

## Data Sources

### The Centers for Medicare and Medicaid Services

To better understand health insurance quality of care, specifically related to Medicare and Medicaid payments and quality measures, we will be utilizing the following five datasets. A complete data dictionary can be found here ([linked](#)). All of the following datasets have a related variable, Facility ID, that we will use to join the data.

- *Hospital Value-Based Purchasing (HVBP) - Efficiency Scores*. (2022, April 27). [Dataset]. <https://data.cms.gov/provider-data/dataset/su9h-3pvi>
  - The Efficiency and Cost Reduction dataset provides details on the Medicare Spending per Beneficiary (MSPB) measure. The MSPB measure is the hospital's MSPB measure ratio calculated as MSPB amount (hospital's Medicare spending per beneficiary dollar amount) divided by the median MSPB amount (national median MSPB dollar amount).
- *Hospital Value-Based Purchasing (HVBP) - Person and Community Engagement Domain Scores (HCAHPS)*. (2022, April 27). [Dataset]. <https://data.cms.gov/provider-data/dataset/avtz-f2ge>
  - The Person and Community Engagement Domain Scores dataset provides details on eight dimensions related to patient experience of care collected from Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) surveys. CMS calculates these scores using the "top-box raw score", which is the unrounded percentage of a hospital's patients who chose the most positive response to the HCAHPS survey items. More information about these scores can be found in the HVBP Step-by-Step Guide ([linked](#)). Higher percentages correspond to higher quality.
- *Hospital Value-Based Purchasing (HVBP) - Safety*. (2022, April 27). [Dataset]. <https://data.cms.gov/provider-data/dataset/dgmq-aat3>
  - The Safety Measures dataset provides details on Healthcare-associated infections (HAI). These measures are standardized infection ratios (SIRs) using the number of observed infections (numerator) divided by the number of predicted infections, calculated by the CDC, (denominator). Lower values correspond to higher quality.
- *Hospital Value-Based Purchasing (HVBP) - Clinical Outcomes Domain Scores*. (2022, April 27). [Dataset]. <https://data.cms.gov/provider-data/dataset/pudb-wetr>
  - The Clinical Outcomes dataset provides details on mortality on Acute Myocardial Infarction (AMI), Heart Failure (HF), Pneumonia (PN), and Total Hip/Knee Arthroplasty (COMP-HIP-KNEE) in a survival ratio. Higher values indicate better outcomes.
- *Payment and value of care - Hospital*. (Jan 26, 2022). [Dataset]. <https://data.cms.gov/provider-data/dataset/c7us-v4mf>
  - The Payment and Value of Care dataset provides details on payments and cost estimates. The dataset also includes payment category (Is the payment greater, no different, or less than the average payment?) and value of care category (Is the mortality better, average, or worse? Are the complications average or worse? Was the payment high, low, or average?). Our ML model will attempt to predict payment price using these features.

## United States Census Bureau

- U.S. Census Bureau, *2015-2019 5-Year American Community Survey*, Table S2704: Public Health Insurance Coverage by Type and Selected Characteristics (All States within United States). (2020) [Data set].  
<https://data.census.gov/cedsci/table?t=Health%20Insurance&g=0100000US%240400000&tid=ACST5Y2020.S2704&tp=true>
- The Public Health Insurance Coverage by Type and Selected Characteristics (All States within United States) will provide population counts for the state, the number of residents on public health insurance alone, and the breakdown populations for each type of public health insurance: Medicare, Medicaid, and VA-only. We will merge on State/State ID to join this dataset with our other data.