

# STATS 506 - Final Project - Proposal

Shenyi Tang

November 2024

## 1 Research Question

The research question is to identify the primary factors driving higher Medicare paid amount for cardiologists in regions with different income levels (categorized as low, middle and high income level). Through seeking key factors, the findings may help guide policies for effective distribution of medicare funds.

## 2 Analysis Plan

### 2.1 Data Integration

Data related with cardiological operations will be extracted from the **Medicare Physician and Other Practitioners - by Providers and Services**, focusing on variables such as services type, provider demographics etc. Tax related data will add income information at the ZIP code of state level, by which I will merge with the Medicare data.

### 2.2 Income Level Classification

Basically, I will categorize regions as low, middle, or high income based on percentiles. Besides, KMeans clustering will help verify and refine these income groupings.

### 2.3 Modeling

The objective variable will be Medicare paid amount. Related features such as service codes, service volume, place of service, providers related information will be selected.

XGBoost will be used for characteristic selection, to evaluate the importance of each features.

### 2.4 Visualization and Interpretation

Graphs will be used to reveal which variable most affect Medicare reimbursement. The outcome of the model will provide insights into how do different features influence medicare payments and specify the differences between regions with different income levels.