<https://medium.com/data-science/evaluating-synthetic-data-the-million-dollar-question-a54701d1b621>

Absolutely! Here's a clean, professional version of the table you can directly adopt into your report. I've added a descriptive title and clarified the column headers to make it more publication-ready:

**Performance Comparison of Synthetic Data Generation Methods on the Wisconsin Breast Cancer Dataset**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Avf** | **Feature Correlation** | **Utility Score** | **Similarity Score** |
| **CopulaGAN** | 0.957 | 0.887 | **0.860** | 0.066 |
| **CTGAN** | 0.957 | 0.893 | **0.877** | 0.334 |
| **GaussianCopula** | 0.957 | 0.950 | **0.940** | 0.985 |
| **synthpop** | 0.957 | 0.946 | **0.939** | 0.990 |
| **TVAE** | 0.957 | 0.946 | **0.946** | 0.990 |
| **UNCRi** | 0.957 | 0.946 | **0.941** | 0.988 |

🟩 *Note: Rows shaded in green in the original table (GaussianCopula, synthpop, TVAE, UNCRi) indicate superior performance across multiple metrics.*

Let me know if you'd like help writing the analysis or discussion section to go with this table!

The cross-set similarity between real and synthetic data shows a maximum of 0.9346 and an average of 0.8215. This indicates that the synthetic data closely resembles real data while maintaining sufficient diversity. The high average similarity demonstrates that the synthetic samples effectively capture the underlying distribution of the real data, making them suitable for downstream tasks, while the gap between the maximum and average values ensures that the model is not simply replicating individual real samples. Overall, these results suggest a realistic yet non-identical synthetic dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Fidelity | | | Utility | Privacy |
| KS Metric (Average) | Intraset similarity  (real) | Intra-set Similarity (Synthetic) | TSTR | Cross-set Similarity (Real vs Synthetic) |
| CTGAN | 0.7634 | Max = 0.9809,  Avg = 0.8528 | Max = 0.9134  Avg = 0.8057 | Accuracy = 0.439  F1 Score = 0.418  AUC = 0.376 | Max = 0.9346,  Avg = 0.8215 |
| CopulaGAN | 0.7444 | Max = 0.9809,  Avg = 0.8528 | Max = 0.9240  Avg = 0.7946 | Accuracy = 0.772  F1 Score = 0.729  AUC = 0.884 | Max = 0.9214,  Avg = 0.8068 |
| Gaussian Copula | 0.9379 | Max = 0.9809,  Avg = 0.8528 | Max = 0.9731  Avg = 0.8109 | Accuracy = 0.895  F1 Score = 0.838  AUC = 0.965 | Max = 0.9772,  Avg = 0.8486 |
| TVAE | 0.9081 | Max = 0.9809,  Avg = 0.8528 | Max = 0.9501  Avg = 0.8147 | Accuracy = 0.965  F1 Score = 0.956  AUC = 0.998 | Max = 0.9690,  Avg = 0.8581 |