What is the problem you want to solve?

Provider Fraud is one of the biggest problems facing Medicare. According to the government, the total Medicare spending increased exponentially due to frauds in Medicare claims. Healthcare fraud is an organized crime which involves peers of providers, physicians, beneficiaries acting together to make fraud claims.

Problem Statement:

Identifying providers engaged in fraudulent activities would increase the chances of fairness and equity by 10% within a year by identifying significant variables that aid in detecting suspicious provider behavior.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

The government would use the analysis to raise awareness of pricing strategy. They would invest in analysis to ensure resources are being used wisely. Identifying significant variables would enhance services for patients.

What data are you going to use for this? How will you acquire this data?

I will use the dataset from Kaggle https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis/data.

In brief, outline your approach to solving this problem (knowing that this might change later).

I will utilize a clustering algorithm to gain a holistic perspective of the dataset. Additionally, I will conduct numerous smaller comparisons between individual features within the dataset.

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

I will provide code for implementing the models, a comprehensive paper detailing the methodology, and an extensive analysis of the results. Additionally, a slide deck will be included to succinctly present the findings.