Big Data Analytics Assignment

Healthcare

Table of Contents

[1. Executive Summary 5](#_Toc42033637)

[2. Background: 6](#_Toc42033638)

[3. Analysis and Results 15](#_Toc42033639)

[4. Justification 31](#_Toc42033640)

[5. Discussion of Results 34](#_Toc42033641)

[6. Conclusion 37](#_Toc42033642)

[7. Recommendations: 37](#_Toc42033643)

[8. References 39](#_Toc42033644)

Table of Figures:

[Figure 1: Descriptive Analysis 1: Gender 10](#_Toc42179654)

[Figure 2: Descriptive Analysis 2; Race 10](#_Toc42179655)

[Figure 3: Descriptive Analysis 3; Patient's Age 11](#_Toc42179656)

[Figure 4: Descriptive Analysis 4; Diagnosis Group 11](#_Toc42179657)

[Figure 5: Descriptive Analysis 5; Diagnosis Subcategory 12](#_Toc42179658)

[Figure 6: Task 2: Average number of ICU days spent with Gender and Diagnosis Group 14](#_Toc42179659)

[Figure 7: Task 3; Most and Least popular diagnosis group for each region 15](#_Toc42179660)

[Figure 8: Task 4; Disease for each diagnosis group, most and least common 15](#_Toc42179661)

[Figure 9: Task 5; Top five departments in terms of No. of patients in each department 16](#_Toc42179662)

[Figure 10: Task 6; Top 3 regions in terms of number of female patients 16](#_Toc42179663)

[Figure 11: Task 7; Top 5 places each patient was discharged to 17](#_Toc42179664)

[Figure 12: Task 8; Top 3 regions for black patients 17](#_Toc42179665)

[Figure 13: Task 9; Top 5 hospitals in terms of No. of visits (average) for asthma patients 18](#_Toc42179666)

[Figure 14: Task 10, a; Trend of admission by month for female patients 18](#_Toc42179667)

[Figure 15: Task 10, b; Trend of admission by month and male patients 19](#_Toc42179668)

[Figure 16: Task 11, a; No. of days spent in hospital calculation 19](#_Toc42179669)

[Figure 17: Task 11, b); No. of days spent in hospital data review 20](#_Toc42179670)

[Figure 18: Task 11, c); No. of days spent in hospital against top 3 regions 20](#_Toc42179671)

[Figure 19: Task 12; Top 10 cities in terms of number of patients 21](#_Toc42179672)

[Figure 20: Task 13; Trend of admissions from October 2011 to June 2012, gender wise 21](#_Toc42179673)

[Figure 21: Task 14; Most and least popular month for asthma patients, 22](#_Toc42179674)

[Figure 22: Task 15; Trend of patient admissions diagnosed with 'CHF' 22](#_Toc42179675)

[Figure 23: Task 16; Trend of patient admissions for each diagnosis group 23](#_Toc42179676)

[Figure 24: Task 17; Number of operations varied across different months, top 5 departments 23](#_Toc42179677)

[Figure 25: Task 18; decision tree analysis 24](#_Toc42179678)

[Figure 26: Task 19; Geo map of hospitals 24](#_Toc42179679)

[Figure 27: Task 20; Cluster analysis of patient data 25](#_Toc42179680)

[Figure 28: Addition visualization 1; Mean ICU days for each race 25](#_Toc42179681)

[Figure 29: Additional Visualization 2; Mean patient age for each disease 25](#_Toc42179682)

[Figure 30: Additional Visualization 3; Mean patient age for each race 26](#_Toc42179683)

[Figure 31: Additional Visualization 4; disease severity levels against race 26](#_Toc42179684)

[Figure 32: Additional Visualization 5; Average number of days spent in ICU against disease subcategory 27](#_Toc42179685)

[Figure 33: Additional Visualization 6; Total number of cases of each disease 27](#_Toc42179686)

# Executive Summary

The study herewith is conducted to analyse and find the usability and applicability of big data analytics and business intelligence in the healthcare industry. The dataset used contains data obtained from hospitals situated in various cities and counties in the US, which is further analysed to draw useful and meaningful insights. These insights can help businesses and economies make effective and quicker decisions. The significance behind the study falls in the essentiality of quick and better decision making during a healthcare crisis, and ways to tackle it using analysed data. Some of the main findings of the research are as follows:

1. A patient’s age is the biggest predictor of the possibility of an individual having heart disease.
2. February, January and May showed the greatest no. of patients in hospitals
3. There appears to be a steep rise in the number of people admitted who were diagnosed with Chronic Heart Failure during December
4. July and August had the least amount of patients and least number of operations
5. Hospital number 13 as well as hospital number 35 had the greatest amount of visitors for asthma patients
6. Regions 3, 8 and 11 had the highest quantity of patients
7. Delray Beach and Miami had the greatest no. of people admitted during the time period specified initially.
8. An individual is likely to spend more time in ICU if one is white, male and diagnosed with ‘AMI’ diagnosis group.
9. Regions 9, 4 and 7 had patients who spent the highest number of days on average, in hospitals.
10. Bronchopneumonia had highest average patient age of 84, while Chronic hosp 46 heart had the lowest average patient age of 70.
11. . Patients diagnosed with Acute Myocardial Infarct, on average, spent the most number of days in ICU than all other patients.

The above results give key insights to policy makers to assist them in making viable decisions about the future of the sector.

# Background:

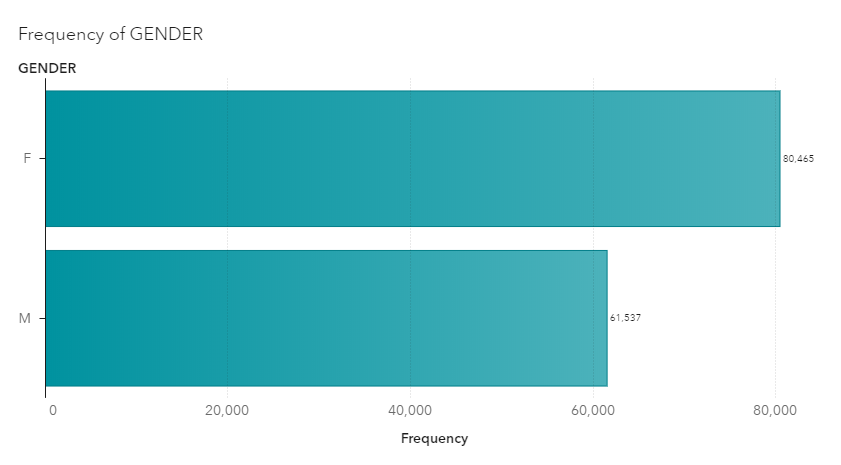
The dataset used for this particular project involves data on hospitals across the US from 2011 to June 2012.

A data dictionary is first created out of the dataset, which can be seen in the following manner:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Variable | Data Type | Data Format | Decription |
| 1. | Encounter Key | Numeric | NNNNNNNNN | 142,000 distinct values for each patient |
| 2. | Patient Number | Numeric | NNNNNNNNNN | 142,000 distinct numbers assigned to each patient |
| 3. | Doctor | Numeric | NNNNNN | 142,000 distinct values assigned to each doctor for every patient |
| 4. | Admit Date | Date | NN/Month/Year | Date of admission for each patient |
| 5. | Discharge Date | Date | NN/Month/Year | Date of discharge for each patient |
| 6. | ICU Days | Numeric | N | Number of days spent in ICU |
| 7. | Department | Categorical | Text | Department concerned for treatment of each patient |
| 8. | Discharged to | Categorical | Text | Place discharged to for every patient |
| 9. | Standard Orders Used | Categorical | Text – Y/N | Whether standard orders were used for the treatment |
| 10 | Num Chronic Cond | Numeric | N | Total number of chronic conditions |
| 11 | Disch Nurse ID | Numeric | NNNNNN | 142,000 distinct numbers assigned to each nurse |
| 12 | Order Set used | Categorical | 0/1 | Which set of orders are used |
| 13 | Order total charges | Numeric | NNNNN | Distinct values |
| 14 | Gender | Categorical | M/F | Whether patient is male or female |
| 15 | ZIP | Numeric | NNNNNN | Distinct zip code values |
| 16 | Statecode | Numeric | NNNNNN | Distinct state code values |
| 17 | City | Categorical | Text | City of orgin for each patient |
| 18 | County name | Categorical | Text | County of origin for each patient |
| 19 | X | Numeric | NN | X coordinates |
| 20 | Y | Numeric | NN | Y coordinates |
| 21 | Region | Categorical | Text | Patient region origin |
| 22 | Race CD | Categorical | Text | Race of Patient |
| 23 | Patient Age | Numeric | NN | Age of each patient |
| 24. | Diagnosis Group | Categorical | Text | Diagnosis group of each patient |
| 25. | ICD9 Target | Categorical | 1/0 | Target variable for whether an individual has heart disease |
| 26. | MS DRG code | Numeric | NNN | DRG code |
| 27. | MS DRG DESC | Categorical | Text | Cause of death |
| 28 | Readmitted | Categorical | 0/1 | Whether the patient was readmitted or not |
| 29 | DRG APR CODE | Numeric | NNN | Distinct codes |
| 30 | DRG APR DESC | Categorical | Text | DRG cause |
| 31 | DRG APR Severity | Categorical | 1 - 4 | DRG level of severity |
| 32 | Diagnosis subcat code | Numeric | NNN | Distinct code for each diagnosis subcategory |
| 33 | Diagnosis subcat desc | Categorical | Text | Diagnosed subcategory |
| 34 | Diagnosis ICD Code | Numeric | NNN | Distinct codes for diagnosis |
| 35 | Diagnosis Long Desc | Categorical | Text | Diagnosed category |
| 35 | Procedure subcat Code | Numeric | NN | Codes for procedures used |
| 36 | Procedure subcat desc | Categorical | Text | Procedure subcategory used for treatment |
| 37 | Procedure ICD code | Numeric | NN | Procedure ICD distinct codes |
| 38 | Procedure Long desc | Categorical | Text | Procedure category used |
| 39 | DX Code | Numeric | NNNNN | Codes assigned for disease |
| 40 | DX Group | Categorical | Text | Disease group |
| 41 | Operation Count | Numeric | NN | Total number of operations |
| 42 | Hospital | Categorical | Text | Hospital number where the patient was admitted |
| 43 | Admit Mth | Categorical | 1 - 12 | Month of admission |
| 44 | Admit Los | Numeric | NN |  |
| 45 | Readmit Mth | Categorical | Text | Month of readmission |
| 46 | No. of visits | Numeric | N | Number of visits of each patient |
| 47 | Readmit date | Date | dd/mm/yyyy | Date of readmission |
| 48 | Readmit discharge date | Date | dd/mm/yyyy | Date of readmission |
| 49 | Readmit Los | Numeric | N |  |
| 50 |  |  |  |  |

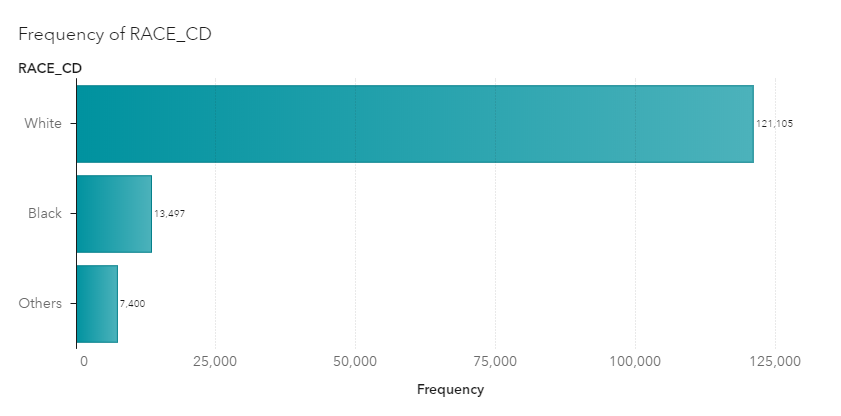
Our analysis first focuses on information about the patients. Each and every patient is assigned a unique no. as soon as they are admitted in a hospital, which is given in the file as 142 thousand distinct values.

Figure : Descriptive Analysis 1: Gender



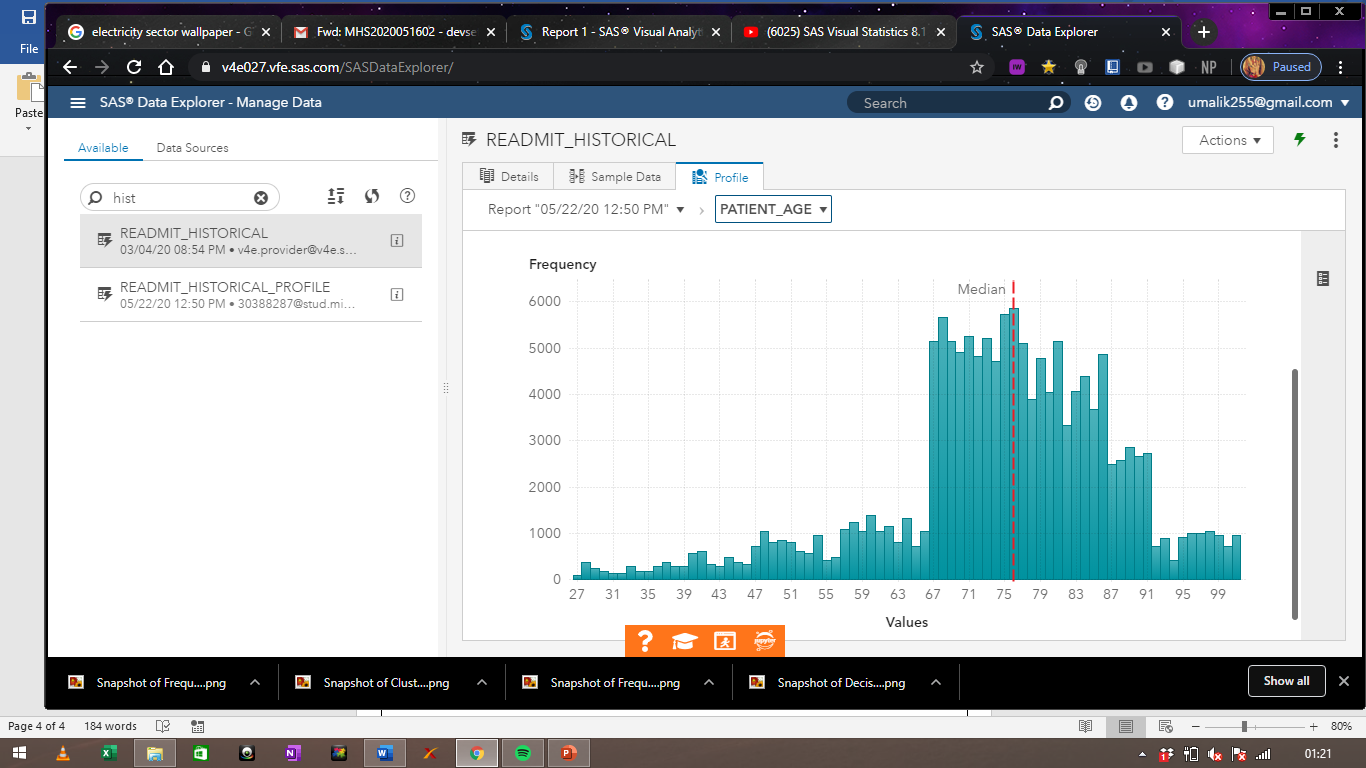
While the total No. of male patients was about 62,000, female patients were far more at 80,000. We then look at distribution according to race:

Figure : Descriptive Analysis2; Race



As shown from the above plot, out of the total number of patients admitted, most were white. The average age of all patients admitted during the time period was 74, while the mode and median were 76 years respectively.

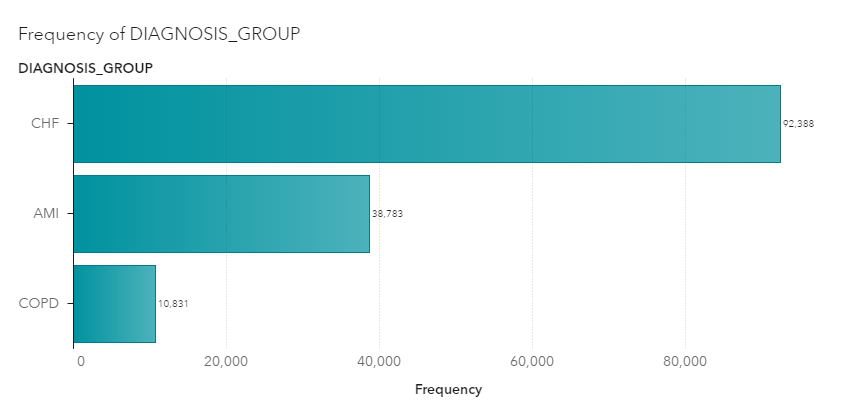
Figure : Descriptive Analysis 3; Patient's Age



As shown from the above plot, most of the patients admitted happened to be between the age group of 67 to 91.

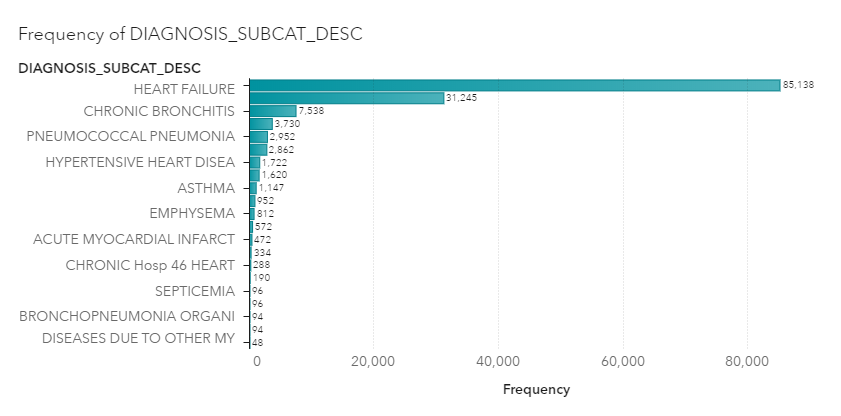
The dataset is divided into 3 distinct diagnosis groups (CHF, COPD and AMI) whose frequencies can be shown as follows:

Figure : Descriptive Analysis 4; Diagnosis Group



The type of disease and their respective frequencies can be seen in the following manner:

Figure : Descriptive Analysis 5; Diagnosis Subcategory



In the above visualization, it is evident that heart failure was the most common cause of admission in all hospitals during the time period specified. Additionally, information on the location of hospitals, counties and regions is also given through X and Y coordinates as well as zip codes.

With the explosion in the transfer and storage of dataset volume, analysing and interpreting such data is becoming more and more important. Big data analytics and business intelligence tools help organisations, firms, businesses, economies analyse humungous amounts of data to draw quick and informative results for effective decision making. It is made up of softwares, techniques and services which transforms raw data into useful insights that lead to efficient and effective decision making (Pratt & Fruhlinger, 2019).

Gone are the days when one needed specialized and skilful analysts to analyse data.. BI and big data analytics tools make it extremely easy to do the same for any business without any prior skill or knowledge. This is revolutionary since data is now the most valuable resource in the world. These tools are easy to use, flexible and effective. Business intelligence is a wider term and contains numerous processes and techniques such as Data Mining, Reporting, Performance metrics, Data warehousing and so on (tableau, 2019).

The major advantages of adopting big data analytics techniques are:

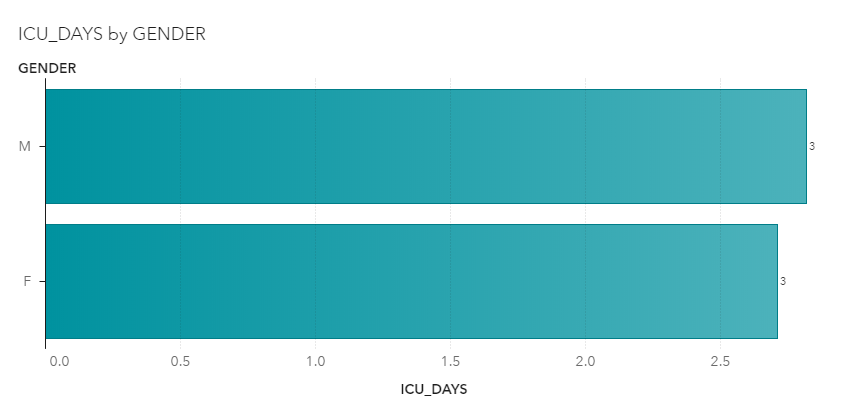
1. Reduction in Cost: using these techniques significantly diminishes the time it takes to analyse and visualize data. These techniques ensure that costs of business are saved by using techniques like (Insights, 2018).
2. Reduction in human errors: Using techniques like machine learning, artificial intelligence, deep learning, etc, errors made by humans are minimized (Kumar, 2018).
3. Dynamic insights and useful information: Data analytics techniques give useful meaningful results which ensure proper understanding of business operations and customers.
4. Interesting visualizations: These techniques give interesting and meaningful visualizations which assist in the explanation of data (James, 2019).
5. New products: Big data analytics helps in redeveloping and selling of new products and also gaining insights about the behaviour of customers for existing products (Simplelearn, 2019).
6. Better quality: Using techniques like data quality management, businesses ensure high quality of decision (Learntek, 2017).

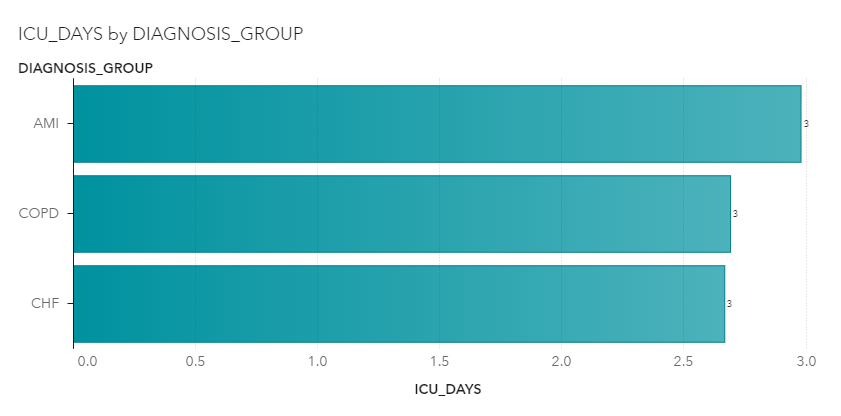
These techniques help organisation make better and faster decisions, which save costs, time and effort, and thus add on to the revenues of any business (King, 2018).

# Analysis and Results

The data obtained for the visualization contains information obtained from multiple hospitals all around the USA from September 2011 to June 2012. After completing the visualization, we find the results given below:

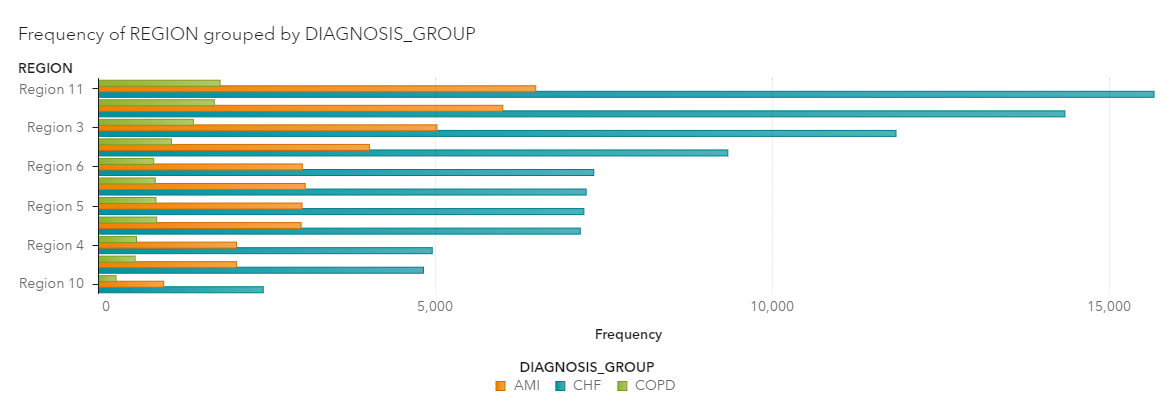
Figure : Task 2: Average number of ICU days spent with Gender and Diagnosis Group





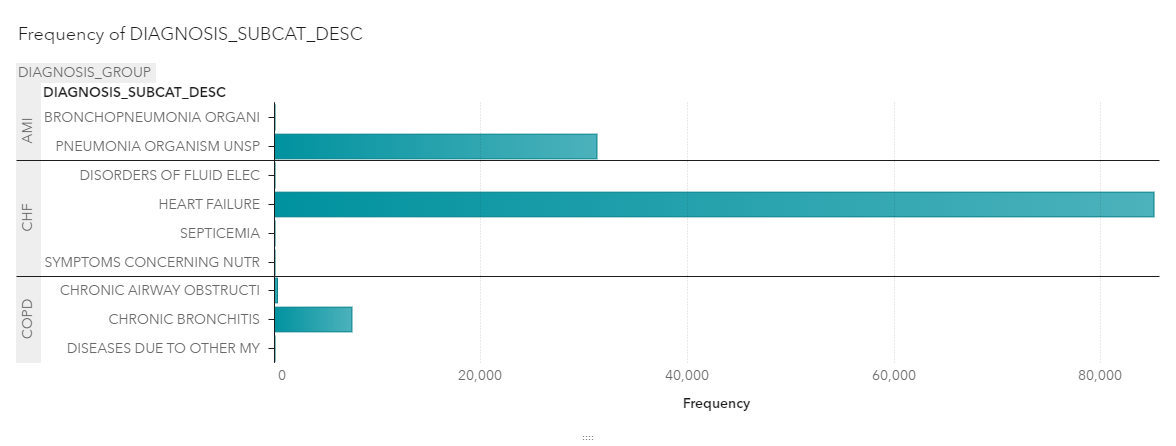
The researchers first plot the mean no. of ICU days spent with gender and diagnosis group of each patient. This shows that the average number of days spent in Icu by male patients was more than that of females. In a similar manner, the mean days spent in ICU for patients in the ‘AMI’ diagnosis group were the highest, as compared to the diagnosis groups of ‘COPD’ and ‘CHF’.The analysis now shifts towards differences among regions:

Figure : Task 3; Most and Least popular diagnosis group for each region



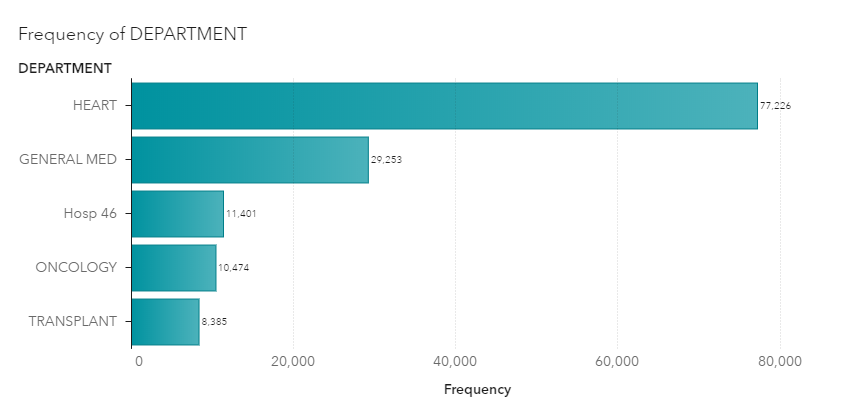
Taking a look at the above chart, we witness that the diagnosis group of COPD was the least common in all regions, whereas the diagnosis group of ‘CHF’ was the most common. Thus, during the time period of September 2011 to June 2012, it is evident that most of the patients belonged to the ‘CHF’ diagnosis group.

Figure : Task 4; Disease for each diagnosis group, most and least common



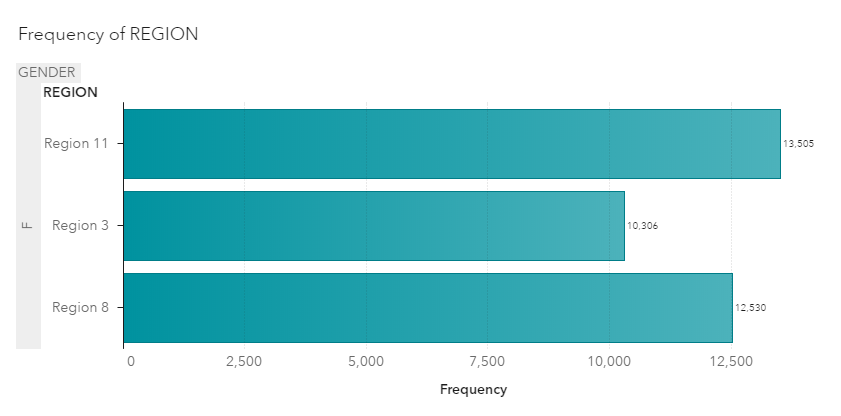
The above plot evidently shows the most and least common disease in each diagnosis group. We find that Pneumonia Organism UNSP is the most common form of disease in the ‘AMI’ diagnosis group with 31,245 patients, whereas Bronchopneumonia is the least common disease with only 94 patients in the same. Similarly, Heart failure is the most common disease in the diagnosis group of ‘CHF’ with 85,138 patients, and disorders of fluid electricity is the least common with only 94 patients. Finally, Chronic bronchitis is the most common disease in ‘COPD’.

Figure : Task 5; Top five departments in terms of No. of patients in each department



The next analysis entails useful information on the top 5 departments in terms of No. of patients admitted, which can be shown from the above plot. It is evident that the department concerning diseases related to heart had the most No. of patients.

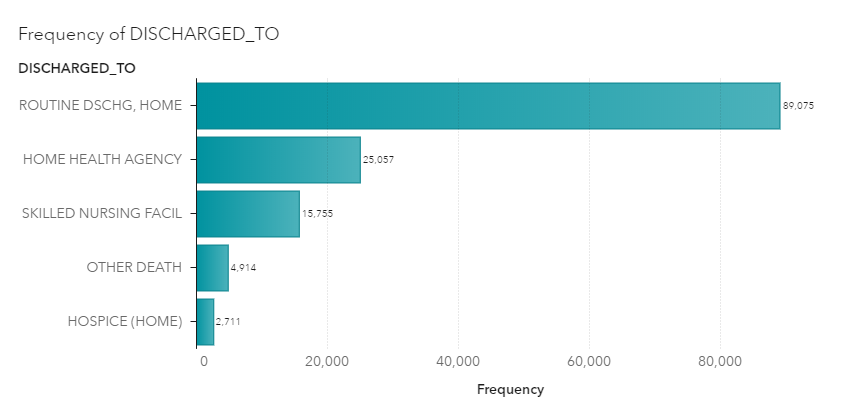
Figure : Task 6; Top 3 regions in terms of number of female patients



The above plot evidently shows that region 11, region 8, region 3 had the highest No. of female patients.

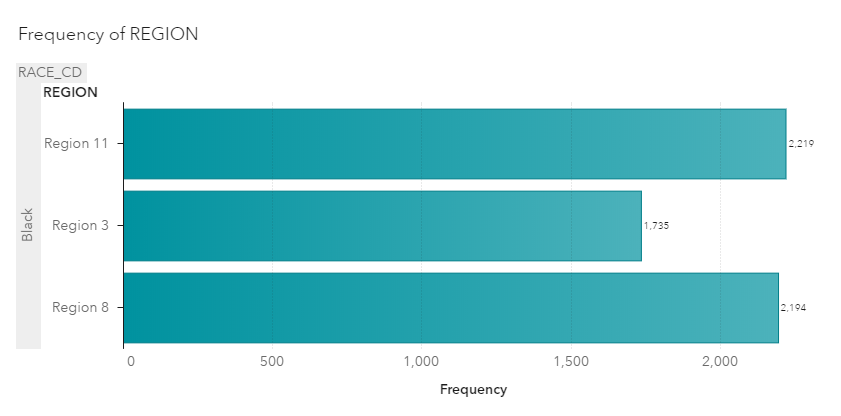
We now focus our attention towards the places that each patient was discharged to, which can be shown as follows:

Figure : Task 7; Top 5 places each patient was discharged to



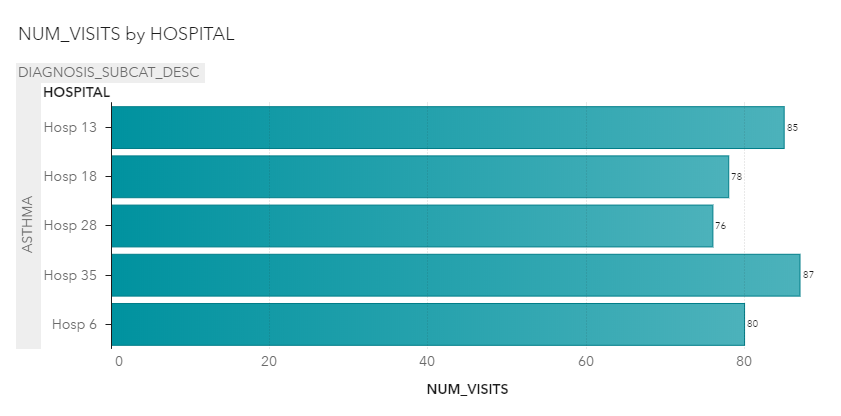
it shows the top 5 places patients were discharged to. It is evident that Routine home discharge was the most common, followed by home health agency.

Figure : Task 8; Top 3 regions for black patients



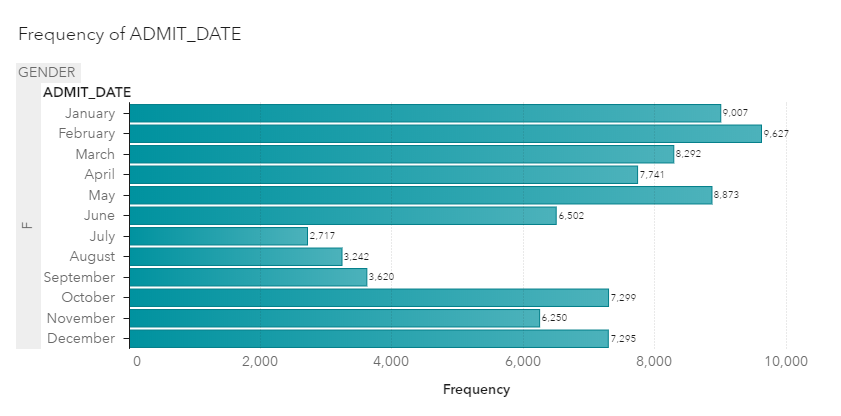
The plot above distinctly shows the top 3 regions for patients which were black by race. These 3 regions are region 11, region 3 and region 8.

Figure : Task 9; Top 5 hospitals in terms of No. of visits (average) for asthma patients



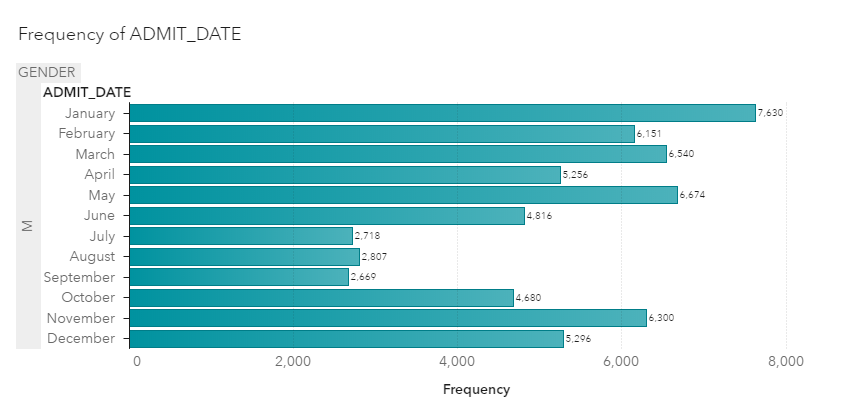
The data entails values on the no. of visits for each patient in every hospital. From the plot above, it is evident that the top 5 hospital in terms of the No. of visits for asthma patients, which were hospital 12, hospital 18, hospital 28, hospital 35 and hospital 6.

Figure : Task 10, a; Trend of admission by month for female patients



The above plot shows the frequency of admit months for female patients only. It is evident that January February and May had the highest No. of male patient admits, whereas July and August were the lowest.

Figure : Task 10, b; Trend of admission by month and male patients



For male patients, January, may and march entailed the highest No. of patients, whereas July August and September had the lowest.

The next task focuses on finding the average number of days spent in hospital, which is calculated by using the “treat as” function in SAS to treat dates as numbers, and returns the difference between date dispatched and date admitted in numeric form.

Figure : Task 11, a; No. of days spent in hospital calculation

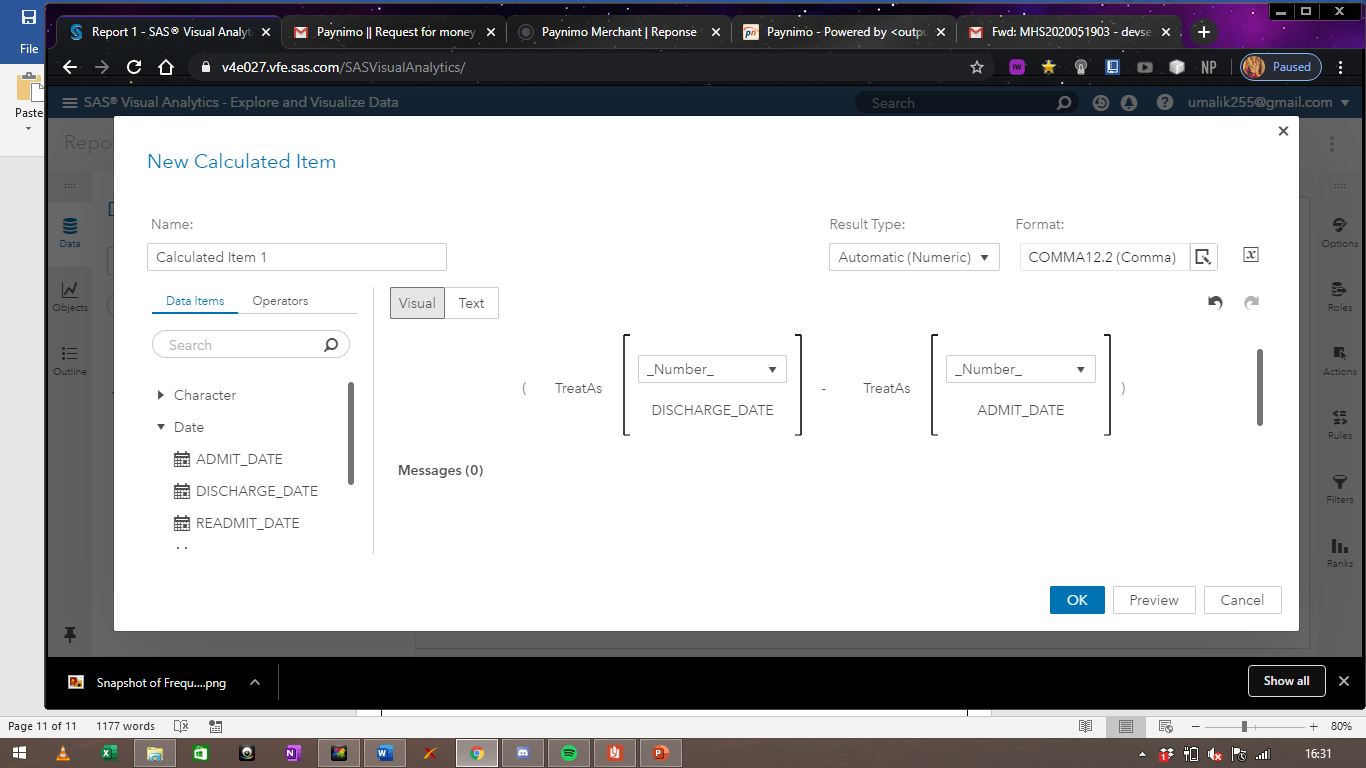
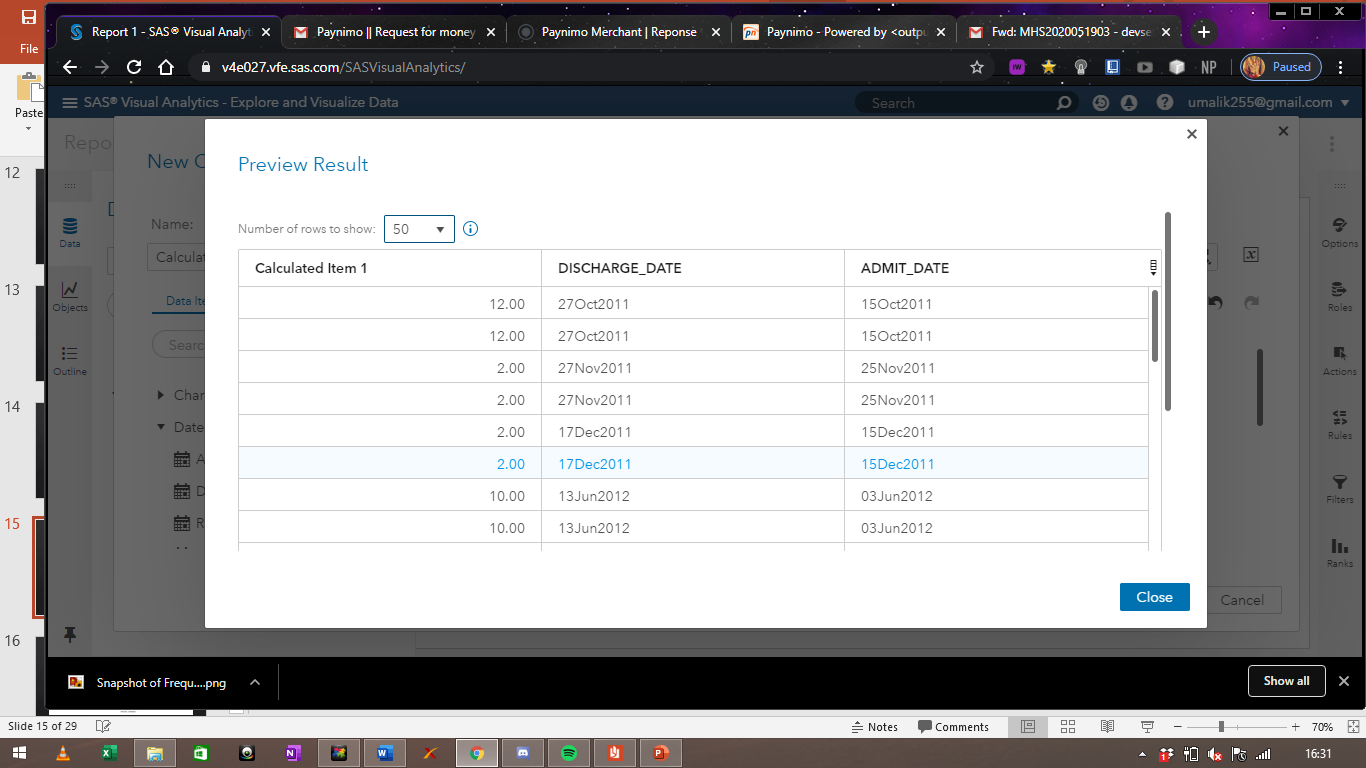
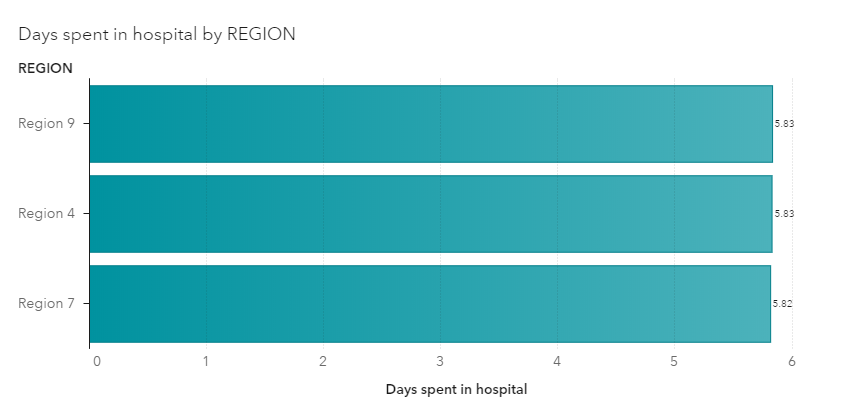


Figure : Task 11, b); No. of days spent in hospital data review



The first column gives us the total No. of days spent in the hospital. We then find the average and plot it against the desired variable.

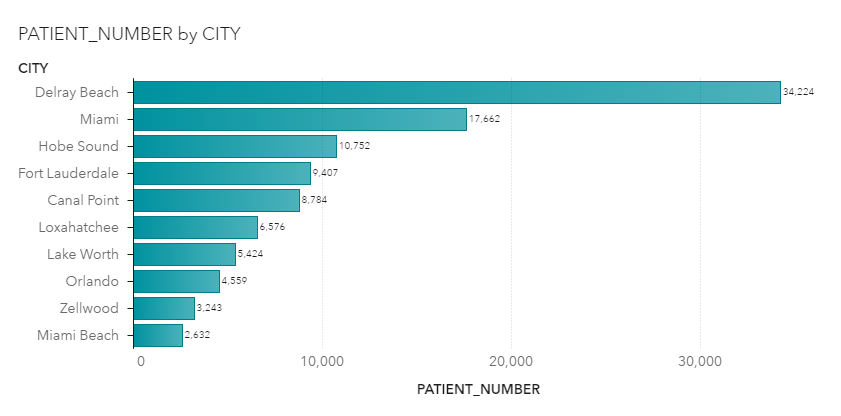
Figure : Task 11, c); No. of days spent in hospital against top 3 regions



The above plot gives us the top 3 regions with respect to the mean No. of days spent in the hospital.

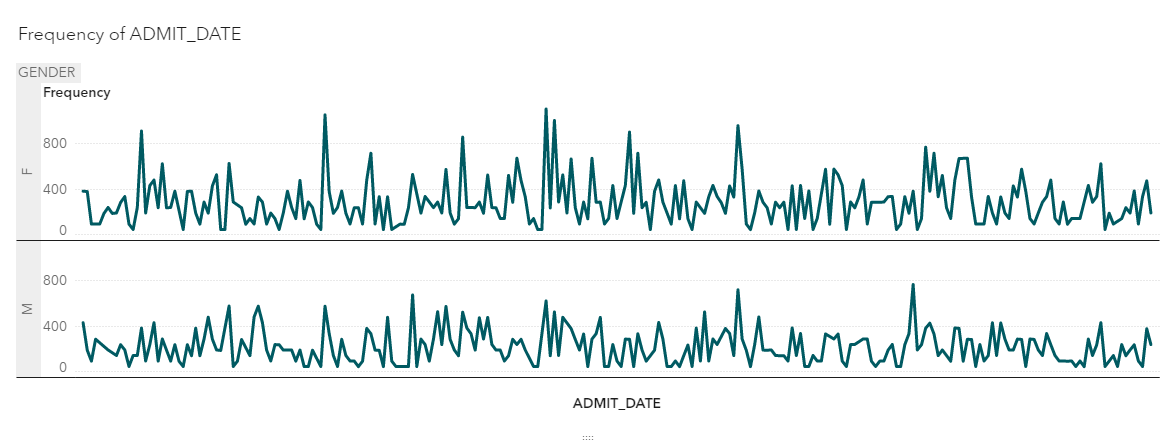
To check for differences among cities, we plot the following plot:

Figure : Task 12; Top 10 cities in terms of number of patients



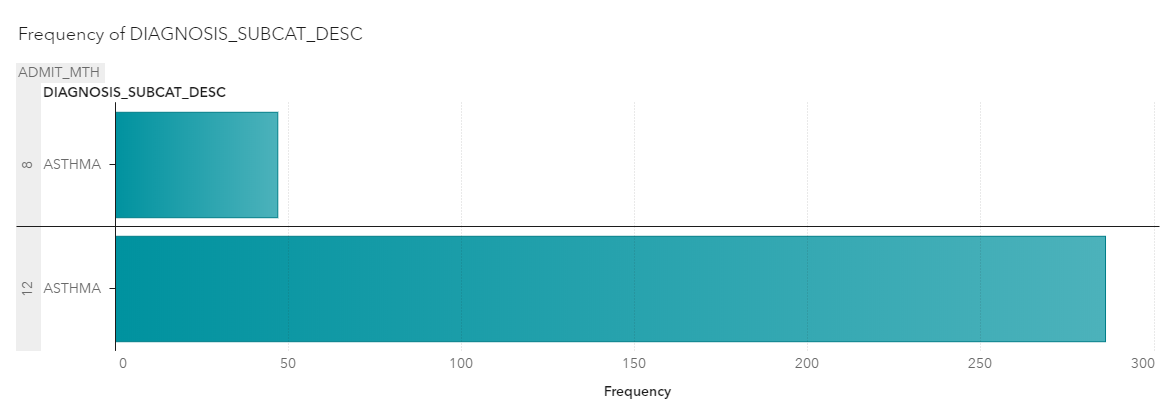
It is evident that most of the patients were from Delray beach, followed by Miami. This plot gives us the top 10 cities in terms of No. of patients admitted.

Figure : Task 13; Trend of admissions from October 2011 to June 2012, gender wise



We now take a look at the trend of admission dates for both genders, which can be seen from the above plot between October 2011 and June 2012.

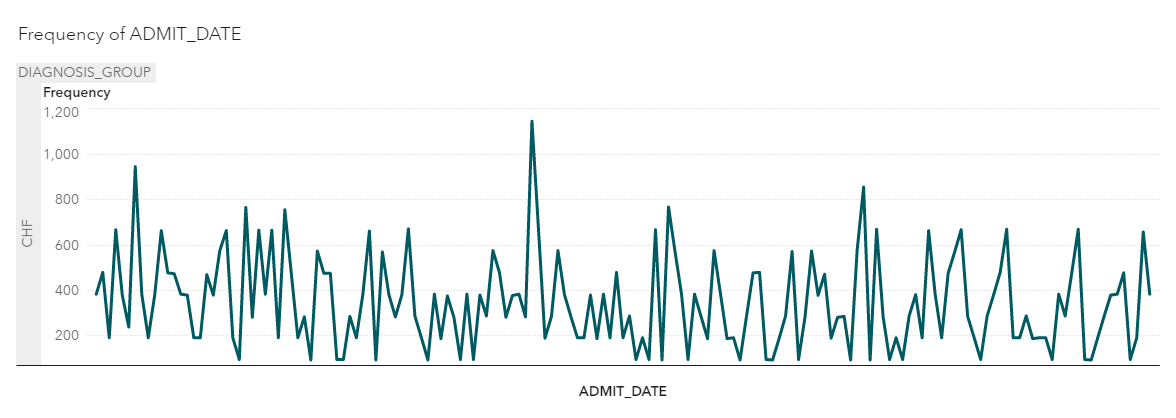
Figure : Task 14; Most and least popular month for asthma patients,



It is evident that December is the most common month for No. of visits for asthma patients, while August is the least common month.

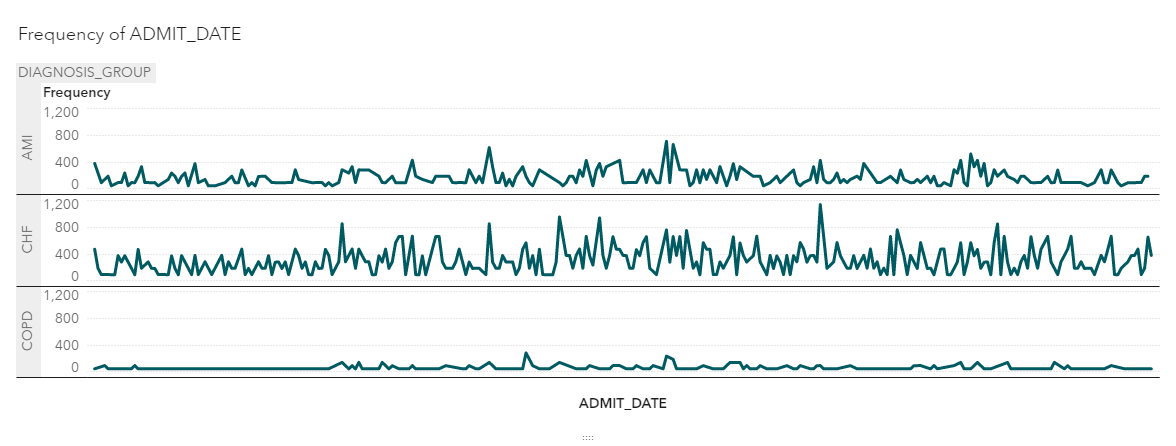
The researchers now emphasize on the trend of patient numbers diagnosed with the group of ‘CHF’ or chronic heart failure. It is evident such trend as follows:

Figure : Task 15; Trend of patient admissions diagnosed with 'CHF'



Trend analysis is useful for finding seasonality in the dataset.

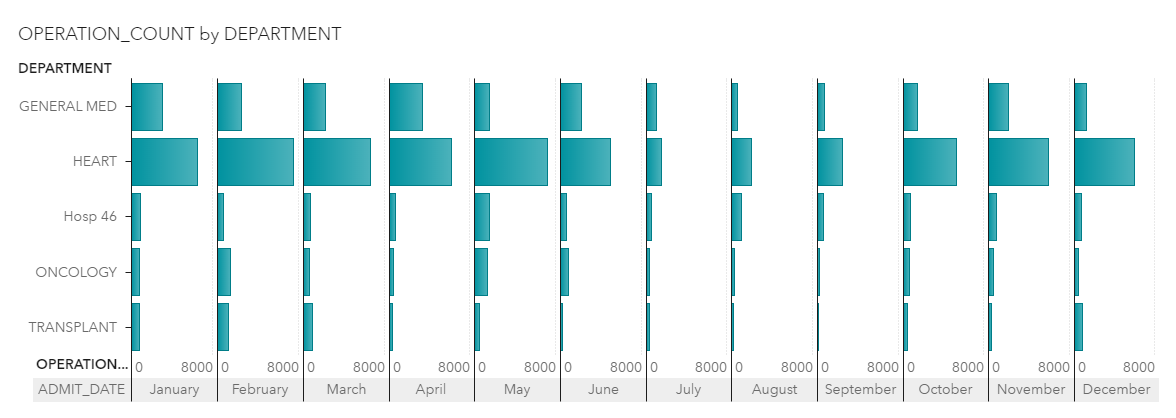
Figure : Task 16; Trend of patient admissions for each diagnosis group



From the above plot, it is evident that the trend of diagnosis groups of AMI, CHF and COPD respectively. It is evident that CHF has the highest volatility, while the trend of COPD is relatively stagnant and stable.

We now have a look at the major 5 departments in terms of operation count, which can be shown in the following manner:

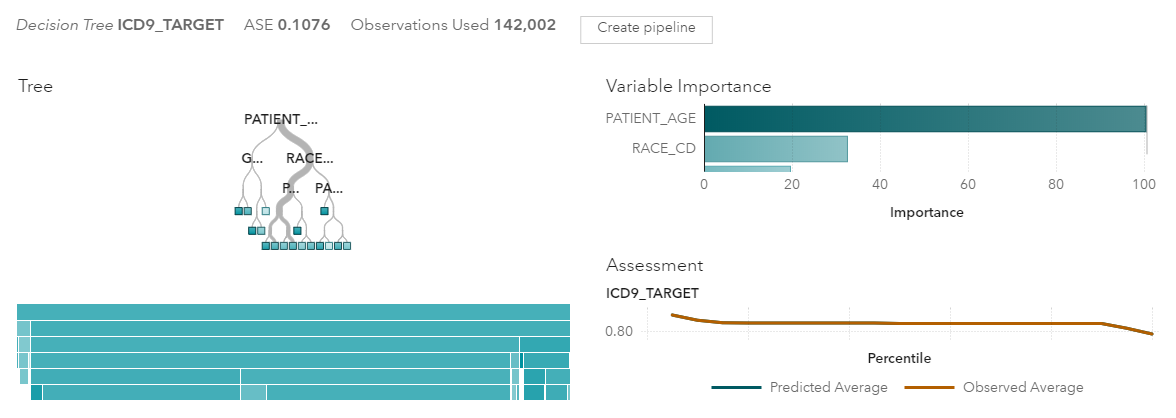
Figure : Task 17; Number of operations varied across different months, top 5 departments



It is evident that the department which deals with heart diseases entails the highest No. of operations in all hospitals, followed by General med. It is also evident that February, January and May had the highest No. of operation counts, whereas July had the lowest.

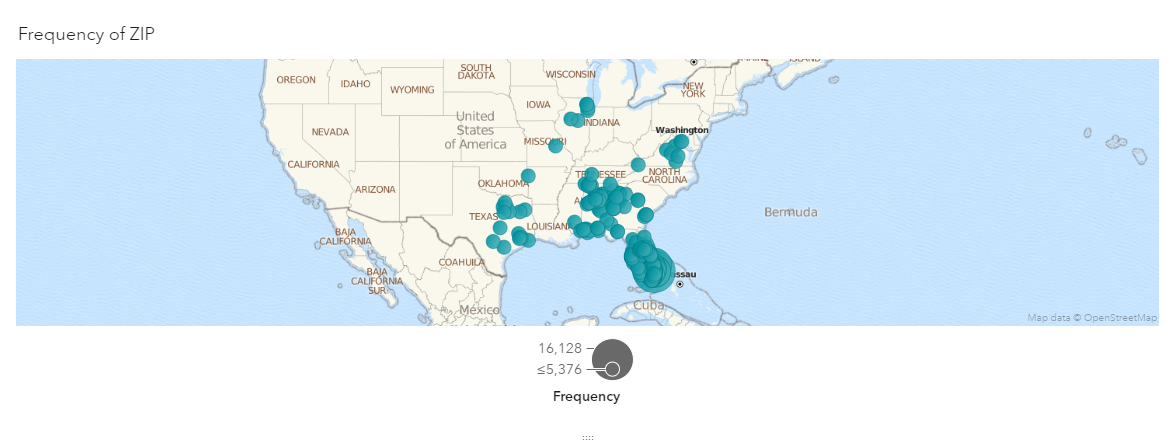
In the next task, we create a decision tree model to predict whether an individual would have chronic heart disease or not. We us the variable ‘ICD9 Target’ as the target variable and Gender, race and Patient’s age as the predictors.

Figure : Task 18; decision tree analysis



We find that a patient’s age is the most important predictor of whether an individual will have heart disease or not. Race is also a significant predictor, but not as much as Patient’s age.

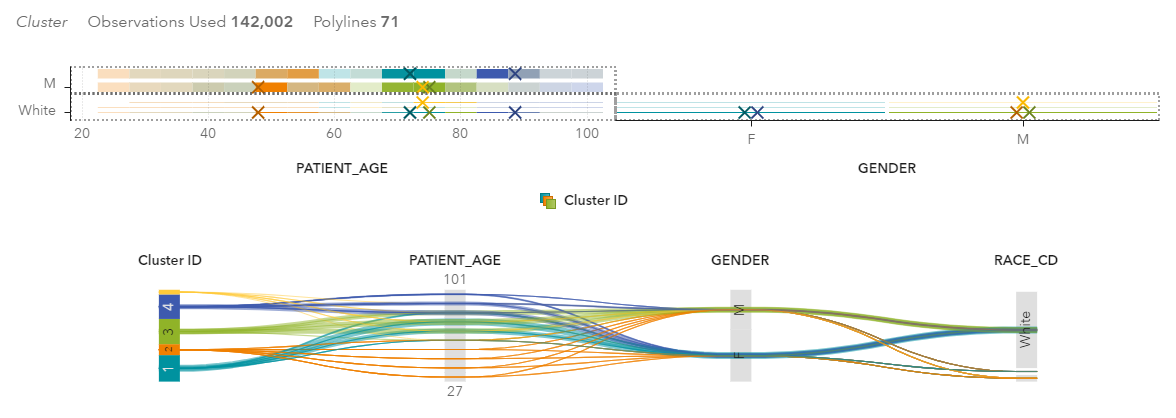
Figure : Task 19; Geo map of hospitals



The above GEO map plots the hospitals across united states which are used for this study.

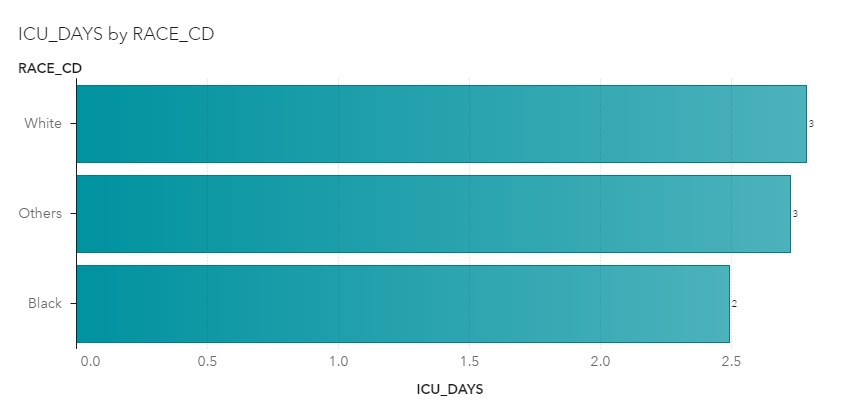
Finally, we create a cluster model on information about patients like gender, race and age, which can be shown as follows:

Figure : Task 20; Cluster analysis of patient data



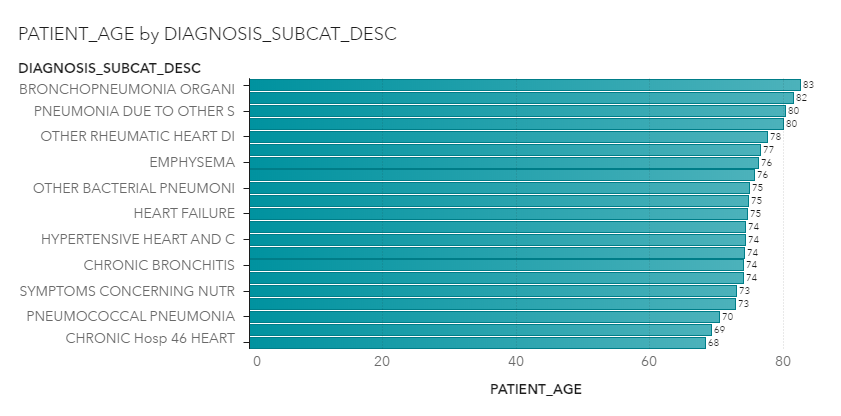
In addition to these tasks, we have other additional visualizations, which can be seen as follows:

Figure : Addition visualization 1; Mean ICU days for each race



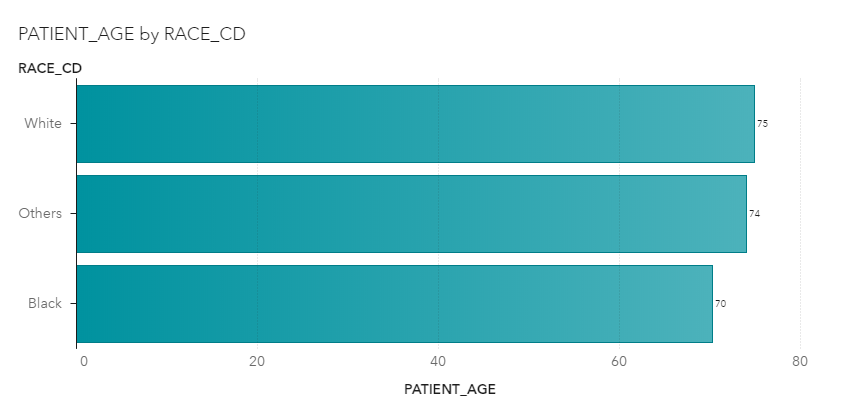
The above plot shows the mean no. of days spent in ICU for each race. It is evident that patients who were white had the highest No. of ICU days, while the least was of black patients.

Figure : Additional Visualization 2; Mean patient age for each disease



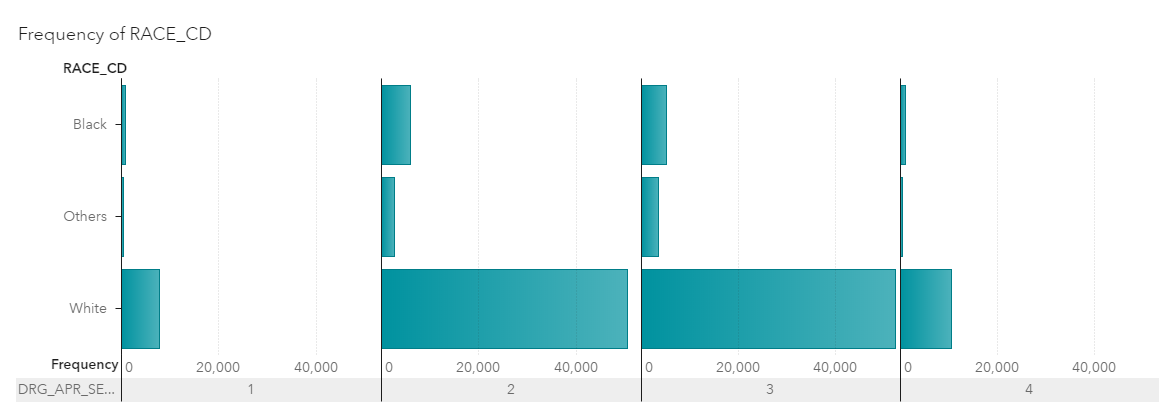
We look at the mean age of patient under each disease. We find the Bronchopneumonia had the highest mean patient age of 83, while chronic heart failure had the lowest (68).

Figure : Additional Visualization 3; Mean patient age for each race



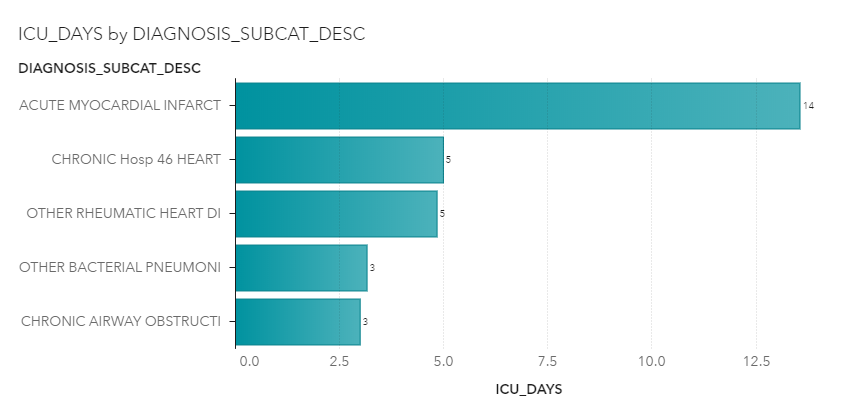
Next, we look at the mean patient age with each race, which tends to be the highest among white patients and lowest among black.

Figure : Additional Visualization 4; disease severity levels against race



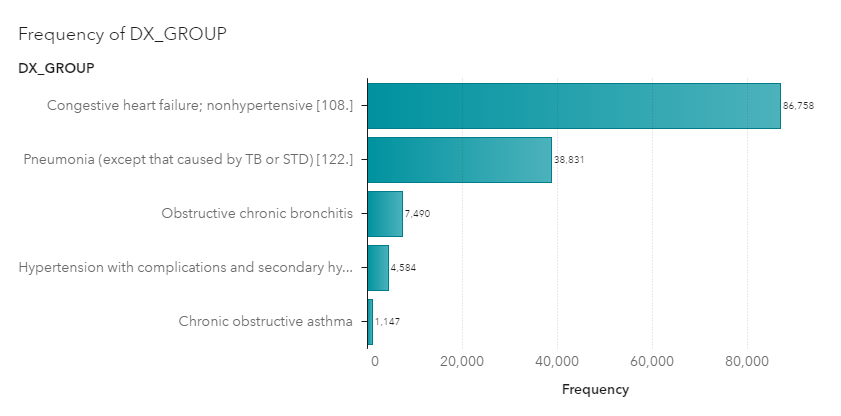
The above graph plots the severity of DGR with race. It shows that white patients tend to have the highest level of severity.

Figure : Additional Visualization 5; Average number of days spent in ICU against disease subcategory



The above graph plots the average number of ICU days for each diagnosed subcategory (top 5). It shows that patients diagnosed with Acute Myocardial Infarct, on average, spent the most number of days in ICU than all other patients

Figure : Additional Visualization 6; Total number of cases of each disease



Finally, The above graph plots the frequency of top 5 DX groups. It shows that congestive heart failure is the most common group of all.

# Justification

The first analysis entails the average number of days spent in ICU by each patient against gender and diagnosis group. For this, a bar graph is used, wherein gender and diagnosis groups are categories and average number of days spent in ICU is the measure. Similarly, the next task involves finding the most and least common diagnosis group for each region. A bar graph is used again, with regions as lattice rows and diagnosis groups as categories. Diagnosis group is plotted as lattice rows and diagnosis subcategory is plotted as frequencies in the bar graph for the next task to find the most and least common disease in each diagnosis group. Total number of patients is plotted against the departments and the top 5 of them are selected in a column chart. For the next task, gender as taken as lattice row and regions are taken as frequencies. The top 3 are then selected to determine the top three regions for female patients.

Frequency analysis of the variable ‘discharged to’ is employed to determine the top 5 places a patient was discharged to. The bar graph plots the frequency of each place, and the top 5 are then filtered out. For the next task, race is plotted along with regions. Then ‘black’ is filtered, and the top 5 regions in terms of frequencies are filtered n the bar graph. The next task requires finding the top 5 hospitals in terms of asthma patients’ number of visits. First, the diagnosis subcategory is plotted as lattice rows, out of which asthma is filtered out. Then hospitals are plotted against number of visits per patients and the top 5 are selected to get the final bar graph.

Admission month is plotted against gender to get the most and least popular months for both male and female patients on the bar graph. For the next task, first a calculated measure is developed which finds the difference between discharge date and admission date in numeric terms. This calculated measure is nothing but number of days spent in the hospital. This measure is then converted into average aggregation and plotted against regions to find the top 3 regions with highest number of average days spent in hospital by patients in the bar graph. Frequency of each city in terms of bar graph is analysed to find the top 10 cities in terms of number of patients.

To determine the trend of admission, a time series analysis is run taking admission date as the date variable. Then data from only October 2011 to June 2012 is filtered out to get the final time series chart. A bar graph is used to find the frequencies of each month, and then the most and least popular months are filtered out. To find the trend of patient numbers between January 2012 to June 2012, time series analysis is run taking only data for the time period specified. Then diagnosis group is added as a category, and CHF is filtered out to get the final trend.

To determine the trend of all three diagnosis groups, time series analysis is run using admission date as the date variable and diagnosis group as the grouping variable, as lattice rows. For the next analysis, first number of operations are plotted against departments. After selecting the top 5 of them, admission date in terms of month is added as lattice columns to find the variation of operations for each month.

Decision tree is formed taking ICD9 target as the target variable and patient’s age, gender and race as predictors. A geomap of hospitals is formed taking the x and y coordinates provided. Finally, cluster analysis is conducted using patient’s information like age, gender and race.

Six additional visualizations were also plotted. The first one plots the average number of ICU days spent by each patient against the patient’s race, to find the differences in the same, using a bar graph. Second, to find the average age for all patients diagnosed with each disease, a bar graph is plotted using average patient age as a measure and diagnosis subcategory as categorical variable. Next, we plot a bar graph taking average patient age against the race of the patient to find a significant difference among the groups. The severity level of disease is plotted as lattice columns while race is plotted as rows to find the variances of each race in each severity level. Then, for every disease, the average number of days spent in ICU is plotted on the bar graph. Finally, to find the 5 most popular diseases, a bar graph is used.

# Discussion of Results

After detailed investigation and visualization of the health care values given or obtained from hospitals throughout the US, many meaningful insights are drawn which can assist in quick and better decision making. We first found the variances in mean No. of days spent in ICU for each gender, by plotting average number of days spent in ICU against Gender. It is evident that male patients spend more time in ICU, on average than female patients. When diagnosis group is plotted against average number of ICU days, we find that patients diagnosed with ‘AMI’ tend to spend the most number of days in ICU than patients diagnosed with other diagnosis groups.

The average number of days spent in ICU was also plotted against race in a bar chart, to find that white individuals spent more time than other races. Thus, an individual is likely to spend more time in ICU if one is white, male and diagnosed with ‘AMI’ diagnosis group.

Plotting diagnosis groups against regions in a grouped bar chart, we find that ‘Chronic Heart Failure’ was the most common diagnosis group in all regions, whereas ‘AMI’ was the least common diagnosis group. COPD lied in the middle range, not as prevalent as CHF.

The next visualization focused on the most and least common diseases in each diagnosis group. It shows that Pneumonia is the most common disease in the diagnosis group of ‘AMI’, whereas heart failure is the most common disease in ‘Chronic heart failure’, and Chronic Bronchitis is the most common disease in ‘COPD’.

Similarly, Bronchopneumonia is the least common disease in the diagnosis group of ‘AMI’, Septicaemia is the least common disease in ‘chronic heart failure’ and diseases due to other factors is the least common disease in ‘COPD’. Chronic heart failure was the most common diagnosis group across all regions, whereas COPD was the least common.

The top 5 departments which had the most No. of patients in the time period were Heart, general med, oncology, transplant and Hosp 46. After plotting regions against gender in the bar graph, we find that regions 3, 11 and 8 had the greatest No. of female patients to be admitted in hospitals.

The areas or Region numbers were also plotted against the race of each patient, to find the top 5 regions for African American patients. We find that region numbers 3, 8 and 11 were again the ones with the highest number of African American patients. Thus, these regions seem to have large numbers of African American as well as female patients admitted in hospitals.

Looking at the mean No. of hospital visits for asthma patients, we see that Hospital number 13, 35, 6, 28 and 18 were the top 5 hospitals in terms of number of visits for asthma patients. Hospital number 13 and 35 had the largest patient count out of all of them.

January, February and May had the highest number of female patients, whereas July and August had the lowest number of patient admissions. Similarly, January may and march had the highest number of male patient admissions, whereas July and September had the lowest number.

The number of days spent in the hospital is calculated and its average is plotted against regions, to find that regions 9, 4 and 7 had patients which spent the highest number of days on average, in hospitals.

Out of all the patients admitted during the time period, more than 60% of the patients were from Delray Beach, followed by Miami, Hobe Sound, and so on. These cities tend to have the highest number of patient count among all cities. The most popular month for asthma patients’ average No. of visits was December, whereas the least popular month was March. Looking at the trend of CHF diagnosed patients, it is visible that a big rise in the month of December. COPD and AMI trends are relatively stagnant and less volatile than that of CHF. The department concerned with heart diseases had the greatest No. of operations out of all departments, followed by General medicine and Oncology. The department of heart diseases had the highest number of operations in January, February and march, and the lowest number in July.

A patient’s age significantly determines the chances of an individual suffering from heart disease, followed by race. The older a patient is, the more likely they are to face chances of heart failure. Plotting the average age of patients along with disease subcategories, we find that Bronchopneumonia had highest average patient age of 84, while Chronic hosp 46 heart had the lowest average patient age of 70. The average age of white patients admitted was 79, while black patients had the average age of 72. Patients diagnosed with Acute Myocardial Infarct, on average, spent the most number of days in ICU than all other patients. Congestive heart failure, pneumonia and obstructive chronic bronchitis were the most common diseases diagnosed to all patients.

Thus, the report gives concrete and useful insights into the healthcare sector of the various counties in the United States of America. These results can help reshape the policy structure of the department of health, and potentially save lives in the country.

# Conclusion

After going through the above report on a detailed study on the healthcare sector, we conclude and find innumerable meaningful and concrete insights which can assist policy makers make effective and quick decisions. The essentiality of these decisions has increased significantly, with such an unstable environment. The insights gathered from the above research can help policy makers adopt these technologies and make decisions in a field which could potentially save lives.

# Recommendations:

In light of the above findings, we make the following suggestions:

1. Since Chronic Heart Failure was the most common diagnosis group in all regions of US and Heart failure was the most common disease, more funds should be allocated for the research and treatment for the same.
2. Regions 3, 8 and 11 had the highest No. of female and African American patients. These regions tend to have a spike, and thus it needs to be determined as to why this spike occurs.
3. January, February and may had the greatest No. of hospital admissions. An investigation needs to be conducted to find out the reason behind the same and whether its coincidental or seasonal. It can be that cold weather can lead to higher chances of heart problems, and thus solutions need to be provided for the same.
4. More than 60% of the patients were from Delray Beach and Miami. Thus, the reason behind the spike in number of cases in Delray beach needs to be determined and solved.
5. There was a sharp rise in the No. of CHF patients in the month of December. This can be due to the increased cold or reduced temperature. Thus, effective decisions should be made to make sure the old population of each region is sufficiently warm.
6. A patient’s age turns out to be the biggest predictor of whether an individual will have heart disease or not. Thus, special care needs to be employed for the elderly for each county.

# References

Insights, V. (2018, April 17). *10 Advantages of Data-Visualization*. Retrieved from Visualr: https://visualrsoftware.com/advantages-data-visualization/

James, L. (2019). *Big Data Analytics: The need for pragmatism, tangible benefits and real-world case studies*. Retrieved from Yellowfin.com: https://www.yellowfinbi.com/blog/2014/06/yfcommunitynews-big-data-analytics-the-need-for-pragmatism-tangible-benefits-and-real-world-case-165305

King, A. (2018, May 4). *7 Benefits to Using Big Data for Small Businesses*. Retrieved from Industriuscfo.com: http://www.industriuscfo.com/7-benefits-using-big-data/

Learntek. (2017, January 17). *What are the advantages of BIG DATA Analytics, and how will it impact the future*. Retrieved from Learntek.com: https://www.learntek.org/blog/advantages-big-data-analytics/

SAS. (2018). *What is Big Data Analytics and Why it Matters*. Retrieved from Sas.com: https://www.sas.com/en\_in/insights/analytics/big-data-analytics.html

Verma, A. (2018, March 19). *Why is Big Data Analytics So Important?* Retrieved from Whizlabs.com: https://www.whizlabs.com/blog/big-data-analytics-importance/