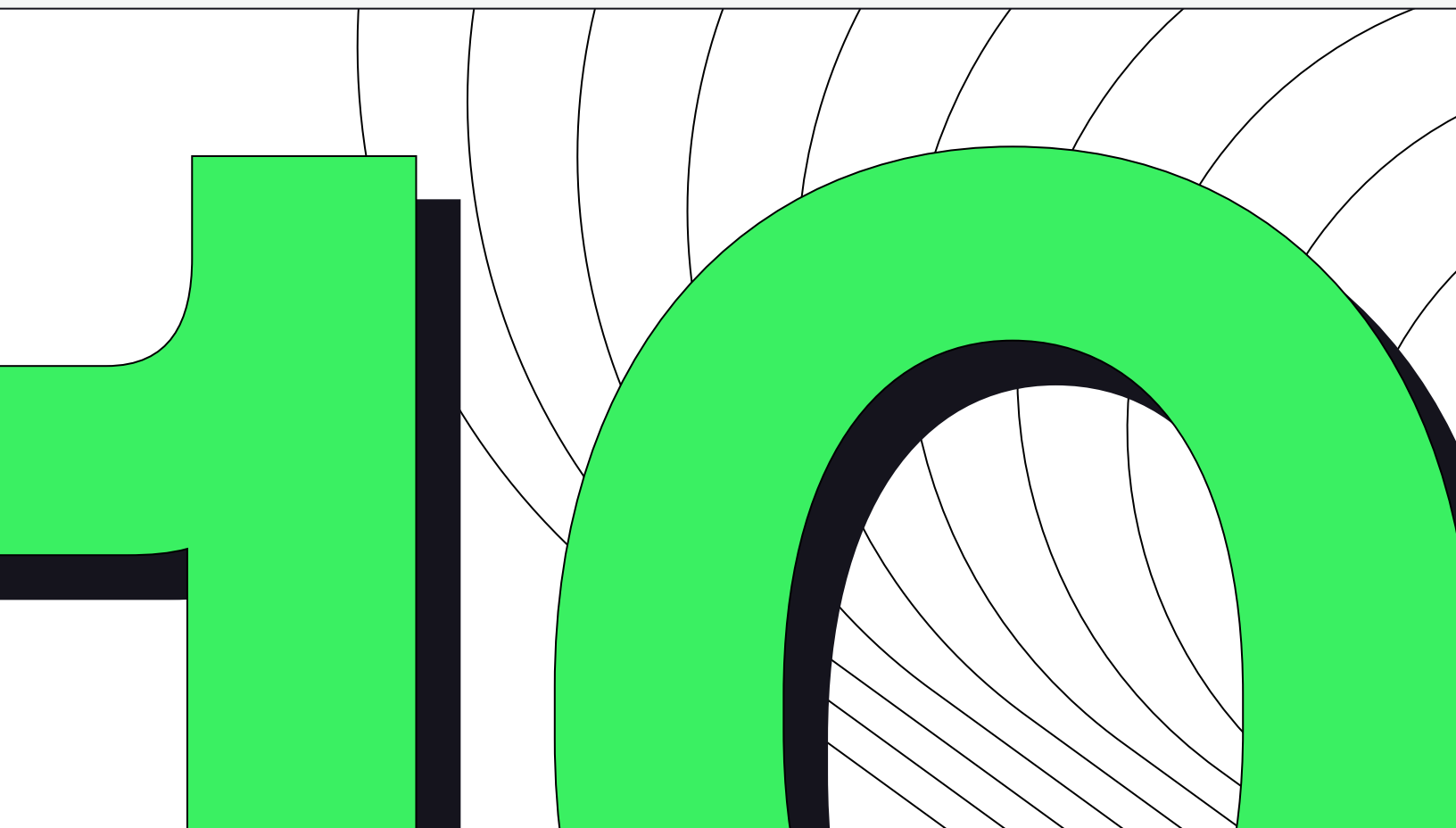




Top 10 synthetic data use cases and applications

MAXIMIZING THE UTILITY AND VALUE OF DATA



EXECUTIVE SUMMARY

The preservation of data privacy and the necessity of maximising data utility has become a high priority for many industries and organizations. Maximizing the utility and value of the data that organizations create whilst ensuring full compliance with increasingly stringent data privacy regulations means that new tools and practices must be adopted. Increasingly organizations are making use of high-quality synthetic data to extract the full value of their data and grow their competitive advantages. This brief explores 10 highly useful and emerging business applications of synthetic data and the value and opportunity that organisations are unlocking by adopting this tool.



Synthetic data for customer data analytics

Businesses are becoming increasingly reliant on market and customer analytics to underpin their growth strategies. To drive data analytics effectively, data teams need access to real-time, high-quality datasets. Traditional masking techniques such as data anonymization limits the ability to draw insights from data, by removing identifiable information from the original data. This restricts the use of data for detailed insights to improve decision making and growth. Synthetic data is used to protect sensitive customer data while providing similar statistical insights to the original data for an analysis of customer behavioural trends, to drive sales growth for businesses, while complying with regional data privacy regulations. Using synthetic data, businesses gain detailed analytical insights into customer behaviours, enabling them to accurately meet customer needs, and potentially facilitate the expansion into new markets.

Applicable Industries: Financial Services, Insurance, Healthcare, Consumer Goods, e-Commerce & Retail



Predicting customer churn using machine learning

Predicting customer churn is important for businesses because it enables preventative measures to be taken to prevent a loss in revenue. Machine-learning modes can be trained to calculate the probability of customer churn using synthetic data, to maximise data utility from sensitive, original datasets, while eliminating the risk of non-compliance with regulations such as GDPR, CCPA, LGPD and more. Synthetic data can be manipulated to balance undersampled data on customer churn, which improves the performance of the ML model. Predicting customer churn using synthetic data benefits businesses by enabling them to reduce the risk of churn, and improve customer satisfaction by tailoring products and offering proactive customer service.

Applicable Industries: **All**



Data monetization in financial services

Data monetization is becoming a pivotal contributor to profitability. And according to Deloitte, 70% of an organization's data goes unused. Using sensitive customer data presents significant privacy and non-compliance risks for banks consumers in the event of data leakage. However, customer data from banks can be used to generate statistically identical, privacy-compliant synthetic data, augmented data products to sell to companies in a range of sectors or to other banks. Despite the dataset not being original, the similarity in statistical composition gives buyers the same insights as to the original dataset. Synthetic data is preferred to anonymized data, as the collection of anonymous data limits the ability to gain insights and value from the data. Synthetic data for monetization can benefit commercial partners that share data by enabling them to develop a better understanding of customer needs and enhance insights into improving products.

Other Applicable Industries: **Insurance, Healthcare, Consumer Goods, e-Commerce & Retail, Manufacturing**



Mitigate bias in recruitment processes using synthetic data

The positive impact of diversity and inclusion is no longer debatable, but a given fact. According to Gartner, inclusive teams improve team performance by up to 30% in high-diversity environments. Research from McKinsey shows that companies with greater gender diversity in their management were 25% more likely to have above-average annual profits. The use of synthetic data in recruitment processes reduces the reliance on limited historical data by supplementing it with high-quality augmented data to eradicate the data gaps that may currently exist. With enriched and balanced data, hiring algorithms can make talent sourcing more equitable and diverse. Synthetic data can also be used to train algorithms to better identify skills which can be inferred from various roles and functions to better match and place talent to career opportunities.

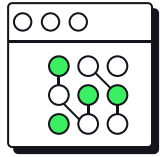
Applicable Industries: All



Driving innovation: sharing data with third parties

Third-party data sharing is an important vehicle for commercial partnerships and innovation within industries. Yet is highly restricted by compliance requirements due to its sensitive nature and the potential risk it brings. Synthetic data enables businesses to share data with third parties, to help address a sector's challenges. While it maintains the same statistical attributes as original data, synthetic data can unlock new opportunities across numerous industries in several industries. It can lead to improved overall cost management between supply chain partners, and faster innovation, resulting in lower costs for customers, and faster time to market for new products and services.

Applicable Industries: All



Software testing and development

Production data is sensitive and either impossible to use or impossibly difficult to extract for testing purposes. Synthetic data automates and speeds up the generation of higher quality datasets for software testing and development with real-time realistic data, while significantly reducing the risk of data leakage in testing environments. The use of synthetic test data eliminates the time and labour associated with setting up software testing environments. You get access to high-quality datasets at any volume, optimized to your coverage needs while maintaining full data utility without any compliance risk.

Applicable Industries: **All**



Training machine learning models to identify fraudulent activity

Millions of consumers fall victim to credit card fraud annually, and the cost is also passed onto financial institutions. In Financial Services, synthetic data is used to detect the probability of fraudulent transactions by training machine-learning models to recognise transactional anomalies. The system of prediction has to be balanced on reducing fraud as accurately as possible. The synthetic financial data generated can be augmented to create balanced datasets, and perform predictive analyses of fraudulent transactions. Improved fraud detection model performance using machine learning and synthetic data can reduce fraud operations costs while saving tens of millions of pounds for consumers.

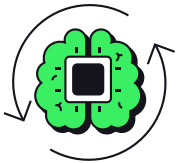
Other Applicable Industries: **Insurance**



Synthetic data generation for clinical trials

Clinical trial research is limited by the access needed to data that could compromise patient data. The volume of data and representativity for a wider population is another challenge practitioners face. Synthetic data can be used to augment skewed data, balance the inclusion of underrepresented groups, and identify broader clinical trends

from the original datasets. The use of balanced synthetic data enables high-quality hypothesis testing much faster and cost-efficient, and its compliance with privacy laws promotes collaboration between researchers and pharmaceutical companies globally, to accelerate the process of drug development that can potentially save numerous lives.



Training machine learning models to increase insurance quote conversion

There is a business opportunity for insurance companies to increase quote conversion rates by accurately identifying early on customers that are interested in purchasing policies. Trained machine-learning models can detect and rebalance skewed datasets, and generate synthetic customer datasets to accurately predict the probability of a customer accepting the product offer. This strengthens the understanding of target customers and reduces potential inaccuracies in insurance quotes. Insurance companies can benefit from an enhanced brand reputation and an increase in customer acquisitions, which can improve profitability.

Other Applicable Industries: **Healthcare, e-Commerce & Retail**



Synthetic data for pattern analysis and trends projections in e-commerce

E-commerce businesses must rapidly understand consumer patterns and behaviours, to determine and occasionally influence a consumer's buying decisions. Synthetic data can be used to enhance data analysis and draw deeper insights into those consumer patterns and future trends. Because synthetic data is immune to traditional statistical issues, such as skip patterns and non-response, it can be used to accurately uncover trends regarding customer behaviours. Access to more information enables e-commerce businesses to accurately segment customer data, and personalise marketing campaigns to target specific customers successfully. The increased effectiveness of promotional efforts can result in greater customer satisfaction and profitability.

Other Applicable Industries: **Healthcare, e-Commerce & Retail**

Synthesized delivers the first API-driven data generation platform that creates data, better than production data — in minutes. QA and ML teams can now easily create and access high-quality data in minutes with easy-to-use YAML configs.

For more information, visit synthesized.io