Jupyter Notebook for generating Synthetic data using the Synthia package

In [1]:
```
import pandas as pd
import synthia as syn

#https://dmey.github.io/synthia/
```

Load data in memory

In [2]:
```
german_credit = pd.read_csv('data/german_credit.csv')
```

In [3]:
```
german_credit
```

Out[3]:

	Creditability	Account.Balance	Duration.of.Credit..month.	Payment.Status.of.Previous.Credit	Purpose	Credit.Amo
0	1	1	18	4	2	
1	1	1	9	4	0	
2	1	2	12	2	9	
3	1	1	12	4	0	
4	1	1	12	4	0	
...	...	...	...	...	...	
995	0	1	24	2	3	
996	0	1	24	2	0	
997	0	4	21	4	0	12
998	0	2	12	2	3	
999	0	1	30	2	2	

1000 rows × 21 columns

Create a Gaussian copula generator and generate 1000 synthetic rows

In [4]:
```
generator = syn.CopulaDataGenerator()
generator.fit(german_credit, copula=syn.GaussianCopula())
N_samples = 1000
synthetic_data = generator.generate(N_samples)
```

In [5]:
```
synthetic_data
```

Out[5]:
```
array([[ 1,  1, 12, ...,  1,  1,  1],
       [ 1,  4, 18, ...,  1,  2,  1],
       [ 0,  1, 21, ...,  1,  2,  1],
       ...,
       [ 0,  4, 36, ...,  1,  2,  1],
       [ 1,  2, 25, ...,  1,  2,  1],
       [ 0,  2, 18, ...,  1,  1,  1]], dtype=int64)
```

Convert to pandas dataframe and then save to CSV format

In [6]:
```
data_frame = pd.DataFrame(synthetic_data.tolist())
data_frame.to_csv("generated_data/synthethic2.csv", index=False)
```