DS Technical Task

- **Problem definition:**

  We have a medical claim dataset, each row represents a patient, while each row represents a medical service or a personal information.

- **The main targets are:**

  1- Apply a technique for classifying whether each patient (row) implies normal or fraud behavior (might be ML pipeline or other statistical approach). Note, we expect only 1 ML algorithm for classification.

  2- Apply un unsupervised ML pipeline for anomaly detection based given the fact that there are no labels for the claims (you can apply other statistical approach instead), we expect only 1 ML algorithm.

  Kindly, find the dataset (https://bit.ly/3wmEaK5) which will be used for building& training& testing your model [note: the file contains two sheets one for dataset and the other for features description].

  Note, sometimes u might need to apply needed preprocessing to generate& split& merge features. Also, regarding the two columns "Claim amount & Paid Amount", you can use them to determine whether a claim is normal or fraud.

  As a potential data scientist, we expect to pay much attention to both:

  - Preprocessing: including any needed further operations that need to be applied firstly.
  - Feature engineering: Whether there is a need for any quantization, merging, generating of new features.
  - Explainability and business impacts: After applying the model, please shed more lights on the business value, the medical impact, and the financial impact you are adding. Visualizations and meaningful graphs are welcome too. Add a section in your code where you show the outcome of your work to a client.

- **Expected feedback:**

  Two .PYNB files, one for each task. Where for each one; the input: one raw (like that in the dataset) / or [csv - excel] file contains multiple rows, the output: Normal or Fraud.