Sadasivuni Sai Chaitanya
AIT580
G01241462

Hospital Charge Data

Describing the data set

The data which I chose will show the hospital-specific charges for more than 3000 hospitals in the U.S that receive Medicare Inpatient Prospective System (IPPS) Payments charged by the hospitals on both the Inpatient and Outpatient for the common services used for the fiscal year 2011. In this dataset, we have the providers i.e. hospitals revealing the data of charges for various categories depending on diagnosis-related groups. This dataset has grabbed my attention with a whopping amount of data which has 163,065 records which can be used in many ways to show different statistics and multiple varieties of visualizations.

Source: (Inpatient Charge Data FY 2011 n.d.)

Importance of dataset and what questions it might address

With this huge amount of data, based on the DRG (Diagnosis-related groups), we can find out what is the most occurred case i.e. with which issue is most of the cases coming to the health care providers. This can be filtered with the states, regions and ZIP codes. Apart from this, the data gathered also reveals how much payment has been done by both groups of patients.
The visualization which can be created by this data might clearly show which diagnosis-related groups stand in which position, the count of discharges from each healthcare provider and also the average covered charges, total payments, and the average Medicare payments.

Source: (Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011 2014)

Method of analysis

The data that has been gathered first will have to go through some data preparation methods where the information will be tested for fraudulency, validated by checking if it is complete and suits the research criteria. Editing of the data will take place if there are any outliers identified while performing the basic data check and issues related to missing data will be addressed by some techniques such as Mean/Median/Mode Imputation, Last Observation Carried Forward (LOCF). The process of coding in R, Python, and SQL will take place and visualization tools like tableau software will be used to create visuals depicting this information which will support the conclusion.

Source: (4 Ways to Deal with Missing Values n.d.)

Sadasivuni Sai Chaitanya
AIT580
G01241462

Works Cited

n.d. "4 Ways to Deal with Missing Values ." *Humans of data .* https://humansofdata.atlan.com/2017/09/4-methods-missing-data/.

n.d. "Hospital Charge Data ." *data.gov.* https://www.data.gov/health/highlights.

n.d. "Inpatient Charge Data FY 2011 ." *cms.gov.* https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient2011.

n.d. "Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011 _Data." *data.cms.gov.* https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3/data.

2014. "Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011." *data.cms.gov.* June 02. https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3.