Large amounts of data in the healthcare industry are being collected every day. How can we use this data, and display it in a way that everyone can easily follow?

How Did We Do This?

• Our team's original focus was to work directly with a company, TeamHealth, to look at their record hospital data, and determine what factors impact the propensity to pay.

• However, the project ended up working on a broader scale. We instead utilized the databases from the following:
  - Healthcare Cost and Utilization Project’s (HCUP)
  - National Inpatient Sample (NIS)

• Our goal was to identify trends and correlations within these sets in order to learn about what patient attributes has the most impact on the healthcare industry. Then put this information in a format that everyone can easily comprehend.

Our Findings

Our team utilized two primary software's: R and Tableau. We used them to

• Clean the data
• To build predictive models for certain attributes within the data
• To visualize all our findings from working with the predictive models.

Model Results/Visualizations

Our Findings

About Our Data Set

• It is the largest publicly available all-payer inpatient database in the U.S.
• It represents a 20% stratified sample of U.S. hospitals
• It contains 7 million hospital stays per year.
• There are 100+ clinical and non-clinical variables.
• There are 3 files: Core, Severity, and Hospital.

These models can be applied in the healthcare industry by:

• Predicting Hospital Capacity
• Predicting a Patient's Propensity to Pay
• Help Inform Where the Hospital Should Focus Manpower & Resources