Is Your Hospital Bamboozling You? Helping Patients Shop for Medical Care in Michigan

Team 54: Nadia Ayub, Nancy Njuguna, Junting Huang, Evelyn Wangai, Amber Johnson, Bri Jordan, Kasole Nyembo

Background

Healthcare costs are one of the largest categories of consumer spending in the U.S., but hospital charges have been historically opaque. Effective 2021, the federal government promulgated a new rule for hospitals requiring them to disclose pricing on a wide range of hospital services. This project explored how this price transparency regulation can help patients understand costs, and examine the relationship between hospital pricing and demographic data. Our goal is twofold:

- To develop a tool which allows patients to research costs for various procedures; this tool will reduce their financial burdens and empower them to take a proactive approach with their health.
- To provide an analysis on the variability of hospital pricing based on demographic and hospital statistics.

Data Sourcing and Cleaning



Figure 1: Data Schema & Dataframes

We pulled our data from three main sources: (1) the pricing data from hospital websites, (2) the hospital data from the American Hospital Directory, and (3) the population and demographic data primarily from the Michigan State Department of Health and Human Services. We developed a python script that extracted relevant information from individual data files.

Overall, our dataset contained a total of 120 hospitals in 18 different hospital systems in Michigan. By the end of the data cleaning phase, we had approximately 5-19.5% of missing data.

clas	ss 'pandas.core.fr	ame.1	Da
Range	eIndex: 590 entrie	s, 0	t
Data	columns (total 25	colu	ım
#	Column	Non-	-N
0	price_id	590	n
1	cash_price	500	n
2	gross_charge	569	n
3	max_ng	435	n
4	min_ng	435	n
5	capacity	590	n
6	revenue	562	n
7	tps_score	590	n
8	pop_density	590	n
9	household_income	590	n
10	unemployment	590	n
11	sex_ratio	590	n
12	children_percent	590	n
13	senior_percent	590	n
14	white_percent	590	n
15	high_school	590	n
16	pcp_ratio	590	n
17	life_exp	590	n
18	uninsured	590	n
19	cpt_code	590	n
20	urban	590	n
21	ownership	590	n
22	system	590	n
23	hospital_name	590	n
24	zip_code	590	n
ltype	es: float64(16), i	nt64	(2



Figure 2: Michigan Hospital Pricing Overview With selected procedures (by CPT code), we started with a pricing comparison of different hospitals and systems.



Figure 4: Grouped Scatter Plots

After standardization, we plotted scatter plots to observe trends in and largely found a negative correlation with hospital data (quality score, capacity and revenue) and a positive correlation with some demographic data (senior population, white population, population density, and unemployment).

			P	red	ict	ive
Model	Explained Variance Score	Max Error	Median Absolute Error	Mean Absolute Error	R^2 Score	Accuracy Rate
Linear Regression	0.218	3.467	0.672	0.465	0.1797	54%
Bayesian Regression	0.229	3.524	0.666	0.484	0.1960	55%
Least Angle Regression	0.198	3.509	0.678	0.486	0.1707	Did not test
Stochastic Gradient Descent	0.226	3.499	0663	0.479	0.1903	Did not test
K- Nearest Neighbors	0.3222	3.447	0.598	0.423	0.2804	59.22%
Support Vector Machine	0.318	3.521	0.596	0.403	0.2728	Did not test
Decision Tree	-0.04	4.004	0.646	0.395	-0.058	Did not test
Random Forest	0.444	3.616	0.525	0.402	0.4243	66.60%

Figure 5: Model Selection & Predicted and Original Prices Comparison We explored various predictive models where Random Forest was the best with a 90% accuracy after optimization (66% baseline) after using the random search and grid search methodology.



taFrame'> o 589 ns): ull Count Dtype float6 float64 on-nu] float64 on-null string on-nul object on-null object on-null object on-null string), object(4), string(3

Highlights

- resources in Michigan.
- pricing data, and improving the ability of patients to use this information.

Figure 3: Correlation Matrix We plotted a correlation matrix to better understand the interactions between variables.



lodeling







Figure 6: Interactive Dashboard determine which hospital to choose for their respective procedures.

Office_Visit



1. Hospital Pricing is Weakly Correlated to A Set of Key Demographic and Hospital Variables:

- Cash price is higher in areas with lower population density.
- Cash price is higher in counties with higher unemployment, higher
- Cash price is higher with hospital with lower capacity and revenue.

2. Demographic Data Provides Insights into the Discrepancies of the **Uneven Distribution of Healthcare Resources in Michigan**

Note: Although we limited our conclusion to the overall state, each county within the state showcased a combination of different trends.

This study ultimately helps the patients by empowering them to go "shopping" and to compare the prices of all selected procedures within their residential area.

Although hospital pricing is weakly correlated to a set of key demographic and hospital variables, demographic data provides insights into the discrepancies of the uneven distribution of healthcare

2. Future policy efforts should focus on enforcing the price transparency law, standardizing hospital

Interactive Dashboard



Our patient dashboard allows users to compare prices across hospitals. Based on the pricing information (original prices, predicted prices, and price indexes), they can

Key Takeaways

Figure 7: Price Index & **Geographical Clustering** We noticed a clear pattern of clustering, which forms distinctive groups of data groups. Although these differences do not seem to heavily determine the prices, the clusters offer insight regarding the distribution of healthcare resources in the state and gather observable conclusions regarding resources and its effect on hospital's pricing strategies.

• Cash price is higher in areas with more seniors and white residents.

uninsured rate, and lower household income.