

Synthetic Financial Data:

An Innovative Approach to Data Scarcity and Limited Data Availability





Content

| Introduction | 2 |
|--|----|
| What is Synthetic Financial Data? | 2 |
| How is Synthetic Data Generated? | 3 |
| Beyond Data Privacy: The Benefits of Synthetic Financial Data | 4 |
| From Stress Testing to Fraud Detection: How Synthetic Financial Data is Changing Finance | 5 |
| Deep Dive Into a Recent Use Case: Synthesizing a Finance Relational Database | 7 |
| Best Practices and Considerations for Using Synthetic Financial Data | 10 |
| The Future of Data-Driven Financial Services | 11 |



Introduction

In finance, data is king. The more data you have, the more insights you can gain and the better decisions you can make. But what if you don't have enough data, or the data you have is biased, incomplete, or outdated? That's where synthetic financial data comes as a game-changer.

Synthetic data is artificial data generated by computer algorithms to mimic the patterns and characteristics of real-world data. Think of it as a "digital twin" of your data but with more flexibility and diversity. By using synthetic data, you can supplement or replace your real-world data and get more accurate and comprehensive results. And that's not just a theory.

Synthetic data has already proven valuable in various finance domains, from asset management to insurance, banking, and fintech. In this comprehensive guide, we will explore the concept of synthetic financial data and its immense potential to address the problem of limited data availability. By the end, you'll understand how synthetic data can help you make better-informed decisions and achieve financial success.

What is Synthetic Financial Data?

At its core, synthetic data is a computer-generated representation of real-world financial data. But unlike real-world data, which is collected from various sources such as stock exchanges, regulatory filings, or surveys, synthetic data is created from scratch based on predefined rules or statistical models. This gives it some distinct advantages over traditional data sources, such as flexibility, scalability, and privacy.

Synthetic financial data aims to provide additional data points similar to real-world data, allowing for more comprehensive analysis and modeling. It supplements existing data sets or creates new ones, enabling a more detailed look at financial trends and patterns.

Gartner

Forbes

<u>Gartner</u> named "Synthetic Data" and "Differential Privacy" as one of its Top Strategic Technology Trends and estimates that 60% of large enterprises will be leveraging one or more of these techniques by 2025.

Forbes named 'Synthetic Data' as one of The 5 Biggest Data Science Trends in the previous year, further highlighting its growing importance.

Since synthetic data isn't considered "Personal Identifiable Information (PII)," according to privacy regulations such as GDPR, all the data can be safely used and collaborated on without worrying about privacy breaches or compliance.





According to the <u>European Data Protection Supervisor (EDPS</u>), "Synthetic data is a technical solution to a legal problem," enhancing technology privacy, mitigating bias, and democratizing access to data. In their regularly published <u>TechSonar</u> report on emerging technologies, EDPS mentions synthetic data as one of the most promising technologies worth monitoring.

With growing cybersecurity concerns, money laundering, increased legislative pressure, and restricted access to transaction data, financial institutions face false positive rates, increased costs, and delays in lending decisions. Synthetic data addresses these challenges by providing financial institutions with a realistic, secure, and compliant alternative to their original data.

How is Synthetic Data Generated?

Synthetic data is created artificially using Generative AI techniques, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), in combination with Differential Privacy. With these techniques, the Syntheticus platform creates synthetic data that realistically mimics the original data while preserving privacy.



Once the original data is collected, the platform orchestrates the proprietary machine learning algorithms to synthesize, secure, validate, augment, and enrich it. This results in synthetic data that looks and behaves like the original one without revealing PII or risking compliance issues.



Beyond Data Privacy: The Benefits of Synthetic Financial Data

When you think of synthetic financial data, what is the first thing that comes to mind? Do you think of it mainly as a fix for data privacy issues, allowing you to share data without compromising sensitive information?

While data privacy is paramount, synthetic financial data is much more than a privacy shield. Generating synthetic data for financial institutions opens up a world of possibilities and improvements to your machine learning (ML) processes and model development. And many of those have nothing to do with privacy or security.

Gartner

<u>According to Gartner</u>, 80% of heritage financial services firms will go out of business, become commoditized or exist only formally but not compete effectively by 2030. Banks that remain in business will need a strategy to stay ahead of the curve, and generating, using, and sharing synthetic financial data is the key to achieving this goal.

Here are some of the main benefits synthetic financial data will bring to your organization:

Improved Data Quality and Diversity

Synthetic data is generated to represent a wide range of scenarios and events, allowing you to create more diverse and representative datasets than those available from traditional sources. This diversity improves the accuracy of your models and enables better predictions and risk assessments.

Enhanced Scalability

Traditional financial data is often limited in scope and volume, making it difficult to scale up models and predictions. With synthetic data, however, you're able to generate as much data as you need to support your ML algorithms, allowing you to expand your operations without constraints.

Improved Risk Management

Synthetic data simulates different risk scenarios, allowing you to test and refine your risk management strategies before implementing them in the real world. By running simulations on synthetic data, you will better understand and anticipate risk and minimize losses while reducing the time and resources needed to develop your models.



Enhanced Collaboration and Knowledge Sharing

Due to its privacy-preserving capabilities, synthetic financial data is easily shared and distributed within and between organizations, enabling better collaboration between departments and teams. By collaborating on synthetic data, you will build on your organization's collective knowledge and expertise, improving the quality of your models and

Improved Regulatory Compliance

Due to its privacy-preserving capabilities, synthetic financial data is easily shared and distributed within and between organizations, enabling better collaboration between departments and teams. By collaborating on synthetic data, you will build on your organization's collective knowledge and expertise, improving the quality of your models and

Better Innovation and Experimentation

Synthetic financial data is your secret weapon for innovation and experimentation. It allows you to prototype new financial products, services, and business models without risking real-world assets. This means you experiment more freely and get creative with your ideas, knowing you can test and validate them in a safe and time-effective way.

From Stress Testing to Fraud Detection: How Synthetic Financial Data is Changing Finance

While digital transformation has become a priority for banks, on that route, they are confronted with a series of challenges, such as privacy regulations, workforce training, new processes, and risks associated with using real-world data. Perhaps the biggest bottleneck is the legacy banking systems' inability to accommodate the needs of modern financial institutions, which require greater flexibility and scalability.

Synthetic financial data provides a solution to these challenges and offers a wide range of use cases across finance domains like asset management, insurance, banking, and fintech. Let's dive deeper into a few practical use cases: Data sharing Credit/lending risk **AI/ML development**

Rare events/ Software/ edge cases system testing

Fraud/financial crime Ensuring fairness/ removing bias ESG Financial inclusion Test data Automation Data ethics policy Personalisation

Stress Testing and Scenario Analysis

Stress testing and scenario analysis involve creating hypothetical scenarios and simulating how a portfolio or financial instrument would perform under those conditions. With synthetic data, institutions generate a diverse range of scenarios that are difficult or impossible to obtain from real-world data.

Credit Scoring and Loan Origination

Synthetic data allows financial institutions to generate digital twins of customers and simulate their credit scores, enabling lenders to make more accurate loan origination decisions, better understand the creditworthiness of their clients, and improve risk management.

Fraud Detection and Risk Management

By generating synthetic data that mimics real-world fraud patterns, institutions improve their fraud detection models and reduce the number of false positives. They can simulate different risk scenarios to fine-tune their risk management strategies and ensure they are operating at optimal levels.

Portfolio Optimization

With synthetic data, institutions generate vast amounts of data on different investment scenarios and evaluate the performance of various portfolios. This helps them identify the most profitable and efficient portfolios, leading to better returns for their clients.

Anti-Money Laundering

By generating large sets of synthetic transactions, organizations train and test their anti-money laundering (AML) models to detect suspicious activity more accurately. Patterns of criminal activity can still be seen in the synthetic data, allowing organizations to hone their AML models and stay ahead of new criminal tactics.

Data Bias Reduction

Even though synthetic data is not immune to bias, it helps reduce the risk of data being used to perpetuate prejudices. By creating datasets more representative of the entire population, including underrepresented groups, institutions will ensure their models are not based on faulty data.



<u>The recent partnership</u> between <u>Basinghall Analytics</u> and <u>Syntheticus</u>, announced in March 2023, is a great example of how synthetic financial data is used for stress testing and scenario-based financial planning. The joint proposition of Basinghall's proprietary risk tools with Syntheticus' privacy-preserving synthetic data capabilities enables financial institutions to extract maximum insights from sensitive data without compromising compliance with data protection and privacy regulations. This partnership provides a game-changing solution for financial institutions to build new models, validate and monitor them, and perform stress testing and scenario-based financial planning.

Deep Dive Into a Recent Use Case: Synthesizing a Finance Relational Database

At Syntheticus, we love tackling challenging projects that push the boundaries of what's possible with data and take pride in our ability to solve complex data challenges for our clients. Recently, we had the opportunity to partner with a client in the finance industry who needed a solution for their large and complex financial data.

They were struggling to manage >8 million rows of data spread across a dozen tables connected through referential integrity.

*Given the client's privacy, we're not disclosing the actual underlying data but only a public dataset for the sake of showcasing the results. See below for the entity relationship diagram.





To tackle the challenge, we leveraged our expertise in data modeling and database design to optimize the tables for performance and scalability while maintaining referential integrity. Using the <u>Syntheticus Hub</u>, we seamlessly connected to the client's database. The platform's out-of-the-box integrations support various services such as AWS, Google, Azure, Snowflake, Databricks, and integration with enterprise databases such as Oracle, MySQL, MongoDB, PostgreSQL, MariaDB, and many others.



After only a few minutes of reviewing the client's underlying data model, the platform generated a synthetic version of the entire database with a dozen tables, each with a specific set of fields to accommodate the client's requirements, and over 8 million records. We take quality seriously, and so the platform conducted a thorough Quality Assurance (QA) report to ensure that the database met the highest standards. The QA report highlighted the main metrics, including data accuracy, completeness, privacy, and consistency, for each single table as well as for the database.

Here are some of the main metrics found within the QA report:





Data Diagnostics:

Column Boundaries (Average Score=1.0)



Data Quality: Column Pair Trends (Average Score=0.79)



Numerical Correlation (Real Data)



account_id amount balance account

Numerical Correlation (Synthetic Data)







Data Diagnostics: Column Coverage (Average Score=0.93)



Metric

RangeCoverage CategoryCoverage

After only a few minutes of reviewing the client's underlying data model, the platform generated a synthetic version of the entire database with a dozen tables, each with a specific set of fields to accommodate the client's requirements, and over 8 million records. We take quality seriously, and so the platform conducted a thorough Quality Assurance (QA) report to ensure that the database met the highest standards. The QA report highlighted the main metrics, including data accuracy, completeness, privacy, and consistency, for each single table as well as for the database.

Best Practices and Considerations for Using Synthetic Financial Data

If you've decided to leverage synthetic financial data for your institution, there are a few things to keep in mind to ensure you're using it effectively and responsibly.

First, it's crucial to select appropriate models that are relevant to your use case. Different models have different strengths and weaknesses, and you should choose one that fits your specific needs. For example, if you're looking to generate data for credit risk analysis, you might choose a model that accurately simulates default probabilities and other credit risk metrics.

But selecting the suitable model is just the first step. You need to validate the quality and reliability of the synthetic data you generate. This means running tests to ensure your data is accurate, unbiased, and representative of real-world financial data. Only then can you have confidence in the insights you extract from it.

10



Of course, ethical and legal considerations also come into play when working with synthetic financial data. Data privacy is a big one, as synthetic data is sometimes derived from real-world data containing personally identifiable information (PII). Intellectual property is another concern, as some synthetic data generation techniques may involve proprietary algorithms or models.

Another important consideration when working with synthetic financial data is ensuring compliance with regulatory requirements. Depending on your industry and location, specific regulations or guidelines may dictate how to generate and use synthetic data. Make sure to do your research and consult with legal experts to ensure you're following all applicable regulations and guidelines.

Finally, the lack of clear standards on privacy can create uncertainty around how to best protect sensitive information in synthetic datasets. <u>Syntheticus</u> recognizes this challenge and actively participates in the IEEE Standards Association, which has set up an IC Expert Group to set a standard for structured privacy-preserving synthetic data.

Transparency, explainability, and accountability are essential principles to keep in mind when using synthetic financial data. You need to be able to explain how the data was generated, what assumptions were made, and what the limitations of the data are. This is important both for internal decision-making and for external stakeholders who may be relying on it.

The Future of Data-Driven Financial Services

The financial services industry is undergoing a significant transformation driven by advances in technology and the growing importance of data. The future of banking will be shaped by the ability to use data-driven decision-making and successful digital transformation. One of the biggest challenges in this context is handling bank-specific and personal data and its processing by artificial intelligence.

Synthetic financial data is likely to become even more important for institutions in the years to come. The increased availability of synthetic data will allow institutions to leverage AI and machine learning to make more informed decisions while complying with data privacy and regulatory requirements. As new technologies like blockchain and distributed ledgers become widespread, the potential applications for synthetic data will only grow. From risk management and compliance to fraud prevention and customer segmentation, synthetic financial data will become an essential tool for banks and other financial institutions.





However, this future has challenges, including data privacy concerns, cybersecurity risks, and regulatory compliance. As the industry continues to evolve and new use cases emerge, financial services companies must balance the opportunities presented by data-driven insights with the potential risks to ensure long-term success. Synthetic financial data will be essential to this equation and will help bridge the gap between traditional services and digital banking.

As synthetic financial data solutions continue to evolve, the possibilities are endless. With careful planning, ethical considerations, and a clear understanding of its potential applications, financial institutions capitalize on this opportunity and create a data-driven future in banking.



Ready to explore the power of synthetic data for your financial institution?

Sign up for a free demo

and learn how Syntheticus advances your data-driven projects while protecting customer privacy.

syntheticus.ai