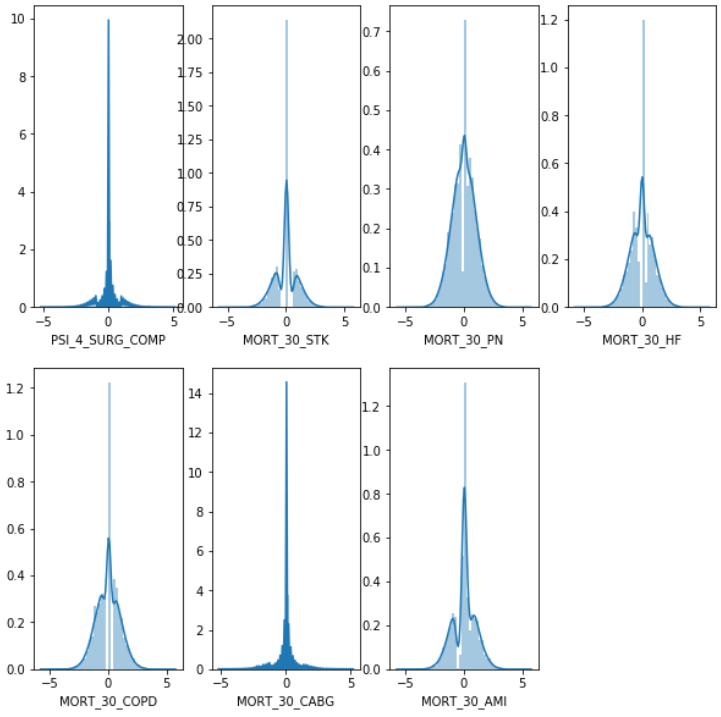
- The total number of measures for timeliness of care group is 7.
- Among 7 the number of hospitals which reported all the 7 measures is 369 hospitals
- Among 7 measure 6 measure are reported by 1568 hospitals

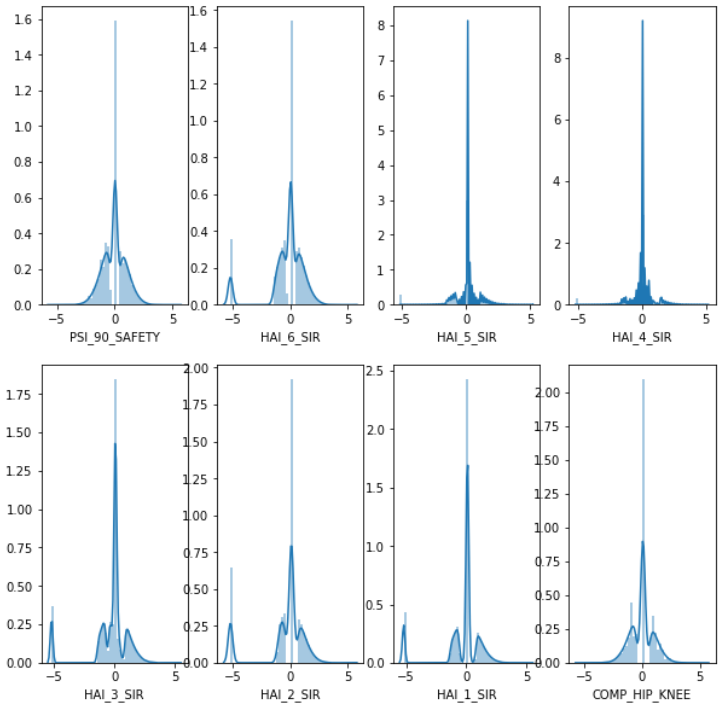
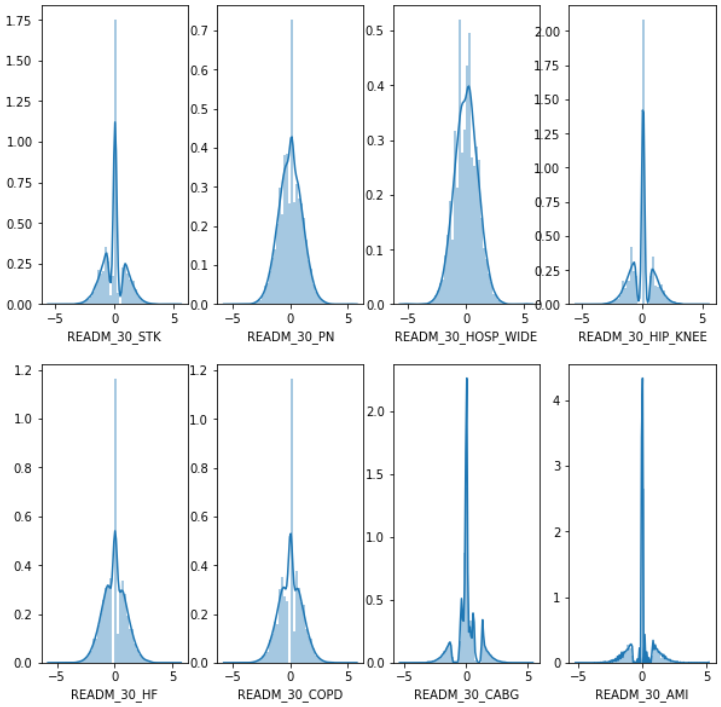
- 871 hospitals reported no measures

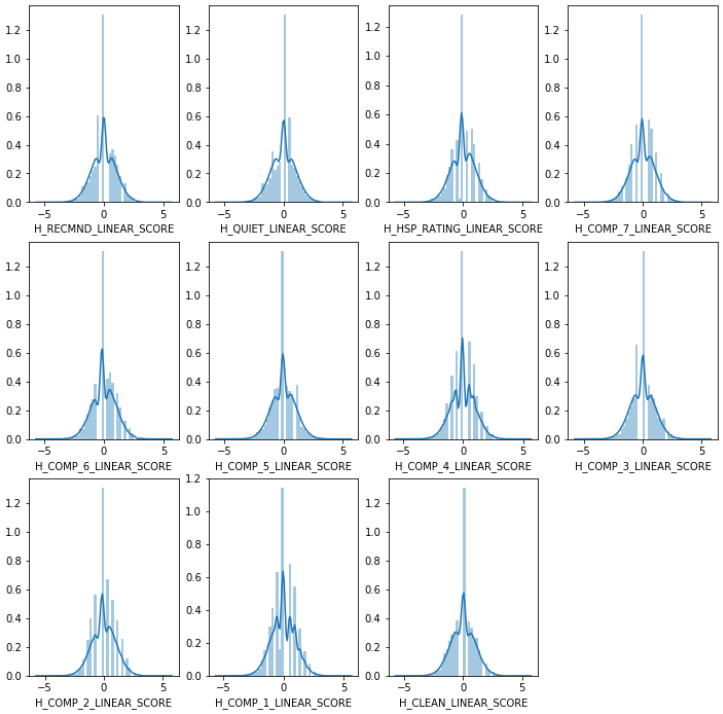
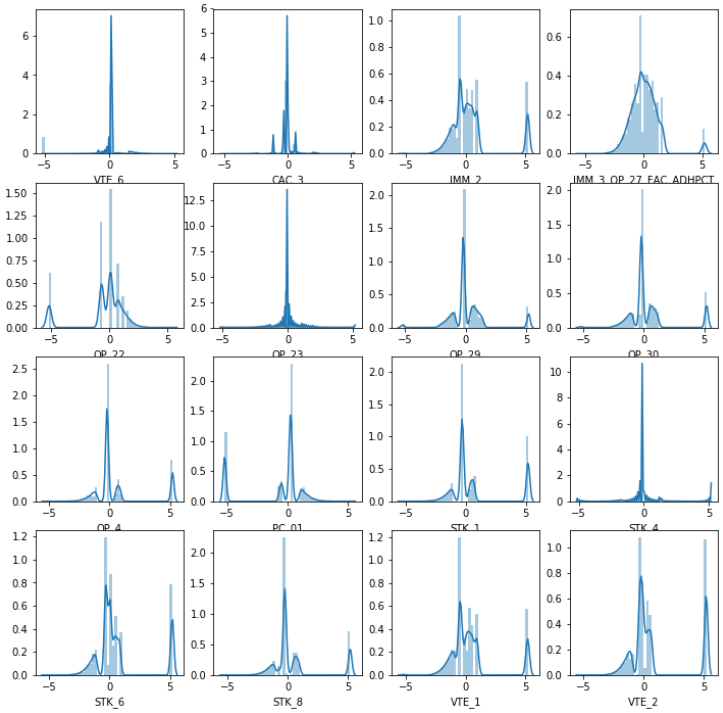


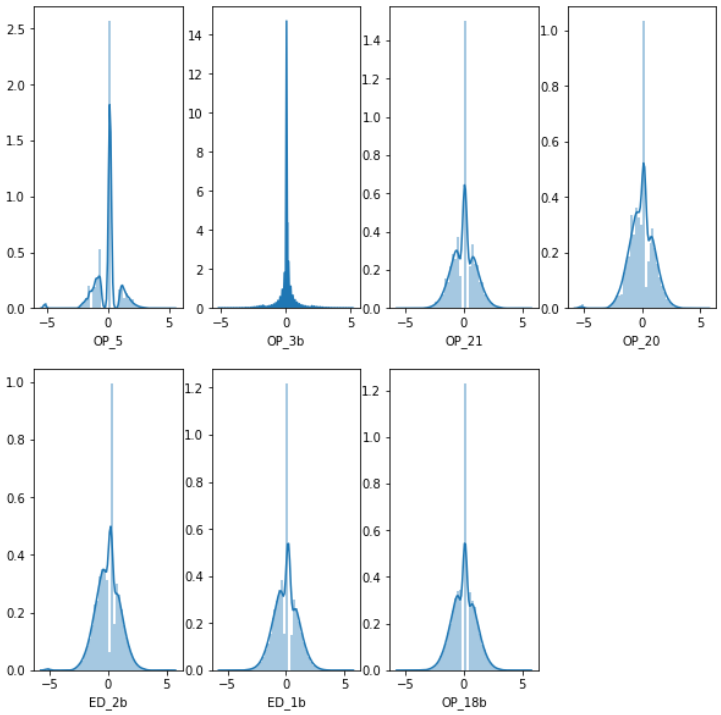
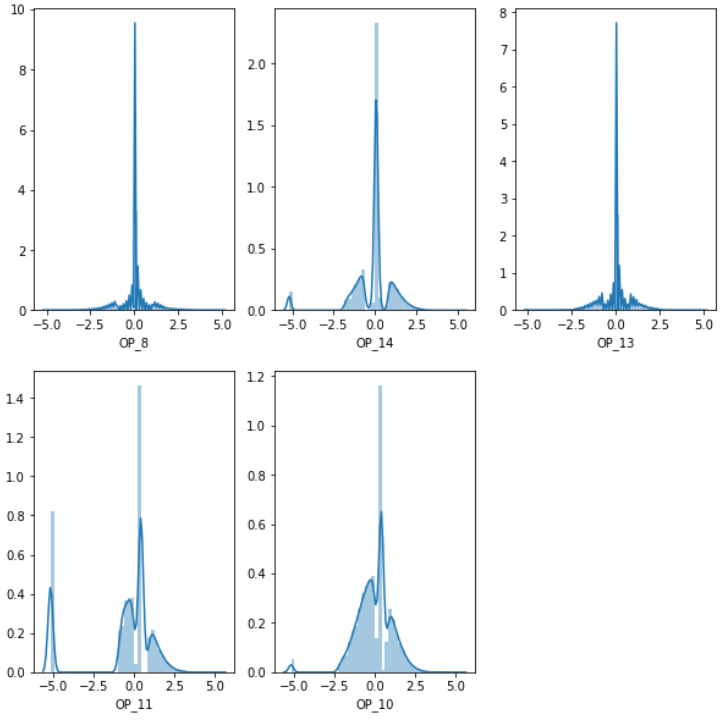
- In this majority of the data is missing i.e the hospitals which does not provide any data are 1083
- Among 5 measures in this group 1076 hospitals reports all the measures

- <u>CHECKING THE DISTRIBUTION OF THE DATA</u>

Mortality distribution	

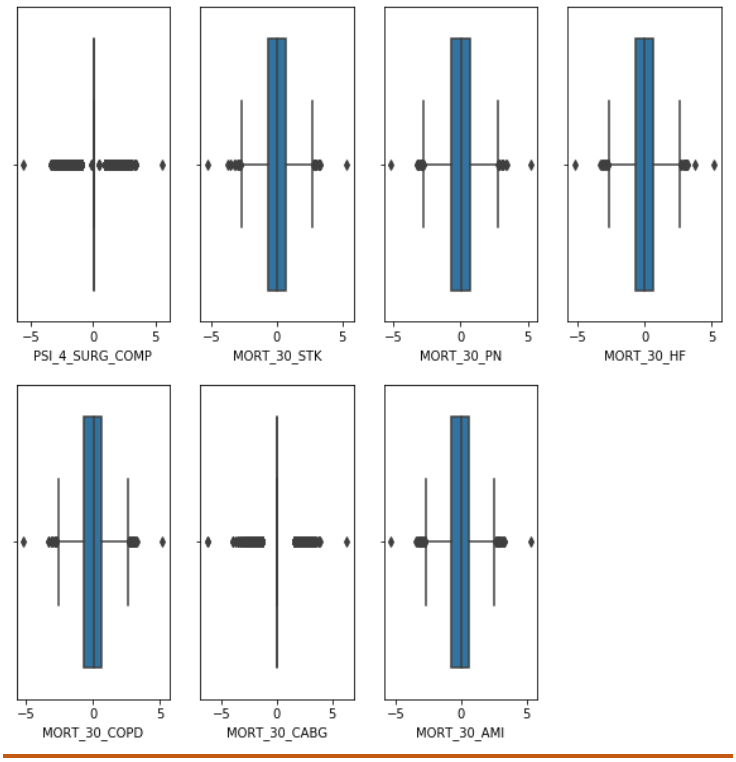
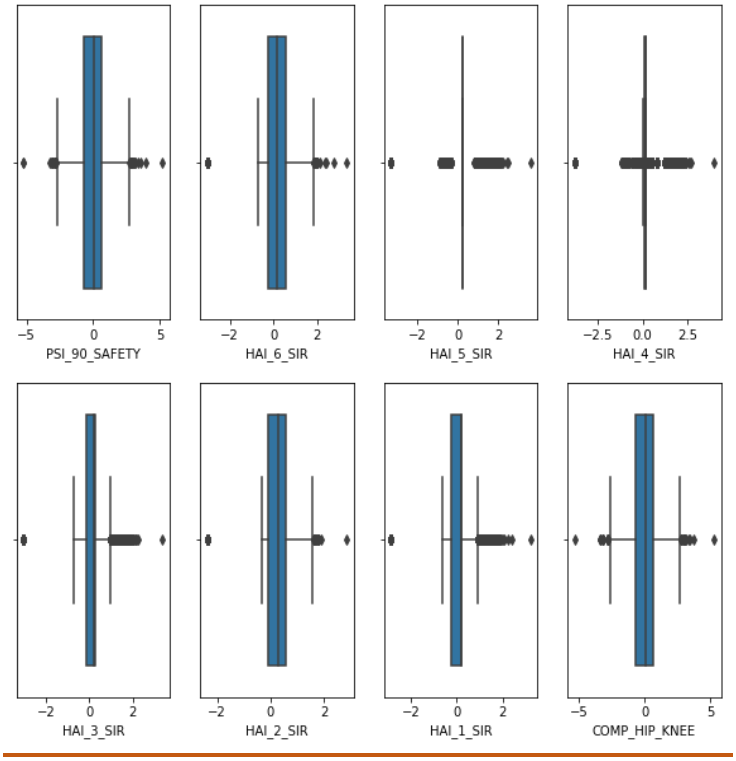
Safety of Care distribution	
Readmission distribution	

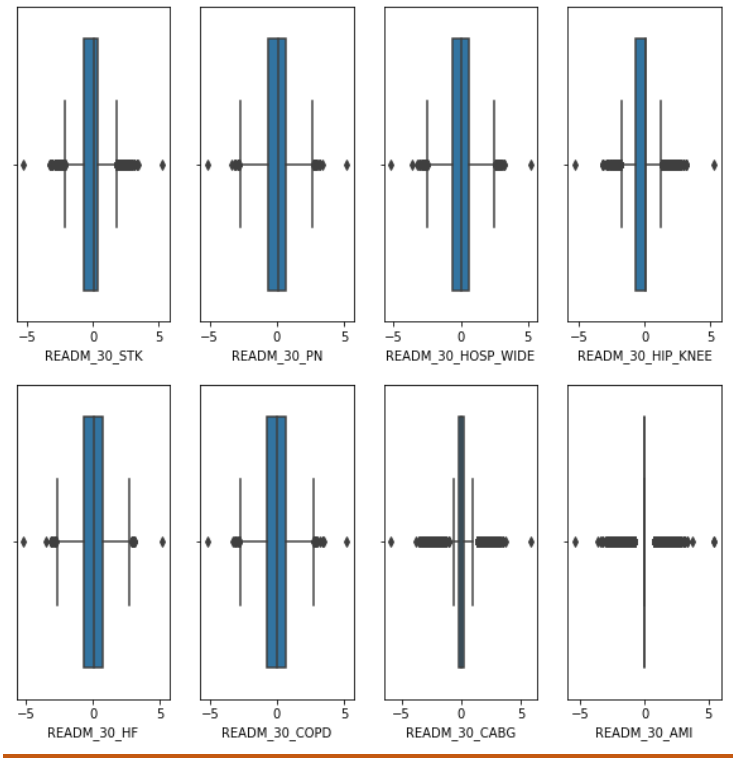
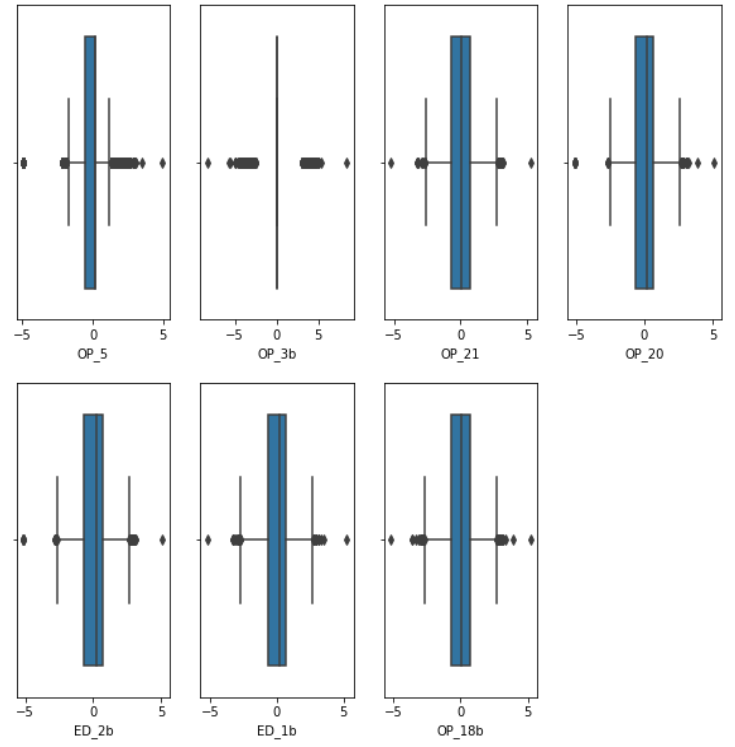
Patient Experience	
Effectiveness of care	

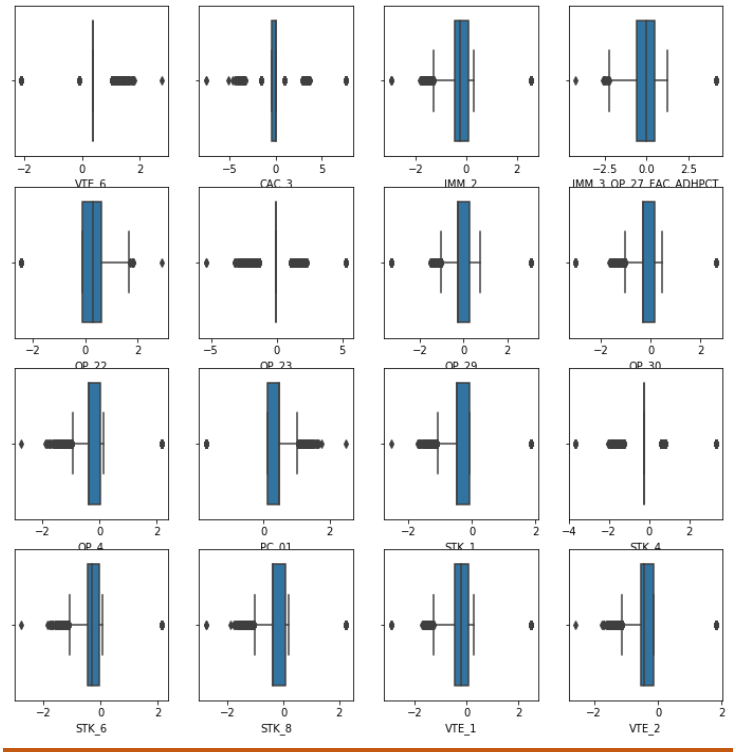
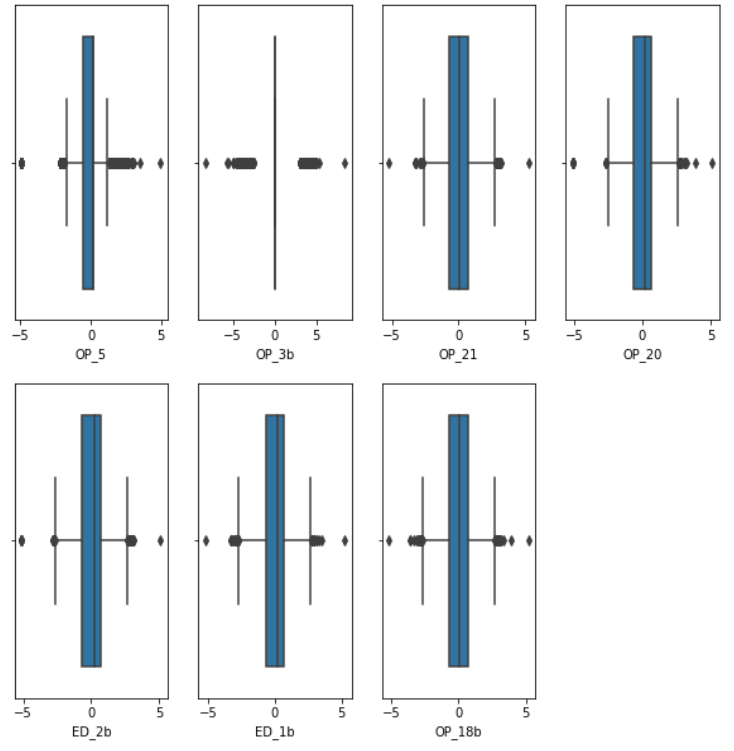
Timeliness of care	
Efficient use of Medical Imaging	

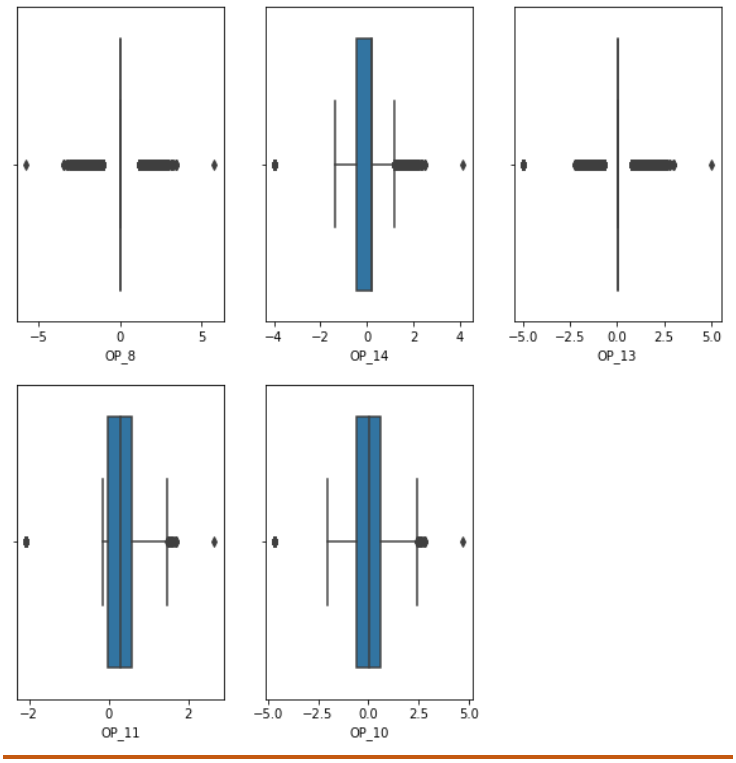
Quantile transformation is applied to normalize the data

- <u>CHECKING FOR OUTLIERS</u>

Mortality	
Safety of Care	

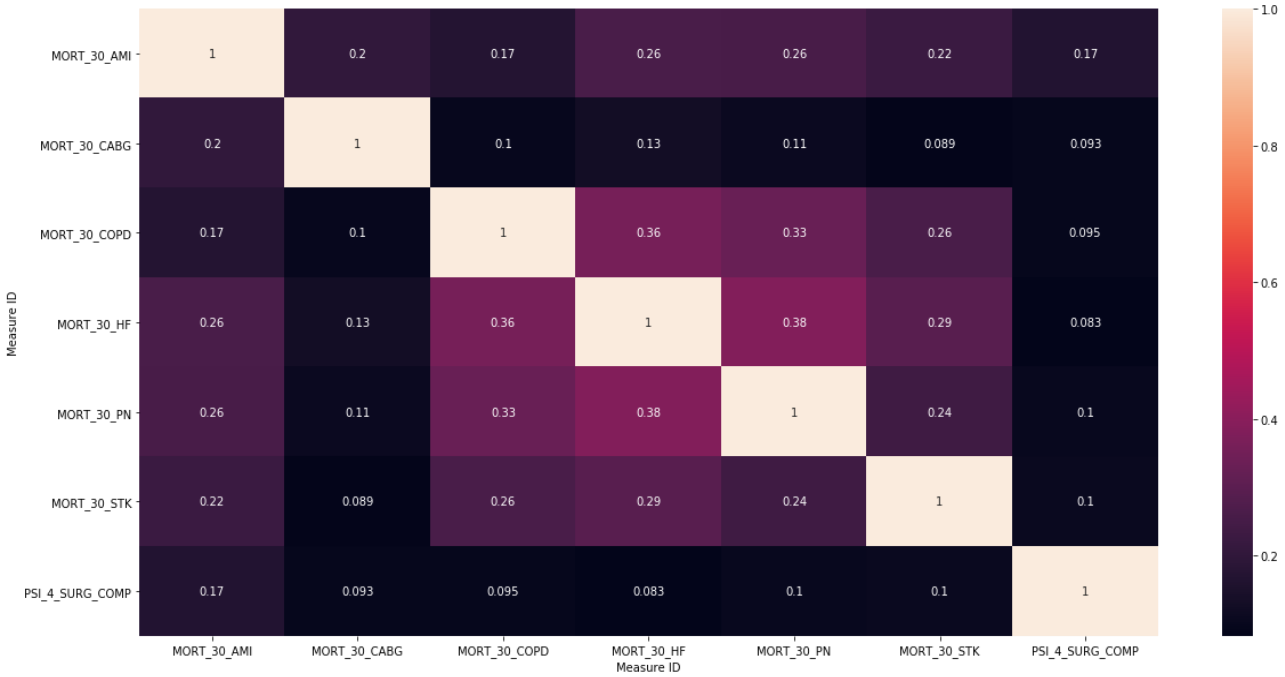
Readmission	
Patient Experience	

Effectiveness of care	
Timeliness of care	

Efficient use of Medical Imaging	

- CORRELATION MATRIX

Mortality

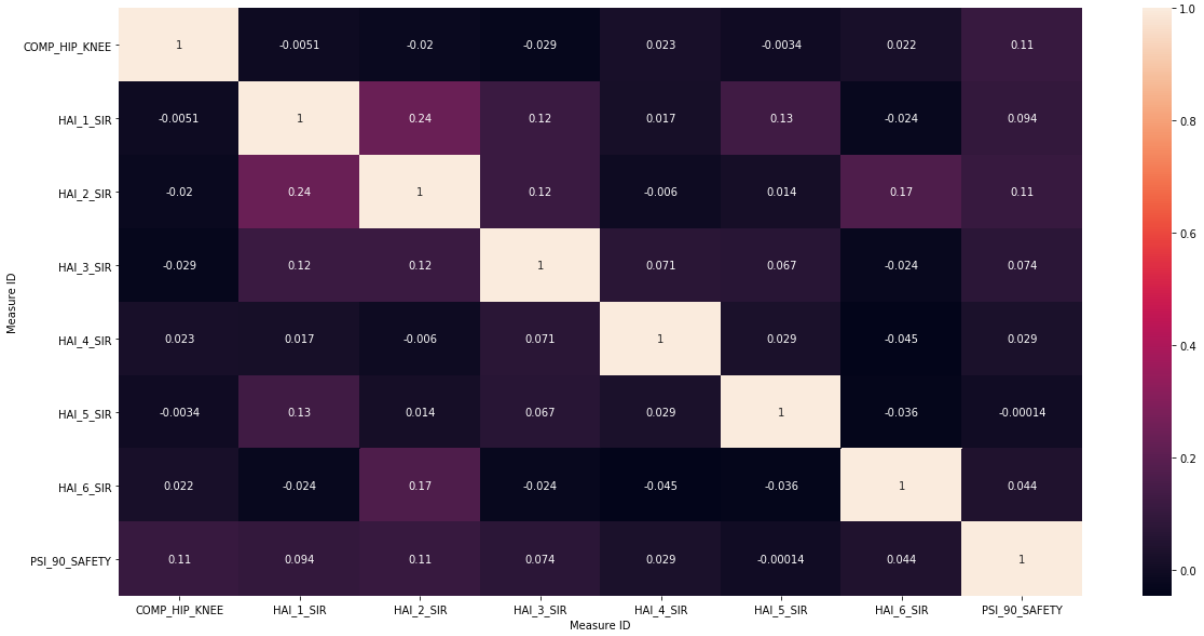


- From the correlation matrix the measure “PSI_4_SURG_COMP” is least correlated to

other measures in the mortality dataframe

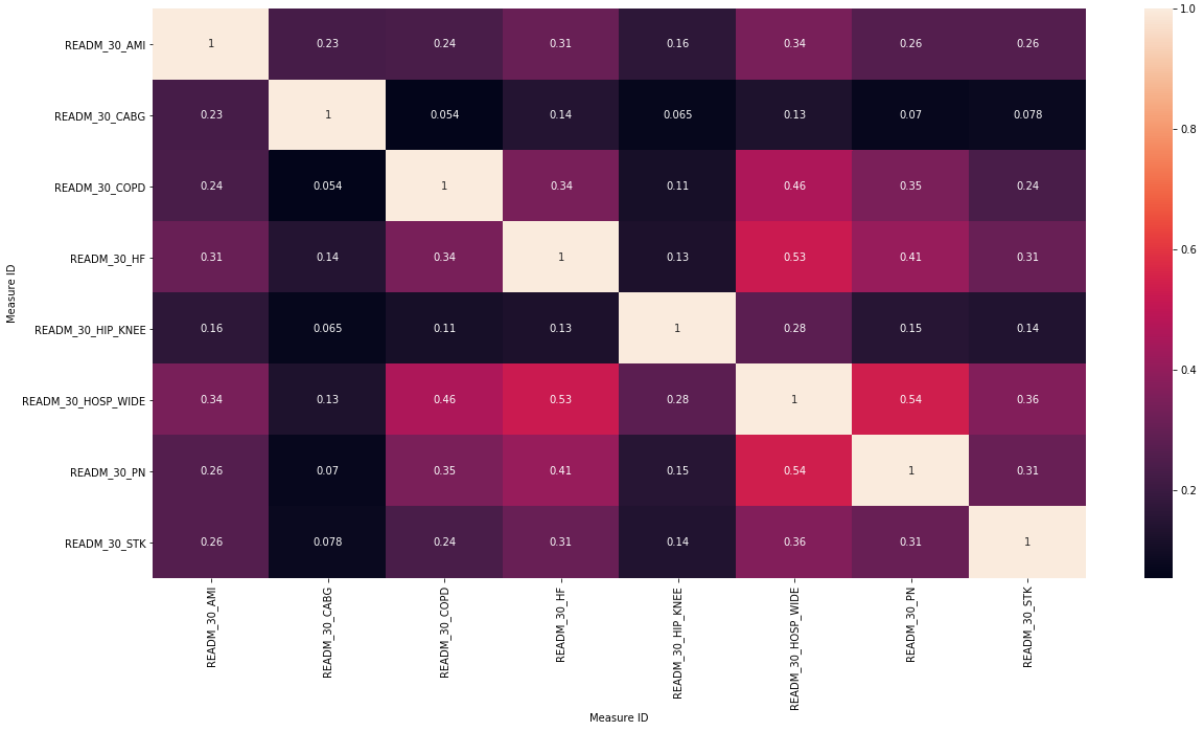
- MORT_30_HF is correlated to MORT_30_COPD
- MORT_30_PN is correlated to MORT_30_HF

Safety of Care



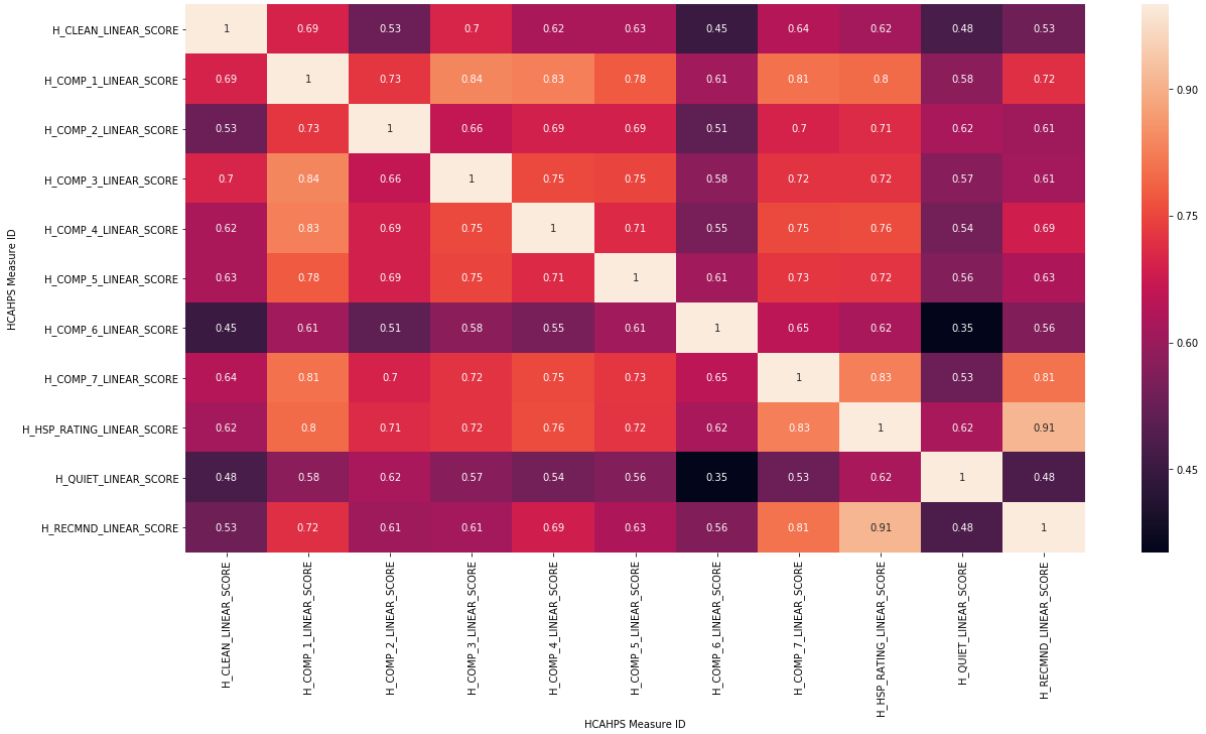
- COMP_HIP_KNEE is least correlated to other measures in this data frame
- The highest correlated variable in this matrix is HAI_2_SIR to HAI_1_SIR.

Readmission



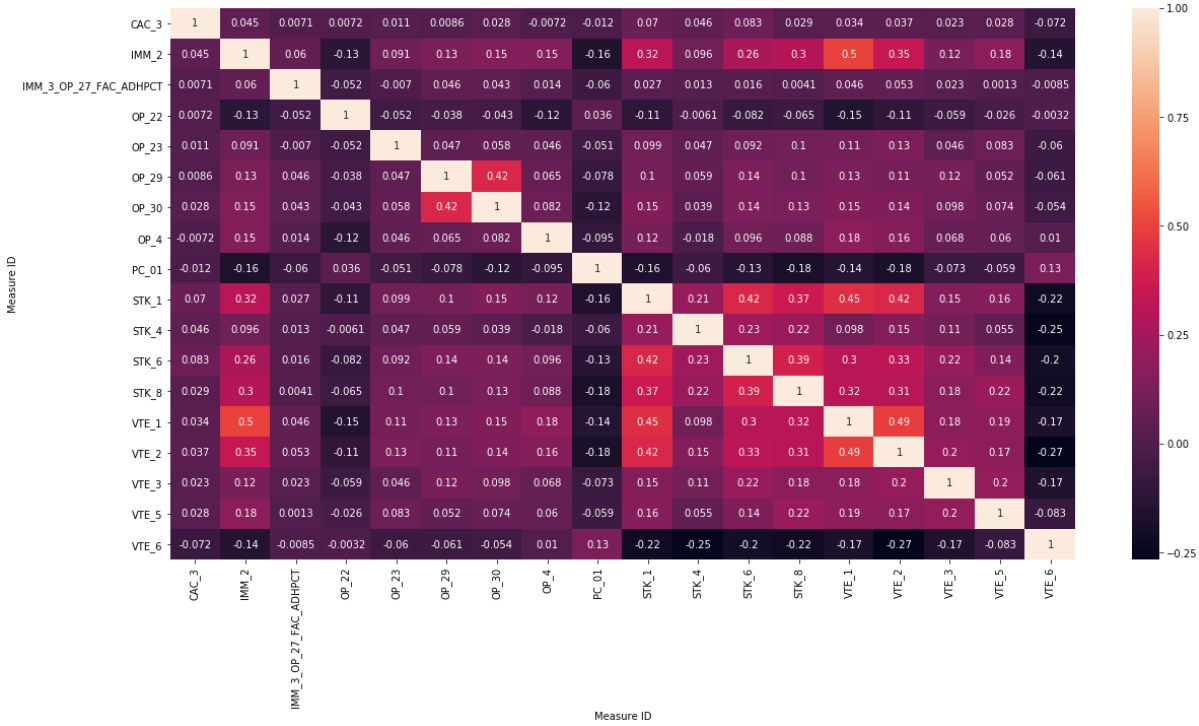
- In readmission data the variable are highly correlated.
- READ_30_HOSP_WIDE is correlated to READ_30_HF
- READM_30_CABG is least correlated to all the other variable in the data frame

Patient Experience



- In this data frame most of the variables are highly correlated to each other
- H_HSP_RATING_LINEAR_SCORE is highly correlated to H_RECMND_LINEAR_SCORE
- H_COMP_1 _LINEAR_SCORE is highly correlated to H_COMP_3_LINEAR_SCORE
- H_COMP_1 _LINEAR_SCORE is correlated to other variable in the data frame
-

Effectiveness of care



- Among 18 features in this data frame STK_1 is the highly correlated to other variable like STK_6, STK_8,VTE_1,VTE_2
- OP_30 is correlated to OP_29
- VTE_1 is correlated to IMM_2
- VTE_2 is correlated to VTE_1

Timeliness of care

- There are 7 measures in this data frame among them the highly correlated variables are ED_2b to ED_1b
- Op_18B is correlated to ED_1b with a correlation factor of 0.62
- Op_18B is correlated to ED_2B with a correlation factor of 0.52

Efficient use of Medical Imaging

- There are 5 important measures in this data frame among them OP_11 is correlated to OP_10 with a correlation factor of 0.35

Future Roadmap:

Modelling:

Unsupervised Learning

- Calculation of loadings using Factor Analysis as this is the best method to get the influence/weight of each measure within a group
- Calculation of Group Scores based on the weights calculated using Factor Analysis.
- Calculation of Summary Score by aggregating the weighted average of group scores.
- Use K-Means Clustering for assigning the star rating to the hospitals based on their summary scores.
- Compare the ratings assigned with that provided by CMS
- Evaluate the explanatory power and stability of the model

Supervised Learning

- The following models are considered for the analysis
    - Random Forests : (For its stability and diversity)
    - Linear Regression: (For its simplicity and ease of implementation)
- Evaluate the results of the random forest and linear regression models

Recommendations for the provider

- Do the provider analysis and suggest the measures to improve the star rating based on the learning from above analysis.