# USE CASE 3: Generative models on different datasets

## Instruction:

1. Start a new notebook (anywhere you like, local machine, Google Colab, DSRI, etc)
2. Try on three different datasets:
   1. <https://drive.google.com/file/d/1053Y9qjnI3GVAnXLe39d0qVJ7Evvftz_/view?usp=drive_link>
   2. <https://drive.google.com/file/d/1M7c-Fb0_SojrazJvld4VtGTQPpXiTqg-/view?usp=drive_link>
   3. <https://drive.google.com/file/d/1ecfibGnPTfXwyo-dgAA4jVDwrpExLTGl/view?usp=drive_link>
3. Select at least 1 generative model (you can try more models if you manage to finish 1)
4. Adjust the hyperparameters for different datasets (if needed, and if you want)
5. Select 2 evaluation metrics (distribution, correlations, ML accuracy, efficiency, etc)
6. Get evaluation results and observe the different results on different datasets

## Answer the following questions:

**Q1: What are the basic characteristics of the three datasets? (size, number of numerical variables, number of categorical variables, etc)**

Second dataset: 30162 rows × 16 columns, 6 numerical variables and 9 categorical variables.

**Q2: What model(s) did you choose and if you adjust hyperparameters for different data?**

CTGANSynthesizer

We changed epochs=50,

**Q3: What evaluation metric(s) did you choose and how did the generative model perform on different datasets?**

**Q4: If the generative model has different performance on different datasets, what do you think could be the reason?**

**Question 5: Any discovery or insights to share?**