

Price/Quality Analysis of Healthcare Providers

Eric Tsai, Andrew Lam, Ashton Chevallier
W205 Project Presentation
Thursday, October 20, 2016

Problem Statement

- With the introduction of High-Deductible Health Plans (HDHPs), consumers will need to pay more before insurance kicks in; as a result, there is increasing demand for accurate pricing
- However, medical procedure costs are often opaque to patients. The cost of procedures is never presented to patients up front and patients only ever find out about the cost after the fact when a bill arrives.
- Even the doctors that prescribe and perform the procedures are often in the dark about the cost.
- Moreover, price is only one of many criteria - consumers need greater transparency in pricing and quality to make informed decisions about their health

Goals

- The goal of the project is to empower consumers with information about the quality of hospitals relative to the price paid
- Analyze and combine large datasets into a user-friendly, consistently updated format
- Automate the difficulty of obtaining and analyzing this data

Data Sets

- Medicare data
 - Hospital compare data
 - Hospital charge data
- Hospital rating sites
 - Yelp
 - Hospital Safety Score
 - Health Grades
 - Consumer Reports
- Geographical data
 - Open Street Maps
 - Census Geocoder

Approach

- Locate available data sources for relevant data on medical procedure prices and metrics on procedure outcomes (HDFS)
- Extract the data on hospital prices and hospital quality of service
 - Scrape data from sites
- Clean and transform the data (Spark)
 - Convert addresses into lat and long using Census Geocoder API and Open Street Maps
- Run analytics on the data to generate aggregate statistics (Spark)
- Store the transformed data in a database (Postgres)
- Build a presentation layer to enable users to query against the database to find information

Challenges

- Locating additional data sources to feed into the system
- Ratings data is non-trivial to normalize and compare across datasets
- Cleaning data into a format that compares correctly between datasets of different origins
- Setting up data to properly scrape updates consistently and format them properly and incorporate into existing data will be hard
- Building a useful serving layer