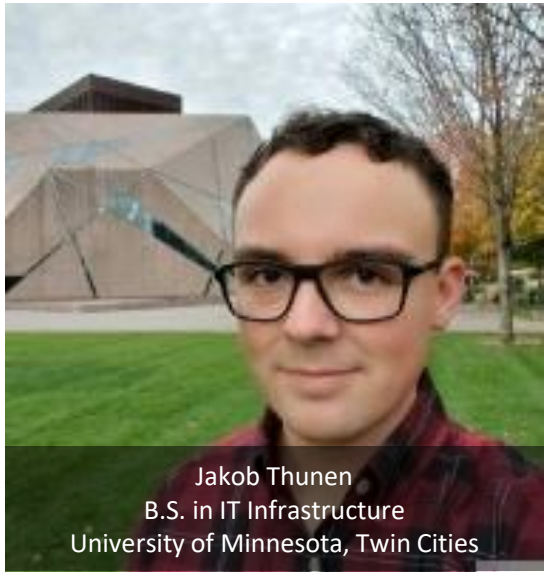
Public Health Insurance, Quality of Care and Payments

Predicting Payments (Price)
Using Azure Products, Kafka,
Python, and Machine Learning

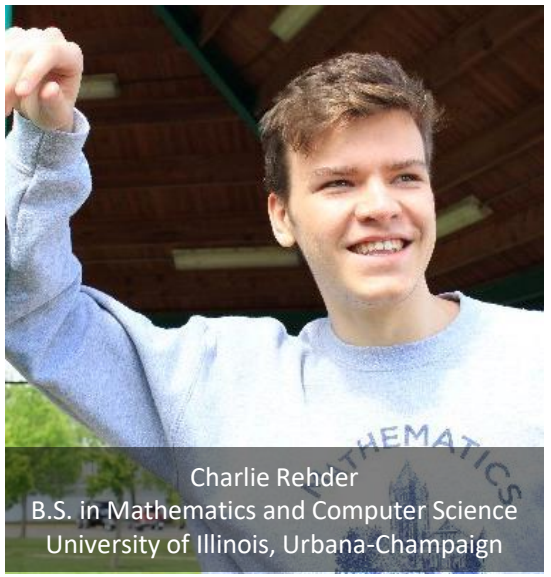
Project Objective

When it comes to healthcare, there are numerous contributors to the overall quality of care given to patients. This project uses data from the United States Census Bureau and the Centers of Medicare and Medicaid to better understand the demographics of the population on public health insurance, the impact of quality of care on outcomes and cost, and ultimately, use this understanding and analysis to predict payment. Our prediction model could be utilized by stakeholders, hospitals, and government agencies to better predict price of hospitalizations and illustrate the importance of linking quality of care to public payment and reimbursement programs.

Group Members



Jakob Thunen
B.S. in IT Infrastructure
University of Minnesota, Twin Cities



Charlie Rehder
B.S. in Mathematics and Computer Science
University of Illinois, Urbana-Champaign

{dev10}
genesis10

Alysia Halverson
B.S. in Family Social Science
University of Minnesota, Twin Cities

Project Goals

Research	Determine which measures have the greatest impact on patient quality of care by investigating possible correlations between value of care scores and payments as well as researching trends across states
Analyze	Compare payments, quality of care, and public health insurance enrollment across states to identify relationships or trends using visualizations and data analysis tools
Predict	**Using machine learning, build a model with high reliability to predict price of a patient's hospital stay using associated features to train and test the model**

Research Process & Data Sources

The Centers for Medicare and Medicaid Services

HVBP Datasets (2022):

- Efficiency and Cost Reduction: *ratio of hospital's Medicare Spending per Beneficiary (MSPB) dollar amount over national median MSPB dollar amount*
- Person and Community Engagement: *percentage of survey respondents who selected "top box" answer for respective domain (communication, care transition, etc.)*
- Safety: *ratio of observed number of infections while hospitalized over expected number of infections while hospitalized*
- Clinical Outcomes: *mortality (survival rate) vs. complication (mortality rate)*

Payment and Value of Care – Hospital (2022):

- Payment and value of care: *payment includes all payments by Medicare, secondary insurance, and patients for services in 30-day hospitalization period or 90-day hospitalization period (depending on hospitalization reason)*

U.S. Census Bureau

5-Year American Community Survey (2020): S2704 Public Health Insurance Coverage by Type and Selected Characteristics (All States)

- *Population counts per state: total, public health insurance only, Medicare only, Medicaid only, VA-only*

CMS: Data Introduction

[CMS: Hospital Value Based Purchasing (HVBP) Program Overview]

"Designed to promote better clinical outcomes"
"Improve patient experience of care during hospital stays"
"Incentivize hospitals to improve safety of care"

[CMS: Linking Quality to Payment]

"Patients who received high-quality care during their hospitalizations and transition to the outpatient setting will likely have improved outcomes"

Table 1. FY 2022 Baseline and Performance Periods

Domain/Measure Description	Baseline Period	Performance Period
Clinical Outcomes: 30-Day Mortality measures for Acute Myocardial Infarction (AMI), Coronary Bypass Graft (CABG) Surgery, Chronic Obstructive Pulmonary Disease (COPD), and Heart Failure (HF)	July 1, 2012–June 30, 2015	July 1, 2017–June 30, 2020
Clinical Outcomes: 30-Day Mortality measure for Pneumonia (PN) (updated cohort)	July 1, 2012–June 30, 2015	September 1, 2017–June 30, 2020
Clinical Outcomes: Total Hip Arthroplasty (THA)/Total Knee Arthroplasty (TKA) Complication measure	April 1, 2012–March 31, 2015	April 1, 2017–March 31, 2020
Person and Community Engagement: Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) dimensions	January 1–December 31, 2018	January 1–December 31, 2020
Safety: Healthcare-Associated Infection (HAI) measures	January 1–December 31, 2018	January 1–December 31, 2020
Efficiency and Cost Reduction: Medicare Spending per Beneficiary (MSPB) measure	January 1–December 31, 2018	January 1–December 31, 2020

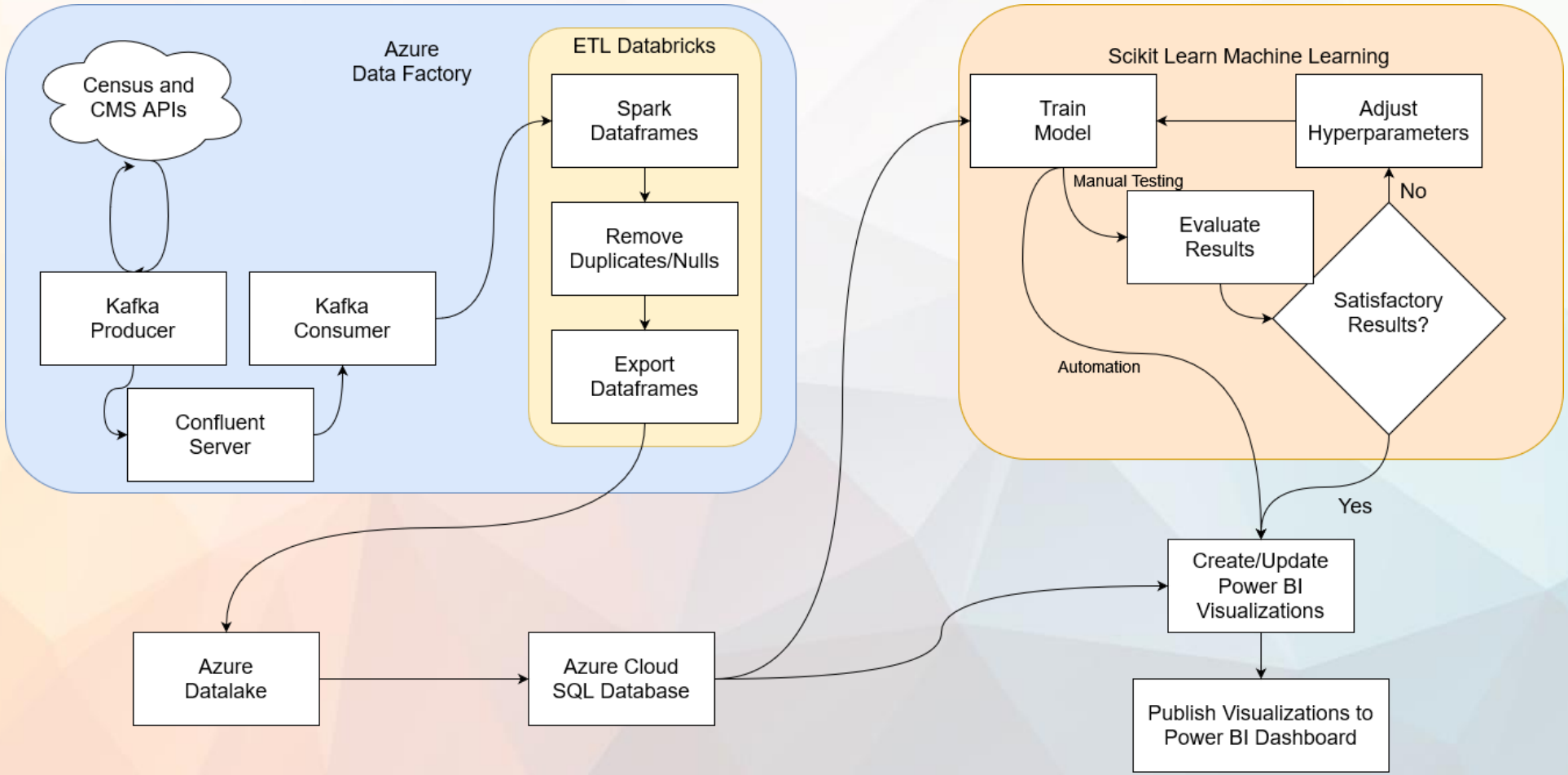
Initial Questions

- Do price and value of care always align?
- Will safety have a large influence on the price of hospitalizations?
- Would higher efficiency scores decrease the price of hospitalizations or increase them?
- How might community engagement influence the price of hospitalizations?
- How might clinical outcomes influence the price of hospitalizations?
- What are trends among U.S. states and those insured on public healthcare?
- Are there patterns between value of care and states with higher public insurance rates?
- Do hospitals with higher safety scores suffer with lower efficiency scores?
- Do states that are more/less urbanized have higher/lower community engagement scores?
- Do hospitals with higher community engagement scores have higher clinical outcomes scores?

Data Handling

Technologies Used

Azure Data Factory · databricks · pandas · scikit learn · CONFLUENT · PySpark · Azure Data Lake Storage Gen2 · Azure SQL · jupyter · Power BI



Power BI Dashboard

Machine Learning & Prediction

"When given associated features and utilizing machine learning regression models, how accurately can we predict the price (payment) of a patient's hospital stay?"

The machine learning data set merges tables to include over 30 features in our ML process with our y-variable being payment.

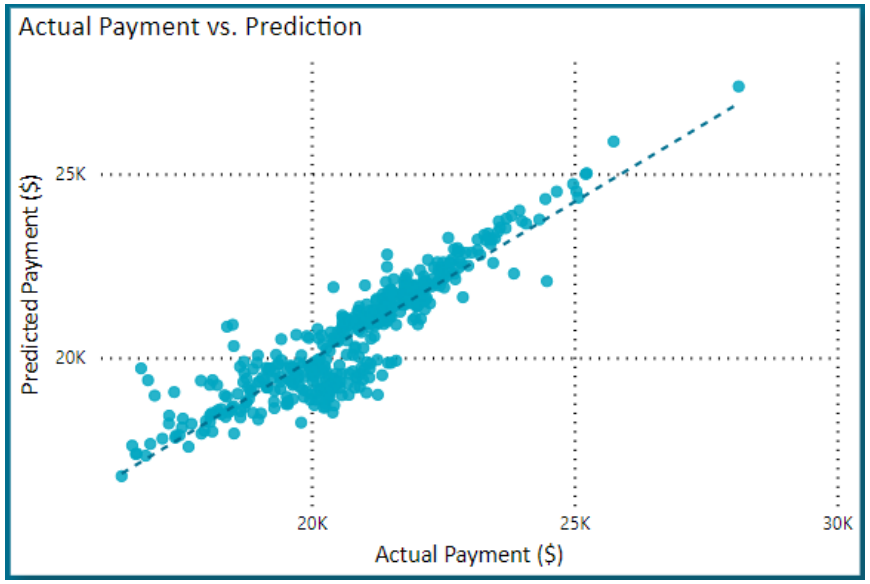
- Baseline and Performance Period scores for each of the four domains and their respective categories make up the 30+ features

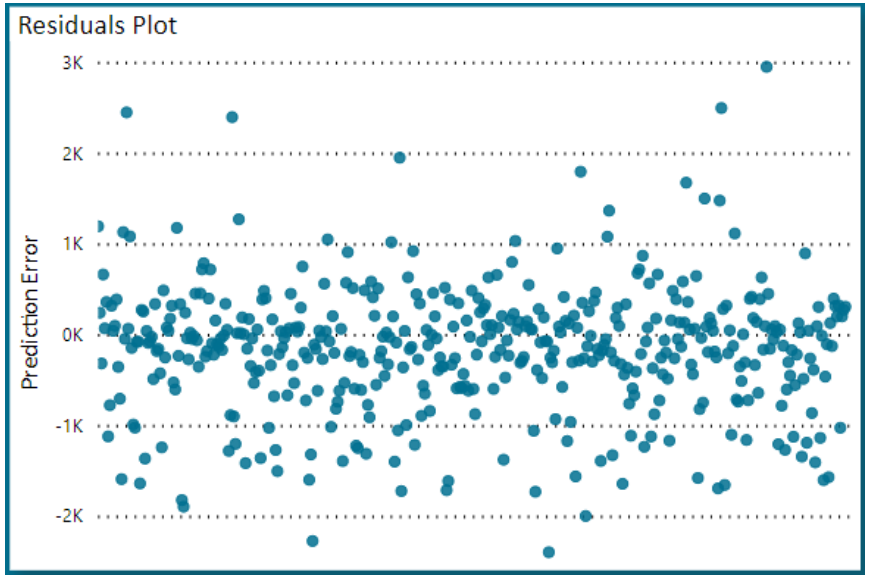
A successful machine learning model with the goal of accurately predicting price for the use of a participating hospital or CMS could provide important insights, confirm assumptions linking payment to quality of care, and present these findings to stakeholders.

Machine Learning Process & Findings

Four models were tested:

- Linear Regression Payment model (baseline)
  - **R^2 score = 0.643**
- Lasso Regression Payment model (trying new variant)
  - **R^2 score = 0.642**
- Random Forest Regressor model (hyperparameter tuning, ensemble method)
  - **R^2 score = 0.676**
- XGBoost model (extensive hyperparameter tuning and CV, maximizing ensemble method)
  - **R^2 score = 0.783**
  - Plots shown





Discussion

Our data analysis found a **strong positive correlation** between states' total population and publicly insured population. However, no strong relationships were found when looking at quality of care patterns by state public insurance enrollment rates or urbanization percentages.

As shown in the Power BI Dashboard Report, we concluded that there was a **lack of a relationship** between clinical outcomes and community scores.

In terms of cost reduction, we found that higher efficiency scores did not lead to higher infection rates in hospitals concluding that there is **not a visible risk to safety** by hospitals trying to match the national median MSPB.

However, because of the pandemic, many measures were not reported or not considered when determining HVBP measures so it would be **critical to re-assess in future** reporting periods to have a more accurate determination of impact between safety scores and efficiency scores.

Overall, we have determined that payment and value of care do not always align, but there is enough of a relationship among the value of care measures to reliably predict price when utilizing our XGBoost model compared to other regression models.

This concludes the presentation. Any questions?