

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

NAME: Anusha Manjappa

CIN: 305851746

CIS5810 Project1

Fall 2017

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

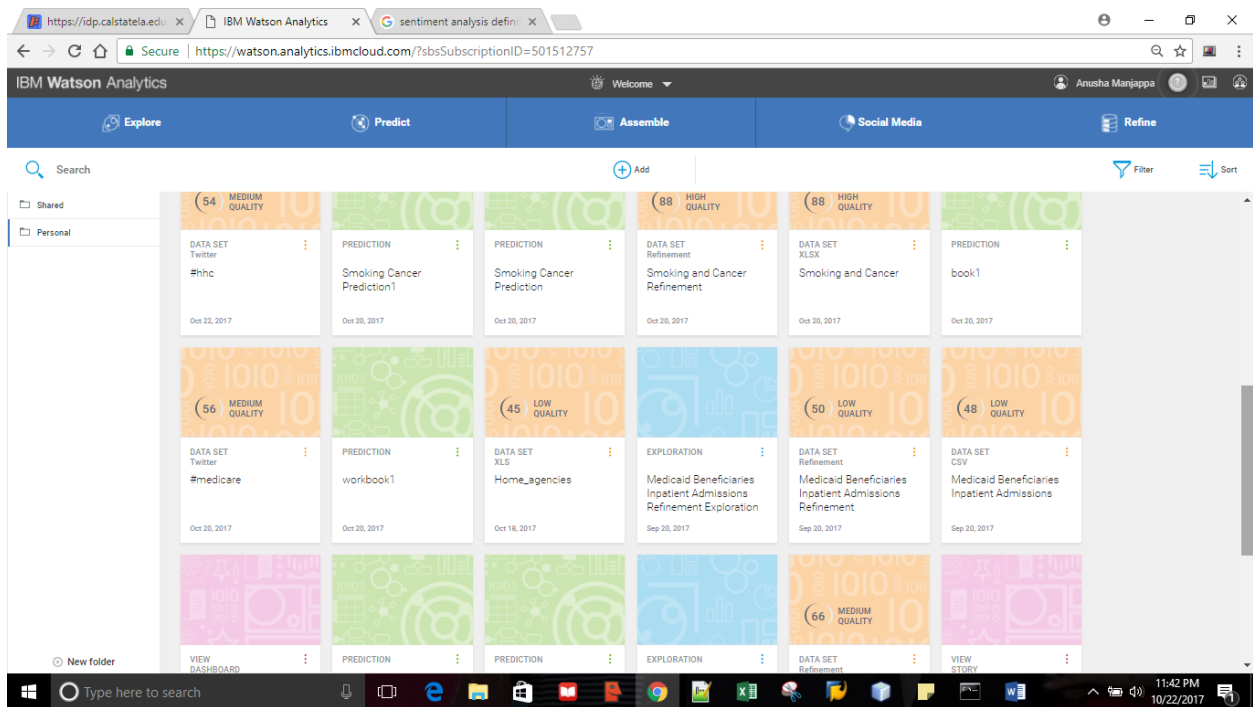
I have used two hashtags from the most recent 5 years of twitter data. I used the hashtags #homehealth #agencies. Homehealth tells us about the home healthcare agencies that are being talked of the most. It also refers to the home health plans are being talked about the most. Agencies hashtag would help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

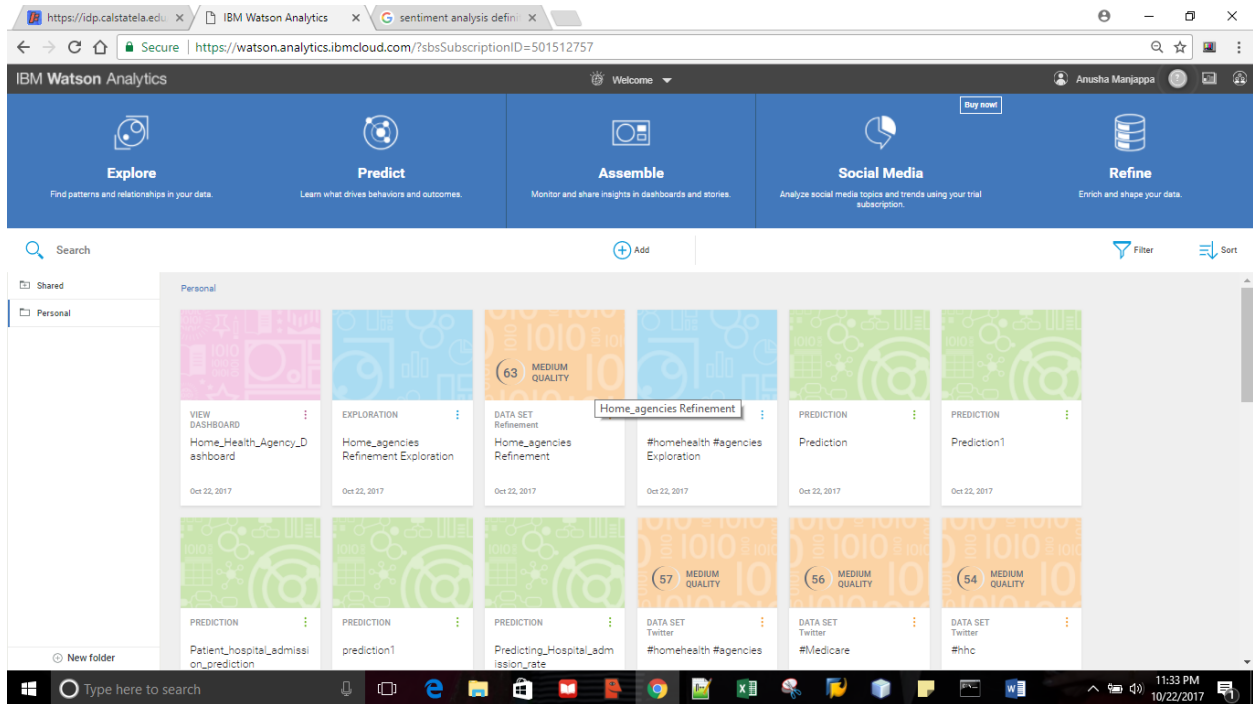
B) Data Quality

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 Expected
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 Expected
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 BAL	COVINA	91723
3250 WILS	LOS ANGE	90010
2095 W VI	VISTA	92083
800 WEST	ESCONDIDO	92025

- After

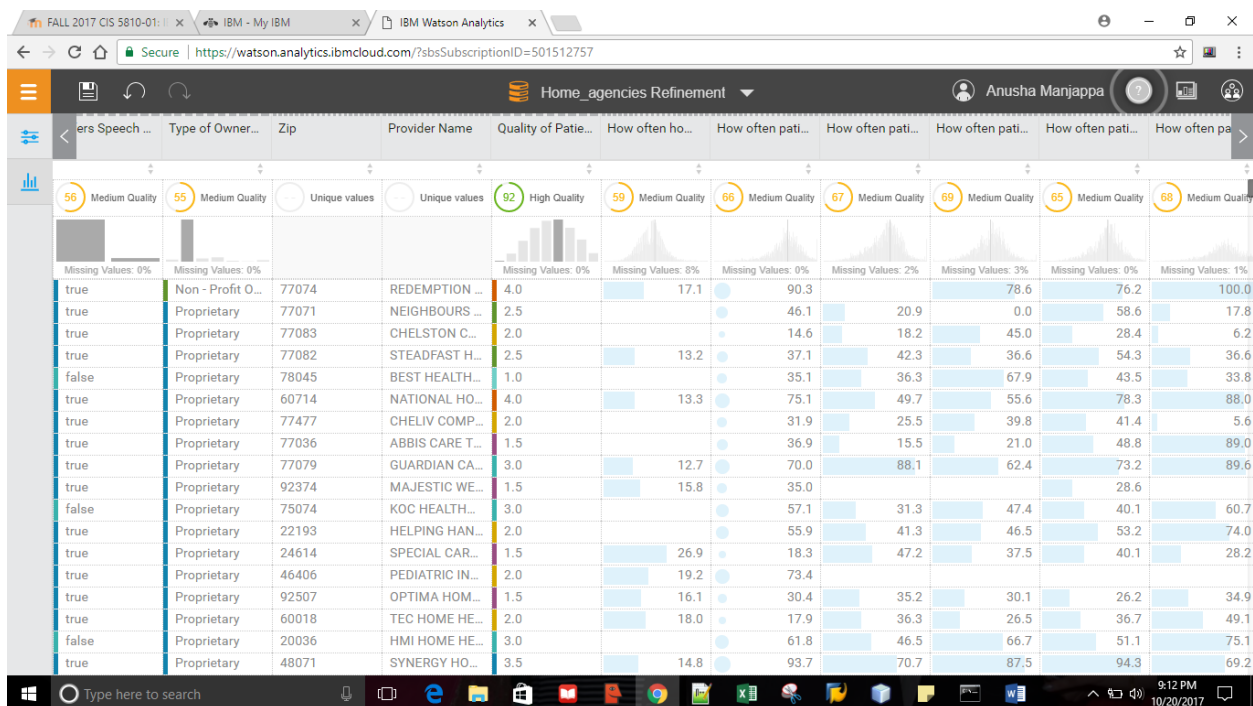
350 TERRA	REDLANDS	92373
270 W BAL	COVINA	91723
3250 WILS	LOS ANGELES	90010
2095 W VI	VISTA	92083
800 WEST	ESCONDIDO	92025

- Refinement in IBM Watson

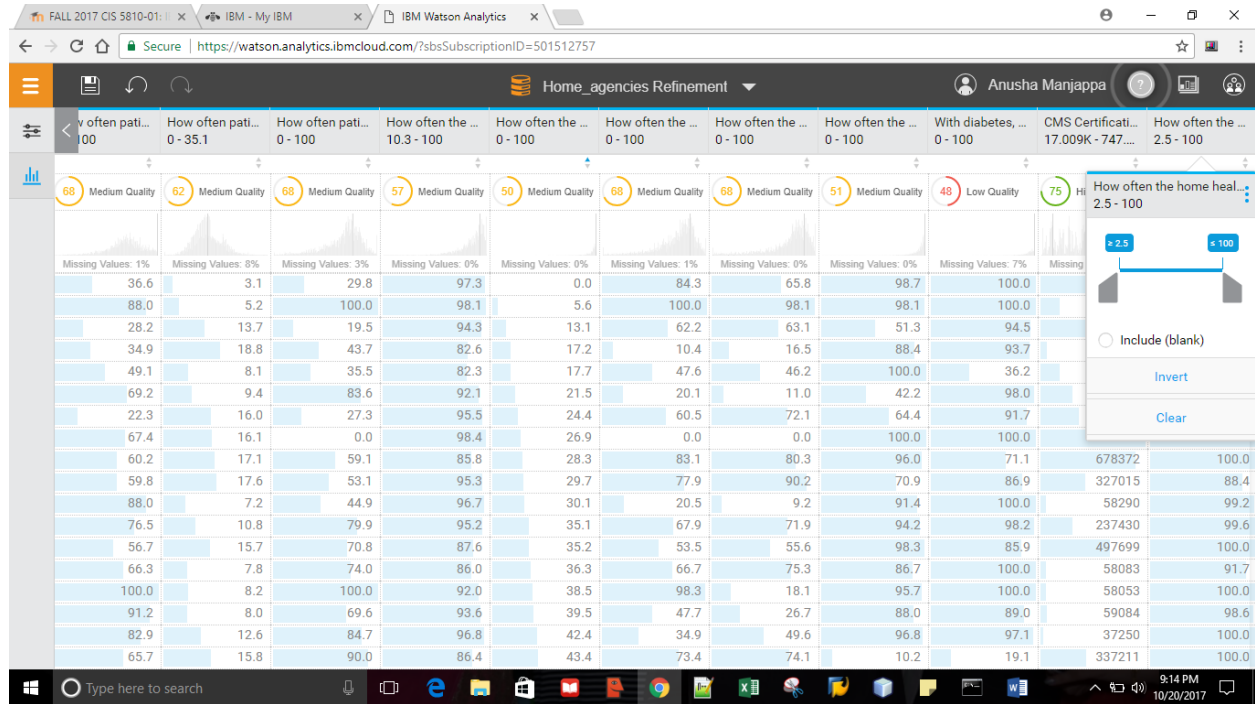
a) Missing Values

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 a few such changes are shown below.

- Before

https://idp.calstatela.edu x IBM Watson Analytics x

Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757

Home_agencies Refinement Anusha Manjappa

Year (Date Cert...)	Month (Date C...	Day (Date Cert...	State	City	Address	Phone	Date Certified	How often ho...	How often ho...	Offers Home H...	Offers Medical ...	Offers Occupat...	Offers Physical...	Offers Sp...
1994	July	21	NM	LAS CRUCES	2293 DIVOT A...	5755212663	1994-07-21	Same As Exp...	Same As Exp...	true	true	true	true	true
1995	February	21	NM	LAS CRUCES	3870 FOOTHL...	5755568409	1995-02-21	Same As Exp...	Worse Than ...	true	true	true	true	true
1995	May	25	NM	ARTESIA	1301 WEST G...	5757462892	1995-05-25	Better Than E...	Worse Than ...	true	false	true	true	true
1995	September	27	NM	CLOVIS	1944 WEST 2...	5757692243	1995-09-27	Same As Exp...	Same As Exp...	true	false	false	true	false
1996	April	9	NM	FARMINGTON	408 NORTH A...	5053266024	1996-04-09	Same As Exp...	Same As Exp...	true	false	false	true	false
1996	July	19	NM	LAS CRUCES	425 S TELSH...	5755248302	1996-07-19	Same As Exp...	Worse Than ...	true	false	true	true	true
1996	September	25	NM	FARMINGTON	727 EAST LUTE	5053262525	1996-09-25	Same As Exp...	Same As Exp...	true	true	true	true	true
1996	November	15	NM	LAS CRUCES	2975 TE...			Better Than E...	true	false	true	true	true	true
1996	October	21	NM	CARLSBAD	1300 N...			Same As Exp...	true	false	true	true	true	true
1997	June	30	NM	ESPANOLA	1010 S...			Worse Than ...	false	true	false	true	true	false
1997	June	30	NM	ROSWELL	315 WE...			Worse Than ...	true	true	true	true	true	true
1998	August	4	NM	CARLSBAD	513 S G...			Worse Than ...	true	true	true	true	true	true
2001	March	27	NM	RUIDOSO	1096 M...			Worse Than ...	true	true	true	true	true	false
2003	March	13	NM	TRUTH OR CO...	1400 N...			Same As Exp...	true	true	true	true	true	true
2003	December	17	NM	ROSWELL	217-A N...			Same As Exp...	true	true	true	true	true	true
2003	November	21	NM	LAS CRUCES	505 SQ...			Same As Exp...	true	true	true	true	true	true
2003	October	15	NM	FARMINGTON	2800 N...			Worse Than ...	true	true	true	true	true	true
2004	June	25	NM	ROSWELL	500 NQ...			Same As Exp...	true	true	true	true	true	true
2004	August	27	NM	ALBUQUERQUE	8725 AL...			Same As Exp...	true	true	true	true	true	true
2004	August	16	NM	ALBUQUERQUE	5700 HU...			Worse Than ...	true	true	true	true	true	true
2004	October	15	NM	ALBUQUERQUE	5981 JE...			Same As Exp...	true	true	true	true	true	false
2004	November	18	NM	HOBBS	1000 N...			Same As Exp...	true	true	true	true	true	true
2004	November	29	NM	PORTALES	1608 E SPRUC...	5752264663	2004-11-29	Worse Than ...	true	true	true	true	true	true
2004	December	29	NM	CLOVIS	2300 NORTH...	5757690049	2004-12-29	Same As Exp...	Worse Than ...	true	true	true	true	true
2005	January	5	NM	SANTA TERESA	5312 RIO BRA...	5758742211	2005-01-05	Same As Exp...	Same As Exp...	true	true	true	true	true
2005	April	29	NM	ALAMOGORDO	1909 CUBA A...	5754346222	2005-04-29	Same As Exp...	Worse Than ...	true	false	true	true	true
2005	May	6	NM	ANTHONY	820 ANTHON...	5758823539	2005-05-06	Same As Exp...	Same As Exp...	true	true	true	true	true
2006	April	14	NM	LAS CRUCES	3780 FOOTHL...	5755252273	2006-04-14	Same As Exp...	Same As Exp...	true	true	true	true	true
2006	April	27	NM	HOBBS	1508 N DAL P...	5753939281	2006-04-27	Same As Exp...	Same As Exp...	true	true	true	true	true
2006	August	18	NM	ALBUQUERQUE	4131 MONTG...	5055634041	2006-08-18	Same As Exp...	Same As Exp...	true	true	true	true	true
2006	November	29	NM	ALBUQUERQUE	4308 CARLIS...	5058810425	2006-11-29	Same As Exp...	Worse Than ...	true	true	true	true	true
2006	November	30	NM	ALBUQUERQUE	300 VALENCI...	5058800400	2006-11-30	Same As Exp...	Same As Exp...	false	true	true	true	true

Properties

hospital stay, had to be re-admitted to the ho...

Change name

Type

Filter

Select value

Default

Aggregation

Count distinct

Type here to search

6:15 PM 10/22/2017

- After

https://idp.calstatela.edu x IBM Watson Analytics x

Secure | https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757

Home_agencies Refinement Anusha Manjappa

Year (Date Cert...)	Month (Date C...	Day (Date Cert...	State	City	Address	Phone	Date Certified	Re-Admission ...	How often ho...	Offers Home H...	Offers Medical ...	Offers Occupat...	Offers Physical...	Offers Sp...
2003	September	19	TX	SAN ANTONIO	10010 ROGER...	2107366100	2003-09-19	Re-Admission Rate	Available	true	true	true	true	true
2003	September	22	TX	SAN ANTONIO	4606 CENTER...	2103417800	2003-09-22	Available	Available	true	true	true	true	true
2003	September	10	TX	BAY CITY	2803 7TH STR...	9792440600	2003-09-10	Available	Available	true	true	true	true	true
2003	September	25	TX	ALLEN	1101 RAINTR...	9723907733	2003-09-25	Available	Available	true	true	true	true	true
2003	September	24	TX	CORPUS CHRIS...	1521 S. STAP...	3618879000	2003-09-24	Available	Available	true	true	true	true	true
2003	September	18	TX	HOUSTON	8306 BALLIN...	2813130508	2003-09-18	Available	Available	true	true	true	true	true
2003	October	2	TX	DEL RIO	2409 VETERA...	8307785566	2003-10-02	Available	Available	true	true	true	true	true
2003	November	21	TX	HOUSTON	9311 MEADO...	2815502928	2003-11-21	Available	Available	true	true	true	true	true
2003	October	30	TX	DE SOTO	1615 OSPREY...	9722249911	2003-10-30	Available	Available	true	true	true	true	true
2003	October	6	TX	SANGER	808 UTILITY R...	2148289991	2003-10-06	Available	Available	true	true	true	true	true
2003	September	26	TX	HOUSTON	4615 SOUTH...	7138401811	2003-09-26	Available	Available	true	true	true	true	true
2003	October	16	TX	MCALLEN	3703 NORTH...	9566879998	2003-10-16	Available	Available	true	true	true	true	true
2003	October	10	TX	TEXARKANA	5604 SUMME...	9032555100	2003-10-10	Available	Available	true	true	true	true	true
2003	November	3	TX	LUBBOCK	4511 UNIVER...	8067985683	2003-11-03	Available	Available	true	true	true	true	true
2003	November	24	TX	ROCKWALL	2305 RIDGE R...	9727725086	2003-11-24	Available	Available	true	true	true	true	true
2003	November	19	TX	EL PASO	1155 WESTM...	9159886602	2003-11-19	Available	Available	true	true	true	true	true
2003	December	1	TX	GARLAND	3617 BROAD...	2142217900	2003-12-01	Available	Available	true	true	true	true	true
2003	December	1	TX	MCALLEN	2214 W NOLA...	9569710981	2003-12-01	Available	Available	true	true	true	true	true
2003	December	3	TX	SAN ANGELO	3103 SOUTH...	3594489116	2003-12-03	Available	Available	true	true	true	true	true
2003	December	5	TX	DALLAS	12750 MERIT ...	2143609090	2003-12-05	Available	Available	true	true	true	true	true
2003	December	19	TX	HOUSTON	2646 SOUTH...	7132187099	2003-12-19	Available	Available	true	true	true	true	true
2003	December	19	TX	SAN ANTONIO	85 NE LOOP 4...	2108220475	2003-12-19	Available	Available	true	true	true	true	true
2004	January	20	TX	SAN ANTONIO	4203 WOODC...	2103490096	2004-01-20	Available	Available	true	true	true	true	true
2009	January	14	TX	EL PASO	1515 CESSNA...	9152319494	2009-01-14	Available	Available	true	true	true	true	true
2003	December	30	TX	PHARR	702 W INTERS...	9562234176	2003-12-30	Available	Available	true	true	true	true	true
2009	April	15	TX	ODESSA	6010 E HWY 1...	4326178125	2009-04-15	Available	Available	true	true	true	true	true
2004	January	8	TX	DESOTO	1801 N HAMP...	9722288500	2004-01-08	Available	Available	true	true	true	true	true
2004	January	13	TX	ADDISON	16750 WESTG...	9724349400	2004-01-13	Available	Available	true	true	true	true	true
2004	January	16	TX	CLEBURNE	1101 W HEND...	8172020617	2004-01-16	Available	Available	true	true	true	true	true
2004	January	20	TX	ROCKPORT	1521 W MARK...	3617290340	2004-01-20	Available	Available	true	true	true	true	true
2004	January	29	TX	HOUSTON	10615 ROCK...	7135540806	2004-01-29	Available	Available	true	true	true	true	true
2004	January	29	TX	SAN ANTONIO	5726 WEST H...	2103490096	2004-01-29	Available	Available	true	true	true	true	true

Re-Admission Rate

Available

Not Available

Worse Than Expected

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)

The screenshot shows the IBM Watson Analytics interface. A 'Data group' dialog box is open, showing the 'Quality of Patient Care Star Rating' field being grouped into four categories: Low, Below Average, Above Average, and High. The dialog box has a search bar and a list of items grouped under each category. The background shows a data table with columns for Year, Month, Day, State, City, Address, Phone, Date Certified, How often ho..., How often ho..., Offers Home H..., and Offer... The table contains data for various home health agencies across different states and cities, with dates ranging from 2003 to 2004.

Year (Date Cert...)	Month (Date C...	Day (Date Certi...	State	City	Address	Phone	Date Certified	How often ho...	How often ho...	Offers Home H...	Offer...
2003	December	19	TX	HOUSTON	2646 SOUTH...	7132187099	2003-12-19	Not Available	Not Available	true	true
2003	December	19	TX	SAN ANTONIO	85 NE LOOP 4...	2108220475	2003-12-19	Same As Exp...	Same As Exp...	true	true
2004	January	20	TX	SAN ANTONIO	4203 WOODC...	2103490096	2004-01-20	Same As Exp...	Same As Exp...	true	true
2009	January	14	TX	EL PASO	1515 CESSNA...	9152319494	2009-01-14	Same As Exp...	Same As Exp...	true	true
2003	December	30	TX	PHARR	702 W INTERS...	9562234176	2003-12-30	Not Available	Not Available	true	true
2009	April	15	TX	ODESSA	6010 E HWY 1...	4326178125	2009-04-15	Same As Exp...	Same As Exp...	true	true
2004	January	8	TX	DESOTO	1801 N HAMP...	9722288500	2004-01-08	Not Available	Not Available	true	true

The screenshot displays the IBM Watson Analytics interface. The main table shows data for various agencies, with columns for different patient care metrics and a calculated field 'Summary_of_services'. The calculated field is defined as (rate_of_admission + improved breathing) / 2. The table includes columns for 'rate_of_admission', 'How often the home health team began their patient care', 'How often the home health team checked patient', 'How often the home health team determined patient care needs', 'How often the home health team made sure patient care was safe', 'How often the home health team taught patient', 'With diabetes, how often the home health team checked patient', 'CMS Certification Number (CCN)*', 'How often the home health team checked patient', 'State - City', 'Year (Date Certified) - Day (Date Certified)', 'Category_of_rating', and 'Summary_of_services'.

Columns	rate_of_admission	How often the home health team began their patient care	How often the home health team checked patient	How often the home health team determined patient care needs	How often the home health team made sure patient care was safe	How often the home health team taught patient	With diabetes, how often the home health team checked patient	CMS Certification Number (CCN)*	How often the home health team checked patient	State - City	Category_of_rating	Summary_of_services
rate_of_admission	29.8	97.3	0.0	84.3	65.8	98.7	100.0	747715	100.0	HOUSTON	Below Average	21.5
How often patients got better at bathing	100.0	98.1	5.6	100.0	98.1	98.1	100.0	147642	94.6	NILES	High	56.65
How often patients got better at getting in and out of bed	19.5	94.3	13.1	62.2	63.1	51.3	94.5	497602	98.6	GRUNDY	Low	23.2
How often patients got better at taking their doctor's advice	43.7	82.6	17.2	10.4	16.5	88.4	93.7	59221	57.0	RIVERSIDE	Low	29.9
How often patients got better at walking or moving around	35.5	82.3	17.7	47.6	46.2	100.0	36.2	148317	100.0	DES PLAINES	Below Average	26.75
How often patients had less pain when moving around	83.6	92.1	21.5	20.1	11.0	42.2	98.0	237607	97.9	MADISON HEIGHTS	Above Average	49.2
Unplanned ER	27.3	95.5	24.4	60.5	72.1	64.4	91.7	287146	85.7	OMAHA	Low	21.0
Improved breathing	0.0	98.4	26.9	0.0	0.0	100.0	100.0	747199	100.0	RICHMOND	Below Average	5.35
How often the home health team began their patient care	59.1	85.8	28.3	83.1	80.3	96.0	71.1	678372	100.0	HOUSTON	Below Average	35.3
How often the home health team checked patient	44.9	96.7	30.1	20.5	9.2	91.4	100.0	58290	99.2	CHINO	Above Average	32.8
How often the home health team determined patient care needs	79.9	95.2	35.1	67.9	71.9	94.2	98.2	237430	99.6	MARINE CITY	Above Average	45.2
How often the home health team made sure patient care was safe	70.8	87.6	35.2	53.5	55.6	98.3	85.9	497699	100.0	HERNDON	Below Average	45.9
How often the home health team taught patient	74.0	86.0	36.3	66.7	75.3	86.7	100.0	58083	91.7	CITY OF INDIANAPOLIS	Below Average	44.45
With diabetes, how often the home health team checked patient	100.0	92.0	38.5	98.3	18.1	95.7	100.0	58053	100.0	WEST COVINA	High	54.1
CMS Certification Number (CCN)*	69.6	93.6	39.5	47.7	26.7	88.0	89.0	59084	98.6	COVINA	Above Average	38.7
How often the home health team checked patient	84.7	96.8	42.4	34.9	49.6	96.8	97.1	37250	100.0	CHANDLER	Below Average	53.3
State - City	90.0	86.4	43.4	73.4	74.1	10.2	19.1	337211	100.0	CROTON ON HUDSON	Below Average	53.8
Year (Date Certified) - Day (Date Certified)	26.2	88.8	45.1	4.5	4.6	7.8	13.4	109414	100.0	MIRAMAR	Low	24.55
Category_of_rating	48.9	78.9	47.2	8.4	92.3	100.0	97.6	557719	100.0	LONG BEACH	Below Average	38.6
Summary_of_services	80.8	92.6	48.7	80.8	80.7	98.7	92.5	227569	97.6	QUINCY	Above Average	48.6
	62.6	78.8	49.0	67.5	59.0	75.6	85.2	107635	100.0	TALLAHASSEE	Low	43.35
	85.7	57.8	50.0	75.8	80.4	86.5	76.0	327113	95.3	LAS VEGAS	Above Average	51.95
	44.6	87.2	50.0	65.7	82.1	50.7	100.0	287136	91.2	OMAHA	Below Average	31.3
	46.3	79.8	50.6	29.1	33.3	88.1	28.0	157067	94.4	SOUTH BEND	Low	29.95
	77.9	83.2	53.7	79.9	92.0	83.1	75.9	67501	99.6	PUEBLO	Below Average	47.55
	31.6	81.1	54.3	6.8	6.3	100.0	98.4	747194	98.3	HOUSTON	Low	26.0
	48.0	78.1	55.0	59.0	60.3	90.1	94.9	237559	100.0	PLYMOUTH	Below Average	35.75
	29.2	94.5	56.6	29.4	63.8	93.4	93.8	197402	96.9	SHREVEPORT	Low	22.3
	27.0	83.2	58.4	12.7	86.1	90.1	95.0	747744	85.2	KATY	Low	18.35
	32.7	84.9	58.9	47.7	53.3	98.1	97.1	459483	96.7	DALLAS	Low	31.75
	51.8	92.4	59.0	5.7	8.6	100.0	100.0	677995	97.9	HOUSTON	Below Average	33.5

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.

The screenshot displays the IBM Watson Analytics interface. At the top, the browser address bar shows the URL: <https://watson.analytics.ibmcloud.com/?sbsSubscriptionID=501512757>. The main workspace is titled "Home_agencies Refinement". On the left, a "Columns" panel lists various data fields, including "Rate_of_admission", "How often patients got better at bathing", "How often patients got better at getting in and out of water", "How often patients got better at taking their drugs", "How often patients got better at walking or moving", "How often patients had less pain when moving", "Unplanned_ER", "Improved_breathing", "How often the home health team began their patient care", "How often the home health team checked patient vital signs", "How often the home health team determined patient needs", "How often the home health team made sure their patient was safe", "How often the home health team taught patient about their condition", "With diabetes, how often the home health team checked patient blood sugar", "CMS Certification Number (CCN)*", "How often the home health team checked patient vital signs", "State - City", "Year (Date Certified) - Day (Date Certified)", "Category_of_rating", and "Summary_of_services". The "Summary_of_services" field is selected. In the center, a "Calculation" dialog box is open, showing the formula: $\text{Summary_of_services} = \text{Rate_of_admission} / 2 + \text{Improved_breathing} / 2$. The dialog box has "Cancel", "Calculation", and "Done" buttons. The background shows a data table with columns: "Offers Speech...", "Type of Owner...", "Zip", "Provider Name", "Quality of Patient Care", "Rate_of_admission", "How often patients got better at bathing", "How often patients got better at getting in and out of water", "How often patients got better at taking their drugs", "How often patients got better at walking or moving", "How often patients had less pain when moving", "Unplanned_ER", "Improved_breathing", "How often the home health team began their patient care", "How often the home health team checked patient vital signs", "How often the home health team determined patient needs", "How often the home health team made sure their patient was safe", "How often the home health team taught patient about their condition", "With diabetes, how often the home health team checked patient blood sugar", "CMS Certification Number (CCN)*", "How often the home health team checked patient vital signs", "State - City", "Year (Date Certified) - Day (Date Certified)", "Category_of_rating", and "Summary_of_services". The table contains data for various home health agencies, including "STEADFAST HOME HEALTH", "NATIONAL HOME HEALTH", "SPECIAL CARE HOME HEALTH", "OPTIMA HOME HEALTH", "TEC HOME HEALTH", "SYNERGY HOME HEALTH", "FIDELITY HOME HEALTH", "STARLITE HOME HEALTH", "NIGHTINGALE HOME HEALTH", "ALAMOGORDO HOME HEALTH", "ULTIMATE HOME HEALTH", "PATHWAYS HOME HEALTH", "CAPITAL HOME HEALTH", "NEBRASKA HOME HEALTH", "VICTORY HOME HEALTH", "CENTER FOR HOME HEALTH", "FRONTIER HOME HEALTH", "DREAMS HOME HEALTH", "HOME HEALTH CARE", "NATIONAL HOME HEALTH", "HUCKEY HOME HEALTH", "NEW HOPE HOME HEALTH", and "UNIFIED MEDICAL SERVICES".

The screenshot displays the IBM Watson Analytics web application. The main area shows a data table with columns for various health metrics, state, and city. The table is filtered by 'Home_agencies Refinement'. The interface includes a search bar, a list of columns, and a sidebar with navigation options like 'Summary of services' and 'Include (blank)'.

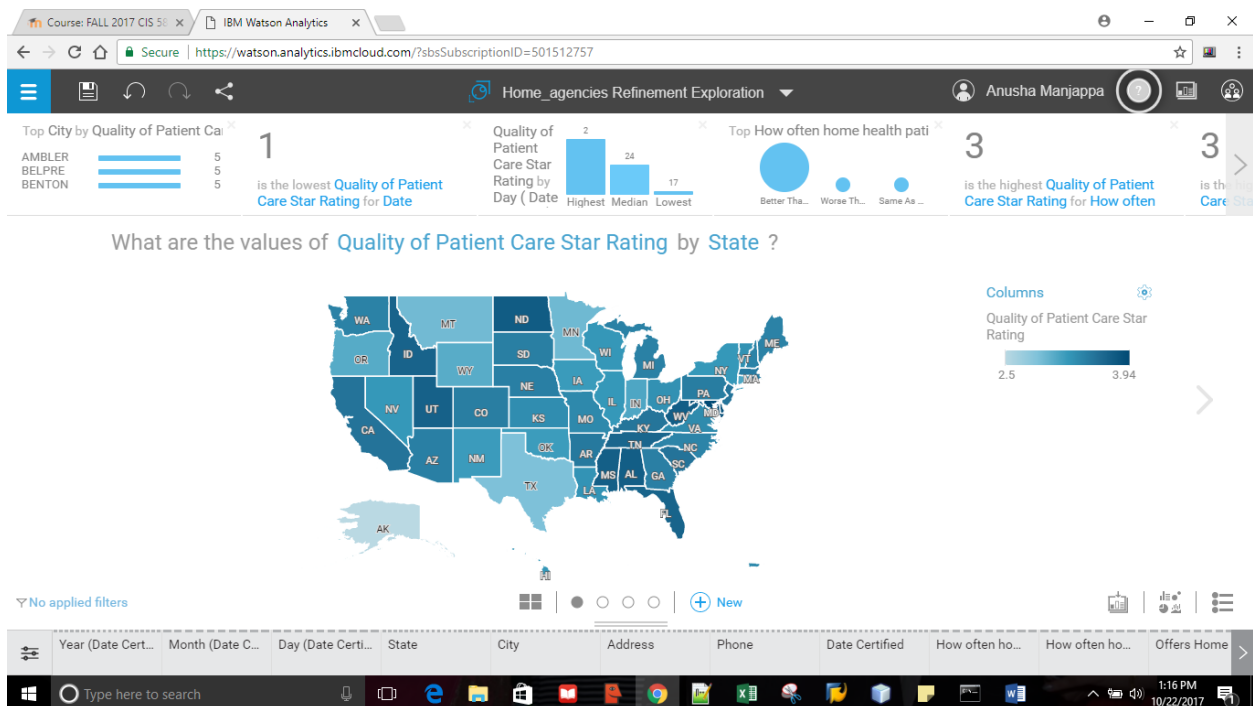
Columns	How often the ... 10.3-100	How often the ... 0-100	How often the ... 0-100	How often the ... 0-100	How often the ... 0-100	With diabetes, ... 0-100	CMS Certificat... 17.059K-747...	How often the ... 2.5-100	State - City	Category of ra...	Summary of...
Search	29.8	97.3	0.0	84.3	65.8	98.7	100.0	747715	100.0	HOUSTON	Below Average
✓ wate_of_admission	100.0	98.1	5.6	100.0	98.1	98.1	100.0	147642	94.6	NILES	High
✓ How often patients got better at bathing	19.5	94.3	13.1	62.2	63.1	51.3	94.5	497602	98.6	GRUNDY	Low
✓ How often patients got better at getting in and ...	43.7	82.6	17.2	10.4	16.5	88.4	93.7	59221	57.0	RIVERSIDE	Low
✓ How often patients got better at taking their d...	35.5	82.3	17.7	47.6	46.2	100.0	36.2	148317	100.0	DES PLAINES	Below Average
✓ How often patients got better at walking or mo...	83.6	92.1	21.5	20.1	11.0	42.2	98.0	237607	97.9	MADISON HEI...	Above Average
✓ How often patients had less pain when moving...	27.3	95.5	24.4	60.5	72.1	64.4	91.7	287146	85.7	OMAHA	Low
✓ Unplanned_ER	0.0	98.4	26.9	0.0	0.0	100.0	100.0	747199	100.0	RICHMOND	Below Average
✓ Improved_breathing	59.1	85.8	28.3	83.1	80.3	96.0	71.1	678372	100.0	HOUSTON	Below Average
✓ How often the home health team began their p...	53.1	95.3	29.7	77.9	90.2	70.9	86.9	327015	88.4	ALAMOGORDO	Below Average
✓ How often the home health team checked pati...	44.9	96.7	30.1	20.5	9.2	91.4	100.0	58290	99.2	CHINO	Above Average
✓ How often the home health team determined ...	79.9	95.2	35.1	67.9	71.9	94.2	98.2	237430	99.6	MARINE CITY	Above Average
✓ How often the home health team made sure th...	70.8	87.6	35.2	53.5	55.6	98.3	85.9	497699	100.0	HERNDON	Below Average
✓ How often the home health team taught patien...	74.0	86.0	36.3	66.7	75.3	86.7	100.0	58083	91.7	CITY OF INDU...	Below Average
✓ With diabetes, how often the home health tea...	100.0	92.0	38.5	98.3	18.1	95.7	100.0	58053	100.0	WEST COVINA	High
✓ CMS Certification Number (CCN)*	69.6	93.6	39.5	47.7	26.7	88.0	89.0	59084	98.6	COVINA	Above Average
✓ How often the home health team checked pati...	84.7	96.8	42.4	34.9	49.6	96.8	97.1	37250	100.0	CHANDLER	Below Average
✓ State - City	90.0	86.4	43.4	73.4	74.1	10.2	19.1	337211	100.0	CROTON ON ...	Below Average
✓ Category of rating	26.2	88.8	45.1	4.5	4.6	7.8	13.4	109414	100.0	MIRAMAR	Low
✓ Summary of services	48.9	78.9	47.2	8.4	92.3	100.0	97.6	557719	100.0	LONG BEACH	Below Average
Year (Date Certified) - Day (Date Certified)	80.8	92.6	48.7	80.8	80.7	98.7	92.5	225659	97.6	QUINCY	Above Average
Calculation	62.6	78.8	49.0	67.5	59.0	75.6	85.2	107635	100.0	TALLHASSEE	Low
Data group	44.6	87.2	50.0	65.7	82.1	50.7	100.0	287136	91.2	OMAHA	Below Average
Hierarchy	85.7	57.8	50.0	75.8	80.4	86.5	76.0	327113	95.3	LAS VEGAS	Above Average
Create a new column	46.3	79.8	50.6	29.1	33.3	88.1	28.0	157067	94.4	SOUTH BEND	Low
Calculation	77.9	83.2	53.7	79.9	92.0	83.1	75.9	67501	99.6	PUEBLO	Below Average
Data group	31.6	81.1	54.3	6.8	6.3	100.0	98.4	747194	98.3	HOUSTON	Low
Hierarchy	48.0	78.1	55.0	59.0	60.3	90.1	94.9	237559	100.0	PLYMOUTH	Below Average
Calculation	29.2	94.5	56.6	29.4	63.8	93.4	93.8	197402	96.9	SHREVEPORT	Low

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?

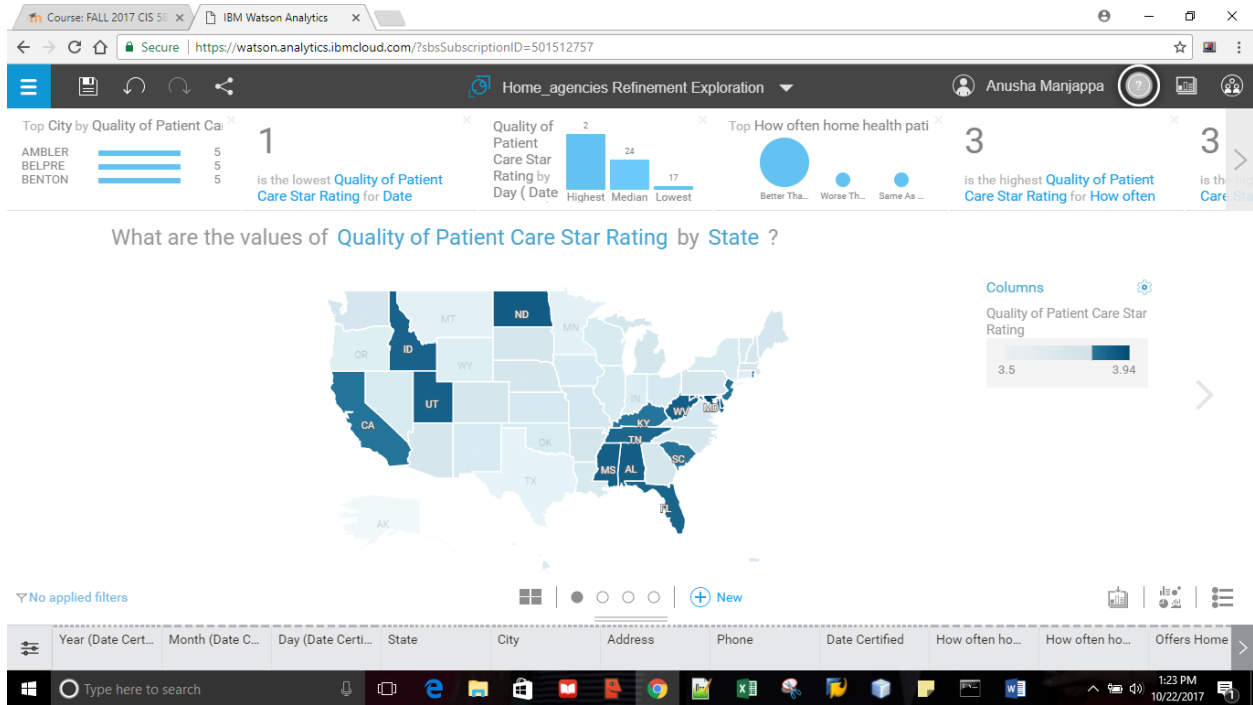
Fields used:

- 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.

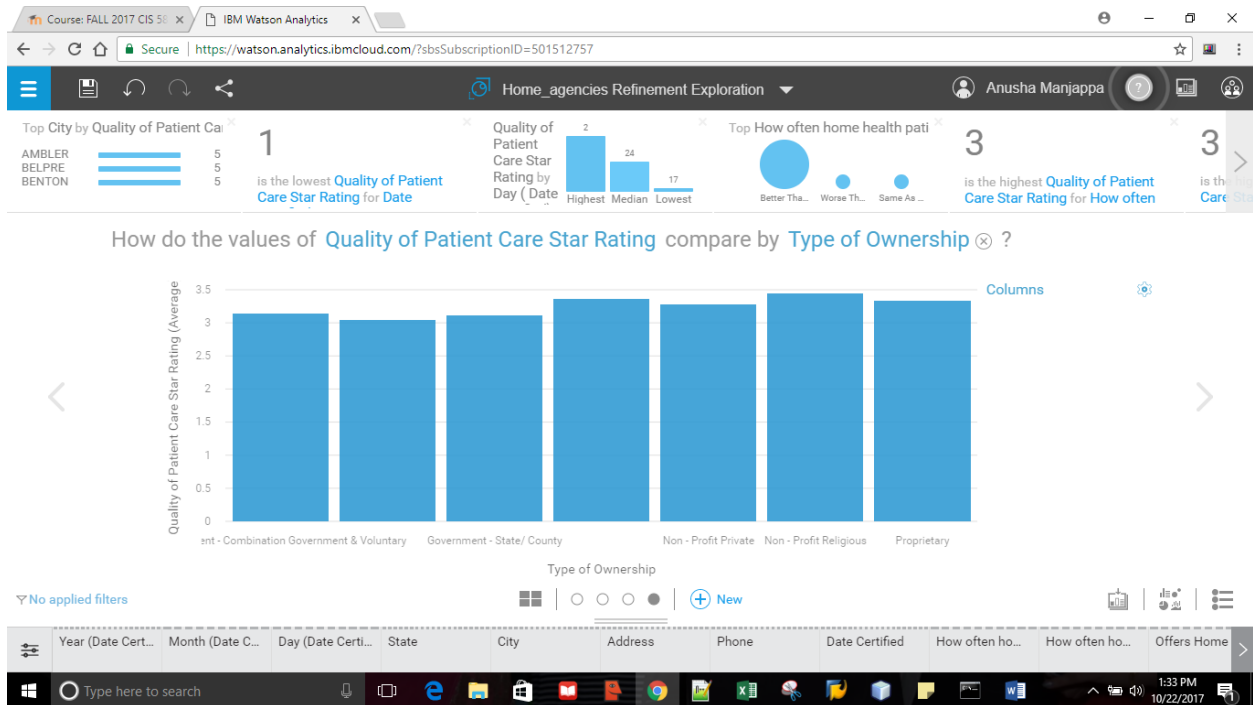


This is a continuation of the above graph which highlights only the states with a more than average star rating of above 3.5. The states with average star rating higher than 3.5 are ND, CA, UT, MS, AL, KY, ID, TN, SC, FL, and NJ.

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



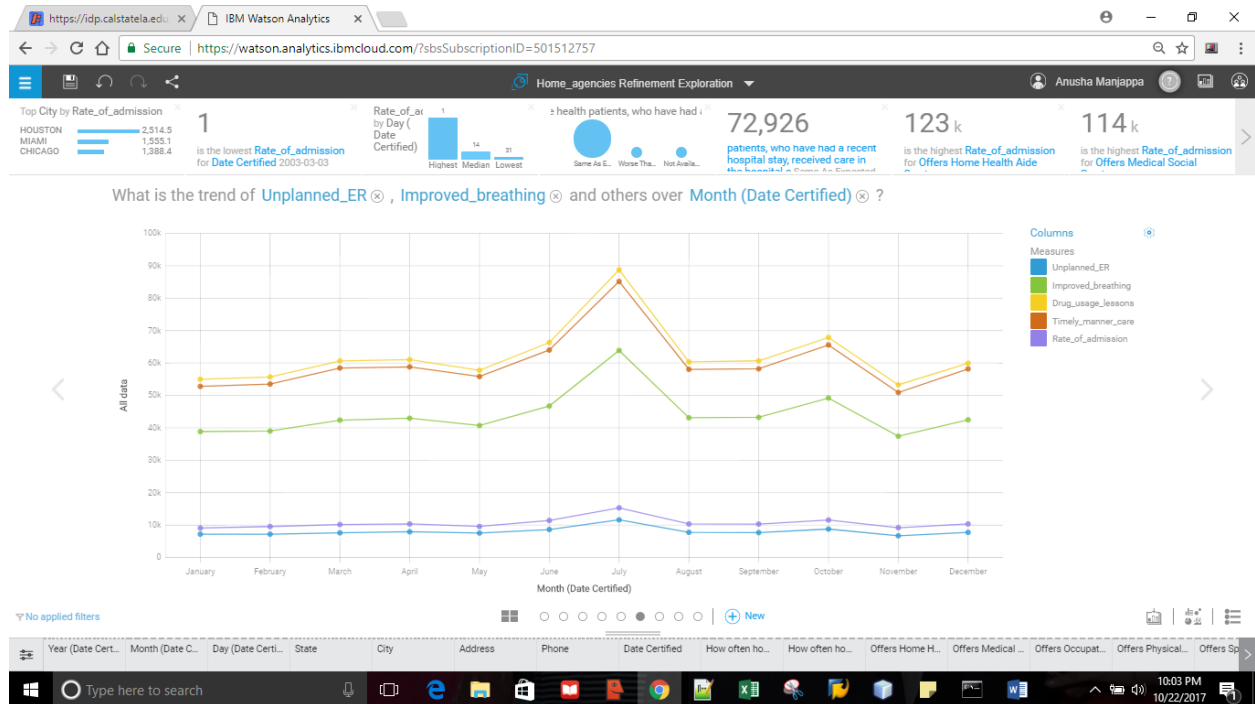
The above bar graph shows the average star rating of every home health care agencies in the USA. The average rating is calculated based on the count of star ratings of every home health agency in the USA. The graph finally shows that all the agencies have made good progress as all of them have their 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_Use 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.

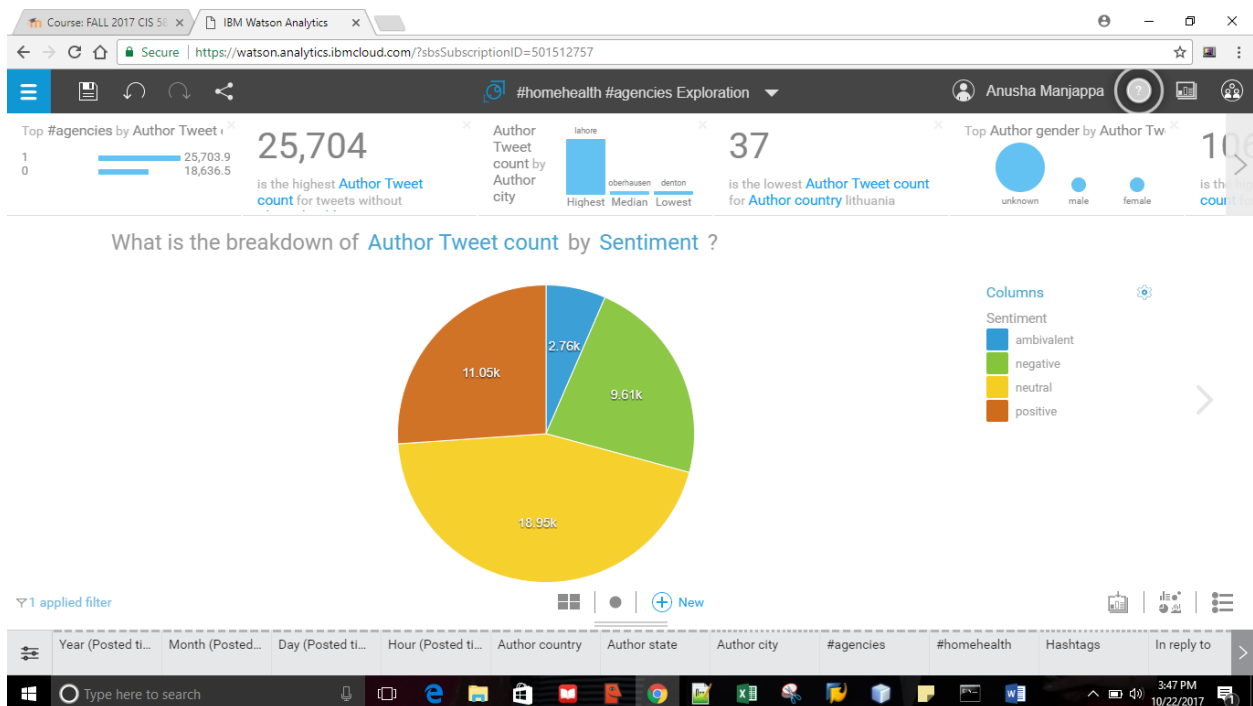
4) What is the breakdown of Author tweet count by sentiment for the hashtag of #homehealth and #agencies?

Fields used:

- Author Tweet Count
- Sentiment

Hashtags Used:

- #homehealth
- #agencies



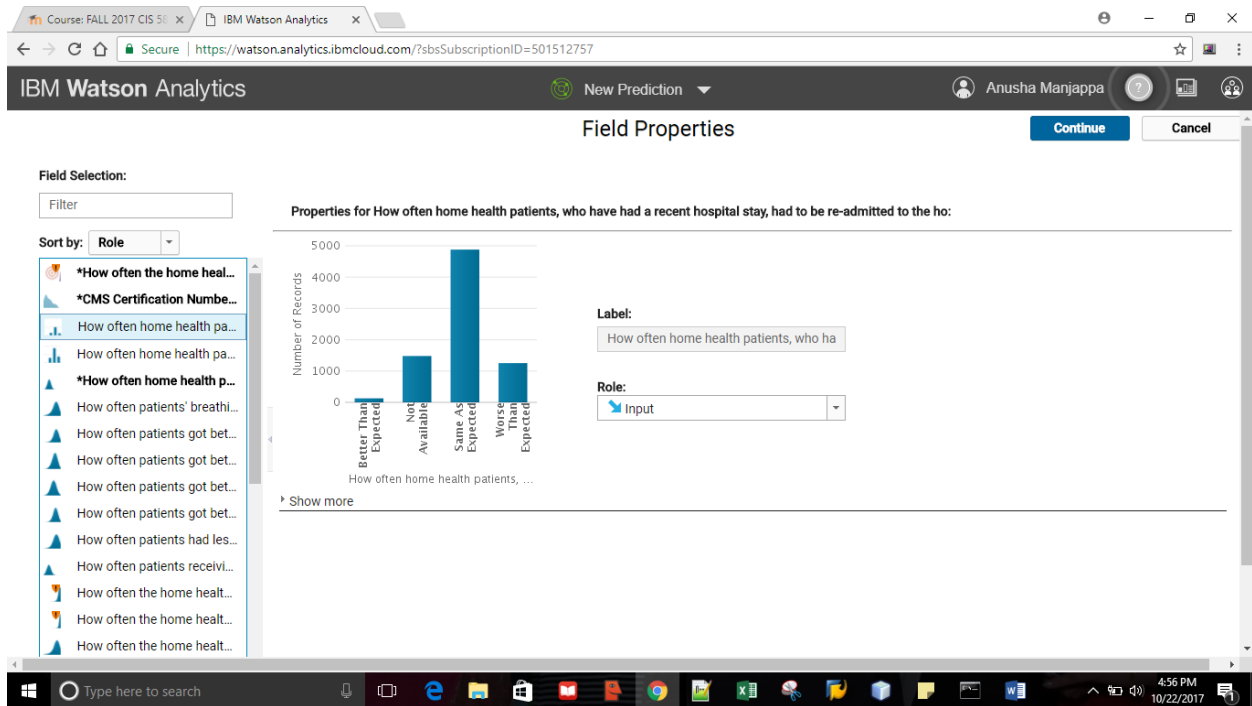
The dataset used for the above analysis is based on two twitter hashtags being #homehealth and #agencies. It relates to the dataset I am analyzing in this project. Sentimental analysis is one important aspect used in identifying and categorizing opinions expressed in a piece of text, especially to determine 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 next in line of the number of counts comes negative and last is ambivalent.

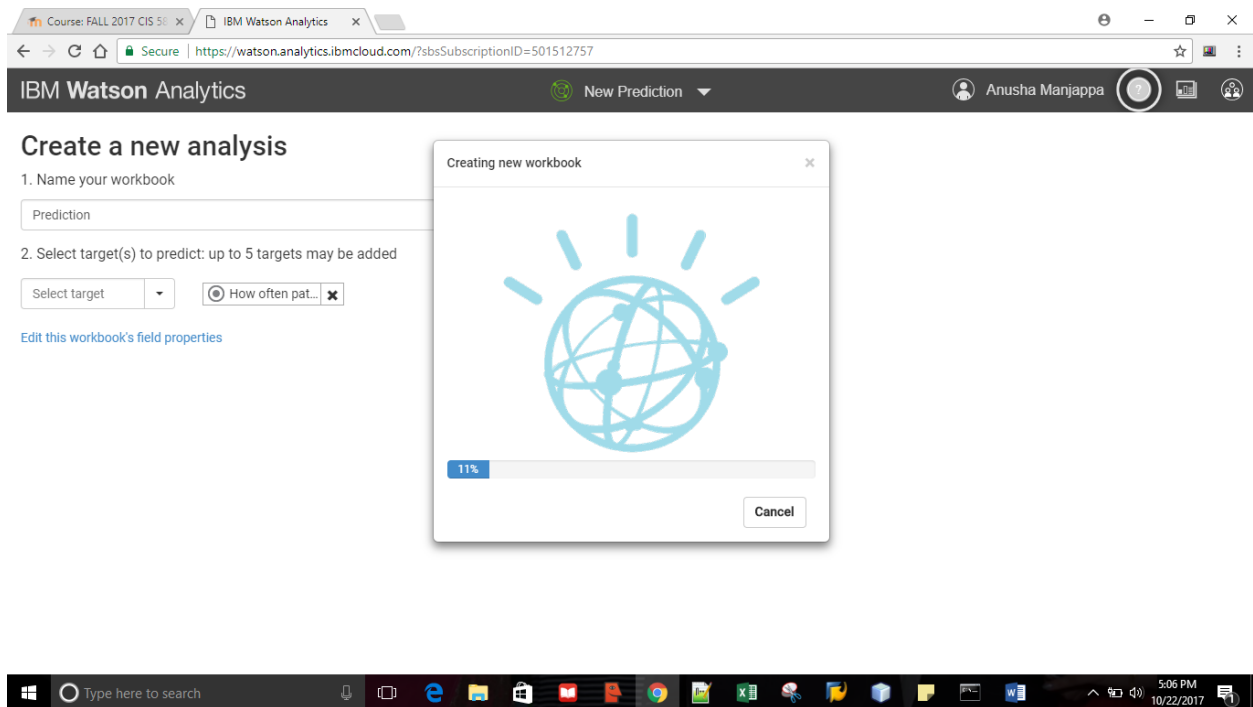
F)Prediction:

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

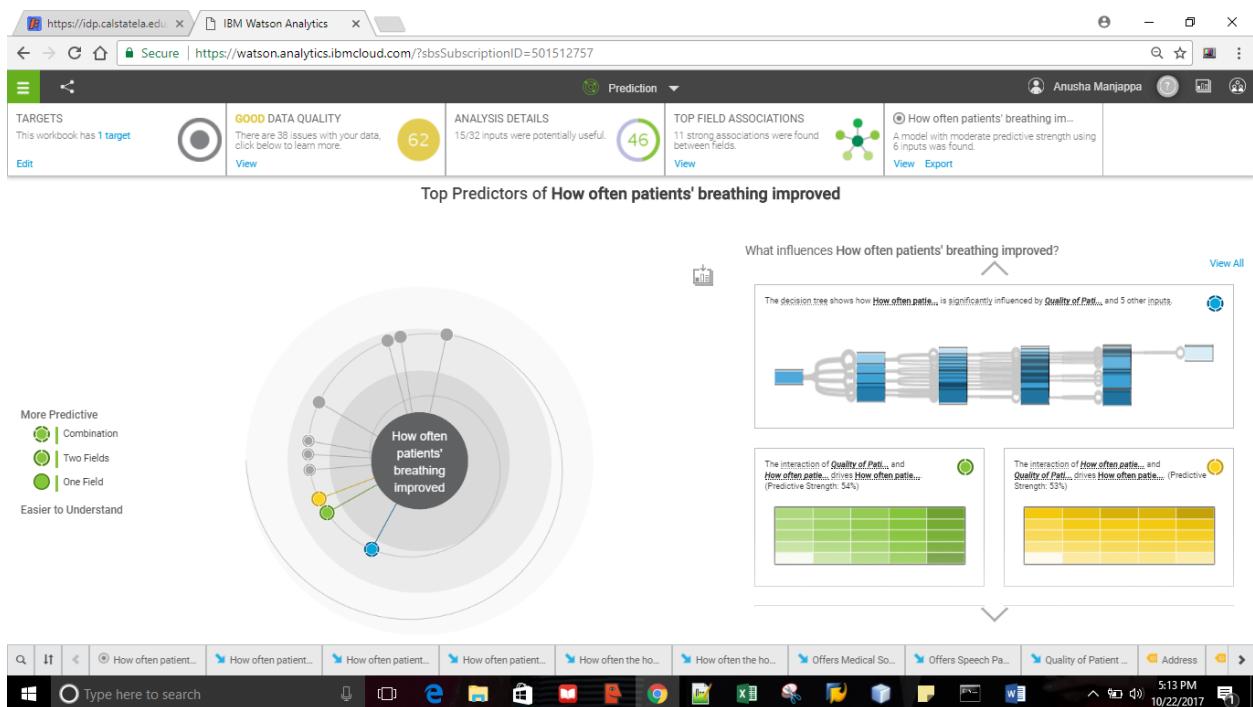
range of different predictions from the factors which gave the highest prediction to the lowest prediction. The screenshots of the detailed steps are given below.



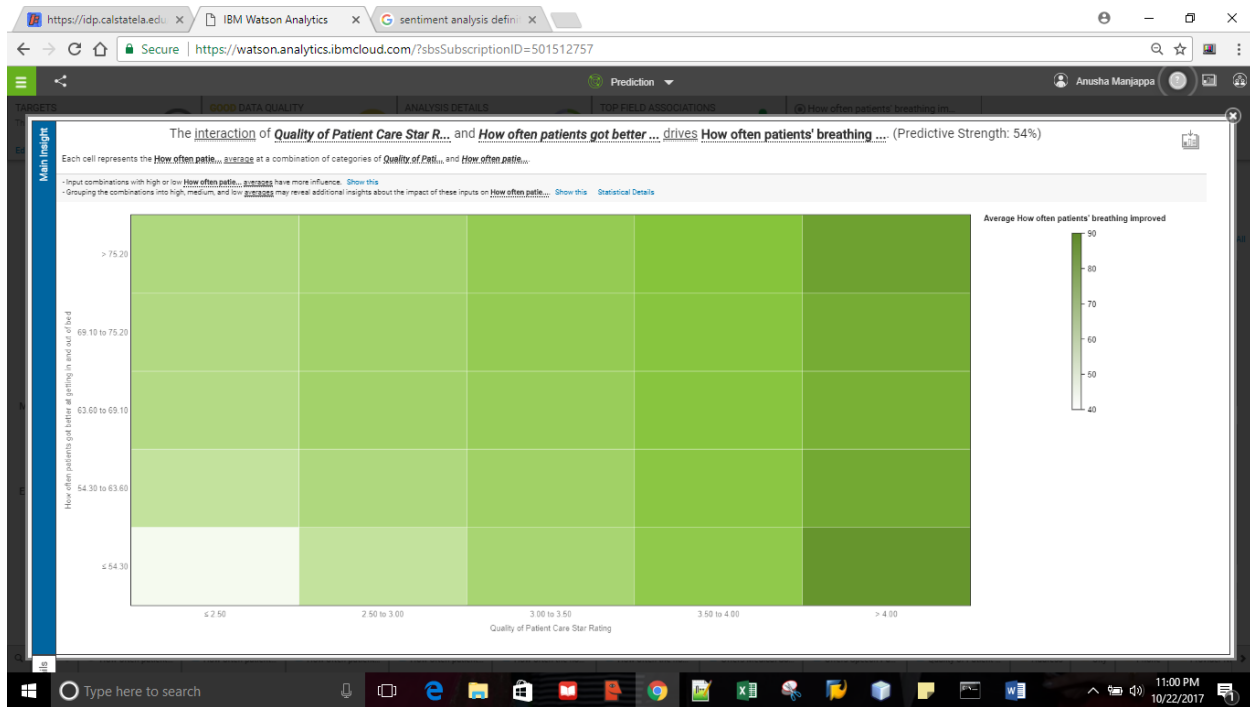
The above screenshot shows the process of selecting the input fields to predict a target dataset field.



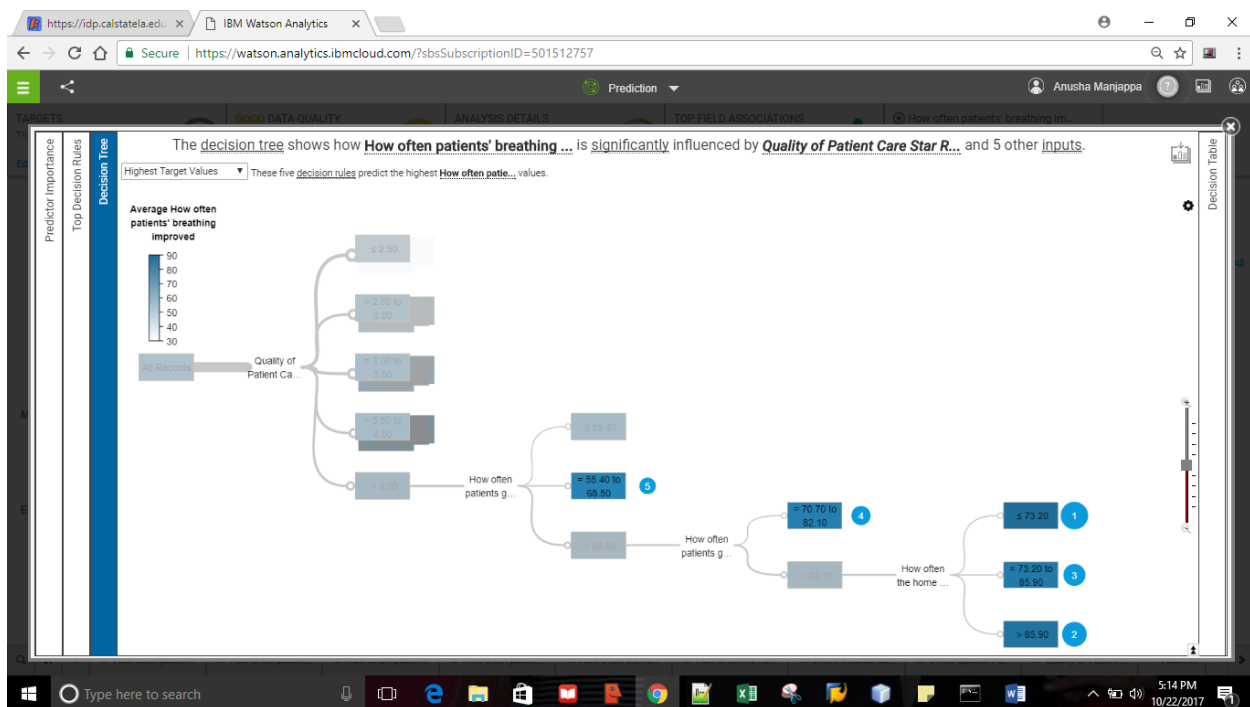
This screenshot shows the process of creating the prediction after selecting the appropriate input and target fields.



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 screenshot displays the IBM Watson Analytics interface. The main panel shows a **Decision Table** for the target variable **How often patients' breathing ...**. The table is a 6x6 grid of orange cells, representing different ranges of input values. The first column is labeled 'Quality of Pati...' and the first row is labeled 'How often pati...'. The table is titled 'How often patients' breathing ... is a continuous target, so a CHAID regression tree is used.' Below the table, a rule is shown: 'IF (Quality of Patient Care Star Rating = 3.00 to 3.50) AND (How often patients got better at taking their drugs correctly by mouth ≤ 46.50) THEN predicted How often patients' breathing improved = 66.56.'

Input	3.00 to 3.50	3.50 to 4.00	3.50 to 4.00	> 4.00	> 4.00	≤ 2.50
How often pati...						> 63.60
How often pati...	≤ 46.50	61.50 to 68.50	> 68.50	≤ 55.40	55.40 to 68.50	
How often pati...						≤ 62.80
How often pati...						
How often pati...						

IF (Quality of Patient Care Star Rating = 3.00 to 3.50) AND (How often patients got better at taking their drugs correctly by mouth ≤ 46.50) THEN predicted How often patients' breathing improved = 66.56.

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

The screenshot displays the IBM Watson Analytics interface. The main panel shows the **Top Decision Rules** for the target variable **How often patients' breathing ...**. The rules are listed in a table, showing the predicted value and the associated decision rules. The first rule has a predicted value of 88.81 (High) and the second rule has a predicted value of 87.43 (High).

How often patients' breathing improved	Decision Rules
88.81 High	Quality of Patient Care Star Rating > 4.00 How often patients got better at taking their drugs correctly by mouth > 68.50 How often patients got better at bathing > 82.10 How often the home health team determined whether patients received a flu shot for the current flu s ≤ 73.20
87.43 High	Quality of Patient Care Star Rating > 4.00 How often patients got better at taking their drugs correctly by mouth > 68.50 How often patients got better at bathing > 82.10 How often the home health team determined whether patients received a flu shot for the current flu s > 85.90
How often patients' breathing improved	Quality of Patient Care Star Rating > 4.00

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.

