HomeCare4Me

Choosing Your Home Health Care Agency

Final Report

March 1, 2016

HomeCare4Me proposes a recommendation engine

to assist consumers of Home Health Care Services

in easily selecting the best and most appropriate

CMS Certified Home Health Care Agency

to meet their individual needs.

Authors: Sandra Duenas, Principal Data Scientist and Project Lead (Founder)

Siddartha Sathyanarayan, Data Scientist and Project Manager

Capstone 498 – SEC 55 – Winter 2016 – Professor Don Wedding, Ph.D.

Contents

Executive Summary

Problem

Home Health Care is a wide set of health care services that can be rendered at the home of the patient for an illness or injury. Usually these services are less expensive, more convenient and effective than the services provided in a skilled nursing facility or hospital.

Normally, an attending physician or the primary care physician of the patient decides if the patient needs home health care based on existing health conditions or a recent surgery. The physician gives the patient a list of agencies that offer home health care services in the home-area of the patient and it is the responsibility of the patient to research the information regarding the agencies in order to make a selection for his or her specific needs.

The task of comparing the performance for quality of care of Home Health Care Agencies using information available can be difficult and confusing involving much research of information and comparison of results. This process can be overwhelming to patients and may result in additional stress with futile attempts to finding a highly rated Home Health Care Agency that would meet their needs.

This project attempts to alleviate this problem for home-bound patients by proposing the creation of a recommendation engine that provides easily available information of CMS Certified Home Health Agencies. The proposed information includes the historical mean of the Star Rating of Agencies, which is a rating based on nine Outcome and Process performance measures, an estimate of future performance for each Agency, and additional information such as the State, City, Services Offered, and Type of Ownership.

The easily available proposed rating information of each CMS certified Agency would allow patients to expedite their decision making when selecting the most appropriate Agency for their needs.

Approach

To meet the above challenge, the principal goal of this project is the development of a recommendation engine for the selection of CMS certified Home Health Care Agencies (aka HHCA or HHA or Provider) that uses public CMS data (3) and the development is performed using an open-source software, R, which is free of cost to the organization. The main objective for this recommendation engine is to provide patients the ability to *easily* select high performing HHCAs with respect to Patient Care Quality Measures (aka PCQM) that meet their individual home health care needs.

The Star Rating methodology utilized by CMS in the 'How the Ratings are Calculated' (5) is applied on 11-quarters of integrated historical public data from CMS Home Health Care Compare (1) by using R. The Data Preparation phase is primarily implemented using R and some of the following tasks are performed: a) standardization of the metadata from different time periods of the historical 11-quarters data set, b) standardization of the categorical values, c) imputation of missing numeric data, and d) removal of observations having Patient Care Quality Measures with values greater than 100% is performed. The original data set contains over 135,000 observations and it is reduced to 107,000 for the Star Scoring of the Providers at the Quarterly level using nine of the twenty four Patient Care Quality Measures. The data set is further reduced to about 19,000 observations at the Provider Name level when applying the historical mean Start Rating to each Provider.

The resulting prepared data set now has a derived target or response variable, Star Rating, for each Provider or HHCA that is in turned used in all predictive models fitted to identify explanatory or important predictors as well as estimates.

Several machine learning algorithms are applied to the data using R in order to identify explanatory variables or variables with predictive influence on the Star Rating.

The first algorithms applied are the Random Forest (RF) for Regression and also for Classification. These applications identify several predictors with explanatory variables, however, these techniques did not provide estimate coefficients for the predictors. Also, the application of these two techniques in R resulted in memory errors when the categorical variable for the Provider Name was included as part of the predictors because this variable has over 13,000 distinct values resulting in the Provider Name having to be excluded from the model fitting.

The Support Vector Machine (SVM) algorithm is applied to the data set using two different methods of the SVM algorithm, the linear method which uses linear lines and the radial method which uses radial lines. Again, these two applications result in a list of important variables without estimates, excluding the Provider Name in the model fitting.

The Lasso is applied to the data set. This technique uses a "*tuning parameter*" that controls the impact of a "*shrinkage penalty*" resulting in a list of important or explanatory variables as well as predictor estimates. This is the only algorithm of all the algorithms applied which is able to fit a model using the Provider Name variable which has over 13,000 categories as well as all of the other categorical and numerical variables resulting in a total of 13,856 variables in the model fitting.

The Lasso application enables the ability to include performance estimates for each Provider or HHCA in the recommendation engine ensuring that the principal goal of this project is accomplished.

Best practices to the application of predictive analytics is adhered to by the utilization of a train data set to fit the models and a test data set to make the predictions and evaluate the models. The evaluation of the residuals or differences between the Observed and Predicted Star Ratings for the observations is performed as well as thorough documentation of the analysis from the results is provided.

Tableau is used to provide fast and insightful views into the data, including data hierarchy, prevalence of some categories, and nested distribution in multiple categorical variables.

A team of two members met weekly during most of the project time and tasks and due dates were agreed on Sunday to be delivered throughout the week. The approach to Project Management is Agile and the team followed the CRISP-DM (8) methodology to structure project deliverables and sequence.

Conclusion

The approach delineated above successfully demonstrates that the development of a recommendation engine using CMS public data and an open-source software, R, is possible.

The project concluded by completing the nine goals it set out to complete. The framework of the recommendation engine has been successfully created and home-bound patients will be able to easily select a Home Health Care Agency certified by CMS in their home State and City with the knowledge of the Agency's historical (11 calendar quarters) Star Rating and estimate for future performance.

The decision to use R as the development platform and Tableau as the Visualization tool proved to be very productive and cost-free. R provides granular information of model fitting results in terms of descriptive statistics, statistical significance, predictor coefficient, residual results, and metadata structure of the objects developed for data preparation and for predictive models that is truly unequal with other Statistical Modeling tools, such as IBM SPSS Modeler and SAS, which tools are also very expensive to purchase.

Tableau provided a quick and beautiful way to visualize the data providing meaningful insights into aspects of the data that would have been very time consuming to perform in R.

Additional factors of success that allowed the accomplishment all of the goals proposed during the inception of this project is the utilization of the CRISP-DM (8) methodology, which ensures that the correct tasks and deliverables are performed in the correct sequence. Also, the Agile Project Management approach is implemented allowing the team to produce small deliverables for quick wins and re-work some aspects of the project as required.

Recommendation

Based on the successful accomplishment of the nine goals this project proposed, the low cost of the development platform, and the expected high value to home-bound patients of the information that would be provided by the recommendation engine, our recommendation is to adopt the findings of this project for a Production release, test the production results with a control group of home-bound patients to ensure its accuracy, and then, if all proves successful again, create a marketing campaign to promote this recommendation engine as a new data product.

Introduction (Problem Statement)

Home Health Care is a wide set of health care services that can be given in a patient's home for an illness or injury. Usually these services are less expensive, more convenient and effective as the care you receive in a skilled nursing facility or hospital. Examples of home health services include Nursing, Physical Therapy, Occupational Therapy, Speech Pathology, Medical and Social Services, and Home Health Aide Services.

During routine patient care, the primary care doctor or referring health care provider decides if the patient needs home health care and provides the patient with a list of agencies that offer home health care services in their area.

The US Government agency Center for Medicare and Medicaid Services or CMS maintains an online tool or website called *Home Health Compare* (1) which provides information about the quality of care provided by "CMS-certified" home health agencies throughout the nation. "CMS-certified" means the home health agency is approved by CMS and meets certain federal health and safety requirements.

The *Home Health Compare* online tool can help patients and their family or friends choose a home health agency in their geographical location that has the specific home health services that meets their health care needs.

Home Health Compare provides information for each home health agency, such as

- Agency name, address, and phone number
- Agency's initial date of Medicare certification

- Type of ownership, such as proprietary, voluntary, non-profit, private, government, local, state, county
- Services offered, such as nursing care, physical therapy, occupational therapy, speech therapy, medical and social services, and home health aide services
- A quality of patient care star rating measure that summarizes selected information about the performance of each home health agency compared to other agencies (only added for third quarter of 2015)
- Information on 24 quality of patient care measures for each home health agency measuring performance with respect to following best practices during patient care (2 of these measure were added on the third quarter of 2015)
- information from patients about experiences with the home health agency (patient survey results)

It is the responsibility of the patient to research the information using the Home Health Compare online tool regarding the agencies that are part of the list provided by their primary physician in order to make a selection for his/her specific needs. Comparing the Home Health Care Agencies based on the data available from this online tool can be a tedious task and involves combing through multiple columns spread across many spreadsheets.

In this project, HomeCare4Me proposes the development of a recommendation engine that automatically analyzes these data to create a score for each Agency that is based on nine patient care quality performance measures that are standard to all Agencies. The Agency Score would enable the patient to *easily* make a selection of the highest scored Agency that meets his or her geographical and service needs.

1. Project Goals and Deliverables

The fundamental goal of this project is a **proof of concept** is to demonstrate a low cost and simple platform that leverages public data from CMS and uses an open source software, R, to create predictive models that score home health care agencies based on patient care quality measures. An aspect of the proof of concept is to develop internal capabilities in the application of data mining and predictive analytics techniques in order to build the recommendation engine.

As a **proof of value**, we want to demonstrate how these public data and open source tool can be used to help individuals when making important decisions regarding their home health care and show that predictive analytics adds great value to the wellbeing of individuals.

1.1 Specific goals for *Proof of Value*

1.1.1. Leverage public CMS home health care data to identify attributes that distinguish high performing Home Health Care Agencies from poor performing agencies. These attributes will be used by the recommendation engine to score the Agencies.

1.1.2. Identify natural groupings or segments of Home Health Care Agencies which can be used by the recommendation engine to guide the consumer in selecting the appropriate Home Health Care Agency for their individual needs.

1.1.3. Provide a list of Home Health Care agencies with their Quality of Care Score based on the geographical location of a patient and the type of service desired.

1.1.4. Provide a Quality Performance Measure **Score** for the Home Health Care Agency selected by the patient.

1.2 Specific goals for the *Proof of Concept*

1.2.1. Evaluate R as a tool to the prepare data, such as combining different data sets, imputing missing values, standardizing metadata data across historical files, standardizing categorical values across historical data, filtering records, and dichotomizing categorical variables.

1.2.2. Evaluate R as a tool to apply initial EDA for descriptive statistics, correlation, reducing the number of variables, data mining techniques to find insights in the data.

1.2.3. Evaluate R as a tool to fit predictive models and evaluate their accuracy.

1.2.4. Evaluate R as a tool to deploy the predictive model to production to score new observations.

1.2.5. Evaluate Tableau as a visualization tool to explain insights about the data.

1.3 Deliverables

The Deliverable for each of the Goals listed above can be found in Table 1 below.

Table 1 - Deliverables

Goal Number	High Level Goal	Goal	Deliverables
1	**Proof of Value**	***Identify attributes*** that distinguish high performing Home Health Care Agencies from poor performing agencies by leverage public CMS home health care data to . These attributes will be used by the recommendation engine to score the Agencies.	***Set of variables*** that are important predictors to identify high performing Agencies from low performing Agencies.
2	**Proof of Value**	***Identify natural groupings or segments*** of Home Health Care Agencies which can be used by the recommendation engine to guide the consumer in selecting the appropriate Home Health Care Agency for their individual needs.	***List of segments or groupings*** for high performing Agencies versus low performing Agencies.
3	**Proof of Value**	***Provide a list of Home Health Care agencies*** with their Quality of Care Score based on the geographical location of a patient and the type of service desired.	***List of Home Health Care agencies*** with their Quality of Care Score based on the geographical location of a patient and the type of service desired.
4	**Proof of Value**	***Provide a Quality Performance Measure* Score** for the Home Health Care Agency selected by the patient.	***List of providers*** and their respective Quality Performance Score
5	**Proof of Concept**	***Evaluate R as a tool to the prepare data***, such as combining different data sets, imputing missing values, standardizing metadata data across historical files, standardizing categorical values across historical data, imputing numeric and character data, filtering records, and dichotomizing categorical variables.	***Clean and conformed data set of historical data*** and is ready for modeling by imputing missing values, removing variables or observations not needed, dichotomizing categorical variables.
6	**Proof of Concept**	***Evaluate R as a tool to apply initial EDA for descriptive statistics***, correlation, reducing the number of variables, data mining techniques to find insights in the data.	***Insights*** obtained from the EDA using R
7	**Proof of Concept**	***Evaluate R as a tool to fit predictive models*** and evaluate their accuracy.	***Predictive models*** and their accuracy based on training and test data sets in order to select the best model.
8	**Proof of Concept**	***Evaluate R as a tool to deploy the predictive model*** to production to score new observations.	***Instructional steps to deploy a predictive*** or scoring model in production using R.
9	**Proof of Concept**	***Evaluate Tableau as a visualization*** tool to explain insights about the data.	***Visualizations the generate insights into the historical data*** being used, descriptive statistics of the data, relationship among variables, dependencies among predictors and variable to be predicted.

2. Data Description

2.1 Data Overview:

The analysis presented in this paper is based on 11 quarterly historical data files from CMS containing aggregated observations for Episodes of Care for home health care agencies. Each observation represents "*home health episodes of care in which the start or resumption of care date was either on the physician-specified date or within 2 days of the referral date or inpatient discharge date, whichever is later*." (2) NOTE: the acronyms of HHCA, HHA are equivalent mean a "Provider".

Each observation includes 24 Patient Care Quality Measures and 12 categorical variables representing clinical and process services provided by an agency to a patient at their home as well as agency descriptors, such as the legal organization and geographical location.

The data also contains twenty six fields for the Footnotes of each measure and general text informational data explaining why a measure is missing. All of the 'Footnote' fields were removed from the data set to be used for the analysis.

The historical data is collected by sourcing or downloading the data from the CMS website (2), a total of 11 files containing data from the first calendar quarter of 2013 through the third calendar quarter of 2015. The files are comma delimited in MS-DOS CVS format with row heading.

The Integration of the data from these 11 quarters resulted in 135,464 observations and it this task has proven challenging due to the differences of the variables names and categorical values in the 2013 files versus the 2014 and 2015 files. A new variable was derived called year.quarter and populated with the Year and the Quarter of the data file from which the observations were sourced. This newly derived variable helps identify the original data file source for each observation in the final integrated data set.

R was used to standardize the data across all 11 data files and Tableau was used to visualize the results of the data standardization and quickly determine whether the related categorical values and variable names were the same across all 11 files. The standardization of the variable names and of the categories is important because it would provide consistent historical data on to which the analysis can be correctly performed. This is also important for the application of machine learning algorithms to be able to correctly compute the coefficients of the Patient Care Quality Measures on the same variables across the historical data set.

2.2 Numerical Variables

There are 24 Patient Care Quality Measures (PCQM), each measure represents the percentage or portion of patients that accomplished some clinical goal or the percentage that a quality care process was performed out of all of the Episodes of Care counted for a given PCQM for a quarter.

For example, the measure "**Improvement in Bed Transferring**" is measured as "*Percentage of home health episodes of care during which the patient improved in ability to get in and out of bed.*" The numerator is the "Number of home health episodes of care where the value recorded on the discharge assessment indicates less impairment in bed transferring at discharge than at start (or resumption) of care." and the denominator is the "Number of home health episodes of care ending with a discharge during the reporting period, other than those covered by generic or measure-specific exclusions."

2.3 Categorical Variables

The data contains 12 categorical variables as shown in Table 2 below. The dichotomization of these 12 variables would result in 22,363 variables as shown in Table 2 below. However, further analysis shows that some of these categorical variables do not need to be dichotomized, such as Provider Name, because the segmentation of data by Provider Name is too narrow to provide any useful insights into the performance of Agencies.

Table 2 – Categorical Variables and resulting number of Dummy variables.

Category #	Categorical Variable	Number of Categories
1	State	55
2	Provider Name	13,119
3	City	3,293
4	Type of Ownership	12
5	Offers Nursing Care Services	2
6	Offers Physical Therapy Services	2
7	Offers Occupational Therapy Services	2
8	Offers Speech Pathology Services	2
9	Offers Medical Social Services	2
10	Offers Home Health Aide Services	2
11	year quarter	11
12	zip	5,861
	Total dummy or dichotomized variables	**22,363**

The data set contains six (6) indicators (categorical variables) representing the type of clinical service the agency offers, these are shown in Table 2 above as Category #s 5 through 10. These categories are used for segmenting HHCAs to provide insights of the data.

Another descriptor of the home health care agency is the Type of Ownership (legal organization type) shown as #4 in Table 2 above. The 12 categories for this variable, listed below, are also used to segment the HHCAs and provide further insights into the data.

1. combination govt and voluntary
2. govt - local
3. govt - state or county
4. local
5. other
6. private
7. proprietary
8. religious affiliations
9. state or county
10. voluntary non-profit - church
11. voluntary non-profit - other
12. voluntary non-profit - private

Each agency has its address and the name also listed. The State and City are used to perform EDA and segmentation of agencies as one way to gain more understanding of the data.

3. Data Preparation

The data preparation was performed entirely using R. In general, data from 11 files were integrated, variable names were standardized, categories were standardized, missing values imputed, removal of observations with inappropriate values for project objectives were removed, binning and scoring was performed to create a target or class variable for the data set in order to enable the creation of predictive models required for the recommendation engine.

3.1 Imputation of Missing Data

The analysis for missing data found two variables having close to 50% missing data. These two variables are shown in Table 3 below highlighted in yellow, "Heart Failure Symptoms Addressed during All Episodes of Care" and" Improvement in Status of Surgical Wounds" .

After close inspection of the observations missing these two measures by looking at the comments in the Footnote fields of the original data set, which state the following '*The number of patient episodes for this measure is too small to report.*', and '*This measure currently does not have data or has less than 6 months of data*', it was decided to remove these two measures from the data set, however, their observations remained.

The other numerical measures with lower rates of missing values are due to the same reasons as explained in the paragraph above, which indicates that it would be appropriate to impute the numerical data with zero (0) value as most of the observations have a percentage > 0.

Table 3 – Numerical Variables and resulting Descriptive Statistics on Original data.

	VAR.NAME	NOT.NA	IS.NA	PCT.NA	MIN	Q01	Q05	Q25	Q50	Q75	Q95	Q99	MAX	MEAN	SD
1	HO.HHT.began.care.in.timely.manner.18	121,741	13,723	10%	1	56	73	88	93	97	199	201	201	97.0684	27.6503
2	HO.HHT.taught.about.their.drugs.20	120,678	14,786	11%	0	45	71	90	96	99	199	201	201	101.6558	33.2385
3	HO.HHT.checked.for.risk.of.falling.22	120,366	15,098	11%	0	77	90	98	100	100	199	201	201	105.3006	26.8837
4	HO.HHT.checked.for.depression.24	121,687	13,777	10%	0	45	84	97	99	100	199	201	201	102.3482	26.7380
5	HO.HHT.ensured.received.flu.shot.26	118,854	16,610	12%	0	4	30	63	76	87	199	201	201	80.2265	39.4818
6	HO.HHT.made.received.pneumonia.shot.28	121,492	13,972	10%	0	2	19	58	76	88	199	201	201	76.6435	39.0089
7	HO.HHT.taught.gave.foot.care.30	111,141	24,323	18%	0	53	77	93	98	100	199	201	201	110.7490	40.1531
8	HO.HHT.checked.for.pain.32	121,829	13,635	10%	0	79	92	98	99	100	199	201	201	104.2345	24.7259
9	HO.HHT.treated.for.pain.34	119,393	16,071	12%	0	77	92	98	100	100	199	201	201	108.3757	31.2751
10	HO.HHT.treated.heart.failure.weakening.of.the.heart.36	77,248	58,216	43%	6	87	94	99	100	199	199	201	201	145.4735	50.6172
11	HO.HHT.took.action.to.prevent.pressure.bed.sores.38	109,228	26,236	19%	0	50	81	96	99	100	199	201	201	113.3620	40.8161
12	HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40	111,941	23,523	17%	0	52	83	97	100	100	199	201	201	108.5746	34.9494
13	HO.HHT.checked.for.risk.of.pressure.bed.sores.42	121,829	13,635	10%	0	62	86	97	99	100	199	201	201	102.8613	25.6938
14	HO.PAT.got.better.at.moving.around.44	115,325	20,139	15%	0	26	37	52	60	68	199	201	201	71.6297	43.1194
15	HO.PAT.got.better.at.getting.in.and.out.of.bed.46	114,098	21,366	16%	0	18	29	46	56	66	199	201	201	68.3534	46.4123
16	HO.PAT..got.better.at.bathing.48	115,737	19,727	15%	0	23	39	58	67	76	199	201	201	76.6668	41.4677
17	HO.PAT.had.less.pain.when.moving.around.50	114,838	20,626	15%	0	15	36	58	68	81	199	201	201	78.4402	42.9034
18	HO.PAT.breathing.improved.52	113,299	22,165	16%	0	7	24	53	66	76	199	201	201	74.8471	46.4745
19	HO.PAT.wounds.improved.healed.54	80,591	54,873	41%	9	70	80	90	96	199	199	201	201	131.1675	52.9912
20	HO.PAT.got.better.at.taking.drugs.correctly.56	112,667	22,797	17%	0	8	22	41	50	60	199	201	201	64.2783	49.9402
21	HO.PAT.needed.urgent.unplanned.ER.wout.admission.58	113,475	21,989	16%	0	3	6	9	12	15	199	201	201	31.1629	57.3199
22	HO.PAT.had.to.be.admitted.to.the.hospital.60	113,475	21,989	16%	0	6	9	13	16	19	199	201	201	34.6353	56.1485

3.2 Standardization of Variable Names

During the integration of 11 files containing historical data for 11 calendar quarters, it was discovered that the variable names in the files for the year 2013 did not all match the variable names to their corresponding variable names in the files for the years 2014 and 2015. Refer to Table 4 below, Measure #7 and #21 in which the variable name for the year 2013 is written different than their corresponding variable names in years 2014 and 2015.

Also, the 3rd quarter for year 2015 introduced two new variables, #23 and #24 shown in Table 4 below.

The variable names across the 11 files were standardized to the names shown the "Target Measure Name and Number" in blue font in Table 4 below.

Table 4 – Variable Name Standardization

Measure #	2013 all Quarters Measures Variable Name in data file	2014 all Quarters and 2015 Q1 and Q2 Measures Variable Name in data file	2015 Q3 Variable Name in data file	Target Measure Name and Number
1	How often the home health team began their patients' care in a timely manner.	How often the home health team began their patients' care in a timely manner	How often the home health team began their patients' care in a timely manner	"HO.HHT.began.care.in.timely.manner.18",
2	How often the home health team taught patients (or their family caregivers) about their drugs.	How often the home health team taught patients (or their family caregivers) about their drugs	How often the home health team taught patients (or their family caregivers) about their drugs	"HO.HHT.taught.about.their.drugs.20",
3	How often the home health team checked patients' risk of falling.	How often the home health team checked patients' risk of falling	How often the home health team checked patients' risk of falling	"HO.HHT.checked.for.risk.of.falling.22",
4	How often the home health team checked patients for depression.	How often the home health team checked patients for depression	How often the home health team checked patients for depression	"HO.HHT.checked.for.depression.24",
5	How often the home health team determined whether patients received a flu shot for the current flu season.	How often the home health team determined whether patients received a flu shot for the current flu season	How often the home health team made sure that their patients have received a flu shot for the current flu season.	"HO.HHT.ensured.received.flu.shot.26",
6	How often the home health team determined whether their patients received a pneumococcal vaccine (pneumonia shot).	How often the home health team determined whether their patients received a pneumococcal vaccine (pneumonia shot)	How often the home health team made sure that their patients have received a pneumococcal vaccine (pneumonia shot).	"HO.HHT.made.received.pneumonia.shot.28",
7	For patients with diabetes, how often the home health team got doctor's orders, gave foot care, and taught patients about foot care.	With diabetes, how often the home health team got doctor's orders, gave foot care, and taught patients about foot care	With diabetes, how often the home health team got doctor's orders, gave foot care, and taught patients about foot care	"HO.HHT.taught.gave.foot.care.30",
8	How often the home health team checked patients for pain.	How often the home health team checked patients for pain	How often the home health team checked patients for pain	"HO.HHT.checked.for.pain.32",
9	How often the home health team treated their patients' pain.	How often the home health team treated their patients' pain	How often the home health team treated their patients' pain	"HO.HHT.treated.for.pain.34",
10	How often the home health team treated heart failure (weakening of the heart) patients' symptoms.	How often the home health team treated heart failure (weakening of the heart) patients' symptoms	How often the home health team treated heart failure (weakening of the heart) patients' symptoms	"HO.HHT.treated.heart.failure.weakening.of.the.heart.36",
11	How often the home health team took doctor-ordered action to prevent pressure sores (bed sores).	How often the home health team took doctor-ordered action to prevent pressure sores (bed sores)	How often the home health team took doctor-ordered action to prevent pressure sores (bed sores)	"HO.HHT.took.action.to.prevent.pressure.bed.sores.38",
12	How often the home health team included treatments to prevent pressure sores (bed sores) in the plan of care.	How often the home health team included treatments to prevent pressure sores (bed sores) in the plan of care	How often the home health team included treatments to prevent pressure sores (bed sores) in the plan of care	"HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40",
13	How often the home health team checked patients for the risk of developing pressure sores (bed sores).	How often the home health team checked patients for the risk of developing pressure sores (bed sores)	How often the home health team checked patients for the risk of developing pressure sores (bed sores)	"HO.HHT.checked.for.risk.of.pressure.bed.sores.42",
14	How often patients got better at walking or moving around.	How often patients got better at walking or moving around	How often patients got better at walking or moving around	"HO.PAT.got.better.at.moving.around.44",
15	How often patients got better at getting in and out of bed.	How often patients got better at getting in and out of bed	How often patients got better at getting in and out of bed	"HO.PAT.got.better.at.getting.in.and.out.of.bed.46",
16	How often patients got better at bathing.	How often patients got better at bathing	How often patients got better at bathing	"HO.PAT..got.better.at.bathing.48",
17	How often patients had less pain when moving around	How often patients had less pain when moving around	How often patients had less pain when moving around	"HO.PAT.had.less.pain.when.moving.around.50",
18	How often patients' breathing improved.	How often patients' breathing improved	How often patients' breathing improved	"HO.PAT.breathing.improved.52",
19	How often patients' wounds improved or healed after an operation.	How often patients' wounds improved or healed after an operation	How often patients' wounds improved or healed after an operation	"HO.PAT.wounds.improved.healed.54",
20	How often patients got better at taking their drugs correctly by mouth.	How often patients got better at taking their drugs correctly by mouth	How often patients got better at taking their drugs correctly by mouth	"HO.PAT.got.better.at.taking.drugs.correctly.56",
21	How often patients receiving home health care needed any urgent, unplanned care in the hospital emergency room – without being admitted to the hospital.	How often patients receiving home health care needed urgent, unplanned care in the ER without being admitted	How often patients receiving home health care needed urgent, unplanned care in the ER without being admitted	"HO.PAT.needed.urgent.unplanned.ER.wout.admission.58",
22	How often home health patients had to be admitted to the hospital	How often home health patients had to be admitted to the hospital	How often home health patients had to be admitted to the hospital	"HO.PAT.had.to.be.admitted.to.the.hospital.60",
23			How often home health patients, who have had a recent hospital stay, had to be re-admitted to the hospital	"HO.PAT.was.re.admitted.to.hospital.62",
24			How often home health patients, who have had a recent hospital stay, received care in the hospital emergency room without being re-admitted to the hospital	"HO.PAT.received.ER.care.wout.re.admission.64",

The variables that were found to have historical data across all 11 files are listed in Table 5 below. Only 22 measures of 24 will remain due to the two measures (#23 and #24 in Table 4 above) being introduced just in the last quarter. These 22 measures are shown in green font in Table 5 below.

The variables in blue font are the categorical variables which will be dichotomized.

The variable in red font, CMS.Cer.Number.CCN, is the identifier of the Home Health Care Agency. The variables in black font will not be used for the analysis.

Table 5 - Common Variables across 11 historical data files

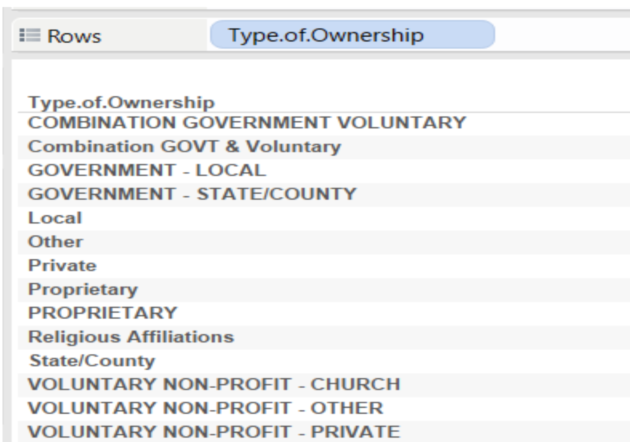
Field #	Common Fields across 11 data files
1	"Address",
2	"City",
3	"CMS.Cer.Number.CCN",
4	"Date.Certified",
5	"HO.HHT.began.care.in.timely.manner.18",
6	"HO.HHT.checked.for.depression.24",
7	"HO.HHT.checked.for.pain.32",
8	"HO.HHT.checked.for.risk.of.falling.22",
9	"HO.HHT.checked.for.risk.of.pressure.bed.sores.42",
10	"HO.HHT.ensured.received.flu.shot.26",
11	"HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40",
12	"HO.HHT.made.received.pneumonia.shot.28",
13	"HO.HHT.taught.about.their.drugs.20",
14	"HO.HHT.taught.gave.foot.care.30",
15	"HO.HHT.took.action.to.prevent.pressure.bed.sores.38",
16	"HO.HHT.treated.for.pain.34",
17	"HO.HHT.treated.heart.failure.weakening.of.the.heart.36",
18	"HO.PAT.breathing.improved.52",
19	"HO.PAT.got.better.at.bathing.48",
20	"HO.PAT.got.better.at.getting.in.and.out.of.bed.46",
21	"HO.PAT.got.better.at.moving.around.44",
22	"HO.PAT.got.better.at.taking.drugs.correctly.56",
23	"HO.PAT.had.less.pain.when.moving.around.50",
24	"HO.PAT.had.to.be.admitted.to.the.hospital.60",
25	"HO.PAT.needed.urgent.unplanned.ER.wout.admission.58",
26	"HO.PAT.wounds.improved.healed.54",
27	"Offers.Home.Health.Aide.Services",
28	"Offers.Medical.Social.Services",
29	"Offers.Nursing.Care.Services",
30	"Offers.Occupational.Therapy.Services",
31	"Offers.Physical.Therapy.Services",
32	"Offers.Speech.Pathology.Services",
33	"Phone",
34	"Provider.Name",
35	"State",
36	"Type.of.Ownership",
37	"Zip",

3.3 Standardization of Categorical Values

During the integration of the historical data files, it was also found that the values or categories for many of the categorical variables were not the same across the 11 data files.

An example of this standardization is shown below in Figures 1 and 2 where the categories of the variable 'Type of Ownership' had to be standardized so that the analysis can be more accurate.

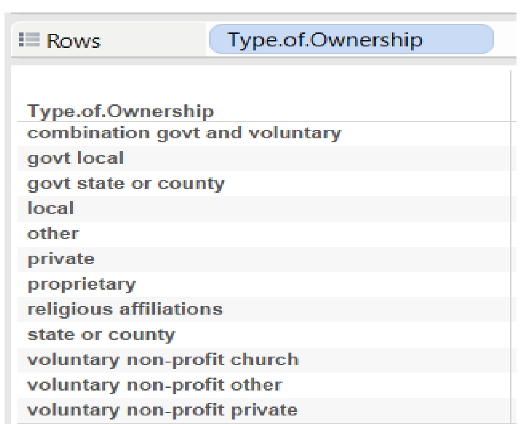
Figure 1 - Categories PRIOR to standardization



Rows	Type.of.Ownership

Type.of.Ownership
COMBINATION GOVERNMENT VOLUNTARY
Combination GOVT & Voluntary
GOVERNMENT - LOCAL
GOVERNMENT - STATE/COUNTY
Local
Other
Private
Proprietary
PROPRIETARY
Religious Affiliations
State/County
VOLUNTARY NON-PROFIT - CHURCH
VOLUNTARY NON-PROFIT - OTHER
VOLUNTARY NON-PROFIT - PRIVATE

Figure 2 - Categories AFTER to standardization



Rows	Type.of.Ownership

Type.of.Ownership
combination govt and voluntary
govt local
govt state or county
local
other
private
proprietary
religious affiliations
state or county
voluntary non-profit church
voluntary non-profit other
voluntary non-profit private

3.4 Dichotomization of Categorical Variables

There are 9 categorical variables, shown in yellow highlight in Table 6 below, that were dichotomized in R (6). The 10th variable, year.quarter was derived to indicate the quarterly data file from which the record was sourced.

The dichotomization of the variables resulted in 3,382 variables which converted the data set into Big Data. Variable reduction will be performed using the machine learning algorithm Random Forest in order to identify important variables for the Recommendation Engine.

The Provider Name and the Zip Code were not dichotomized for two reasons, R kept running out of memory and also, the categories in these two variables have a very limited coverage of the data to convey a trend or any other insights except their own small data space.

Table 6 - Categorical Variables – yellow highlights represent variables that were dichotomized.

Category #	Categorical Variable	Number of Categories
1	State	55
2	Provider.Name	13,119
3	City	3,293
4	Type.of.Ownership	12
5	Offers.Nursing.Care.Services	1
6	Offers.Physical.Therapy.Services	2
7	Offers.Occupational.Therapy.Services	2
8	Offers.Speech.Pathology.Services	2
9	Offers.Medical.Social.Services	2
10	Offers.Home.Health.Aide.Services	2
11	year.quarter	11
12	Zip	5,861
	Total categories	**22,362**
	Total dummy or dichotomized variables	***3,382***

3.5 Removal of Observations with any of the Nine Ranking-Based Measures > 100%

The quality patient care measures are in percentage thus the correct values should be between 0 and 100; however, the original data from CMS has some observations with values in the measure between 101 and 201. Refer to columns Q95, Q99, and MAX in Table 3 shown in the Numerical Variables section above. For the nine measures that are the basis of the HHCA Scoring, the observations having a value > 100 for anyone of these nine measures were removed from the data set. These nine measures are:

1. Timely Initiation of Care
2. Drug Education on All Medications Provided to Patient/Caregiver during All Episodes of Care
3. Influenza Immunization Received for Current Flu Season
4. Improvement in Bathing
5. Improvement in Bed Transferring
6. Improvement in Ambulation- Locomotion
7. Improvement in Dyspnea
8. Improvement in Pain Interfering with Activity
9. Acute Care Hospitalization (Claims based)

The historical data set has 135,464 rows, in this data preparation step, there were 15,238 observations removed, resulting in a working data set of 120,226 observations. This was an 11.24% decrease in the number of observations in the data set due to any of these nine measures having a value > 100.

The working data set from this data preparation step, ‘Incorrect Values’, is 120,226 observations.

3.6 Removal of Observations with all of the Nine Ranking-Based Measures = 0

Further preparation of the data with respect to the nine measures used for the HHCA Scoring was performed in order to ensure that at least one of these nine measures had a value greater than zero and thus the observation represents an Agency that participated in the application of these care management processes.

The data preparation required for the evaluation of the summation of the nine measures to be greater than zero. If the summation was greater than zero for an observation or HHCA, then that observation was preserved in the data set; otherwise, that observation was removed from the data set.

This data preparation step is based on the observations left in the ‘Incorrect Values’ step above that has 120,226 observations. This data preparation step resulted in the removal of 12,755 observations leaving a working data set of 107,471 observations with 3,409 numeric variables.

The 3,409 variables include 3,386 dichotomized variables, the CMS Certification ID for the HHCA, and 22 Patient Care Quality Measures of which 9 measures are the basis for the HHCA Scoring.

3.7 Bin the Nine Ranking-Based Measures into 10 Star Bins for each Observation

Beginning on the third quarter of 2015, CMS has begun to rate Home Health Care Agencies via a measure labeled the *‘HHA Quality Star Rating’*. T

The Star Rating is based on completed quality episodes and Claims data. In order for a HHCA to receive a rating, each HHCA must have data for at least 20 complete quality episodes and have reported on at least 5 of the 9 measures used in the calculation of the Star Rating within the 12 month reporting period. The methodology behind the Star Rating is documented in a letter to each provider which shows the rating and calculation behind it (5).

The first step to calculate the Star Rating is to bin the score received by the HHCA for each measure into bins. The range of values for each bin is documented in the 'CMS Methodology on Binning and Scoring' (5). For each of the 9 quality measures, all HHCAs' scores are sorted low to high and divided into 10 approximately equal sized groups (deciles) of HHAs. For all measures, except acute care hospitalization, a higher measure value means a better score. The cut points apply to all agencies and do not vary by agency and those exact same values are used in our calculation (5). A sample scorecard for one of the providers is shown in the figure 3 below.

Figure 3 - Home Health Care Star Rating Score Card

HHC Star Rating Scorecard[1]

		Measure Score Cut Points by Initial Decile Rating								
1	Initial Decile Rating	Timely initiation of care	Drug education on all medications	Received flu shot for current season	Improved walking or moving around	Improved getting in and out of bed	Improved bathing	Had less pain moving around	Breathing improved	Admitted to hospital
2	0.5	0.0-79.6	0.0-79.7	0.0-44.0	0.0-43.3	0.0-34.9	0.0-45.7	0.0-43.7	0.0-33.0	20.1-100.0
3	1.0	79.7-85.4	79.8-87.7	44.1-58.1	43.4-50.0	35.0-42.8	45.8-54.5	43.8-53.7	33.1-46.1	18.3-20.0
4	1.5	85.5-88.8	87.8-91.6	58.2-66.1	50.1-54.6	42.9-48.1	54.6-59.5	53.8-59.2	46.2-54.3	17.1-18.2
5	2.0	88.9-91.0	91.7-94.0	66.2-71.3	54.7-57.8	48.2-52.3	59.6-63.2	59.3-63.0	54.4-59.9	16.2-17.0
6	2.5	91.1-92.8	94.1-95.7	71.4-75.4	57.9-60.4	52.4-55.4	63.3-66.3	63.1-66.4	60.0-64.1	15.3-16.1
7	3.0	92.9-94.4	95.8-97.0	75.5-79.0	60.5-62.7	55.5-58.5	66.4-69.0	66.5-69.8	64.2-67.7	14.4-15.2
8	3.5	94.5-95.9	97.1-98.0	79.1-82.4	62.8-65.5	58.6-61.6	69.1-71.9	69.9-73.7	67.8-71.1	13.4-14.3
9	4.0	96.0-97.2	98.1-98.9	82.5-86.5	65.6-68.7	61.7-65.2	72.0-75.3	73.8-78.7	71.2-75.0	11.9-13.3
10	4.5	97.3-98.6	99.0-99.9	86.6-92.2	68.8-74.0	65.3-70.9	75.4-80.7	78.8-86.6	75.1-80.3	10.0-11.8
11	5.0	98.7-100.0	100.0-100.0	92.3-100.0	74.1-100.0	71.0-100.0	80.8-100.0	86.7-100.0	80.4-100.0	0.0-9.9
12	Your HHA Score	94.6	95.6	75.6	57.8	51.8	63.5	70.1	57.9	17.3
13	Your Initial Decile Rating	3.5	2.5	3.0	2.0	2.0	2.5	3.5	2.0	1.5
14	Your Number of Cases (N)	4,919	4,860	2,966	3,397	3,246	3,420	2,309	2,883	1,881
15	National (All HHA) Median	93.1	96.3	75.5	60.7	56.2	66.4	67.6	64.8	15.9
16	Your Statistical Test Probability Value (p-value)	0.000	0.264	0.397	0.080	0.000	0.000	0.062	0.000	0.010
17	Your Statistical Test Results (Is the p-value < 0.050?)	Yes	No	No	No	Yes	Yes	No	Yes	Yes
18	Your HHA Adjusted Rating	3.5	2.5	3.0	2.5	2.0	2.5	3.0	2.0	1.5

19	Your Average Adjusted Rating	2.5
20	Your Average Adjusted Rating Rounded	2.5
21	**Your Overall Star Rating (1.0 to 5.0)**	***** (3.0 stars)**

The HHCA's score on each measure is then assigned its decile location as a preliminary rating. Each decile is assigned an initial ranking from 0.5 to 5 in 0.5 increments. Rows 12 and 13, labeled "Your HHA Score" and "Your Initial Decile Rating" on the HHC Star Rating Scorecard shown in Figure 3 above, show the HHCA's score for each of the nine measures and the corresponding initial decile rating based on the score, respectively.

3.8 Score Agencies based on the Average Bin of the Nine Ranking-Based Measures

The initial decile rating assigned to each HHCA using the method described in section 3.7 above is subsequently adjusted according to a statistical test using the difference between the Agency's individual measure score and the national median score across all Agencies for that measure. (5)

The overall HHCA median score is shown in Row 15, labeled "National (All HHA) Median," on the HHC Star Rating Scorecard. The resulting probability value from the statistical test is shown in HHC Star Rating Scorecard Row 16, "Your Statistical Test Probability Value (p-value)". A probability value greater than 0.050 indicates that the HHCA is not significantly different from the overall national median (at a standard 5 percent significance level). (5)

On row 17, "Your Statistical Test Results", indicates "Yes" if the p-value is equal to or less than 0.050 and "No" if the p-value is greater than 0.050. If HHCA's initial decile rating for a measure is anything other than a 2.5 or 3 (the two middle decile categories), and the statistical test results show a p-value greater than 0.050 (indicating a "No" for being significantly different from the national median), the initial rating is adjusted to the next half star level closer to the middle categories of 2.5 or 3. The adjusted ratings are shown in Row 18 "Your HHA Adjusted Rating" on the HHA Star Rating Scorecard. (5)

To obtain an overall score for each HHCA, the adjusted ratings are averaged across the nine measures and rounded to the nearest 0.5. These results are shown in Row 19 "Your Average Adjusted Rating" and Row 20 "Your Average Adjusted Rating Rounded" on the HHCA Star Rating Scorecard shown in Figure 3 above.

An Overall HHCA Star Rating (Row 21) is then assigned to your HHCA incorporating an additional adjustment made so that ratings will range from 1.0 to 5.0 in half star increments (see Table 7 below). Thus, there are nine Star Ratings, with 3.0 stars being the middle category in this distribution.

Table 7 - Star Rating Scale

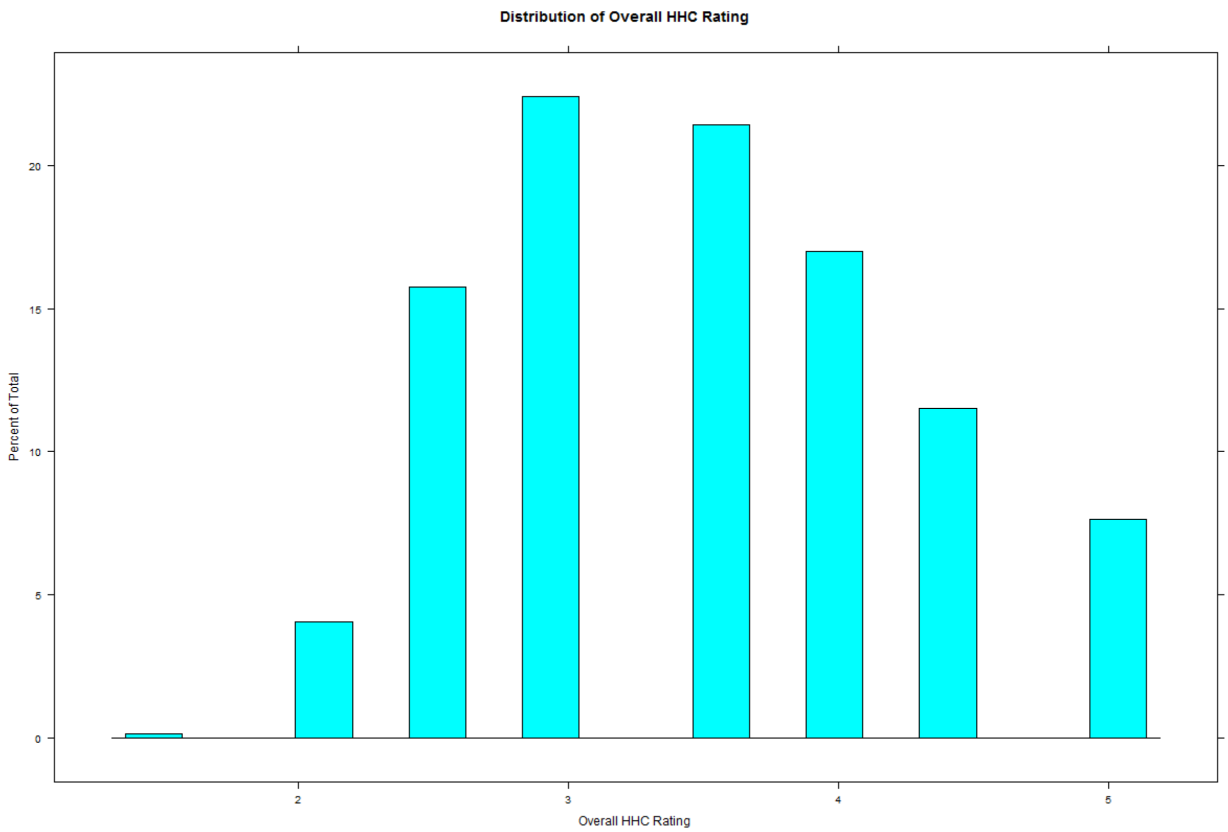
Average Adjusted Rating Rounded	Overall HHC Star Rating
4.5 and 5.0	(5.0)
4.0	(4.5)
3.5	(4.0)
3.0	(3.5)
2.5	(3.0)
2.0	(2.5)
1.5	(2.0)
1.0	(1.5)
0.5	(1.0)

After the HHCAs are assigned an Overall Star Rating, the distribution of ratings for the HHCAs is shown in Figure 4 below, the distribution bar chart. A summary of the distribution of Overall Star Rating for the HHACs is given in the Table 8 below. The median value is 3.5 and mean is 3.482 and are approximately equal when rounded off to one decimal point, which means the distribution can be assumed to be approximately symmetrical. The first and third quartiles fall between 3.0 and 4.0.

Table 8 – Summary of Distribution of values for the Overall HHC Rating for providers

Min	1st Qu.	Median	Mean	3rd Qu.	Max
1.500	3.000	3.5000	3.482	4.000	5.000

Figure 4 - Distribution of Star Ratings for all Home Health Care Agencies (HHCAs)



A 4 or 5 Star rating means that the agency performed better than other agencies on the measured care practices and outcomes. A 1 or 2 star rating means that the agency's average performance on the 9 measured care practices and outcomes are below the averages of other agencies. Across the country, most agencies fall in the middle with 3 or 3.5 stars.

The quality of patient care star rating of different home health agencies to identify how the agencies perform compared to each other. However, since star rating calculation ranks all agencies from lowest to highest, some agencies will be ranked below others even though they're providing good quality care.

The quality of patient care star ratings are updated quarterly, at the same time the data on the individual quality measures are updated. This means the quality of patient care star rating for each home health agency may change from one quarter to the next.

3.9 Assign Historical Overall Mean of Star Rating to each Provider

After each of the Providers or HHCA are assigned a Star Rating for each quarter in the process in 3.8 above, the final steps to preparing the data predictive modeling are to calculate and assign the historical mean Star Rating for each Provider.

This calculation was performed in R by taking the mean of the overall.hcc.rating by CMS Number (Provider or HHCA) for the complete historical data set across all the quarters. This mean value was then assigned to the respective CMS Number without regards to year.quarter, creating in effect a scored data set at the Provider level excluding the year.qurater field or data grain. This data set now contains a target or response variable that can be used for predictive modeling.

The assignment of the historical mean Star Rating by CMS Number was performed on the data set that contains the original categorical variables rather than the dichotomized variables. The reason for this is so that we could have category values to perform EDA on the scored data.

Also, the same data preparation that was performed to create the dichotomized data set was performed in the resulting data set to ensure the same analytical result. This data preparation was completed in Excel and included the removal of observations with measure values > 100%, the removal of two measures, HO.PAT.was.re.admitted.to.hospital.62 and

HO.PAT.received.ER.care.wout.re.admission.64 because they only have data for one quarter, the removal of the Quality.of.Patient.Care.Star.Rating variable because it only has data for one quarter, and finally the removal of all observations which sum of the 22 measure = 0.

The resulting data set from the above process has 19,371 observations with 35 categorical and measure variables, which is down from 135,464 observations and 40 categorical and numerical measures.

This data preparation is the culmination of the initial data preparation and historical Star Rating (scoring) for all of the HHCAs resulting in the FINAL data set to be used for the predictive models (Lasso, Random Forest) in order to identify the explanatory variables to be used for the Recommendation Engine.

4. Data Analysis and Insights

The data analyzed in sections 4.1 (using Tableau) and 4.2 (using R for Descriptive Statistics) uses prepared data up to section *3.3 Standardization of Categorical Values*. The intent of the initial analysis is to understand the distribution of the nine Score-Based Measures with respect to categorical variables to ensure the appropriate modeling technique is utilized for the recommendation engine.

Another reason for analyzing the data is to better understand the hierarchy or functional relationships among variables. For example, an Agency has observations under different States. Also, an Agency can provide services in the same Zip Code but yet have several Types of Ownership.

Clustering Analysis was performed using data prepared up to section *3.9 Assign Historical Overall Mean of Star Rating to each Provider* to extract patterns/features within the data set and make logic sense of why the features appear together. Clustering analysis yielded some useful insights like the States of California and Florida seem to be doing a good job of managing patient's mobility like getting in and out of bed or moving around whereas State of Texas did poorly. Also, it was found that services like occupational services and speech pathology were mainly offered only if the ownership type was not proprietary.

4.1 Insights Using Tableau

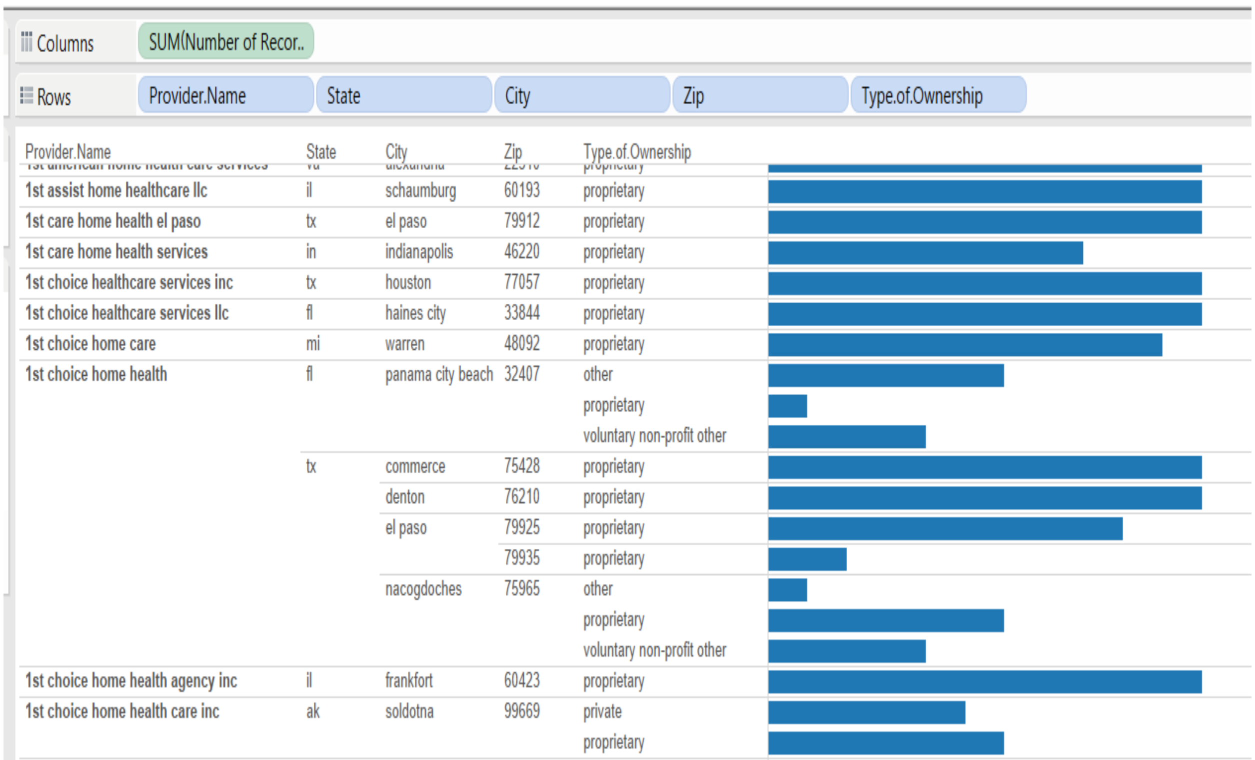
Tableau is a visualization tool that can quickly source data and load it in a well formatted table structure. This tool also offers a GUI that allows for quick creation of bar charts, histograms, box plots, and bubble charts, and tables. This visualization speeds the assessment of data quality in terms of standard categorical values across historical data sets and also distribution of the data in terms of outliers or missing data.

Tableau provides a quick way to understand the hierarchy of the data.

Analysis of the data using Tableau yielded the following insights based on Figure 5 shown below.

- A Provider can have locations in multiple States.
- A Provider can have different Types of Ownership, in the same Zip Code.

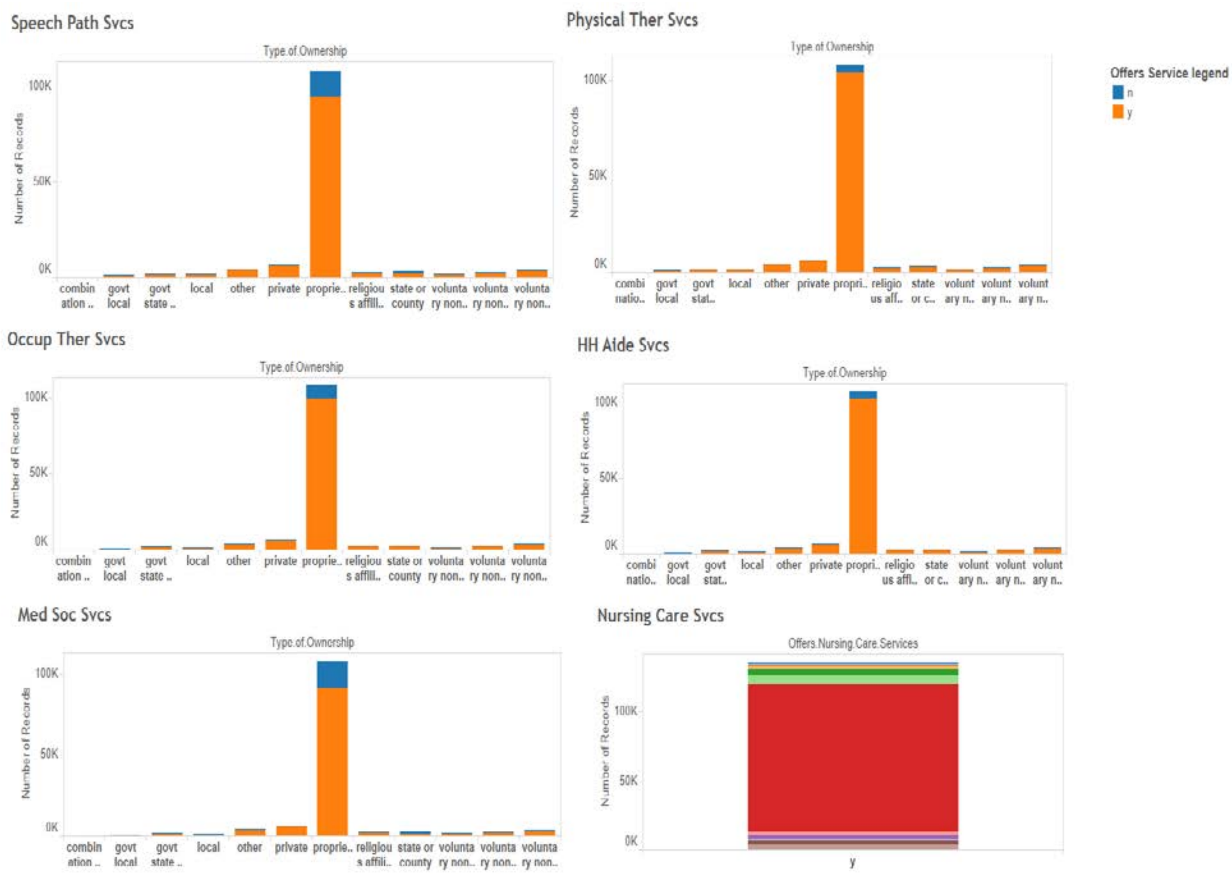
Figure 5 - Number of observations showing an Agency practicing across State boundaries as well as having multiple Types of Ownership within the same Zip Code.



Provider.Name	State	City	Zip	Type.of.Ownership
1st assist home healthcare llc	il	schaumburg	60193	proprietary
1st care home health el paso	tx	el paso	79912	proprietary
1st care home health services	in	indianapolis	46220	proprietary
1st choice healthcare services inc	tx	houston	77057	proprietary
1st choice healthcare services llc	fl	haines city	33844	proprietary
1st choice home care	mi	warren	48092	proprietary
1st choice home health	fl	panama city beach	32407	other
				proprietary
				voluntary non-profit other
	tx	commerce	75428	proprietary
		denton	76210	proprietary
		el paso	79925	proprietary
			79935	proprietary
		nacogdoches	75965	other
				proprietary
				voluntary non-profit other
1st choice home health agency inc	il	frankfort	60423	proprietary
1st choice home health care inc	ak	soldotna	99669	private
				proprietary

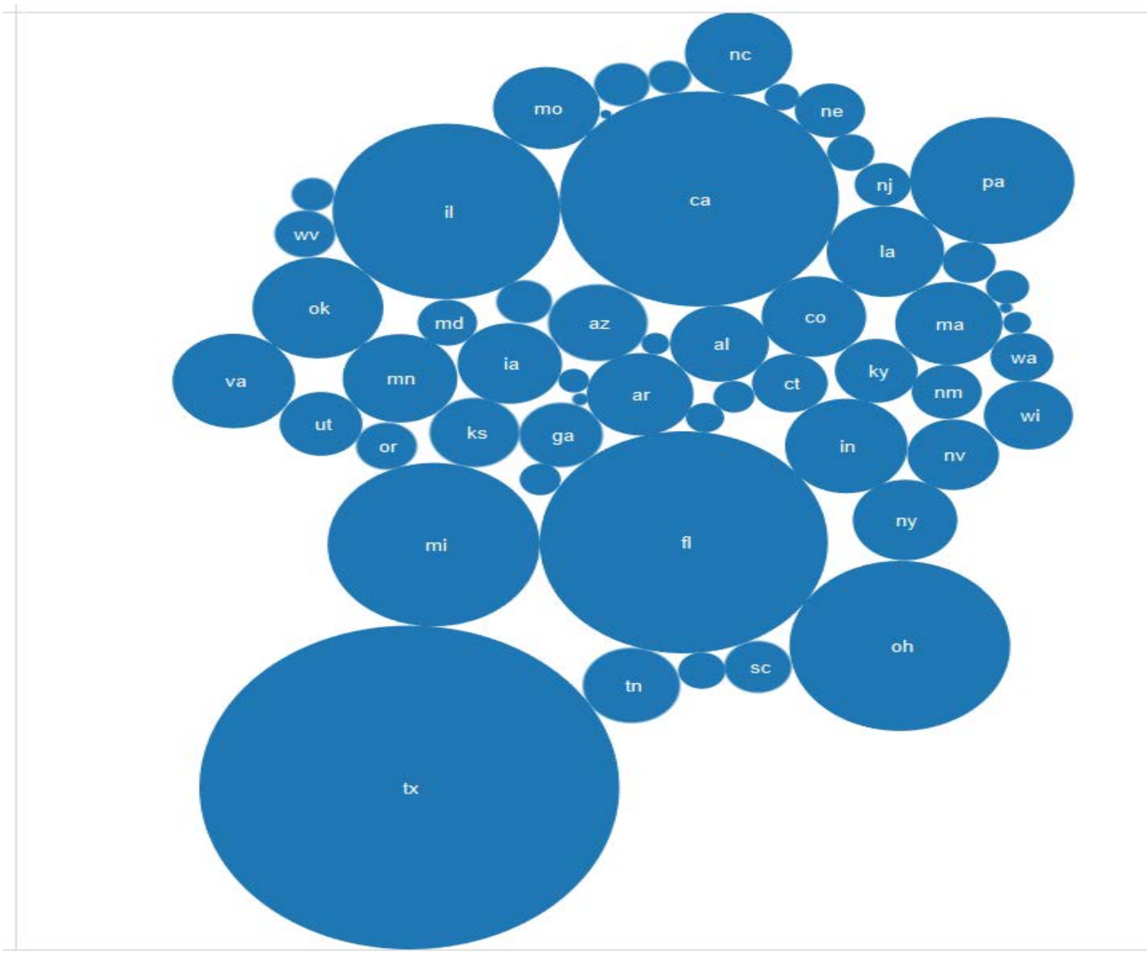
➢ Most of the observations are for 'Proprietary' Type of Ownership as shown in Figure 6 below.

Figure 6 - Number of observations by 'Type of Ownership' segmented by the type of Services Provided.



➢ Most of the observations are in Texas, Florida, California, Illinois, Ohio, Michigan, and Pennsylvania as can be seen in Figure 7 below by the larger sized bubbles.

Figure 7 - Bubble Chart – size of bubble indicates the number of observations by State.



4.2 Insights from Descriptive Statistics Using R

R provides many statistical and graphical packages that allow a transparent and granular understanding of the data. The distribution of the data or each of the nine Patient Care Quality Measure (PCQM) that are the basis for the Scoring can be intuitively understood using the Density Plot, Box Plot, Histogram, and the Q-Q Plot, as will be shown in this section in the graphs that follow.

The visualization that follows is based on the data set after it was thoroughly prepared by the removal of observations with percentage values greater than 100, imputation of missing measures to zero, retention of observations with at least one of the nine Scoring-based measures having a value > 0.

The way to interpret the percentage value of each of the nine measure is as follows:

- For measures 1 through 8 in Table 9 below, a higher value means 'better' because these measures represent that a Process was performed (measures 1, 2, and 3) or it means that the clinical Outcome of a Patient improved (measures 4, 5, 6, 7, and 8)
- For measure 9, a lower value means 'better' because this measure represents whether a patient had to be admitted to a hospital despite the attempt to take care of this patient at home.

Additional information on the Measures naming convention:

- The prefix HO.HHT means 'How Often the Health Home Team…'
- The prefix HO.PAT means 'How Often the Patient …'

- The integer suffix ties this measure to the original CMS data dictionary field number listed for the HHC_SOCRATA_PRVDR.csv file, which is the source data file. The data dictionary file is bundled in the zip file when downloading the data files from CMS (2). The data dictionary file is called "*HomeHealthCompare_Revised_FlatFiles.pdf*"

We start the analysis of the insights from R with the Summary of the Descriptive Statistics as shown in Table 9 below. In this table, we can clearly see that most of the observations have percentage values for the nine measures between quartiles 25 and 99 (1,725 + 1,526 = 3,251 observations).

However, given that the Mean is just below the Median (Q50) for all the measures, this is an indication that there is a slight skewness to the left. The left skewness may have two reasons, first, the imputation of the missing values (which were at around 10% of the observations for many of the measures) and second, the observations that remained in the final data set may not have had values > 0 for all of the nine measure thus causing the measure to have more observations with a value of 0 than > 0. However, this preparation of the data is appropriate because the observations should be part of the population even if it only has one of the nine measures with a value > 0.

Table 9 - Descriptive Statistics - Summary

	VAR.NAME	NOT.NA	IS.NA	PCT.NA	MIN	Q01	Q05	Q25	Q50	Q75	Q95	Q99	MAX	*MEAN*	SD
1	HO.HHT.began.care.in.timely.manner.18	107,471	0	0	0	30	72	87	**93**	96.8	99	100	100	*89.83*	12.32
2	HO.HHT.taught.about.their.drugs.20	107,471	0	0	0	0	64	89	**95**	98	100	100	100	*89.99*	16.57
3	HO.HHT.ensured.received.flu.shot.26	107,471	0	0	0	0	9	60	**74**	83	96	100	100	*67.99*	23.30
4	HO.PAT.got.better.at.moving.around.44	107,471	0	0	0	0	0	49	**58**	65	78	93	100	*54.50*	19.02
5	HO.PAT.got.better.at.getting.in.and.out.of.bed.46	107,471	0	0	0	0	0	42	**53**	62	75	89	100	*49.37*	19.94
6	HO.PAT..got.better.at.bathing.48	107,471	0	0	0	0	0	55	**65**	72	85	96	100	*60.35*	20.77
7	HO.PAT.had.less.pain.when.moving.around.50	107,471	0	0	0	0	0	54	**65**	75	91	98	100	*60.88*	23.30
8	HO.PAT.breathing.improved.52	107,471	0	0	0	0	0	45	**62**	72	84	93	100	*55.09*	24.32
9	HO.PAT.had.to.be.admitted.to.the.hospital.60	107,471	0	0	0	0	0	12	**15**	18	22	27	54	*14.31*	5.93
						30	145	493	580	641.8	730	796			
						175		1715			1526				

The next nine sub-sections describe each of the nine Scoring-Based Measure with respect to their distribution, outliers, and normality.

The resulting statistical descriptions confirm that these nine measures are not normally distributed and will require scale transformation in order to normalize them for a more accurate fit in regression modeling.

4.2.1. Insights for the ***Timely Initiation of Care*** Quality Measure Visualization

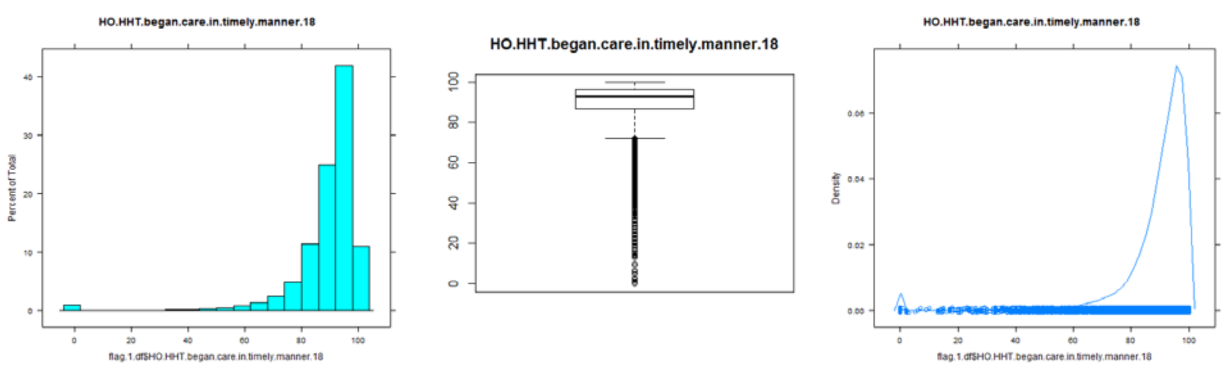
This measure is the percentage of home health episodes of care in which the start or resumption of care date was either on the physician- specified date or within 2 days of the referral date or inpatient discharge date, whichever is later.

The **histogram** shown in Figure 8 below clearly indicates that most of the observations have a high percentage value for the ***Timely Initiation of Care***, with a Mean of about 89%, a Median of 93%, and a Standard Deviation of 12.32%. In fact the difference between the Mean and the Median is only 4 points, not highly skewed. This is indeed great performance for most of the Agencies as this indicates that they started the Home Care on a timely manner.

The **boxplot** in Figure 8 below indicates that there are quite a few outliers skewed to the bottom or below the first quartile. This visualization would need further investigation to identify whether these outlier Agencies were indeed late starting the Home Care or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

The **Density** plot in Figure 8 below, shows that the distribution of the data is normally distributed, left-skewed, and bimodal.   However, the very small left peak of the bimodal distribution on the zero value indicates that these observations make very little of the proportion of the data.

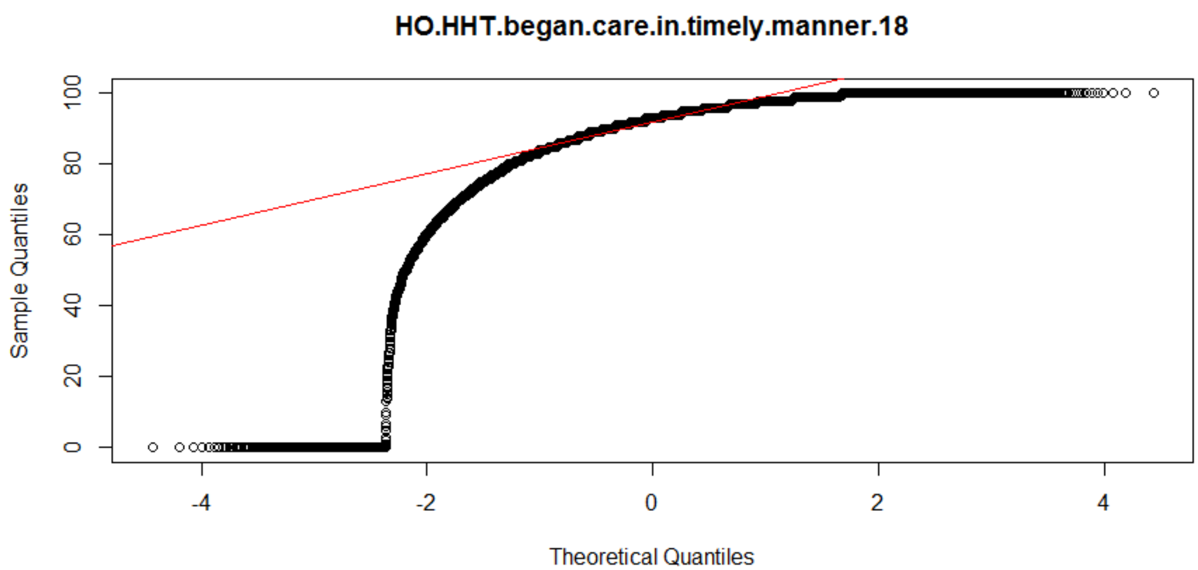
Figure 8 - Process - *Timely Initiation of Care* – Histogram, Box Plot, Density Plot



Judging global features of this variable to ensure a good fit, the analysis requires the comparison of the univariate continuous variable against a theoretical distribution, in this case the normal distribution. The tool to use for this comparison is the Q-Q Plot, which is shown in Figure 9 below.

The Q-Q Plot shown in Figure 9 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 9 - Process - *Timely Initiation of Care* – Q-Q Plot



4.2.2. Insights for the ***Drug Education*** Quality Measure Visualization

This measure the percentage of home health episodes of care during which patient/caregiver was instructed on how to monitor the effectiveness of drug therapy, how to recognize potential adverse effects, and how and when to report problems.

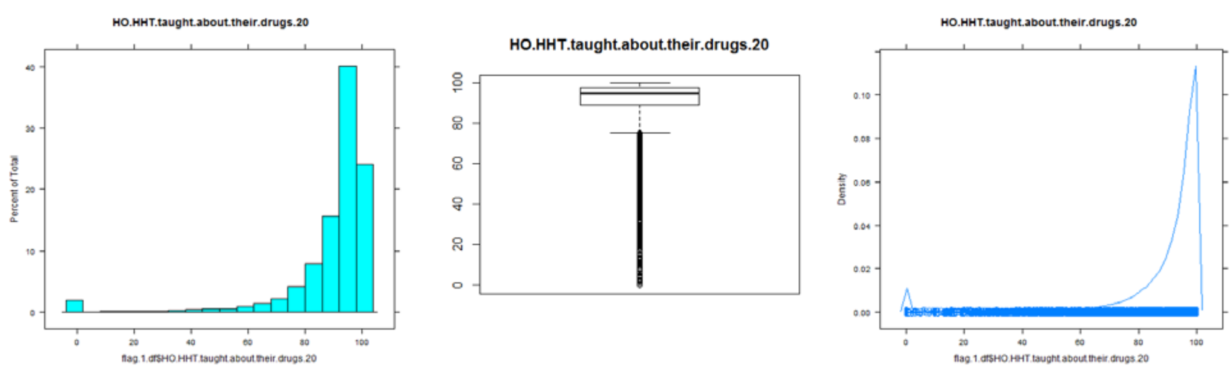
The **histogram** shown in Figure 10 below clearly indicates that most of the observations have a high percentage value for the ***Drug Education***, with a Mean of about 90%, a Median of

95%, and a Standard Deviation of 16.57%. In fact, the difference between the Mean and the Median is only 5 points, not highly skewed. This is indeed great performance for most of the Agencies as this indicates that they completed Drug Education for their home bound patients.

The **boxplot** in Figure 10 below indicates that there are quite a few outliers skewed to the bottom or below the first quartile. This visualization would need further investigation to identify whether these outlier Agencies did not perform *Drug Education* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

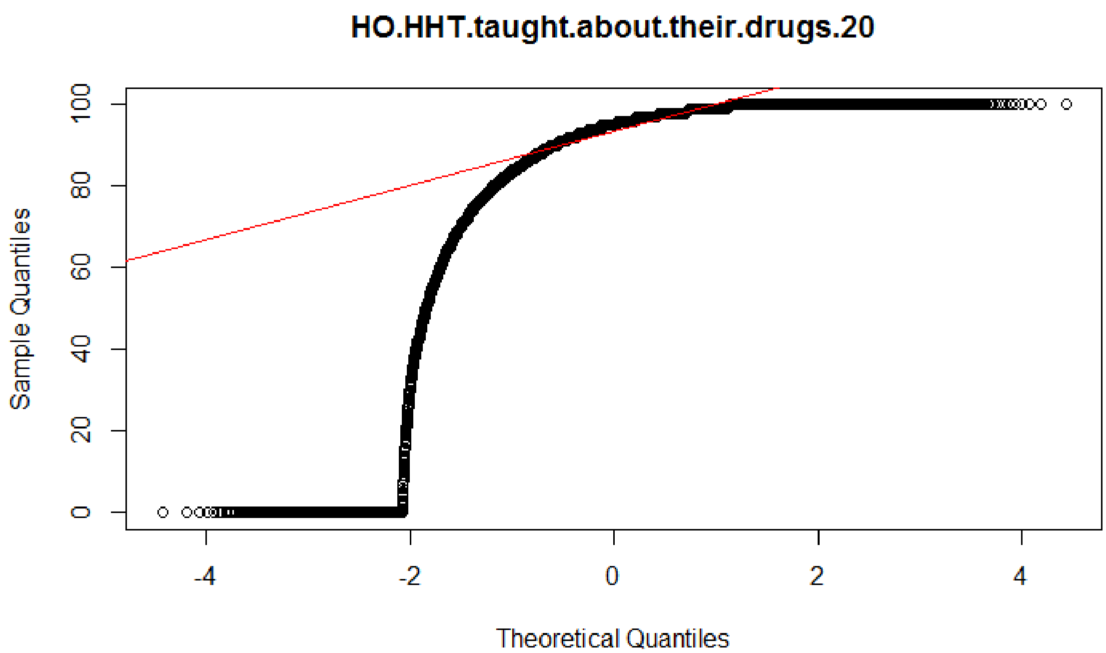
The **Density** plot in Figure 10 below, shows that the distribution of the data is normally distributed, left-skewed, and bimodal.   However, the very small left peak of the bimodal distribution on the zero value indicates that these observations make very little of the proportion of the data.

Figure 10 - Process - *Drug Education* on All Medications Provided to Patient/Caregiver during All Episodes of Care - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 11 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 11 - Process - *Drug Education* – Q-Q Plot



4.2.3. Insights for the ***Influenza Immunization*** Quality Measure Visualization

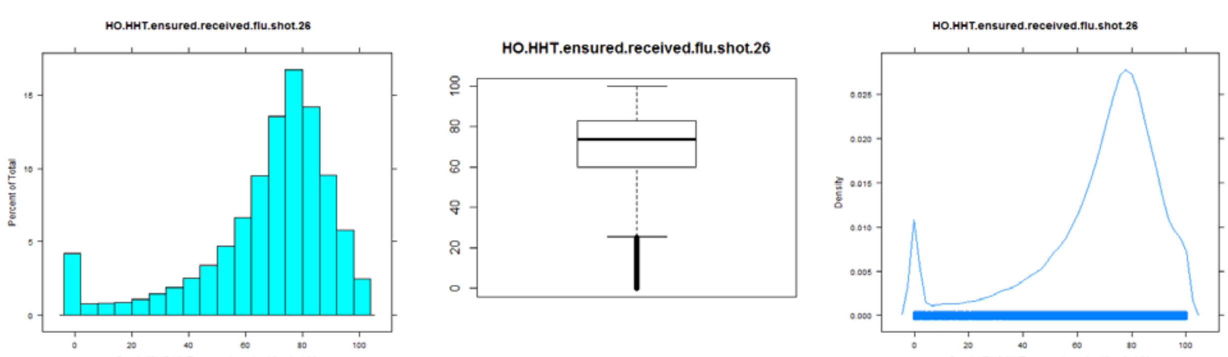
This measure the percentage of home health episodes of care during which patients received influenza immunization for the current flu season.

The **histogram** shown in Figure 12 below clearly indicates that most of the observations did not excel in providing ***Influenza Immunization*** to their home bound patients, with a Mean of about 68%, a Median of 74%, and a Standard Deviation of 23.30%. The difference between the Mean and the Median is only 6 points, not highly skewed but the Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum. This is **not** a great performance for many of the Agencies as this indicates that many of the Agencies did not complete Influenza Immunization for their home-bound patients.

The **boxplot** in Figure 12 below indicates that there are quite a few outliers skewed to the bottom or below the first quartile. This visualization would need further investigation to identify whether these outlier Agencies did not perform *Influenza Immunization* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

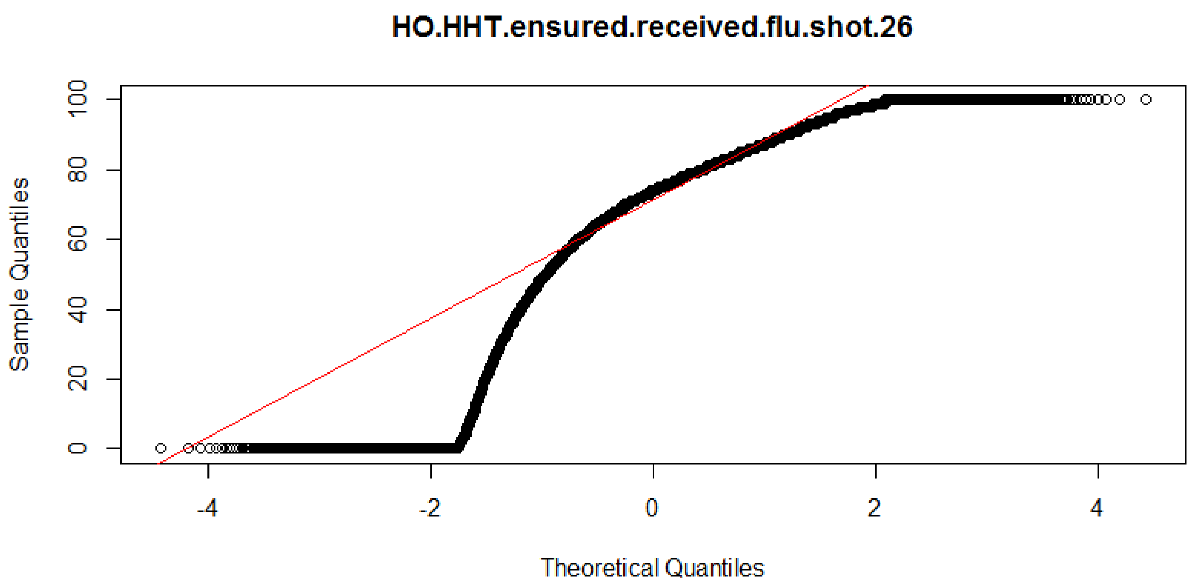
The **Density** plot in Figure 12 below, shows that the distribution of the data is normally distributed, left-skewed, and bimodal. However, the larger left peak of the bimodal distribution on the zero value indicates that these observations make a significant size of the proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing Influenza Immunizations to their home bound patients.

Figure 12 - Process – Prevention - *Influenza Immunization* Received for Current Flu Season - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 13 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 13 - Process - *Influenza Immunization* – Q-Q Plot



4.2.4. Insights for the ***Improvement in Ambulation- Locomotion*** Quality Measure Visualization

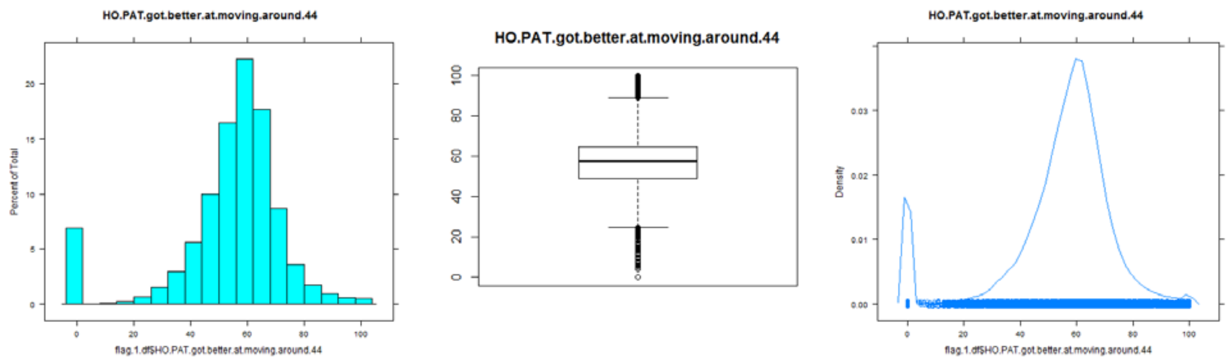
This measure is the percentage of home health episodes of care during which the patient improved in ability to ambulate.

The **histogram** shown in Figure 14 below clearly indicates that most of the observations did not excel in the *Improvement in Ambulation- Locomotion* of their home bound patients, with a Mean of about 55%, a Median of 58%, and a Standard Deviation of 19%. The difference between the Mean and the Median is only 3 points, not highly skewed but the Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum. This is **not** a great performance for many of the Agencies as it indicates that many of the Agencies did not make Improvements on the clinical outcome measure for *Improvement in Ambulation- Locomotion* for their home-bound patients.

The **boxplot** in Figure 14 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile and toward the top above the 75th quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Improvement in Ambulation- Locomotion* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

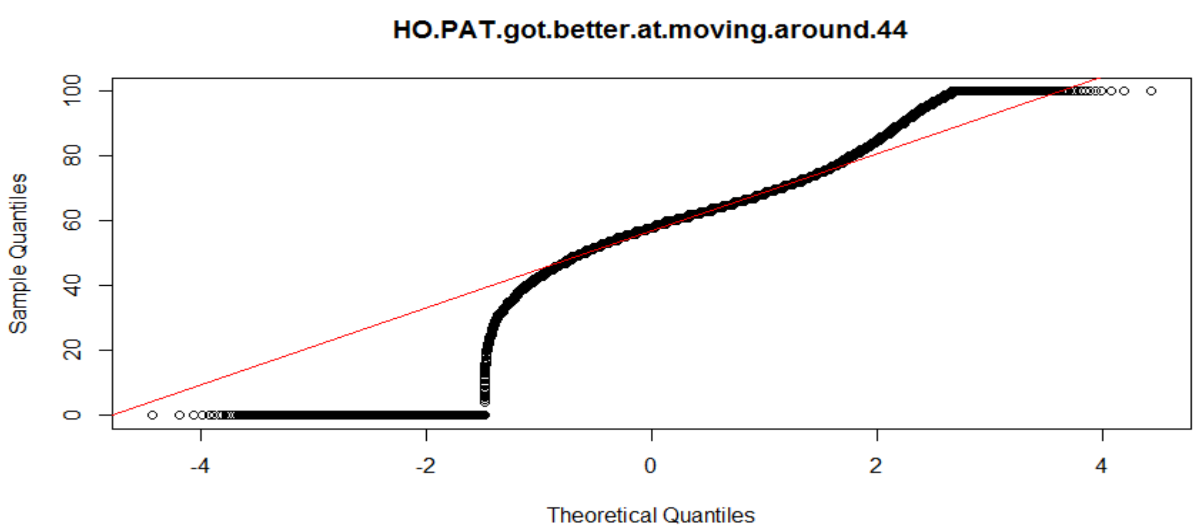
The **Density** plot in Figure 14 below, shows that the data is normally distributed, left-skewed, and tri-modal. However, the larger left peak of the tri-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing *Improvement in Ambulation-Locomotion* to their home-bound patients. The right-most peak of the tri-modal distribution may indicate that many agencies performed quite well in improving the outcome for their home-bound patients to move around by themselves or it may also indicate that there is some data issue with this measure having such large values for few of the observations. Further investigation needs to be performed to better understand why there is a third mode in this distribution.

Figure 14 - End Result Outcome – Functional - *Improvement in Ambulation-Locomotion* - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 15 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 15 - Process - *Improvement in Ambulation- Locomotion* – Q-Q Plot



4.2.5. Insights for the ***Improvement in Bed Transferring*** Quality Measure Visualization

This measure is the percentage of home health episodes of care during which the patient improved in ability to get in and out of bed.

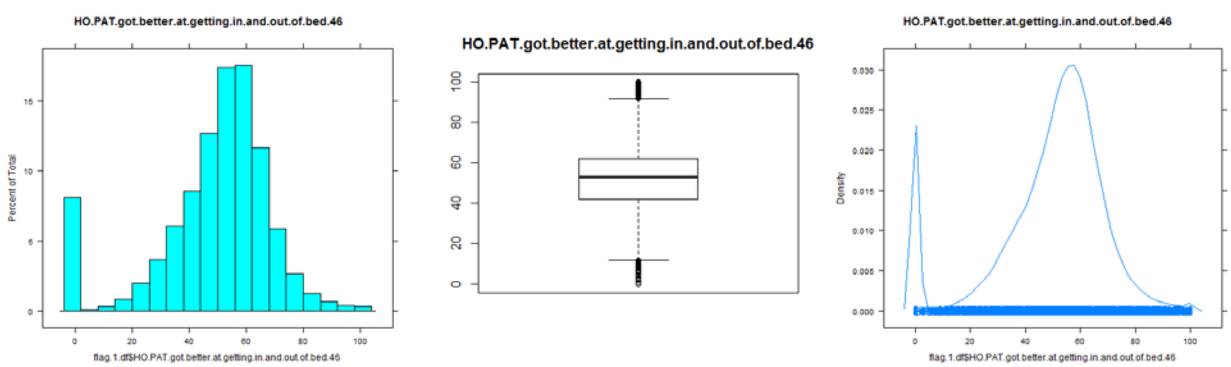
The **histogram** shown in Figure 16 below clearly indicates that most of the observations did not excel in the *Improvement in Bed Transferring* of their home bound patients, with a Mean of about 49%, a Median of 53%, and a Standard Deviation of 20%. The difference between the Mean and the Median is only 3 points, not highly skewed but the Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum. This is **not** a great performance for many of the Agencies as it indicates that many of the Agencies did not make Improvements on the clinical outcome measure for *Improvement in Bed Transferring* for their home-bound patients.

The **boxplot** in Figure 16 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile and toward the top above the 75th quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Improvement in Bed Transferring* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

The **Density** plot in Figure 16 below, shows that the data is normally distributed, left-skewed, and tri-modal. However, the larger left peak of the tri-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing *Improvement in Bed*

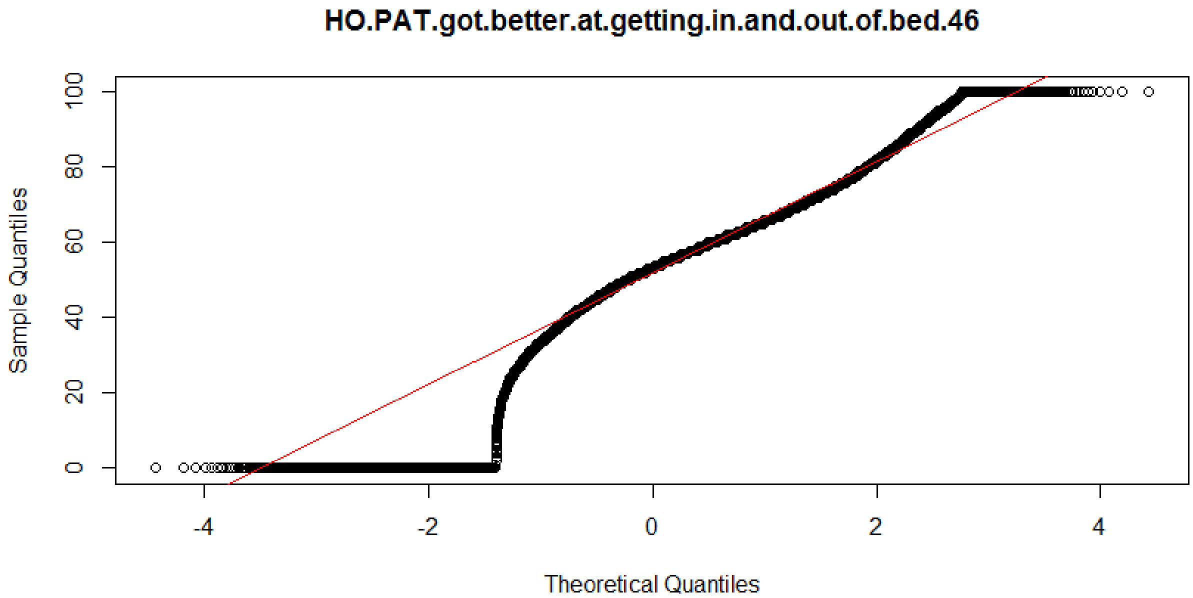
Transferring to their home-bound patients. The right-most peak of the tri-modal distribution may indicate that many agencies performed quite well in improving the outcome for their home-bound patients to get out of bed by themselves or it may also indicate that there is some data issue with this measure having such large values for few of the observations. Further investigation needs to be performed to better understand why there is a third mode in this distribution.

Figure 16 - End Result Outcome – Functional - *Improvement in Bed Transferring* - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 17 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 17 - Process - *Improvement in Bed Transferring* – Q-Q Plot



4.2.6. Insights for the *Improvement in Bathing* Quality Measure Visualization

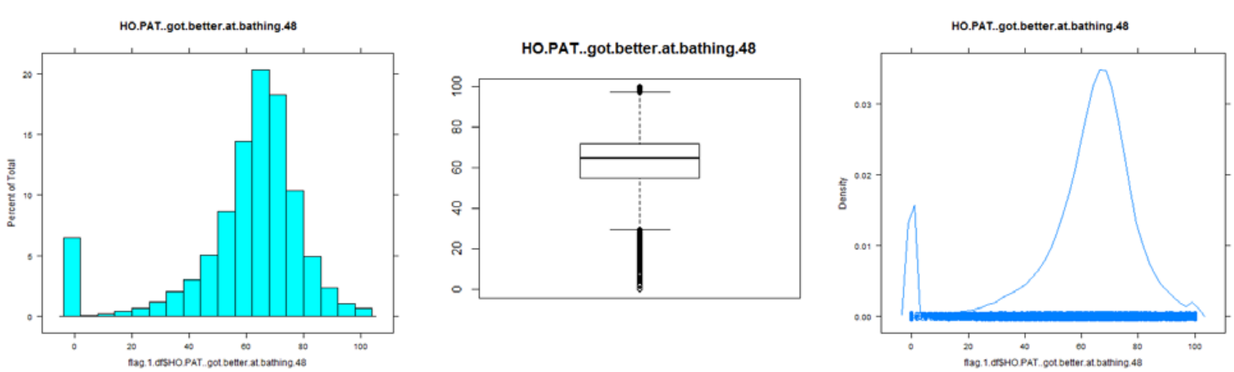
This measure is the percentage of home health episodes of care during which the patient got better at bathing self.

The **histogram** shown in Figure 18 below clearly indicates that many of the observations bettered the *Improvement in Bathing* of their home bound patients, with a Mean of about 60%, a Median of 65%, and a Standard Deviation of 21%. The difference between the Mean and the Median is 5 points, somewhat skewed to the left; however the Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum.

The **boxplot** in Figure 18 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile and toward the top above the 75th quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Improvement in Bathing* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

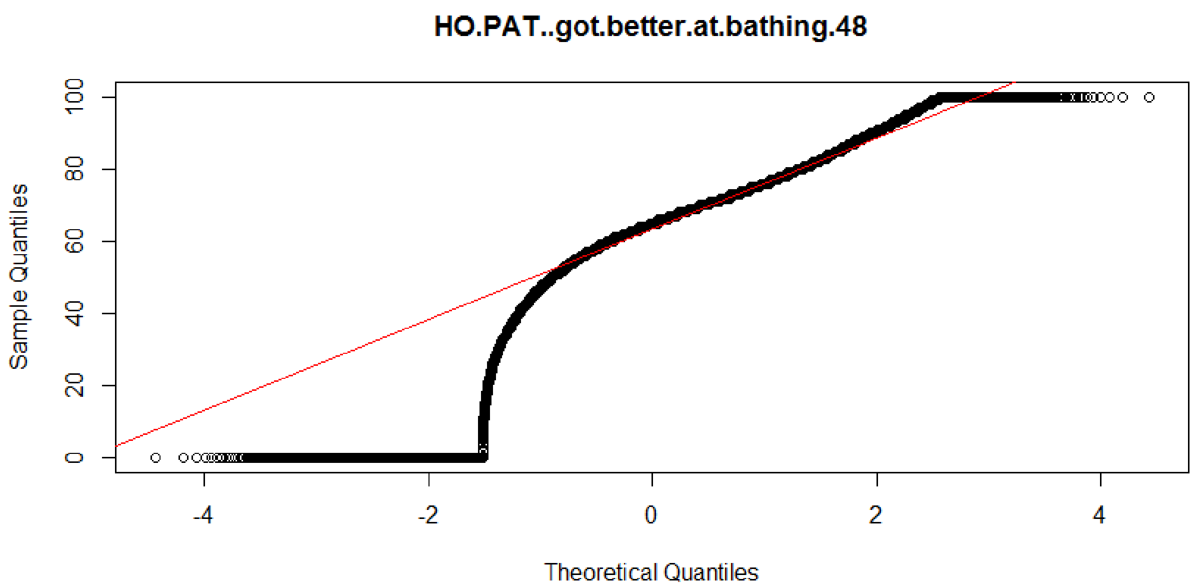
The **Density** plot in Figure 18 below, shows that the data is normally distributed, left-skewed, and tri-modal. However, the larger left peak of the tri-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing *Improvement in Bathing* to their home-bound patients. The right-most peak of the tri-modal distribution may indicate that many agencies performed quite well in improving the outcome for their home-bound patients to get bathe by themselves or it may also indicate that there is some data issue with this measure having such large values for few of the observations. Further investigation needs to be performed to better understand why there is a third mode in this distribution.

Figure 18 - Outcome – Functional - *Improvement in Bathing* - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 19 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 19 - Process - *Improvement in Bathing* – Q-Q Plot



4.2.7. Insights for the ***Improvement in Pain Interfering with Activity*** Quality Measure Visualization

This measure is the percentage of home health episodes of care during which the patient's frequency of pain when moving around improved.

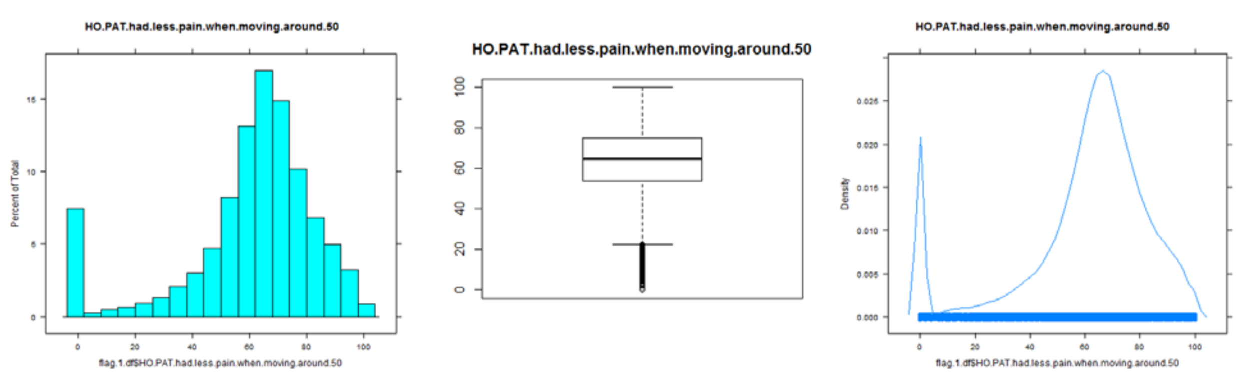
The **histogram** shown in Figure 20 below clearly indicates that many of the observations bettered the *Improvement in Pain Interfering with Activity* of their home bound patients, with a Mean of about 61%, a Median of 65%, and a Standard Deviation of 23%. The difference between the Mean and the Median is 5 points, somewhat skewed to the left; however the

Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum.

The **boxplot** in Figure 20 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Improvement in Pain Interfering* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

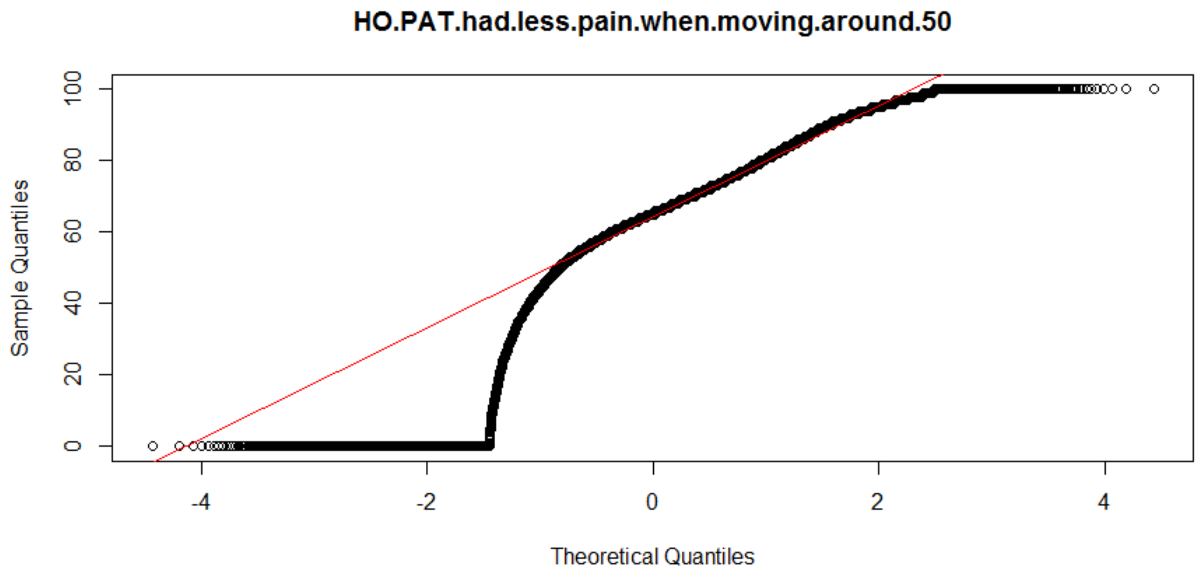
The **Density** plot in Figure 20 below, shows that the data is normally distributed, left-skewed, and bi-modal. However, the larger left peak of the bi-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing *Improvement in Pain Interfering* to their home-bound patients.

Figure 20 - End Result Outcome – Health - *Improvement in Pain Interfering with Activity* - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 21 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 21 - Process - *Improvement in Pain Interfering with Activity* – Q-Q Plot



4.2.8. Insights for the ***Improvement in Dyspnea (Breathing)*** Quality Measure Visualization

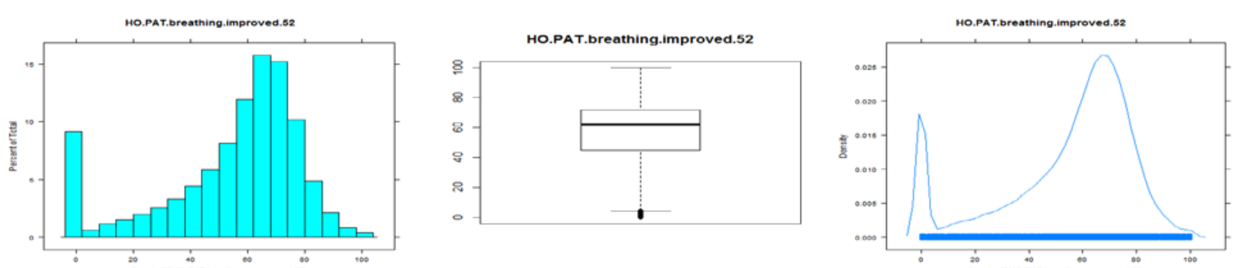
This measure is the percentage of home health episodes of care during which the patient became less short of breath or dyspneic.

The **histogram** shown in Figure 22 below clearly indicates that many of the observations bettered the *Improvement in Dyspnea (Breathing)* of their home bound patients, with a Mean of about 55%, a Median of 62%, and a Standard Deviation of 24%. The difference between the Mean and the Median is 7 points, highly skewed to the left compared to the other measures; also the Standard Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum.

The **boxplot** in Figure 22 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Improvement in Dyspnea (Breathing)* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

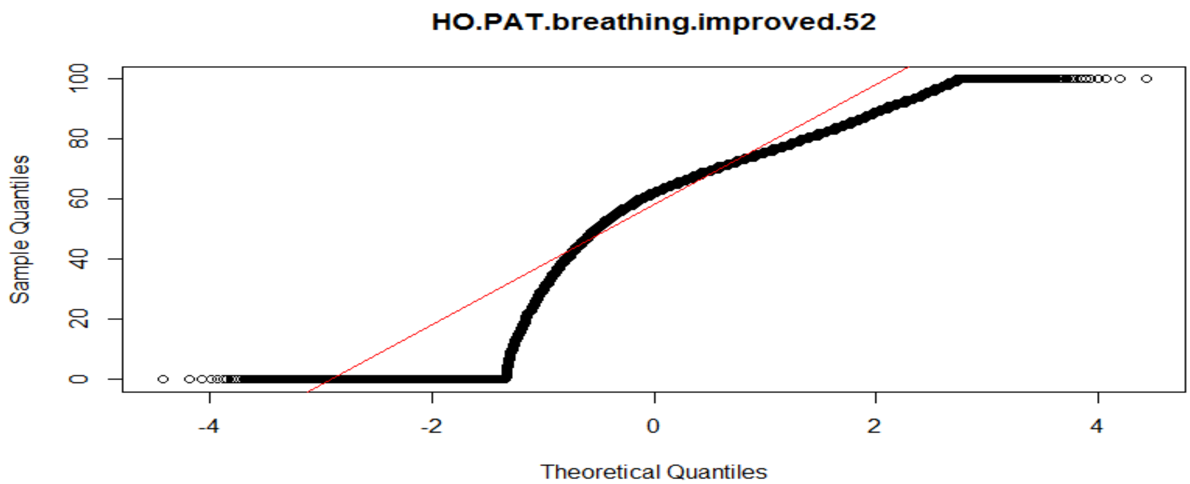
The **Density** plot in Figure 22 below, shows that the data is normally distributed, left-skewed, and bi-modal. However, the larger left peak of the bi-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not providing *Improvement in Dyspnea (Breathing)* to their home-bound patients.

Figure 22 - End Result Outcome – Health - *Improvement in Dyspnea (Breathing)* - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 23 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 23 - Process - *Improvement in Dyspnea (Breathing)* – Q-Q Plot



4.2.9. Insights for the *Acute Care Hospitalization* Quality Measure Visualization

This measure is the percentage of home health stays in which patients were admitted to an acute care hospital during the 60 days following the start of the home health stay.

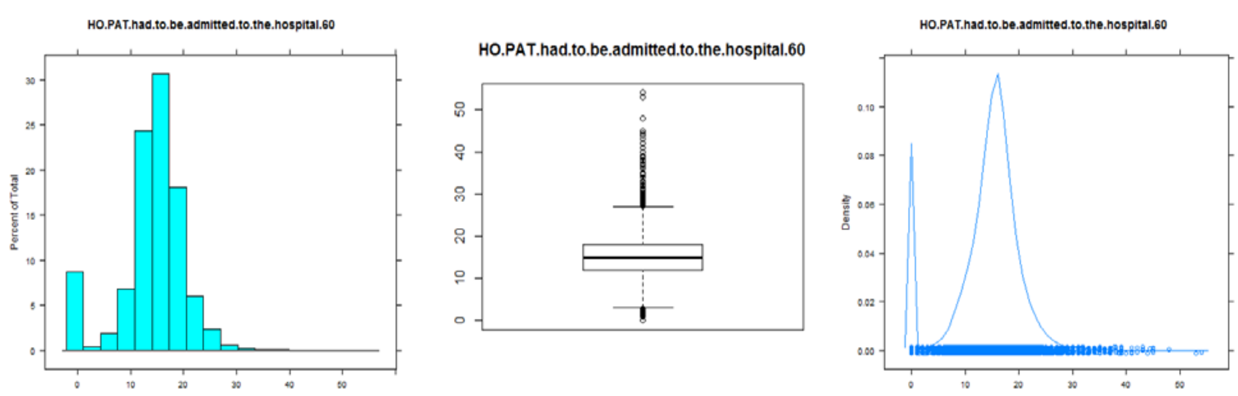
The **histogram** shown in Figure 24 below clearly indicates that many of the observations had lower *Acute Care Hospitalization* of their home bound patients, with a Mean of about 14%, a Median of 15%, and a Standard Deviation of 6%. The difference between the Mean and the Median is 1 point, however, the distribution is highly skewed to the right; also the Standard

Deviation indicates a much wider spread of the data indicating outliers at both end of the spectrum.

The **boxplot** in Figure 24 below indicates that there are quite a few outliers skewed toward the bottom or below the first quartile and toward the top of the 75h quartile. This visualization would need further investigation to identify whether the outlier Agencies at the lower-end did not perform *Acute Care Hospitalization* or whether the Agency has a missing value in this measure for a valid reason and their value was imputed to zero.

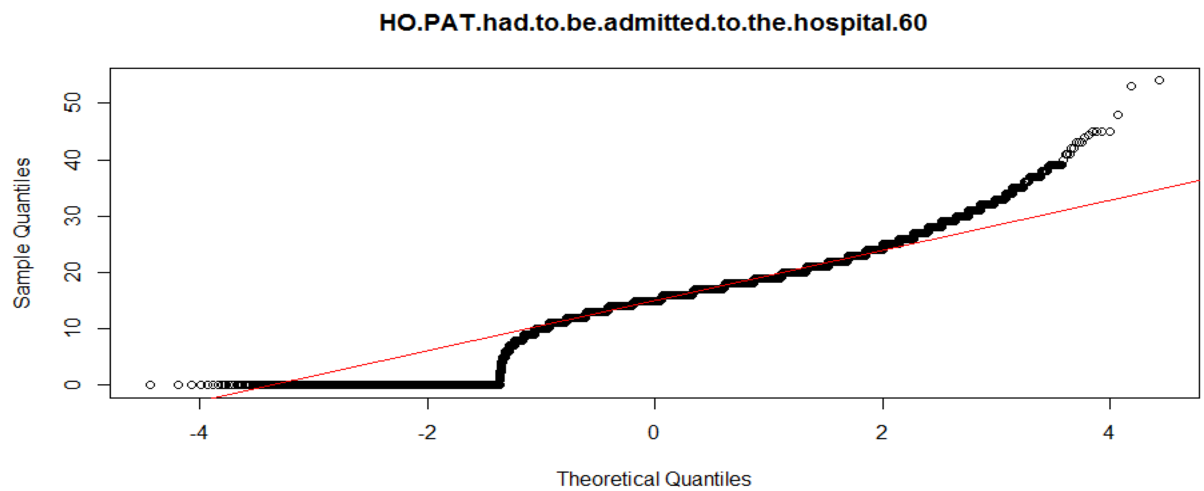
The **Density** plot in Figure 24 below, shows that the data is normally distributed, right-skewed, and bi-modal. However, the larger left peak of the bi-modal distribution on the zero value indicates that these observations make a significant proportion of the data. Further research needs to be performed to understand whether these low values are due to missing data or are due to low performance by Agencies in not lowering the events for *Acute Care Hospitalization* to their home-bound patients.

Figure 24 - Utilization Outcome - *Acute Care Hospitalization* (Claims based) - Histogram, Box Plot, Density Plot



The Q-Q Plot shown in Figure 25 below clearly demonstrates that normality and equal variance cannot be assumed for this measure because there are deviations in the plotted black data points from the red straight line that represents normality. For a good fit of this measure, the plotted black data points need to be roughly linear. Before fitting **this measure** in a regression model, it **will require scale transformation** to become Normal.

Figure 25 - Process - *Acute Care Hospitalization* – Q-Q Plot



4.3 Insights from Clustering Using R

Clustering Analysis was performed on the Home Health Care data set which had been dichotomized and imputed and several quarters of data had been combined into one data set. The entire data set had around 107,000 rows and 3430 variables. A majority of the 3,430 variables were the zip codes of locations in which the providers had provided home health care services. Assuming that the zip codes would not provide much value in this analysis, we decided that it would be wise to eliminate them from our analysis. Also, the number of cases in states other than CA, FL, MI, PA, OH, and TX was very small to impact our analysis. So, we eliminated the variables representing other states. This left us with 28 variables which broadly included Types of Ownership, Services offered, nine Patient Care Quality measures used in the calculation of the Quality Star Rating, and three States.

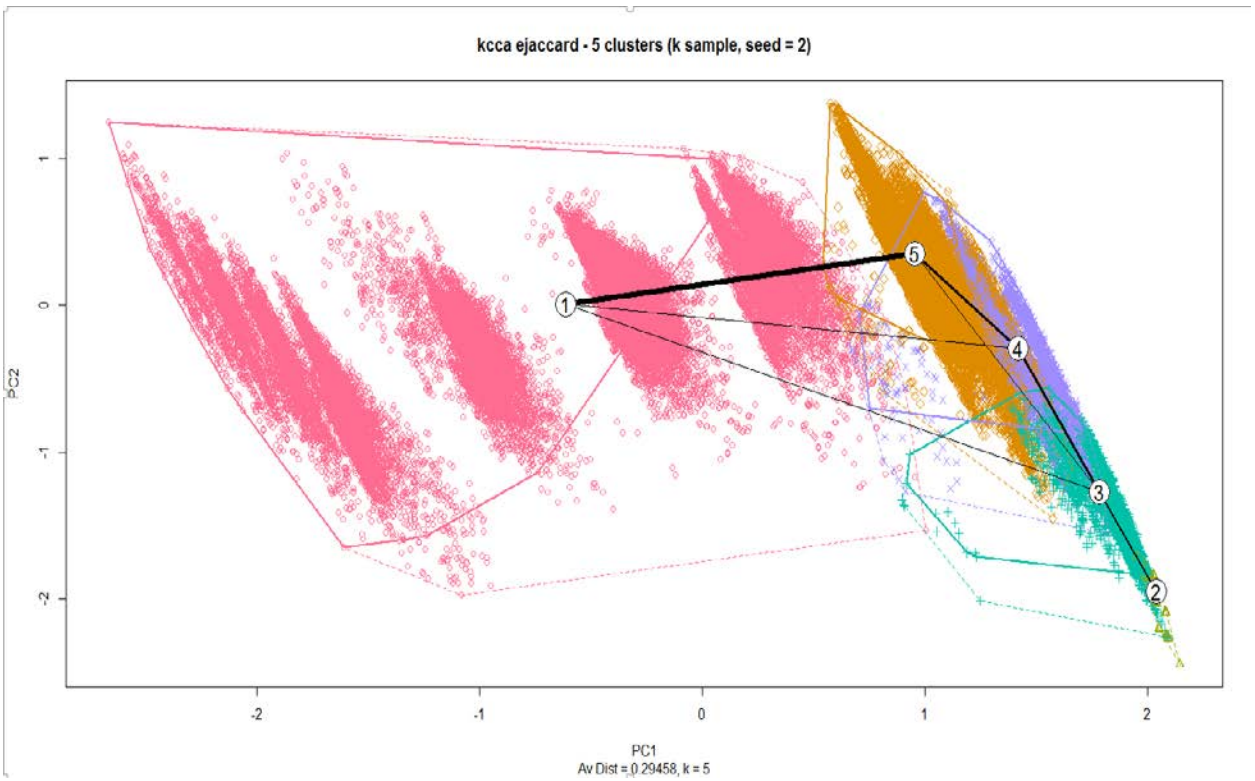
We utilized the ‘flexclust’ package in R to build the k-means model and ran the analysis for k values of 4, 5, 6 & 7.

We applied Principal Component Analysis to reduce the dimensionality of the data and help us obtain better clusters. We tried capping the number of iterations to 30 but did not see any convergence. Then we increased the max iterations value to 100, still could not find any convergence. The plot of PC1 v/s PC2 did not show clear clusters.

We noticed that the scale of the variables could be an issue. The scale of variables for the nine Patient Care Quality measures was 0-100 and for rest of the variables were 0 or 1. So, we divided the values for all the nine PCQ measures by 100 to bring their scale to be between 0 and 1. A rerun of the clustering function produced convergence at 13 iterations and the plot showed clear separation between clusters.

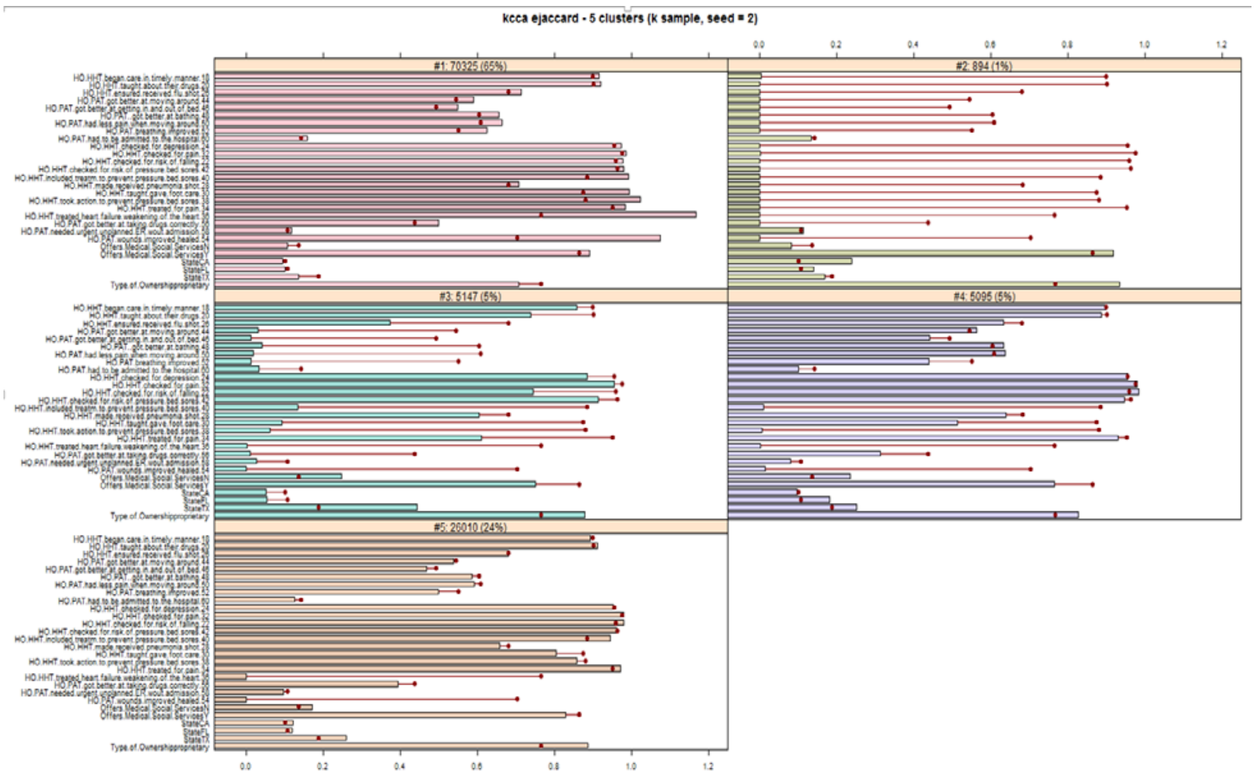
After applying the dimensionality reduction using Principal Component Analysis we tried several iterations with different number of clusters. In the end the iteration with 5 clusters provided the best separation of clusters. Figure 26 below shows the graph with 5 clusters and each color represents a different cluster. In Figure 26, points tagged in little circle as 1, 2, 3, 4, and 5 are the centroids of clusters 1 through 5. Clusters are identified by colors: 1 – pink, 2 – green, 3 - blue, 4 – violet, and 5 – orange.

Figure 26 - K-Means Cluster with PCA



Also, we plotted the significance of each variable within each cluster. Figure 28 below shows a lattice of 5 bar graphs each representing one cluster and the color scheme follows the cluster plot explained for Figure 26 above. Each bar represents the significance of each variable within each cluster and for reference a mark on each bar suggests the overall mean value from the entire data set.

Figure 28 – Significance of Variables from Cluster Analysis



The value of separation between cluster and the significance of the variables in each cluster can be used to drive conclusions about the characteristics of each cluster. For example, **cluster # 1** shows comparatively higher values for the State of CA and values of mobility measures like getting in and out of bed and moving around seem significant when compared to the other clusters. Hence we can assume that the HHCAs in the State of CA and FL probably do a good job at working on their patient's mobility compared to the national average.

Similarly in **cluster # 2** values of performance measures seem to dip while the values of the providers in state of Texas are significantly higher. We can infer that the providers in the state of Texas seem to do poorly in comparison to the national average. Perhaps focusing on the processes on the providers might improve the overall rating of care in the state of TX. However, the size of this cluster is too small (1%) to make a significant impact on the overall rating.

Also, from **cluster # 5** it seems that in the absence of HHCAs owned by proprietors, there seems to be a drop in number of speech pathology and occupational therapy services. Although it has not been found that there is a direct correlation between offering these services improves the rating of providers but the coincidence that in this cluster the state of TX is also a factor and providers in the state of TX have a less than favorable rating compared to other states, providers in the states of TX may also want to look at beginning to offer speech pathology and occupational therapy services.

Cluster # 3 & 4 seem to show that while variable for providers from the State of FL is significantly higher than in other clusters, there is a big drop in offering of Medical Social services. There is also a sharp decline in the number of times patient had to be taken to ER without being admitted to Hospital in cluster # 3 and 4. The values for hospital admissions are also low in these two clusters. This does not mean that the quality of care is good because this particular measure, different for the other measures, a lower value means a better provider since the patient did not have to be taking to the ER.

These findings from the Cluster analysis will be used with the findings from other sections of this study to draw our final conclusions.

4.4 Insights from Random Forest (Regression) on Important Variables using R

The resulting data set from the data preparation phase resulted in 3,409 variables creating in effect a Big Data problem. (*no* pun intended)

In order to reduce the number of variables, the importance of each variable was evaluated using the Random Forest package in R. This technique identifies the variables with the greatest impact in variability on the target variable.

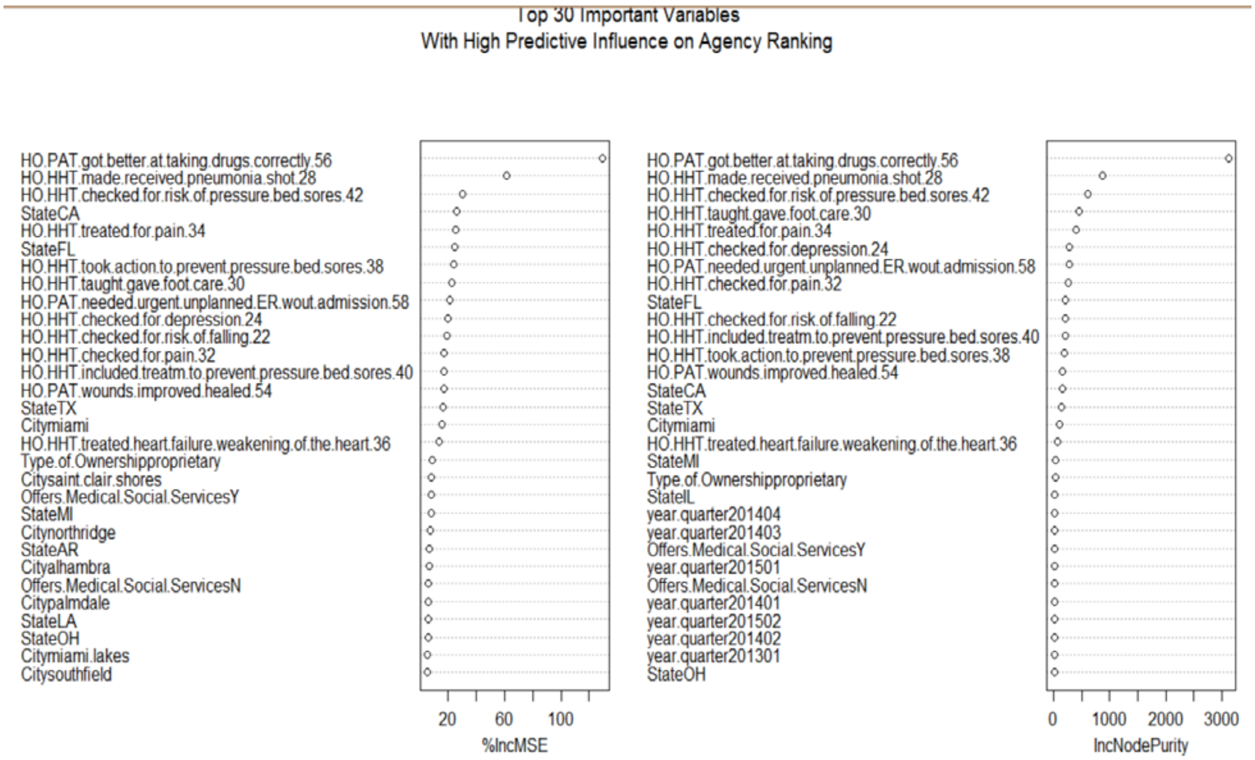
The original historical data set did not include a Target or Class variable. The creation of the Target variable was performed during the Scoring of the data in the Data Preparation phase. This Scoring provided each observation or HHCA with an Overall Rating based on nine Patient Care Quality Measures, please refer to section *3.8 Score Agencies based on the Average Bin of the Nine Ranking-Based Measures* for specifics.

The top 30 Important Variables are shown in Table 10 below. These variables have the highest *%IncMSE measure*, which is the measurement of disturbance the variable added to the tree when it was introduced during the random selection of variables as the algorithm created new trees and compared them. Also, these variables have the highest *IncNodePurity measure*, which is the ability of the variable to decrease the Node Purity and split the Tree.

These two measures are accumulated by each variable during the Random Forest execution and the most important variables are those whose measures are the highest because they have the ability to impact the Target variable more heavily than non-important variables and thus are the most predictive variables.

Figure 27 below shows these top 30 variables ordered from top as the most important to bottom as the least important. The left pane uses the *%IncMSE* measure to rank the variables and the right pane uses the IncNodePurity.

Figure 27 – Random Forest (Regression) Top 30 Important Variables



Some of the variables in Figure 27 above are repeated in both panes. In order create a distinct list of these variable, the 'importance' function was used and the results exported to Excel. The top 30 variables from each pane were unioned and then twenty two duplicates removed resulting in the list of thirty eight variables shown in Table 10 below.

Table 10 – Random Forest (Regression) Distinct list of 38 Important Variables

Sorted Alphabetically

#	Variable Name
1	Cityalhambra
2	Citymiami
3	Citymiami.lakes
4	Citynorthridge
5	Citypalmdale
6	Citysaint.clair.shores
7	Citysouthfield
8	HO.HHT.checked.for.depression.24
9	HO.HHT.checked.for.pain.32
10	HO.HHT.checked.for.risk.of.falling.22
11	HO.HHT.checked.for.risk.of.pressure.bed.sores.42
12	HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40
13	HO.HHT.made.received.pneumonia.shot.28
14	HO.HHT.taught.gave.foot.care.30
15	HO.HHT.took.action.to.prevent.pressure.bed.sores.38
16	HO.HHT.treated.for.pain.34
17	HO.HHT.treated.heart.failure.weakening.of.the.heart.36
18	HO.PAT.got.better.at.taking.drugs.correctly.56
19	HO.PAT.needed.urgent.unplanned.ER.wout.admission.58
20	HO.PAT.wounds.improved.healed.54
21	Offers.Medical.Social.ServicesN
22	Offers.Medical.Social.ServicesY
23	StateAR
24	StateCA
25	StateFL
26	StateIL
27	StateLA
28	StateMI
29	StateOH
30	StateTX
31	Type.of.Ownershipproprietary
32	year.quarter201301
33	year.quarter201401
34	year.quarter201402
35	year.quarter201403
36	year.quarter201404
37	year.quarter201501
38	year.quarter201502

These *thirty eight* important variables provide valuable insights as follows:

- States of Texas, California, Florida, Ohio, Illinois provide high predictability for Scoring. These states also the largest representation in the data set as shown in the *Tableau EDA in section Insights Using Tableau*. However, Pennsylvania has a large representation but for some reason is not predictive of Scoring.
- With respect to the six different type of Services Offered, only one Service Offered is listed as important or predictive. This is the Medical and Social Services, both the Yes and the No categories are important.
- With respect the Year and Quarter, the year 2014 has all of its quarters as Important, but for 2013 only the first quarter is Important. Likewise for 2015, not all the quarters have been identified as important.
- As expected, the "Proprietary" Type of Ownership is identified as important or predictive. This is most prevalent type of ownership in the data as shown in Figure 6 of the *Insights Using Tableau*.
- There are seven Cities listed as Important predictors. Shown in Table 11 below, the city of Miami exists in three different States, however, given that the city in the State of Florida has the largest number of observations, it can be presumed that it is Miami, Florida the city with high predictive value.

  Three cities are in California but their number of observations is not high relative to Miami; however, those Cities have predictive influence over the Score or Target variable.

Table 11 - Important Cities with high predictive value

City	State	
alhambra	ca	101
miami	az	3
	fl	3,475
	ok	11
miami lakes	fl	387
northridge	ca	146
palmdale	ca	22
saint clair shores	mi	45
southfield	mi	1,192

4.5 Insights from Random Forest (Classification) on Important Variables using R

In this section, several models to predict the Quality Star Rating for Providers are considered. The models are multi-modal, meaning the target variable, Star Rating, has separate categories or distinct values ranging from 0 to 5.

Three classification models are fitted in this section: Random Forest, Support Vector Machine (Linear method), and Support Vector Machine (Radial Method).

Random Forest (RF) technique is a very widely used ensemble method that builds several classification trees on different subsets of data and uses majority voting for prediction.

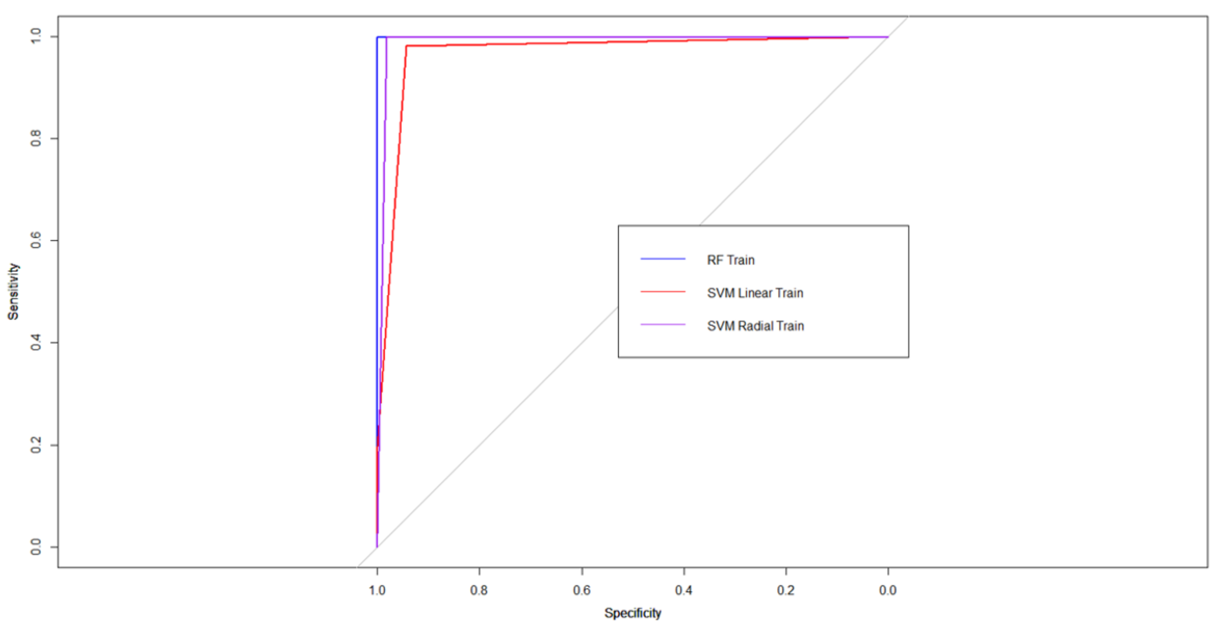
Support Vector Machines (SVM) utilize a technique which builds a hyperplane or a set of hyperplanes in a multi-dimensional space such that the margin between the data points is greatest. Two methods of SVM are explored: the linear and the Radial methods, which build linear lines and radial lines to separate the different classes.

The data of 107,000 rows are split into a training set and testing set on a 70/30 row-count split.

The results show that the RF model did exceedingly well when it was fitted on the Train data set to predict the ratings from the same data set. It had a 99.99% accuracy. The fitted model was evaluated on the test data set that we had withheld from training which meant that model had not seen this data. The accuracy of the RF model while predicting using the test data was 85.4%, which is quite remarkable. Another measure of the prediction is the Area-Under-the-Curve (AUC) which resulted at approximately between 0.9899 and 0.9686 for the training and testing data set, respectively.

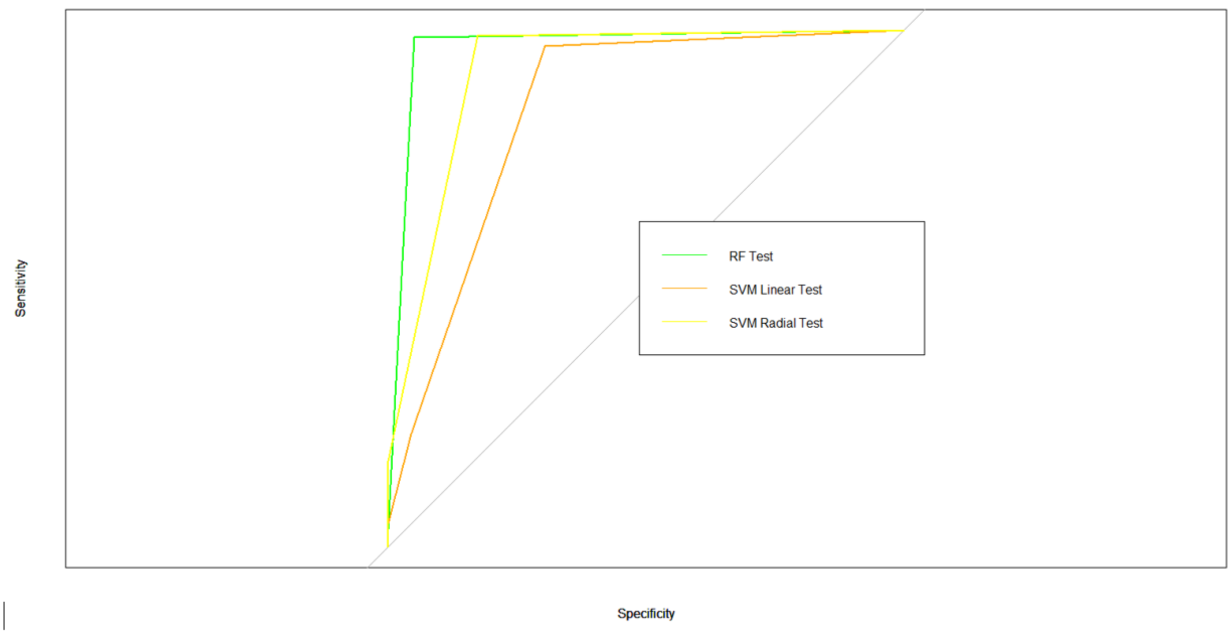
Additional evaluation of the model performance is based on how well the model does on a Receiving Operator Characteristics (ROC) curve plot. Referring to Figure 28 below, on the Training data set all three models are very close.

Figure 28 - ROC Curve based on training data set



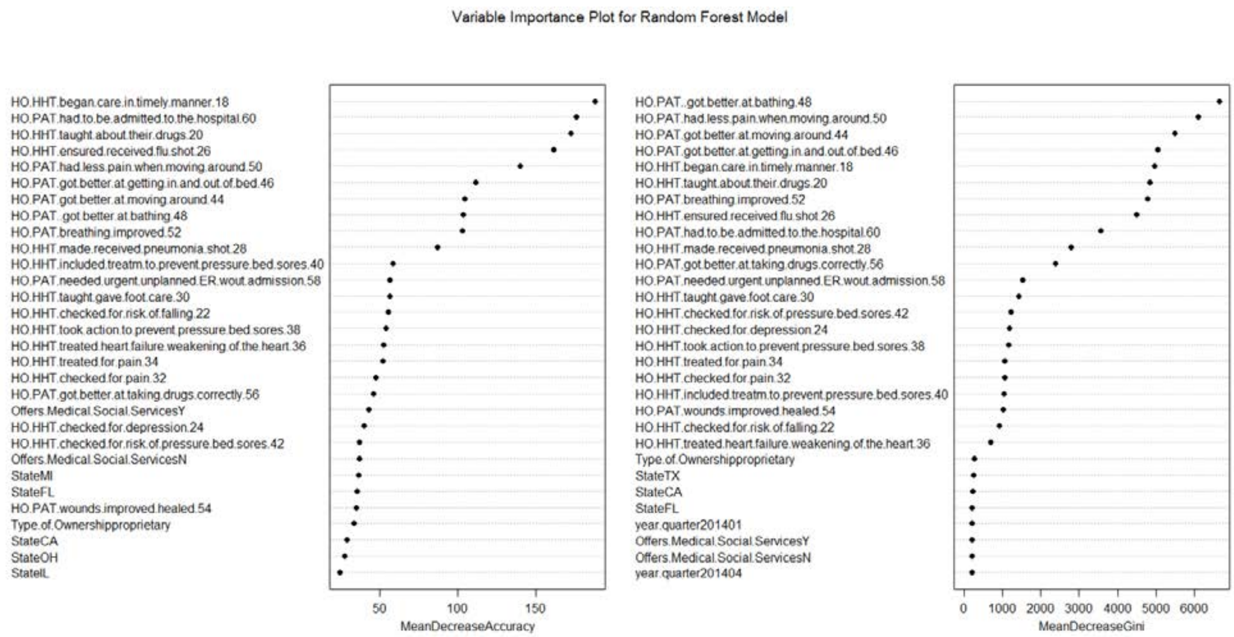
The three models are evaluated on the Test data set and the results of the ROC plots are shown in Figure 29 below. It can be clearly seen that the RF models performs better than the SVM models, with the Radial method SVM doing better than the Linear methods SVM.

Figure 29 - ROC Curve based on training data set



Judging by the accuracy and ROC plots we picked the Random Forest model to be the best among the three models built because it has the largest area above the diagonal line. The plot of variables by order of importance as adjudicated by the Random Forest Classification model is shown in Figure 30 below.

Figure 30 - Top 30 important variables from Random Forest (Classification) Model



Some of the variables are repeated across both panes. In order to obtain the distinct variables that have been deemed important by the RF model, the list of these variables was exported to an Excel file, the two lists were combined and extracted the distinct variable names. This process resulted in 47 distinct variables or predictors identified by the RF, which are shown in Table 12 below.

Table 12 – Distinct Important Variables from Random Forest (Classification) model

Row #	Predictor.RF.Classification
1	HO.HHT.began.care.in.timely.manner.18
2	HO.HHT.taught.about.their.drugs.20
3	HO.HHT.ensured.received.flu.shot.26
4	HO.PAT.got.better.at.moving.around.44
5	HO.PAT.got.better.at.getting.in.and.out.of.bed.46
6	HO.PAT..got.better.at.bathing.48
7	HO.PAT.had.less.pain.when.moving.around.50
8	HO.PAT.breathing.improved.52
9	HO.PAT.had.to.be.admitted.to.the.hospital.60
10	Cityalhambra
11	Citymiami
12	Citymiami.lakes
13	Citynorthridge
14	Citypalmdale
15	Citysaint.clair.shores
16	Citysouthfield
17	HO.HHT.checked.for.depression.24
18	HO.HHT.checked.for.pain.32
19	HO.HHT.checked.for.risk.of.falling.22
20	HO.HHT.checked.for.risk.of.pressure.bed.sores.42
21	HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40
22	HO.HHT.made.received.pneumonia.shot.28
23	HO.HHT.taught.gave.foot.care.30
24	HO.HHT.took.action.to.prevent.pressure.bed.sores.38
25	HO.HHT.treated.for.pain.34
26	HO.HHT.treated.heart.failure.weakening.of.the.heart.36
27	HO.PAT.got.better.at.taking.drugs.correctly.56
28	HO.PAT.needed.urgent.unplanned.ER.wout.admission.58
29	HO.PAT.wounds.improved.healed.54
30	Offers.Medical.Social.ServicesN
31	Offers.Medical.Social.ServicesY
32	StateAR
33	StateCA
34	StateFL
35	StateIL
36	StateLA
37	StateMI
38	StateOH
39	StateTX
40	Type.of.Ownershipproprietary
41	year.quarter201301
42	year.quarter201401
43	year.quarter201402
44	year.quarter201403
45	year.quarter201404
46	year.quarter201501
47	year.quarter201502

These *forty seven* important variables provide valuable insights as follows:

- If we look the important variables in decreasing order of Gini index, we will find 4 of the top four important variables are outcomes related. As we have found with distribution of Quality Star Rating for these outcomes related measures, patients and their relatives tend to rate the providers higher if they see better outcomes.
- States of TX, CA and FL have the highest number of cases of Home Health Care Providers so they are important variables to consider in the prediction model.
- Among all the services offered, offering medical social services in considered the most influential in determining a good provider.
- The type of ownership – Proprietary which is deemed important is also the largest type of ownership among the providers we analyzed in this data set. During EDA we found that providers of this type of ownership do a very good job of taking care of the process related measures. The large size of these types of providers could be the reason why this is an important variable but we certainly do not have proof that this type of ownership contributes to distinguishing between a good and excellent provider.

4.6 Insights on Historical Overall Mean of Star Rating for each Provider.

The EDA performed up to this point and illustrated in the sections above has been based on the data preparation in section *3.8 Score Agencies based on the Average Bin of the Nine Ranking-Based Measures*.

The Final preparation of the data resulted in the assignment of the historical mean for the Star Rating to each HHCA. This process is described in section *3.9 Assign Historical Overall Mean of Star Rating to each Provider*. The analysis and insights on this FINAL data preparation are explained below.

We created a data set that has distinct rows per CMS Certificate Number and any combination of Type of Ownership, Zip and other descriptors. We used Tableau to visualize the characteristics of providers within the context of their Overall Star Rating.

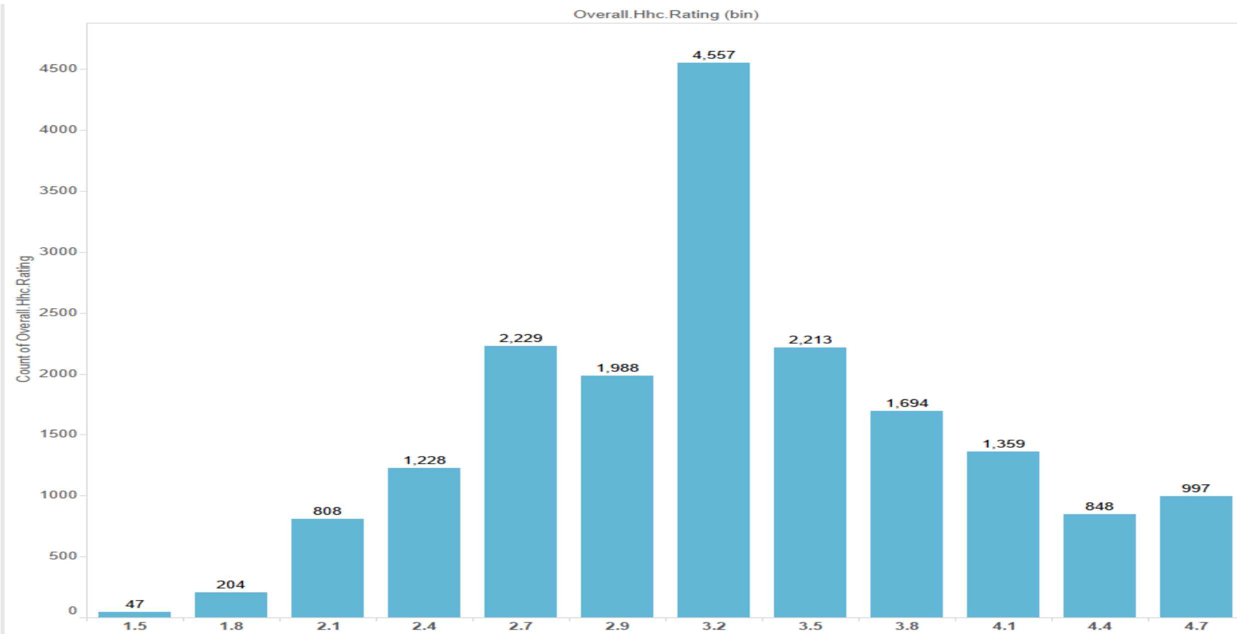
The Descriptive Statistics for the assignment of the Historical Mean Star Rating for each Provider, which is based on the historical rating from 11 quarters and the nine measures shown in Table 13 below, indicate that the Average Rating at 3.43 is slightly lower than the Median at 3.50, indicating that Ratings are not skewed and the distribution can be considered normal. However, the Variance is close to the 0.50 point, which is the unit of increase to obtain a half a Star Rating. Also, this distribution Table 8 above, which is based on the individual Quarterly Ratings.

Table 13 – Descriptive Statistics of the Distinct Provider Names and their Historical Mean Star Rating

Observations	Min.	Percentile (25)	Avg.	Median	Percentile (75)	Max.	Variance	Std. dev.
18,172	1.50	3.00	3.43	3.50	4.00	5.00	0.48	0.70

The distribution of the Star Rating for all Providers shown in Figure 30.5 below is normally distributed with no outliers.

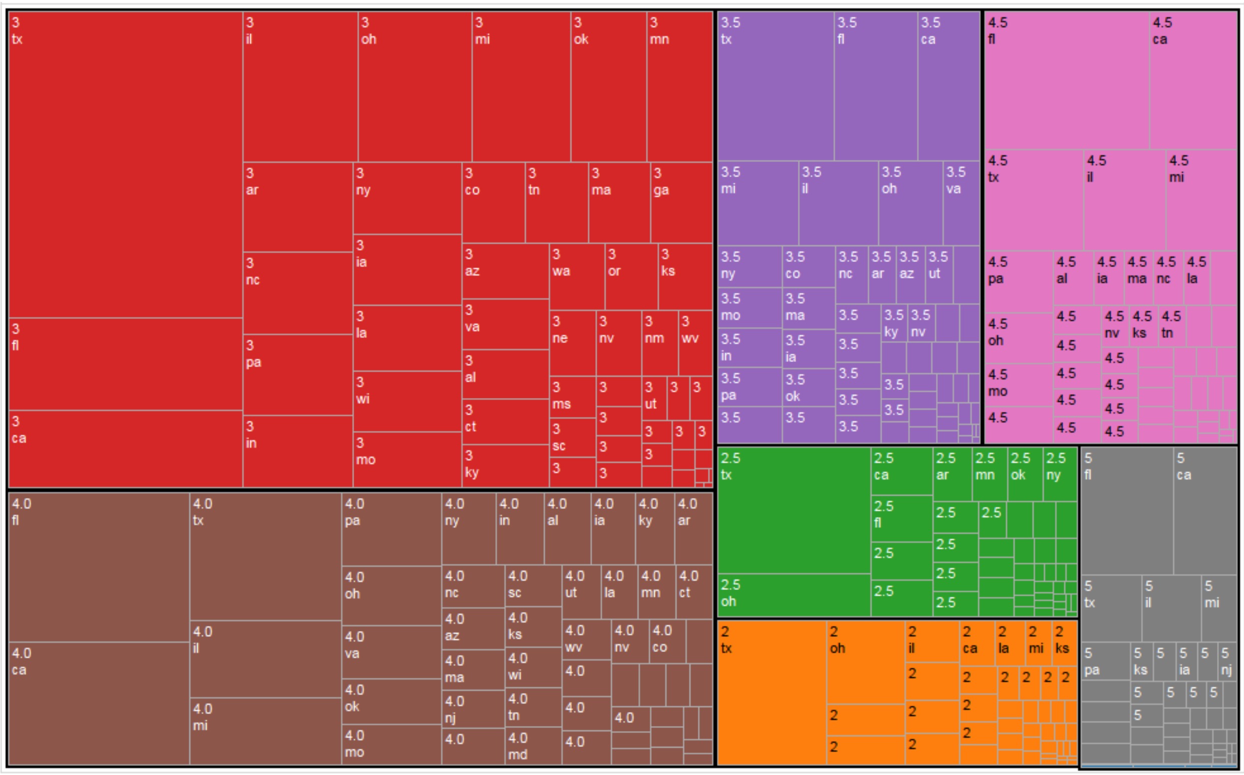
Figure 30.5 – Distribution of Overall Star Rating by Star Rating



The heat-map shown in Figure 30.6 below visualizes the proportion of observations in the data set by Star Rating and State. It can be clearly seen that the Star Ratings 3.0 and 4.0 have the highest proportion of the observations in colors red and brown. The least number of observations are for Star Rating of 5.0 in the grey block.

Also, it can be clearly seen that Texas dominates the lower Star Ratings of 2.0, 2.5, 3.0, 3.5 and Florida dominates the higher Star Ratings of 4.0, 4.5, and 5.0 and California is second to Florida in the Star Ratings 4.5 and 5.0.

Figure 30.6 – Proportion of the Number of Observations by 'State' within Each Star Rating



The stack-bar plot created in Tableau and shown in Figure 30.7 below visualizes the proportion of observations in the data set by Star Rating and 'Type of Ownership'. It can be clearly seen that 'proprietary' (red) HHCAs dominate all Star Ratings, however, the next high proportions are for 'private' (light-green), 'voluntary non-profit' (light-brown), and finally 'other' (dark-green).

Beginning on Star Rating 3.2 and increasing, it can be seen that the 'proprietary' portion of the Star Rating is reduced due to the higher Star Ratings of the other Types of Ownership and of particular interest is the 'Voluntary non-profit' group. This legal type of HHCA seems to be more prevalent in the higher Star Ratings than in the lower ones. It would be of interest to make a similar comparison of the Types of Ownership by Cost and see how they stack up. I would suspect that the 'Voluntary non-profit' would have the least proportion of the Cost yet hold a relatively high proportion of the high performing Star Rating HHCAs, but that hypothesis can be answered in another project.

Figure 30.7 – Proportion of the Number of Observations by 'Type of Ownership' within Each Star Rating

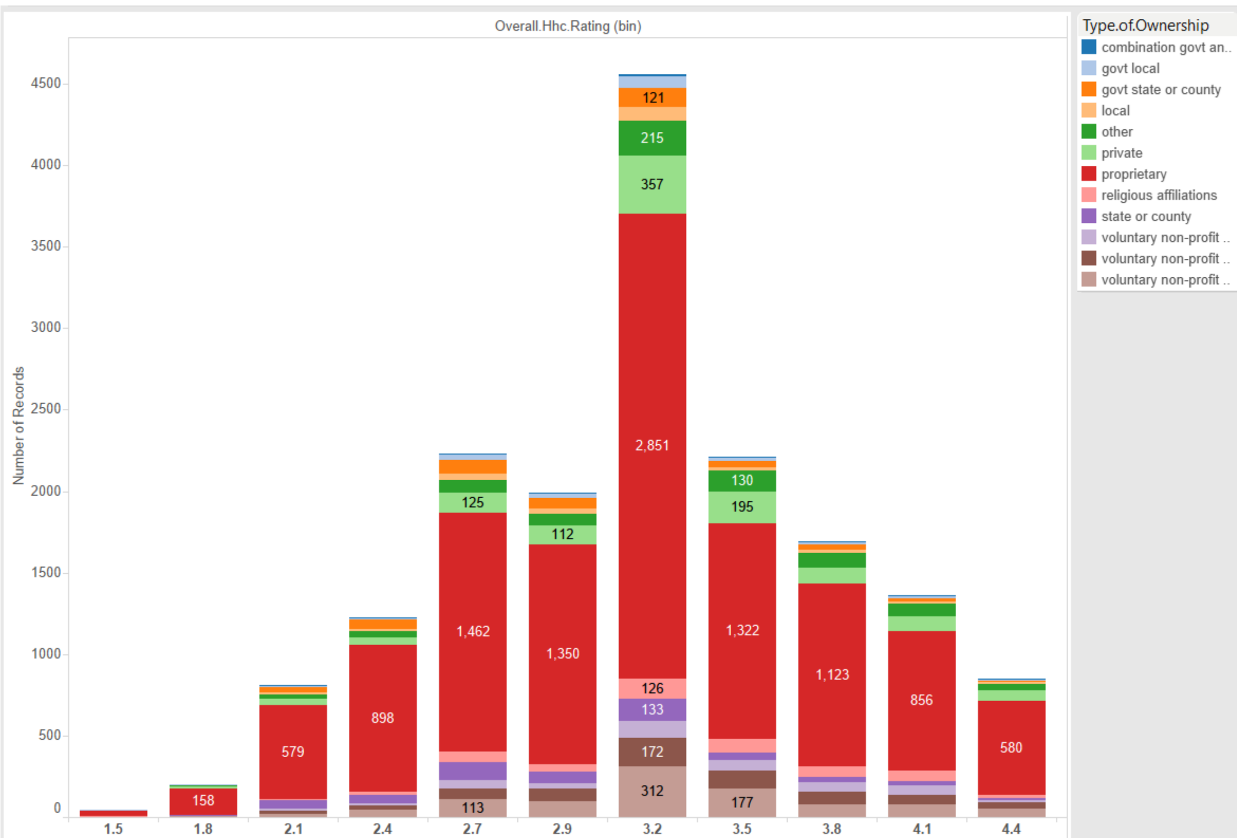
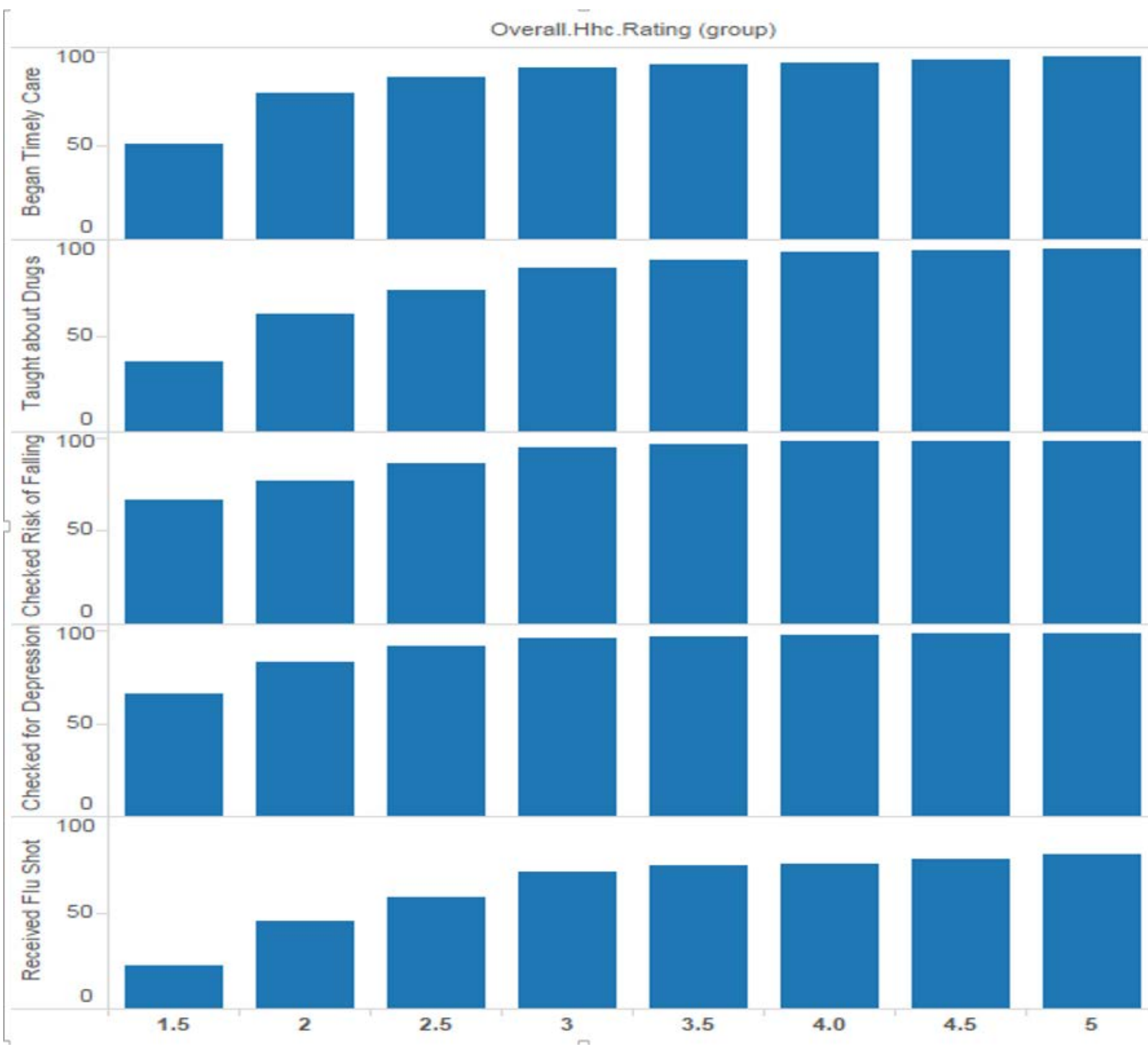


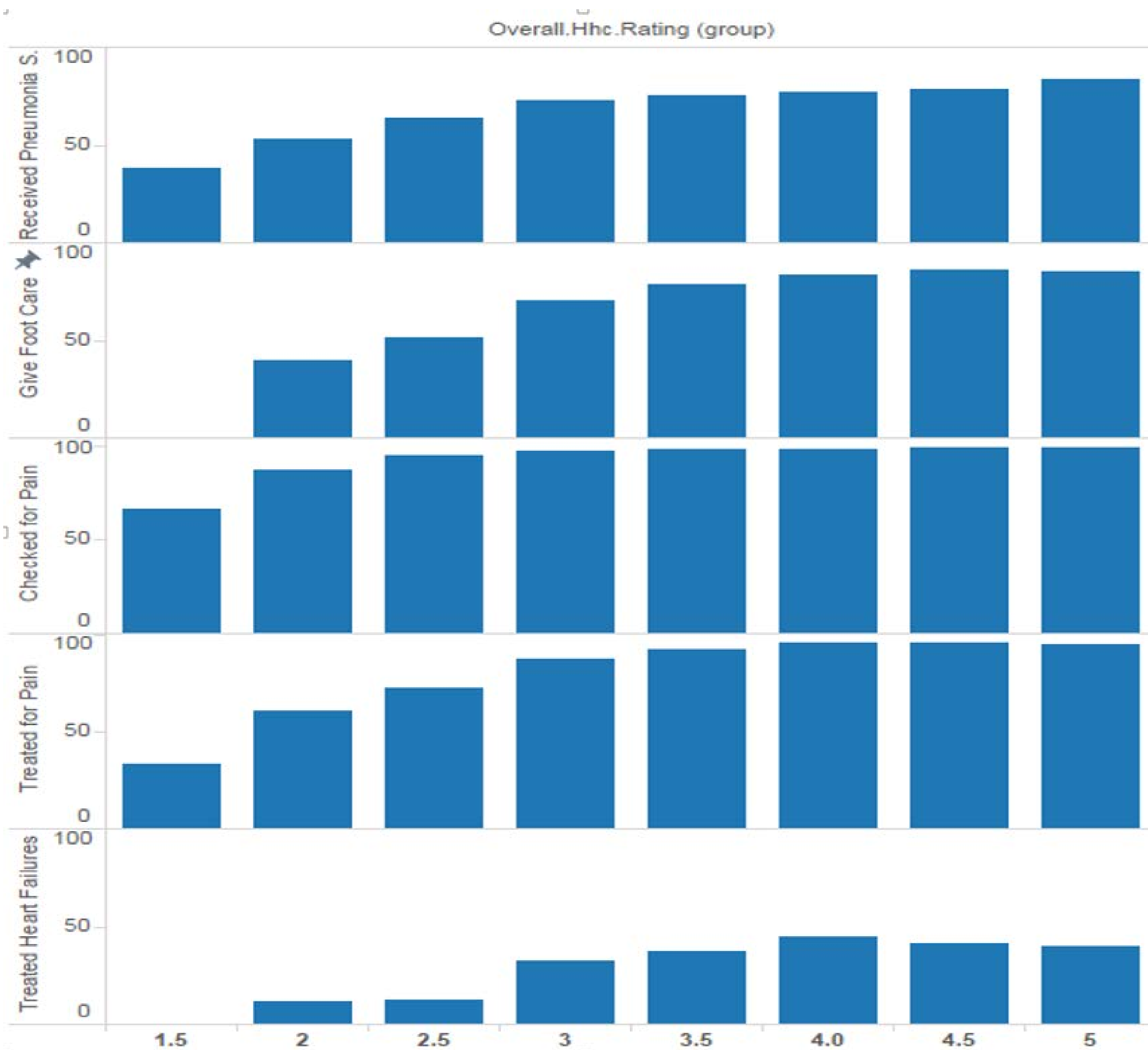
Figure 31 below, shows the distribution of the Overall Star Rating for 5 variables which measure processes – Timely Care, taught about their drugs, check for risk of falling, checked for depression, and ensured received flu shot. Providers who received an overall star rating of 3 or lower seem to doing a bad job of managing their processes. However, these five variables do not provide much insight to distinguish between providers who earned more than 3 stars.

Figure 31 – Distribution of Overall Star Rating by measures - Timely Care, Taught about Drugs, check for risk of falling, checked for depression, and received flu shot



In figure 32 below, we plot five other measures which also represent processes – received pneumonia shot, taught how to give foot care, checked for pain, and treated for heart failures. Similar to figure 31 above, figure 32 below shows that providers need to improve their processes if they intend to improve their star rating. It seems that these variables are not significant in explaining how a provider can improve if they have a rating of 3.5 or 4.

Figure 32 – Distribution of Overall Star Rating by measures - received Pneumonia shot, taught how to give foot care, checked for pain, treated for pain, and treated for heart failure.



In figure 33 below, we plot the distribution of the star rating for three other process driven and two outcome driven measures – action to prevent bed sores, include treatment to prevent bed sores, checked for risk of pressure bed sores, got better at moving, and got better at getting in and out of bed. We see that providers who received good scores in the two outcome driven measures had better overall star rating. Patients and their relatives probably rate a provider well if the outcomes are good.

Figure 33 – Distribution of Overall Star Rating by measures - Action to prevent bed sores, treatment to prevent bed sores, checked for pressure bed sores, got better at moving and got better at getting in and out of bed.

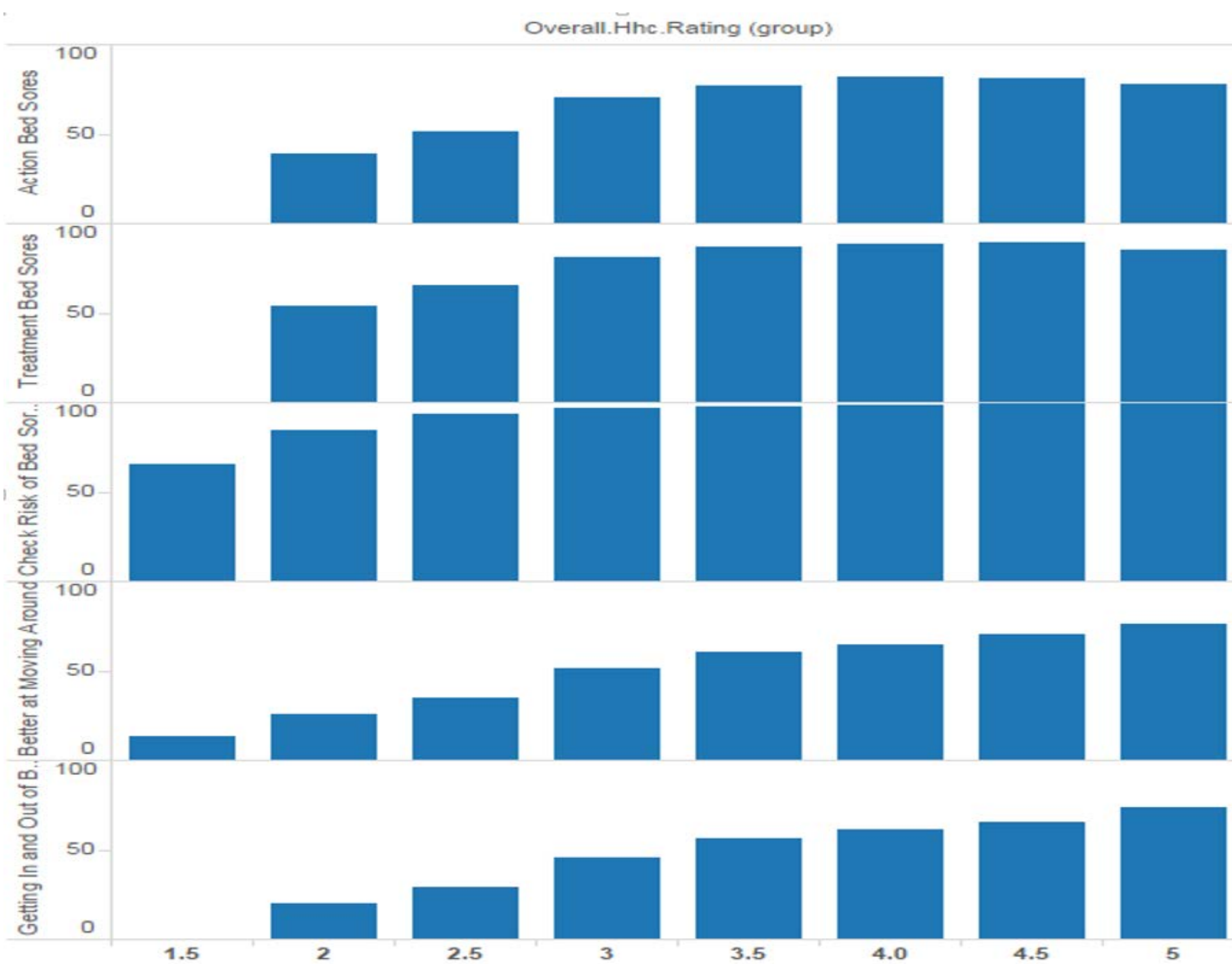


Figure 34 below is a plot of more outcomes driven measures – got better at bathing, had less pain moving around, breathing improved, wounds improved healed, and got better at taking drugs. Just as we had noticed with two outcomes related measures in figure 33 above, these outcome related measures also show steady increase in rating as the providers did well in these measures.

Figure 34 – Distribution of Overall Star Rating by measures – got better at bathing, had less pain moving around, breathing improved, wounds improved, and got better at taking drugs.

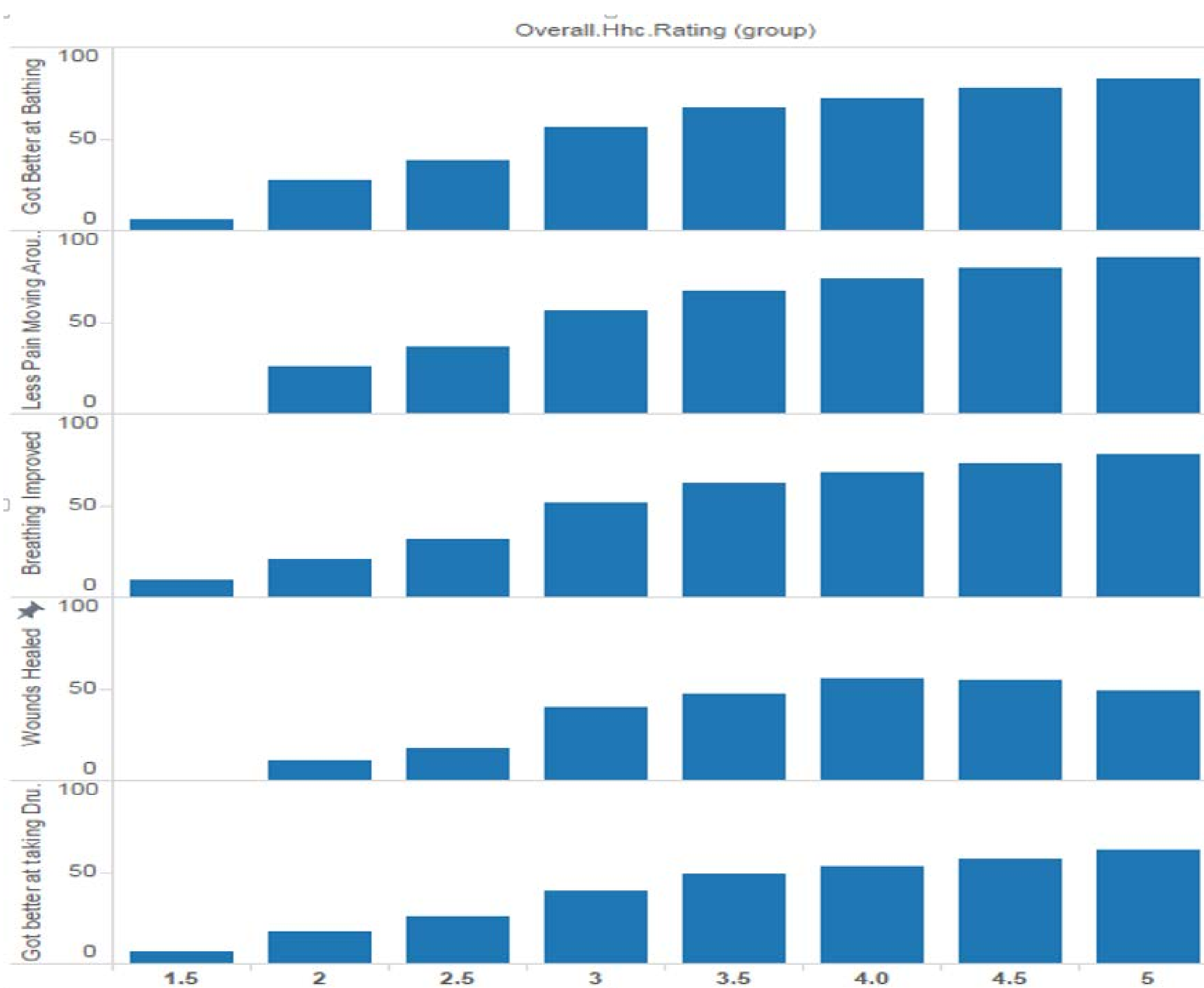
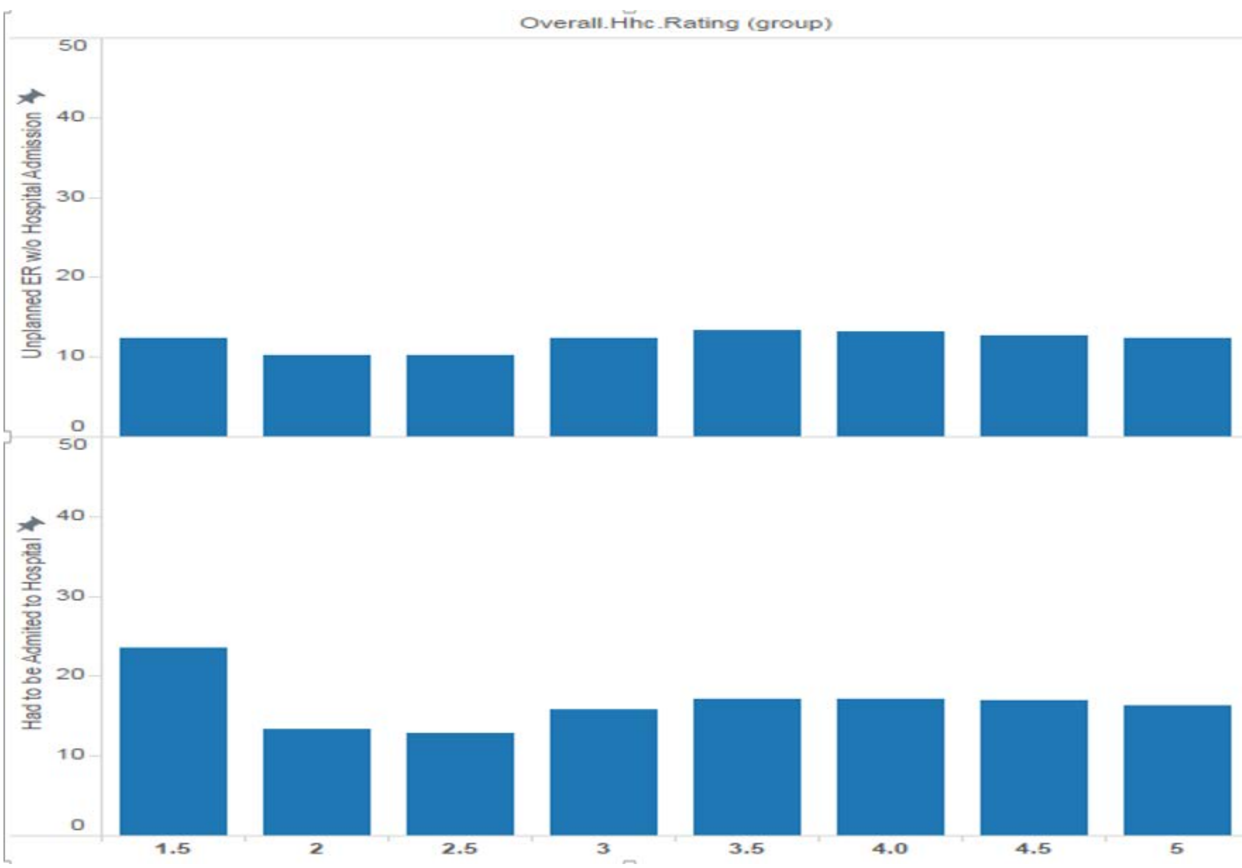


Figure 35 below shows the distribution of the overall star rating for outcomes driven measures which measure if the patient had to be taken to the ER or had to be admitted to the hospital. So, a low measure means higher quality rating for the provider because patients did not need to be taken to the ER. Patients who had to be taken to the ER or had to be admitted to the hospital seem to rate the provider poorly.

Figure 35 – Distribution of Overall Star Rating by measures – had to be taken to ER without admission to hospital and had to be admitted to the hospital.



4.7 Insights from Lasso (Regression) on Explanatory Variables using R

The Lasso algorithm available in the 'glmnet' R package has been the only machine learning technique implemented for this paper that is capable of processing the 'Provider Name' categorical variable together with the rest of the variables, except for the Zip Code which was excluded for analysis reasons.

The Lasso performed beautifully with 14,714 dichotomized and measure variables and fitting a model with about 19,000 observations in less than 1 minute. In addition, the Lasso provided estimates that can be used by the recommendation engine (unlike the other techniques used in this paper) for each of the predictors it found to have explanatory influence on the target variable, which is the Star Rating for the Provider. This is truly a remarkable accomplishment and discovery in terms of real practical value of a machine learning algorithm and for the recommendation engine.

The Lasso method seeks coefficient estimates that minimize the difference between the observed outcome and the predicted outcome for each variable. However, this method uses a "tuning parameter" that controls the impact of a "shrinkage penalty" on the estimated association between each variable and the response or outcome variable making it fundamentally different than the 'least squares' method when it is applied by itself in traditional statistical learning regression and classification methods. (7)

The Lasso regression also provides flexibility in the reduction of the variance that is introduced to modeling when working with high-dimensional data. The tuning parameter acts as a 'dialing-knob' to control the variance in the model and therefore the accuracy of the estimated coefficients for each variable. Depending on the tuning parameter value, the Lasso can produce a model involving any number of variables so it is closely related to best-subset variable selection resulting in models that are easy to interpret. (7)

Also, the Lasso method reduces the time and effort in data exploration and data preparation because it automatically standardizes the data, takes in all of the dummy variables from the dichotomization of categorical variables including the reference or base category, and it computes coefficient for data sets where the number of variables exceeds the number of observations (high-dimensional data) in a computational feasible way using the tuning parameter and the shrinkage penalty.

For this analysis, the modeling using Lasso was performed using the final data set created in section *3.9 Assign Historical Overall Mean of Star Rating to each Provider*, which has the derived target variable representing the historical mean of the Star Rating at the HHCA level. This data set contains 19,371 observations and 35 categorical and numerical variables.

The Lasso model requires the data to be dichotomized in a matrix structure. The function build.x of the package 'useful' in R was used to create this matrix. This process resulted in 14,714 variables due to the dichotomization of the categorical variables, the reference dummy variable is not removed from the data set.

This is a high-dimensional data set and based on the mathematical techniques of the Lasso, the model fitting executes very quickly, in less than 1 minute with a TRAIN data set based on a random sample of 75% of the 19,371 observations.

4.7.1. Lasso Model - Evaluation

At the best lambda or tuning parameter, the Lasso model identified 201 explanatory variables with coefficients <> 0 out of the 14,714 dichotomized and numeric variables. These variables were identified due to their highest explanatory value on the target variable, the historical mean Star Rating of the Provider.

The evaluation of the model using the TEST set resulted in in a Mean Square Root Error of 0.198832 using the 25% random sample TEST including all variables.

Table 14 below shows the Descriptive Statistics from the Evaluation for the Lasso fitting and Predictions. The *Rate Change* (highlighted in yellow below), clearly shows that there is not much difference between the *Observed Outcomes* and the *Predicted Outcomes* using the TEST data set, in fact, the Predictions statistics are closer to Table 8 or Table 13, the descriptive statistics for the entire population, than are the Observed statistics.

The *mean Rate Change* is only 0.0021 or 0.2%, which is extremely low. The *median Rate Change* is 8% higher for the Predictions, which is also very low. However, the *Min* and the *Max Prediction* values are outside of the range, meaning that a Star Rating above 5.0 or below 1.50 is not part of the original deciles, so those predictions would have to be capped.

Table 14 - Lasso Model Evaluation, Descriptive Statistics

	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Predictions	1.2150	3.2000	3.5090	3.4130	3.7300	5.2500
Observed	1.5000	2.7500	3.2500	3.4060	4.0000	5.0000
Delta (pred - obser)	-0.2850	0.4500	0.2590	0.0070	-0.2700	0.2500
Rate Change of the Predictions *over* the Observed:(Delta / Observed)	**-0.1900**	**0.1636**	**0.0797**	**0.0021**	**-0.0675**	**0.0500**

These findings indicate that this Lasso model seems to be an **extremely USEFUL model** to the understanding of how each of the 201 explanatory variables can be used to predict the Start Rating for Providers and thus can be recommended to be used for the predictive model of the Recommendation Engine.

4.7.2. Lasso Model – Analysis and Insights

We start our analysis of the Lasso coefficients by first analyzing the value it provides for individual categorical values such as the Provider Names, followed by the State analysis, then the City, and lastly the Types of Ownership. In this way, we can provide a good sense of how the Lasso results provide value for different practical aspects of the data.

Of *NOTE* in this segmented analysis is that the specific coefficients discussed below are evaluated with the premise that the rest of the predictors are zero, in this way there is no interaction of other predictors influencing the target variable, Star Rating, except for the predictors analyzed in the commentary.

With respect to the technical implementation of the Lasso results, we provide the complete list of the Explanatory Variables in *Addendum 2 – Lasso Estimates for all 201 Explanatory Predictors*, which shows the 201 predictors and their coefficients at the best lambda.

Of note, the Zip Code was processed as a numerical variable by the Lasso technique and it returned a coefficient of zero or not important. The Zip Code was not fitted as a categorical variable because it would have been too granular to provide insight across a wider range of observations. The 'Provider Name' is the lowest level fitted.

4.7.2.1. Lasso Model – Analysis on Provider Name

The original data set downloaded from CMS contained 13,119 distinct Provider Names, however after the data preparation process, only 10,371 distinct Provider Names remained in the FINAL data set as prepared in section *3.9 Assign Historical Overall Mean of Star Rating to each Provider*. Of these, there were 103 distinct Provider Names identified by the Lasso model which provide explanatory value on the Star Rating.

The Lasso coefficient or estimate for the Star Rating of each Provider Name was merged with the FINAL data set prepared in section 3.9, which contains the CMS Number, the Overall Historical Mean Star Rating for each Provider, the State, City, and the Type of Ownership to create a data set that can be the basis for the Recommendation Engine.

The proposed data set for the Recommendation is shown below in two tables. The first table, Table 15, shows the top HHCAs based on the largest value of the Lasso coefficient or estimate. The second table, Table 16, shows the bottom HHCAs based on the lowest Lasso coefficient.

The complete list of Lasso estimates for all of the HHCas or Providers is located in *Addendum 1 - Lasso Estimates for all HHCas or Providers*.

In Table 15, we can provide the patient with the Historical Mean Star Rating for a Provider together with an estimate (green bold font) indicating that the Provider can be estimated to have a Star Rating <estimate value> higher. For example, for the first Provider listed, Cari Home Care Inc, it is estimated to have a 0.77 Star Rating higher, given that all other predictors remain are zero, which is almost two Stars.

Table 15 - Home Health Care Agency (Provider) with *highest* Lasso estimates

Provider_Name	CMS_Cer_Number_CCN	Overall_Hhc_Rating	Lasso Coefficient or Estimate	City	State	Type_of_Ownership
CARI HOME CARE INC	108034	4.75	0.7732	homestead	FL	proprietary
VOLUNTEERS OF AMERICA HOME HEALTH AT WESTCHESTER	37265	4.50	0.5804	tempe	AZ	proprietary
DYNAMIC HOME HEALTH CARE	497655	4.25	0.5114	chantilly	VA	proprietary
SIERRA HOME HEALTH CARE	239066	4.25	0.4920	dearborn	MI	proprietary
SIERRA HOME HEALTH CARE	239066	4.25	0.4920	westland	MI	proprietary
SIERRA HOME HEALTH CARE	58022	4.17	0.4920	fontana	CA	proprietary

Likewise, but in the reverse, Table 16 below shows the Provider who have a negative estimate, indicating that, all other predictors being 0, the given Provider is expected to have a Star Rating <estimate value> lower. For example, for the last Provider listed, Angel Care Home Health Services Inc., it is estimated to have a -0.80 Star Rating lower, given that all other predictors remain are zero, which is almost two Stars.

Table 16 - Home Health Care Agency (Provider) with *lowest* Lasso estimates

Provider_Name	CMS_Cer_Number_CCN	Overall_Hhc_Rating	Lasso Coefficient or Estimate	City	State	Type_of_Ownership
PRUDENT HEALTHCARE AGENCY	747352	2.25	-0.4987	mesquite	TX	proprietary
WISDOM HEALTH CARE SERVICES INC	59060	2.50	-0.5034	gardena	CA	private
WISDOM HEALTH CARE SERVICES INC	59060	2.50	-0.5034	gardena	CA	voluntary non-profit private
REGISTERED NURSES CARE LTD	368071	2.75	-0.5214	westerville	OH	proprietary
GOOD SAMARITAN HOME HEALTH CARE INC	679478	2.50	-0.7342	dallas	TX	proprietary
ANGEL CARE HOME HEALTH SERVICES INC	459412	2.25	-0.8029	grand prairie	TX	proprietary

4.7.2.2. Lasso Model – Analysis on State

Of 55 distinct State categories in the data set, the Lasso model identified 21 States as having explanatory value on the Star Rating for Home Health Agencies, refer to Table 17 below.

The State of PR, which is Puerto Rico, provides the best State for selecting a Home Health Agency with 0.41 higher Star Rating estimate than all other States, provided that all other predictors do not change. Puerto Rico rates extremely high compared to the next lower estimates identified by the Lasso, which are for Florida at 0.15, California at 0.12. The other States with increasing estimates are not very high at less than 0.10.

The worse State to select a Home Health Agency is Minnesota at -0.16 of an estimated Star Rating, given that all other predictors do not change. However, the negative estimates for the other States are not high in magnitude, ranging from -0.00539 to -0.07170, or less than 8%.

Table 17 - Lasso Coefficients on State

	Predictor Name	Coefficient	When the Predictor is 1, the Star Rating increases or decreases by this quantity, given that all other predictors remaining the same
1	State-PR	0.41289	0.41289
2	State-FL	0.14701	0.14701
3	State-CA	0.11675	0.11675
4	State-NJ	0.08856	0.08856
5	State-SD	0.04571	0.04571
6	State-UT	0.04473	0.04473
7	State-MI	0.03949	0.03949
8	State-MD	0.01082	0.01082
9	State-IL	0.00872	0.00872
10	State-PA	0.00253	0.00253
11	State-OK	-0.00539	-0.00539
12	State-AK	-0.02965	-0.02965
13	State-LA	-0.03065	-0.03065
14	State-AR	-0.03250	-0.03250
15	State-IN	-0.03947	-0.03947
16	State-NC	-0.04275	-0.04275
17	State-TX	-0.04351	-0.04351
18	State-WA	-0.05904	-0.05904
19	State-OH	-0.06460	-0.06460
20	State-OR	-0.07170	-0.07170
21	State-MN	-0.16024	-0.16024

4.7.2.3. Lasso Model – Analysis on City

Of 3,293 distinct City categories in the data set, the Lasso model identified 53 Cities as having explanatory value on the Star Rating for Home Health Agencies, refer to Table 18 below.

The cities of Hayesville, NC and Sikeston, MO provide the best cities for selecting a Home Health Agency with 0.40 and 0.25 higher Star Rating estimate, respectively, than all other cities, provided that all other predictors do not change. There are 17 cities in 11 different

States providing HHCAs with 0.20 and 0.10 higher Star Rating estimates, which means that these explanatory predictors can be useful in many States.

The two worse cities are Osceola, AR or IA, and Brooklyn, NY at -0.17 and -0.15, respectively. Further analysis needs to be performed for the city of Osceola to determine which of these two States is the estimate referring.

Table 18 - Lasso Coefficients on City

	Predictor Name	Coefficient	When the Predictor is 1, the Star Rating increases or decreases by this quantity, given that all other predictors remaining the same
1	City - HAYESVILLE, NC	0.39558	0.39558
2	City - SIKESTON, MO	0.25132	0.25132
3	City - SIOUX CITY, IA	0.18569	0.18569
4	City - MIAMI, FL	0.18258	0.18258
5	City - RHINELANDER, WI	0.17789	0.17789
6	City - JEFFERSON CITY, MO	0.16672	0.16672
7	City - CREIGHTON, NE	0.16161	0.16161
8	City - MARSHFIELD, WI	0.15322	0.15322
9	City - ARDMORE, OK	0.14758	0.14758
10	City - STURGEON BAY, WI	0.13934	0.13934
11	City - BEVERLY, MA	0.13365	0.13365
12	City - SPRING HILL, FL	0.13323	0.13323
13	City - SANTA CRUZ, CA	0.12235	0.12235
14	City - WAVERLY, IA or OH	0.12007	0.12007
15	City - PANORAMA CITY, CA	0.11964	0.11964
16	City - TUSCALOOSA, AL	0.11887	0.11887
17	City - MANITOWOC, WI	0.11600	0.11600
18	City - NEW BOSTON, OH	0.10996	0.10996
19	City - NORRISTOWN, PA	0.10743	0.10743
20	City - FESTUS, MO	0.07219	0.07219
21	City - HIALEAH, FL	0.06832	0.06832
22	City - GALESBURG, IL	0.06269	0.06269
23	City - CARLSBAD, CA or NM	0.05528	0.05528
24	City - WAUSAU, WI	0.05183	0.05183
25	City - DEL CITY, OK	0.04893	0.04893
26	City - LACONIA, NH	0.03696	0.03696
27	City - GRAND RAPIDS, MI or MN	0.02595	0.02595
28	City - TYLER, MN or TX	0.02534	0.02534
29	City - GLENDALE, AZ or CA or CO	0.02448	0.02448
30	City - SWEETWATER, TN or TX	0.02358	0.02358
31	City - BEATRICE, NE	0.02298	0.02298
32	City - PIERRE, SD	0.01158	0.01158
33	City - EGG HARBOR TOWNSHIP, NJ	0.00786	0.00786
34	City - MIAMI LAKES, FL	0.00498	0.00498
35	City - HATTIESBURG, MS	0.00406	0.00406
36	City - SOUTHFIELD, MI	0.00171	0.00171
37	City - LYNBROOK, NY	0.00077	0.00077
38	City - POMPTON PLAINS, NJ	0.00061	0.00061
39	City - NORTH KANSAS CITY, MO	0.00050	0.00050
40	City - KENOSHA, WI	0.00049	0.00049
41	City - BROWNFIELD, TX	0.00034	0.00034
42	City - CHISAGO CITY, MN	0.00029	0.00029
43	City - CORTLAND, NY	0.00023	0.00023
44	City - BELFAST, ME	0.00005	0.00005
45	City - ROCKINGHAM, NC	0.00000	0.00000
46	City - SAN ANTONIO, TX	-0.00674	-0.00674
47	City - SHERMAN, TX	-0.00939	-0.00939
48	City - CHANDLER, AZ or OK	-0.02435	-0.02435
49	City - CLEVELAND, MS or OH or OK or TN or TX	-0.03949	-0.03949
50	City - HUDSON, FL or NY or WI	-0.06403	-0.06403
51	City - VANCOUVER, WA	-0.07317	-0.07317
52	City - OSCEOLA, AR or IA	-0.14902	-0.14902
53	City - BROOKLYN, NY	-0.17130	-0.17130

4.7.2.4. *Lasso Model – Analysis on Type of Ownership*

Of twelve distinct ‘Type of Ownership’ values in the data, the Lasso model identified five Types of Ownership as having explanatory value on the Star Rating for Home Health Agencies, refer to Table 19 below.

Of the five Types of Ownership identified by the Lasso, three are Voluntary and those three were identified by the Lasso model as having positive or increasing explanatory value on the Star Rating for a Provider. This is indeed insightful as it conveys that the best caring for humans is from the heart and this seems to be ‘better’ or ‘more positively’ accomplished in a Voluntary legal setting.

The ‘caring from the heart’ aspect is confirmed by seeing that Proprietary and State or County owned HHCAs have a negative or decreasing explanatory influence on the Star Rating for Providers.

Table 19 - Lasso Coefficients on Type of Ownership

	Predictor Name	Coefficient	When the Predictor is 1, the Star Rating increases or decreases by this quantity, given that all other predictors remaining the same
1	Type.of.Ownership - voluntary non-profit church	0.04867	0.04867
2	Type.of.Ownership - voluntary non-profit other	0.03494	0.03494
3	Type.of.Ownership - voluntary non-profit private	0.01239	0.01239
4	Type.of.Ownership - proprietary	-0.02328	-0.02328
5	Type.of.Ownership - state or county	-0.02756	-0.02756

4.8 Explanatory Predictors for the Recommendation Engine

We now consolidate, rationalize, and rank the explanatory variable or important variables to create a list of variables for the recommendation engine.

The application of the different predictive models resulted in a unique list of 217 explanatory predictors out of 13,856.

The three principal models that provided these explanatory predictors are Random Forest Classification, Random Forest Regression, and Lasso. However, Lasso was the only model that incorporated the Provider Name in the fitting.

These 217 predictors are ranked to reflect the number of models in which they were identified to be predictive of the Star Rating of the Provider.

The Ranking process required the following steps:

1. Create the list of explanatory predictors from the 3 different models. A new column created for each list called "Model.Source" to identify the model the predictor is from.

2. Union the 3 lists. They were put in a union data structure way.

3. Pivot the Unioned List of predictors and their model source to identify the different models each predictor is from. The prefix of "Max.of." for some of the predictors from the Lasso model are removed.

4. Each predictor is assigned a number from 1 to 3 to indicate how many models that predictor is found in. This is basically Ranking each predictor with the number of models it appears in, the more models, the higher the ranking.

5. Pivot the data from #4 above to list the Ranking and their respective predictors.

6. Do a final sorting by the Rank from 3 to 1 and showing the Predictors sorted in alpha order.

Table 20 below shows the results of the Ranking. There are 217 distinct predictors identified out of 13,856 that have explanatory value on the Star Rating of Home Health Care Agencies. These are the predictors that are recommended to be used for the recommendation engine.

Out of the 217 explanatory variables, 19 resulted from 3 different models, RF Regression, RF Classification, and Lasso.

Out of the 217 explanatory variables, 28 resulted from 2 different models. And the rest, 171 predictors, resulted from only one model. However, please note that the Provider Name was fitted in only one model, the Lasso, because the other models kept running into memory issues by having so many dichotomized dummy variables from the Provider Name.

Table 20 - Ranked Explanatory Predictors

	Predictor	**In how many models does the predictor appear?**
1	HO.HHT.checked.for.pain.32	**3**
2	HO.HHT.checked.for.risk.of.falling.22	**3**
3	HO.HHT.checked.for.risk.of.pressure.bed.sores.42	**3**

4	HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40	**3**
5	HO.HHT.made.received.pneumonia.shot.28	**3**
6	HO.HHT.treated.for.pain.34	**3**
7	HO.HHT.treated.heart.failure.weakening.of.the.heart.36	**3**
8	HO.PAT.got.better.at.taking.drugs.correctly.56	**3**
9	HO.PAT.needed.urgent.unplanned.ER.wout.admission.58	**3**
10	HO.PAT.wounds.improved.healed.54	**3**
11	StateAR	**3**
12	StateCA	**3**
13	StateFL	**3**
14	StateIL	**3**
15	StateLA	**3**
16	StateMI	**3**
17	StateOH	**3**
18	StateTX	**3**
19	Type.of.Ownershipproprietary	**3**
20	Cityalhambra	**2**
21	Citymiami	**2**
22	Citymiami.lakes	**2**
23	Citynorthridge	**2**
24	Citypalmdale	**2**

25	Citysaint.clair.shores	**2**
26	Citysouthfield	**2**
27	HO.HHT.began.care.in.timely.manner.18	**2**
28	HO.HHT.checked.for.depression.24	**2**
29	HO.HHT.ensured.received.flu.shot.26	**2**
30	HO.HHT.taught.about.their.drugs.20	**2**
31	HO.HHT.taught.gave.foot.care.30	**2**
32	HO.HHT.took.action.to.prevent.pressure.bed.sores.38	**2**
33	HO.PAT..got.better.at.bathing.48	**2**
34	HO.PAT.breathing.improved.52	**2**
35	HO.PAT.got.better.at.getting.in.and.out.of.bed.46	**2**
36	HO.PAT.got.better.at.moving.around.44	**2**
37	HO.PAT.had.less.pain.when.moving.around.50	**2**
38	HO.PAT.had.to.be.admitted.to.the.hospital.60	**2**
39	Offers.Medical.Social.ServicesN	**2**
40	Offers.Medical.Social.ServicesY	**2**
41	year.quarter201301	**2**
42	year.quarter201401	**2**
43	year.quarter201402	**2**
44	year.quarter201403	**2**
45	year.quarter201404	**2**

46	year.quarter201501	**2**
47	year.quarter201502	**2**
48	Cityardmore	**1**
49	Citybeatrice	**1**
50	Citybelfast	**1**
51	Citybeverly	**1**
52	Citybrooklyn	**1**
53	Citybrownfield	**1**
54	Citycarlsbad	**1**
55	Citychandler	**1**
56	Citychisago city	**1**
57	Citycleveland	**1**
58	Citycortland	**1**
59	Citycreighton	**1**
60	Citydel city	**1**
61	Cityegg harbor township	**1**
62	Cityfestus	**1**
63	Citygalesburg	**1**
64	Cityglendale	**1**
65	Citygrand rapids	**1**
66	Cityhattiesburg	**1**

67	Cityhayesville	**1**
68	Cityhialeah	**1**
69	Cityhudson	**1**
70	Cityjefferson city	**1**
71	Citykenosha	**1**
72	Citylaconia	**1**
73	Citylynbrook	**1**
74	Citymanitowoc	**1**
75	Citymarshfield	**1**
76	Citymiami lakes	**1**
77	Citynew boston	**1**
78	Citynorristown	**1**
79	Citynorth kansas city	**1**
80	Cityosceola	**1**
81	Citypanorama city	**1**
82	Citypierre	**1**
83	Citypompton plains	**1**
84	Cityrhinelander	**1**
85	Cityrockingham	**1**
86	Citysan antonio	**1**
87	Citysanta cruz	**1**

88	Citysherman	**1**
89	Citysikeston	**1**
90	Citysioux city	**1**
91	Cityspring hill	**1**
92	Citysturgeon bay	**1**
93	Citysweetwater	**1**
94	Citytuscaloosa	**1**
95	Citytyler	**1**
96	Cityvancouver	**1**
97	Citywausau	**1**
98	Citywaverly	**1**
99	Provider.Nameangel care home health services inc	**1**
100	Provider.Nameangels care home health of the emerald coast	**1**
101	Provider.Nameann's choice visiting nurse services	**1**
102	Provider.Nameaspirus vna home health inc	**1**
103	Provider.Nameasset health services inc	**1**
104	Provider.Namebaptist health home health dash heber springs	**1**
105	Provider.Namebaycare home care inc	**1**
106	Provider.Namebeatitudes home health	**1**
107	Provider.Namebest home healthcare network inc	**1**
108	Provider.Namebethany home health services	**1**

109	Provider.Namebig bend home health	**1**
110	Provider.Namebreeze health care inc.	**1**
111	Provider.Namebrmc home care	**1**
112	Provider.Namebrookdale home health holland	**1**
113	Provider.Namecareall	**1**
114	Provider.Namecari home care inc	**1**
115	Provider.Namecaring like family inc	**1**
116	Provider.Namecarter healthcare inc	**1**
117	Provider.Namecedar crest village inc home health department	**1**
118	Provider.Namecentral florida quality care services inc	**1**
119	Provider.Namechristus homecare	**1**
120	Provider.Namecortland regional medical center inc lthhcp	**1**
121	Provider.Namedale medical ctr home health	**1**
122	Provider.Namedeaconess homecare	**1**
123	Provider.Namedoctors home care	**1**
124	Provider.Namedoniphan co health dept hha	**1**
125	Provider.Namedpn homecare	**1**
126	Provider.Namedynamic home health care	**1**
127	Provider.Nameel passion home health agency inc	**1**
128	Provider.Nameeverwell health agency llc	**1**
129	Provider.Nameexceed home health inc	**1**

130	Provider.Namefairview lakes homecaring hosp	**1**
131	Provider.Namefamily care of texas	**1**
132	Provider.Namefamily hospice and pallative care	**1**
133	Provider.Namefidelity health care inc	**1**
134	Provider.Namefirst health home care dash richm	**1**
135	Provider.Namefloyd memorial home health care	**1**
136	Provider.Nameforrest general home care hha	**1**
137	Provider.Namefranklin hospital medical center chha	**1**
138	Provider.Namegentiva health services	**1**
139	Provider.Namegirling home health texas by harden healthcare	**1**
140	Provider.Nameglobal home health i llc	**1**
141	Provider.Namegolden years home care inc.	**1**
142	Provider.Namegood samaritan home health care inc	**1**
143	Provider.Namegreenwood county hospital hha	**1**
144	Provider.Namehazard arh hha	**1**
145	Provider.Namehealing to heal another home healthcare inc	**1**
146	Provider.Namehealth first home care	**1**
147	Provider.Namehealthwatch home health of weatherford llc	**1**
148	Provider.Namehome care of the grand valley	**1**
149	Provider.Namehumanity home health inc	**1**
150	Provider.Nameinnovative senior care home health	**1**

151	Provider.Nameintegris home care enid	**1**
152	Provider.Nameintrepid usa healthcare services	**1**
153	Provider.Namejm homecare solutions inc	**1**
154	Provider.Namejordan health services	**1**
155	Provider.Namekaiser foundation tri central hha	**1**
156	Provider.Namekenosha vna	**1**
157	Provider.Namekind hands inc	**1**
158	Provider.Namekings home healthcare inc	**1**
159	Provider.Namelakeview home health care	**1**
160	Provider.Namelong life home care inc	**1**
161	Provider.Namemacy's health services inc	**1**
162	Provider.Namemount auburn home health	**1**
163	Provider.Namenational church residences home and community serv	**1**
164	Provider.Namenemaha county home care	**1**
165	Provider.Namenorth kansas city hospital home health services	**1**
166	Provider.Namenu dash era home health agency inc	**1**
167	Provider.Namenurses choice home care	**1**
168	Provider.Nameomhs home care services	**1**
169	Provider.Nameomni home care agency inc	**1**
170	Provider.Nameoneida county hospital home care	**1**
171	Provider.Nameottawa co health center hha	**1**

172	Provider.Nameperry county health department	**1**
173	Provider.Namepinnacle senior care	**1**
174	Provider.Nameprudent healthcare agency	**1**
175	Provider.Namepulse homecare	**1**
176	Provider.Namequality care home health services inc	**1**
177	Provider.Namequality life home health agency corp	**1**
178	Provider.Nameregional home care helena	**1**
179	Provider.Nameregistered nurses care ltd	**1**
180	Provider.Nameriverside shore home health	**1**
181	Provider.Namesampson home health	**1**
182	Provider.Nameselect home care llc	**1**
183	Provider.Namesierra home health care	**1**
184	Provider.Namesouth florida home health care inc	**1**
185	Provider.Namessm home care at st francis hospital	**1**
186	Provider.Namessm home care at st mary's health center	**1**
187	Provider.Namest francis home health care	**1**
188	Provider.Namesupreme patient care inc	**1**
189	Provider.Nameteam select home care	**1**
190	Provider.Nametehc llc	**1**
191	Provider.Nametrinity home care	**1**
192	Provider.Nameunion regional home care	**1**

193	Provider.Nameunitypoint at home	**1**
194	Provider.Namevalley of the sun home health care llc	**1**
195	Provider.Nameveritas home care inc	**1**
196	Provider.Nameviva home health care inc	**1**
197	Provider.Namevna nazareth home care	**1**
198	Provider.Namevolunteers of america home health at westchester	**1**
199	Provider.Namewaldo county home healthcare services	**1**
200	Provider.Namewisdom health care services inc	**1**
201	Provider.Nameyuma district hospital home health care	**1**
202	Stateak	**1**
203	Statein	**1**
204	Statemd	**1**
205	Statemn	**1**
206	Statenc	**1**
207	Statenj	**1**
208	Stateok	**1**
209	Stateor	**1**
210	Statepa	**1**
211	Statepr	**1**
212	Statesd	**1**
213	Stateut	**1**

214	Statewa	**1**
215	Type.of.Ownershipstate or county	**1**
216	Type.of.Ownershipvoluntary non-profit church	**1**
217	Type.of.Ownershipvoluntary non-profit other	**1**
218	Type.of.Ownershipvoluntary non-profit private	**1**

5. Instructional Approach to Deploy a Predictive Model to Production

In this section we suggest a high level approach to replicating or moving to production the findings from this project.

5.1 Obtain new data set and data preparation

Ensure new data set has the same variables as the one used for this project.

Follow each of the steps in the Data Preparation section sequentially to ensure that the new data is prepared according to the findings from this work. Reuse the R code for the data preparation.

5.2 Model fitting and Evaluation with New Data

Use the predictors found to have explanatory impact. Refer to the *4.8 Ranking of Explanatory Predictors.* All 217 predictors can be used or a subset.

The Lasso technique is the only technique used in this project that resulted in estimates as well as variable reduction. All the other predictive modeling techniques were used only for variable reduction or Exploratory Data Analysis (EDA).

Create a Train and Test data set to fit and evaluate the model. Reuse the R Code provided for the Lasso model.

Evaluate the fitted model with new data and ensure the predictors provided are still providing a USEFUL model.

Document the predictors and their resulting estimates.

5.3 Update Production Star Ratings and Estimates

The results from the steps above provides new Star Scoring for each Provider and new predictive estimates that can be used to replace existing ones in Production.

6. Final Conclusion

The two-fold goals of this project have been successfully accomplished and are expected to provide great value for the organization.

The development of the recommendation engine helps consumers select the best performing and most appropriate Home Health Care Agency (CMS certified) for their needs as well as successfully accomplishing the internal capabilities for the development of a low-cost technology platform enabling the organization to apply data mining and predictive analytics techniques that enables the creation of this new data product.

The accomplishment of the **proof of value** goal, which is to develop a **recommendation engine** to help consumers select the best and most appropriate Home Health Care Agency, was completed by the application of the Lasso algorithm on the prepared data set, including the Provider Name. The resulting estimate of performance for each Provider Name as well as other categories such as State, City, and Type of Ownership demonstrate that they reflect reality to a great extent. For example, the Lasso estimates on the Type of Ownership shows that the HHCAs with positive estimates are those that are 'Voluntary' in their legal setting, this correlates to the EDA results of the stack bar in Figure 30.7 showing that the proportion of 'Voluntary non-profit' HHCAs increases as the Star Rating increases, confirming the positive coefficients from the Lasso estimates.

Another example that the Lasso estimates seem to reflect reality is with respect to the prevalence of the State of Texas in the lower Star Ratings of 2.0, 2.5, and 3.5, as shown in Figure 30.6, and the prevalence of the State of Florida in the higher Star Ratings of 4.0, 4.5, and 5.0 and also California as second to Florida in the Star Ratings 4.5 and 5.0, refer to Figure 30.6. This findings coincide with the Lasso estimates for States showing Texas having a negative estimate of -0.043, Florida having a positive estimate of 0.145 and California having a positive estimate of 0.117.

Based on this analysis, the Lasso estimates seem extremely useful in providing predictions for the future performance of HHCAs and can be the basis for the predictive aspect of the recommendation engine.

Also, the identification of 217 explanatory variables is based on the robust application of different machine learning algorithms using best practices, based on 11 quarter of historical data and considers the many services provided as well as geographical attributes of the Providers or HHCAs. This approach concludes that these 217 explanatory variables can be used in the model fitting of new data and predictions for the recommendation engine.

The accomplishment of the **proof of concept** goal, which is to develop internal capabilities in terms of technology platform and in terms of applying data mining and predictive analytics techniques, was completed by the identification of a machine learning algorithm available in R, the Lasso, which can compute performance estimates at the Provider Name level, thus allowing the provisioning of predictions for consumers at this level of detail, which is what they would require in order to easily select a CMS certified Home Health Care Agency or Provider.

Also, the completion of all aspects for the Data Preparation in R demonstrated that data manipulation can be done on the same platform as the one used for the fitting of predictive models, for the performance of statistical analysis, and the visualizing of model performance.

For the **proof of value** goal, which is to develop a recommendation engine to help consumers select the best and most appropriate Home Health Care Agency was completed by the identification of a set of important variables that help us distinguish between bad, good and excellent providers. We were able to adopt the CMS methodology for assigning Quality Star Rating for each Provider from 2015-Q3 and extend it to historical quarterly data sets from 2014 and 2013. The recommendation engine is able include the Star Rating of how the Providers have been performing historically. The estimates or predictive coefficient for Provider performance based on Process and on Outcome measures help consumers distinguish between bad, good and excellent Providers. This can also useful for future research if there is an interest in looking at why a provider's performance improved or declined over time.

The utilization of Tableau expedited the practical insights on the data enabling the team to complete commentary and analysis much quicker than having to have done this in R.

In conclusion, this project demonstrates that using the low-cost open-source statistical modeling tool R, using a low-cost visualization tool Tableau, and using public data from the Centers of Medicare and Medicaid Services (CMS) enable the successful creation of a valuable data product or recommendation engine that has great potential in helping tens of thousands of home-bound patients to easily select a CMS certified Home Health Care Agency that is high performing and meets their personal home health care needs.

7. Project Status:

The Project has been successfully completed with all the deliverables associated with the Goals being met at the end of the 8 weeks since the beginning of the project, refer to Table 21 below which shows the detailed list of the Project Goals, Deliverables and their Status.

The project has successfully developed a recommendation engine that can help home-bound patients easily select CMS certified Home Health Care Agencies by comparing them based on performance and outcomes.

The project successfully utilized the open-source statistical modeling software R to perform all data preparation, model fitting, and predictions.

The project team encountered a few instances of memory limits while running the programs in R but they were overcome without negative impact to the project goals or timeline.

Table 21 – HomeCare4Me Project Goals, Deliverables and ***Final Report Status***

Goal Number	High Level Goal	Goal	Deliverables	Final Report Status
1	Proof of Value	***Identify attributes*** that distinguish high performing Home Health Care Agencies from poor performing agencies by leverage public CMS home health care data to . These attributes will be used by the recommendation engine to score the Agencies.	***Set of variables*** that are important predictors to identify high performing Agencies from low performing Agencies.	This goal has been accomplished by fitting several predictive models, such as Random Forest Regression, Random Forest Classification, Support Vector Machine (Linear method), and Support Vector Machine (Radial Method), and Lasso. There were 217 distinct predictors identified out of 13,856 that have explanatory value on the Star Rating of Home Health Care Agencies. 19 of the 217 explanatory predictors resulted from 3 different models, RF Regression, RF Classification, and Lasso. 28 of the 217 explanatory predictors resulted from 2 different models. And the rest, 171 predictors, resulting from only one model. These findings provides the Set of Variables for this deliverable.
2	Proof of Value	***Identify natural groupings or segments*** of Home Health Care Agencies which can be used by the recommendation engine to guide the consumer in selecting the appropriate Home Health Care	***List of segments or groupings*** for high performing Agencies versus low performing Agencies.	The Lasso findings show the list of Providers or Agencies with their historical mean Star Rating that provides the consumer the performance information for selecting an Angency. It also provides the estimate or predictive value of the Agency for an additional data point to making a decision on an Agency.
3	Proof of Value	***Provide a list of Home Health Care agencies*** with their Quality of Care Score based on the geographical location of a patient and the type of service desired.	***List of Home Health Care agencies*** with their Quality of Care Score based on the geographical location of a patient and the type of service desired.	The final preparation of the data in section 3.9 creates a list of the Providers with their historical mean Star Rating, State, City, and Type of Ownwership.
4	Proof of Value	***Provide a Quality Performance Measure*** **Score** for the Home Health Care Agency selected by the patient.	***List of providers*** and their respective Quality Performance Score	Same as goal #3 but without the geographical variables.
5	Proof of Concept	***Evaluate R as a tool to the prepare data*** , such as combining different data sets, imputing missing values, standardizing metadata data across historical files, standardizing categorical values across historical data, imputing numeric and character data, filtering records, and dichotomizing categorical variables.	***Clean and conformed data set of historical data*** and is ready for modeling by imputing missing values, removing variables or observations not needed, dichotomizing categorical variables.	About 95% of the data preparation was performed successfully using R. There were some aspects of the data preparation that were performed faster using Excel and MS Access, but they could have been completed in R.
6	Proof of Concept	***Evaluate R as a tool to apply initial EDA for descriptive statistics*** , correlation, reducing the number of variables, data mining techniques to find insights in the data.	***Insights*** obtained from the EDA using R	All of the EDA for descriptive statistitics and variable reduction was performed in R. The summary function was used several time to provide the quartiles. Scatter plots, and histograms were used to assess statistical characteristics of the data.
7	Proof of Concept	***Evaluate R as a tool to fit predictive models*** and evaluate their accuracy.	***Predictive models*** and their accuracy based on training and test data sets in order to select the best model.	All of the predictive modeling fitting, prediction, and evaluations were performed in R. Train and Test data sets were created and used in the fitting of the models and evaluation.
8	Proof of Concept	***Evaluate R as a tool to deploy the predictive model*** to production to score new observations.	***Instructional steps to deploy a predictive*** or scoring model in production using R.	The deployment instructions based on the findings from this project are delineated in the section: Instructional Approach to Deploy a Predictive Model in Production.
9	Proof of Concept	***Evaluate Tableau as a visualization*** tool to explain insights about the data.	***Visualizations the generate insights into the historical data*** being used, descriptive statistics of the data, relationship among variables, dependencies among predictors and variable to be predicted.	Tableau was used successfully to provide insights using Bubble charts as to the prevalence of observations by State. Tableau's Hierarchical table view of the data also provided a fast insight as to how the same Provider Name has data across several States, or several Cities, or even within the same Zip Code. Also, the same Provider Name can have different CMS Certification Numbers. Tableau's Stack Bar chart also provided quick insights as to the most prevalent Type of Ownership, 'Proprietary", is embedded in all of the types of Services Offered.

Table 22 below shows the Project Status Dashboard and Timeline for the Final Report submission. The focus for the project team after the submission of the Final Report is on the preparation of the Oral presentation and also to incorporate the feedback from the CEO about the Final Report into the Oral Presentation.

Table 22 – HomeCare4Me Project Status Dashboard

(Credit for the template of the Project Status Dashboard is acknowledged to Scott Lancasters who provided this template in the DB)

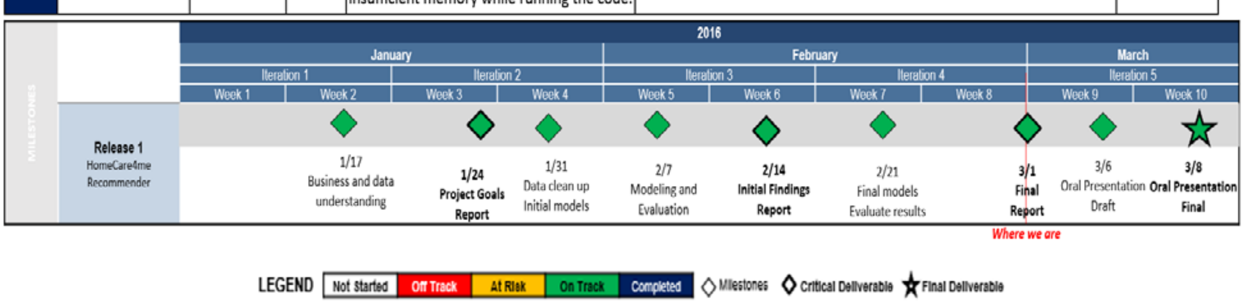
HomeCare4me Team	REPORT DATE	CURRENT STATUS	PRIOR STATUS	TREND
	Period Ending 2016-02-28	On Track	On Track	↔

PROJECT OVERVIEW		
Project Sponsor	Don Wedding, CEO	Project Team: Sandra Duenas and Siddartha Sathyanarayan
Project Objective	• To develop a recommendation engine to help consumers select the best and most appropriate Home Health Care Agency for their individual needs. • To develop internal capabilities in terms of technology platform and in terms of applying data mining and predictive analytics techniques in order to build the recommendation engine.	

STATUS	Schedule			Resourcing			Business Alignment		
Dimension	Schedule			Resourcing			Business Alignment		
Status	Current: G	Prior: G	↔	Current: G	Prior: G	↔	Current: G	Prior: G	↔
Status Description	Schedule set and on target			Team formed and roles assigned			Goals set but need to be agreed to by sponsor		

Key Milestones	Progress Since Last Period	Target Accomplishments Next Period
Completed: 01/24 – Project Goals report 02/14 – Initial Findings report 03/01 – Final report **Upcoming:** 3/8 –Ora Presentation	• Expanded Executive Summary as requested • Re-worked Clustering with PCA reduction as requested • Built predictive models and evaluated them • The Lasso model provided the performance estimates at the Provider Name level, which is required for the recommendation engine. • Completed Final Report	• Prepare for Oral Presentation • Incorporate feedback from CEO about the Final Report into Oral presentation.

Name	Owner	Type	Description & Impact Detail	Mitigation/Contingency Plan	Res. Target
Size of variables	Sandra & Siddartha	Risk	After creating dummy variables on a set of 22 variables, the number of variables in the data set has grown exponentially to around 22,000. Team has run into errors due to insufficient memory while running the code.	Apply dimensionality reduction techniques or filter data down the number of provider names. Status: Resolved – The issue has been addressed by eliminating the Provider names variable from the data set.	1/24/2016



References

1. What is Home Health Compare?

https://www.medicare.gov/homehealthcompare/About/What-Is-HHC.html

2. Data Sources:

Archived Data: from 2004 to 2014

https://data.medicare.gov/data/archives/home-health-compare

Current Data: 2015

https://data.medicare.gov/data/home-health-compare

3. Using Home Health Compare data:

https://www.medicare.gov/HomeHealthCompare/About/Using-The-Data.html

4. Measuring agency performance:

https://www.medicare.gov/HomeHealthCompare/Data/Measuring-Agency-Performance.html

5. Star Rating and Agency Scoring Methodology:

https://www.medicare.gov/HomeHealthCompare/Data/Patient-Care-Star-Ratings.html

6. R code provided by Peng in the Sharing DB that dichotomizes categorical variables. The specific R code is **dummy.df <- data.frame(model.matrix(~.-1, data = my.df))** This code saved my many hours.

7. James, G. et al, 2013. An Introduction to Statistical Learning with Applications in R. Springer Science+Business Media. New York.

8. CRISP-DM Methodology

https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Gossary

CMS - Center for Medicaid and Medicare Services

HHCA or HHA - Home Health Care Agencies

PCQM - Patient Care Quality Measures

EDA - Exploratory Data Analysis

CRISP-DM – Cross Industry Standard Process for Data Mining

Addendum 1 - Lasso Estimates for all HHCAs or Providers

The list shown in this Addendum comprise of unique rows for the combination of the values for Provider Name, CMS Cert Number, City, State, and Type of Ownership. However, the Lasso estimate is assigned at the CMS Cert Number. For example, the Provider Name of 'SIERRA HOME HEALTH CARE' is shown in rows 4, 5, and 6; however, in rows 4 and 5, it has the same CMS Cert Number but different cities in the same state, and in row 6, this Provider Name has a different CMS Cert Number and the location is in another state and city.

The estimates with positive or increasing explanatory impact on the Star Rating are highlighted in light green. The estimates with negative or decreasing explanatory impact on the Star Rating are highlighted in light orange.

	Provider Name	CMS Cert Number	Historical Mean of Star Rating	Lasso Estimate	City	State	Type of Ownership
1	CARI HOME CARE INC	108034	4.75	**0.7732**	homestead	FL	proprietary
2	VOLUNTEERS OF AMERICA HOME HEALTH AT WESTCHESTER	37265	4.50	**0.5804**	tempe	AZ	proprietary
3	DYNAMIC HOME HEALTH CARE	497655	4.25	**0.5114**	chantilly	VA	proprietary
4	SIERRA HOME HEALTH CARE	239066	4.25	**0.4920**	dearborn	MI	proprietary
5	SIERRA HOME HEALTH CARE	239066	4.25	**0.4920**	westland	MI	proprietary

6	SIERRA HOME HEALTH CARE	58022	4.17	**0.4920**	fontana	CA	proprietary
7	HEALING TO HEAL ANOTHER HOME HEALTHCARE INC	148115	2.75	**0.3781**	matteson	IL	proprietary
8	HEALING TO HEAL ANOTHER HOME HEALTHCARE INC	148115	2.75	**0.3781**	olympia fields	IL	proprietary
9	INNOVATIVE SENIOR CARE HOME HEALTH	107656	3.50	**0.3332**	sun city center	FL	proprietary
10	INNOVATIVE SENIOR CARE HOME HEALTH	107766	4.00	**0.3332**	fort myers	FL	proprietary
11	INNOVATIVE SENIOR CARE HOME HEALTH	108066	4.75	**0.3332**	pompano beach	FL	proprietary

12	INNOVATIVE SENIOR CARE HOME HEALTH	108119	3.75	**0.3332**	west palm beach	FL	proprietary
13	INNOVATIVE SENIOR CARE HOME HEALTH	108133	4.75	**0.3332**	west melbourne	FL	proprietary
14	INNOVATIVE SENIOR CARE HOME HEALTH	108467	5.00	**0.3332**	seminole	FL	proprietary
15	INNOVATIVE SENIOR CARE HOME HEALTH	109317	4.25	**0.3332**	jacksonville	FL	proprietary
16	INNOVATIVE SENIOR CARE HOME HEALTH	109581	3.75	**0.3332**	the villages	FL	proprietary
17	INNOVATIVE SENIOR CARE HOME HEALTH	109595	4.17	**0.3332**	fort walton beach	FL	proprietary
18	INNOVATIVE SENIOR CARE HOME HEALTH	147878	4.25	**0.3332**	naperville	IL	proprietary

19	INNOVATIVE SENIOR CARE HOME HEALTH	157582	4.00	**0.3332**	indianapolis	IN	proprietary
20	INNOVATIVE SENIOR CARE HOME HEALTH	157642	3.50	**0.3332**	portage	IN	proprietary
21	INNOVATIVE SENIOR CARE HOME HEALTH	177187	4.25	**0.3332**	salina	KS	proprietary
22	INNOVATIVE SENIOR CARE HOME HEALTH	178085	4.50	**0.3332**	overland park	KS	proprietary
23	INNOVATIVE SENIOR CARE HOME HEALTH	227302	4.25	**0.3332**	quincy	MA	proprietary
24	INNOVATIVE SENIOR CARE HOME HEALTH	237530	5.00	**0.3332**	coloma	MI	proprietary
25	INNOVATIVE SENIOR CARE HOME HEALTH	237540	4.00	**0.3332**	farmington	MI	proprietary

26	INNOVATIVE SENIOR CARE HOME HEALTH	267603	4.25	**0.3332**	creve coeur	MO	proprietary
27	INNOVATIVE SENIOR CARE HOME HEALTH	267609	3.75	**0.3332**	raymore	MO	proprietary
28	INNOVATIVE SENIOR CARE HOME HEALTH	327201	4.33	**0.3332**	albuquerque	NM	proprietary
29	INNOVATIVE SENIOR CARE HOME HEALTH	347011	4.50	**0.3332**	durham	NC	proprietary
30	INNOVATIVE SENIOR CARE HOME HEALTH	347098	4.50	**0.3332**	charlotte	NC	proprietary
31	INNOVATIVE SENIOR CARE HOME HEALTH	347098	4.50	**0.3332**	charlotte	NC	voluntary non-profit private
32	INNOVATIVE SENIOR CARE HOME HEALTH	347120	4.25	**0.3332**	greensboro	NC	private

33	INNOVATIVE SENIOR CARE HOME HEALTH	347120	4.25	**0.3332**	greensboro	NC	proprietary
34	INNOVATIVE SENIOR CARE HOME HEALTH	347120	4.25	**0.3332**	greensboro	NC	voluntary non-profit private
35	INNOVATIVE SENIOR CARE HOME HEALTH	367711	4.75	**0.3332**	dayton	OH	proprietary
36	INNOVATIVE SENIOR CARE HOME HEALTH	367758	4.50	**0.3332**	westlake	OH	proprietary
37	INNOVATIVE SENIOR CARE HOME HEALTH	367777	4.00	**0.3332**	columbus	OH	proprietary
38	INNOVATIVE SENIOR CARE HOME HEALTH	37255	2.75	**0.3332**	chandler	AZ	proprietary
39	INNOVATIVE SENIOR CARE HOME HEALTH	37256	3.75	**0.3332**	tucson	AZ	proprietary

40	INNOVATIVE SENIOR CARE HOME HEALTH	377126	4.25	**0.3332**	jenks	OK	proprietary
41	INNOVATIVE SENIOR CARE HOME HEALTH	377666	3.60	**0.3332**	edmond	OK	proprietary
42	INNOVATIVE SENIOR CARE HOME HEALTH	387127	3.75	**0.3332**	wilsonville	OR	proprietary
43	INNOVATIVE SENIOR CARE HOME HEALTH	398036	4.75	**0.3332**	elkins park	PA	proprietary
44	INNOVATIVE SENIOR CARE HOME HEALTH	398036	4.75	**0.3332**	king of prussia	PA	proprietary
45	INNOVATIVE SENIOR CARE HOME HEALTH	417059	4.50	**0.3332**	lincoln	RI	proprietary
46	INNOVATIVE SENIOR CARE HOME HEALTH	497625	3.75	**0.3332**	richmond	VA	proprietary

47	INNOVATIVE SENIOR CARE HOME HEALTH	557696	3.50	**0.3332**	covina	CA	proprietary
48	INNOVATIVE SENIOR CARE HOME HEALTH	59028	4.00	**0.3332**	dublin	CA	proprietary
49	INNOVATIVE SENIOR CARE HOME HEALTH	67469	4.25	**0.3332**	greenwood village	CO	proprietary
50	INNOVATIVE SENIOR CARE HOME HEALTH	679313	4.00	**0.3332**	houston	TX	proprietary
51	INNOVATIVE SENIOR CARE HOME HEALTH	679424	4.25	**0.3332**	san antonio	TX	proprietary
52	INNOVATIVE SENIOR CARE HOME HEALTH	679606	4.50	**0.3332**	fort worth	TX	proprietary
53	INNOVATIVE SENIOR CARE HOME HEALTH	679637	4.00	**0.3332**	corpus christi	TX	proprietary

54	INNOVATIVE SENIOR CARE HOME HEALTH	679682	3.50	**0.3332**	austin	TX	proprietary
55	INNOVATIVE SENIOR CARE HOME HEALTH	77243	4.50	**0.3332**	farmington	CT	proprietary
56	FRANKLIN HOSPITAL MEDICAL CENTER CHHA	337189	4.50	**0.3162**	lynbrook	NY	other
57	FRANKLIN HOSPITAL MEDICAL CENTER CHHA	337189	4.50	**0.3162**	lynbrook	NY	private
58	BEST HOME HEALTHCARE NETWORK INC	148081	4.75	**0.3145**	chicago	IL	proprietary
59	BEST HOME HEALTHCARE NETWORK INC	148081	4.75	**0.3145**	lombard	IL	proprietary

60	BEATITUDES HOME HEALTH	37265	4.50	**0.3105**	phoenix	AZ	proprietary
61	BEATITUDES HOME HEALTH	37265	4.50	**0.3105**	phoenix	AZ	religious affiliations
62	HUMANITY HOME HEALTH INC	109676	4.75	**0.2630**	miami	FL	proprietary
63	BAYCARE HOME CARE INC	107212	4.50	**0.2155**	largo	FL	other
64	BAYCARE HOME CARE INC	107212	4.50	**0.2155**	largo	FL	proprietary
65	BAYCARE HOME CARE INC	107260	4.75	**0.2155**	new port richey	FL	private
66	BAYCARE HOME CARE INC	107260	4.75	**0.2155**	new port richey	FL	voluntary non-profit private

67	BAYCARE HOME CARE INC	107282	3.75	**0.2155**	lakeland	FL	other
68	BAYCARE HOME CARE INC	107285	4.75	**0.2155**	tampa	FL	other
69	BAYCARE HOME CARE INC	107285	4.75	**0.2155**	tampa	FL	voluntary non-profit other
70	BAYCARE HOME CARE INC	107398	4.75	**0.2155**	fruitland park	FL	private
71	BAYCARE HOME CARE INC	107398	4.75	**0.2155**	fruitland park	FL	proprietar y
72	BAYCARE HOME CARE INC	107398	4.75	**0.2155**	fruitland park	FL	voluntary non-profit private
73	BAYCARE HOME CARE INC	107415	4.25	**0.2155**	dunedin	FL	other

74	BAYCARE HOME CARE INC	107415	4.25	**0.2155**	dunedin	FL	voluntary non-profit other
75	BAYCARE HOME CARE INC	107417	4.50	**0.2155**	spring hill	FL	private
76	LONG LIFE HOME CARE INC	109303	4.25	**0.2094**	coral gables	FL	proprietar y
77	LONG LIFE HOME CARE INC	109303	4.25	**0.2094**	miami	FL	proprietar y
78	CEDAR CREST VILLAGE INC HOME HEALTH DEPARTMENT	317092	5.00	**0.2061**	pompton plains	NJ	private
79	CEDAR CREST VILLAGE INC HOME HEALTH DEPARTMENT	317092	5.00	**0.2061**	pompton plains	NJ	voluntary non-profit private

80	QUALITY LIFE HOME HEALTH AGENCY CORP	109603	4.50	**0.2031**	cape coral	FL	proprietary
81	BREEZE HEALTH CARE INC.	109405	3.17	**0.2023**	dania	FL	proprietary
82	BREEZE HEALTH CARE INC.	109405	3.17	**0.2023**	dania	FL	voluntary non-profit private
83	BAPTIST HEALTH HOME HEALTH DASH HEBER SPRINGS	47097	4.75	**0.1680**	heber springs	AR	private
84	BAPTIST HEALTH HOME HEALTH DASH HEBER SPRINGS	47097	4.75	**0.1680**	heber springs	AR	voluntary non-profit church
85	BAPTIST HEALTH HOME HEALTH DASH HEBER SPRINGS	47097	4.75	**0.1680**	heber springs	AR	voluntary non-profit private

86	BRMC HOME CARE	677415	4.75	**0.1583**	brownfiel d	TX	govt local
87	BRMC HOME CARE	677415	4.75	**0.1583**	brownfiel d	TX	local
88	BRMC HOME CARE	677415	4.75	**0.1583**	brownfiel d	TX	other
89	VNA NAZARETH HOME CARE	157584	5.00	**0.1559**	clarksville	IN	religious affiliations
90	VNA NAZARETH HOME CARE	157584	5.00	**0.1559**	clarksville	IN	voluntary non-profit church
91	VNA NAZARETH HOME CARE	187000	5.00	**0.1559**	louisville	KY	proprietar y
92	VNA NAZARETH HOME CARE	187000	5.00	**0.1559**	louisville	KY	voluntary non-profit church
93	DOCTORS HOME CARE	47025	5.00	**0.1460**	camden	AR	private
94	DOCTORS HOME CARE	47025	5.00	**0.1460**	camden	AR	voluntary non-profit other

95	DOCTORS HOME CARE	47025	5.00	**0.1460**	camden	AR	voluntary non-profit private
96	ST FRANCIS HOME HEALTH CARE	237239	4.75	**0.1426**	escanaba	MI	proprietary
97	ST FRANCIS HOME HEALTH CARE	237239	4.75	**0.1426**	escanaba	MI	religious affiliations
98	ST FRANCIS HOME HEALTH CARE	247213	4.50	**0.1426**	breckenridge	MN	religious affiliations
99	ST FRANCIS HOME HEALTH CARE	247213	4.50	**0.1426**	breckenridge	MN	voluntary non-profit church
100	UNION REGIONAL HOME CARE	347210	4.75	**0.1354**	monroe	NC	other
101	UNION REGIONAL HOME CARE	347210	4.75	**0.1354**	monroe	NC	voluntary non-profit other

102	HEALTHWATCH HOME HEALTH OF WEATHERFORD LLC	377748	5.00	**0.1337**	mcalester	OK	proprietary
103	HEALTHWATCH HOME HEALTH OF WEATHERFORD LLC	377748	5.00	**0.1337**	weatherford	OK	proprietary
104	OMHS HOME CARE SERVICES	187112	5.00	**0.1308**	owensboro	KY	private
105	OMHS HOME CARE SERVICES	187112	5.00	**0.1308**	owensboro	KY	voluntary non-profit private
106	MACY'S HEALTH SERVICES INC	747440	4.17	**0.1289**	houston	TX	proprietary
107	MACY'S HEALTH SERVICES INC	747440	4.17	**0.1289**	sugar land	TX	proprietary
108	SSM HOME CARE AT ST	267065	5.00	**0.1224**	maryville	MO	religious affiliations

	FRANCIS HOSPITAL						
109	SSM HOME CARE AT ST FRANCIS HOSPITAL	267065	5.00	**0.1224**	maryville	MO	voluntary non-profit church
110	BIG BEND HOME HEALTH	457686	4.75	**0.1192**	alpine	TX	proprietar y
111	WALDO COUNTY HOME HEALTHCARE SERVICES	207028	4.75	**0.1147**	belfast	ME	other
112	WALDO COUNTY HOME HEALTHCARE SERVICES	207028	4.75	**0.1147**	belfast	ME	private
113	WALDO COUNTY HOME HEALTHCARE SERVICES	207028	4.75	**0.1147**	belfast	ME	voluntary non-profit private

114	HOME CARE OF THE GRAND VALLEY	67133	4.75	**0.1119**	grand junction	CO	private
115	HOME CARE OF THE GRAND VALLEY	67133	4.75	**0.1119**	grand junction	CO	proprietary
116	HOME CARE OF THE GRAND VALLEY	67133	4.75	**0.1119**	grand junction	CO	voluntary non-profit private
117	FAIRVIEW LAKES HOMECARING HOSP	247211	4.75	**0.1100**	chisago city	MN	private
118	FAIRVIEW LAKES HOMECARING HOSP	247211	4.75	**0.1100**	chisago city	MN	voluntary non-profit private
119	ASSET HEALTH SERVICES INC	457935	3.25	**0.1078**	houston	TX	proprietary

120	TRINITY HOME CARE	167149	5.00	**0.1068**	fort dodge	IA	voluntary non-profit other
121	TRINITY HOME CARE	167149	5.00	**0.1068**	fort dodge	IA	voluntary non-profit private
122	GOLDEN YEARS HOME CARE INC.	108349	4.25	**0.1059**	miami lakes	FL	proprietar y
123	KINGS HOME HEALTHCARE INC	398134	5.00	**0.1011**	ashland	PA	proprietar y
124	KINGS HOME HEALTHCARE INC	398134	5.00	**0.1011**	pottsville	PA	proprietar y
125	NORTH KANSAS CITY HOSPITAL HOME HEALTH SERVICES	267462	4.75	**0.1006**	north kansas city	MO	local

126	NORTH KANSAS CITY HOSPITAL HOME HEALTH SERVICES	267462	4.75	**0.1006**	north kansas city	MO	other
127	NORTH KANSAS CITY HOSPITAL HOME HEALTH SERVICES	267462	4.75	**0.1006**	north kansas city	MO	voluntary non-profit other
128	CARING LIKE FAMILY INC	58285	4.75	**0.0983**	marina del rey	CA	other
129	CARING LIKE FAMILY INC	58285	4.75	**0.0983**	marina del rey	CA	proprietary
130	CARING LIKE FAMILY INC	58285	4.75	**0.0983**	santa monica	CA	proprietary
131	CARING LIKE FAMILY INC	58285	4.75	**0.0983**	santa monica	CA	voluntary non-profit other
132	REGIONAL HOME CARE HELENA	47112	4.50	**0.0939**	helena	AR	proprietary

133	TEAM SELECT HOME CARE	37234	4.25	**0.0923**	phoenix	AZ	proprietary
134	TEAM SELECT HOME CARE	377273	4.75	**0.0923**	tulsa	OK	proprietary
135	TEAM SELECT HOME CARE	67490	4.75	**0.0923**	colorado springs	CO	proprietary
136	KIND HANDS INC	59396	5.00	**0.0900**	los angeles	CA	proprietary
137	MOUNT AUBURN HOME HEALTH	107542	4.75	**0.0862**	doral	FL	proprietary
138	CORTLAND REGIONAL MEDICAL CENTER INC LTHHCP	337198	3.00	**0.0813**	cortland	NY	private
139	CORTLAND REGIONAL MEDICAL CENTER INC LTHHCP	337198	3.00	**0.0813**	cortland	NY	voluntary non-profit private

140	OMNI HOME CARE AGENCY INC	109497	4.75	**0.0807**	miami	FL	proprietary
141	SUPREME PATIENT CARE INC	109780	4.25	**0.0744**	dania beach	FL	proprietary
142	SUPREME PATIENT CARE INC	109780	4.25	**0.0744**	oakland park	FL	proprietary
143	BROOKDALE HOME HEALTH HOLLAND	237530	5.00	**0.0713**	coloma	MI	proprietary
144	BROOKDALE HOME HEALTH HOLLAND	237530	5.00	**0.0713**	saint joe	MI	proprietary
145	UNITYPOINT AT HOME	167002	4.25	**0.0674**	dubuque	IA	private
146	UNITYPOINT AT HOME	167002	4.25	**0.0674**	dubuque	IA	voluntary non-profit private

147	UNITYPOINT AT HOME	167005	4.00	**0.0674**	cedar rapids	IA	other
148	UNITYPOINT AT HOME	167005	4.00	**0.0674**	cedar rapids	IA	voluntary non-profit other
149	UNITYPOINT AT HOME	167059	4.50	**0.0674**	atlantic	IA	other
150	UNITYPOINT AT HOME	167059	4.50	**0.0674**	atlantic	IA	voluntary non-profit other
151	UNITYPOINT AT HOME	167147	4.00	**0.0674**	waterloo	IA	other
152	UNITYPOINT AT HOME	167147	4.00	**0.0674**	waterloo	IA	voluntary non-profit other
153	UNITYPOINT AT HOME	167148	4.50	**0.0674**	urbandale	IA	private
154	UNITYPOINT AT HOME	167149	5.00	**0.0674**	fort dodge	IA	other
155	UNITYPOINT AT HOME	167159	4.75	**0.0674**	sioux city	IA	other

156	UNITYPOINT AT HOME	167159	4.75	**0.0674**	sioux city	IA	voluntary non-profit other
157	UNITYPOINT AT HOME	167180	4.00	**0.0674**	storm lake	IA	other
158	UNITYPOINT AT HOME	167180	4.00	**0.0674**	storm lake	IA	state or county
159	CHRISTUS HOMECARE	677544	4.25	**0.0654**	new braunfels	TX	religious affiliations
160	CHRISTUS HOMECARE	677544	4.25	**0.0654**	new braunfels	TX	voluntary non-profit church
161	CHRISTUS HOMECARE	677544	4.25	**0.0654**	new braunfels	TX	voluntary non-profit private
162	CHRISTUS HOMECARE	743197	4.50	**0.0654**	coppell	TX	proprietar y
163	PULSE HOMECARE	457884	3.25	**0.0633**	beaumon t	TX	proprietar y

164	GREENWOOD COUNTY HOSPITAL HHA	177243	4.75	**0.0609**	eureka	KS	govt state or county
165	GREENWOOD COUNTY HOSPITAL HHA	177243	4.75	**0.0609**	eureka	KS	state or county
166	FLOYD MEMORIAL HOME HEALTH CARE	157152	4.75	**0.0582**	new albany	IN	govt state or county
167	FLOYD MEMORIAL HOME HEALTH CARE	157152	4.75	**0.0582**	new albany	IN	local
168	FLOYD MEMORIAL HOME HEALTH CARE	157152	4.75	**0.0582**	new albany	IN	state or county
169	RIVERSIDE SHORE HOME HEALTH	497507	4.75	**0.0539**	onley	VA	other

170	RIVERSIDE SHORE HOME HEALTH	497507	4.75	**0.0539**	onley	VA	voluntary non-profit other
171	FORREST GENERAL HOME CARE HHA	257127	4.75	**0.0516**	hattiesbu rg	MS	govt state or county
172	FORREST GENERAL HOME CARE HHA	257127	4.75	**0.0516**	hattiesbu rg	MS	local
173	DALE MEDICAL CTR HOME HEALTH	17124	4.75	**0.0435**	ozark	AL	other
174	DALE MEDICAL CTR HOME HEALTH	17124	4.75	**0.0435**	ozark	AL	voluntary non-profit other
175	NEMAHA COUNTY HOME CARE	287120	5.00	**0.0427**	auburn	NE	govt state or county

176	NEMAHA COUNTY HOME CARE	287120	5.00	**0.0427**	auburn	NE	state or county
177	SSM HOME CARE AT ST MARY'S HEALTH CENTER	267448	5.00	**0.0350**	jefferson city	MO	religious affiliations
178	SSM HOME CARE AT ST MARY'S HEALTH CENTER	267448	5.00	**0.0350**	jefferson city	MO	voluntary non-profit church
179	INTEGRIS HOME CARE ENID	377208	4.75	**0.0322**	enid	OK	religious affiliations
180	INTEGRIS HOME CARE ENID	377208	4.75	**0.0322**	enid	OK	voluntary non-profit church
181	KENOSHA VNA	527024	5.00	**0.0303**	kenosha	WI	private
182	KENOSHA VNA	527024	5.00	**0.0303**	kenosha	WI	voluntary non-profit private

183	EL PASSION HOME HEALTH AGENCY INC	743128	3.00	**0.0299**	arlington	TX	proprietary
184	ONEIDA COUNTY HOSPITAL HOME CARE	137077	4.50	**0.0250**	malad	ID	state or county
185	ONEIDA COUNTY HOSPITAL HOME CARE	137077	4.50	**0.0250**	malad city	ID	govt state or county
186	ONEIDA COUNTY HOSPITAL HOME CARE	137077	4.50	**0.0250**	malad city	ID	state or county
187	DONIPHAN CO HEALTH DEPT HHA	177063	4.75	**0.0243**	troy	KS	govt state or county
188	DONIPHAN CO HEALTH DEPT HHA	177063	4.75	**0.0243**	troy	KS	state or county

189	NATIONAL CHURCH RESIDENCES HOME AND COMMUNITY SERV	368083	4.75	**0.0209**	waverly	OH	other
190	NATIONAL CHURCH RESIDENCES HOME AND COMMUNITY SERV	368083	4.75	**0.0209**	waverly	OH	voluntary non-profit other
191	NATIONAL CHURCH RESIDENCES HOME AND COMMUNITY SERV	368147	4.50	**0.0209**	columbus	OH	other
192	NATIONAL CHURCH RESIDENCES HOME AND	368147	4.50	**0.0209**	columbus	OH	voluntary non-profit other

	COMMUNITY SERV						
193	NATIONAL CHURCH RESIDENCES HOME AND COMMUNITY SERV	368357	3.75	**0.0209**	cuyahoga falls	OH	other
194	NATIONAL CHURCH RESIDENCES HOME AND COMMUNITY SERV	368357	3.75	**0.0209**	cuyahoga falls	OH	religious affiliations
195	ANN'S CHOICE VISITING NURSE SERVICES	398072	5.00	**0.0163**	warminster	PA	proprietary
196	ANN'S CHOICE VISITING NURSE SERVICES	398072	5.00	**0.0163**	warminster	PA	voluntary non-profit private

197	VALLEY OF THE SUN HOME HEALTH CARE LLC	37405	4.00	**0.0160**	phoenix	AZ	proprietary
198	ASPIRUS VNA HOME HEALTH INC	527016	4.50	**0.0108**	wausau	WI	private
199	ASPIRUS VNA HOME HEALTH INC	527016	4.50	**0.0108**	wausau	WI	voluntary non-profit private
200	CARTER HEALTHCARE INC	377478	4.25	**0.0084**	oklahoma city	OK	proprietary
201	CARTER HEALTHCARE INC	377636	4.50	**0.0084**	lawton	OK	private
202	CARTER HEALTHCARE INC	377636	4.50	**0.0084**	lawton	OK	voluntary non-profit private

203	CARTER HEALTHCARE INC	377637	4.50	**0.0084**	tulsa	OK	proprietary
204	CARTER HEALTHCARE INC	377676	4.75	**0.0084**	ardmore	OK	proprietary
205	SELECT HOME CARE LLC	368056	4.00	**0.0013**	columbus	OH	private
206	SELECT HOME CARE LLC	368056	4.00	**0.0013**	columbus	OH	voluntary non-profit private
207	SELECT HOME CARE LLC	37414	5.00	**0.0013**	scottsdale	AZ	proprietary
208	FIRST HEALTH HOME CARE DASH RICHM	347041	4.75	**0.0004**	rockingham	NC	private
209	FIRST HEALTH HOME CARE DASH RICHM	347041	4.75	**0.0004**	rockingham	NC	voluntary non-profit private
210	HAZARD ARH HHA	187027	2.25	**-0.0028**	hazard	KY	other

211	HAZARD ARH HHA	187027	2.25	**-0.0028**	hazard	KY	voluntary non-profit other
212	FAMILY CARE OF TEXAS	679069	2.25	**-0.0042**	waxahachie	TX	other
213	FAMILY CARE OF TEXAS	679069	2.25	**-0.0042**	waxahachie	TX	voluntary non-profit other
214	INTREPID USA HEALTHCARE SERVICES	117073	3.00	**-0.0184**	valdosta	GA	proprietary
215	INTREPID USA HEALTHCARE SERVICES	117141	2.75	**-0.0184**	albany	GA	proprietary
216	INTREPID USA HEALTHCARE SERVICES	117144	2.75	**-0.0184**	brunswick	GA	proprietary
217	INTREPID USA HEALTHCARE SERVICES	157213	3.38	**-0.0184**	indianapolis	IN	proprietary

218	INTREPID USA HEALTHCARE SERVICES	157530	2.75	**-0.0184**	terre haute	IN	proprietary
219	INTREPID USA HEALTHCARE SERVICES	167270	2.75	**-0.0184**	clive	IA	proprietary
220	INTREPID USA HEALTHCARE SERVICES	17063	2.50	**-0.0184**	birmingham	AL	proprietary
221	INTREPID USA HEALTHCARE SERVICES	17063	2.50	**-0.0184**	mountain brook	AL	proprietary
222	INTREPID USA HEALTHCARE SERVICES	17067	3.25	**-0.0184**	montgomery	AL	proprietary
223	INTREPID USA HEALTHCARE SERVICES	187022	2.50	**-0.0184**	somerset	KY	proprietary
224	INTREPID USA HEALTHCARE SERVICES	187085	3.25	**-0.0184**	elizabethtown	KY	proprietary

225	INTREPID USA HEALTHCARE SERVICES	187110	4.00	**-0.0184**	henderson	KY	proprietary
226	INTREPID USA HEALTHCARE SERVICES	187117	3.50	**-0.0184**	murray	KY	private
227	INTREPID USA HEALTHCARE SERVICES	187117	3.50	**-0.0184**	murray	KY	proprietary
228	INTREPID USA HEALTHCARE SERVICES	197579	3.80	**-0.0184**	eunice	LA	proprietary
229	INTREPID USA HEALTHCARE SERVICES	247093	2.75	**-0.0184**	roseville	MN	proprietary
230	INTREPID USA HEALTHCARE SERVICES	247154	3.25	**-0.0184**	saint louis park	MN	proprietary
231	INTREPID USA HEALTHCARE SERVICES	267168	3.25	**-0.0184**	springfield	MO	proprietary

232	INTREPID USA HEALTHCARE SERVICES	347009	2.75	**-0.0184**	raleigh	NC	proprietary
233	INTREPID USA HEALTHCARE SERVICES	367161	2.50	**-0.0184**	elyria	OH	proprietary
234	INTREPID USA HEALTHCARE SERVICES	37112	3.50	**-0.0184**	oro valley	AZ	proprietary
235	INTREPID USA HEALTHCARE SERVICES	37112	3.50	**-0.0184**	tucson	AZ	proprietary
236	INTREPID USA HEALTHCARE SERVICES	397716	3.75	**-0.0184**	lemoyne	PA	proprietary
237	INTREPID USA HEALTHCARE SERVICES	427009	3.25	**-0.0184**	n charleston	SC	proprietary
238	INTREPID USA HEALTHCARE SERVICES	447226	2.75	**-0.0184**	memphis	TN	private

239	INTREPID USA HEALTHCARE SERVICES	447226	2.75	**-0.0184**	memphis	TN	proprietary
240	INTREPID USA HEALTHCARE SERVICES	447226	2.75	**-0.0184**	memphis	TN	voluntary non-profit private
241	INTREPID USA HEALTHCARE SERVICES	447299	3.25	**-0.0184**	nashville	TN	proprietary
242	INTREPID USA HEALTHCARE SERVICES	447442	2.50	**-0.0184**	cookeville	TN	proprietary
243	INTREPID USA HEALTHCARE SERVICES	447454	3.25	**-0.0184**	mc minnville	TN	proprietary
244	INTREPID USA HEALTHCARE SERVICES	447489	4.00	**-0.0184**	sweetwater	TN	proprietary
245	INTREPID USA HEALTHCARE SERVICES	447565	3.50	**-0.0184**	jackson	TN	proprietary

246	INTREPID USA HEALTHCARE SERVICES	47070	2.75	**-0.0184**	little rock	AR	proprietary
247	INTREPID USA HEALTHCARE SERVICES	47088	4.25	**-0.0184**	mc crory	AR	proprietary
248	INTREPID USA HEALTHCARE SERVICES	497060	3.50	**-0.0184**	chester	VA	proprietary
249	INTREPID USA HEALTHCARE SERVICES	497093	3.75	**-0.0184**	onley	VA	proprietary
250	INTREPID USA HEALTHCARE SERVICES	497292	3.25	**-0.0184**	staunton	VA	proprietary
251	INTREPID USA HEALTHCARE SERVICES	497407	4.00	**-0.0184**	radford	VA	proprietary
252	INTREPID USA HEALTHCARE SERVICES	497456	2.50	**-0.0184**	abingdon	VA	proprietary

253	INTREPID USA HEALTHCARE SERVICES	507109	3.75	**-0.0184**	spokane	WA	proprietary
254	INTREPID USA HEALTHCARE SERVICES	517034	3.25	**-0.0184**	weirton	WV	proprietary
255	INTREPID USA HEALTHCARE SERVICES	517034	3.25	**-0.0184**	wheeling	WV	proprietary
256	INTREPID USA HEALTHCARE SERVICES	677267	3.00	**-0.0184**	dallas	TX	proprietary
257	INTREPID USA HEALTHCARE SERVICES	677297	2.75	**-0.0184**	wichita falls	TX	proprietary
258	INTREPID USA HEALTHCARE SERVICES	677616	4.50	**-0.0184**	nederland	TX	proprietary
259	INTREPID USA HEALTHCARE SERVICES	679211	2.75	**-0.0184**	san angelo	TX	proprietary

260	KAISER FOUNDATION TRI CENTRAL HHA	57794	3.00	**-0.0233**	downey	CA	private
261	KAISER FOUNDATION TRI CENTRAL HHA	57794	3.00	**-0.0233**	downey	CA	proprietary
262	KAISER FOUNDATION TRI CENTRAL HHA	57794	3.00	**-0.0233**	downey	CA	voluntary non-profit private
263	BETHANY HOME HEALTH SERVICES	459368	2.25	**-0.0400**	tyler	TX	proprietary
264	BETHANY HOME HEALTH SERVICES	673116	2.50	**-0.0400**	bay city	TX	proprietary
265	BETHANY HOME HEALTH SERVICES	673128	2.75	**-0.0400**	weatherford	TX	proprietary

266	BETHANY HOME HEALTH SERVICES	673128	2.75	**-0.0400**	weatherford	TX	voluntary non-profit private
267	BETHANY HOME HEALTH SERVICES	677598	2.25	**-0.0400**	webster	TX	proprietary
268	BETHANY HOME HEALTH SERVICES	678431	2.25	**-0.0400**	beaumont	TX	proprietary
269	BETHANY HOME HEALTH SERVICES	679139	2.25	**-0.0400**	nacogdoches	TX	proprietary
270	BETHANY HOME HEALTH SERVICES	679139	2.25	**-0.0400**	nacogdoches	TX	voluntary non-profit other
271	BETHANY HOME HEALTH SERVICES	679529	2.25	**-0.0400**	carthage	TX	proprietary
272	BETHANY HOME HEALTH SERVICES	747766	2.00	**-0.0400**	corpus christi	TX	proprietary

273	BETHANY HOME HEALTH SERVICES	747886	2.50	**-0.0400**	texarkana	TX	proprietary
274	ANGELS CARE HOME HEALTH OF THE EMERALD COAST	108181	2.50	**-0.0408**	fort walton beach	FL	private
275	ANGELS CARE HOME HEALTH OF THE EMERALD COAST	108181	2.50	**-0.0408**	fort walton beach	FL	proprietary
276	GENTIVA HEALTH SERVICES	107017	3.75	**-0.0424**	tampa	FL	proprietary
277	GENTIVA HEALTH SERVICES	107098	3.50	**-0.0424**	davie	FL	proprietary

278	GENTIVA HEALTH SERVICES	107098	3.50	**-0.0424**	plantation	FL	proprietary
279	GENTIVA HEALTH SERVICES	107100	3.50	**-0.0424**	pensacola	FL	proprietary
280	GENTIVA HEALTH SERVICES	107132	3.75	**-0.0424**	sarasota	FL	proprietary
281	GENTIVA HEALTH SERVICES	107136	3.75	**-0.0424**	bradenton	FL	proprietary
282	GENTIVA HEALTH SERVICES	107166	4.00	**-0.0424**	lakeland	FL	proprietary
283	GENTIVA HEALTH SERVICES	107171	3.50	**-0.0424**	clearwater	FL	proprietary
284	GENTIVA HEALTH SERVICES	107171	3.50	**-0.0424**	saint petersburg	FL	proprietary

285	GENTIVA HEALTH SERVICES	107200	3.25	**-0.0424**	tallahasse e	FL	proprietar y
286	GENTIVA HEALTH SERVICES	107234	4.00	**-0.0424**	orlando	FL	proprietar y
287	GENTIVA HEALTH SERVICES	107240	3.25	**-0.0424**	fort walton beach	FL	proprietar y
288	GENTIVA HEALTH SERVICES	107275	3.50	**-0.0424**	panama city	FL	proprietar y
289	GENTIVA HEALTH SERVICES	107290	3.50	**-0.0424**	jacksonvill e	FL	proprietar y
290	GENTIVA HEALTH SERVICES	107330	3.75	**-0.0424**	spring hill	FL	proprietar y
291	GENTIVA HEALTH SERVICES	107338	3.75	**-0.0424**	daytona beach	FL	proprietar y

292	GENTIVA HEALTH SERVICES	107420	3.50	**-0.0424**	lake city	FL	proprietary
293	GENTIVA HEALTH SERVICES	107486	3.75	**-0.0424**	gainesville	FL	proprietary
294	GENTIVA HEALTH SERVICES	107512	3.75	**-0.0424**	ocala	FL	proprietary
295	GENTIVA HEALTH SERVICES	107514	3.75	**-0.0424**	palatka	FL	proprietary
296	GENTIVA HEALTH SERVICES	107565	2.75	**-0.0424**	riverview	FL	proprietary
297	GENTIVA HEALTH SERVICES	107586	4.25	**-0.0424**	port saint lucie	FL	proprietary
298	GENTIVA HEALTH SERVICES	108045	3.25	**-0.0424**	leesburg	FL	proprietary

299	GENTIVA HEALTH SERVICES	109102	3.00	**-0.0424**	marianna	FL	proprietary
300	GENTIVA HEALTH SERVICES	117027	3.25	**-0.0424**	atlanta	GA	private
301	GENTIVA HEALTH SERVICES	117027	3.25	**-0.0424**	marietta	GA	private
302	GENTIVA HEALTH SERVICES	117033	3.00	**-0.0424**	statesboro	GA	proprietary
303	GENTIVA HEALTH SERVICES	117056	4.00	**-0.0424**	columbus	GA	proprietary
304	GENTIVA HEALTH SERVICES	117090	3.00	**-0.0424**	savannah	GA	proprietary
305	GENTIVA HEALTH SERVICES	117136	3.75	**-0.0424**	cumming	GA	proprietary

306	GENTIVA HEALTH SERVICES	117137	3.25	**-0.0424**	griffin	GA	proprietary
307	GENTIVA HEALTH SERVICES	117137	3.25	**-0.0424**	peachtree city	GA	proprietary
308	GENTIVA HEALTH SERVICES	117151	3.25	**-0.0424**	bainbridge	GA	proprietary
309	GENTIVA HEALTH SERVICES	147428	4.25	**-0.0424**	rock island	IL	proprietary
310	GENTIVA HEALTH SERVICES	157115	3.00	**-0.0424**	indianapolis	IN	proprietary
311	GENTIVA HEALTH SERVICES	157180	3.50	**-0.0424**	muncie	IN	proprietary
312	GENTIVA HEALTH SERVICES	167124	3.75	**-0.0424**	hiawatha	IA	proprietary

313	GENTIVA HEALTH SERVICES	167187	2.75	**-0.0424**	des moines	IA	proprietary
314	GENTIVA HEALTH SERVICES	17013	3.25	**-0.0424**	enterprise	AL	proprietary
315	GENTIVA HEALTH SERVICES	17018	4.25	**-0.0424**	rainbow city	AL	proprietary
316	GENTIVA HEALTH SERVICES	17048	3.25	**-0.0424**	geneva	AL	proprietary
317	GENTIVA HEALTH SERVICES	17052	3.75	**-0.0424**	foley	AL	proprietary
318	GENTIVA HEALTH SERVICES	17055	4.75	**-0.0424**	prattville	AL	proprietary
319	GENTIVA HEALTH SERVICES	17056	4.25	**-0.0424**	pike road	AL	proprietary

320	GENTIVA HEALTH SERVICES	17058	3.25	**-0.0424**	daphne	AL	proprietary
321	GENTIVA HEALTH SERVICES	17071	4.00	**-0.0424**	moulton	AL	proprietary
322	GENTIVA HEALTH SERVICES	17090	3.25	**-0.0424**	gilbertown	AL	proprietary
323	GENTIVA HEALTH SERVICES	17121	3.50	**-0.0424**	andalusia	AL	proprietary
324	GENTIVA HEALTH SERVICES	17151	4.25	**-0.0424**	anniston	AL	proprietary
325	GENTIVA HEALTH SERVICES	17152	4.50	**-0.0424**	sylacauga	AL	proprietary
326	GENTIVA HEALTH SERVICES	17161	3.50	**-0.0424**	huntsville	AL	proprietary

327	GENTIVA HEALTH SERVICES	17164	3.75	**-0.0424**	dothan	AL	proprietary
328	GENTIVA HEALTH SERVICES	17166	3.75	**-0.0424**	clanton	AL	proprietary
329	GENTIVA HEALTH SERVICES	17167	3.75	**-0.0424**	pell city	AL	proprietary
330	GENTIVA HEALTH SERVICES	17168	3.50	**-0.0424**	mobile	AL	proprietary
331	GENTIVA HEALTH SERVICES	17170	3.75	**-0.0424**	jasper	AL	proprietary
332	GENTIVA HEALTH SERVICES	17171	4.25	**-0.0424**	cullman	AL	proprietary
333	GENTIVA HEALTH SERVICES	17307	3.75	**-0.0424**	phenix city	AL	proprietary

334	GENTIVA HEALTH SERVICES	178056	4.25	**-0.0424**	wichita	KS	proprietary
335	GENTIVA HEALTH SERVICES	187058	3.75	**-0.0424**	fort mitchell	KY	proprietary
336	GENTIVA HEALTH SERVICES	187090	3.00	**-0.0424**	louisville	KY	proprietary
337	GENTIVA HEALTH SERVICES	187095	3.00	**-0.0424**	lexington	KY	proprietary
338	GENTIVA HEALTH SERVICES	187124	4.00	**-0.0424**	hopkinsville	KY	proprietary
339	GENTIVA HEALTH SERVICES	197443	3.00	**-0.0424**	lake charles	LA	proprietary
340	GENTIVA HEALTH SERVICES	207038	3.75	**-0.0424**	portland	ME	proprietary

341	GENTIVA HEALTH SERVICES	207045	3.25	**-0.0424**	bangor	ME	proprietary
342	GENTIVA HEALTH SERVICES	227219	3.75	**-0.0424**	pittsfield	MA	proprietary
343	GENTIVA HEALTH SERVICES	227224	3.00	**-0.0424**	springfield	MA	proprietary
344	GENTIVA HEALTH SERVICES	227260	3.25	**-0.0424**	fall river	MA	proprietary
345	GENTIVA HEALTH SERVICES	227426	4.50	**-0.0424**	hyannis	MA	proprietary
346	GENTIVA HEALTH SERVICES	237085	2.75	**-0.0424**	kalamazoo	MI	proprietary
347	GENTIVA HEALTH SERVICES	237135	3.00	**-0.0424**	muskegon	MI	proprietary

348	GENTIVA HEALTH SERVICES	237135	3.00	**-0.0424**	muskegon heights	MI	proprietary
349	GENTIVA HEALTH SERVICES	237136	3.50	**-0.0424**	flint	MI	proprietary
350	GENTIVA HEALTH SERVICES	237222	3.25	**-0.0424**	grand rapids	MI	proprietary
351	GENTIVA HEALTH SERVICES	247145	4.50	**-0.0424**	roseville	MN	proprietary
352	GENTIVA HEALTH SERVICES	247241	3.00	**-0.0424**	duluth	MN	proprietary
353	GENTIVA HEALTH SERVICES	257110	3.50	**-0.0424**	columbus	MS	proprietary
354	GENTIVA HEALTH SERVICES	257126	2.75	**-0.0424**	flowood	MS	proprietary

355	GENTIVA HEALTH SERVICES	257300	3.00	**-0.0424**	meridian	MS	proprietary
356	GENTIVA HEALTH SERVICES	257303	2.75	**-0.0424**	tupelo	MS	proprietary
357	GENTIVA HEALTH SERVICES	257304	3.00	**-0.0424**	hazlehurst	MS	proprietary
358	GENTIVA HEALTH SERVICES	257309	4.25	**-0.0424**	calhoun city	MS	proprietary
359	GENTIVA HEALTH SERVICES	267098	3.75	**-0.0424**	rolla	MO	proprietary
360	GENTIVA HEALTH SERVICES	267290	3.00	**-0.0424**	creve coeur	MO	proprietary
361	GENTIVA HEALTH SERVICES	267584	3.25	**-0.0424**	columbia	MO	proprietary

362	GENTIVA HEALTH SERVICES	267639	3.00	**-0.0424**	independence	MO	proprietary
363	GENTIVA HEALTH SERVICES	267639	3.00	**-0.0424**	lees summit	MO	proprietary
364	GENTIVA HEALTH SERVICES	287038	3.50	**-0.0424**	omaha	NE	proprietary
365	GENTIVA HEALTH SERVICES	287060	2.75	**-0.0424**	lincoln	NE	proprietary
366	GENTIVA HEALTH SERVICES	327070	3.25	**-0.0424**	albuquerque	NM	proprietary
367	GENTIVA HEALTH SERVICES	327182	3.00	**-0.0424**	las cruces	NM	proprietary
368	GENTIVA HEALTH SERVICES	347031	3.00	**-0.0424**	greenville	NC	proprietary

369	GENTIVA HEALTH SERVICES	347075	3.00	**-0.0424**	greensboro	NC	proprietary
370	GENTIVA HEALTH SERVICES	347124	3.00	**-0.0424**	kinston	NC	proprietary
371	GENTIVA HEALTH SERVICES	347171	3.25	**-0.0424**	morehead city	NC	proprietary
372	GENTIVA HEALTH SERVICES	347178	2.75	**-0.0424**	raleigh	NC	proprietary
373	GENTIVA HEALTH SERVICES	347217	3.25	**-0.0424**	youngsville	NC	proprietary
374	GENTIVA HEALTH SERVICES	347225	3.50	**-0.0424**	boone	NC	private
375	GENTIVA HEALTH SERVICES	347226	2.75	**-0.0424**	pink hill	NC	proprietary

376	GENTIVA HEALTH SERVICES	347235	3.50	**-0.0424**	asheville	NC	proprietary
377	GENTIVA HEALTH SERVICES	347236	3.25	**-0.0424**	durham	NC	proprietary
378	GENTIVA HEALTH SERVICES	347300	3.25	**-0.0424**	hickory	NC	proprietary
379	GENTIVA HEALTH SERVICES	347317	3.25	**-0.0424**	shelby	NC	proprietary
380	GENTIVA HEALTH SERVICES	347328	3.25	**-0.0424**	rocky mount	NC	proprietary
381	GENTIVA HEALTH SERVICES	347329	3.25	**-0.0424**	washington	NC	proprietary
382	GENTIVA HEALTH SERVICES	347330	3.25	**-0.0424**	king	NC	proprietary

383	GENTIVA HEALTH SERVICES	347331	3.25	**-0.0424**	goldsboro	NC	proprietary
384	GENTIVA HEALTH SERVICES	347333	2.75	**-0.0424**	pollocksville	NC	proprietary
385	GENTIVA HEALTH SERVICES	367159	4.00	**-0.0424**	akron	OH	proprietary
386	GENTIVA HEALTH SERVICES	367520	3.75	**-0.0424**	maumee	OH	proprietary
387	GENTIVA HEALTH SERVICES	37036	4.00	**-0.0424**	tucson	AZ	proprietary
388	GENTIVA HEALTH SERVICES	37037	3.50	**-0.0424**	phoenix	AZ	proprietary
389	GENTIVA HEALTH SERVICES	377170	2.75	**-0.0424**	oklahoma city	OK	proprietary

390	GENTIVA HEALTH SERVICES	377442	3.25	**-0.0424**	tulsa	OK	proprietary
391	GENTIVA HEALTH SERVICES	397237	4.00	**-0.0424**	lancaster	PA	proprietary
392	GENTIVA HEALTH SERVICES	397422	3.75	**-0.0424**	wilkes barre	PA	proprietary
393	GENTIVA HEALTH SERVICES	398005	3.75	**-0.0424**	stroudsburg	PA	proprietary
394	GENTIVA HEALTH SERVICES	427017	3.25	**-0.0424**	greenville	SC	proprietary
395	GENTIVA HEALTH SERVICES	427035	3.00	**-0.0424**	north charleston	SC	proprietary
396	GENTIVA HEALTH SERVICES	427045	3.50	**-0.0424**	myrtle beach	SC	proprietary

397	GENTIVA HEALTH SERVICES	427061	3.00	**-0.0424**	columbia	SC	proprietary
398	GENTIVA HEALTH SERVICES	427117	3.50	**-0.0424**	gaffney	SC	proprietary
399	GENTIVA HEALTH SERVICES	447151	3.00	**-0.0424**	tullahoma	TN	proprietary
400	GENTIVA HEALTH SERVICES	457264	3.00	**-0.0424**	humble	TX	proprietary
401	GENTIVA HEALTH SERVICES	457416	3.25	**-0.0424**	san antonio	TX	proprietary
402	GENTIVA HEALTH SERVICES	467015	4.00	**-0.0424**	st george	UT	proprietary
403	GENTIVA HEALTH SERVICES	47029	3.25	**-0.0424**	little rock	AR	proprietary

404	GENTIVA HEALTH SERVICES	47038	3.50	**-0.0424**	fort smith	AR	proprietary
405	GENTIVA HEALTH SERVICES	47105	3.00	**-0.0424**	hot springs	AR	proprietary
406	GENTIVA HEALTH SERVICES	497429	3.50	**-0.0424**	roanoke	VA	proprietary
407	GENTIVA HEALTH SERVICES	507021	3.25	**-0.0424**	liberty lake	WA	proprietary
408	GENTIVA HEALTH SERVICES	507031	3.00	**-0.0424**	tacoma	WA	proprietary
409	GENTIVA HEALTH SERVICES	507071	3.00	**-0.0424**	everett	WA	proprietary
410	GENTIVA HEALTH SERVICES	507075	3.25	**-0.0424**	spokane	WA	proprietary

411	GENTIVA HEALTH SERVICES	507081	2.75	**-0.0424**	vancouver	WA	proprietary
412	GENTIVA HEALTH SERVICES	507082	3.00	**-0.0424**	kent	WA	proprietary
413	GENTIVA HEALTH SERVICES	507106	3.25	**-0.0424**	bremerton	WA	proprietary
414	GENTIVA HEALTH SERVICES	517048	3.25	**-0.0424**	charleston	WV	proprietary
415	GENTIVA HEALTH SERVICES	517060	3.50	**-0.0424**	chapmanville	WV	proprietary
416	GENTIVA HEALTH SERVICES	517087	4.00	**-0.0424**	parkersburg	WV	proprietary
417	GENTIVA HEALTH SERVICES	517120	2.75	**-0.0424**	summersville	WV	proprietary

418	GENTIVA HEALTH SERVICES	517125	3.00	**-0.0424**	beckley	WV	proprietary
419	GENTIVA HEALTH SERVICES	517126	2.75	**-0.0424**	huntington	WV	proprietary
420	GENTIVA HEALTH SERVICES	527098	4.00	**-0.0424**	racine	WI	proprietary
421	GENTIVA HEALTH SERVICES	527207	2.75	**-0.0424**	west allis	WI	proprietary
422	GENTIVA HEALTH SERVICES	557052	3.25	**-0.0424**	san jose	CA	proprietary
423	GENTIVA HEALTH SERVICES	557104	3.75	**-0.0424**	irvine	CA	proprietary
424	GENTIVA HEALTH SERVICES	557104	3.75	**-0.0424**	santa ana	CA	proprietary

425	GENTIVA HEALTH SERVICES	557139	4.75	**-0.0424**	el centro	CA	proprietary
426	GENTIVA HEALTH SERVICES	57036	3.50	**-0.0424**	santa rosa	CA	proprietary
427	GENTIVA HEALTH SERVICES	57143	3.75	**-0.0424**	san diego	CA	proprietary
428	GENTIVA HEALTH SERVICES	57203	3.25	**-0.0424**	san luis obispo	CA	proprietary
429	GENTIVA HEALTH SERVICES	67129	3.75	**-0.0424**	colorado springs	CO	proprietary
430	GENTIVA HEALTH SERVICES	67196	3.50	**-0.0424**	grand junction	CO	proprietary
431	GENTIVA HEALTH SERVICES	677166	3.00	**-0.0424**	austin	TX	proprietary

432	GENTIVA HEALTH SERVICES	77162	2.75	**-0.0424**	farmington	CT	proprietary
433	GENTIVA HEALTH SERVICES	77218	3.75	**-0.0424**	stratford	CT	proprietary
434	GENTIVA HEALTH SERVICES	77218	3.75	**-0.0424**	trumbull	CT	proprietary
435	DEACONESS HOMECARE	187174	2.75	**-0.0434**	lexington	KY	proprietary
436	DEACONESS HOMECARE	257085	2.75	**-0.0434**	hattiesburg	MS	proprietary
437	DEACONESS HOMECARE	257135	2.25	**-0.0434**	brookhaven	MS	proprietary
438	DEACONESS HOMECARE	447109	2.50	**-0.0434**	fayetteville	TN	proprietary
439	DEACONESS HOMECARE	447498	2.75	**-0.0434**	oneida	TN	proprietary
440	SAMPSON HOME HEALTH	347064	2.25	**-0.0451**	clinton	NC	govt state or county

441	SAMPSON HOME HEALTH	347064	2.25	**-0.0451**	clinton	NC	other
442	SAMPSON HOME HEALTH	347064	2.25	**-0.0451**	clinton	NC	state or county
443	HEALTH FIRST HOME CARE	107233	3.00	**-0.0543**	merritt island	FL	other
444	HEALTH FIRST HOME CARE	107233	3.00	**-0.0543**	merritt island	FL	private
445	HEALTH FIRST HOME CARE	107233	3.00	**-0.0543**	merritt island	FL	voluntary non-profit other
446	HEALTH FIRST HOME CARE	109443	3.50	**-0.0543**	sebastian	FL	other
447	HEALTH FIRST HOME CARE	109443	3.50	**-0.0543**	sebastian	FL	voluntary non-profit other
448	VIVA HOME HEALTH CARE INC	103129	2.50	**-0.0546**	tampa	FL	proprietar y

449	NU DASH ERA HOME HEALTH AGENCY INC	58222	2.75	**-0.0583**	torrance	CA	proprietary
450	PERRY COUNTY HEALTH DEPARTMENT	147161	2.75	**-0.0621**	pinckneyville	IL	govt state or county
451	PERRY COUNTY HEALTH DEPARTMENT	147161	2.75	**-0.0621**	pinckneyville	IL	state or county
452	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457050	2.25	**-0.0669**	austin	TX	proprietary
453	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457096	2.50	**-0.0669**	temple	TX	govt state or county
454	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457096	2.50	**-0.0669**	temple	TX	proprietary

455	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457096	2.50	**-0.0669**	temple	TX	state or county
456	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457126	2.75	**-0.0669**	corpus christi	TX	proprietar y
457	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457128	3.00	**-0.0669**	eastland	TX	proprietar y
458	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457222	2.75	**-0.0669**	bellaire	TX	proprietar y
459	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	457223	3.00	**-0.0669**	dallas	TX	proprietar y

460	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	459496	3.00	**-0.0669**	lubbock	TX	proprietary
461	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	678039	3.50	**-0.0669**	beaumont	TX	proprietary
462	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	679074	2.25	**-0.0669**	san antonio	TX	proprietary
463	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	679096	2.50	**-0.0669**	fort worth	TX	proprietary
464	GIRLING HOME HEALTH TEXAS BY HARDEN HEALTHCARE	747708	2.75	**-0.0669**	bryan	TX	proprietary

465	FIDELITY HEALTH CARE INC	148171	2.25	**-0.0697**	evergreen park	IL	proprietary
466	FIDELITY HEALTH CARE INC	148171	2.25	**-0.0697**	evergreen park	IL	voluntary non-profit private
467	FIDELITY HEALTH CARE INC	148171	2.25	**-0.0697**	orland park	IL	voluntary non-profit private
468	FAMILY HOSPICE AND PALLATIVE CARE	397517	2.00	**-0.0808**	pittsburgh	PA	other
469	FAMILY HOSPICE AND PALLATIVE CARE	397517	2.00	**-0.0808**	pittsburgh	PA	voluntary non-profit other
470	PINNACLE SENIOR CARE	108117	3.50	**-0.0829**	jacksonville	FL	proprietary
471	PINNACLE SENIOR CARE	108117	3.50	**-0.0829**	jacksonville beach	FL	proprietary

472	PINNACLE SENIOR CARE	237202	2.75	**-0.0829**	ann arbor	MI	proprietary
473	PINNACLE SENIOR CARE	237676	3.00	**-0.0829**	grand rapids	MI	proprietary
474	PINNACLE SENIOR CARE	237677	3.00	**-0.0829**	bay city	MI	proprietary
475	PINNACLE SENIOR CARE	237678	2.50	**-0.0829**	okemos	MI	proprietary
476	PINNACLE SENIOR CARE	267643	2.75	**-0.0829**	saint louis	MO	proprietary
477	PINNACLE SENIOR CARE	368140	2.75	**-0.0829**	mentor	OH	proprietary
478	PINNACLE SENIOR CARE	368192	2.25	**-0.0829**	boardman	OH	proprietary
479	PINNACLE SENIOR CARE	368210	3.00	**-0.0829**	cincinnati	OH	proprietary
480	PINNACLE SENIOR CARE	368289	2.75	**-0.0829**	columbus	OH	proprietary
481	PINNACLE SENIOR CARE	457840	2.50	**-0.0829**	corpus christi	TX	private

482	PINNACLE SENIOR CARE	677950	4.00	**-0.0829**	austin	TX	proprietary
483	PINNACLE SENIOR CARE	677950	4.00	**-0.0829**	dripping springs	TX	proprietary
484	PINNACLE SENIOR CARE	678318	2.75	**-0.0829**	irving	TX	proprietary
485	PINNACLE SENIOR CARE	679132	3.00	**-0.0829**	houston	TX	proprietary
486	PINNACLE SENIOR CARE	747021	2.50	**-0.0829**	san antonio	TX	proprietary
487	TEHC LLC	107704	2.50	**-0.0847**	jacksonville	FL	proprietary
488	TEHC LLC	107717	3.25	**-0.0847**	orlando	FL	proprietary
489	TEHC LLC	107717	3.25	**-0.0847**	winter park	FL	proprietary
490	TEHC LLC	107718	2.75	**-0.0847**	cocoa	FL	proprietary
491	TEHC LLC	107718	2.75	**-0.0847**	rockledge	FL	proprietary

492	TEHC LLC	107773	3.25	**-0.0847**	doral	FL	proprietary
493	TEHC LLC	107773	3.25	**-0.0847**	miami	FL	proprietary
494	EXCEED HOME HEALTH INC	59156	2.75	**-0.0907**	studio city	CA	proprietary
495	EXCEED HOME HEALTH INC	59156	2.75	**-0.0907**	studio city	CA	voluntary non-profit private
496	LAKEVIEW HOME HEALTH CARE	387141	2.25	**-0.0992**	lakeview	OR	other
497	LAKEVIEW HOME HEALTH CARE	387141	2.25	**-0.0992**	lakeview	OR	voluntary non-profit other
498	YUMA DISTRICT HOSPITAL HOME HEALTH CARE	67161	2.25	**-0.1119**	yuma	CO	govt local
499	YUMA DISTRICT HOSPITAL	67161	2.25	**-0.1119**	yuma	CO	local

	HOME HEALTH CARE						
500	JORDAN HEALTH SERVICES	453166	3.00	**-0.1138**	arlington	TX	proprietary
501	JORDAN HEALTH SERVICES	457015	3.25	**-0.1138**	texarkana	TX	proprietary
502	JORDAN HEALTH SERVICES	457087	3.00	**-0.1138**	waxahachie	TX	proprietary
503	JORDAN HEALTH SERVICES	457507	2.25	**-0.1138**	mount vernon	TX	proprietary
504	JORDAN HEALTH SERVICES	457882	2.75	**-0.1138**	conroe	TX	proprietary
505	JORDAN HEALTH SERVICES	457882	2.75	**-0.1138**	lufkin	TX	proprietary

506	JORDAN HEALTH SERVICES	458346	2.25	**-0.1138**	palestine	TX	other
507	JORDAN HEALTH SERVICES	458346	2.25	**-0.1138**	palestine	TX	proprietary
508	JORDAN HEALTH SERVICES	47114	3.25	**-0.1138**	texarkana	AR	proprietary
509	JORDAN HEALTH SERVICES	673113	3.25	**-0.1138**	addison	TX	proprietary
510	JORDAN HEALTH SERVICES	677660	2.25	**-0.1138**	jasper	TX	proprietary
511	JORDAN HEALTH SERVICES	677720	2.75	**-0.1138**	sherman	TX	proprietary
512	JORDAN HEALTH SERVICES	679221	3.00	**-0.1138**	denton	TX	proprietary

513	JORDAN HEALTH SERVICES	679275	3.80	**-0.1138**	san antonio	TX	proprietary
514	JORDAN HEALTH SERVICES	747046	2.50	**-0.1138**	fairfield	TX	proprietary
515	JORDAN HEALTH SERVICES	747046	2.50	**-0.1138**	waco	TX	proprietary
516	CENTRAL FLORIDA QUALITY CARE SERVICES INC	109607	2.50	**-0.1769**	orlando	FL	proprietary
517	NURSES CHOICE HOME CARE	58362	1.75	**-0.2102**	sacramento	CA	proprietary
518	SOUTH FLORIDA HOME HEALTH CARE INC	108118	2.00	**-0.2266**	miami	FL	proprietary
519	VERITAS HOME CARE INC	109722	2.83	**-0.2489**	boynton beach	FL	private

520	VERITAS HOME CARE INC	109722	2.83	**-0.2489**	boynton beach	FL	proprietary
521	VERITAS HOME CARE INC	109722	2.83	**-0.2489**	boynton beach	FL	voluntary non-profit private
522	CAREALL	447225	2.25	**-0.2743**	martin	TN	proprietary
523	CAREALL	447503	2.25	**-0.2743**	brownsville	TN	proprietary
524	CAREALL	447503	2.25	**-0.2743**	covington	TN	proprietary
525	PRUDENT HEALTHCARE AGENCY	747352	2.25	**-0.4987**	mesquite	TX	proprietary
526	WISDOM HEALTH CARE SERVICES INC	59060	2.50	**-0.5034**	gardena	CA	private
527	WISDOM HEALTH CARE SERVICES INC	59060	2.50	**-0.5034**	gardena	CA	voluntary non-profit private

528	REGISTERED NURSES CARE LTD	368071	2.75	**-0.5214**	westerville	OH	proprietary
529	GOOD SAMARITAN HOME HEALTH CARE INC	679478	2.50	**-0.7342**	dallas	TX	proprietary
530	ANGEL CARE HOME HEALTH SERVICES INC	459412	2.25	**-0.8029**	grand prairie	TX	proprietary

Addendum 2 – Lasso Estimates for all 201 Explanatory Predictors

The list shown below are for the 201 explanatory variables on the Star Rating identified by the Lasso model fitting. The list is sorted by Coefficient from largest value to smallest value.

This list provides a practical approach to provide predictive estimates for each Provider and it can be used by the Recommendation Engine.

The estimates with positive or increasing explanatory impact on the Star Rating are highlighted in light green. The estimates with negative or decreasing explanatory impact on the Star Rating are highlighted in light orange.

	Predictor Name	Coefficient	When the Predictor is 1, the Star Rating increases or decreases by this quantity, given that all other predictors remaining the same
	(Intercept)	2.0585733	
1	Provider.Namecari home care inc	0.7731745	0.7731745
2	Provider.Namevolunteers of america home health at westchester	0.5803818	0.5803818
3	Provider.Namedynamic home health care	0.5113720	0.5113720
4	Provider.Namesierra home health care	0.4919851	0.4919851
5	Statepr	0.4128922	0.4128922
6	Cityhayesville	0.3955761	0.3955761
7	Provider.Namehealing to heal another home healthcare inc	0.3780465	0.3780465
8	Provider.Nameinnovative senior care home health	0.3331348	0.3331348
9	Provider.Namefranklin hospital medical center chha	0.3161551	0.3161551

10	Provider.Namebest home healthcare network inc	0.3144870	0.3144870
11	Provider.Namebeatitudes home health	0.3104478	0.3104478
12	Provider.Nameglobal home health i llc	0.2911843	0.2911843
13	Provider.Namehumanity home health inc	0.2629150	0.2629150
14	Citysikeston	0.2513205	0.2513205
15	Provider.Namebaycare home care inc	0.2154090	0.2154090
16	Provider.Namelong life home care inc	0.2093651	0.2093651
17	Provider.Namecedar crest village inc home health department	0.2060878	0.2060878
18	Provider.Namequality life home health agency corp	0.2030257	0.2030257
19	Provider.Namebreeze health care inc.	0.2022800	0.2022800
20	Citysioux city	0.1856905	0.1856905
21	Citymiami	0.1825823	0.1825823
22	Cityrhinelander	0.1778902	0.1778902
23	Provider.Namebaptist health home health dash heber springs	0.1679078	0.1679078
24	Cityjefferson city	0.1667217	0.1667217
25	Provider.Namequality care home health services inc	0.1628156	0.1628156

26	Citycreighton	0.1616121	0.1616121
27	Provider.Namebrmc home care	0.1582233	0.1582233
28	Provider.Namevna nazareth home care	0.1558560	0.1558560
29	Citymarshfield	0.1532191	0.1532191
30	Cityardmore	0.1475824	0.1475824
31	Statefl	0.1470144	0.1470144
32	Provider.Namedoctors home care	0.1459970	0.1459970
33	Provider.Namest francis home health care	0.1425642	0.1425642
34	Citysturgeon bay	0.1393358	0.1393358
35	Provider.Nameunion regional home care	0.1353807	0.1353807
36	Provider.Namehealthwatch home health of weatherford llc	0.1336714	0.1336714
37	Citybeverly	0.1336515	0.1336515
38	Cityspring hill	0.1332345	0.1332345
39	Provider.Nameomhs home care services	0.1307711	0.1307711
40	Provider.Namemacy's health services inc	0.1288125	0.1288125
41	Citysanta cruz	0.1223537	0.1223537
42	Provider.Namessm home care at st francis hospital	0.1223415	0.1223415
43	Citywaverly	0.1200731	0.1200731
44	Citypanorama city	0.1196433	0.1196433

45	Provider.Namebig bend home health	0.1191466	0.1191466
46	Citytuscaloosa	0.1188728	0.1188728
47	Stateca	0.1167496	0.1167496
48	Citymanitowoc	0.1160000	0.1160000
49	Provider.Namewaldo county home healthcare services	0.1146922	0.1146922
50	Provider.Namehome care of the grand valley	0.1118425	0.1118425
51	Provider.Namefairview lakes homecaring hosp	0.1099815	0.1099815
52	Citynew boston	0.1099598	0.1099598
53	Provider.Nameasset health services inc	0.1077478	0.1077478
54	Citynorristown	0.1074255	0.1074255
55	Provider.Nametrinity home care	0.1067470	0.1067470
56	Provider.Namegolden years home care inc.	0.1058148	0.1058148
57	Provider.Namedpn homecare	0.1047823	0.1047823
58	Provider.Namekings home healthcare inc	0.1010493	0.1010493
59	Provider.Namenorth kansas city hospital home health services	0.1005996	0.1005996
60	Provider.Namecaring like family inc	0.0982659	0.0982659
61	Provider.Nameregional home care helena	0.0938067	0.0938067
62	Provider.Nameteam select home care	0.0922132	0.0922132

63	Provider.Namekind hands inc	0.0899705	0.0899705
64	Statenj	0.0885633	0.0885633
65	Provider.Namemount auburn home health	0.0861266	0.0861266
66	Provider.Namecortland regional medical center inc lthhcp	0.0812097	0.0812097
67	Provider.Nameomni home care agency inc	0.0806357	0.0806357
68	Provider.Namesupreme patient care inc	0.0743465	0.0743465
69	Cityfestus	0.0721868	0.0721868
70	Provider.Namebrookdale home health holland	0.0712420	0.0712420
71	Cityhialeah	0.0683208	0.0683208
72	Provider.Nameunitypoint at home	0.0673562	0.0673562
73	Provider.Namechristus homecare	0.0653344	0.0653344
74	Provider.Namepulse homecare	0.0632915	0.0632915
75	Citygalesburg	0.0626912	0.0626912
76	Provider.Namegreenwood county hospital hha	0.0608329	0.0608329
77	Provider.Namefloyd memorial home health care	0.0581612	0.0581612
78	Citycarlsbad	0.0552801	0.0552801
79	Provider.Nameriverside shore home health	0.0538229	0.0538229

80	Citywausau	0.0518303	0.0518303
81	Provider.Nameforrest general home care hha	0.0515493	0.0515493
82	Citydel city	0.0489293	0.0489293
83	Type.of.Ownershipvoluntary non-profit church	0.0486694	0.0486694
84	Statesd	0.0457074	0.0457074
85	Stateut	0.0447321	0.0447321
86	Provider.Namedale medical ctr home health	0.0434594	0.0434594
87	Provider.Namenemaha county home care	0.0426767	0.0426767
88	Statemi	0.0394908	0.0394908
89	Citylaconia	0.0369632	0.0369632
90	Provider.Namessm home care at st mary's health center	0.0349488	0.0349488
91	Type.of.Ownershipvoluntary non-profit other	0.0349366	0.0349366
92	Provider.Nameintegris home care enid	0.0321602	0.0321602
93	Provider.Namekenosha vna	0.0302877	0.0302877
94	Provider.Nameel passion home health agency inc	0.0298047	0.0298047
95	Citygrand rapids	0.0259494	0.0259494
96	Citytyler	0.0253397	0.0253397

97	Provider.Nameoneida county hospital home care	0.0249744	0.0249744
98	Cityglendale	0.0244789	0.0244789
99	Provider.Namedoniphan co health dept hha	0.0242080	0.0242080
100	Citysweetwater	0.0235829	0.0235829
101	Citybeatrice	0.0229809	0.0229809
102	Provider.Namenational church residences home and community serv	0.0208903	0.0208903
103	Provider.Nameann's choice visiting nurse services	0.0162760	0.0162760
104	Provider.Namevalley of the sun home health care llc	0.0159635	0.0159635
105	Type.of.Ownershipvoluntary non-profit private	0.0123852	0.0123852
106	Citypierre	0.0115789	0.0115789
107	Max.of.HO.HHT.began.care.in.timely.manner.18	0.0112865	0.0112865
108	Statemd	0.0108166	0.0108166
109	Provider.Nameaspirus vna home health inc	0.0107088	0.0107088
110	Stateil	0.0087165	0.0087165
111	Provider.Namecarter healthcare inc	0.0083450	0.0083450

112	Cityegg harbor township	0.0078620	0.0078620
113	Max.of.HO.PAT.got.better.at.getting.in.and.out.of.bed.46	0.0067549	0.0067549
114	Max.of.HO.PAT.had.less.pain.when.moving.around.50	0.0053250	0.0053250
115	Max.of.HO.PAT..got.better.at.bathing.48	0.0050605	0.0050605
116	Citymiami lakes	0.0049840	0.0049840
117	Cityhattiesburg	0.0040600	0.0040600
118	Max.of.HO.HHT.made.received.pneumonia.shot.28	0.0036813	0.0036813
119	Max.of.HO.HHT.taught.about.their.drugs.20	0.0033500	0.0033500
120	Max.of.HO.PAT.breathing.improved.52	0.0032995	0.0032995
121	Statepa	0.0025349	0.0025349
122	Max.of.HO.PAT.got.better.at.taking.drugs.correctly.56	0.0022960	0.0022960
123	Max.of.HO.PAT.got.better.at.moving.around.44	0.0018371	0.0018371
124	Citysouthfield	0.0017110	0.0017110
125	Provider.Nameselect home care llc	0.0012933	0.0012933
126	Max.of.HO.HHT.ensured.received.flu.shot.26	0.0009322	0.0009322
127	Citylynbrook	0.0007667	0.0007667

128	Citypompton plains	0.0006123	0.0006123
129	Max.of.HO.PAT.wounds.improved.healed.54	0.0006093	0.0006093
130	Citynorth kansas city	0.0005017	0.0005017
131	Citykenosha	0.0004902	0.0004902
132	Provider.Namefirst health home care dash richm	0.0003817	0.0003817
133	Citybrownfield	0.0003447	0.0003447
134	Citychisago city	0.0002862	0.0002862
135	Citycortland	0.0002271	0.0002271
136	Max.of.HO.HHT.treated.heart.failure.weakening.of.the.heart.36	0.0001708	0.0001708
137	Citybelfast	0.0000484	0.0000484
138	Cityrockingham	0.0000044	0.0000044
139	Max.of.HO.HHT.checked.for.risk.of.pressure.bed.sores.42	-0.0001907	-0.0001907
140	Max.of.HO.HHT.included.treatm.to.prevent.pressure.bed.sores.40	-0.0009499	-0.0009499
141	Max.of.HO.HHT.treated.for.pain.34	-0.0026823	-0.0026823
142	Provider.Namehazard arh hha	-0.0027872	-0.0027872
143	Provider.Namefamily care of texas	-0.0041703	-0.0041703
144	Max.of.HO.HHT.checked.for.risk.of.falling.22	-0.0043072	-0.0043072

145	Stateok	-0.0053929	-0.0053929
146	Max.of.HO.PAT.needed.urgent.unplanned.ER .wout.admission.58	-0.0053983	-0.0053983
147	Citysan antonio	-0.0067416	-0.0067416
148	Max.of.HO.HHT.checked.for.pain.32	-0.0070938	-0.0070938
149	Citysherman	-0.0093937	-0.0093937
150	Max.of.HO.PAT.had.to.be.admitted.to.the.ho spital.60	-0.0155853	-0.0155853
151	Provider.Nameintrepid usa healthcare services	-0.0183168	-0.0183168
152	Provider.Namejm homecare solutions inc	-0.0191238	-0.0191238
153	Provider.Namekaiser foundation tri central hha	-0.0232461	-0.0232461
154	Type.of.Ownershipproprietary	-0.0232827	-0.0232827
155	Citychandler	-0.0243504	-0.0243504
156	Type.of.Ownershipstate or county	-0.0275595	-0.0275595
157	Stateak	-0.0296517	-0.0296517
158	Statela	-0.0306521	-0.0306521
159	Statear	-0.0324968	-0.0324968
160	Statein	-0.0394663	-0.0394663
161	Citycleveland	-0.0394928	-0.0394928

162	Provider.Namebethany home health services	-0.0399297	-0.0399297
163	Provider.Nameangels care home health of the emerald coast	-0.0407535	-0.0407535
164	Provider.Namegentiva health services	-0.0423720	-0.0423720
165	Statenc	-0.0427486	-0.0427486
166	Provider.Namedeaconess homecare	-0.0433897	-0.0433897
167	Statetx	-0.0435126	-0.0435126
168	Provider.Namesampson home health	-0.0450674	-0.0450674
169	Provider.Namehealth first home care	-0.0542560	-0.0542560
170	Provider.Nameviva home health care inc	-0.0545455	-0.0545455
171	Provider.Namenu dash era home health agency inc	-0.0582623	-0.0582623
172	Statewa	-0.0590405	-0.0590405
173	Provider.Nameperry county health department	-0.0620527	-0.0620527
174	Cityhudson	-0.0640343	-0.0640343
175	Stateoh	-0.0645984	-0.0645984
176	Provider.Namegirling home health texas by harden healthcare	-0.0668042	-0.0668042
177	Provider.Namefidelity health care inc	-0.0696653	-0.0696653
178	Stateor	-0.0717047	-0.0717047

179	Cityvancouver	-0.0731741	-0.0731741
180	Provider.Namefamily hospice and pallative care	-0.0807074	-0.0807074
181	Provider.Namepinnacle senior care	-0.0828205	-0.0828205
182	Provider.Nametehc llc	-0.0846982	-0.0846982
183	Provider.Nameexceed home health inc	-0.0906206	-0.0906206
184	Provider.Namelakeview home health care	-0.0991580	-0.0991580
185	Provider.Nameyuma district hospital home health care	-0.1118308	-0.1118308
186	Provider.Namejordan health services	-0.1137930	-0.1137930
187	Cityosceola	-0.1490237	-0.1490237
188	Statemn	-0.1602387	-0.1602387
189	Citybrooklyn	-0.1713038	-0.1713038
190	Provider.Namecentral florida quality care services inc	-0.1768048	-0.1768048
191	Provider.Namenurses choice home care	-0.2101921	-0.2101921
192	Provider.Namesouth florida home health care inc	-0.2265337	-0.2265337
193	Provider.Nameottawa co health center hha	-0.2268843	-0.2268843
194	Provider.Nameveritas home care inc	-0.2488064	-0.2488064
195	Provider.Namecareall	-0.2742150	-0.2742150

196	Provider.Nameeverwell health agency llc	-0.3110203	-0.3110203
197	Provider.Nameprudent healthcare agency	-0.4986622	-0.4986622
198	Provider.Namewisdom health care services inc	-0.5033612	-0.5033612
199	Provider.Nameregistered nurses care ltd	-0.5213908	-0.5213908
200	Provider.Namegood samaritan home health care inc	-0.7341695	-0.7341695
201	Provider.Nameangel care home health services inc	-0.8028250	-0.8028250

R Code

This section contains the list of fourteen (14) files created in R

List of R Code Files:

1. A0 - my.summary.func.r
2. A1 - Reading multiple Files - 2013 - Sandra.r
3. A2 - Reading multiple Files - 2014 - Sid.r
4. A3 - Reading multiple Files - 2015 Q1 Q2 - Sid.r
5. A4 - Reading multiple Files - 2015 Q3 - Sid.r
6. B1 - Munging all data - Sandra.r
7. B2 - Munging 9 PCQM For Scoring - Sandra.r

8. C1 - EDA for Munged Data - Sandra.r
9. C1 - EDA Variable Selection - Sid.r
10. C2 - EDA of Process vs Outcomes - Sid.r
11. C3 - EDA Variable Selection - Sandra.r
12. D1 - Binning and Star Rating of HHC providers – Sid.r
13. D2 - EDA for Binning - Sid.r
14. E1 - Cluster Analysis of Providers - Sid.r
15. E2 - Identification of Important Variables with Random Forest - Sandra.r
16. F1 - Linear Regression - Sandra.r
17. F2 - Predictive Modeling - Sid.r
18. F3 - Lasso - Sandra.r