**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.

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.

You can read the details of the CMS rating project [here](https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html).

In this project, you will approach the ‘CMS rating problem’ from different angles. You can think of it as ‘re-engineering the CMS rating system’.

This project has three main parts, which are described below. Note that you are expected to conduct suitable EDA for each part of the problem and clearly mention the insights in your report.

The **template for the report**to be submitted is provided at the bottom of this page. Please download it.

**Data understanding**

The data can be downloaded from the [hospital compare website](https://data.medicare.gov/data/archives/hospital-compare). You need to download 'Hospital\_Revised\_FlatFiles\_20161110'.

**Modelling**

**Part 1 - Supervised Learning-Based Rating**

1. Think about the types of supervised models you will choose for predicting the star ratings (1-5, five-class classification). Consider all the practical as well as technical aspects involved. In the report, mention the types of models you will consider, the pros and cons of each, and the reasons for choosing certain models for this problem. Note that you have to **build two types of model**(as mentioned below).
2. Create a**Random Forest** **Model** and a**Linear Regression Model**( consider the ratings as ordinal variables) to predict the hospital ratings (1-5) using the 64 measures specified by the CMS.    
   Evaluate your model using relevant techniques. Report the relative importance of the measures according to their predictive power. Finally, report the overall accuracy of predictions using the actual star ratings for both the models.

**Part 2 - Clustering-Based Rating (Unsupervised)**

The methodology followed by CMS to assign ratings is described in[the report here](https://u.osu.edu/korzen.1/files/2016/08/Star_Rtngs_CompMthdlgy_052016-148w094.pdf). It is an unsupervised method. Create an unsupervised clustering model to assign hospital ratings (1-5) using the 64 measures specified by the CMS. The main problem is to calculate the 7 group scores for each hospital. Various methods exist to perform this analysis, two of which are mentioned below. The second method is similar to the one used by CMS. You can choose **either of these ways** to complete this part of the project.

* **Method 1:** Figure out your own way to assign weights to the measures in the respective groups(You can do dimensionality reduction using **PCA**which is similar to Method 2 mentioned below to perform this analysis) The method may or may not be similar to the one used by the CMS, though it should be based on sound reasoning. Using the weights, calculate 7 group scores for each hospital. You can then take the weighted average of group scores to calculate a final score of each hospital, perform clustering, and assign a star rating to each cluster.

* **Method 2:** This method uses factor analysis (or latent variable analysis) to find the weight of each measure in the respective groups. Using factor analysis, find the group score for each hospital. You can then take the weighted average of group scores to calculate the final score of each hospital, perform clustering, and assign a star rating to each cluster.

Using the method of your choice, develop a procedure to:

* Calculate and report the weights of all the measures within the respective groups
* Compare the weights obtained using this method with the ones obtained from the random forest and the linear model.
* Report the group score and the final score of each hospital
* Report the final star rating of each hospital
* Compare the rating assigned by you with the actual ratings assigned by CMS and report the analysis using relevant evaluation/comparison metrics, i.e., evaluate your unsupervised model

**Part 3 - Recommendations for Hospitals**

Suppose you work for a healthcare consulting company. The hospital with Provider ID = 140010 (EVANSTON HOSPITAL) is your client.

The hospital’s current star rating is 3, and it wants to improve the rating to at least 4 next year. The data required for this analysis is already available to you.

Using your understanding of the star rating system, recommend ways to improve the rating. The structure you follow for this part is up to you.

**Mentorship**

Given below is the schedule for the mentorship sessions and the milestone you are expected to achieve before each session.

|  |  |
| --- | --- |
| **Mentorship Session** | **Agenda** |
| **6th - 12th**  **August** | Understanding of the business problem and the data;  Overall structure to solve the problem. May discuss CMS’s approach to award ratings. |
| **13th - 19th**  **August** | Results of supervised learning (random forest and linear regression) model, identifying the important variables and improving the model results. May discuss the approach to attempt the eventual unsupervised tasks. |
| **27th - 2nd September** | Results of the unsupervised clustering/factor analysis/PCA tasks, ways to improve the same and identification of important variables. |
| **3rd -9th September** | Structure for the provider analysis problem, presentation of results to the client — important variables, results (in simple terms), etc. |

**Do's and Don'ts for Mentorship Sessions**

* Be prepared for the session as per the agenda prescribed above.
* Ask relevant queries that help you clarify your doubts.
* Do not push the mentor to provide the solution.
* Do not ask the mentor to review your Python codes.
* Respect the time given to you by the mentor: If any mentorship session goes unattended by all the members of a group, no further mentor support will be available to the group.