



Syntheticus



Unlocking New Possibilities in Banking and Finance with Synthetic Data

The Story of SIX



Content

Meet SIX	2
The Challenges	2
The Synthetic Data Approach	3
How Synthetic Data Benefits SIX and Other Banking and Financial Institutions?	4
The Proof of Concept	5
The Results	8

Meet

SIX is a leading global financial company providing services related to transaction security, financial information processing, payment transactions, and building a digital infrastructure. SIX has stood for innovation and stability in the global financial markets since 2008 and operates the infrastructure for the financial centers in Switzerland and Spain, thus ensuring the flow of information and money between financial market players.

SIX offers exchange services, financial information, and banking services, aiming to increase efficiency, quality, and innovative capacity along the entire value chain. The company is owned by around 120 national and international financial institutions. SIX's close relationship with them guarantees the financial infrastructure and processes stability, proximity to clients' evolving business needs, and competitive prices.

The Challenge

To unlock the full potential of their data, derive data-driven insights, and create business value, SIX employees need constant access to a wide range of data from every corner of the company.

However, data scientists at SIX struggle to have constant access to the most up-to-date original data due to complex and evolving regulations that limit access to specific datasets, thus hindering their ability to unlock its full potential. This results in an inability to deliver data-driven insights on time and lost opportunities for the company.

The barriers that lead to the challenges mentioned before are, amongst others:

Legal & Compliance

Data is subject to complex and evolving legislation, making it difficult to unlock and use.

Privacy regulations

Regulations often impose strict limits on the ability to share data and the extent to which it can be used.

IT silos

Data from different business units is often stored in disconnected silos, making it difficult to access.

Finally, their analyses are performed locally and not on the cloud. This further limits the speed and scalability of data-driven activities at SIX; the computation performance is limited, resulting in cumbersome and inefficient data analytics.

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.

The Synthetic Data Approach

To overcome these challenges, **SIX joined forces with Syntheticus**, a powerful synthetic data platform that leverages advanced Privacy-Enhancing Technologies, such as Generative AI and Differential Privacy, to generate realistic and accurate synthetic data.

About Synthetic Data

Synthetic data is artificially generated data that mimics the original data while protecting the privacy of individuals. One of its many use cases includes helping banks and financial institutions overcome compliance concerns, decrease false positive rates, and generate new revenue streams by providing access to AI-generated datasets with almost identical statistical properties as their original data.

 Gartner

[Gartner](#) 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

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

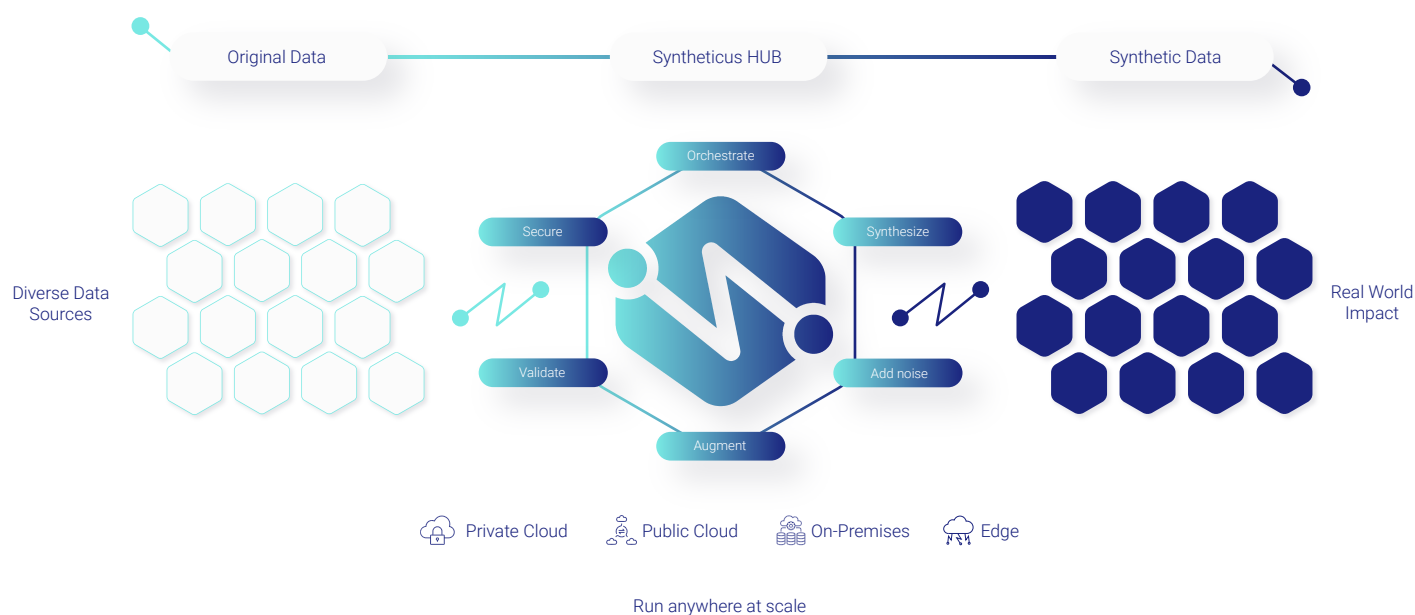
Since synthetic data isn't considered "Personal Identifiable Information (PII)," according to privacy regulations such as GDPR, all the data can be safely used and collaborated on without worrying about privacy breaches or compliance.



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

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 realistic data that 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.

How Synthetic Data Benefits SIX and Other Banking and Financial Institutions?

Synthetic data provides banks and financial institutions with ways to overcome data compliance and usage limitations while still providing data-driven insights.

Financial institutions use synthetic data to:

● **Improve compliance with data privacy regulations and cybersecurity requirements**

With strict data privacy laws in place, a rise in data breaches, and growing cybersecurity concerns, synthetic data can provide a safe and compliant alternative to using original data. It reduces compliance costs and overcomes usage limitations by providing secure, anonymized datasets for modeling and testing without exposing sensitive customer information.

● **Reduce the risk of fraud and lending errors with accurate predictive analytics**

In addition to compliance and privacy concerns, banks and financial institutions struggle with accurately predicting risk, fraud, and money laundering. As these events are rare, it can be difficult to build accurate models based on real-world data. Synthetic data helps overcome these challenges by providing a more robust dataset for testing and refining predictive models, identifying and analyzing trends, assessing risk, and reducing the likelihood of money laundering and fraud.

● Improve access to critical data for better decision-making and revenue opportunities

Banks and financial institutions depend on accurate data to make critical business decisions. Silos and compliance barriers often prevent them from accessing the data they need to gain valuable insights, resulting in missed opportunities and revenue loss. Synthetic data provides a cost-effective and secure solution for accessing critical data, enabling banks and financial institutions to increase their competitive edge and generate new revenue streams.

● Streamline data sharing for software development and testing

Digital products with personalized services, such as mobile banking apps or customer relationship management software, require data sharing with third parties. Synthetic data is the perfect replacement for original data, enabling banks and financial institutions to safely share data needed for software development and testing without compromising security or risking compliance issues.

The Proof of Concept

Syntheticus and SIX

Using synthetic data generated by the Syntheticus platform, SIX's data analytics team collaborates and computes at scale while being 100% compliant with applicable regulations. Additionally, synthetic data enables cloud infrastructure usage and ensures data security while allowing the execution of complex analysis and predictive models. This results in speeded-up development cycles, a higher degree of scalability, increased overall data literacy, and increased efficiency throughout the organization.

At SIX, synthetic data is used to:

1 Populate data warehouses or sandboxes

2 Run all sorts of analytics like Machine or Deep Learning

3 Serve as test data for product development

4 Share internally or externally for secure collaboration

Deep Dive Into the Proof of Concept

To better understand the use cases for synthetic data and its potential impact on SIX, Syntheticus conducted a Proof of Concept (POC) as a part of project collaboration with [Constructor Learning](#) and [SIX](#).

The POC is divided into two parts: the first part highlights the original data and the generated differentially private synthetic data. In the second part, the original and synthetic data are compared and evaluated according to specific metrics.

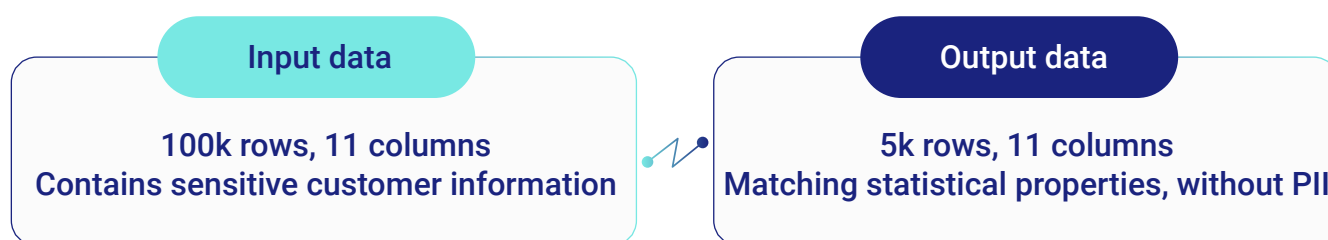
Part One: Data

The underlying financial transaction data consists of customers' personal and financial information. The data scientist's task is to thoroughly understand the customer's loan default risk to better predict future loans.

The original data consists of 100k rows and 11 columns. However, this data is sensitive and contains sensitive information such as the customer's name, address, and bank account number. As a result, the original data cannot be shared outside or within the organization. This poses a challenge for financial institutions looking to collaborate with external partners and data scientists who can't gain full access to the real-world data they need to perform their analysis.

To overcome this problem, Syntheticus was used to generate synthetic data with almost identical statistical properties as the original data but **without containing any PII**.

We truncated the original dataset to 5k rows for the sake of simplicity. The generated differentially private synthetic data consists of 5k rows and 11 columns.



Part Two: Evaluation

The evaluation is based on two different methods - **utility and privacy**.

Utility evaluation was based on comparing the models built on the original and synthetic data. We used standard classification algorithms to compare the performance of both datasets. The results show four different metrics as well as the final score:

Basic statistics

0.9974

Correlation column correlations

0.9218

Mean Correlation between synthetic and real columns

0.8903

1 - MAPE Estimator results

0.9467

Similarity Score

0.9390

The similarity score aggregates different evaluation metrics to cover all aspects of data. It is calculated by taking the mean of these different metrics. The goal is to get a single value representing a synthetic dataset's proximity to a real dataset.

Using the **privacy evaluation**, the results show two metrics:

Duplicate rows between sets (real/synthetic)	0/0
Nearest neighbor mean	2.7591
Nearest neighbor std	0.5285

Two metrics were used to assess the level of privacy. The first is a simple analysis of whether any rows in the synthetic dataset are identical to corresponding rows in the real dataset. Generally, this is not desired, and possible regularisation or checks might be required to prevent it. The second metric is the mean and standard deviation distance between each synthetic record and the most similar real record. The desired outcome is a high mean with a low standard deviation, showing that all synthetic records have a sufficiently large distance from their closest real record while still forming a valid dataset.

The Results

The results of the POC have shown that the synthetic data generated by Syntheticus is of high quality in terms of privacy and utility: it preserves the original statistical structure while protecting sensitive data. This opens up new opportunities for SIX, allowing their analytics team to leverage synthetic data for advanced analytics and collaborate in a secure and privacy-preserving manner.

Using the Syntheticus platform to seamlessly generate high-quality datasets, SIX was able to run predictive models, testing, and other analytical tasks while ensuring consistent results and reducing overall time to data.

The Syntheticus platform leverages Generative AI, Differential Privacy, and Confidential Computing to protect sensitive financial data while providing accuracy and integrity to the generated data. This enabled SIX to generate large-scale synthetic data their analytics team used for advanced analytics while preserving customer privacy and maintaining compliance with industry regulations.

Overall, the promising results of the SIX POC have shown that other banking and finance companies should leverage the Syntheticus platform to create secure and accurate synthetic data that will allow them to:



Securely collaborate and compute on synthetic data at scale to derive data-driven insights.



Stay compliant and de-risk potential privacy fines and reputation damages.



Reduce wait times and administrative burdens, leading to a positive impact on data-driven attitude.



Enable usage of robust cloud infrastructure.



Democratize data, resulting in increased overall data literacy, business value, and a competitive edge.



Syntheticus

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