**IMPLEMENTATION IDEA FOR SOLVING STRUCTURED DATA ASSIGNMENT**

**The code is not completed, but here is an approach to how we can solve it**

* The input train and test files are shared as parquet files.
* These files are loaded to python-pandas using the pyarrow package.
* The data contains historical data of 27k unique patients, out of which 9k patients are treated with the target drug.
* This problem can be considered as time series data.
* There are two ways to approach the problem,

1. Time series analysis
2. Frequency-based analysis

* I choose the Frequency-based approach first, in which we will check the cumulative sum count of drugs a patient is given until he was treated with the target drug.
* Then dummies will be created for individual incidents, which will be considered as feature variable
* The Patient is given a target drug or not will be encoded as 1 and 0
* This will be considered the target variable
* Then a classification model will be applied between features and the target variable

**With given time and resources I am positive that a model can be developed for this problem**