

Harnessing Conditional Generative Models for Synthetic Non-Life Insurance Premium Data

**Insurance Data Science Conference, London
19th June, 2025**

Claudio G. Giancaterino

Synthetic Data



[AI Image Generator](#)

Synthetic data generation represents a transformative approach to addressing critical data challenges in non-life insurance, where privacy regulations, limited sample sizes, or incomplete coverage across risk segments often constrain traditional datasets.

Synthetic data offers a solution by creating artificial datasets that preserve original statistical properties and correlational structures.

Process

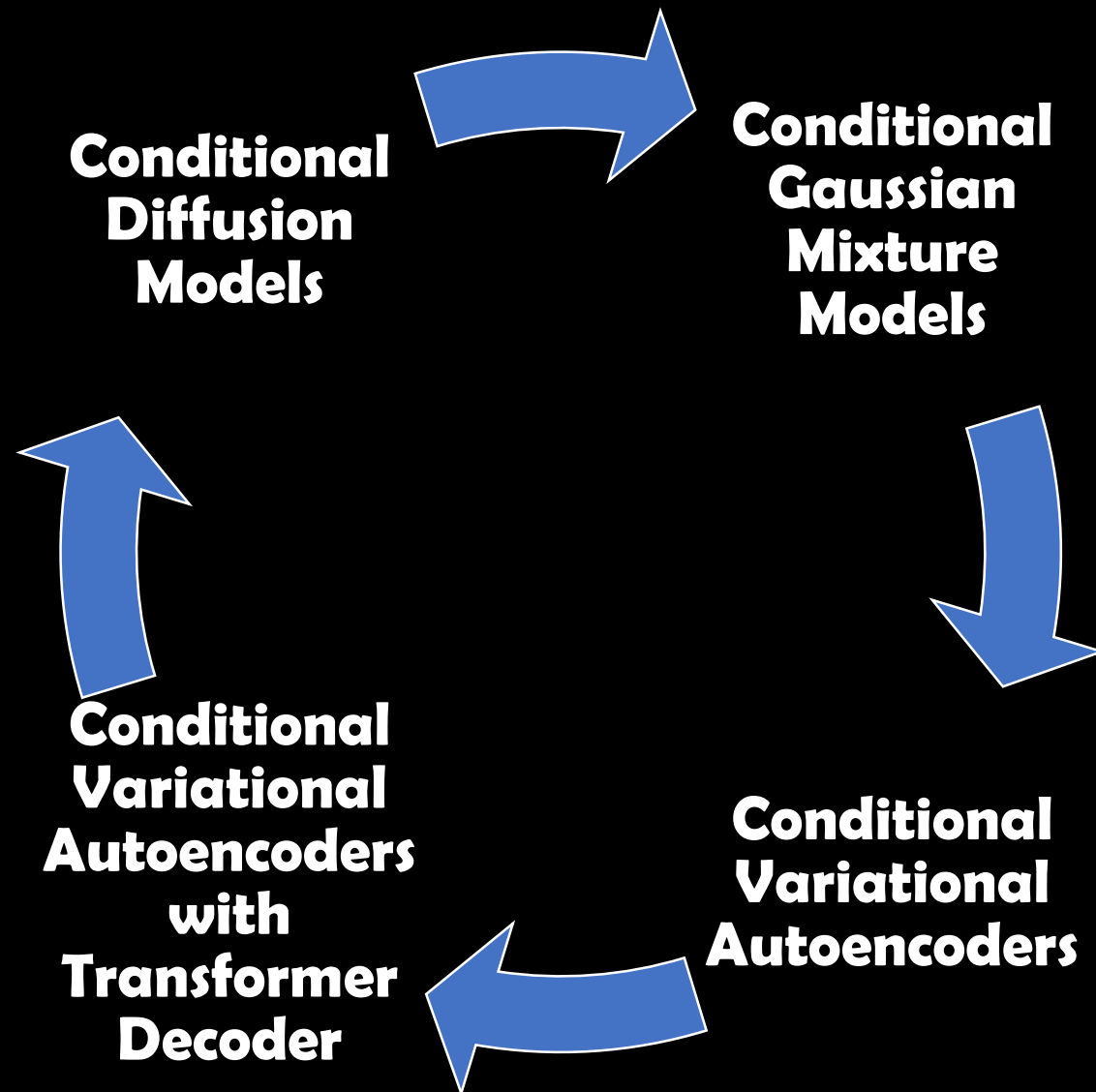
2 Datasets to test

**Encoding
Categorical
Variables**

**4 Conditional
Generative
Models**

Evaluation

Synthetic Data with Conditional Generative Models



Evaluation

**Visualization
Comparison**

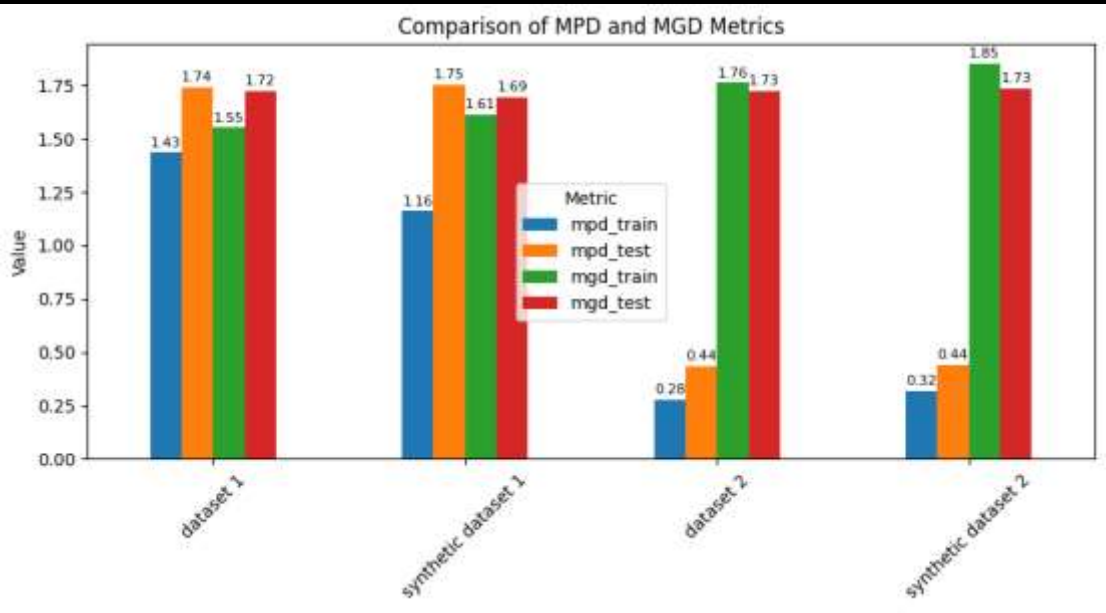
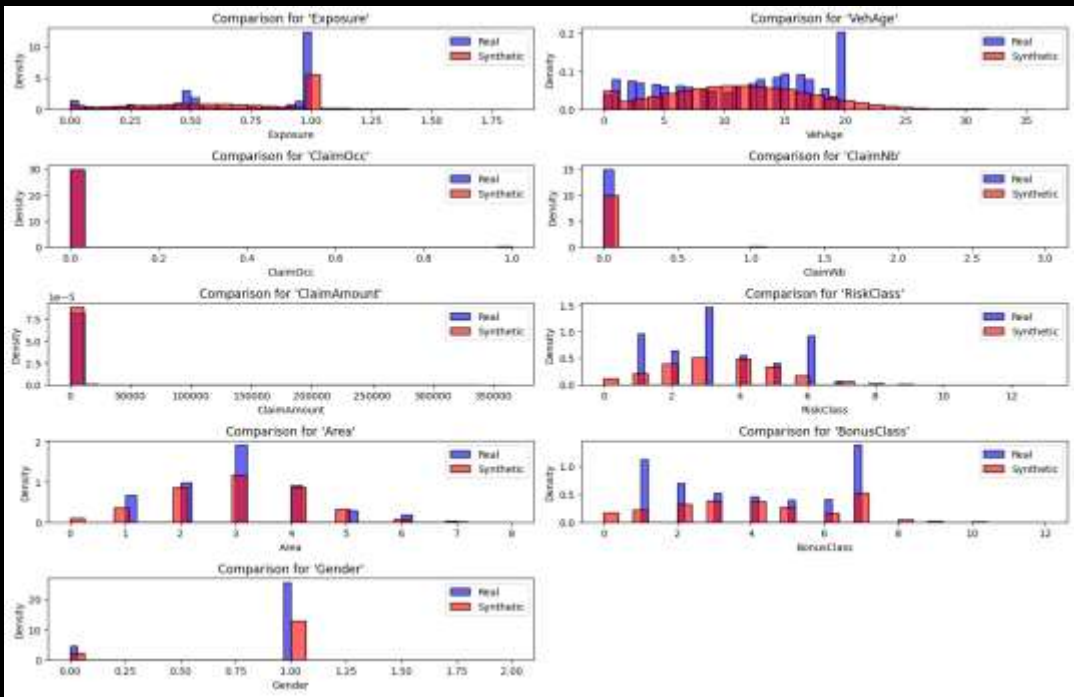
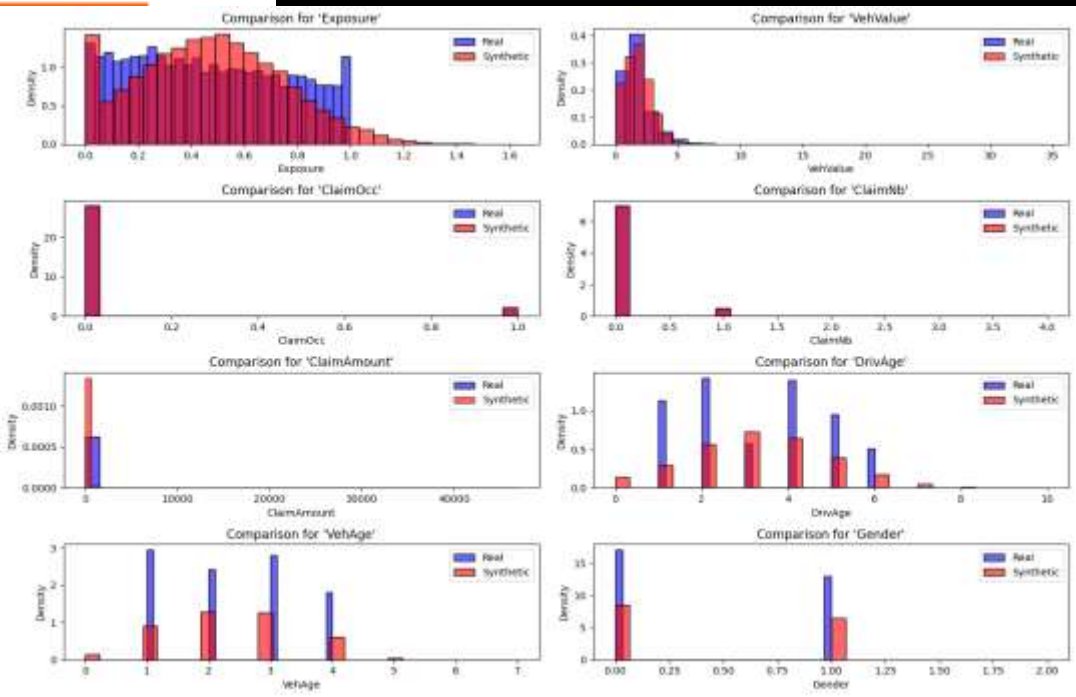
**Kolmogorov
-Smirnov
Test**

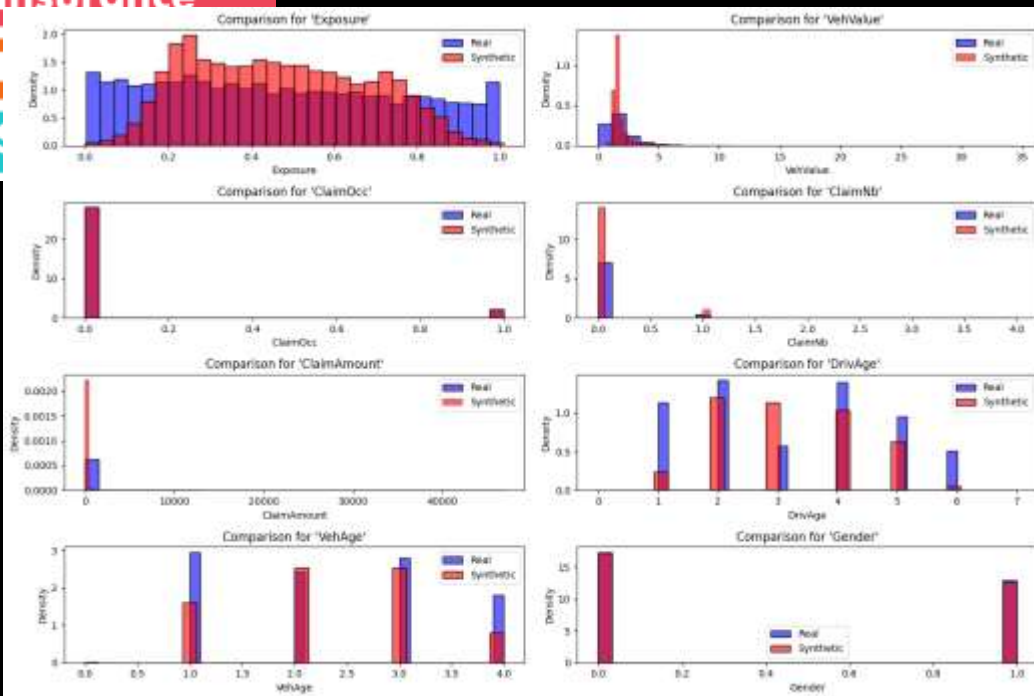
**PCA/UMAP
Analysis**

**GLM Models
Comparison**

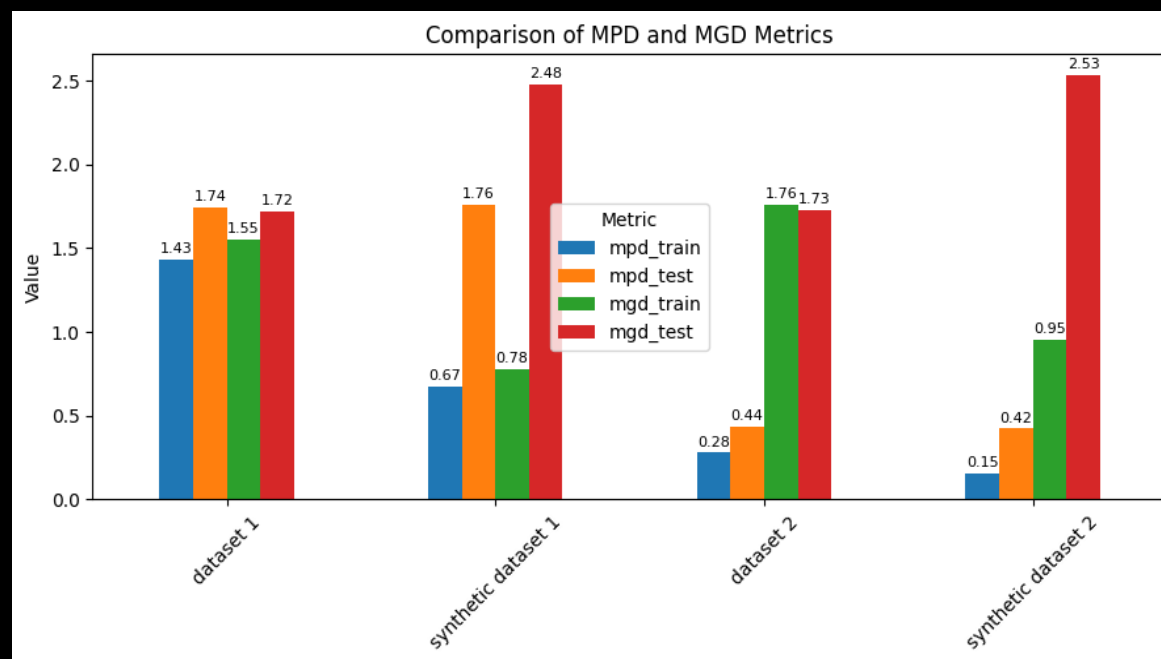
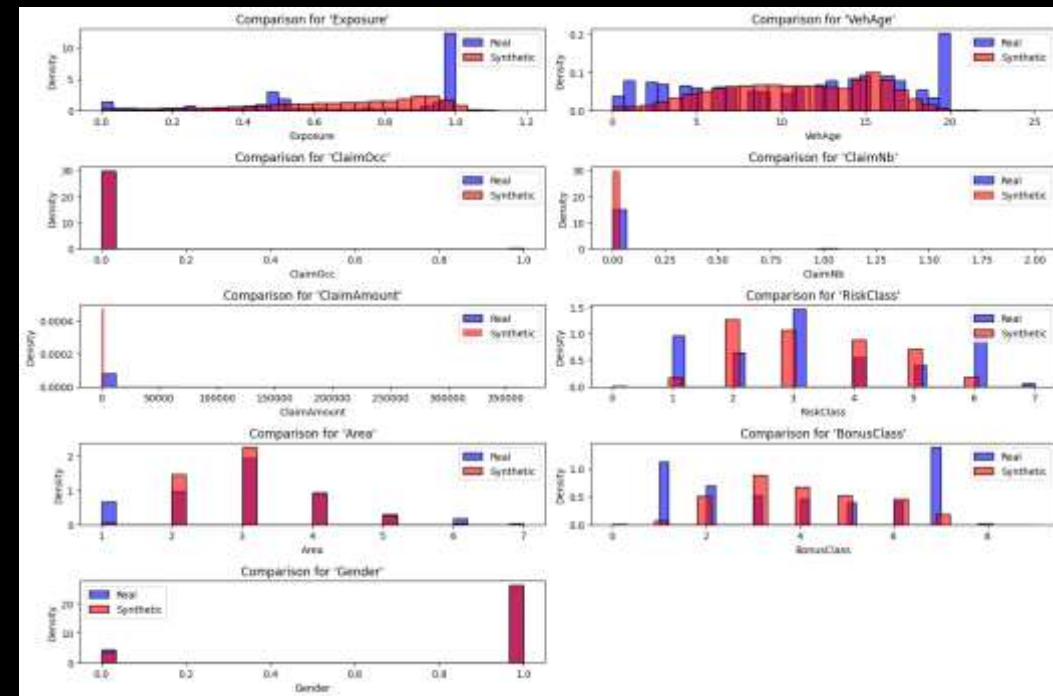
**Feature
Importance
Comparison**

CGMM Results

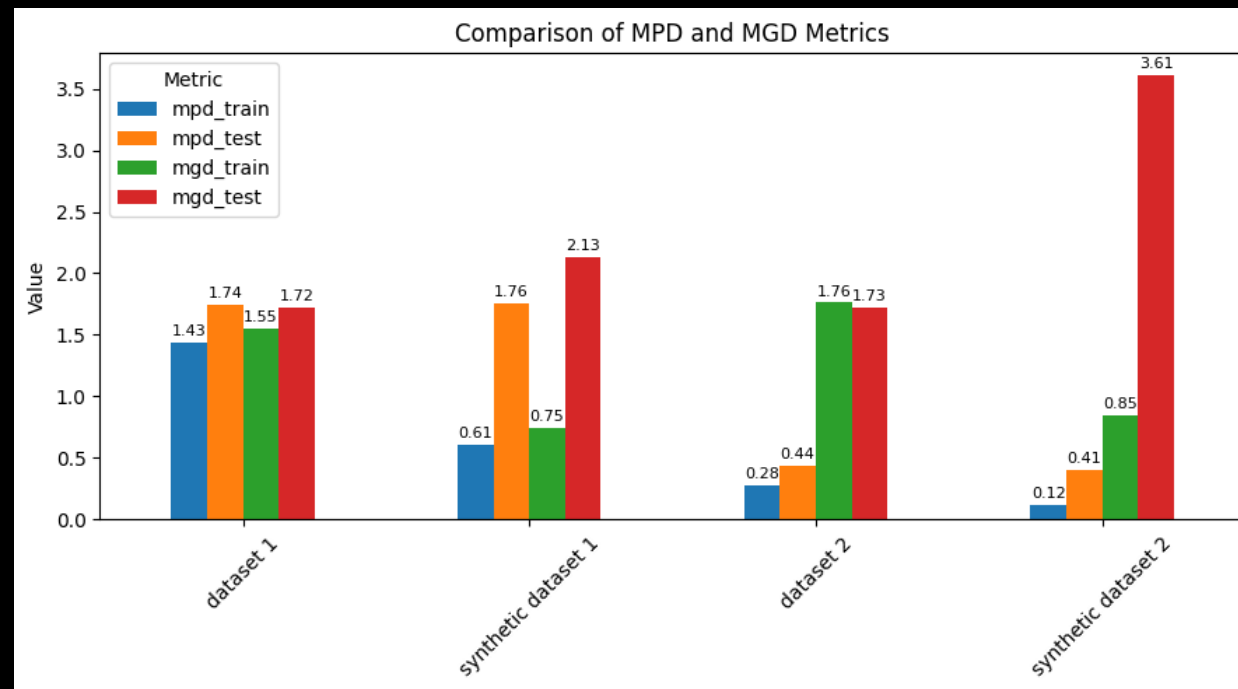
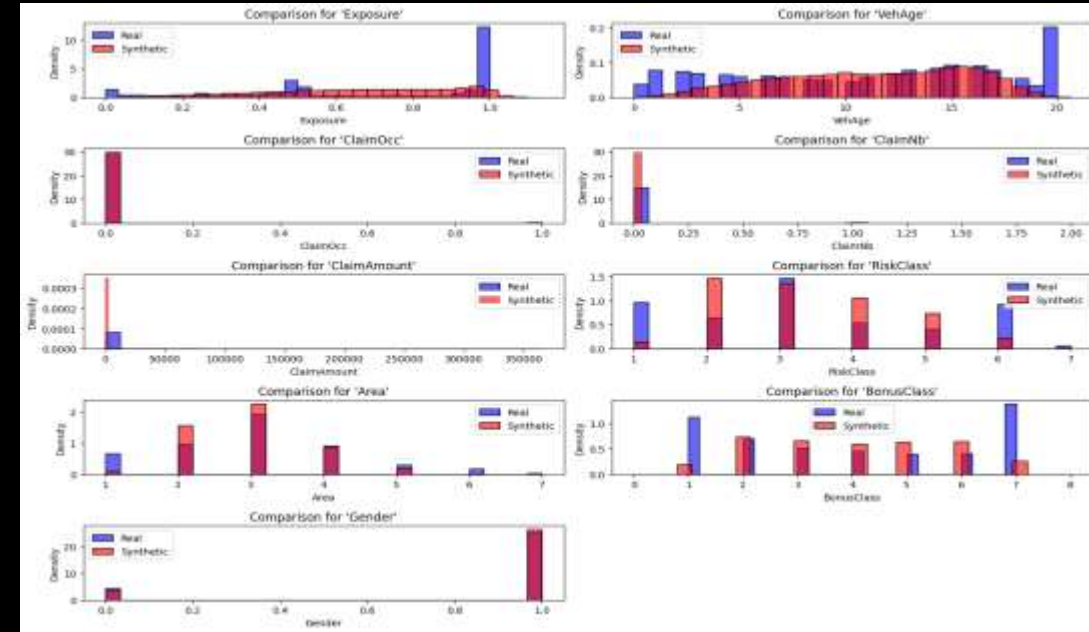
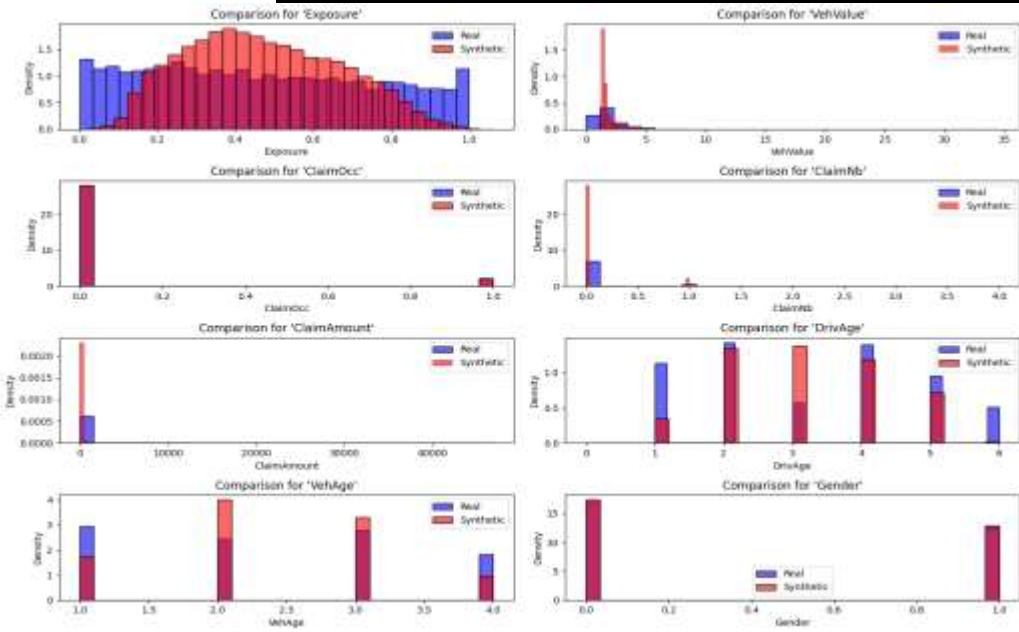




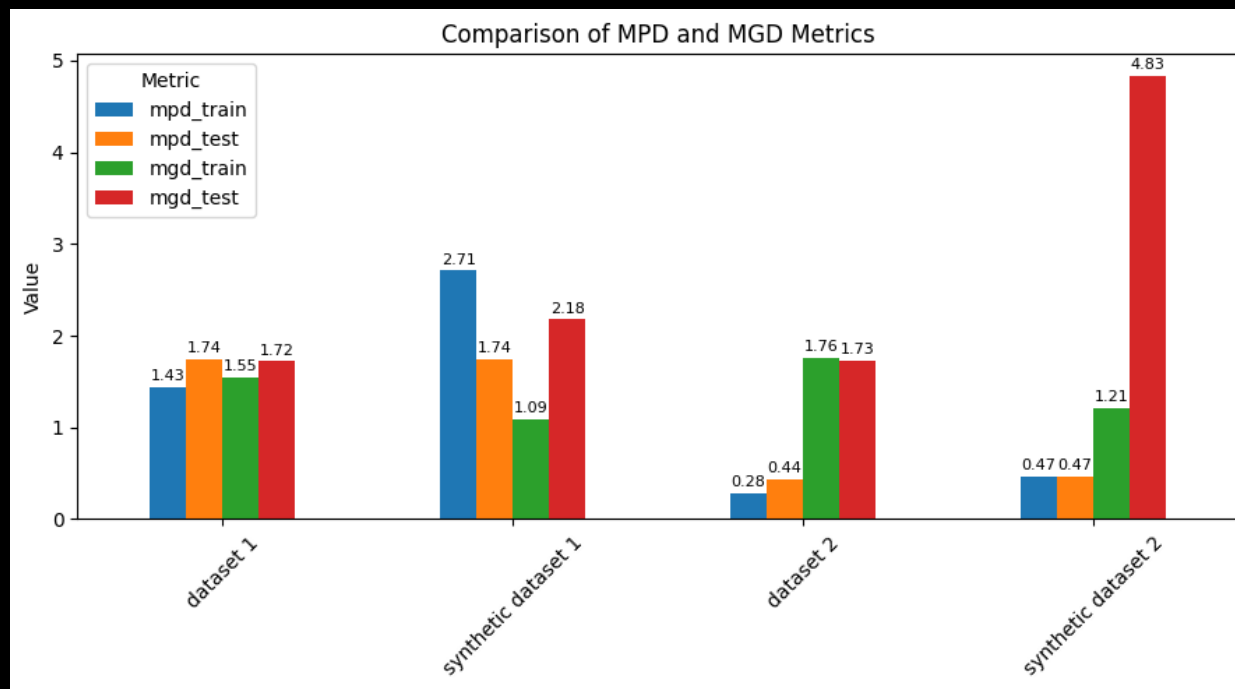
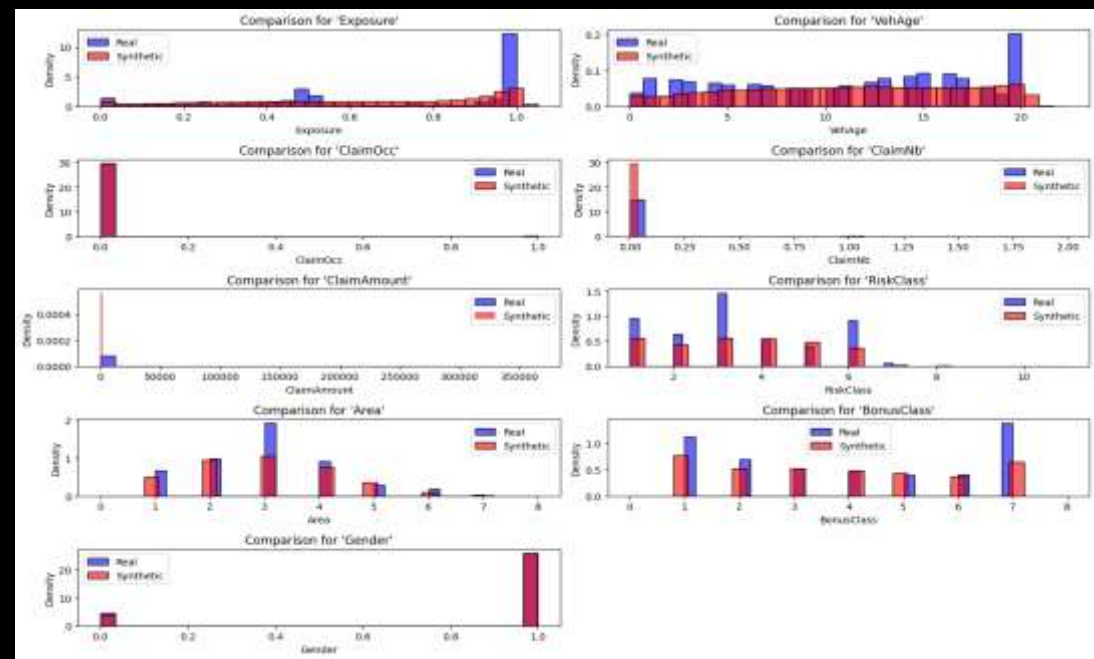
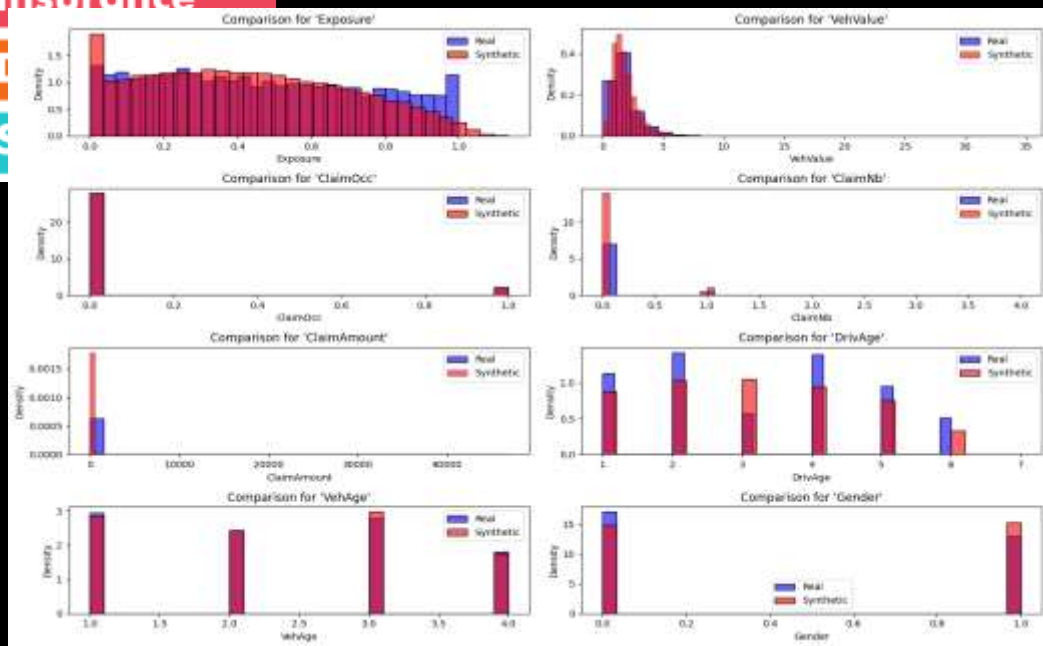
CVAE Results



CTVAE Results



CDF Results



Conclusions

- ✓ **The CGMM benchmark remains a good solution to follow for simplicity.**
- ✓ **Data from Deep Learning Generative Models requires architecture expert judgment.**
- ✓ **GANs were not used because they did not show a valuable output; they can be a better solution for classification tasks.**
- ✓ **LLMs were dropped as a solution because they can be computationally expensive and are prone to memorise a consistent portion of the dataset despite understanding patterns.**
- ✓ **Interesting solution is the personalisation of the CVAE with the Transformer Decoder.**
- ✓ **The next steps involve masking sensitive data and reducing the portion of the training set used to train the generative models.**

References

- Ian Goodfellow, Yoshua Bengio, Aaron Courville, 2016, *Deep Learning*, MIT Press.
- Mario V. Wuthrich, Ronald Richman, Benjamin Avanzi, Mathias Lindholm, Michael Mayer, Jürg Schelldorfer, Salvatore Scognamiglio, 2025, *AI Tools for Actuaries*, SSRN.
- David Foster, 2023, *Generative Deep Learning, 2nd Edition*, O'Reilly.
- Jake VanderPlas, 2016, *Python Data Science Handbook*, O'Reilly.
- Jamotton, Charlotte ; Hainaut, Donatien, 2023, *Variational autoencoder for synthetic insurance data*, ISBA.
- Data: [dutangc/CASdatasets: Datasets for the book Computational Actuarial Science with R](#)
- Github Repository: [claudio1975/Generative_Modelling](#)

?

?

?

?

?

?

Thank you

Keep in touch:

- [Linkedin](#)
- [Newsletter](#)
- [Medium](#)
- [Website](#)