Project Proposal

INFO 201 – Technical Foundations

Team Lays

ZOSHUA COLAH, JUSTIN WANGINGTH, STEVEN BISHOP, YUK HEI NG

PROJECT DESCRIPTION

About Our Dataset

As part of our project we will be working with the *"Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011"* data set of information created and maintained by the US Government.

A Diagnosis Related Group is a statistical system of classifying any inpatient stay into groups for the purpose of payments. The DRG classification system divides possible diagnoses into more than 20 major body systems and subdivides them into almost 500 groups for the purpose of Medicare reimbursement.

The data set we are using in our investigation is a collection which contains a summary for each of the top 100 DRGs for over 3000 hospitals in the United States of America that receive Medicare Inpatient Prospective Payment System Payment paid under Medicare based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for Fiscal Year (FY) 2011. These DRGs represent more than 7 million discharges or 60 percent of total Medicare IPPS discharges. These summaries contain information about the total discharges, the average covered charges, the average total payments and the average Medicare payments made by the patients who have been discharged.

Link to dataset:

https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3

Link to information about the dataset:

https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3/about

Sources:

https://www.healthlawyers.org/hlresources/Health%20Law%20Wiki/Diagnosis-related%20group%20(DRG).aspx

Our Investigation

As part of our investigation of this data we want to compare the average total payments and average Medicare payments for each DRG of each state and city and for each hospital to try and identify the areas where it is costlier to receive medical treatment.

We also wish to investigate the total discharges for a particular DRG and compare the total discharges for each DRG of each state so as to identify accordingly if a particular state suffers more than the other.

Through this investigation of ours we want to help

- those seeking medical treatment and are facing problems in financing heavy medical bills

- soon to be retiring elderly people because as they turn old they will need more and more medical procedures as medical issues increase with age. Hence they will want to find an area which best suits their current financial status so as to ensure their medical issues do not harass them in the future.

- people who are looking to ensure they receive good health insurance. Through our investigation they will be able to identify which areas receive better average Medicare payments than others.

From our investigation of this dataset our audience would like to learn:

1. Which state provides good hospital service in order to find a nice place to settle down after they retire
2. People would like to know how much will cost them to pay for basic health services in area they are thinking about relocating to
3. People would like to know the state resident health condition in the state.

TECHNICAL DESCRIPTION

About Our Analysis

As part of our investigation we will be making use of Shiny App to document our investigation and analysis of the data. The data we will be using will be exported from the data.cms.gov website in the CSV format and will then be imported into R and used as a data frame.

We will be mutating new columns to the data frame to calculate the amount Medicare reimburses a patient with and accordingly with the new found data be able to try and identify for which city, state and hospital is Medicare better.

During these calculations we will be using the group by and summarize function of DPLYR as well so as to restructure our data. We will also be reformatting this data frame such as to find out the total discharges for each procedure for a particular state.

As part of our analysis we are making use of the Plotly library to visualize our findings so as to make our analysis easier to understand for a user and so as to be able to identify trends in data through our analysis.

The major challenges we will be facing would be in identifying the correct method by which we should restructure our data and what calculations we should do to be able to ideally identify which state is better to go to as each calculation may give different results and identifications.