Hospital Inpatient Charges for top 10 DRGs in the U.S.

Problem

The Centers for Medicare and Medicaid Services (CMS), which is a division of the U.S. Department of Health and Human Services, is responsible for medicare reimbursement to hospitals/providers in the U.S. The purpose of this project is to examine the distribution of payments across all states for specific Diagnosis-related groups (DRGs).

Client

As the agency monitoring payments to hospitals for medicare eligible inpatient expenses, the CMS would potentially be interested in knowing whether certain hospitals are overcharging for the same diagnoses as compared to others. Charges and payments may differ based on provider state, number of discharges, income groups of the patients as well as whether the hospital is a teaching hospital or not. Based on the analysis of billing and payment information, the CMS would determine whether the payment rates need to be modified for particular providers.

Data

The hospital inpatient charges dataset for the top 10 DRGs in the U.S. is owned by the US government and is freely available on Data.gov. It is also available on Kaggle.com.
The dataset consists of the following information:
- DRG Code and definition
- Provider ID and Name
- Provider Street Address
- Provider City
- Provider State
- Provider Zip Code
- Hospital Referral Region Description
- Total Discharges
- Average Covered Charges
- Average Total Payments
- Average Medicare Payments

Analysis

The goal of the project is to gain insights into the inpatient billing and payment information so as to aid the CMS in making policy decisions related to payment rates assigned per DRG for each provider.

The specific objectives of the analysis are:
1. To determine the distribution of charges for particular DRGs across states
2. Visualize the differences in charges for DRGs across states
3. To assess the differences in Medicare reimbursement for the top 10 billed DRGs across states

Data will be accessed and cleaned. Data visualization techniques will be utilized for data exploration. Clustering will be performed in order to identify patterns within the dataset. Linear regression will be performed to discover associations between charges and other variables in the dataset.

Deliverables

The deliverables for this project are:
1) R Code

2) Project Report
3) Presentation Slide Deck