EFFECTIVENESS OF HOME HEALTHCARE AGENCIES REGISTERED WITH MEDICARE USING IBM WATSON ANALYTICS

NAME: Anusha Manjappa

CIN: 305851746

CIS5810 Project1

Fall 2017

Anusha Manjappa

A) Dataset URL's:

Dataset:https://data.medicare.gov/Home-Health-Compare/Home-Health-CareAgencies/6jpm-sxkc/data

The dataset contains is drawn from the Medicare.gov's compare websites and directories. This site provides direct access to the official data from the Centers for Medicare & Medicaid Services (CMS) that are used on the Medicare.gov Compare Websites and Directories. This dataset contains a list of all Home Health Agencies that have been registered with Medicare. The list includes addresses, phone numbers, and quality measure ratings for each agency. The dataset also contains the list of the different services offered by each of the agencies along with how effective each of the services were individually for all the agencies. It consists of a total of 54 columns and 11,802 rows which provides the most detailed information

Hashtags dataset

- #homehealth
- agencies

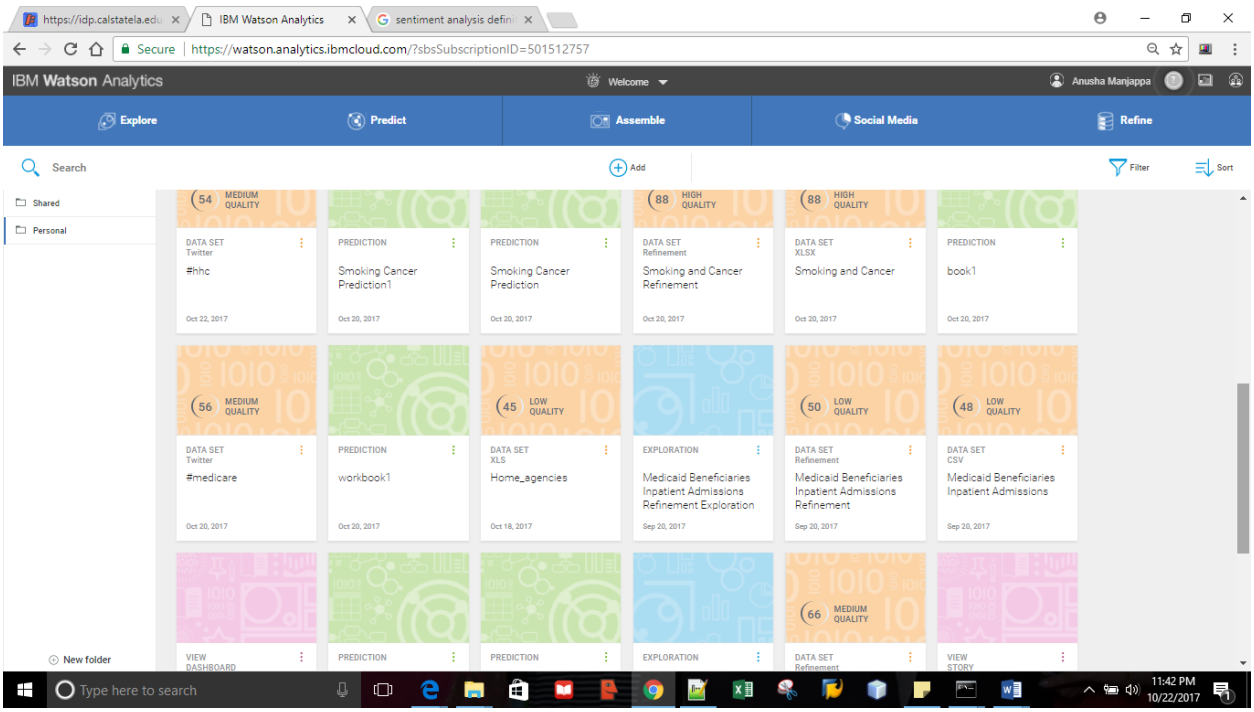
I have used two hashtags from the most recent 5 years of twitter data. I used the hashtags #homehealth #agencies. Homehealth tells us about the home healthcare agencies that are being talked of the most. It also refers to the home health plans are being talked about the most. Agencies hashtag would help us decide the agencies that are considered the most for the home health

insurance plans to be bought. These two tags would help people decide the popularity , good and bad of the agencies and healthcare plans.

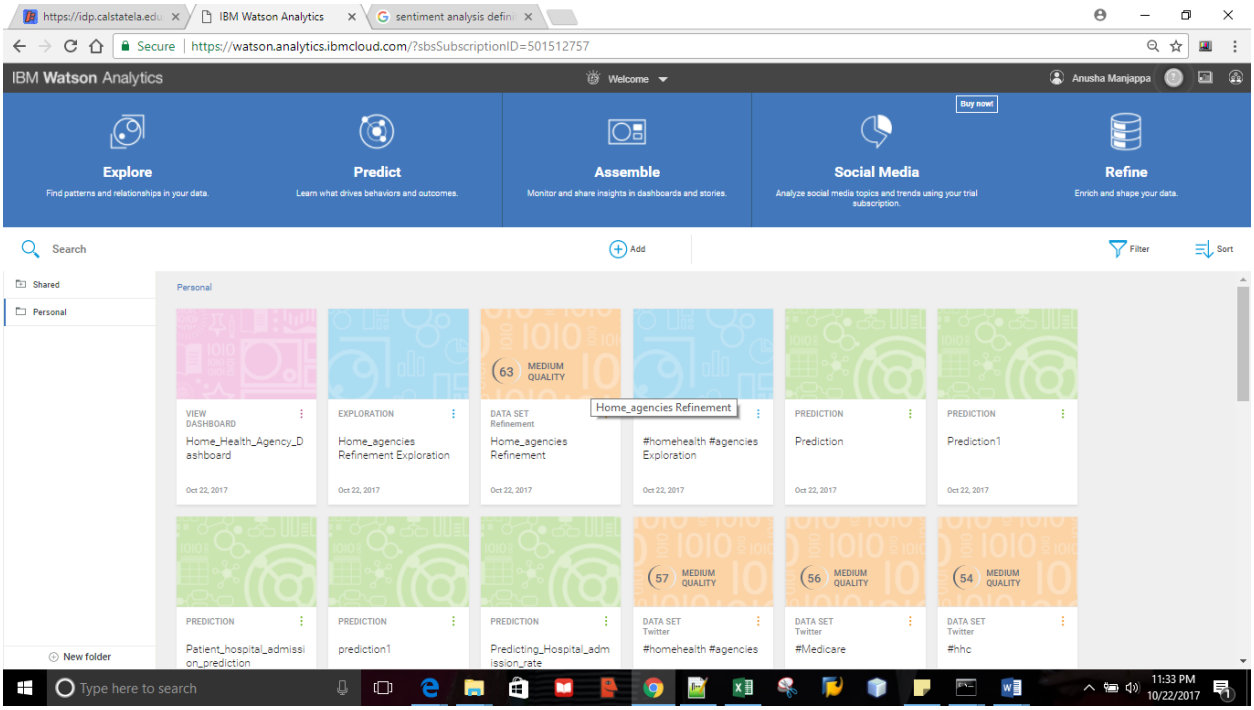
B) Data Quality

The data initially was very poor. It was found to be 45% . After cleaning up of the data as shows in the following sections , it is now a percentage of 62%.

- Before



- After



c) Data Cleaning

- Refinement in Excel

a) Missing Values

Some blocks in the dataset were empty as the value could not be determined or was not known. Therefore, I have replaced such values with the key word “Not Available”. This would make future calculations, explorations and predictions much easier and accurate.

- Before

Same As Expected	Worse Than Expecte
Worse Than Expected	Same As Expected
Same As Expected	Same As Expected
	Not Available
Same As Expected	Same As Expected
Same As Expected	Same As Expected
Same As Expected	Same As Expected

- After

Same As Expected	Worse Than Expecte
Worse Than Expected	Same As Expected
Same As Expected	Same As Expected
Not Available	Not Available
Same As Expected	Same As Expected
Same As Expected	Same As Expected
Same As Expected	Same As Expected

b) Bulk Missing Values

Some rows had many blocks of data in the dataset blank as the value could not be determined or was not known. As there were many missing values , no predictions could be made for that entity using the values given. Therefore, I have deleted such rows of data. This would make future calculations, explorations and predictions much easier and accurate.

- Before

552	TRUE	TRUE	TRUE	TRUE	########	3.5	95.2	97.6	100	100	96.3	92.7		52.1	68	66	72.5	68.7		66.5
553	TRUE	TRUE	TRUE	TRUE	9/4/1984	2.5	91.5	98.5	99.7	98.8	71	82.3	98.9	65	57.3	66.6	58.4	65.2	95.3	53.5
554	TRUE	TRUE	TRUE	TRUE	########	2.5	87	97.2	99.3	97.6	64.9	68	95.3	63.1	63.6	70.4	62	70	74.9	47.6
555	TRUE	TRUE	TRUE	TRUE	########	3.5	91.7	99.3	99.2	100	85.4	89.8	96.9	77.7	71.3	77.7	75	66	94.9	62.8
556	FALSE	TRUE	TRUE	TRUE	########															
557	TRUE	TRUE	TRUE	TRUE	########	3	92.8	97.9	97.6	96.9	83.3	86.3	92	66.3	62.3	69.8	73	72.4	90.2	53.9
558	TRUE	TRUE	TRUE	TRUE	########	3.5	97.7	99.4	99.8	100	78.2	86.9	97.2	68	62.1	73.2	95.7	76.5	100	50.7

- After

552	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	3.5	95.2	97.6	100	100	96.3	92.7		52.1	68	66	72.5	68.7
553	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	9/4/1984	2.5	91.5	98.5	99.7	98.8	71	82.3	98.9	65	57.3	66.6	58.4	65.2
554	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	2.5	87	97.2	99.3	97.6	64.9	68	95.3	63.1	63.6	70.4	62	70
555	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	3.5	91.7	99.3	99.2	100	85.4	89.8	96.9	77.7	71.3	77.7	75	66
556	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	3	92.8	97.9	97.6	96.9	83.3	86.3	92	66.3	62.3	69.8	73	72.4
557	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	3.5	97.7	99.4	99.8	100	78.2	86.9	97.2	68	62.1	73.2	95.7	76.5
558	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	########	4	91.3	98.9	97.9	97.1	79.6	86.1	96.7	72.3	70.2	78.8	77.3	78.6

c) Misspelt words

Some words in the dataset were spelt wrong. The correct spellings were known as they were some common misspelt words. I have corrected such spellings as shown below.

- Before

350 TERRA	REDLANDS	92373
270 W BAI	COVINA	91723
3250 WILS	LOS ANGE	90010
2095 W VI	VISTA	92083
800 WEST	ESCONDIDO	92025

- After

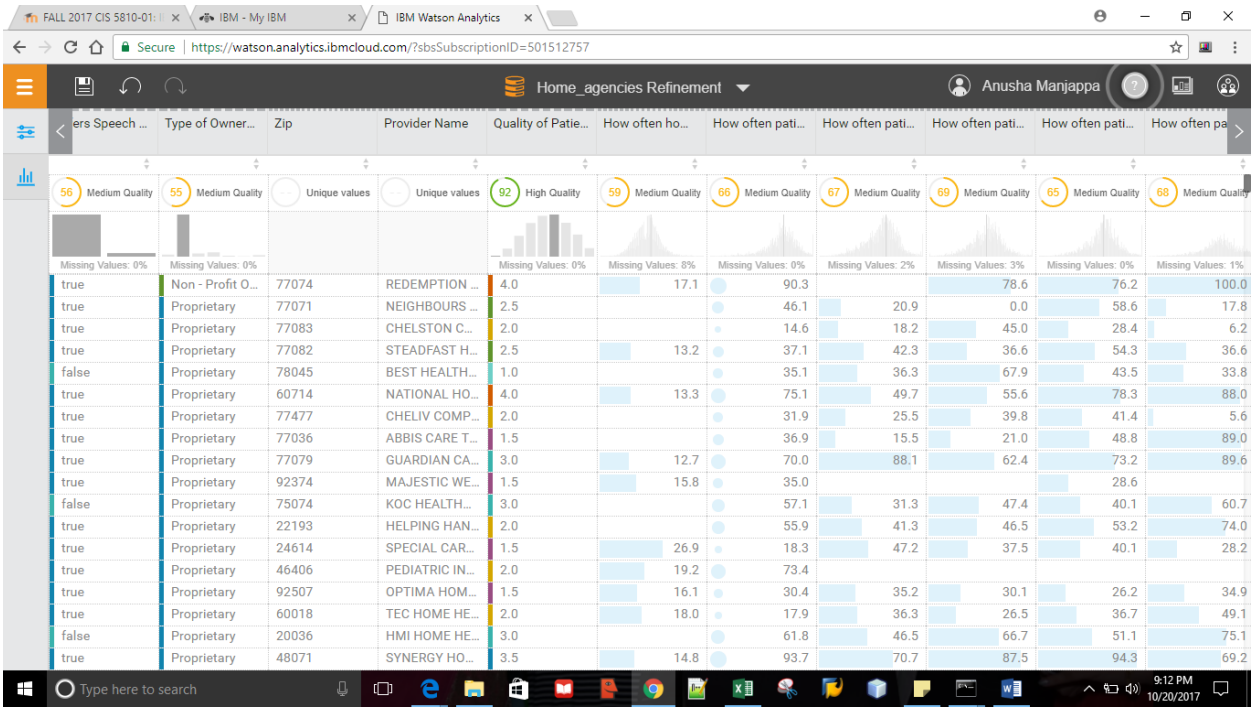
350 TERRA	REDLANDS	92373
270 W BAI	COVINA	91723
3250 WILS	LOS ANGELES	90010
2095 W VI	VISTA	92083
800 WEST	ESCONDIDO	92025

- Refinement in IBM Watson

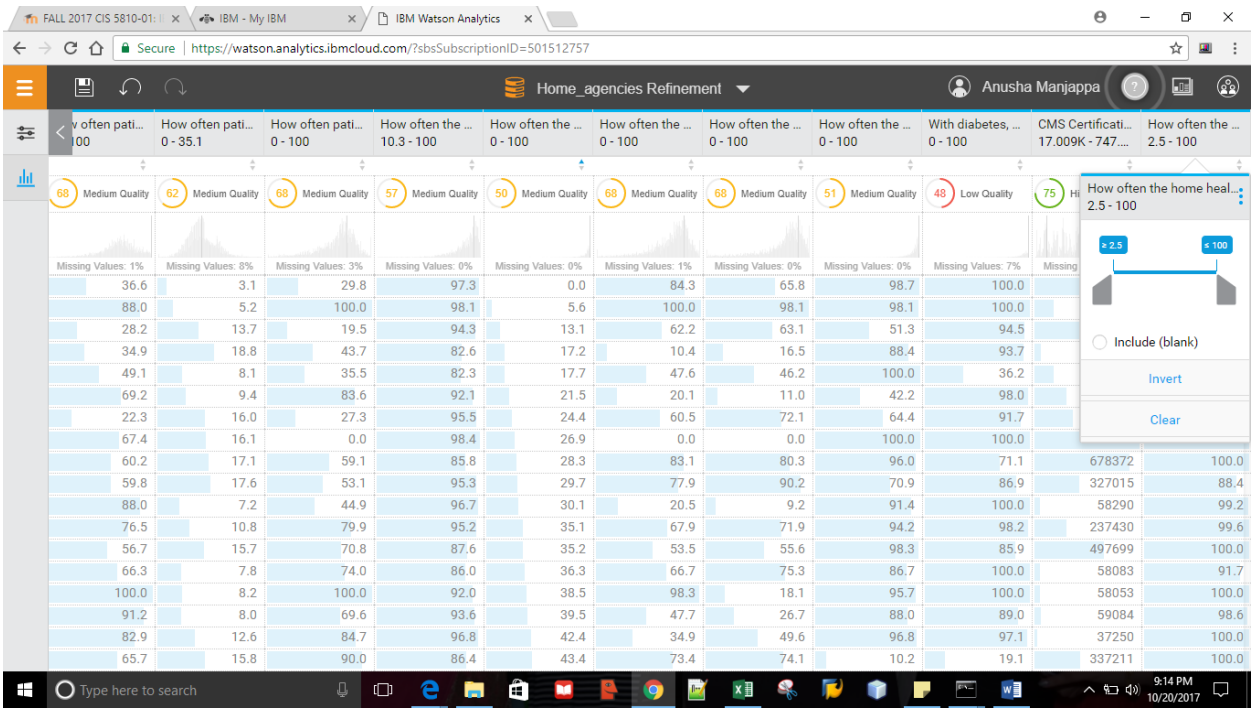
a) Missing Values

There were many missing values in my dataset which was a hindrance for my analysis. IBM Watson provides a method to remove all the blank values as shown below. To remove the blank values, we have to click on the column name and uncheck the option which says "Include (Blank)".

- Before



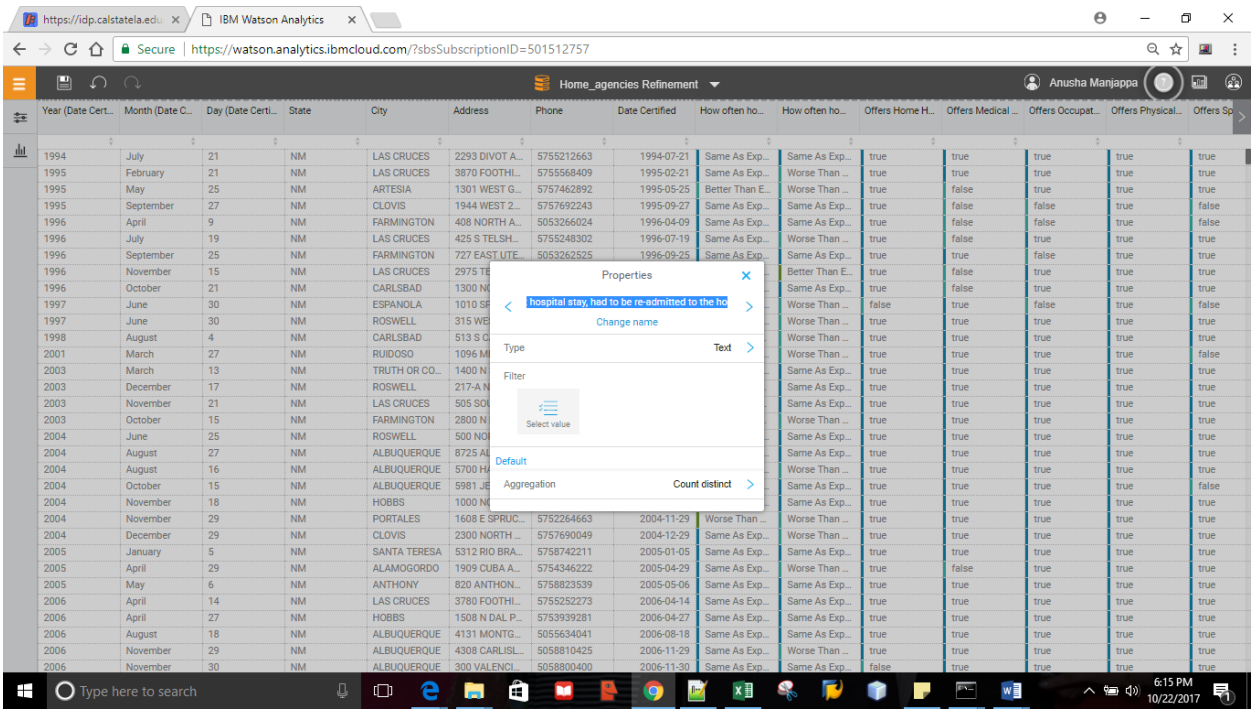
- After



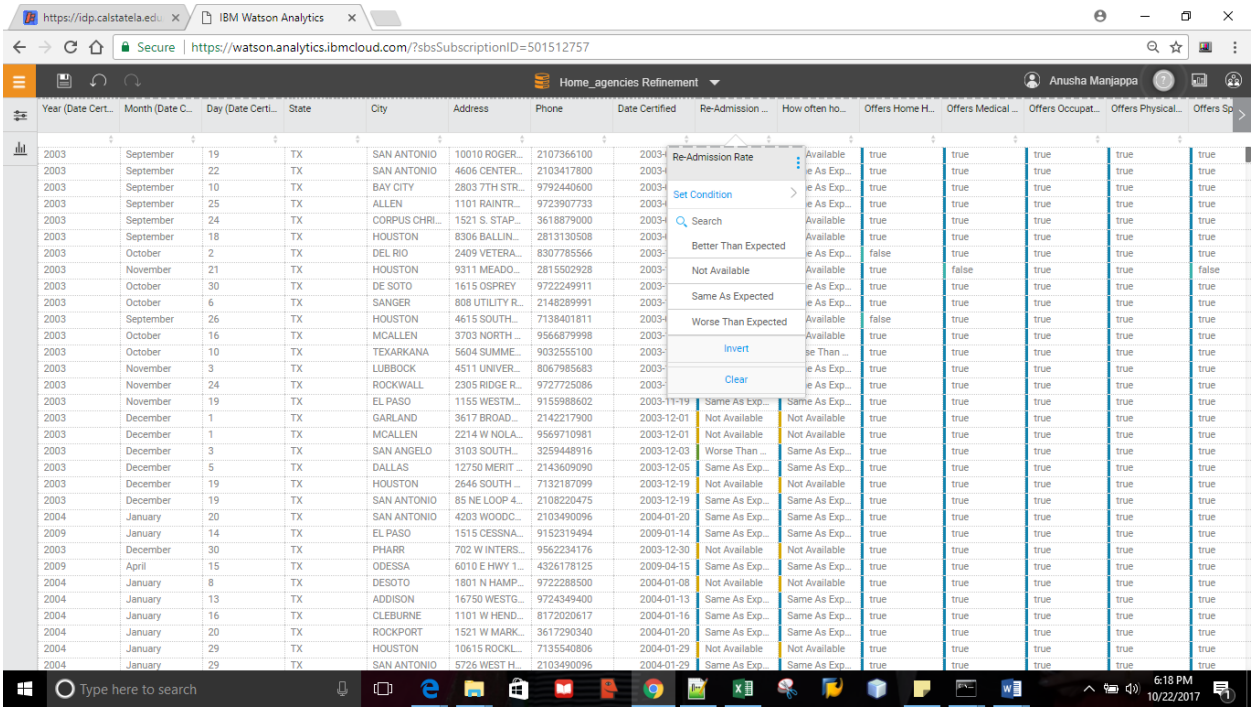
b) Inconvenient long column names

The column names in the dataset were in the form of long sentences which was very inconvenient to refer to. Therefore, I converted such long names into short meaningful names. This is for simple and convenient reference in the future. A screenshot of e few such changes are shown below.

- Before


IBM Watson Analytics
Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757
Home_agencies Refinement
Anusha Manjappa
Properties
hospital stay, had to be re-admitted to the ho
Change name
Type
Text
Filter
Select value
Default
Aggregation
Count distinct
Type here to search
6:15 PM
10/22/2017

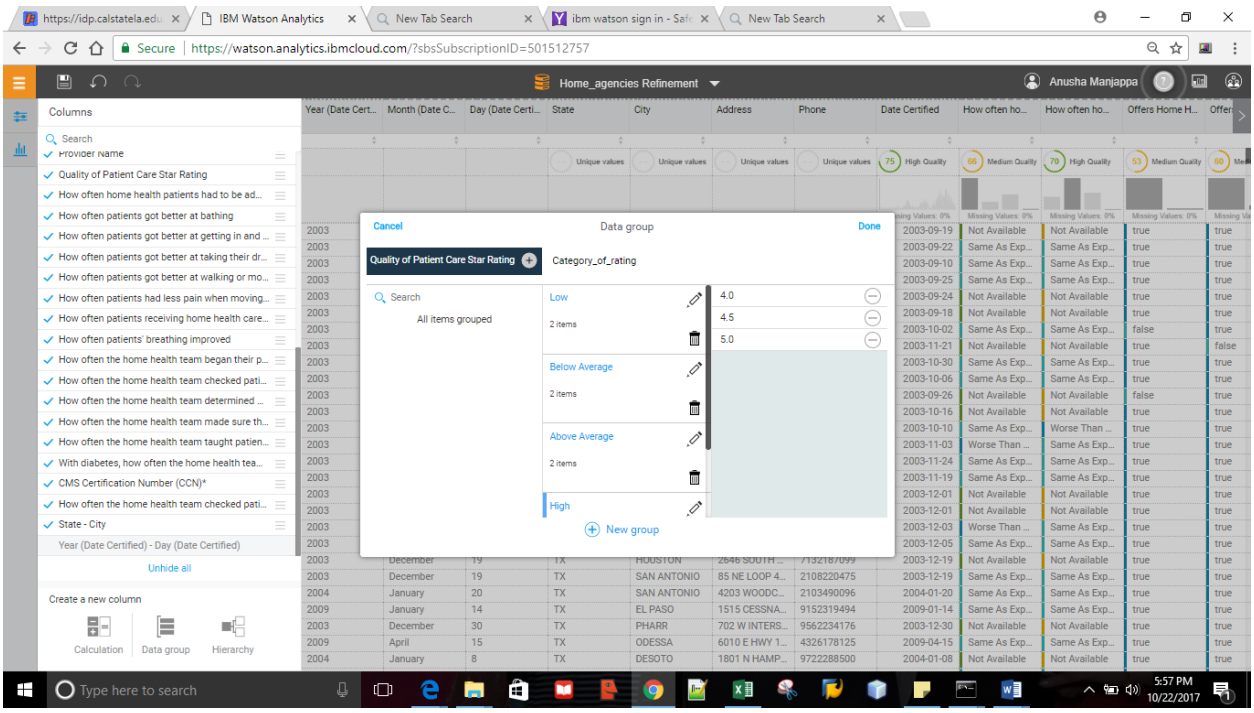
- After

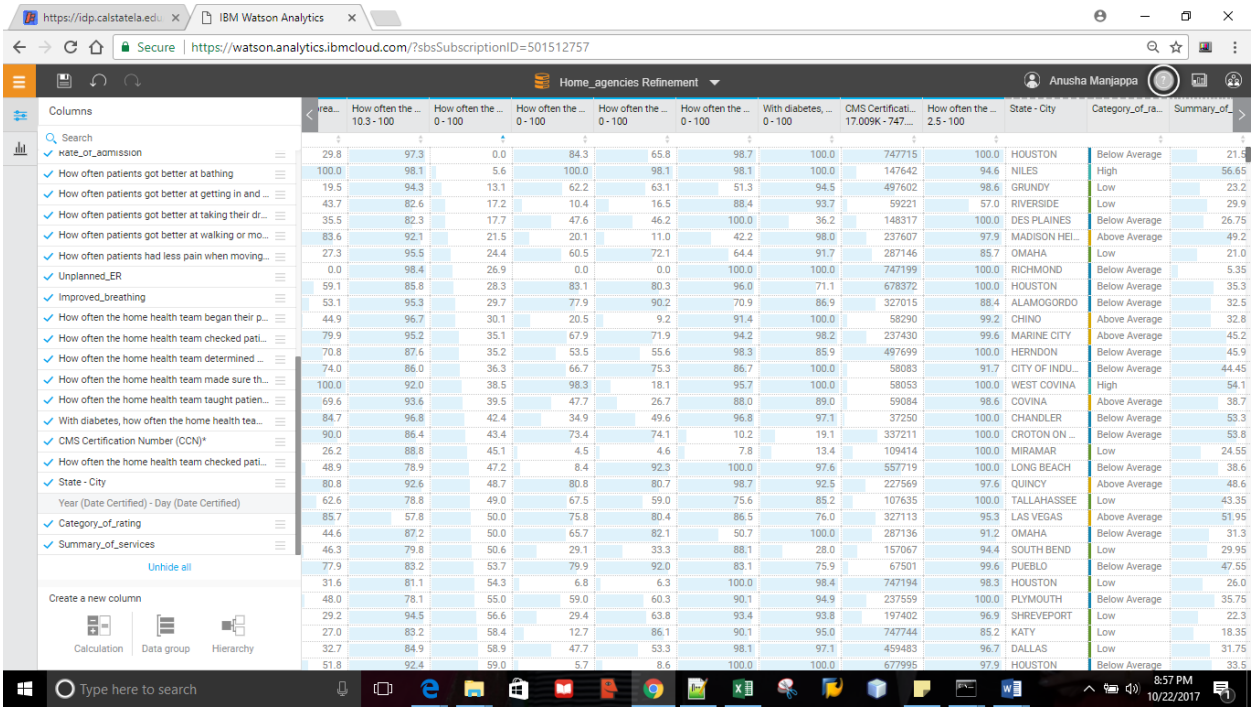

IBM Watson Analytics
Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757
Home_agencies Refinement
Anusha Manjappa
Re-Admission Rate
Set Condition
Search
Better Than Expected
Not Available
Same As Expected
Worse Than Expected
Invert
Clear
Type here to search
6:18 PM
10/22/2017

D) Data Hierarchy/ Data Grouping

One of the prominent fields that portray the effectiveness of the agency is the 'Quality of Patient Care Star Rating 'field. Home Health Compare uses a star rating between 1 and 5 to show people how a home health agency compares to other home health agencies on measurements of their performance. I have grouped these ratings into 4 groups as shown below for better understanding. They are: -

- Low (Rating 1 to Rating 2)
- Below Average (Rating 2 to Rating 3)
- Above Average (Rating 3 to Rating 4)
- High (Rating 4 to Rating 5)



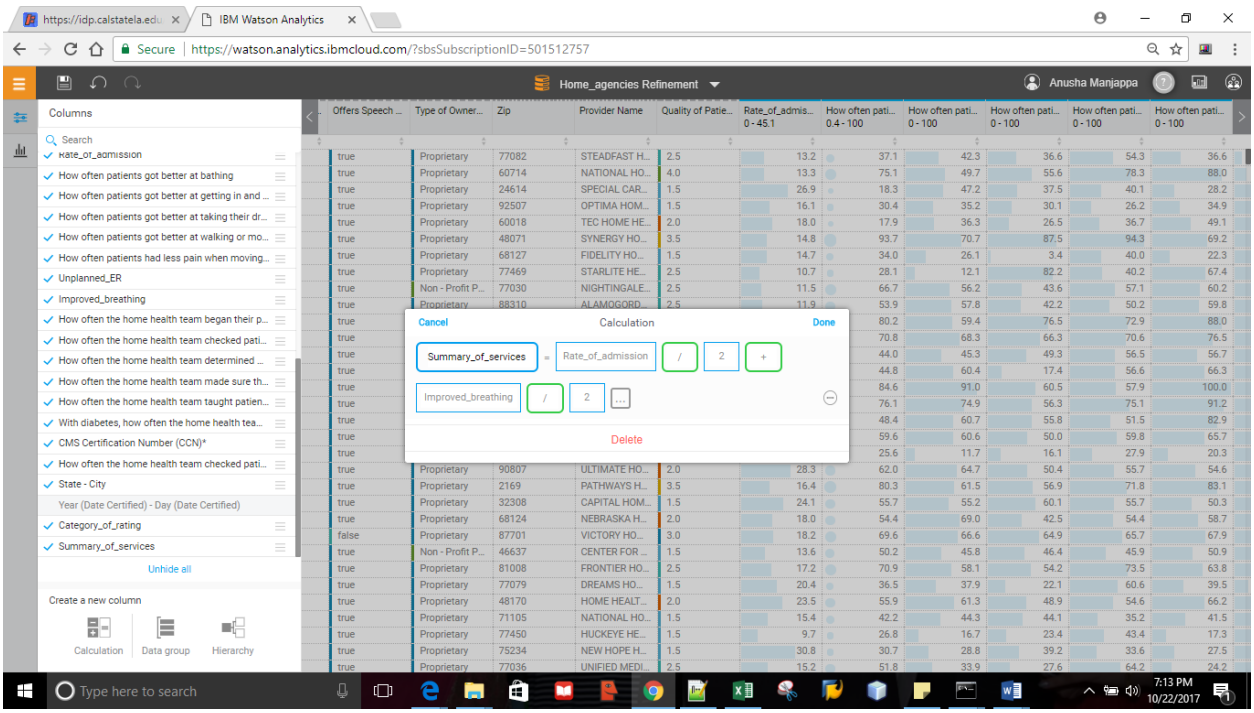


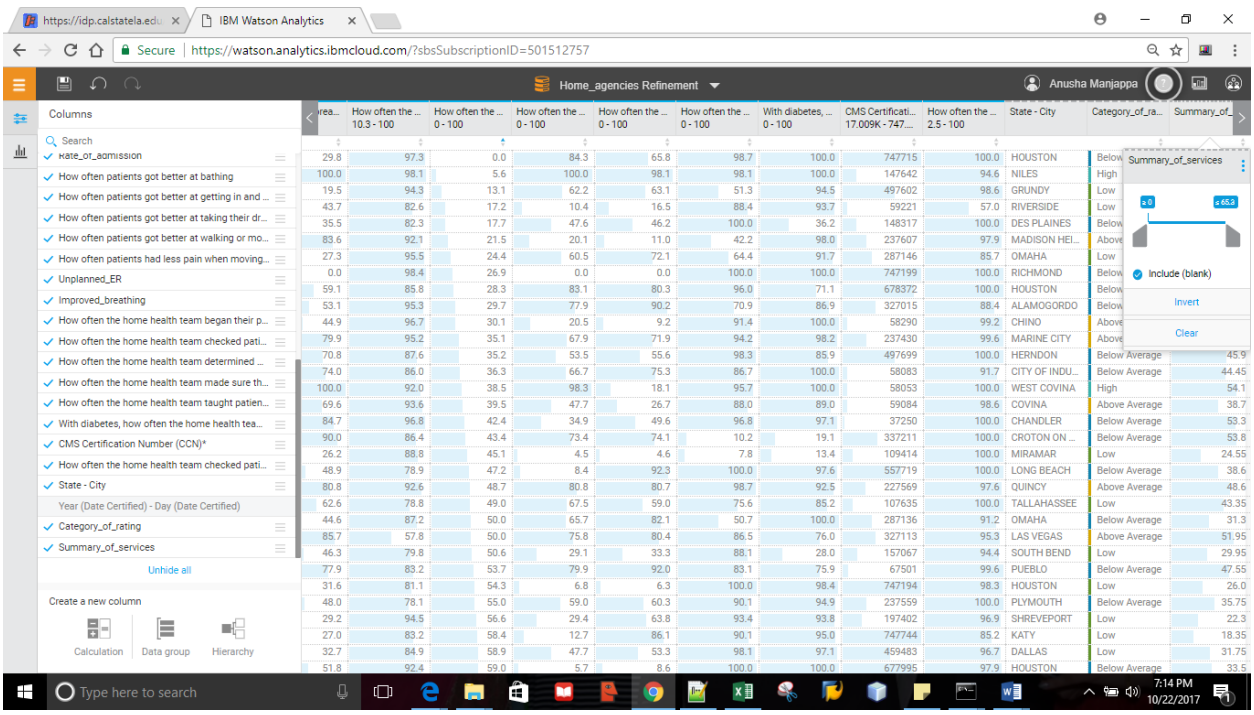
Data Calculation

I have used two fields, rate_of_admission and improved breathing to calculate the field, Summary_of_services.

Summary of services= (rate_of_admission + improved breathing) / 2

This field calculates and portrays the improvement or the quality of services offered by each agency in terms of percentage by combining the percentage of the two fields rate_of_admission and improved breathing.


IBM Watson Analytics
https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757
Home_agencies Refinement
Anusha Manjappa
Columns
Search
How often patients got better at bathing
How often patients got better at getting in and ...
How often patients got better at taking their dr...
How often patients got better at walking or mo...
How often patients had less pain when moving...
Unplanned_ER
Improved_breathing
How often the home health team began their p...
How often the home health team checked pati...
How often the home health team determined ...
How often the home health team made sure th...
How often the home health team taught patien...
With diabetes, how often the home health tea...
CMS Certification Number (CCN)*
How often the home health team checked pati...
State - City
Year (Date Certified) - Day (Date Certified)
Category_of_rating
Summary_of_services
Unhide all
Create a new column
Calculation
Data group
Hierarchy
Cancel
Calculation
Done
Summary_of_services
Rate_of_admission
Improved_breathing
Delete
Type here to search
7:13 PM
10/22/2017

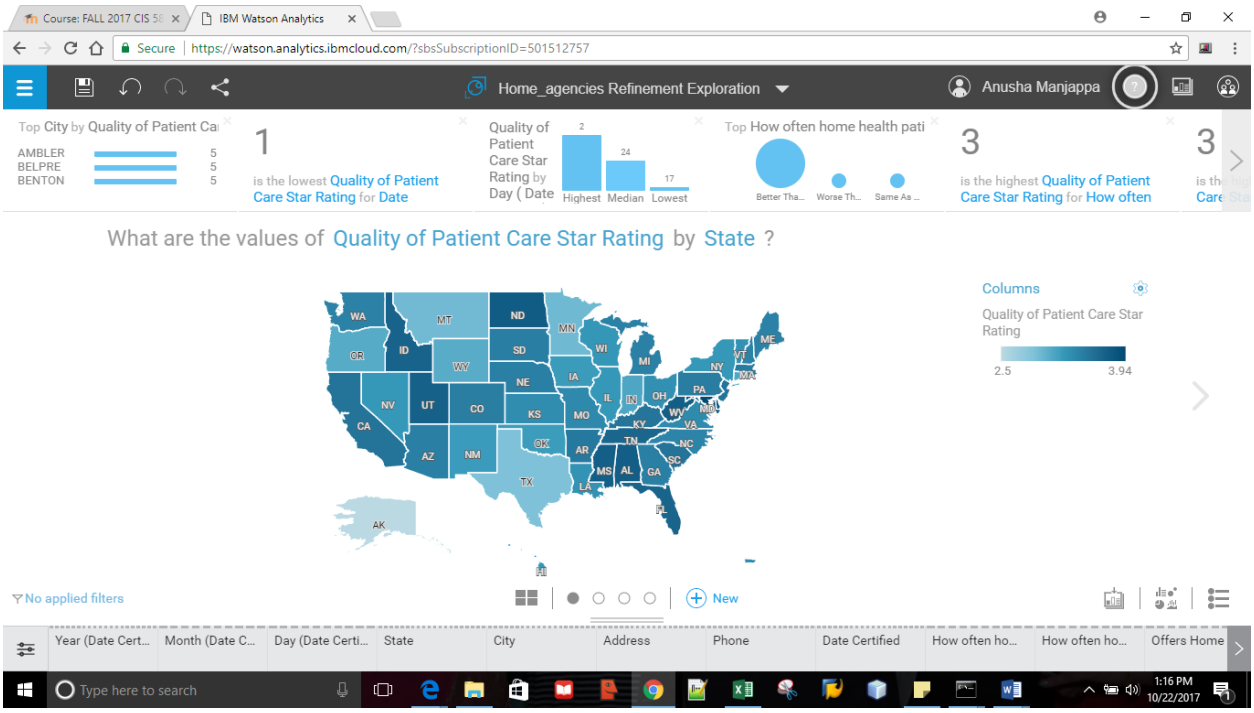

IBM Watson Analytics
https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757
Home_agencies Refinement
Anusha Manjappa
Columns
Search
How often patients got better at bathing
How often patients got better at getting in and ...
How often patients got better at taking their dr...
How often patients got better at walking or mo...
How often patients had less pain when moving...
Unplanned_ER
Improved_breathing
How often the home health team began their p...
How often the home health team checked pati...
How often the home health team determined ...
How often the home health team made sure th...
How often the home health team taught patien...
With diabetes, how often the home health tea...
CMS Certification Number (CCN)*
How often the home health team checked pati...
State - City
Year (Date Certified) - Day (Date Certified)
Category_of_rating
Summary_of_services
Unhide all
Create a new column
Calculation
Data group
Hierarchy
Summary_of_services
Include (blank)
Invert
Clear
Type here to search
7:14 PM
10/22/2017

E) Data Visualizations:

1) What are the values of Quality of Patient Care Star Rating by State on the whole and which are the states with Quality of Patient Care Star Rating of 3.5 and above?

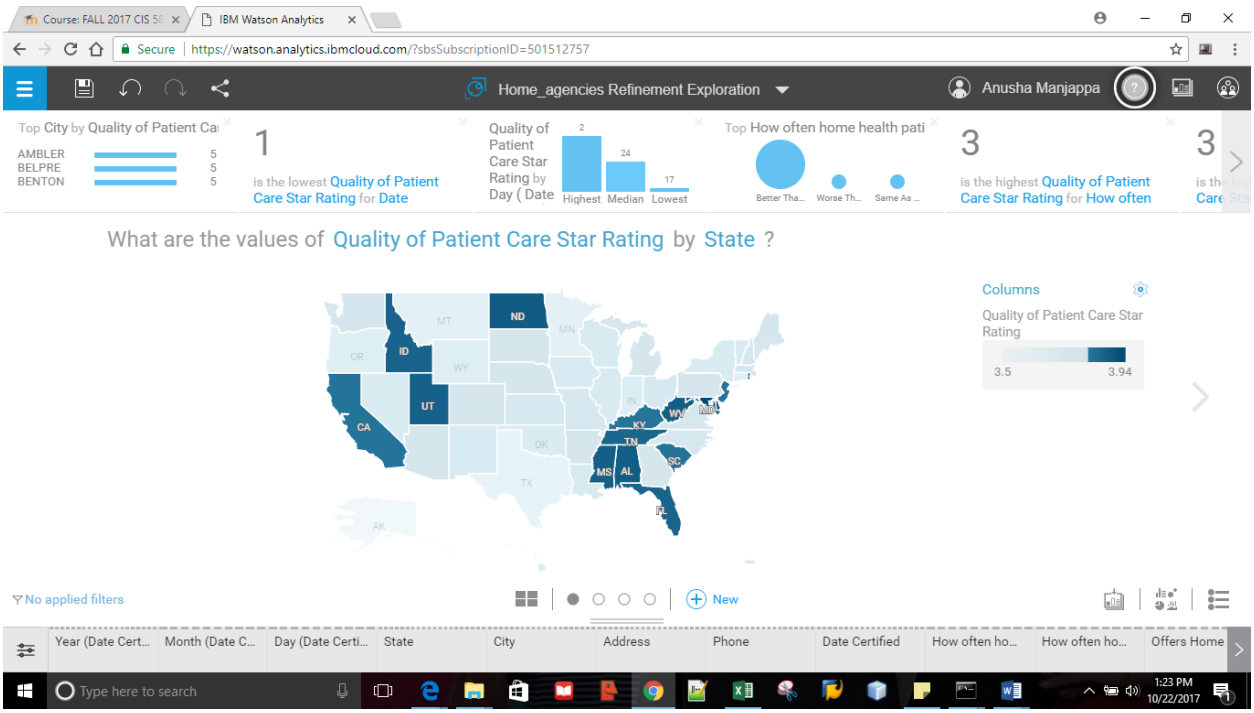
Fields used:

- Quality of Patient Care Star Rating
- State



The above graph shows the average star rating of every state in the USA. The average rating is calculated based on the star ratings of every home health agency in every state. It is found that the lowest average star rating is 2.5 and the highest average star rating is 3.94. The distribution is shown on the map with the color shades of blue from light to dark. The light colors shows a

lower average star rating and the darker shades show a higher average star rating. This would help us determine the quality of the home health agencies in various states.

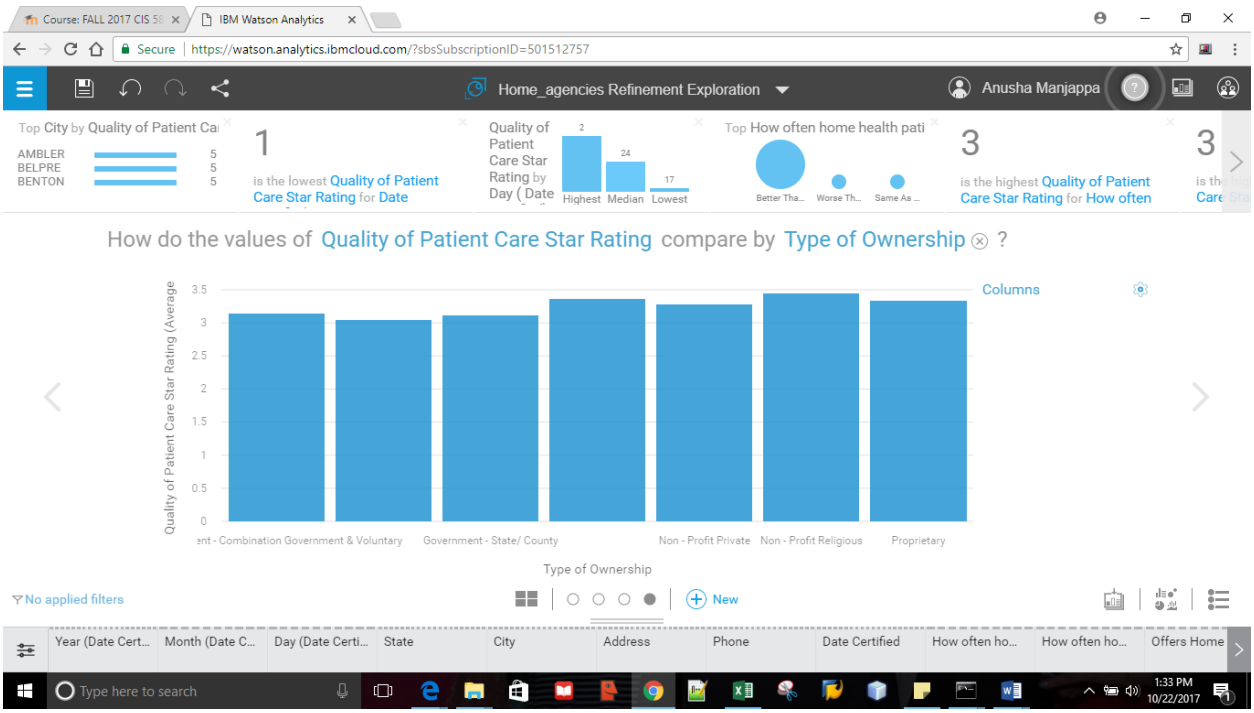


This is a continuation of the above graph which highlights only the states with a more than average star rating of above 3.5. The states with average star rating higher than 3.5 are ND, CA, UT, MS, AL, KY, ID, TN, SC, FL, and NJ.

2) Infer the effectiveness of home health care agencies by comparing the average quality star rating?

Fields used:

- Quality of patient care star rating
- Type of Ownership



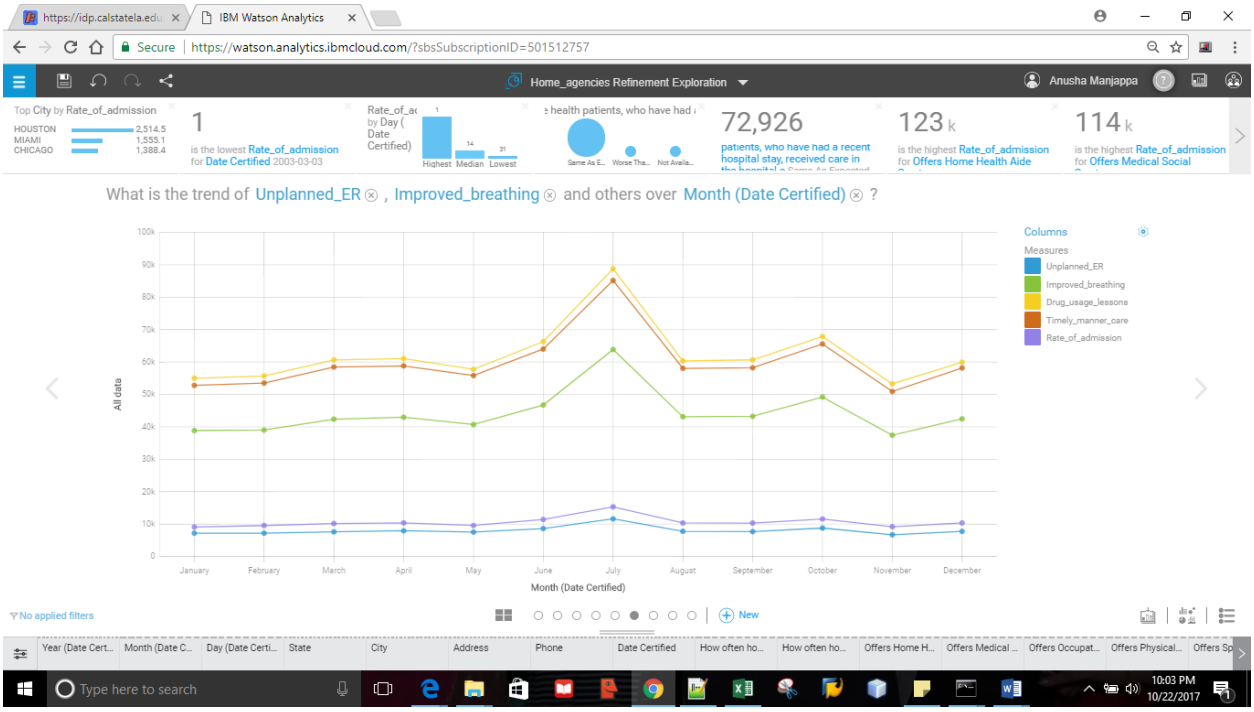
The above bar graph shows the average star rating of every home health care agencies in the USA. The average rating is calculated based on the count of star ratings of every home health agency in the USA. The graph finally shows that all the agencies have made good progress as all of them have their average star rating higher than or equal to 3. The highest average star rating is for the Non-Profit Religious home health care agencies with a rating of 3.45. This shows that all the home health care agencies in the USA are trust worthy and can be relied upon easily.

3) Compare the quality of services provided by home health care agencies of USA by considering a few services like Unplanned ER, improved breathing, Drug usage lessons given to patients and the timely care month wise?

Fields used:

- Unplanned ER
- Improved breathing
- Drug_Usage Lesson

- Timely manner care
- Rate_of_admission
- Month



The following graph consists of the growth of various facilities provided by the health care agencies like Unplanned ER, improved breathing, Drug usage lessons given to patient and the timely care provided to various patients on a month wise scale for the most recent year of 2016. We can infer that the service that was used the most was the Drug usage lessons. This service was asked for the most and it also shows a good percentage of response provided by the health care agencies. Second comes the timely care which is also one of the most important services that tell us about the effectiveness of each agency. The unplanned ER was found the least. This shows the areas on which the home health agencies should concentrate and improve themselves to provide good services to the public.

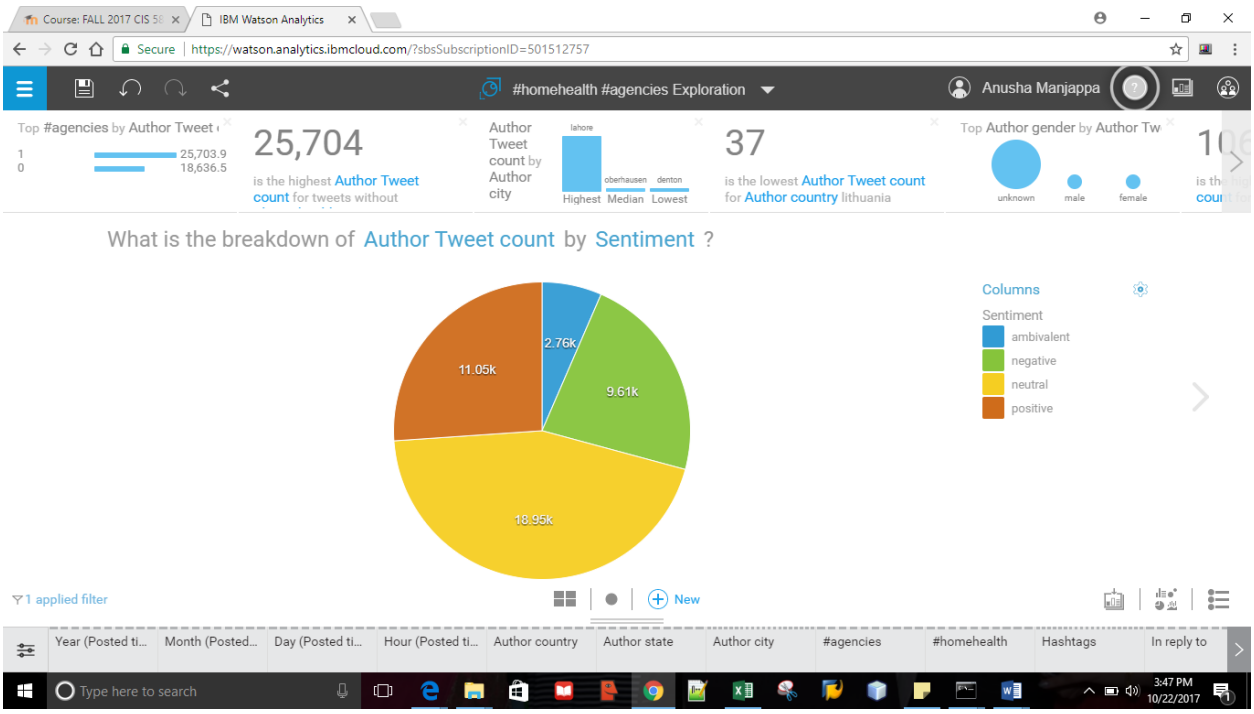
4) What is the breakdown of Author tweet count by sentiment for the hashtag of #homehealth and #agencies?

Fields used:

- Author Tweet Count
- Sentiment

Hashtags Used:

- #homehealth
- #agencies



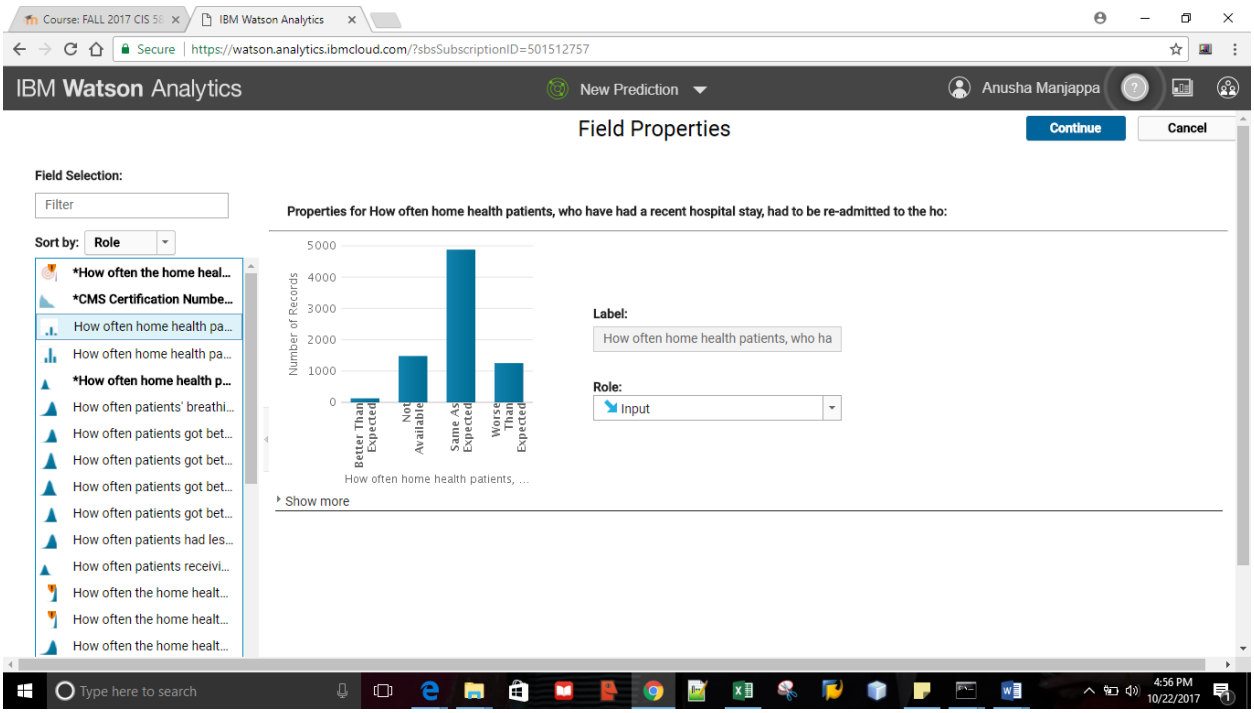
The dataset used for the above analysis is based on two twitter hashtags being #homehealth and #agencies. It relates to the dataset I am analyzing in this project. Sentimental analysis is one important aspect used in identifying and categorizing opinions expressed in a piece of text, especially to determine whether the writer's attitude towards a topic. Here the pie chart shows the

total count of the tweets based on the categories of sentiments. The categories considered here are ambivalent, negative, neutral and positive. We can infer from the chart that the most count is for the category of neutral. The second most count is for the category positive which is a good sign and nest in line of the number of counts comes negative and last is ambivalent.

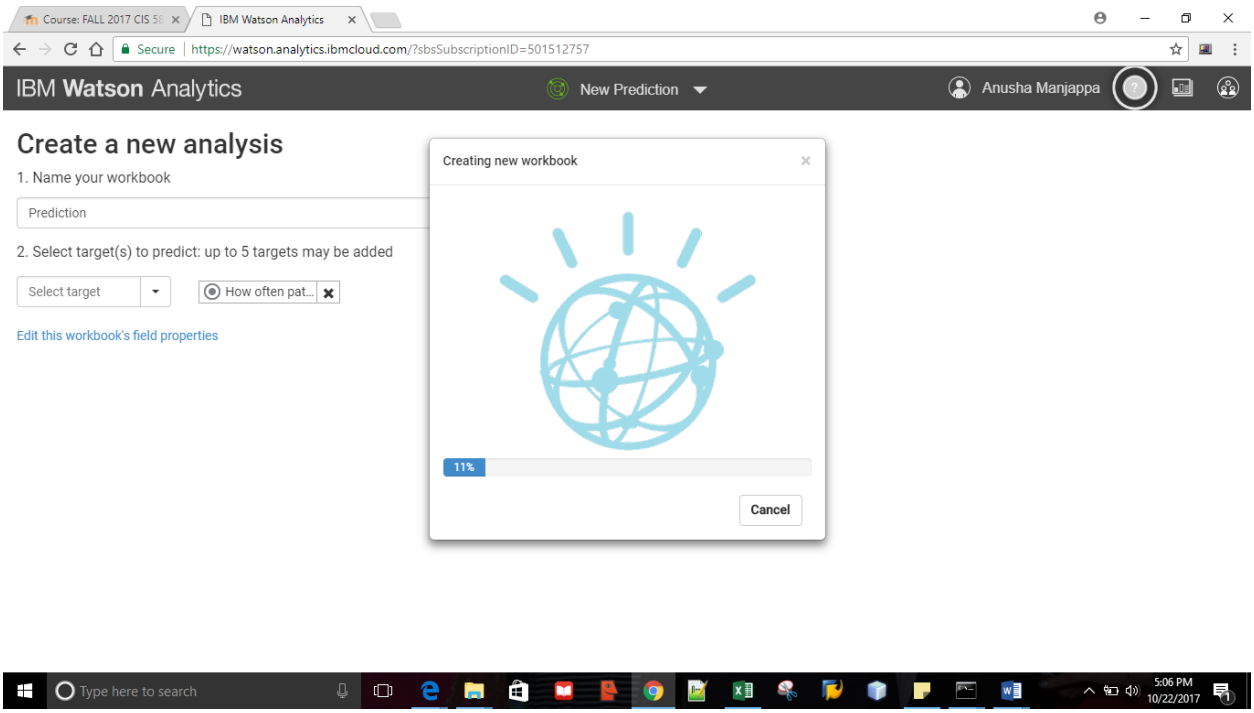
F)Prediction:

IBM Watson has a unique feature of prediction in it. This helps us to predict one of the fields of our dataset using few other fields of the dataset as input. It also shows the percentage of the predictive strength along with the fields that best predict the target. I have tried to predict how often the patients breathing has improved by considering various fields that would probably help in predicting that field correctly. I have considered a total of 6 fields for the input. On creating the prediction it was seen that the field that contributed the most in predicting how often the patients breathing has improved was the “Quality of patient Star Rating “ . The predictive strength field which predicts how correctly the prediction was made was 54% and 53%. The main insight shows that the combination of “Quality of patient Star Rating “ and “ how often patients get better ” lead to a good prediction of how often the patients breathing has improved which is 54%. This is good predictive strength as more than half of the fields were predicted correctly. The main condition on which the prediction was made was if the quality of patient care star rating was either 3 or 3.5 and above and how often the patients got better after taking drugs was more than 46% then the prediction of how often the patients breathing has improved was predicted correctly to a percentage of 66%. All the predictions on the whole are considered and the average prediction rate was found at the highest of 54%. The decision tree shows the flow of the prediction step by step considering each input field to see which field would predict the best outcome. The bulls eye shows us the

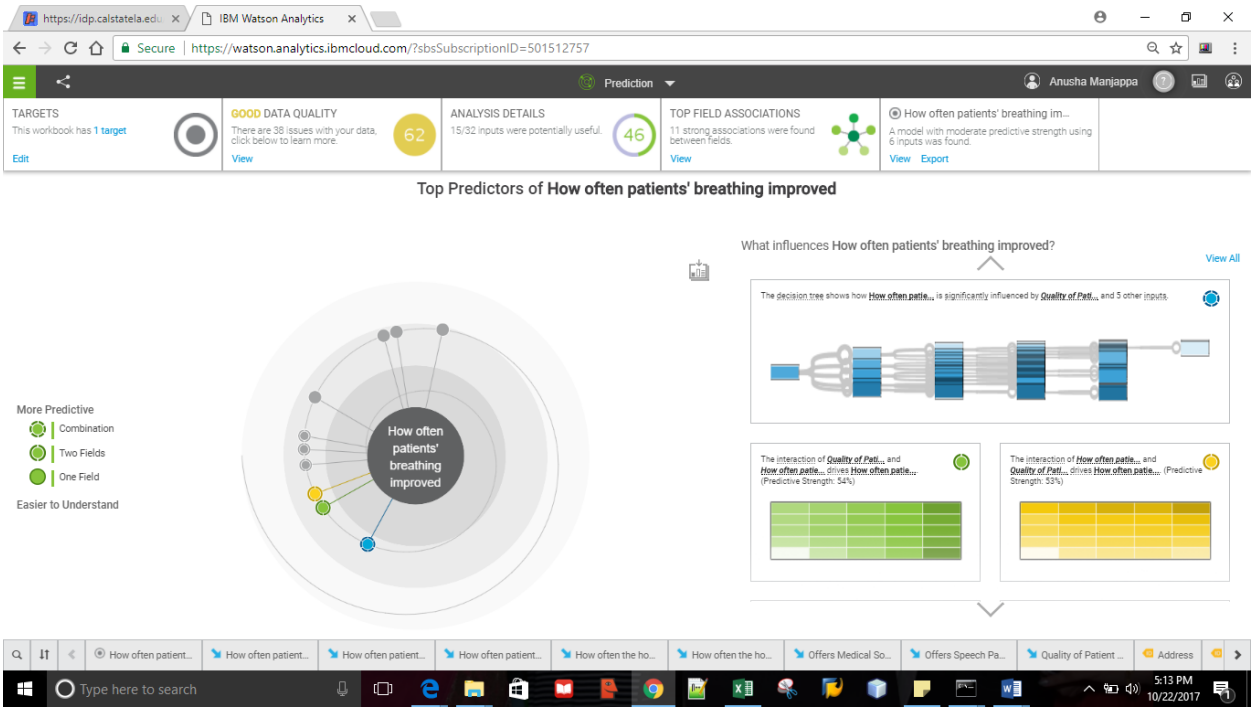
range of different predictions from the factors which gave the highest prediction to the lowest prediction. The screenshots of the detailed steps are given below.



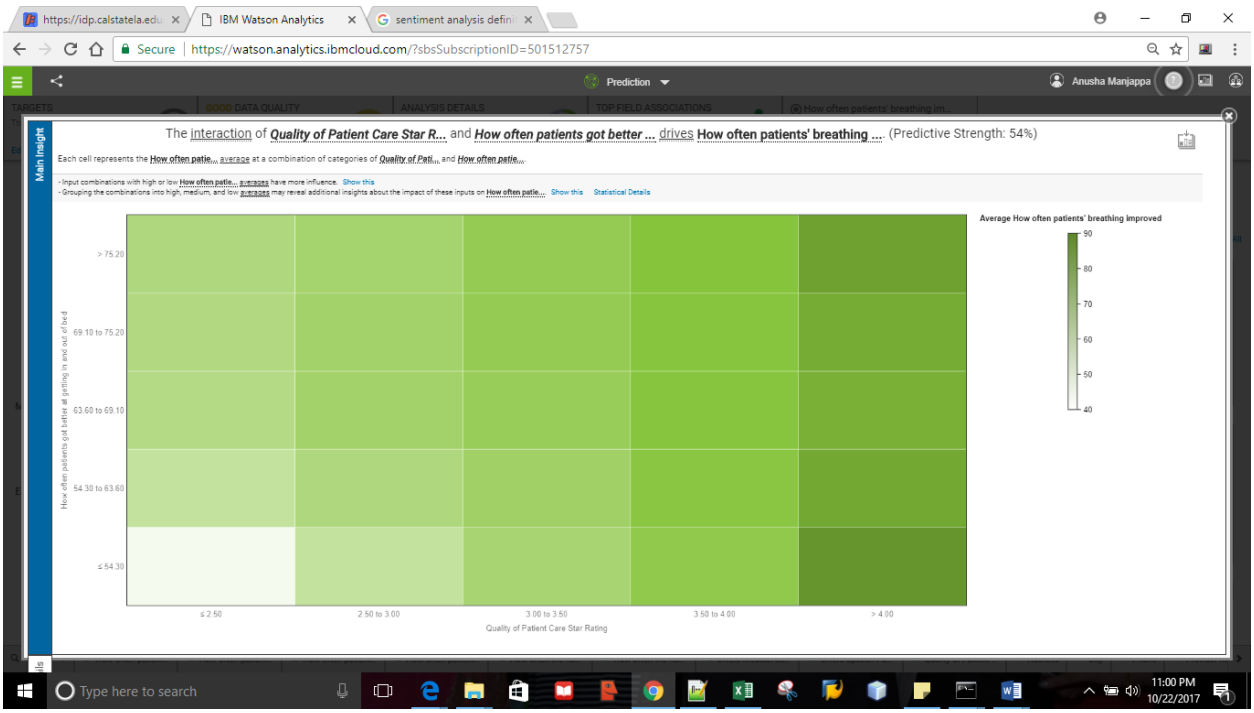
The above screenshot shows the process of selecting the input fields to predict a target dataset field.



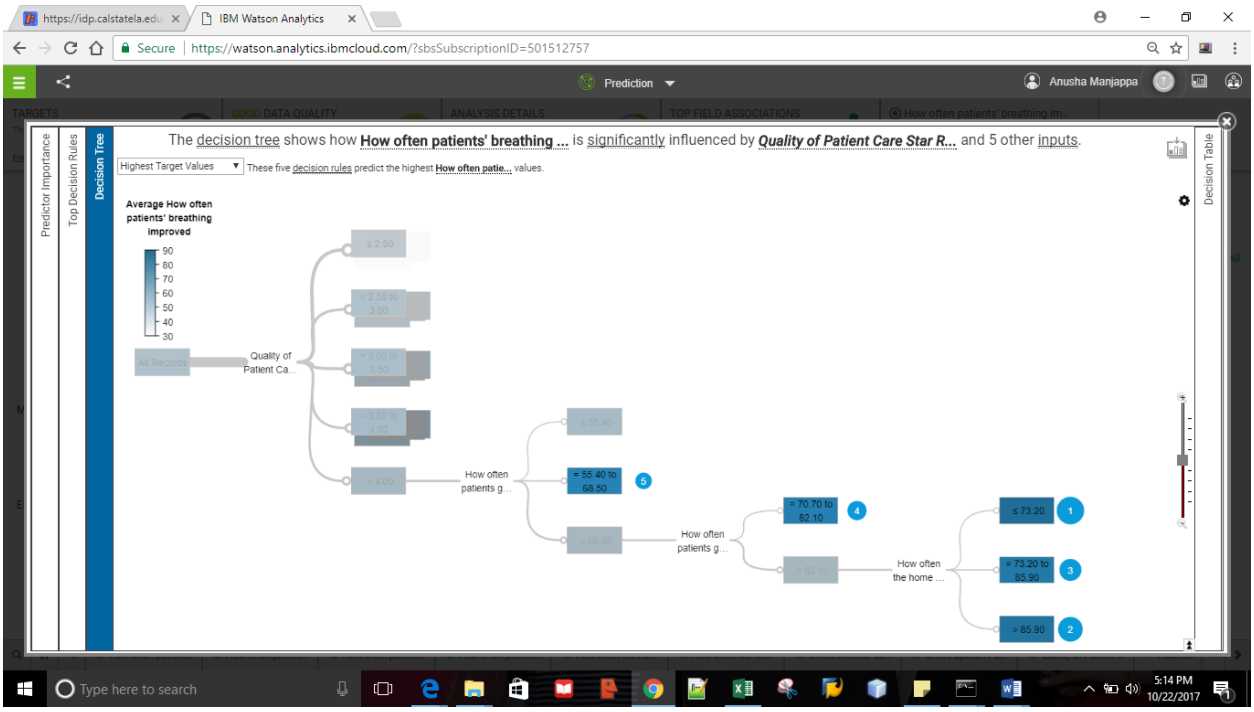
This screenshot shows the process of creating the prediction after selecting the appropriate input and target fields.



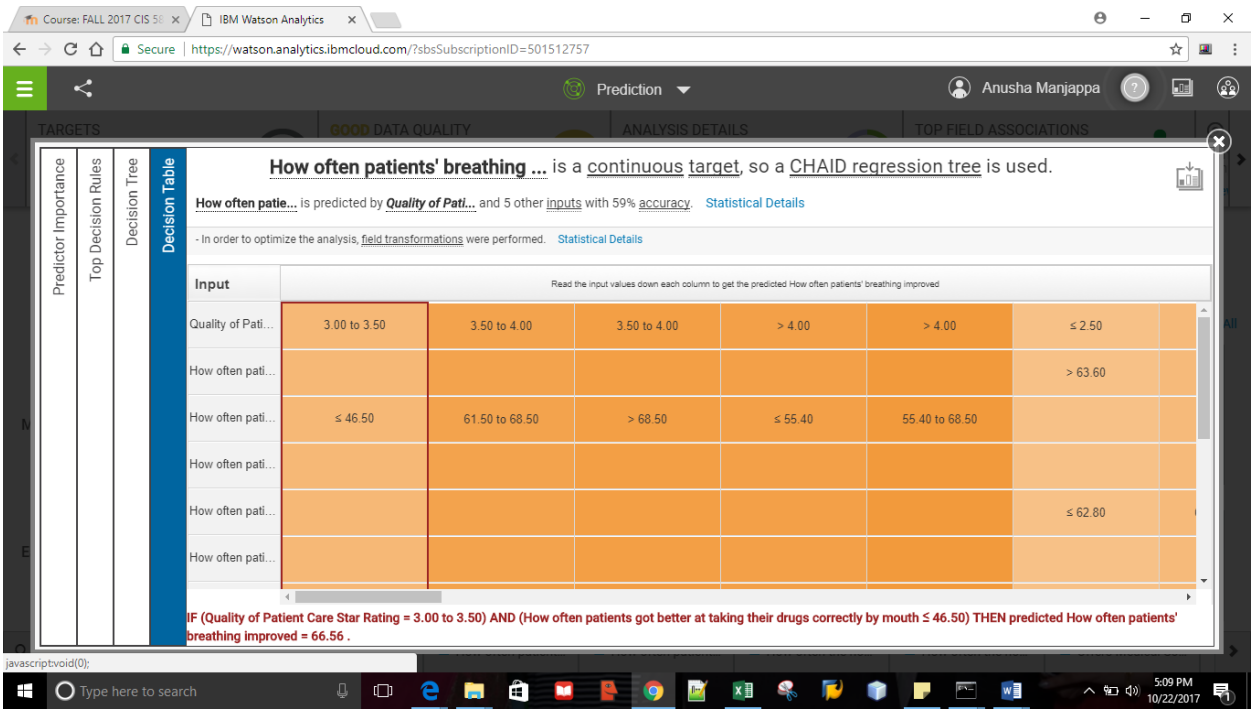
The above screenshot shows the bull's eye representation of the prediction.



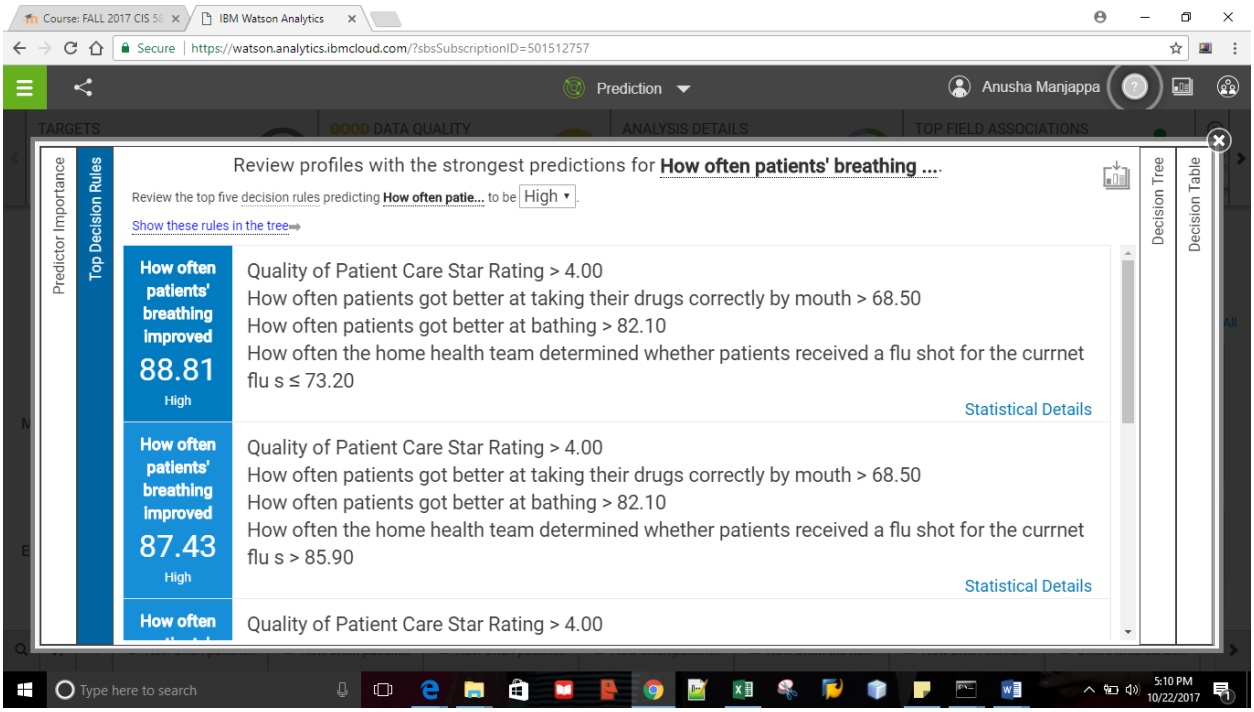
The above screenshot shows the main insight where the combination of " Quality of patient Star Rating " and " how often patients get better " lead to a good prediction of 54%.



The above screenshot shows the Decision tree involved in the prediction process.



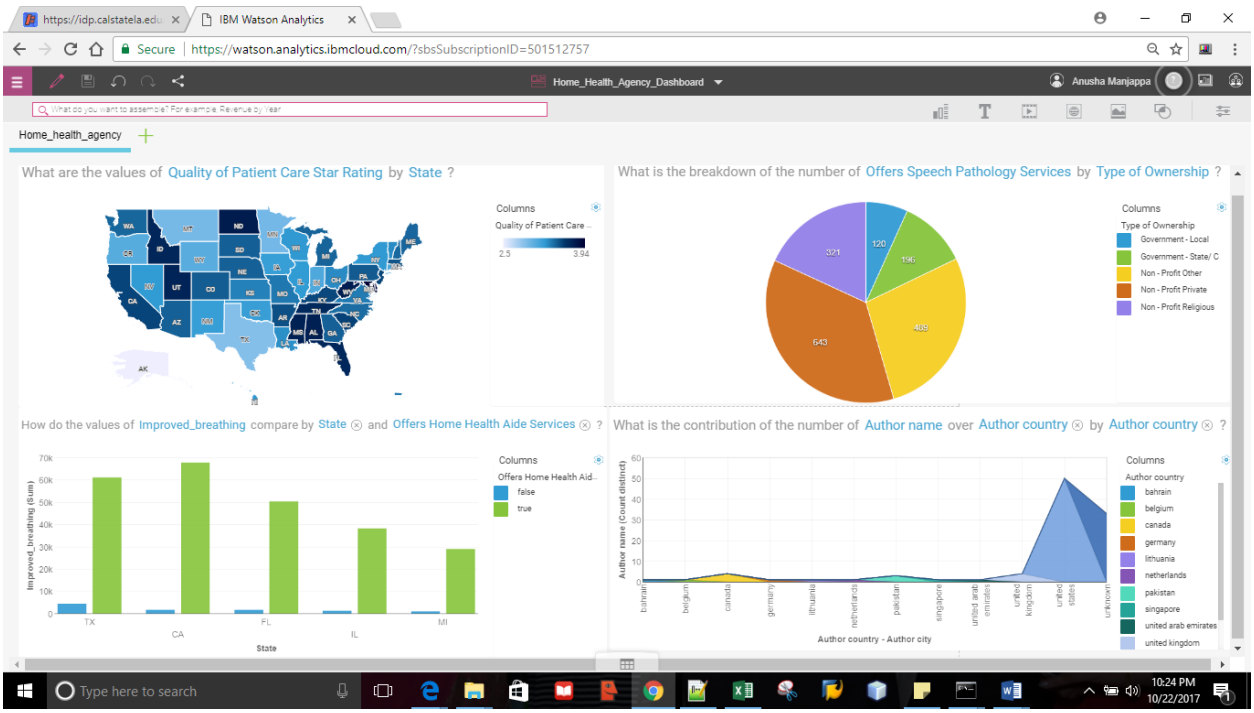
The above screenshot shows the Decision table involved in the prediction process.



G)Dashboard :

The following dashboard has a summary of all the analysis done .This portrays a good story to the viewer as a step by step procedure to understand the analysis.

- The first graph shows the average star rating of every state in the USA. The average rating is calculated based on the star ratings of every home health agency in every state. It is found that the lowest average star rating is 2.5 and the highest average star rating is 3.94.
- The second graph shows the count of the different types of agencies that offer the service of speech pathology in a pie chart. The non-profit religious group contributes to the most.
- The third Graph shows how the improved breathing field compares by state and home health aide services by a stacked bar graph.
- The last graph is of the twitter hashtags #homehealth and #agencies. It shows the contribution of every author to the count of this hashtag by country names.


IBM Watson Analytics
https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757
Home_Health_Agency_Dashboard
Anusha Manjappa
Home_health_agency
What are the values of Quality of Patient Care Star Rating by State ?
Columns
Quality of Patient Care ...
2.5
3.94
What is the breakdown of the number of Offers Speech Pathology Services by Type of Ownership ?
Columns
Type of Ownership
Government - Local
Government - State/ C
Non - Profit Other
Non - Profit Private
Non - Profit Religious
120
196
485
643
321
How do the values of Improved_breathing compare by State and Offers Home Health Aide Services ?
Improved_breathing (Sum)
State
TX
CA
FL
IL
MI
Columns
Offers Home Health Aid...
false
true
What is the contribution of the number of Author name over Author country by Author country ?
Author name (Count distinct)
Author country - Author city
Columns
Author country
bahrain
belgium
canada
germany
lithuania
netherlands
pakistan
singapore
united arab emirates
united kingdom
Type here to search
10:24 PM
10/22/2017