**Problem Statement**

**Here's a brief background of the business problem**

**Problem Background**

The Centers for Medicare & Medicaid Services ([CMS](http://www.cms.gov/)) is an agency of the U.S. Department of Health and Human Services (HHS) which administers the major government health programs in the U.S. This includes insurance programs such as Medicare, Medicaid, and the Children’s Health Insurance Program (CHIP). CMS also plays an important role in collecting and analyzing data with the goal of improving efficiency and equity in the healthcare system. In the current project, you will be analyzing CMS data on hospital quality.

CMS rates hospitals in the U.S. 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. CMS conducts surveys each year and keeps on updating the ratings to reflect the current services being delivered by each of the hospitals.

The ratings directly influence the choice of hospitals made by consumers and may significantly impact hospitals' revenues. This project is focused on developing an approach to calculate hospital ratings and using it to identify areas of improvement for hospitals.

**Problem Statement**

You are working as a consultant for a non-profit organization in the US that provides recommendations to hospitals on areas they need to improve so that they can get a better rating from CMS. This would affect the overall hospital revenue as well as improve the quality of life for the patients. Also, for hospitals that are yet to be rated, you can pre-emptively help them by providing a tentative rating based on their current services and suggesting where they need to invest more in improvement.

**Data Sets**

In this segment, you'll be introduced to the datasets that you'll be working on in this capstone project as well as a few additional reading materials.

**Data Set and Data Description**

For conducting the analysis, you have been provided with the following data. Check the 'hospital-info' CSV file:

Hospital Rating Classification

Download

**[The 'not\_yet\_rated.csv' data set will be helpful during Task 3 and you can ignore the same for now. The detailed list of tasks is present in the next segment]**

**Here's a quick overview of the columns available in the data set 'hospital-info.csv':**

* **Hospital Demographic Information**: This includes Provider ID (or, in other words, the Hospital ID), Hospital Name, Zip, Area, State, etc.
* **Hospital overall rating**: Overall rating of the hospital. By default, CMS rates hospitals on ratings 1-5.
* **Hospital Quality Information**: These are the columns that describe the scores received by hospitals based on different quality measures. These quality measures as defined by CMS (as per the year when the data was collected) are as follows:
  + Mortality
  + Safety of Care
  + Readmission
  + Patient Experience
  + Effectiveness of Care
  + Timeliness of Care
  + Efficient Use of Medical Imaging
* Each qualitative measure has been quantified in 2 ways:
  + **Measure national comparison**: For e.g., for a hospital's Mortality score, you can refer to the column 'Mortality National Comparison'. The score information can be understood as follows:
    - A score of 2 indicates the hospital performs better than the national average
    - A score of 1 indicates the hospital's performance is similar to the national average
    - A score of 0 indicates the hospital's performance is worse than the national average
  + **Sub-Measure level information**: For each of the measures, you also have the information of their sub-measures and their respective scores as well. This has been encoded in the following format: For e.g., for a hospital's Mortality sub-measure, you can find them in all the columns which have a MORT prefix in them (like MORT\_30\_AMI\_Score). Each of the sub-measures has a  combined effect on the overall measure rating of a hospital.

**Understanding the Measures/Sub-Measures**

**For this project, we have designed the scope so that a detailed understanding of the measures or sub-measures is not necessary for completing the tasks or providing recommendations. However, the measure and sub-measure names may seem unintuitive at first glance. And to help you understand the same, here's a brief explanation:**

**For example, take the "Efficient Use of Medical Imaging" measure. It measures the overall capability of a hospital to perform specific imaging procedures such as CT scans, MRI scans, etc. It has the following sub-measures:**

* MED\_OP\_10\_Score
* MED\_OP\_11\_Score
* MED\_OP\_13\_Score
* MED\_OP\_14\_Score
* MED\_OP\_8\_Score
* MED\_OP\_9\_Score

**Each of the above scores is mapped to one of the individual procedures. For e.g. MED\_OP\_10\_Score refers to CT scan, MED\_OP\_11\_Score may refer to MRI scans, and so on.**

**For this measure, a hospital would be:**

* scored for each of its individual sub-measures
* and also have a "national comparison rating"  where its facilities are compared to the national average for that measure.

**An intuitive way to understand the relationship might be that a hospital not having most of the imaging procedure facilities would have lower individual sub-measure scores, and there's a high chance tat its overall measure's national comparison rating would also be low.**

**Additional Reading Material:**

**If you want a detailed deep-dive into the process through CMS itself awards ratings, as well as a detailed understanding of the different measures, please go through the link provided**[**here**](https://bpb-us-w2.wpmucdn.com/u.osu.edu/dist/c/28860/files/2016/08/Star_Rtngs_CompMthdlgy_052016-148w094.pdf)

**Tasks**

**In this segment, you'll get a detailed list of tasks that you must perform in this Capstone Project. Towards the end, we have also provided you with a starter notebook that will help in completing this project.**

**Tasks**

**Task 1: Understand the data**

**Take some time to familiarize yourself with the data. What are the key variables?**

**Specifically, answer the following questions:**

1. Perform a few basic data quality checks to understand the different columns and prepare descriptive statistics for some of the important columns.
2. What is the distribution of "hospital overall ratings"? How are they varying across other parameters like ***State***? Create a few visualizations that provide some insights into the data.

**Task 2: Build machine learning models**

**Use your knowledge of classification models to create three models that predict hospital ratings. You should follow these steps:**

1. Prepare the data for the machine-learning model.
   * Remove all the demographic columns as well as any other unnecessary features from the data set
   * For simplification, instead of having five ratings, we will convert them to 0 and 1. Here 0 indicates that the hospital has been rated 3 or below, and 1 indicates that the hospital has been rated 4 or 5.  Encode the Hospital columns as follows
     + 1,2,3 : 0
     + 4,5: 1
   * Store the predictors and the target variable in variables X and y.
   * Create the dummy variables for categorical columns.
   * Split the data into train and test sets (70-30 split with random state 0. This random state is recommended, though you can use any other random state of your choice).
   * Scale the numerical columns using StandardScaler.
2. Build the 3 classification models that you have learned in this course on your dataset. Carefully apply regularization and hyperparameter tuning techniques to improve your model performance for each of the models.
3. Summarize the classification performance in terms of the necessary metrics such as accuracy, sensitivity, specificity, etc.

**Task 3: Providing Recommendations**

**You have now built three machine-learning models. Choose the best model according to your metrics and provide the following recommendations.**

1. **Hospital Rating Predictor**: Using the best model of your choice, predict the ratings of a few new hospitals which are yet to be assigned a rating by CMS. The information for these hospitals has been provided in a separate CSV file named 'not\_yet\_rated.csv'.
2. **Hospital Improvement Plan**: Let's say a few of the hospitals were rated low (0) by the model that you chose. Provide recommendations on how these hospitals can improve their ratings
   * **Hint**: To solve this question, you can identify the measures and sub-measures which are highly significant toward a hospital's overall rating. This can be done in several ways, for e.g., by using the coefficient signs in a logistic regression model. If a hospital has a lower rating, say for 'Mortality,' and it comes out a significant measure as per your model as compared to other measures, you can provide recommendations that the hospital needs to improve that measure.
   * In case none of the hospitals are rated low by your model, you can choose a few hospitals from the original dataset which have been rated 3 or below and provide similar recommendations.
   * Your recommendation should have a clear structure. You should provide the overall measure level recommendations first and then dive deeper into each of the sub-measures.

**Task 4: Executive Summary**

**Provide a detailed overview of the entire capstone project in a pdf document. Include the following points**

1. Summarize your entire analysis process
2. Prepare and share your final recommendations.

**Starter Notebook**

**We have shared additional resources in the starter notebook for this Capstone project. The notebook would contain detailed comments on performing each task with additional hints. Please download the same below:**

**Starter\_Code\_Hospital\_Rating\_Classification\_Capstone**

**Download**

**Evaluation Methodology**

* Note that you have been provided the option of choosing any ML model of your choice, just that you must build at least 3.
* Also, make sure that the final model that you choose is able to help you in providing recommendations for the last task of this overall assignment.
* You also have the freedom of choosing a random state and other parameters of your choice, which can be different from the ones you studied in the module material.
* There is no single solution for the above Capstone project, and you shall be awarded grades as long as your overall methodology is sound and you are able to provide recommendations that are soundly based on facts and evidence (on the basis of the final model that you select based on your analysis).