**Project Objectives**

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

Rigorous analysis of population data has yielded many physicians who indulge in fraud. They adopt ways in which an ambiguous diagnosis code is used to adopt costliest procedures and drugs. Insurance companies are the most vulnerable institutions impacted due to these bad practices. Due to this reason, insurance companies increased their insurance premiums and as result healthcare is becoming costly matter day by day.

Healthcare fraud and abuse take many forms. Some of the most common types of frauds by providers are:

a) Billing for services that were not provided.

b) Duplicate submission of a claim for the same service.

c) Misrepresenting the service provided.

d) Charging for a more complex or expensive service than was actually provided.

e) Billing for a covered service when the service actually provided was not covered.

Problem Statement  
The goal of this project is to " predict the potentially fraudulent providers " based on the claims filed by them.along with this, we will also discover important variables helpful in detecting the behaviour of potentially fraud providers. further, we will study fraudulent patterns in the provider's claims to understand the future behaviour of providers.

Introduction to the Dataset  
For the purpose of this project, we are considering Inpatient claims, Outpatient claims and Beneficiary details of each provider. Lets s see their details :

A) Inpatient Data

This data provides insights about the claims filed for those patients who are admitted in the hospitals. It also provides additional details like their admission and discharge dates and admit d diagnosis code.

B) Outpatient Data

This data provides details about the claims filed for those patients who visit hospitals and not admitted in it.

C) Beneficiary Details Data

This data contains beneficiary KYC details like health conditions , region they belong to etc.

Data Set --

* Training Data Set is provided to you on which you have to build the predictions.
* Test/Unseen data is also provided on which you have to submit your predictions.

Your tasks and Evaluation Criteria

1. **Data Management**
2. **Exploratory Data Analysis**
3. **Feature Extraction and Engineering**
4. **Feature Selection Process**
5. **Modelling**
6. **Evaluation of Model built**
7. **Business Recommendation and Improvements.**

Domain Knowledge – You can learn more about US healthcare insurance system and fraud from any open source / search engines.

Submission

1. Submit the file with name “Your Full Name\_Submission”.csv with following columns
   1. Probability (if you choose this kind of model).
   2. Predicted Class as per your best model.
2. Submit a notebook with the printed results of your code which is readable in a HTML.
3. (Optional ) Submit a word document or a PPT as per your comfort with your analysis, approach taken, recommendations.

Timelines -- 72 Hours

Tech Stack – You can use any stack preferred by you.

Privacy and Citations –

1. You need to ensure that this data is not shared or case study is not published or cited any where.
2. If you have referred to any link, article, paper, pl cite it in your notebook submission file.