Healthcare Hospital Rating

Business understanding CMS rates hospitals in the US on a scale of 1-5 with the objective to make it easier for patients and consumers to compare the quality of hospitals.

The ratings directly influence the choice of hospital made by consumers and may have a significant impact on the revenue earned by hospitals. Thus, it is extremely important for hospitals to understand the methodology used by CMS for calculating the ratings so that they can work on improving the factors that influence them.

This project is focused on developing an approach to calculate hospital ratings and using it to identify areas of improvement for certain hospitals. It will also require a thorough understanding of the rating system developed by CMS.

The analysis is divided into four parts:

- Data Understanding – Groups and Measures
- Identifying important measures affecting star ratings
- Predictive modelling of star ratings
- Provider analysis: Recommending ways for Evanston Hospital to improve their rating

Data understanding

The data can be downloaded from the hospital compare website.

The main problem is to calculate the 7 group scores for each hospital.

Overall Star Rating Groups are:

- Mortality
- Safety of Care
- Readmission
- Patient Experience
- Effectiveness of Care
- Timeliness of Care

- Efficient Use of Medical Imaging

The rationale for these seven groups is as follows:

- The seven groups are aligned with the CMS Hospital Value-Based Purchasing (HVBP) program, the current categories on the Hospital Compare website, and other national quality initiatives.
- The groups are clinically reasonable in that they capture common components of quality for which hospital quality is likely linked across measures.
- The groups allow for future measures to be added or removed from the Star Ratings.

Data Cleaning/Updating:

The three main data quality issues in the raw data provided by hospital compare are:

Data format

- The original data is in 'wide-format' in approx. 60 files which was converted into one 'long' master file such that each row represents a provider and each column a measure
- Each cell is a numeric score of a measure

Standardization of Measures

- Measures need to be standardized such that 'higher value indicates better performance'

Measures (some examples)

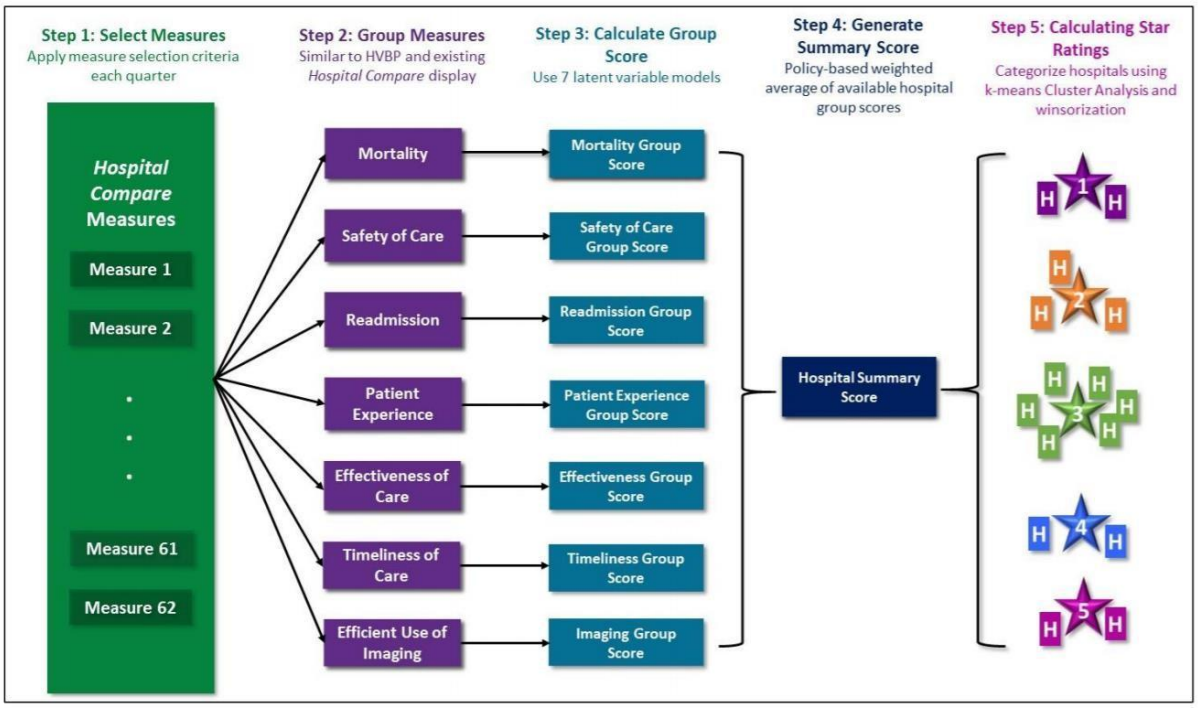
- Positive measures: Patients given appropriate vaccines; Patients given timely treatment etc.
- Negative measures: All mortality measures, readmission measures, timeliness measures (avg. time taken to provide emergency care etc.)

Missing values

- About 50% measures have a large number of missing values – they have been imputed as per the guidelines provided by CMS

Within each group, measure weights are proportional to the correlation of the measure within the group.

Approach to calculate Star Ratings of Hospitals:

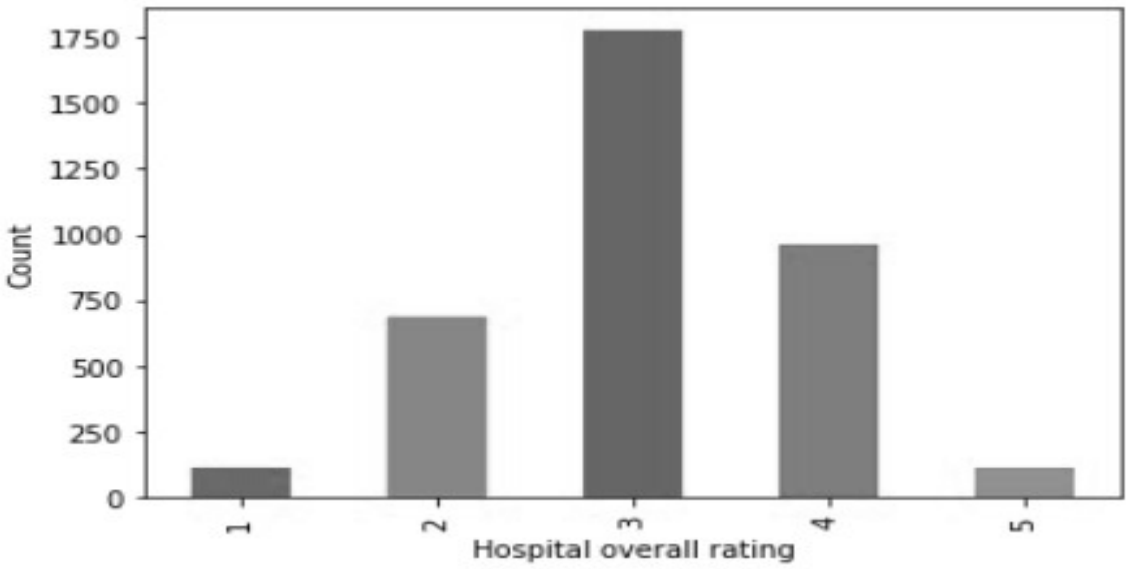


Steps followed for data cleaning:

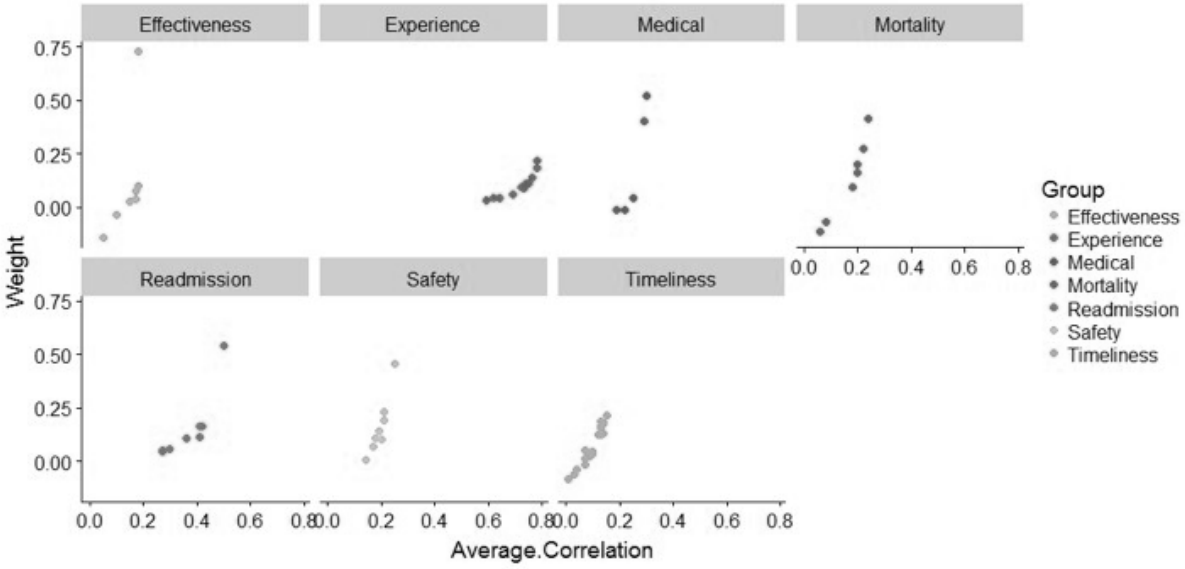
1. Read the respective CSV files for Measure indicators
2. Remove the un-necessary columns and keep Provider ID, Measure ID, Score and Hospital Rating
3. Update the Measure values in the table as per the listed 62 measures
4. Filter the table and keep only records where Measure value is in the list of 62 measures
5. Remove the duplicate rows and Null rows
6. Pivot the tables individually to get the unique rows for unique provider id
7. After working independently on all the tables, do outer merge on the all the measure tables
8. Filter the Hospital general information table and keep only Provider ID and Rating column
9. Finally merge the Hospital information table and combine tables of measure indicator
10. Update the values of Not applicable and not available as 0
11. Remove the rows which has Hospital Rating as 0
12. Final Shape of the table comes out to be **3648 X 63**

Exploratory Data Analysis:

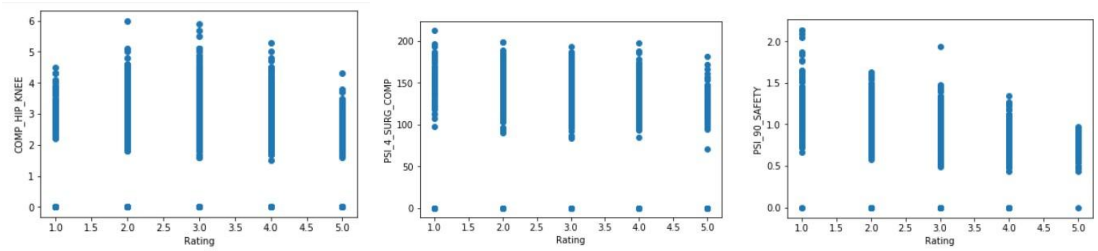
1. Count of Provider with Hospital Rating:



2. Co-relation matrix as per group:



3. Verify the outliers and remove those values – Please find below few of the charts for reference



4. Remove the columns for which 75% of the values are 0
5. Verify the individual factors against the Hospital Rating
6. Replace NaN values with 0

Model preparation

- Supervised Learning
- Un-Supervised Learning

Get the important measures, which has huge impact on ratings.

- Get the average score of hospitals (where rating is 4) and EVANSTON HOSPITAL for those important measure
- Compare the score against the EVANSTON HOSPITAL
- Identify the measure areas where the score of EVANSTON HOSPITAL needs improvement, which will be vital to improve the rating to at-least 4 by next year.