Installing the package by this:  
<https://github.com/lucidrains/tab-transformer-pytorch>

First, loading, cleaning the dataset and pretraining on this synthetic dataset

It worked with a synthetic dataset for pretraining and random data for fine tuning, but with high loss values around 1 for pretraining and around 0.6 for fine tuning.

It has a big problem, the feature names for tables in pretraining and fine tuning must be the same, even the order must be the same.

only the categorical features are embedded in a contextualized fashion using a Transformer. The contextualized categorical embeddings are then concatenated with the (raw) numeric features and subsequently processed by an MLP for the final downstream task.