# *ANALYSIS ON REDUCING THE MEDICARE PAYMENTS FOR INPATIENT PPS HOSPITALS WITH EXCESS RE-ADMISSIONS*



California State University, Los Angeles

CIS-5210 Healthcare Data Analytics

By: Ritika Joshi

**[A] Data Set URLs**

|  |  |  |
| --- | --- | --- |
| 1. | Medicare Spending Per Beneficiary – Hospital | <https://catalog.data.gov/dataset/medicare-hospital-spending-per-patient-hospital-da578> |
| 2. | Hospital Readmissions Reduction Program | <https://catalog.data.gov/dataset/hospital-readmissions-reduction-program> |

**[B] Data set description**

1. Medicare Spending Per Beneficiary – Hospital

This data set is about MSPB Measure, which shows whether Medicare spends more, less or about the same for an episode of care at a specific hospital compared to all hospitals nationally. Each MSPB episode includes medicare payments for services provided by hospitals and other healthcare providers the 3 days prior to, during, and 30 days following a patient’s inpatient stay. This measure evaluates hospitals’ costs compared to the costs of the national median (or midpoint) hospital.

Purpose of using this dataset is to analyse which all states face the over spending on patients in Medicare. Medicare is for 65 and above aged citizens of USA and are more likely towards being admitted to the hospitals. We can also check providers who are affected due to overspending. More focus in this study shall be put on these states and providers.

This measure considers important factors like patient age and health status (risk adjustment) and geographic payment differences (payment-standardization).

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Names** | **Data type** | **Type of values** | **Description** |
| Provider ID | Numeric | Nominal | a 5-digt national identification number of the Medicare providers |
| Hospital Name | String | Nominal | from where the records have been gathered |
| Address | String | Nominal | Address of the hospital |
| City | String | Nominal | City name of the hospital |
| State | String | Nominal | State name of the hospital |
| Zip Code | Numeric | Nominal | Zip Code name of the hospital |
| County Name | String | Nominal | County name of the hospital |
| Phone number | Numeric | Nominal | Phone number of the hospital |
| Measure Name | String | Categorical | Type of measure for which data is been collected i.e. spending per patient. |
| Measure ID | String | Nominal | Identification number of the measure type. |
| Score | Float | Ratio | Fraction of spending on one patient compared to all hospitals nationally. |
| Footnote | String | Text | Comment on from where and how the data was considered in this dataset. |
| Measure Start Date | Date | Ordinal | Date of the start of taking counts i.e. Jan 2017 |
| Measure End Date | Date | Ordinal | Date of the end of taking counts i.e. Dec 2017 |
| Location | String | Text | Street of the hospital |

# 2. Hospital Readmissions Reduction Program

In October 2012, CMS began reducing Medicare payments for Inpatient Prospective Payment System hospitals with excess readmissions. Excess readmissions are measured by a ratio, calculated by dividing a hospital’s number of “predicted” 30-day readmissions for by the number that would be “expected,” based on an average hospital with similar patients.

Below are the measure types for which hospitals have collected data for reduction program for readmissions. These measure types were used through out the analysis done under this project.

heart attack - AMI

heart failure - HF

pneumonia - PN

chronic obstructive pulmonary disease - COPD

hip/knee replacement - THA/TKA

coronary artery bypass graft surgery – CABG

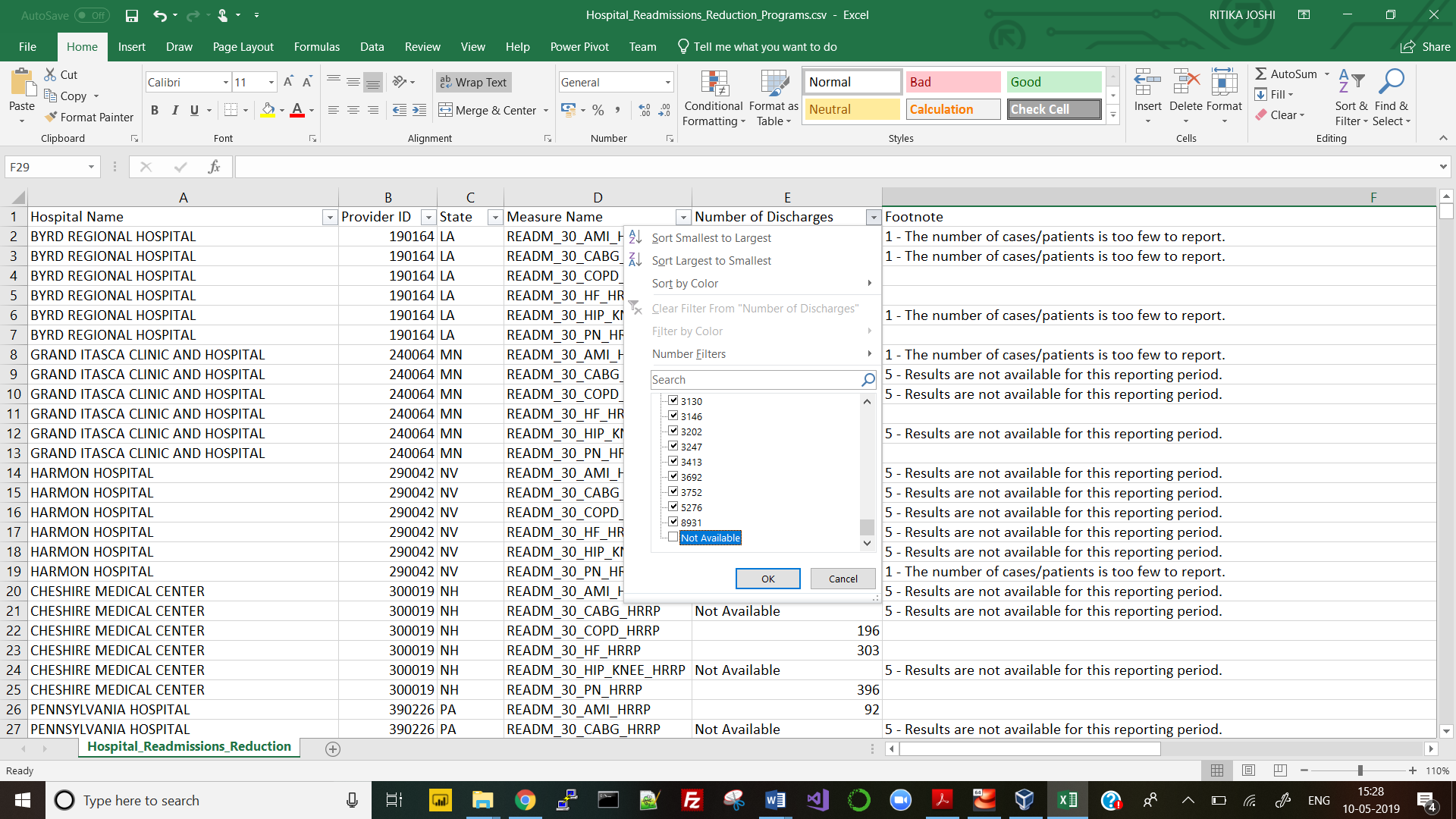
|  |  |  |  |
| --- | --- | --- | --- |
| **Column Names** | **Data type** | **Type of values** | **Description** |
| Provider ID | Numeric | Nominal | Identification number of the service provider |
| State | String | Nominal | State of residence of the service provider |
| Measure Name | String | Categorical | Categories of health issues for which patients have been admitted into the hospitals. |
| Number of Discharges | Numeric | Ratio | Total number of discharge count made by hospital per measure type |
| Excess Readmission Ratio | Float | Ratio | Ratio of a hospital’s number of “predicted” by the number that would be “expected” based on an average hospital |
| Predicted Readmission Rate | Float | Rate |  |
| Expected Readmission Rate | Float | Rate |  |
| Number of Readmissions | Numeric | Ratio | Count of readmissions per hospital per measure type |
| Start Date | Date | Ordinal | Date of the start of taking counts i.e. Jan 7, 2014 |
| End Date | Date | Ordinal | Date of the end of taking counts i.e. June 30, 2017 |

**[C] Data Refinement**

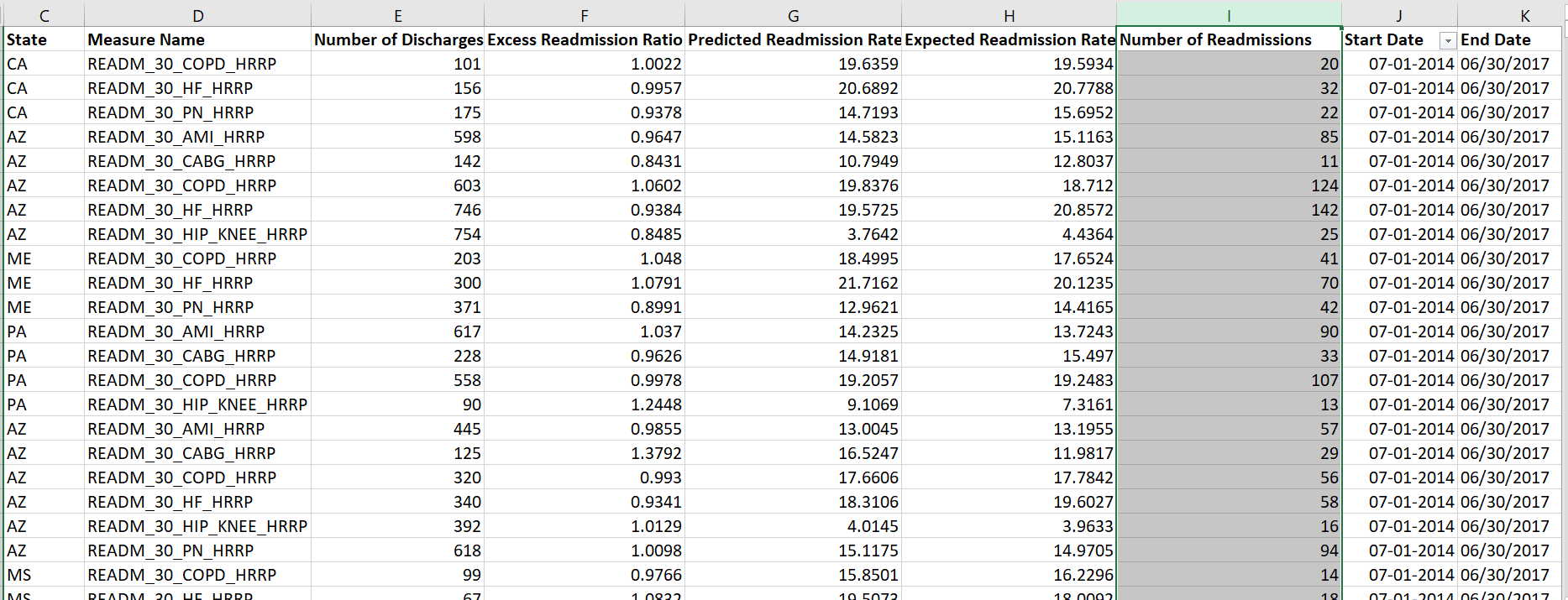
1. **Missing Values:** Values in few of the rows were missing under ‘Footnote’ and ‘Number of Discharges’ from Medicare Spending Per Beneficiary – Hospital and Reductions of re-admissions from 2014 to 2017 respectively.

Using excel missing values were removed as shown in the screenshot.

Pre

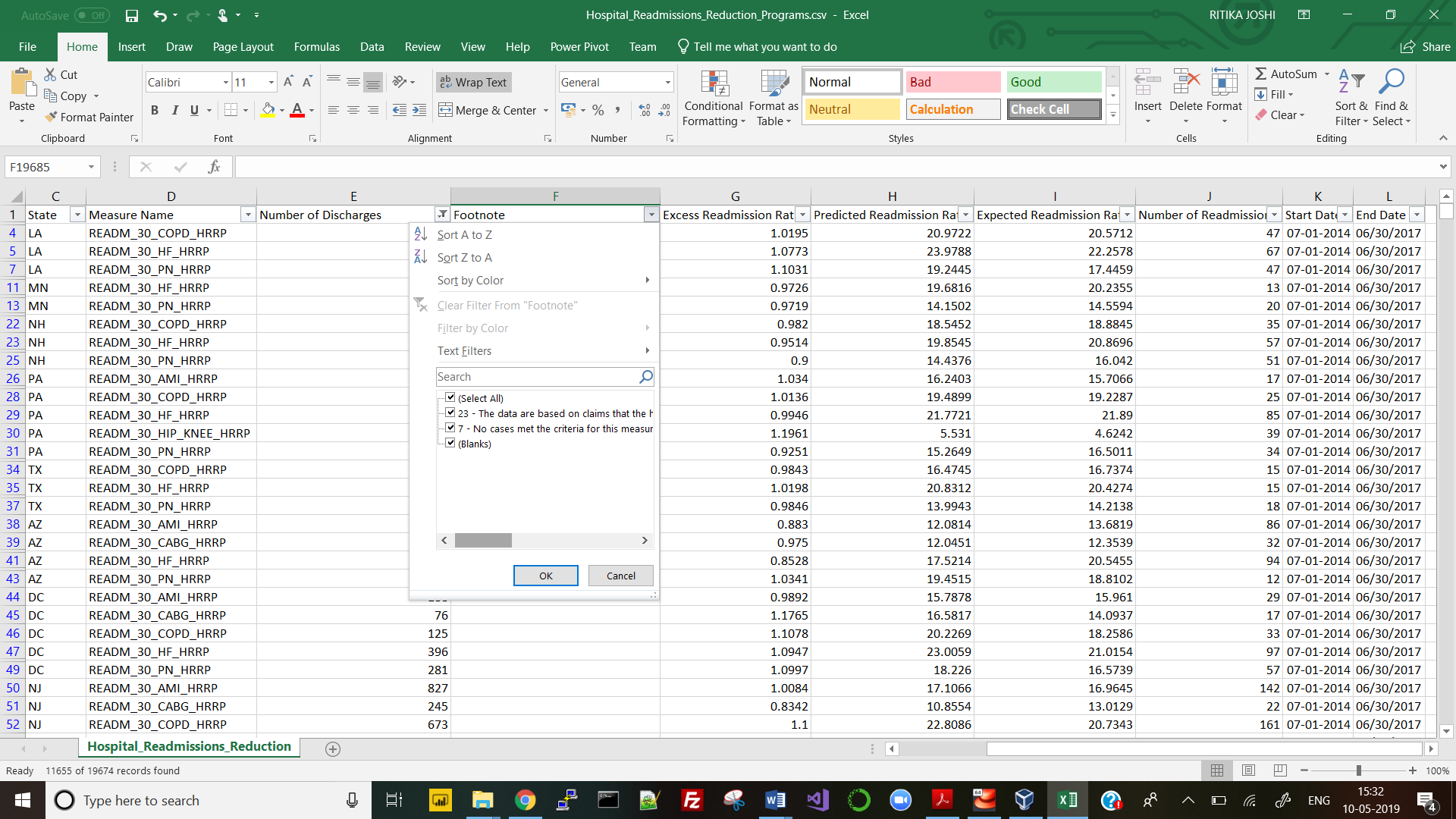


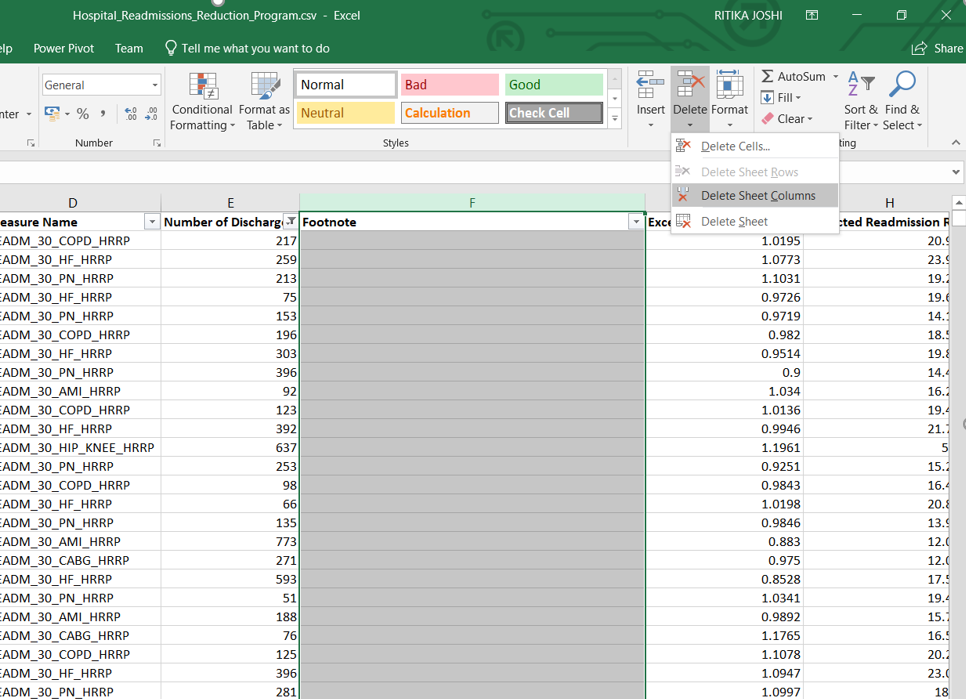
Post



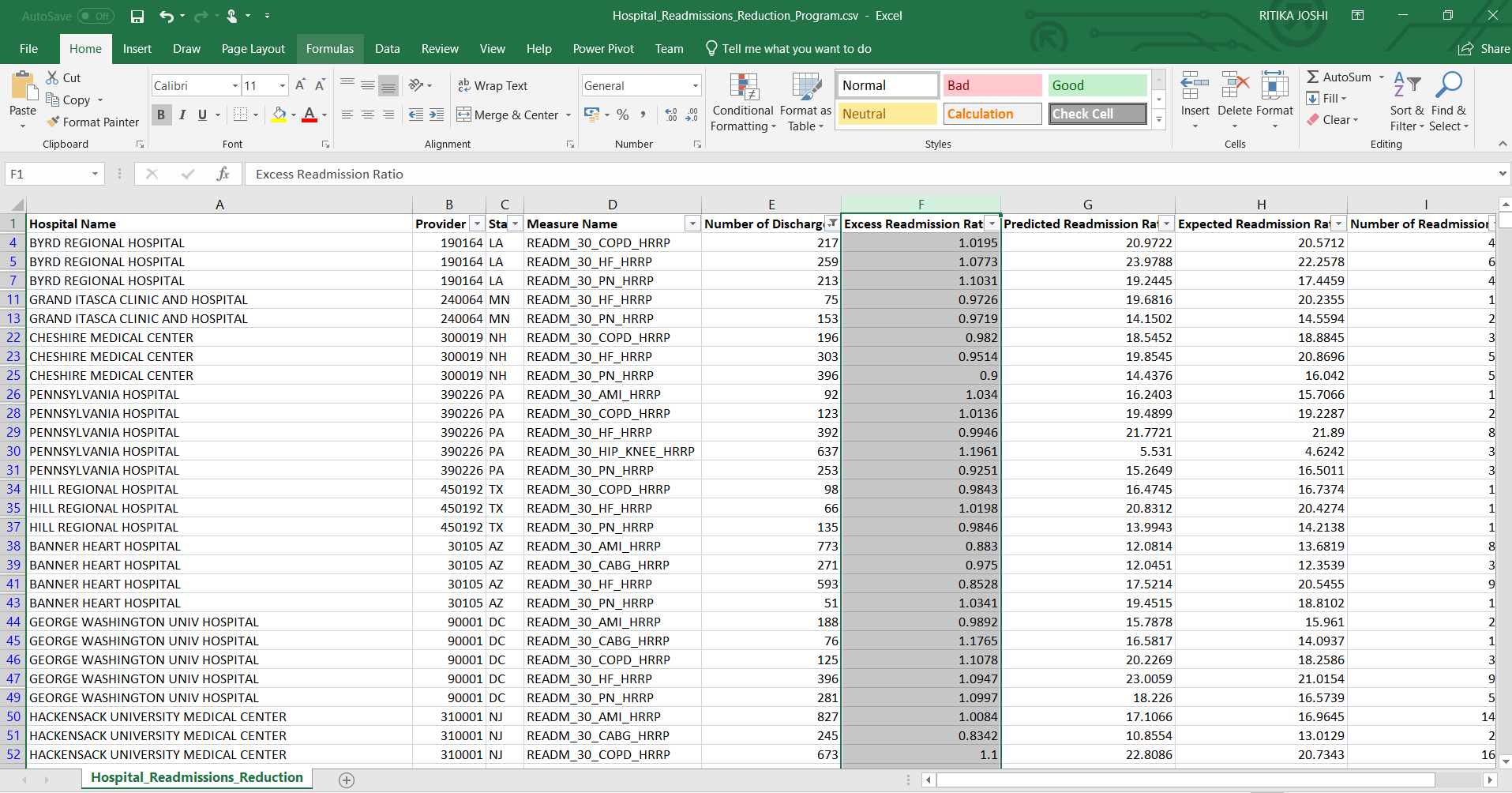
1. **Illegible columns/values:** In any of the columns under both data sets there were no illegible values. But values under ‘Footnote’ column were going to be of no use. It contained two types of comments about the data. This column also had blank values. To increase the execution speed of SAS tool, this column was removed.

Pre

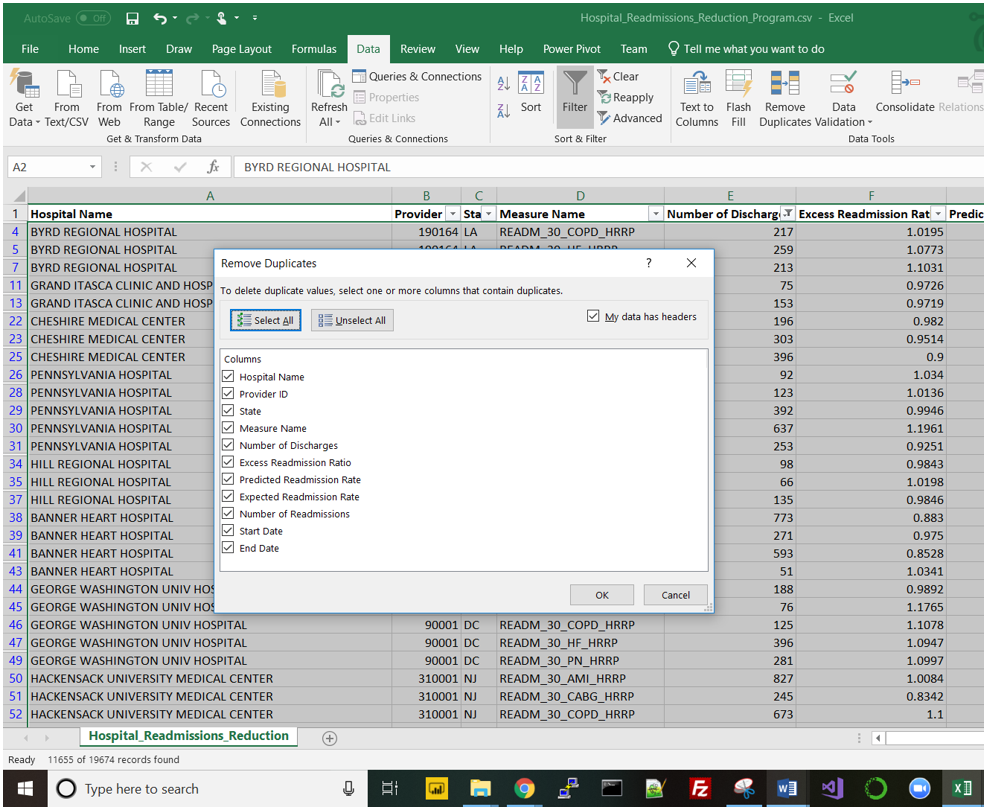


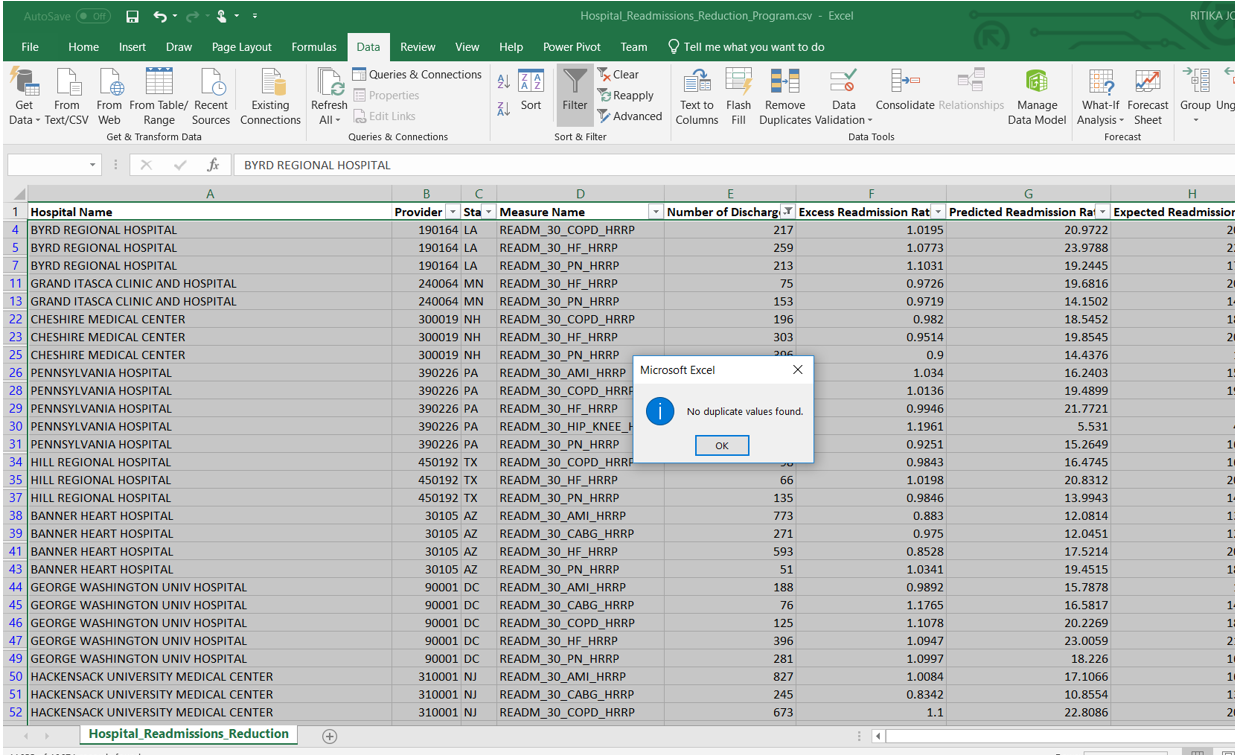


Post

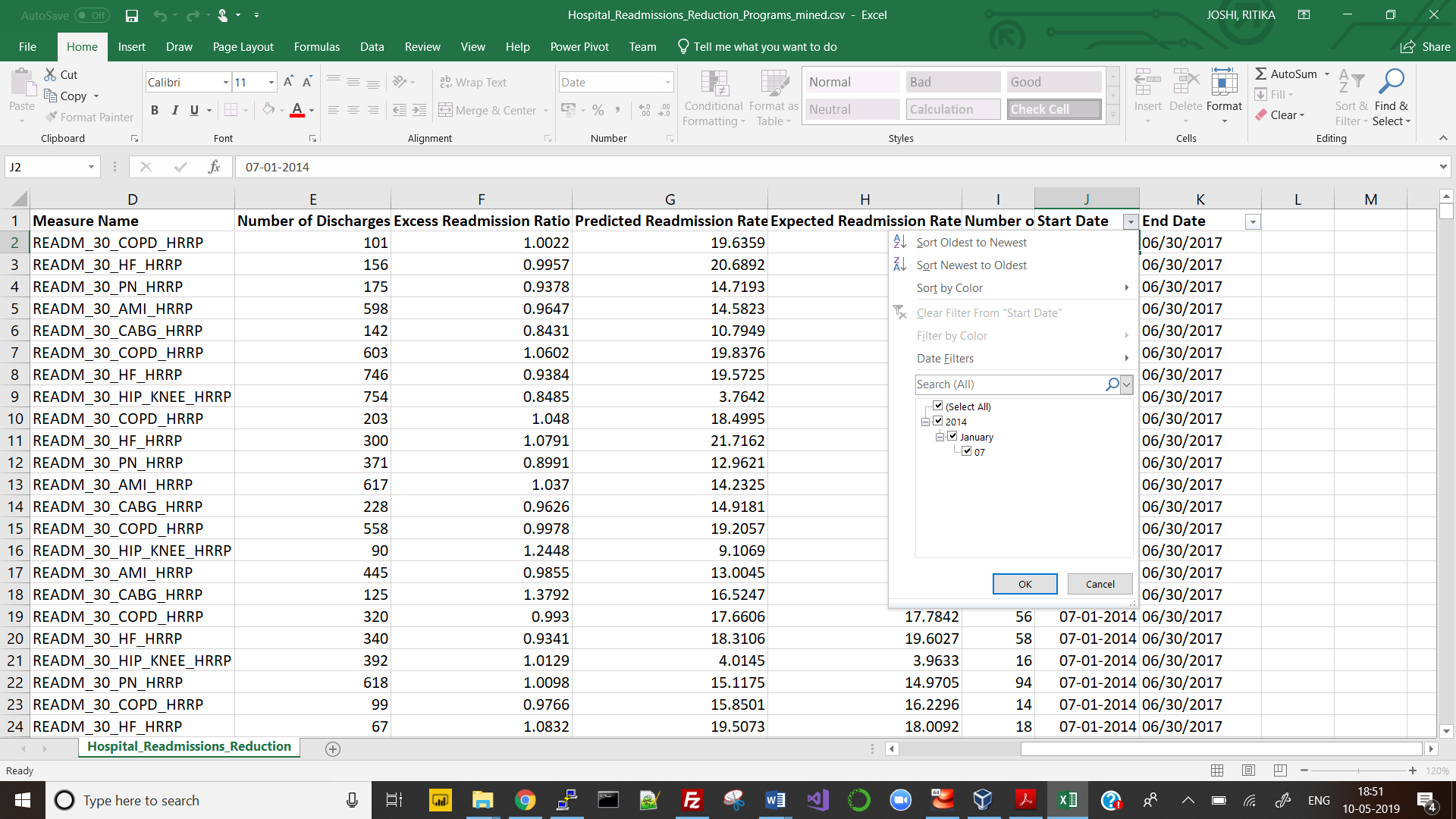


**3) Duplicate Values:** Checked for duplicate rows but none was found.

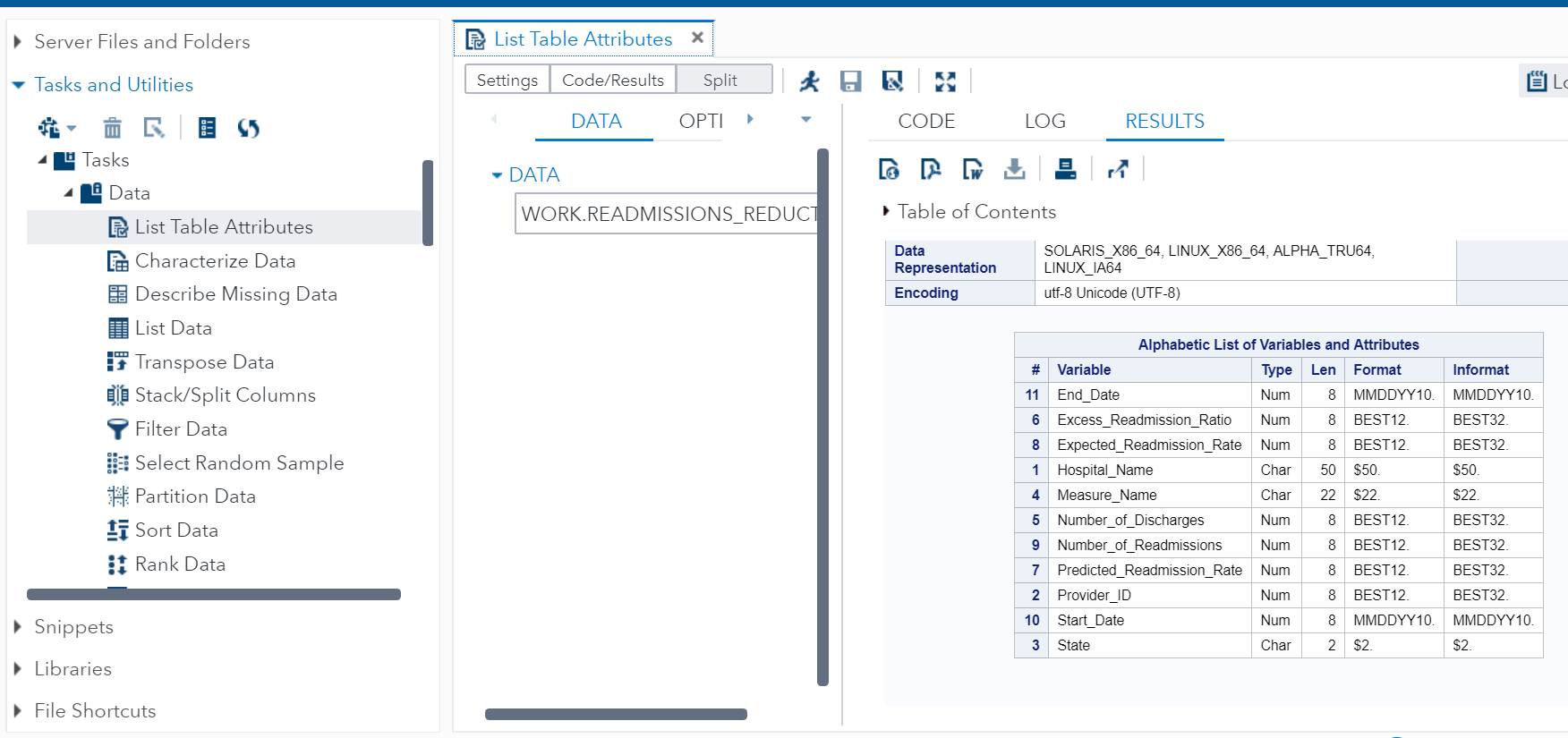




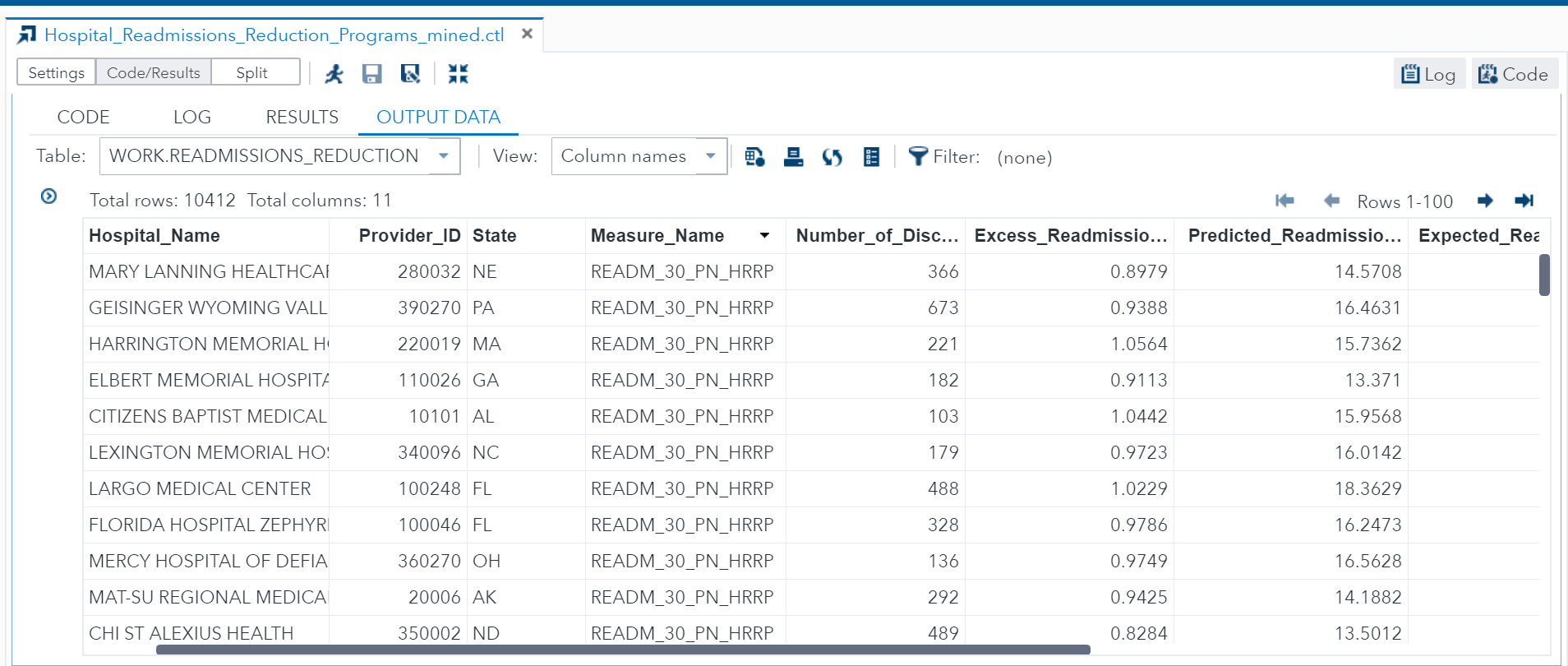
**4) Format of dates:** Checked for format of all columns using excel and SAS Visual. All were found to be correct. As shown in the screenshot below, format of StartDate and EndDate was checked using filter to look what all values are present. Format of the date was correct. For other types of columns, ‘List Table attributes’ feature of SAS was used.



In SAS studio, all of the data types are in correct format.



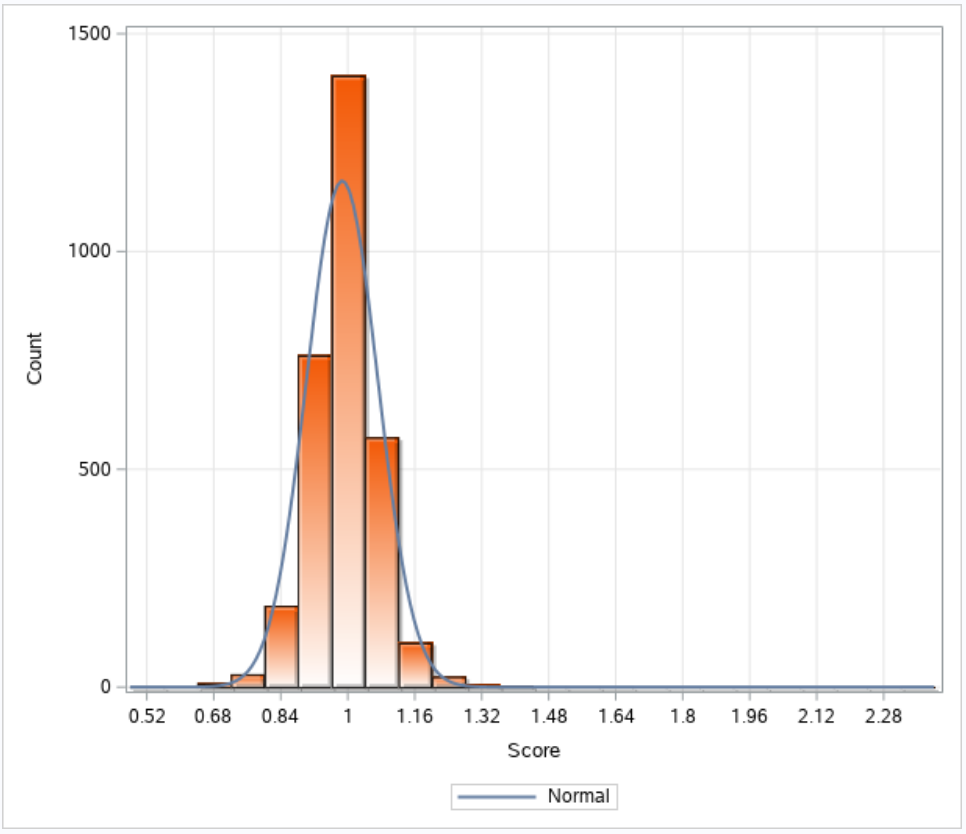
**Visual of Cleaned Data in SAS**



**[D] Analysis & Visualizations**

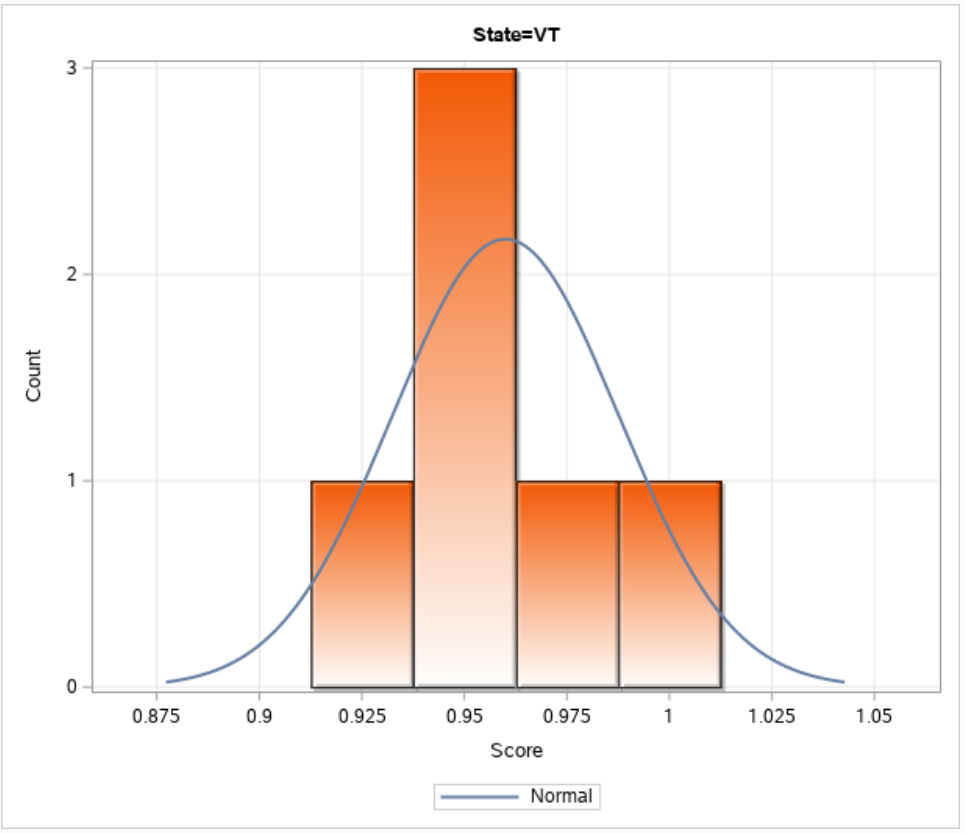
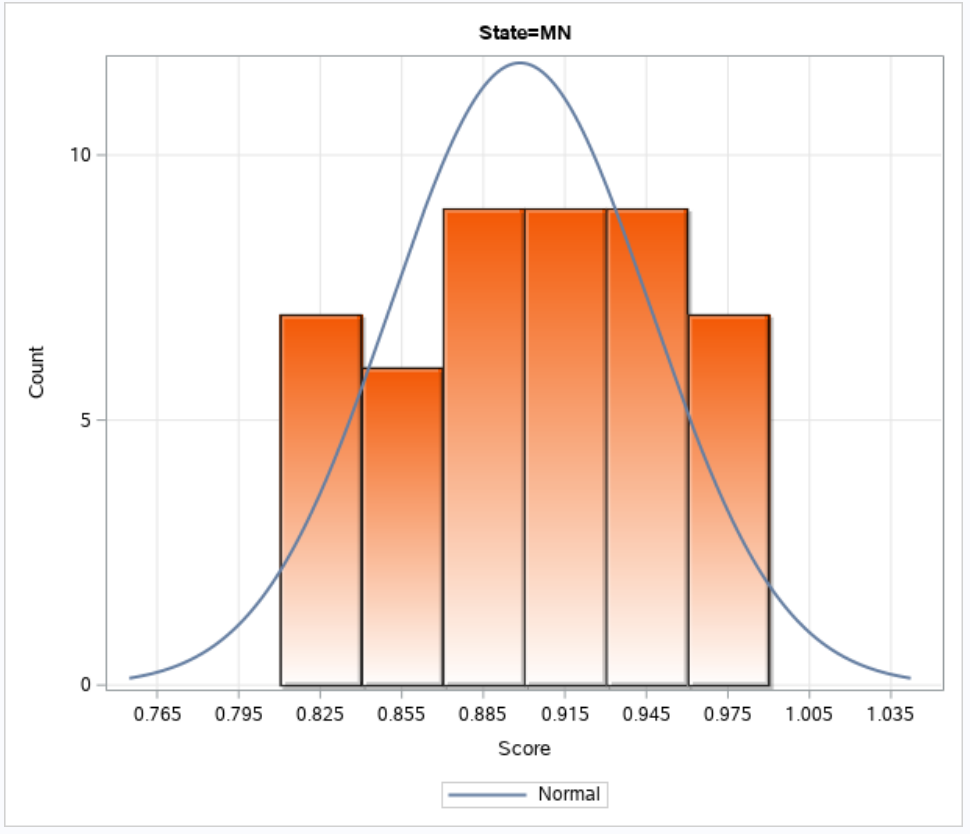
**1. What is the basis of need to reduce the number of admissions?**

Using the first data set on per patient spending by the hospital, score column represents whether the hospital has spent more, less or equal amount on patient. For example, if score is < 1, amount spent is lesser compared to overall hospitals nationally. If >1, then expense is on the higher side.



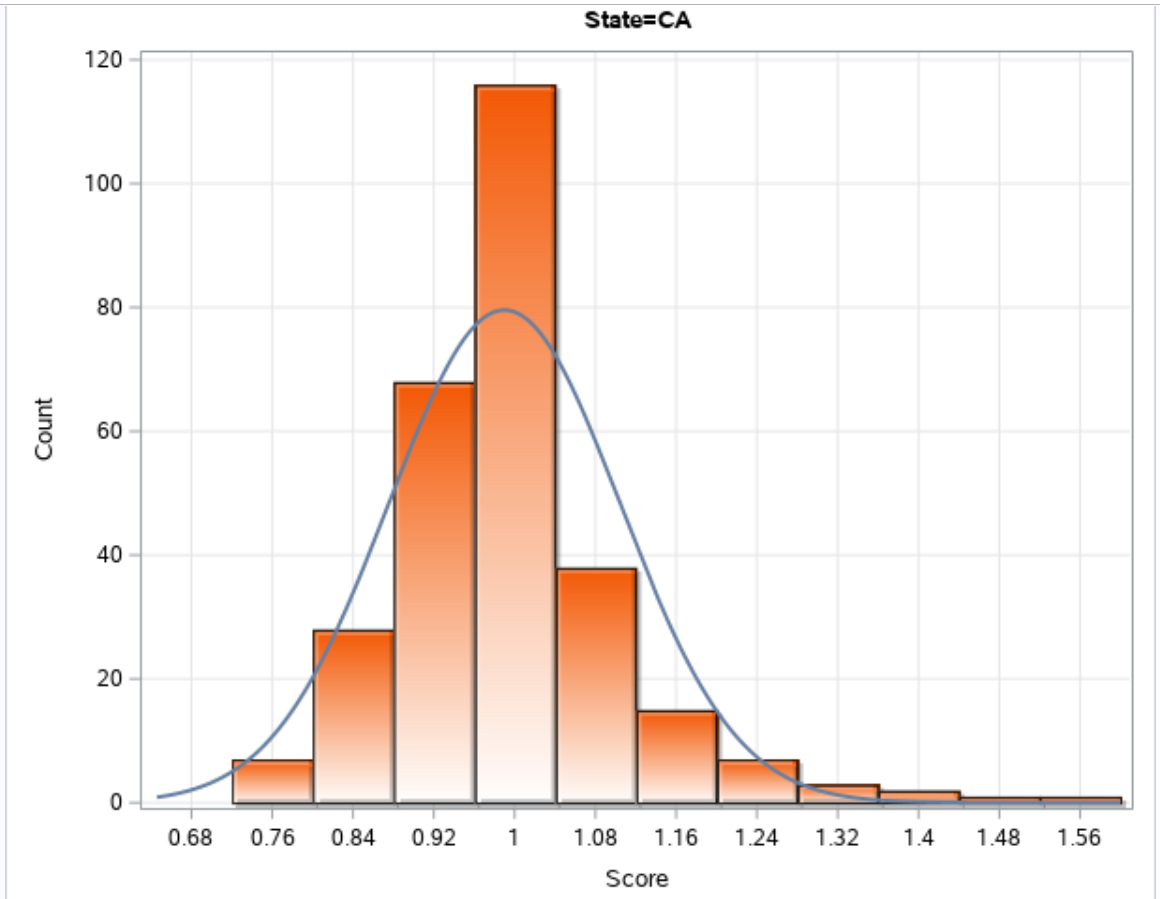
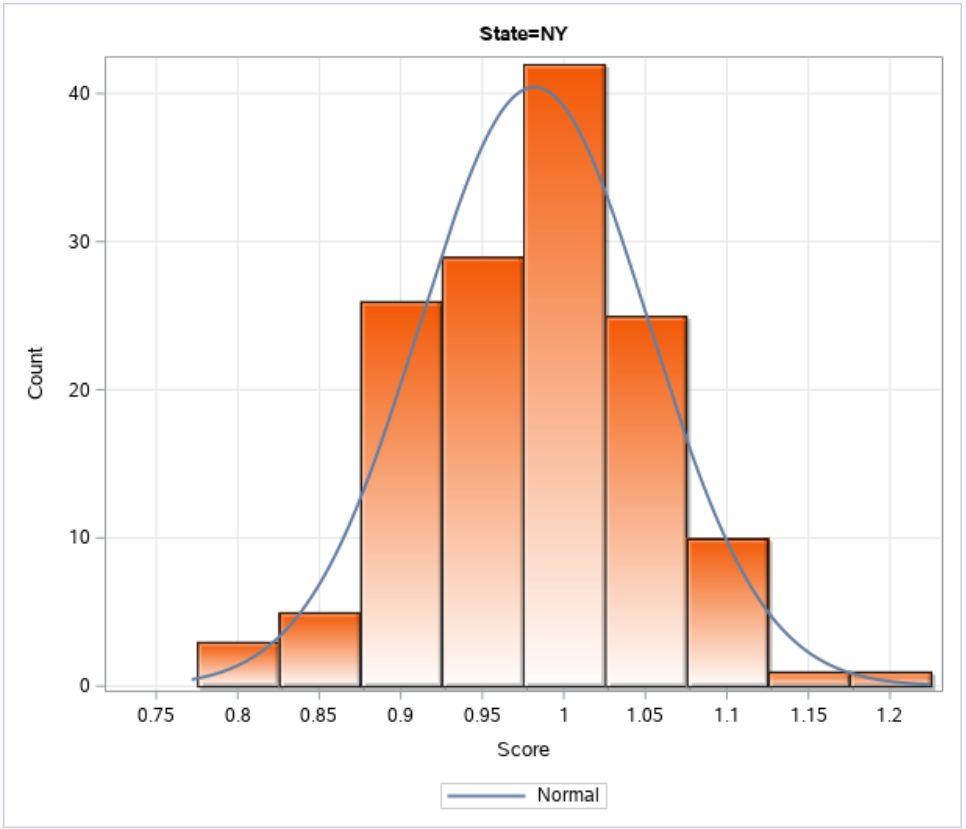
Visual 1.1 Hospital count of each score

**Interpretation:** Adding the approximate count of three bins (600 + 100 + 50) having score values greater than 1, there are 750 hospitals that spent heavily on medicare patients. Normal distribution reference line reaches the highest point of ~1200. To get more insights on whether any state or group of states spend highly on medicare patients, state wise visuals were created.



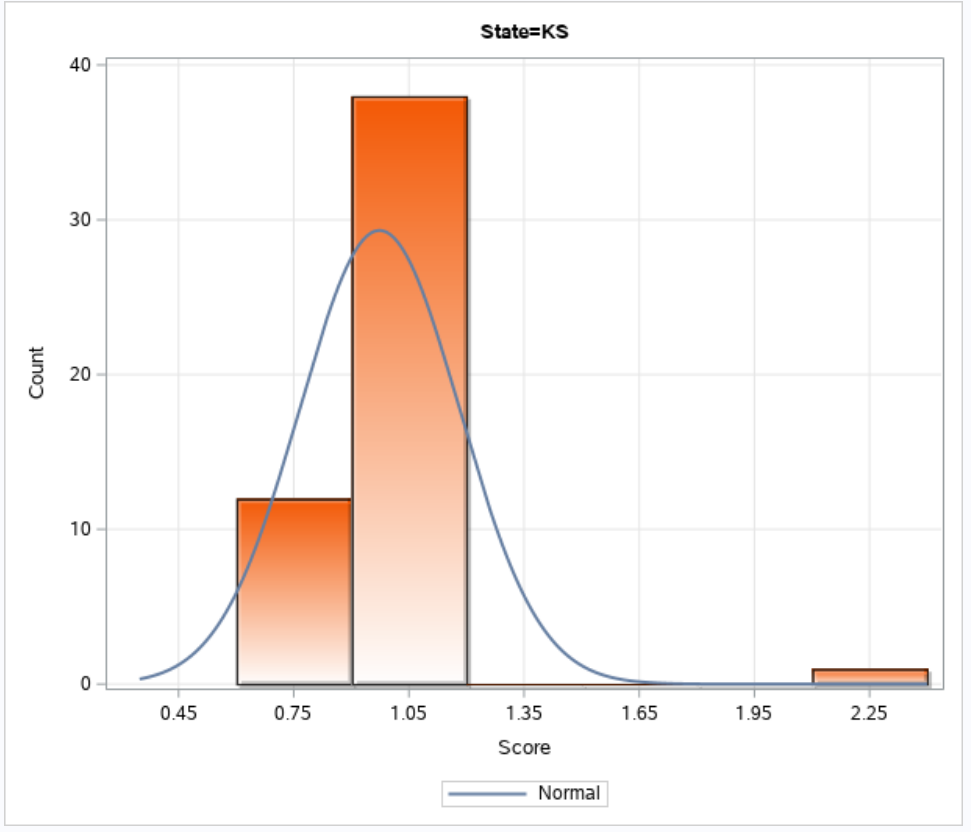
Visual 1.2 State-wise hospital count on least expense score

**Interpretation:** On drilling down these counts statewise, it was found that Minnesota (MN) and Vermont (VT) have no counts of over expense. Their overall expense is very low. None of the hospitals in MN have reached the score of 1 and only one hospital in VT has reached the expense score of 1. That means, their fraction of expense on national basis is below the standard amount of expense.

Visual 1.3 State-wise maximum hospital count on distributed score

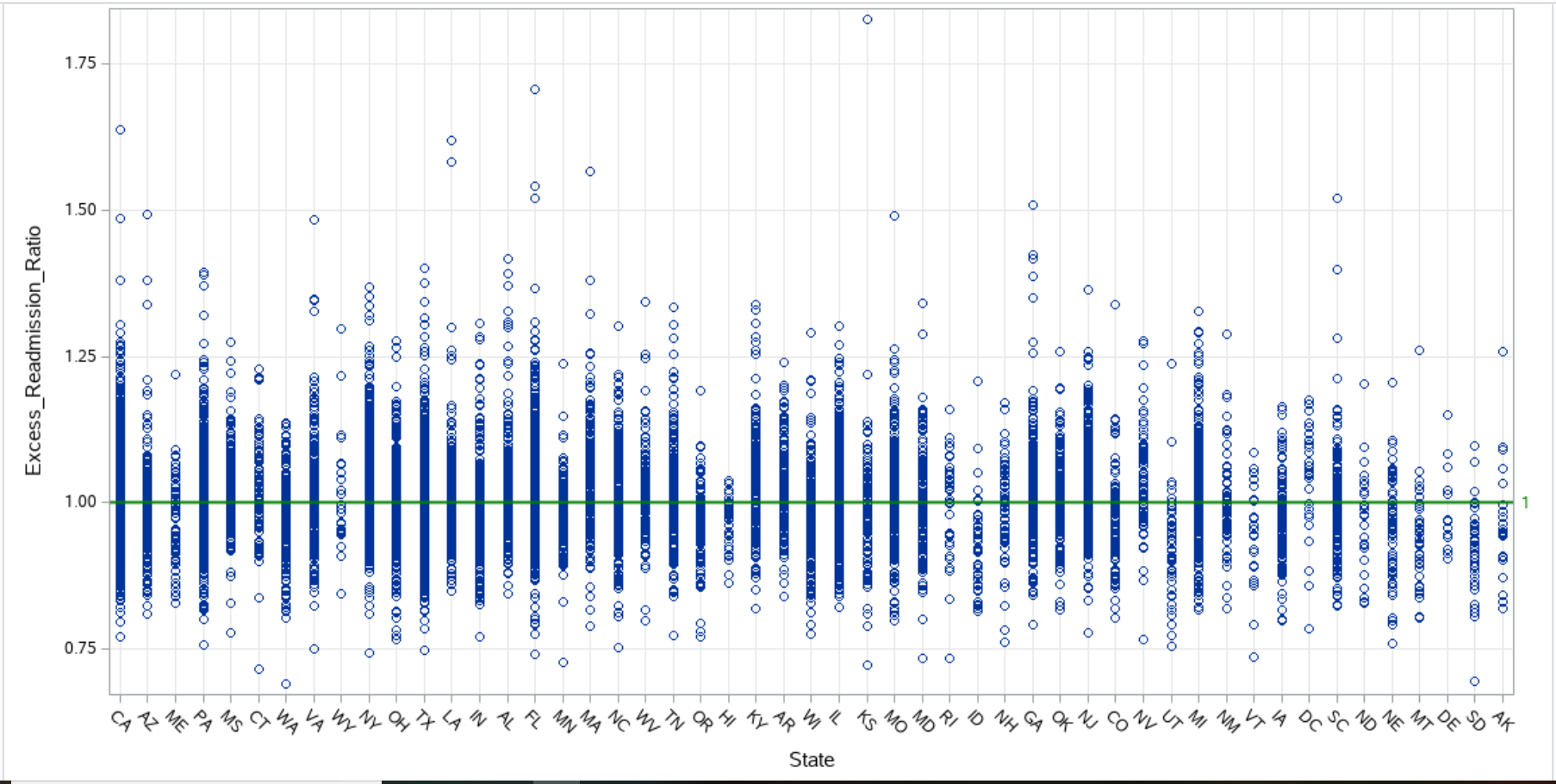
**Interpretation:** This was no surprise that California and New York which are two most of the dense populated states of US have shown the maximum counts. Which means 115 hospitals in California treats medicare patients with score value 1 and reached the total count (38 + 15 + 8 + 2 + 1) of 64 hospitals which over expends on medicare patients. Even for NY, 42 hospitals have shown over expense pattern.



Visual 1.4 State-wise hospital count on most expense score

**Interpretation:** Kansas state has shown an outstanding pattern of 2 hospitals with score of 2.25. Point of focus here is not the count of hospitals but the level of score. 2.25 score means 225% of national score. There definitely is a need to analyse how this over expense be minimized. Using other data set, we can analyse more on medicare expenses.

1. **What are the factors that influence hospital expense and can be minimized?**



Visual 2.1 Excess Readmissions per state

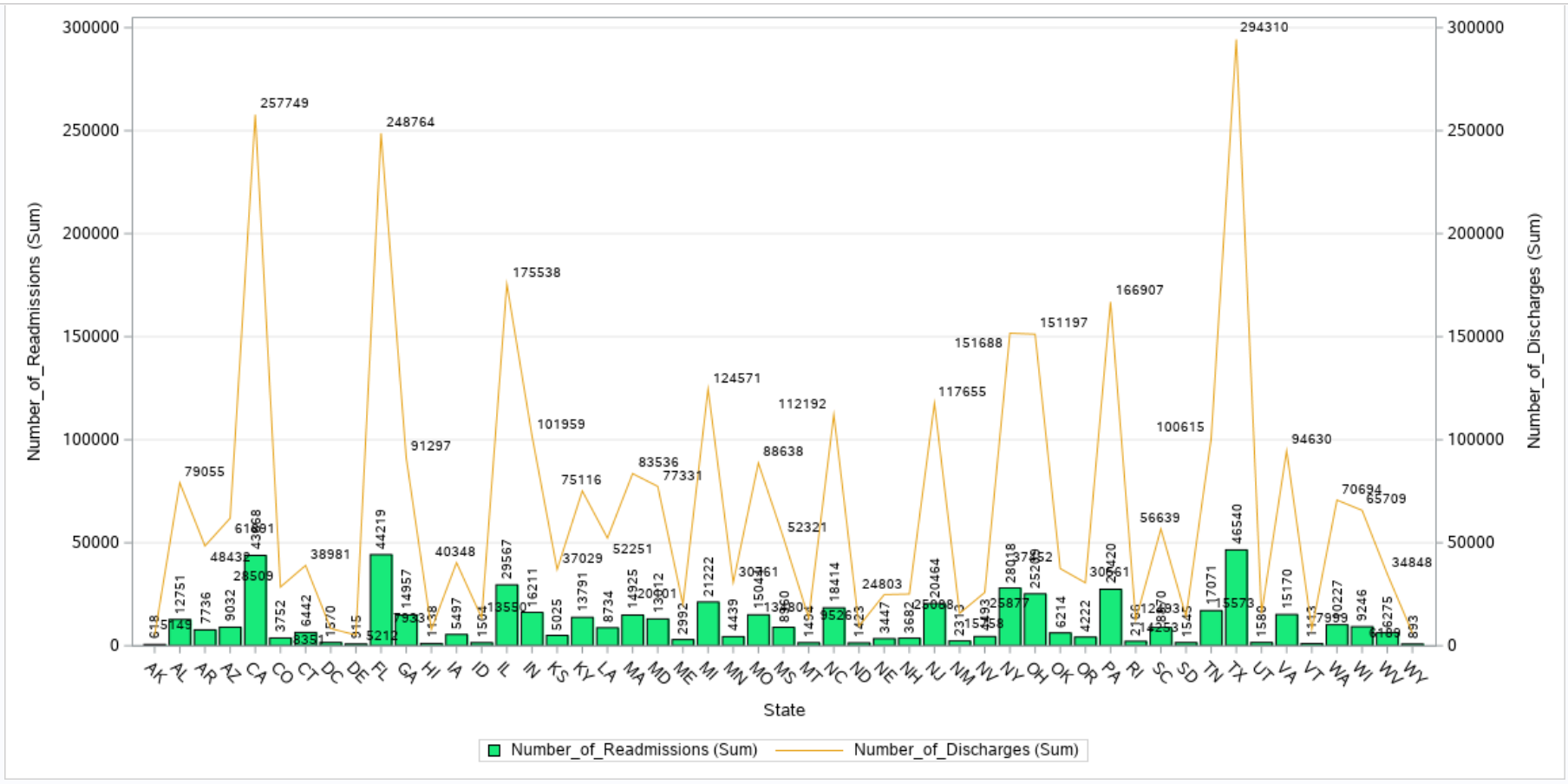
**Interpretation:** Excess readmissions are measured by a ratio, calculated by dividing a hospital’s number of “predicted” 30-day readmissions for heart attack (AMI), heart failure (HF), pneumonia, chronic obstructive pulmonary disease (COPD), hip/knee replacement (THA/TKA), and coronary artery bypass graft surgery (CABG) by the number that would be “expected,” based on an average hospital with similar patients.

Considering the reference line at ratio value 1, a few states shows the plot density dense above reference level. Kansas (KS) alone shows over excess ratio value above 1.75. Since, it may be an indicator of over and above expense score of this state in previous visual.

Four states CA, MA, LA, FL not only shows very dense ratio values between 1.00 – 1.50 but also between 1.50 – 1.75. That means these states have very high number of readmissions.

PA and NY are also falling into the category of states that faces high number of readmissions.

1. **Compare the number of readmissions with first discharge count.**



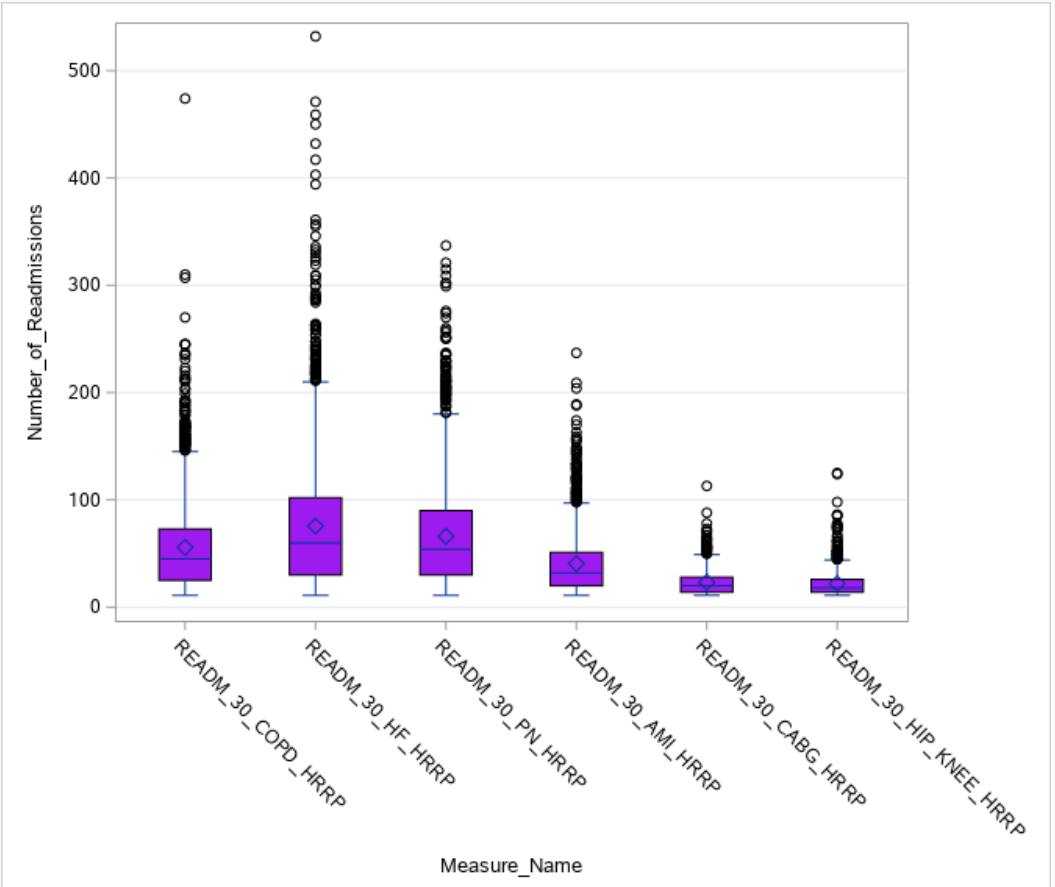
Visual 3.1 Comparison on number of discharges and readmissions

**Interpretation:** One level deeper insight on comparing whether level of discharge counts show similar pattern with count of readmissions. The higher the number of discharges, the higher level of readmission. Not a single state shows other way around. States to be noticed here are California, Florida and Texas as they have over grown peak values for the discharges. Their level for readmissions are also the highest.

1. **Which measures (disease control programs) cause high readmissions?**

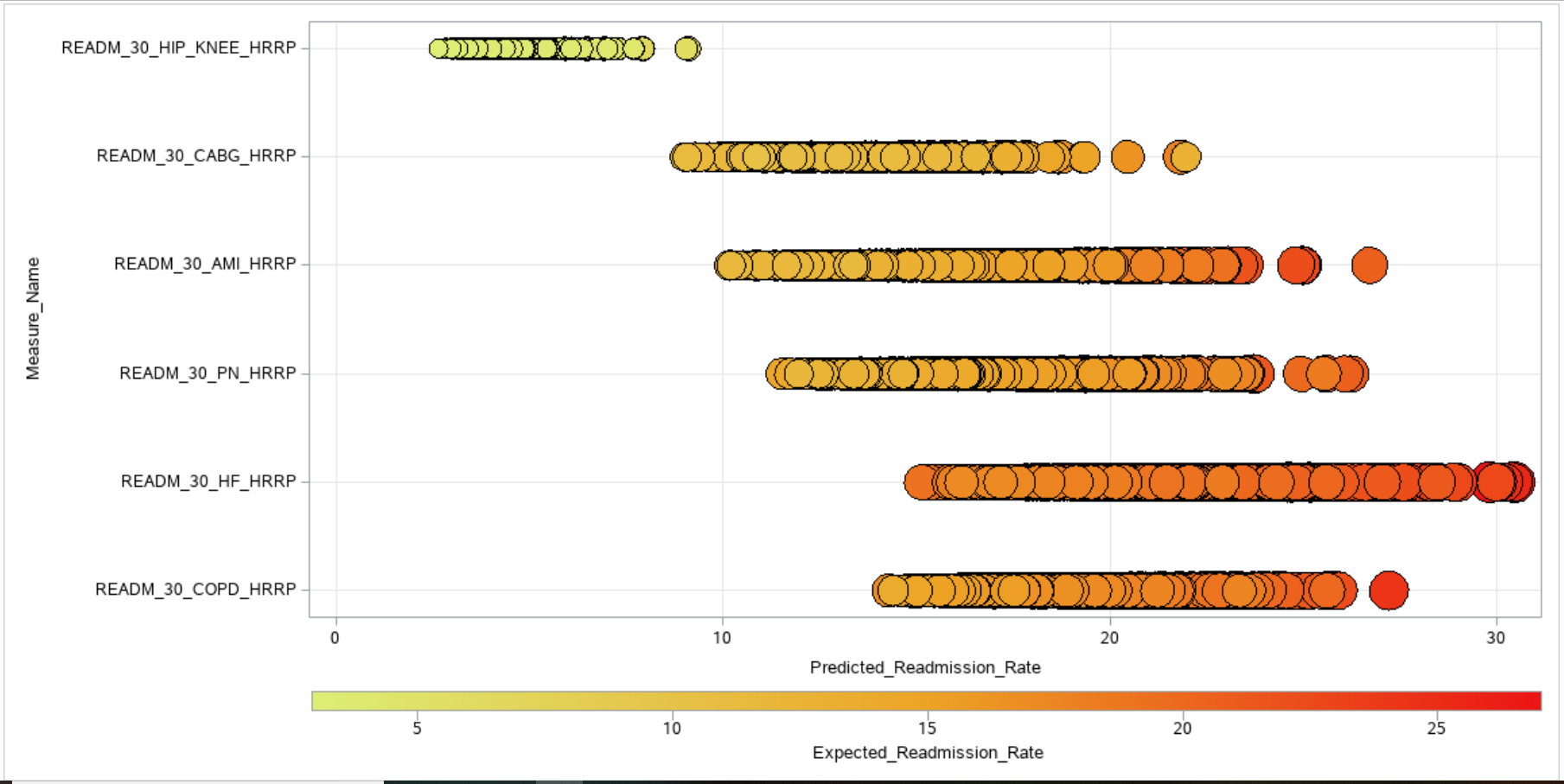
**Interpretation:** Number of readmissions is highest (~510) for Heart failure. Second highest (~ 480) is for chronic obstructive pulmonary disease (COPD). Pneumonia takes third place with readmission count ~350.

Least number of readmissions is shown by coronary artery bypass graft surgery (CABG)



Visual 4.1 Comparison on number of discharges and readmissions

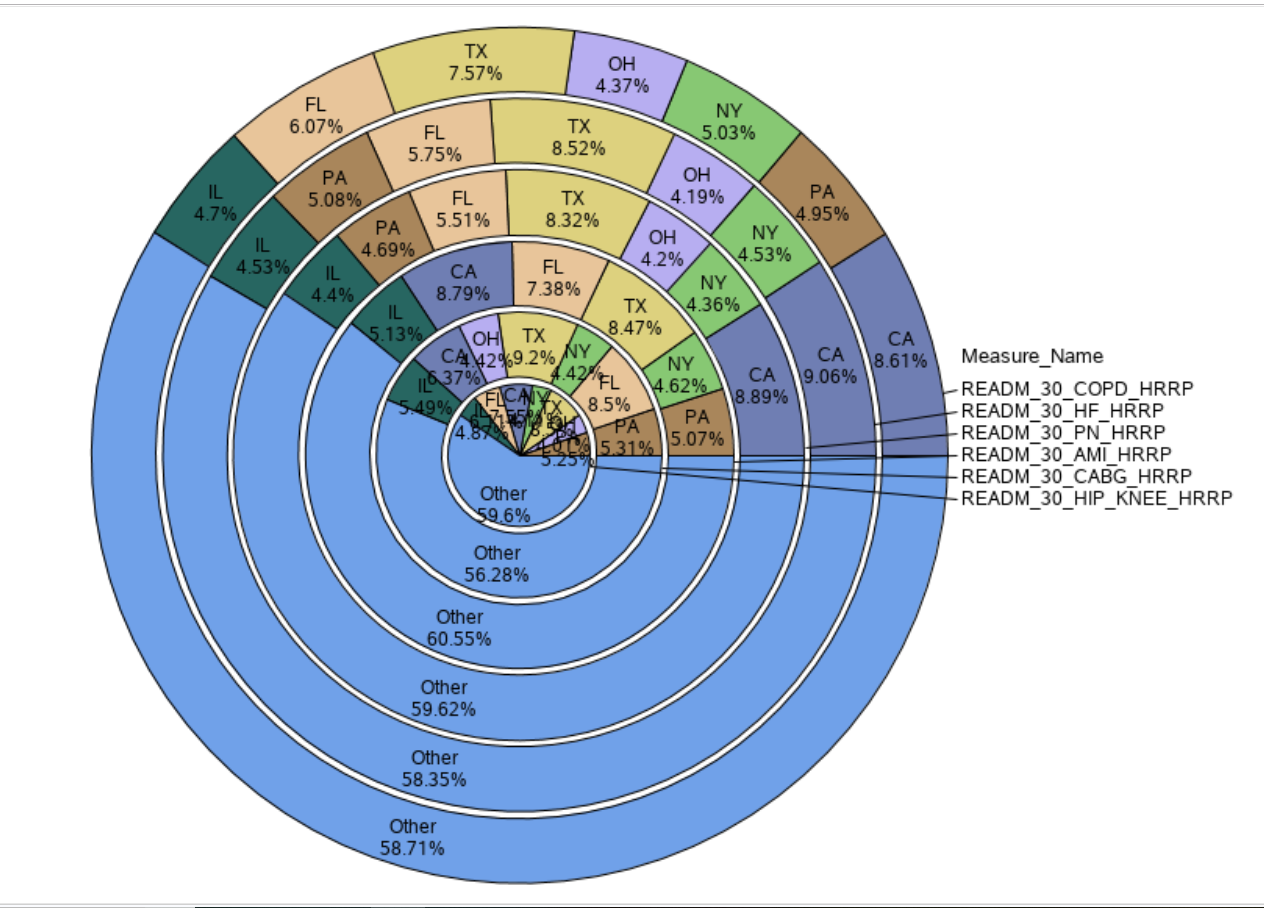
1. **What is the difference in predicted and actual rate of admissions for each measure type?**



Visual 5.1 Comparing predicted and actual rate of admissions for each measure type

**Interpretation:** Heart failure (HF) and pneumonia, chronic obstructive pulmonary disease (COPD) have shown the deepest coloured bubbles becausehigher the colour darkness, more the expected readmission rate**.** This visual also showspredicted admission rate is also highest for the measure type Heart Failure. Least rate of readmissions of both expected and predicted is shown by the measure type knee replacement.

1. **Which state needs to focus on which measure type in order to reduce readmissions and hence over expense.**



Visual 6.1 Occupancy percent of measure type per state

**Interpretation:** This visual in a way confirms the interpretations of earlier visuals, that, California is one of the most struggling state in terms of readmissions and expense. For heart failure (HF), pneumonia (PN), heart attack (AMI) and chronic obstructive pulmonary disease (COPD) California has shown maximum percent of measures.

Texas yet alone has shown 9.2% of occupancy for coronary artery bypass graft surgery (CABG).

**E. Statistical Summary**

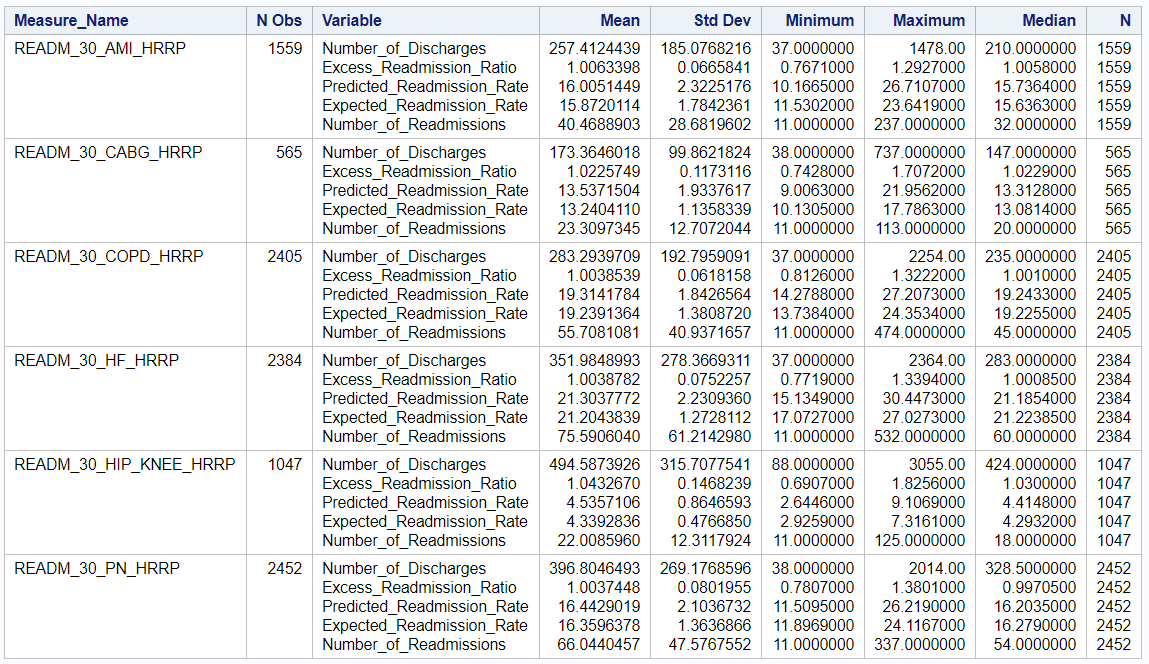


Looking at the mean and standard deviation for number of readmissions i.e. 55.2 and 47.39 respectively, means hospitals show very high variation in the count for each type of measure. Same is seen with the number of discharges, minimum count is 37 and maximum has gone up to 3055. It means, generalizing the statistics of one hospital with another would not be correct.

On the contrary, ratio of excess readmissions doesn’t seem to vary a lot. But one has to keep in mind that it is a ratio. Even a small variation of .5 will make a big difference. Standard deviation of .08 doesn’t look high.

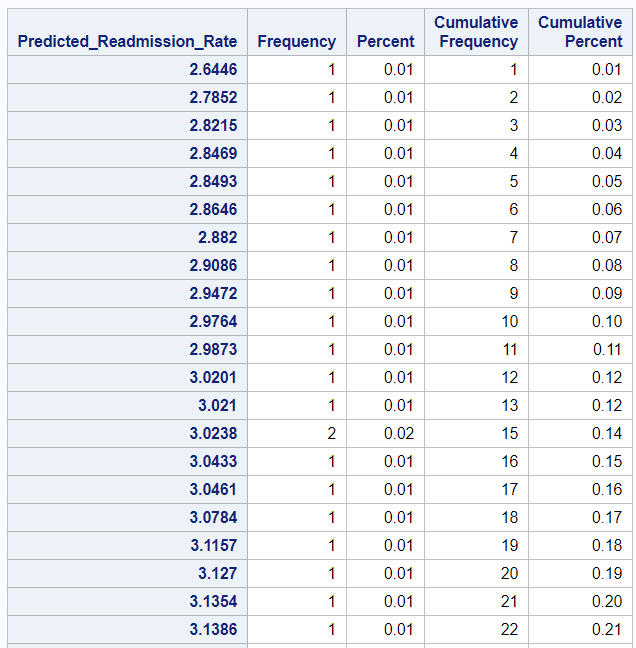
Drilling down the statistics per measure type shows, the highest number of records are present for heart failure. It is also the having the maximum mean (75) for number of readmissions and median of 60. So, heart failure could be the type of measure that needs improvement.

Second in line is pneumonia which has mean 66 for readmission count. Its median is also very high of 54 so this measure should also be taken in priority for the improvement.



**F. Statistical Tests**

1. **One-way frequency:** Performed the one-way frequency test on the predicted readmission rate to check how many hospitals have same predicted readmission rate. Surprisingly, readmission rate value is unique for every hospital per measure type. In the data set there are 10412 records in total and frequency of each rate is 1. Rate cannot be generalized on the basis of hospital name or state or county. It is different for each record in the data set. And because there were no duplicates found, rate is also unique.



.

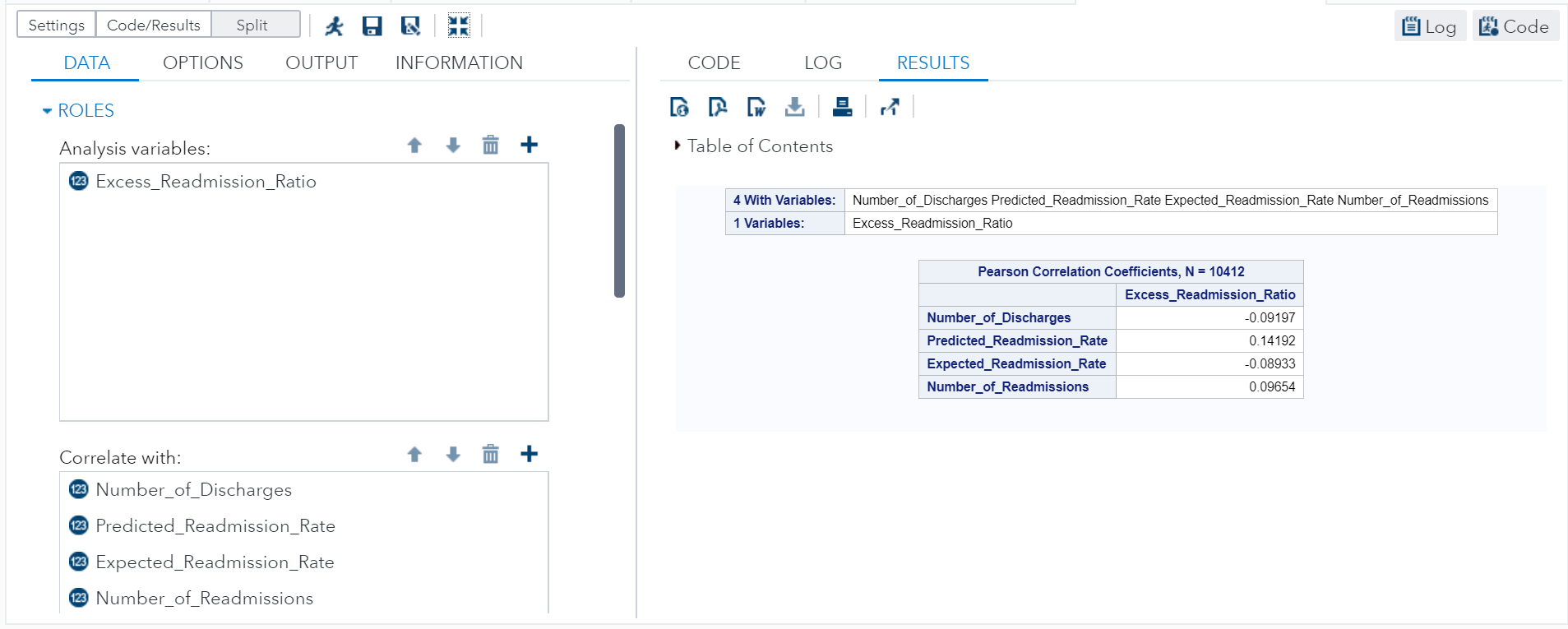
.

.in between records have been skipped

.



1. **Correlation analysis:** To know what features in the dataset have strong relation with excess readmission ratio, correlation analysis was performed. Strongest correlation is shown by predicted admission rate.



1. **t-test:** In the t-test below consider Kolmogorov-Smirnov test type.

P value = .01

consider a (alpha) = 0.05

i.e. p < a

so, reject null hypothesis.

Hence, predicted readmission rate has significance difference.

