



Synthetic data Real impact

Unlocking Privacy-Sensitive Data with AI-Generated
Synthetic Data in Healthcare & Pharma

Wim Kees Janssen
CEO & Founder Syntho



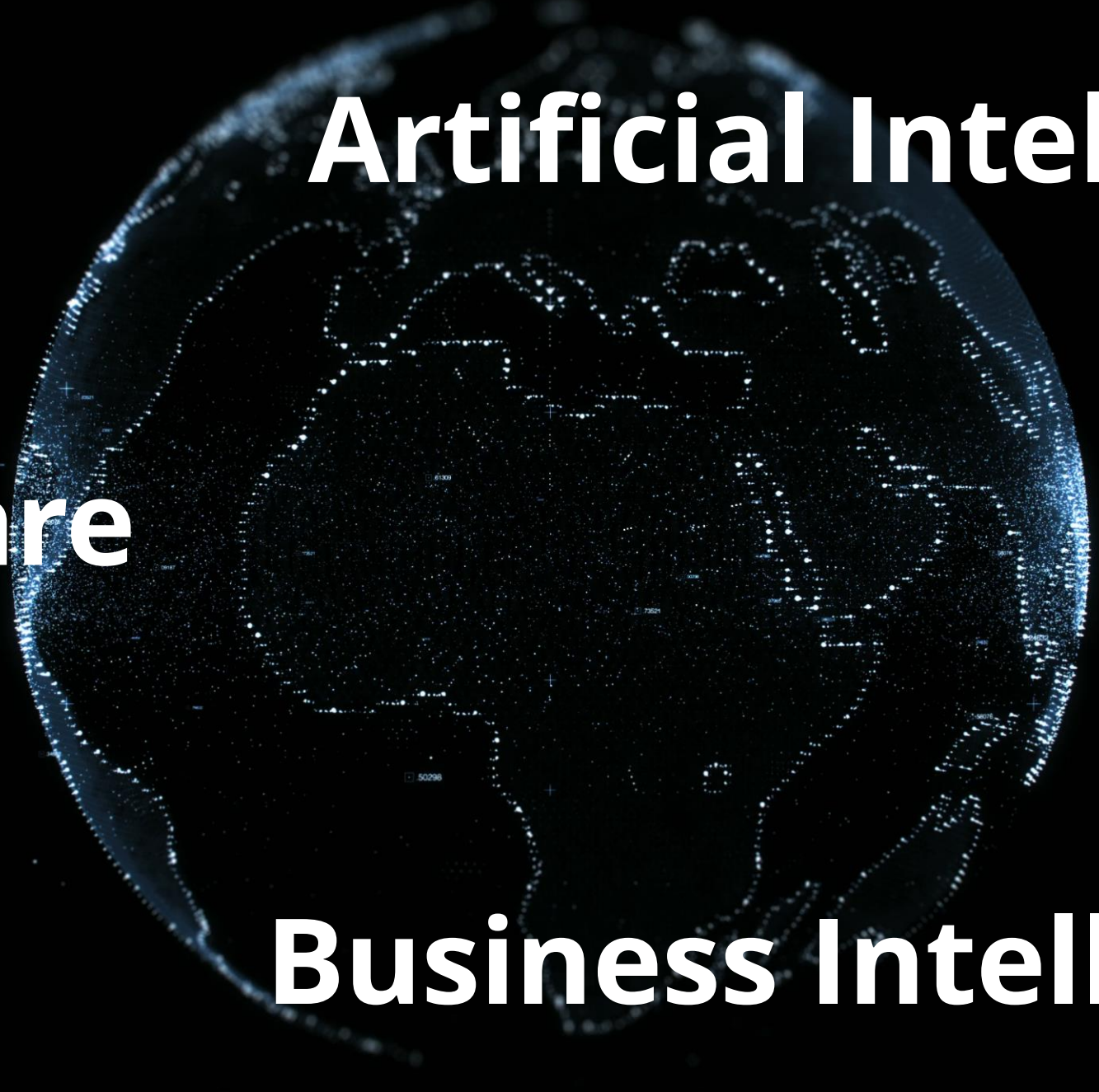
Wim Kees Janssen – Founder & CEO Syntho

- Contact: **kees@syntho.ai**
- **www.syntho.ai**



About Syntho





Artificial Intelligence

Software

Business Intelligence

50%

Of data is locked

\$4T

Missed data opportunities

Life science data is the most privacy-sensitive data

- Cannot simply be used
- Time-consuming
- Extensive Paperwork
- Classic 'anonymization' is no solution

Result: no data, no insights!

Turn data access into a competitive advantage

70%

An increase in industry collaborations is expected with the **use of privacy tools**

30%

More revenue for organizations that **earn and maintain digital trust** with customers and patients

50%

Of data will be **unlocked by privacy-enhancing techniques**

Classic 'anonymization' is no solution

Original data					
Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...	...	...	...	...	...
George	41	Male	Shirt	€25	9 March



Classic anonymization					
Name	Age	Gender	Item	Price	Data
xxx	25-30	Female	Cloth	€100 - €200	March
xxx	70-75	Male	IT	€600 - €700	March
xxx	40-45	Male	Drink	<€5	March
...	...	...	...	...	...
xxx	40-45	Male	Cloth	€20 - €30	March

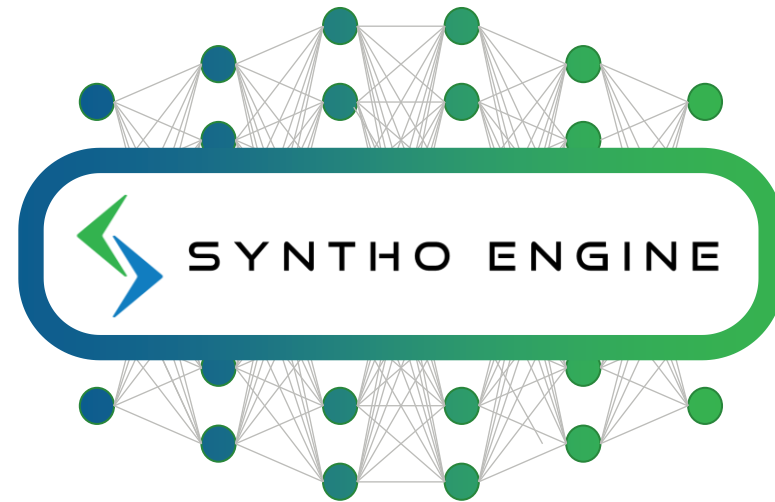
N=100k

× Privacy risks

× Low accuracy

× Low scalability

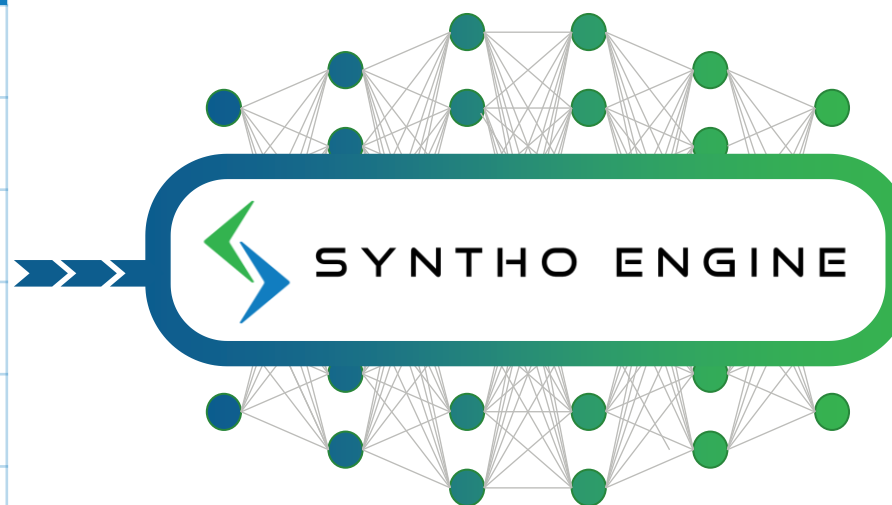
Generative AI For Tabular Data



A synthetic data twin of your original data

Original data					
Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...	...	...	...	...	...
George	41	Male	Shirt	€25	9 March

N=100k



Generate a synthetic data twin of your original data

Original data					
Name	Age	Gender	Item	Price	Data
Olivia	26	Female	Shoes	€125	4 March
John	75	Male	Laptop	€695	5 March
George	41	Male	Beer	€4	7 March
...	...	...	...	...	...
George	41	Male	Shirt	€25	9 March



Synthetic Data Twin					
Name	Age	Gender	Item	Price	Data
NewID1	23	Male	Sofa	€790	1 March
NewID2	23	Female	Scarf	€40	3 March
NewID3	52	Male	Razor	€5	9 March
...	...	...	...	...	...
NewIDn	35	Male	Wine	€7	7 March

N=100k

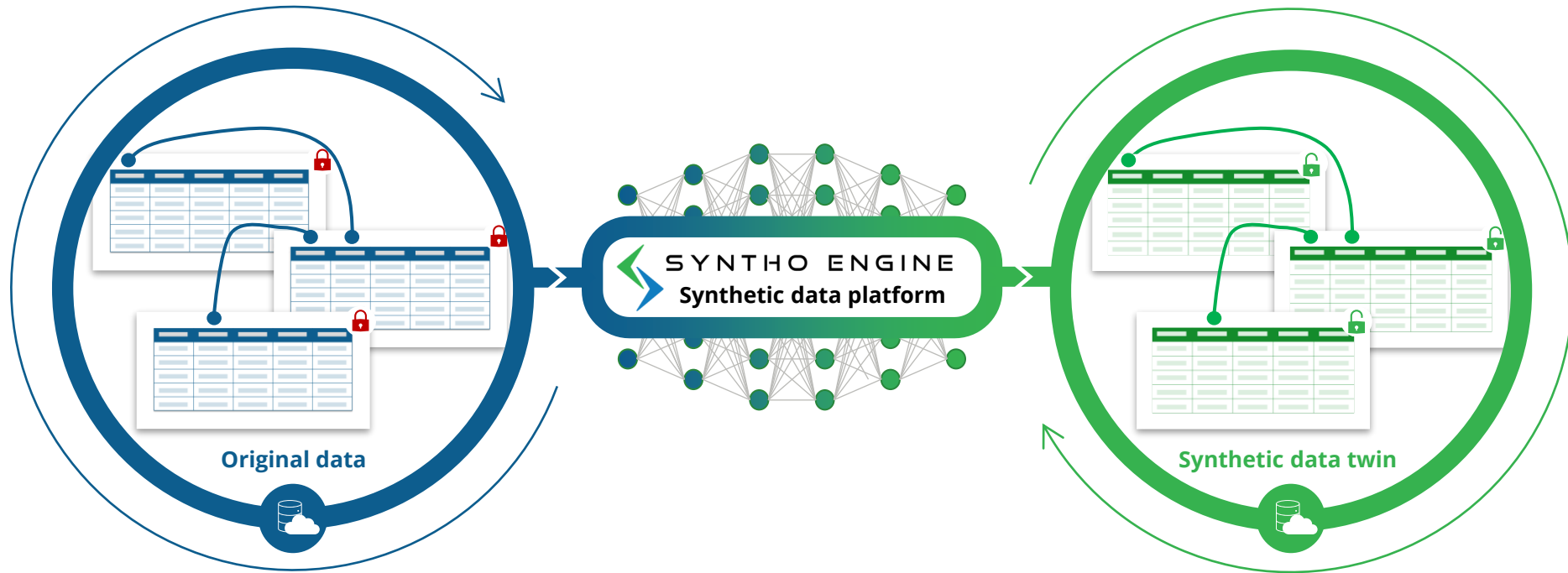
N=100k

✓ Privacy by design

✓ High accuracy

✓ High scalability

AI-generated synthetic twin of your original data

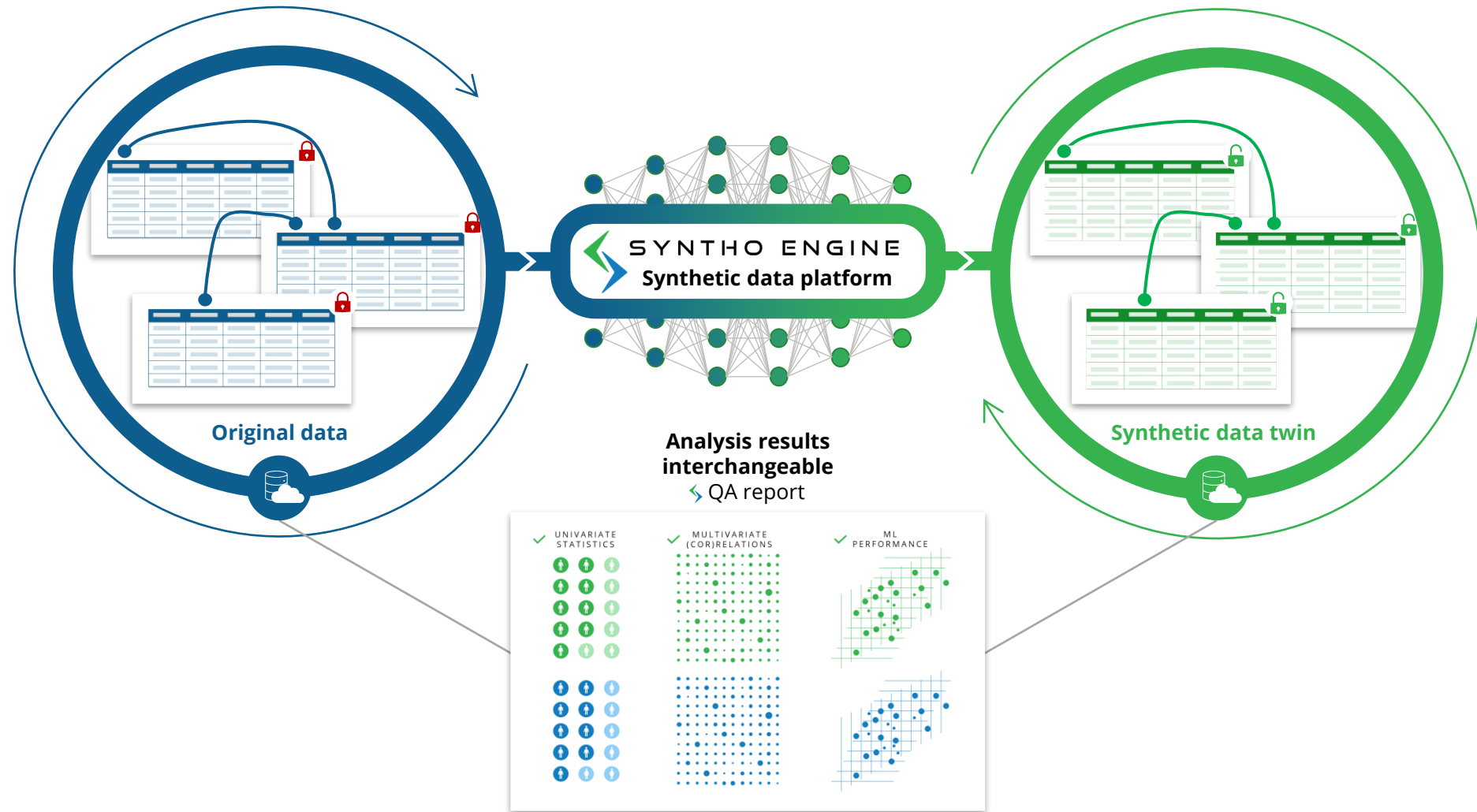


✓ **Privacy by design**

✓ **High accuracy**

✓ **High scalability**

AI-generated synthetic twin of your original data



What do the privacy regulators say about synthetic data?



Under the **EU General Data Protection Regulation**, data synthesis would benefit from the same reasoning as the case for **not needing legal basis for anonymization**.

Article 29, EU Data Protection Work Party
[IAPP 1 & 2 \(p.7\)](#)



Under the **U.S. Health Insurance Portability and Accountability Act**, synthesis would similarly be deemed to be a form of treatment, planning and operations and therefore does **not require additional authorization by a covered entity**.

US Department of Health & Services [HHS](#)

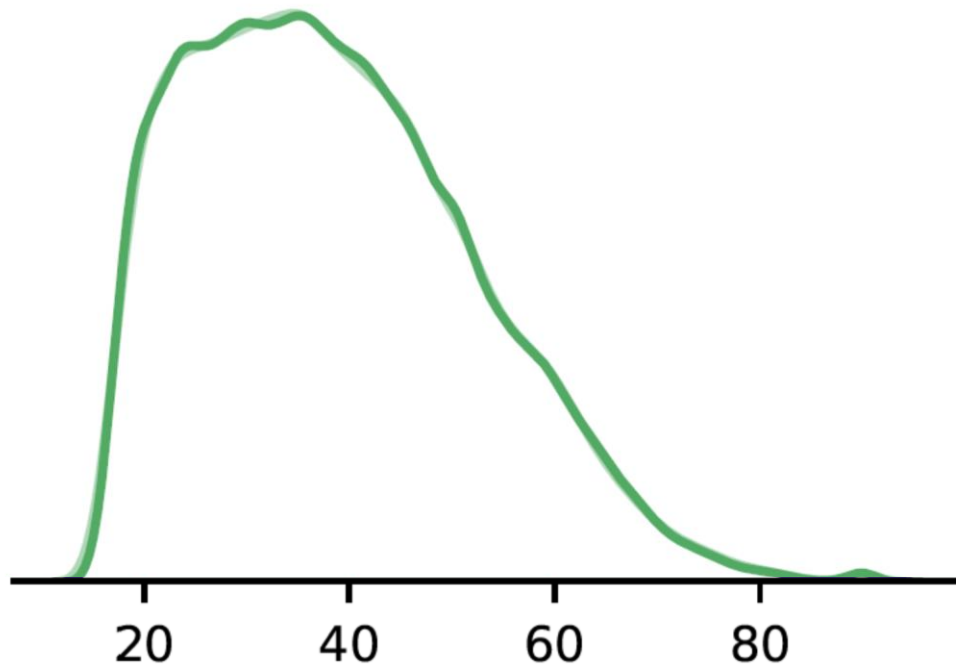


EU DPAs are open **to labeling synthetic data anonymous** under data protection standards.

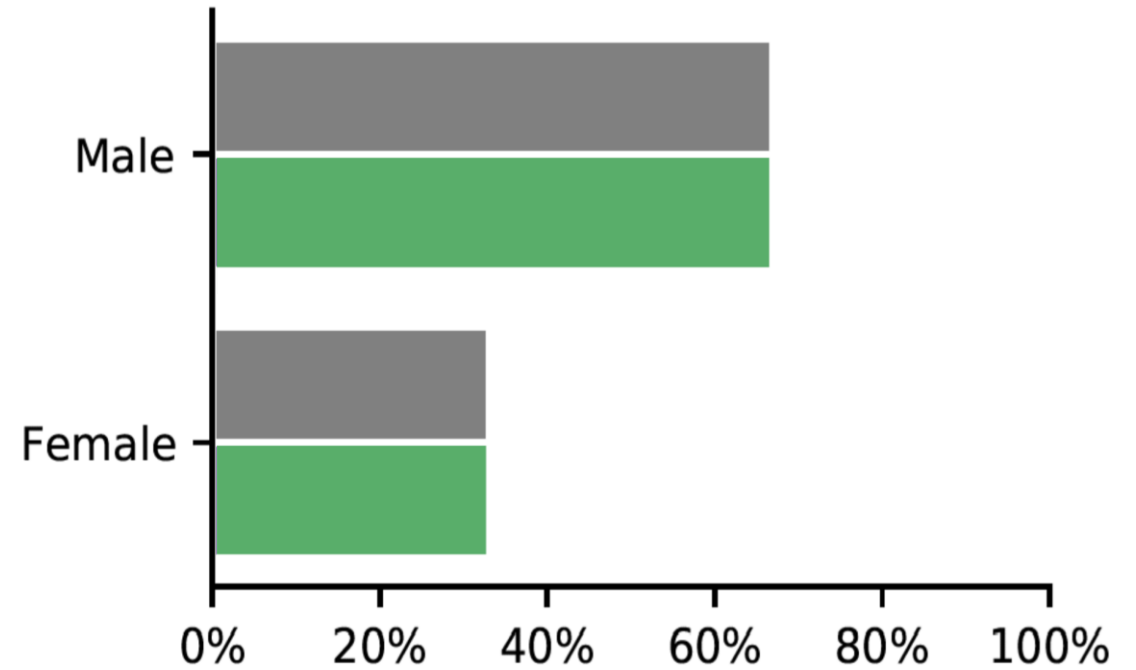
[A guide to the EU's unclear anonymization standards \(2021\)](#)

Distributions are preserved

Age



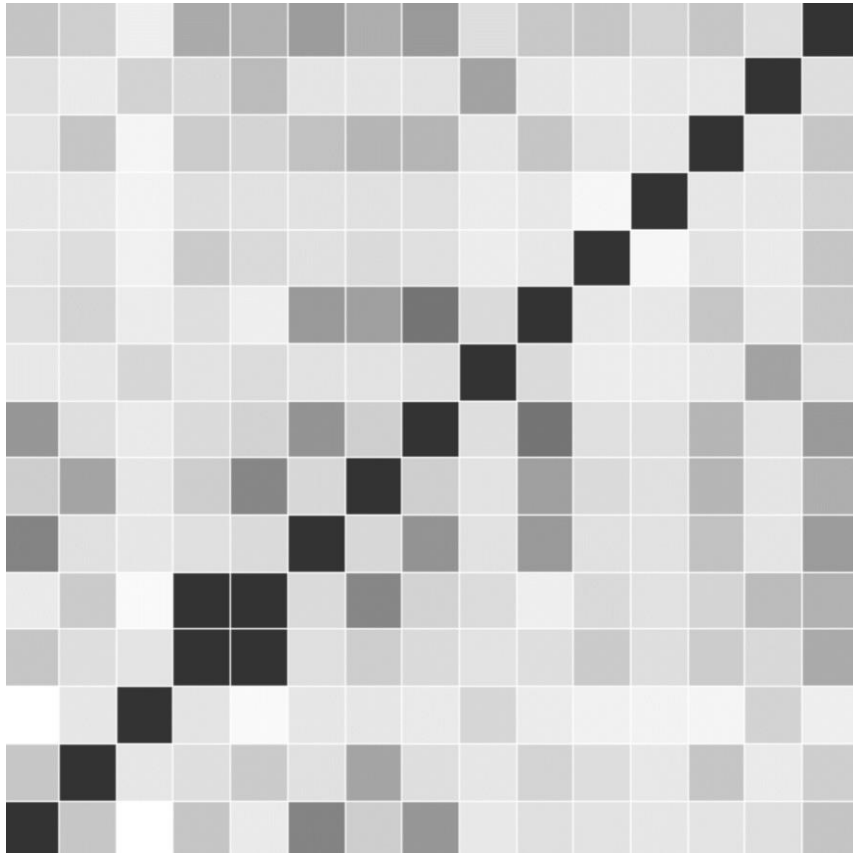
Gender



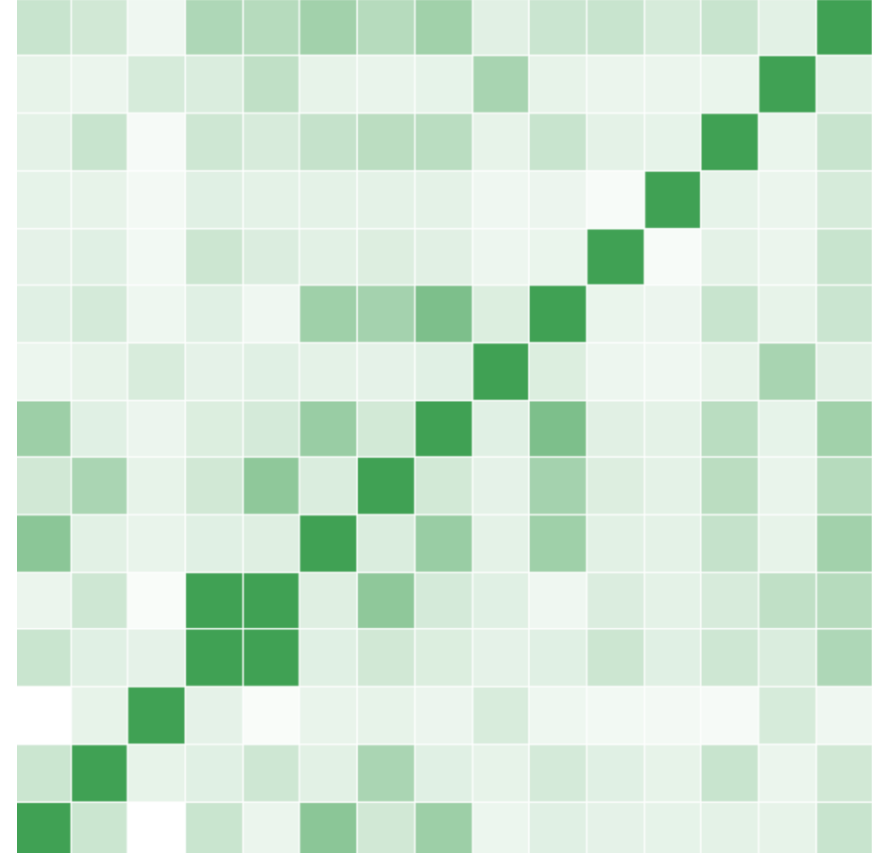
Original data

Synthetic data

Correlations are preserved

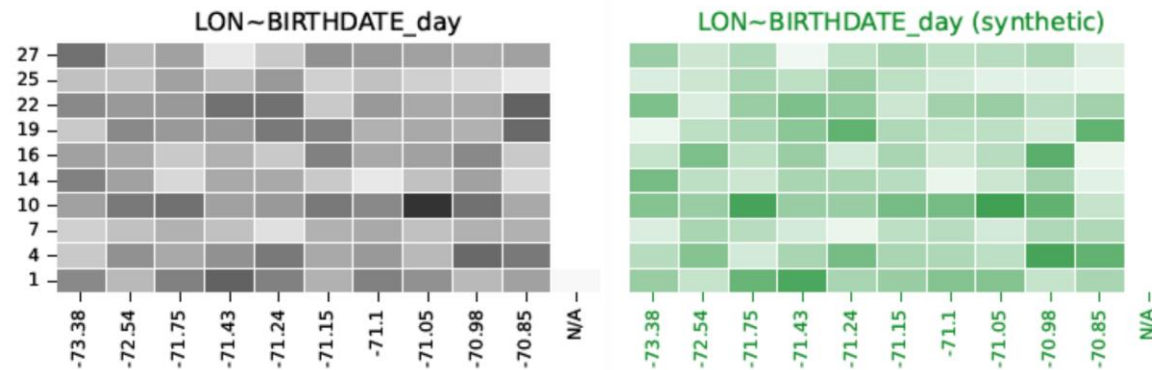
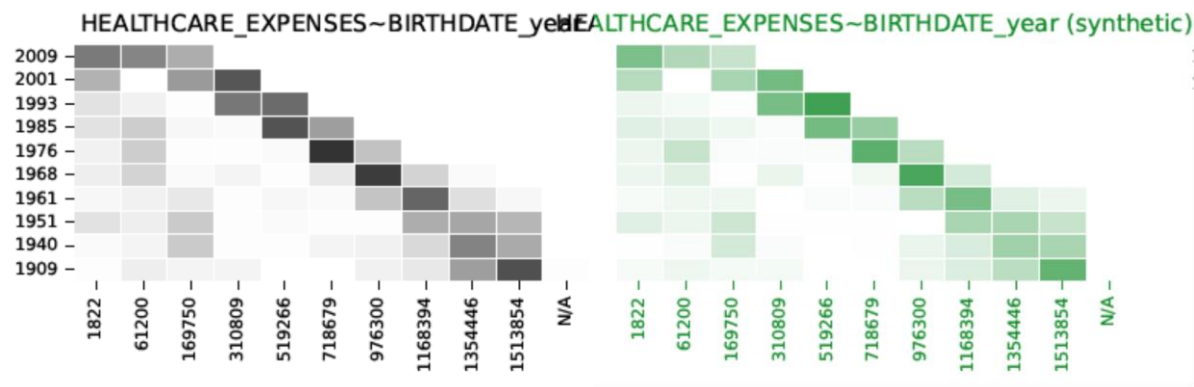


Original data

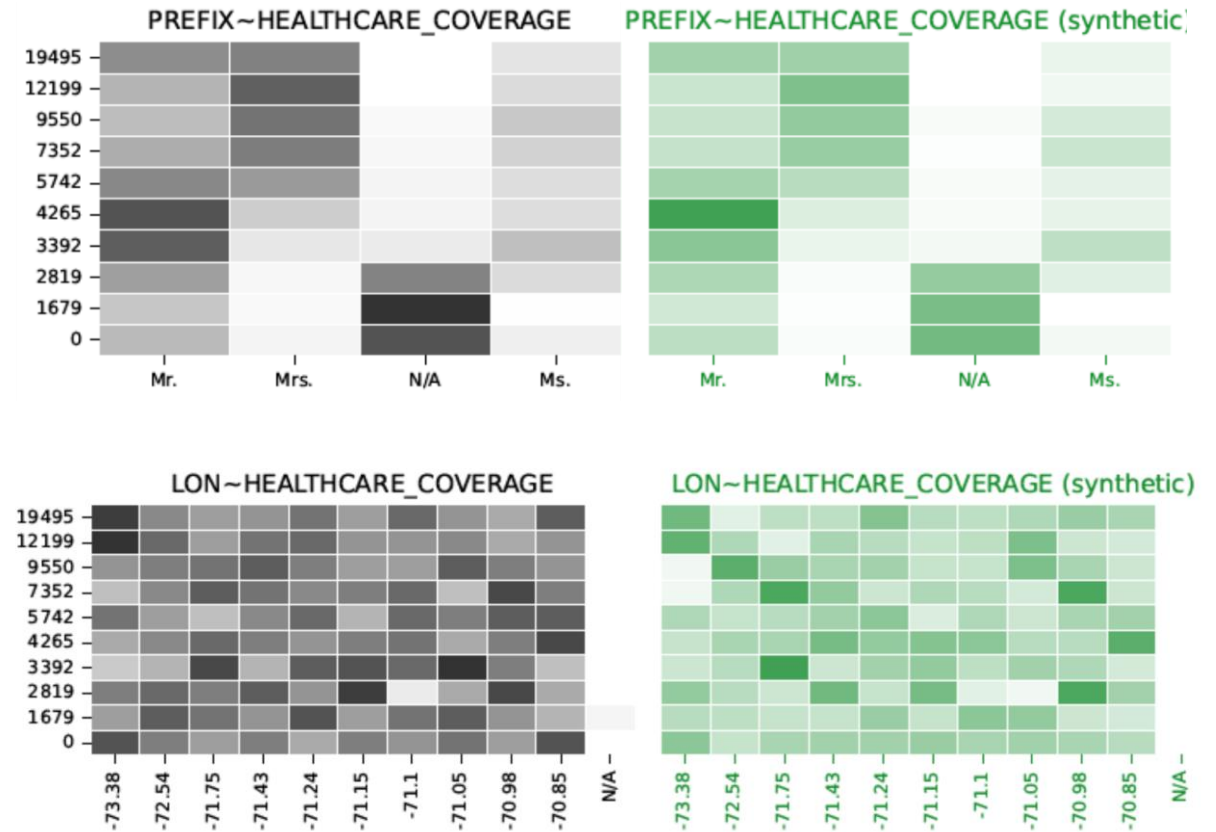


Synthetic data

BI-variates are preserved

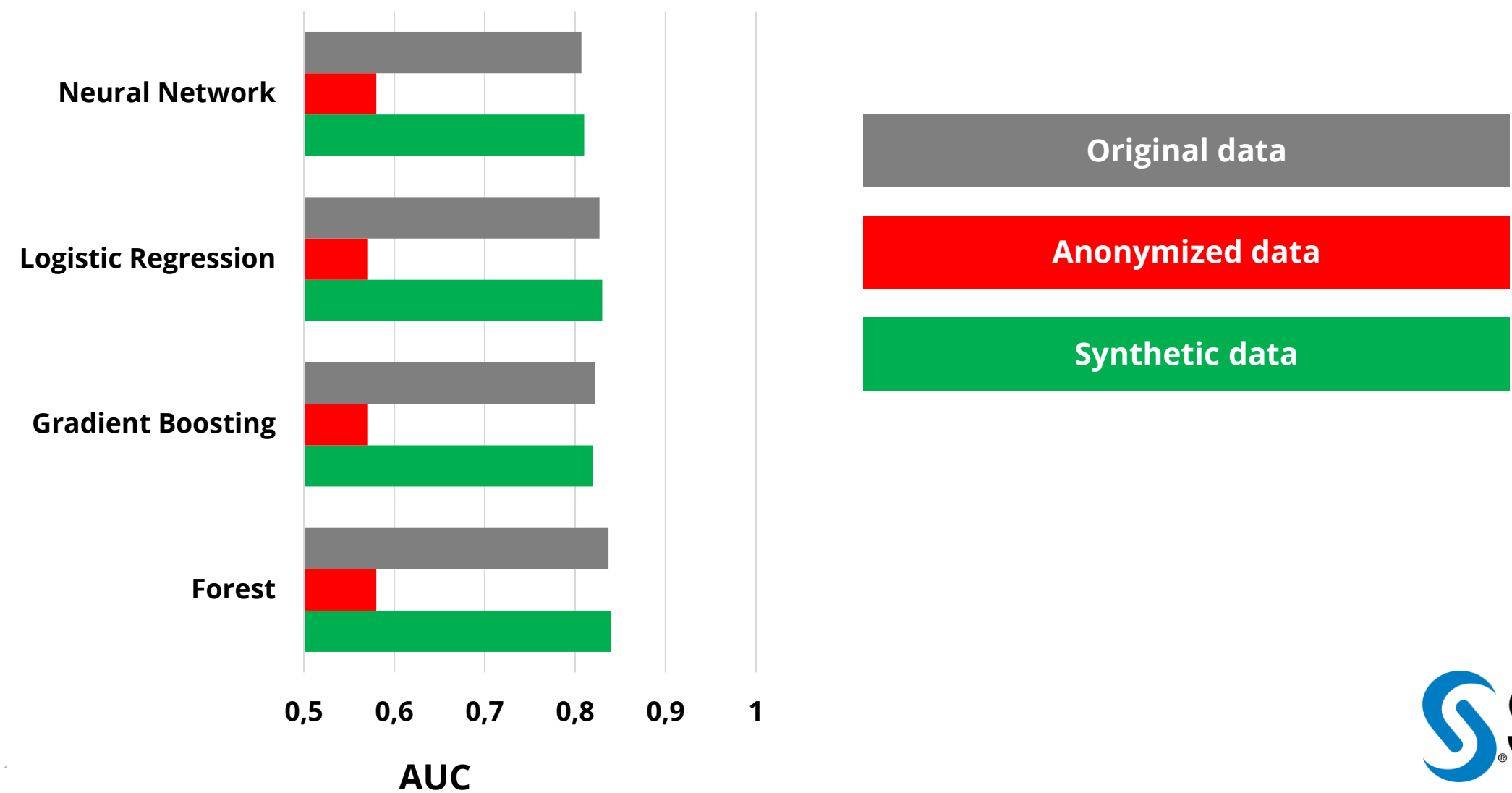


Original data



Synthetic data

Synthetic data **approved** by the data experts of **SAS**

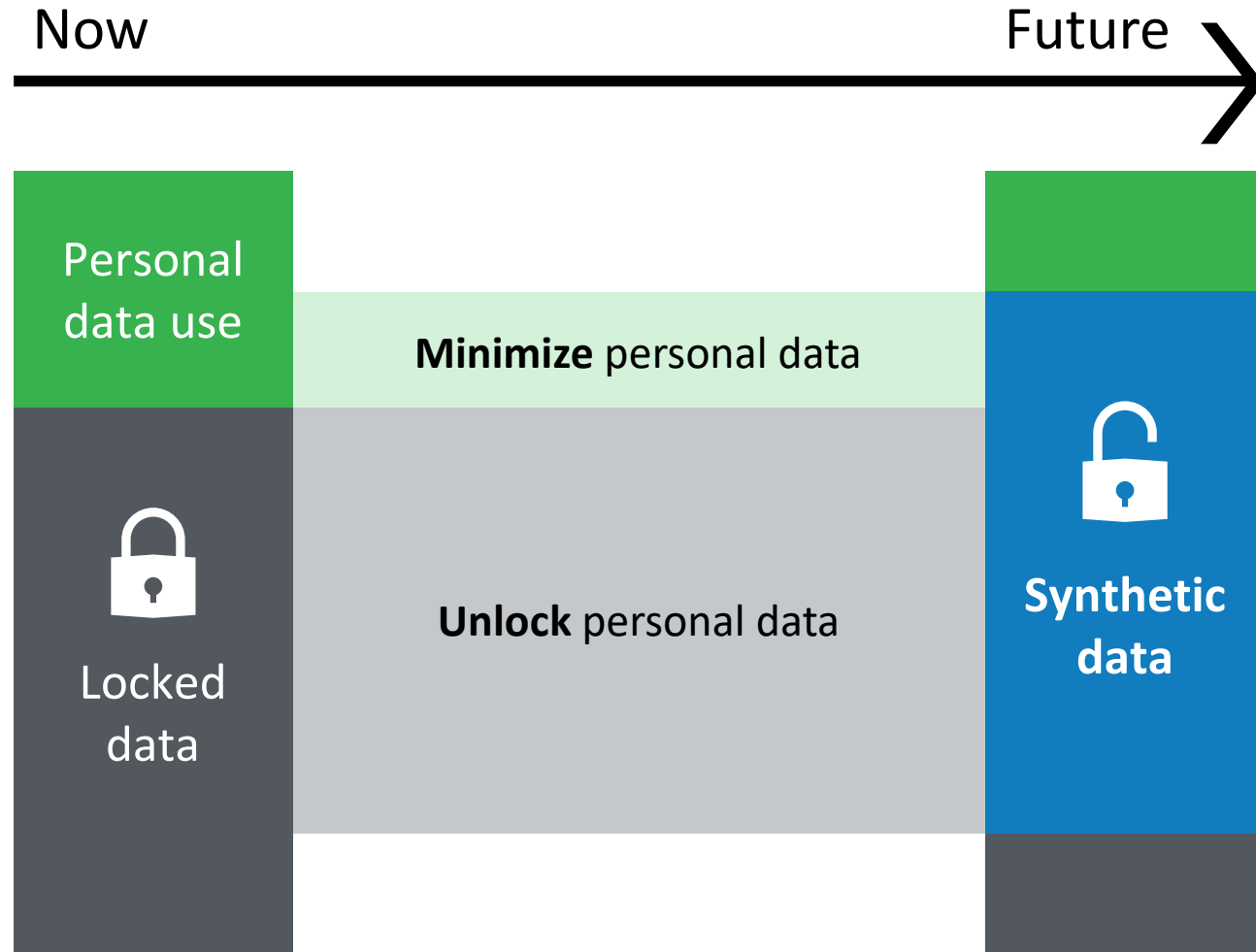


Synthetic data **approved** by the data experts of **SAS**



- **Synthetic data** is as-good-as the original data
- **Anonymized data** shows the worst performance
- Easy, fast, and scalable

Why use real data when you could use synthetic data?



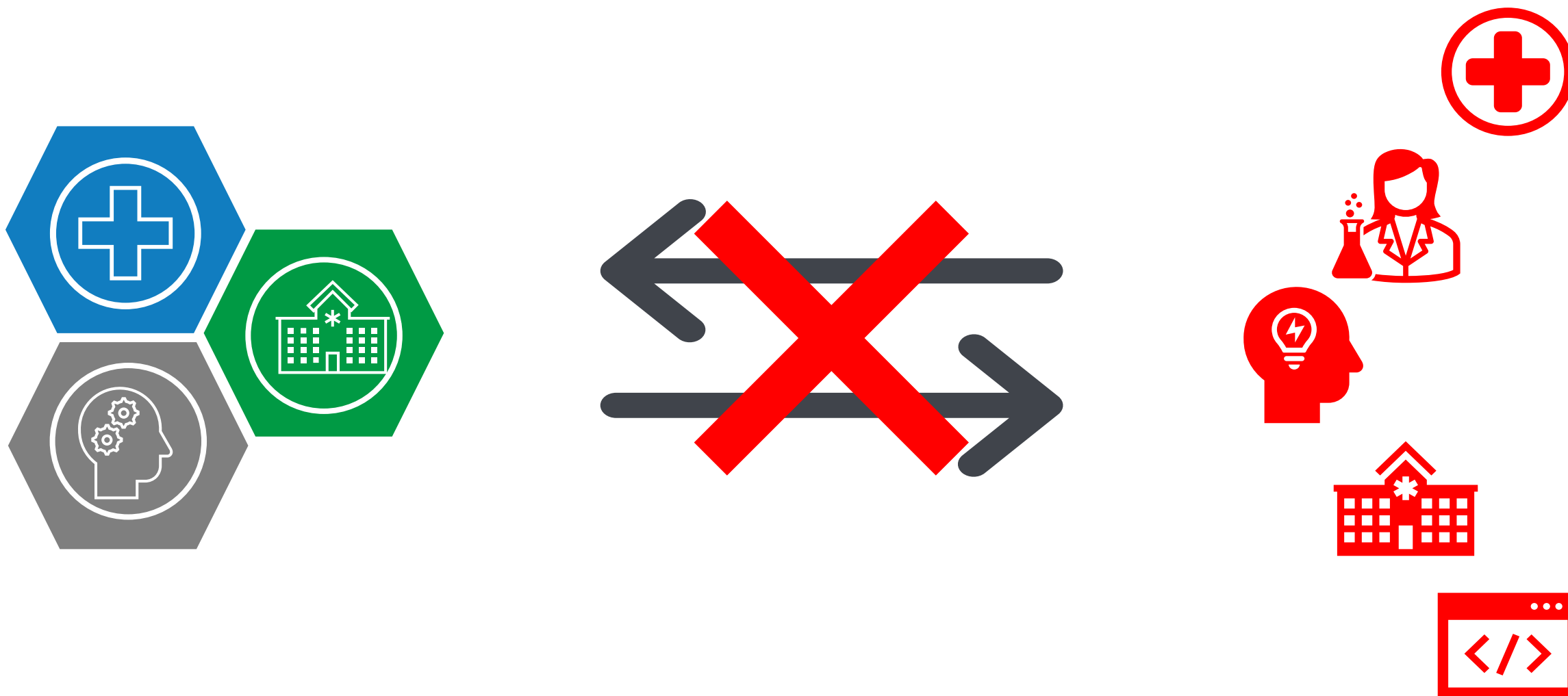
Key benefits

1. Less risk
2. More data
3. Faster data access

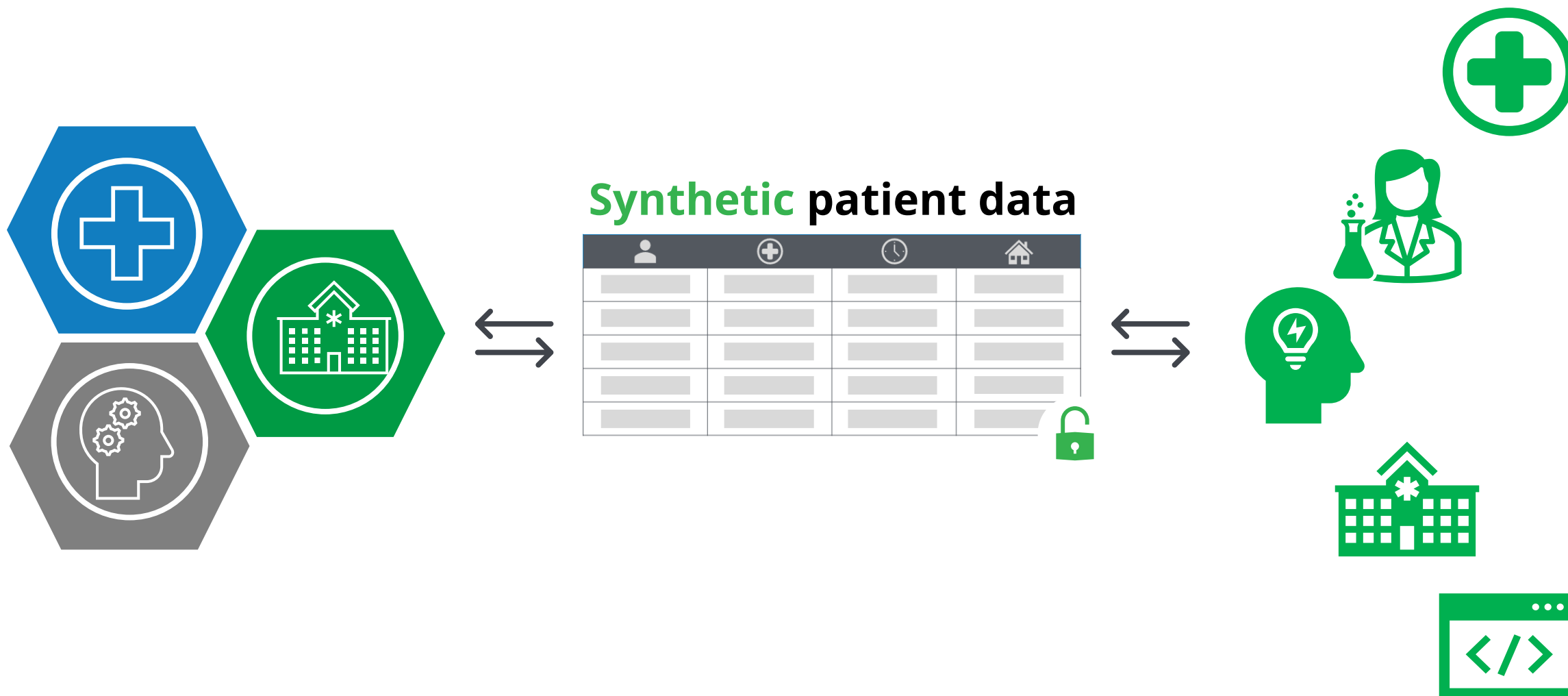


Strong data foundation

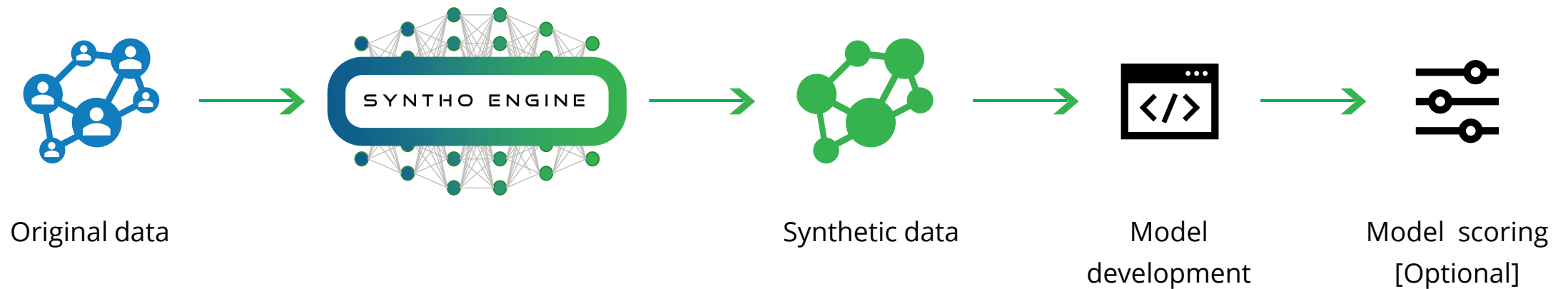
Data sharing: using and sharing data is challenging



Data sharing: freely use and share synthetic data



Analytics: Synthetic Data for Analytics



✓ Ranging from **business intelligence** to **artificial intelligence** solutions

HEALTH CARE & LIFE SCIENCES **WINNER**

AI Generated Synthetic Data in Healthcare

Syntho, Netherlands

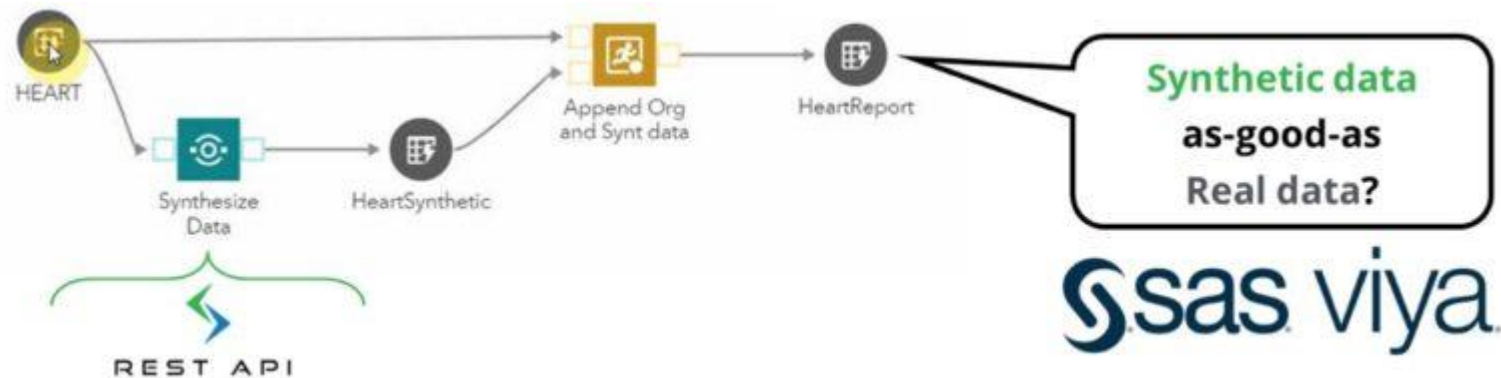


Using AI-generated synthetic data in hospitals can be a powerful tool for utilizing data faster, without privacy concerns.

SAS HACKATHON
#SASHackathon



Synthetic data as-good-as-real?



Correlations

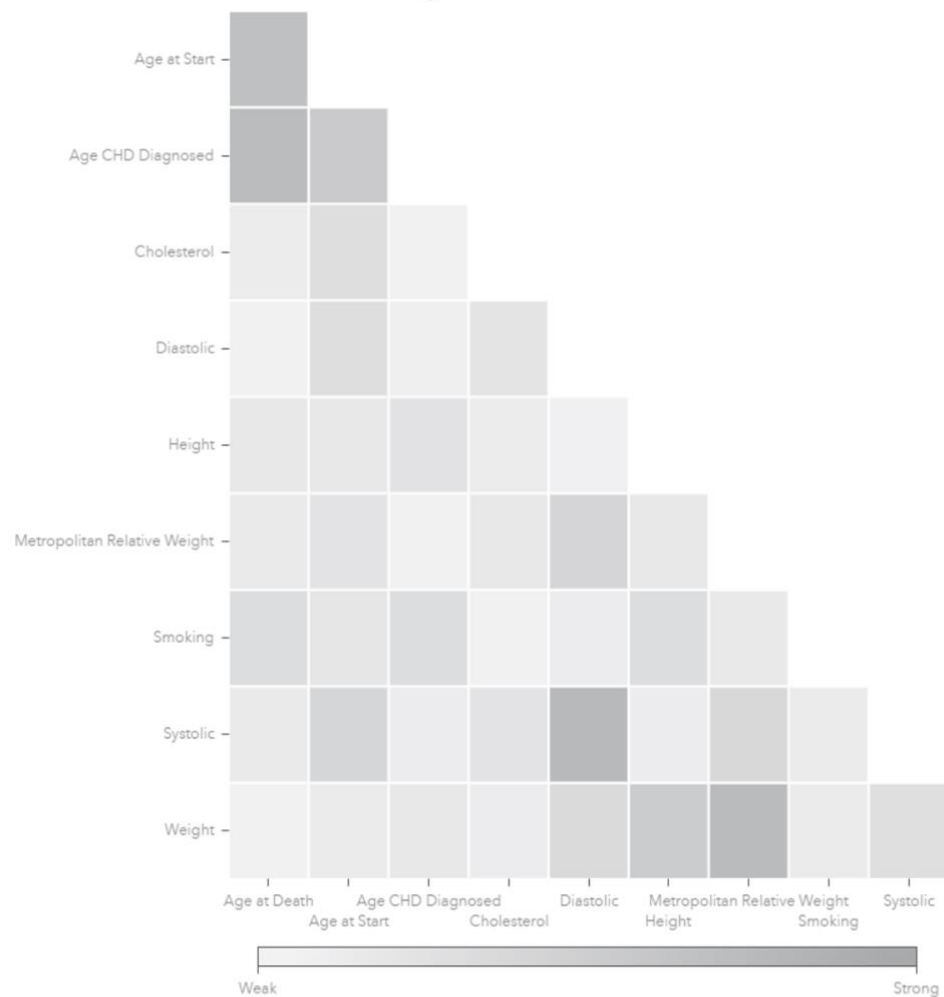


Original data

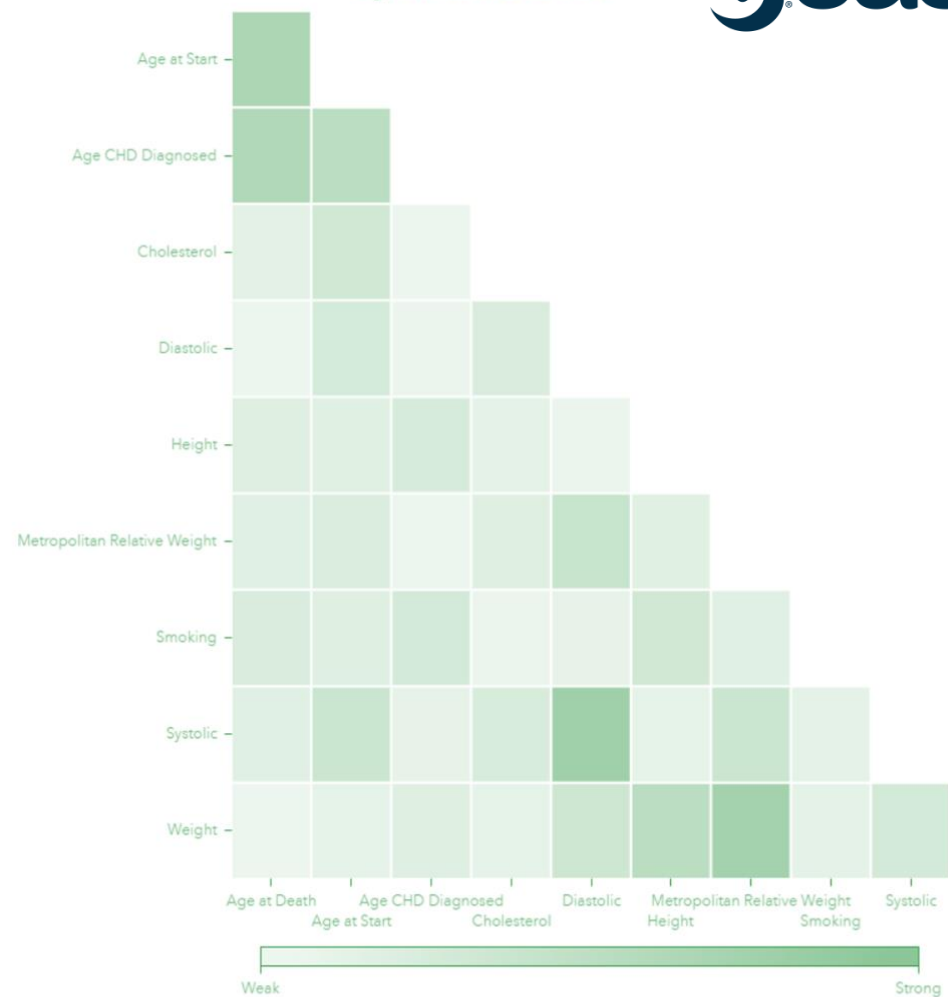
Synthetic data



Original data



Synthetic Data



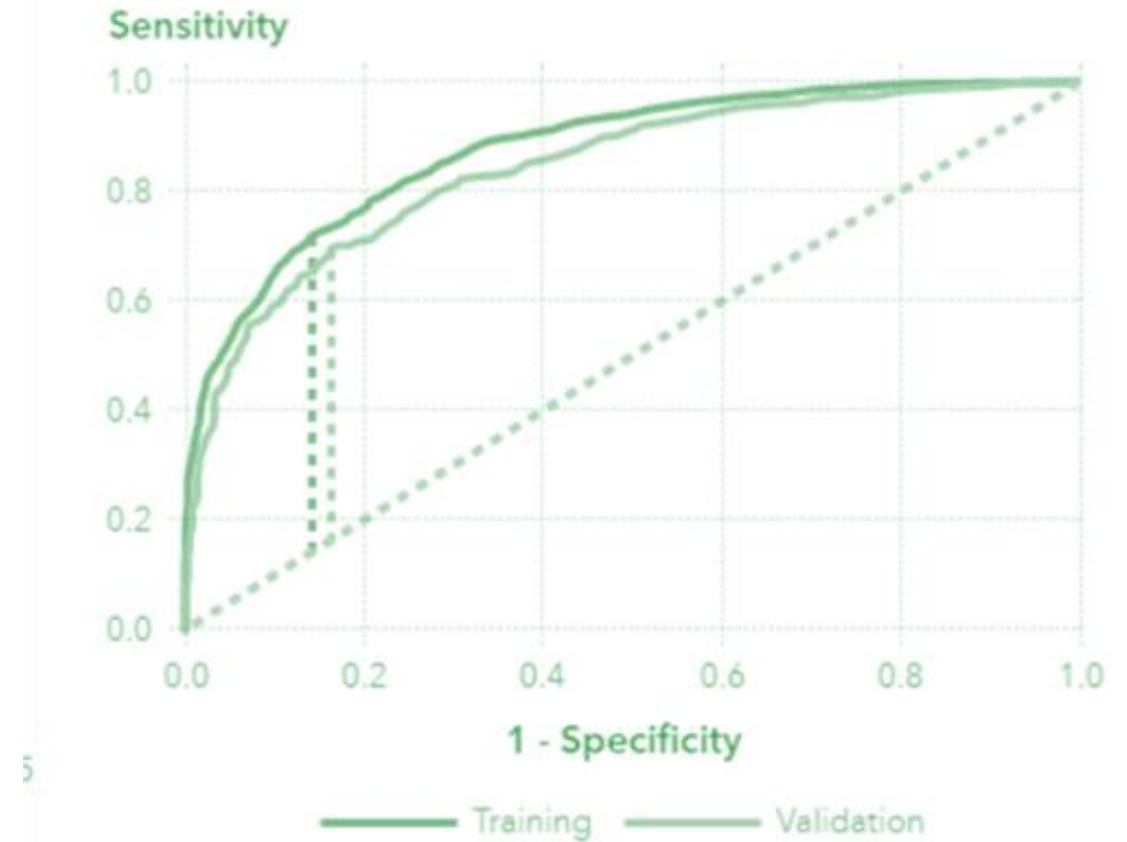
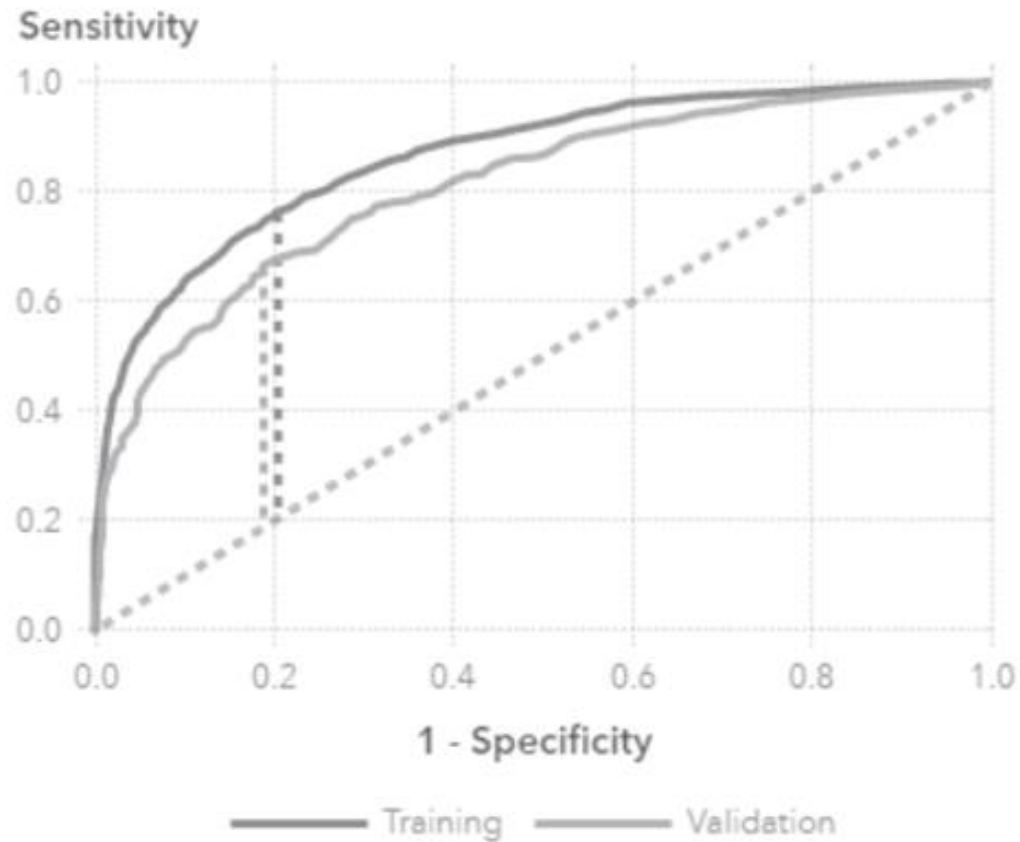
ROC Curve



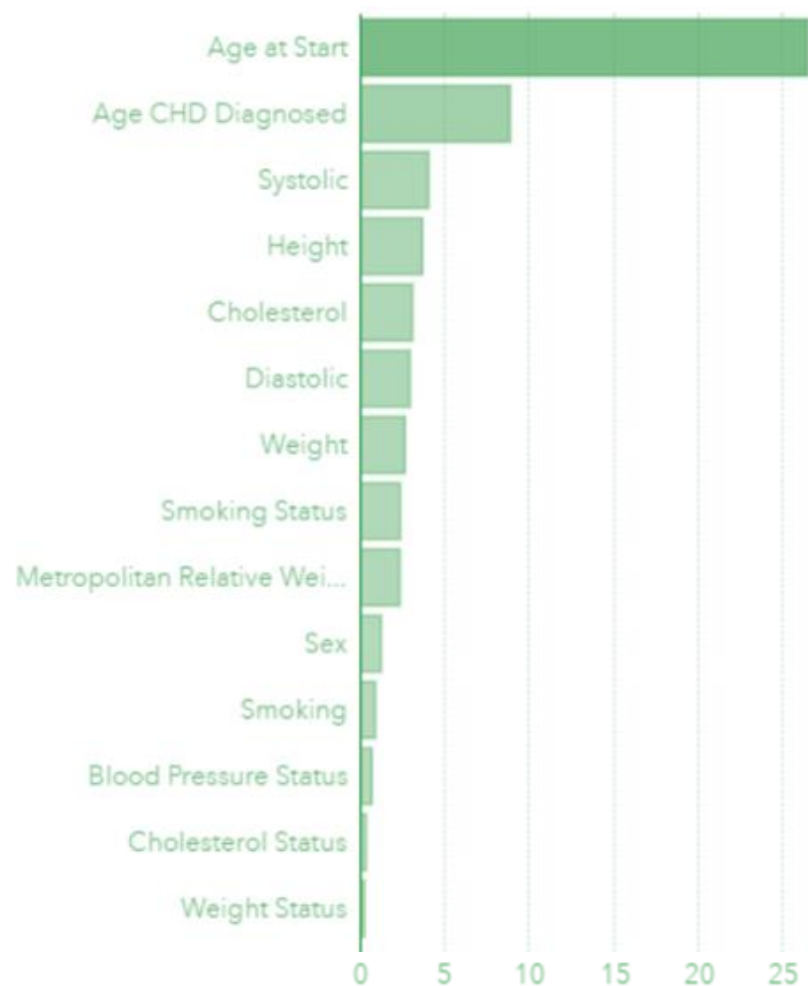
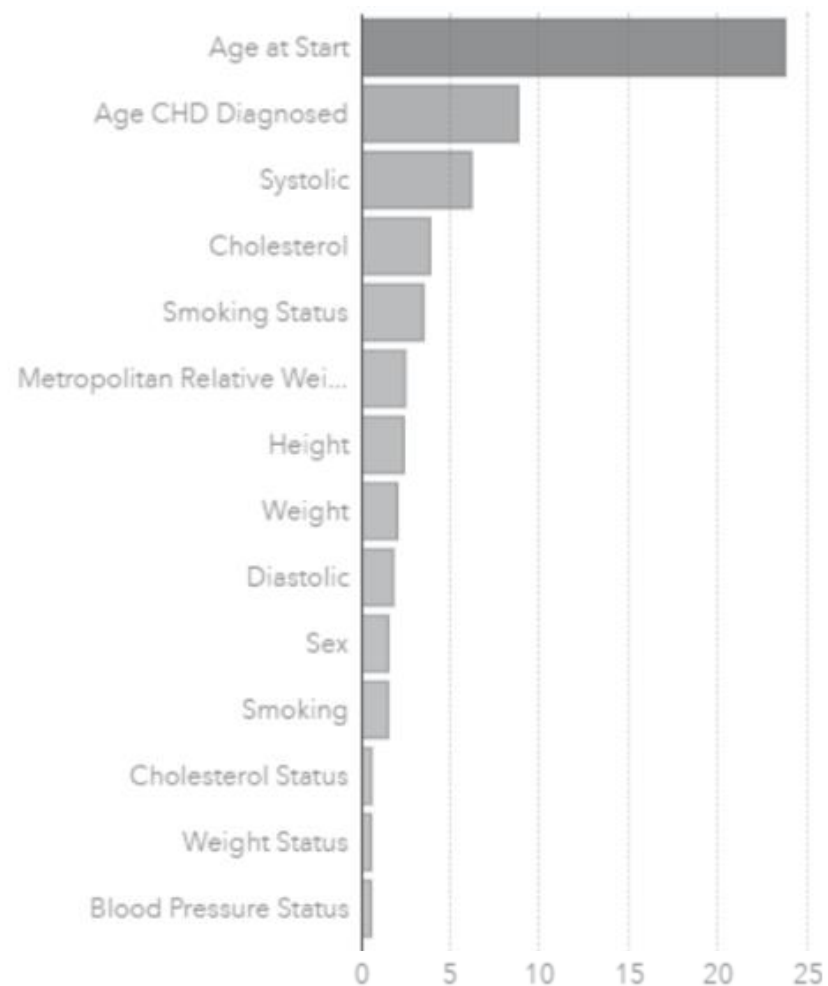
Original data

Synthetic data

sas viya



Variable importance

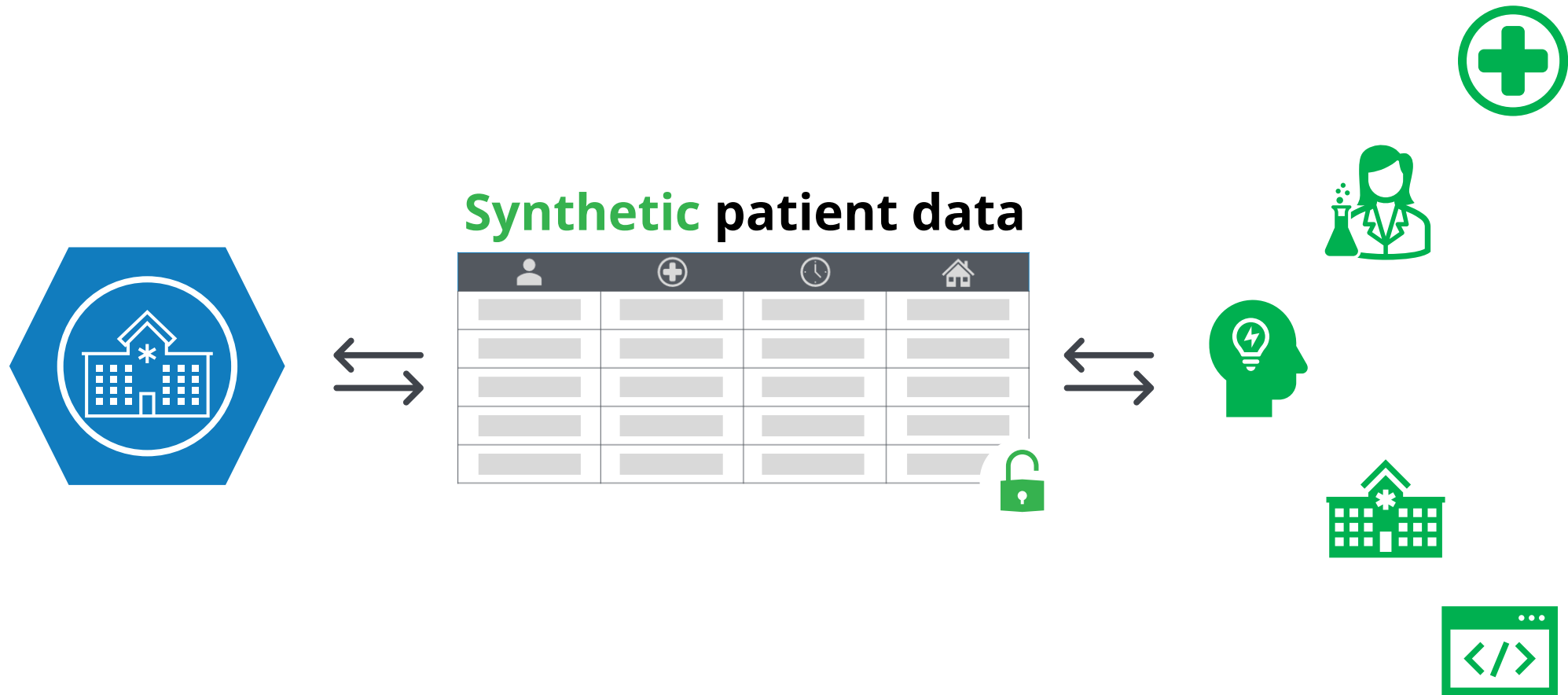


Original data

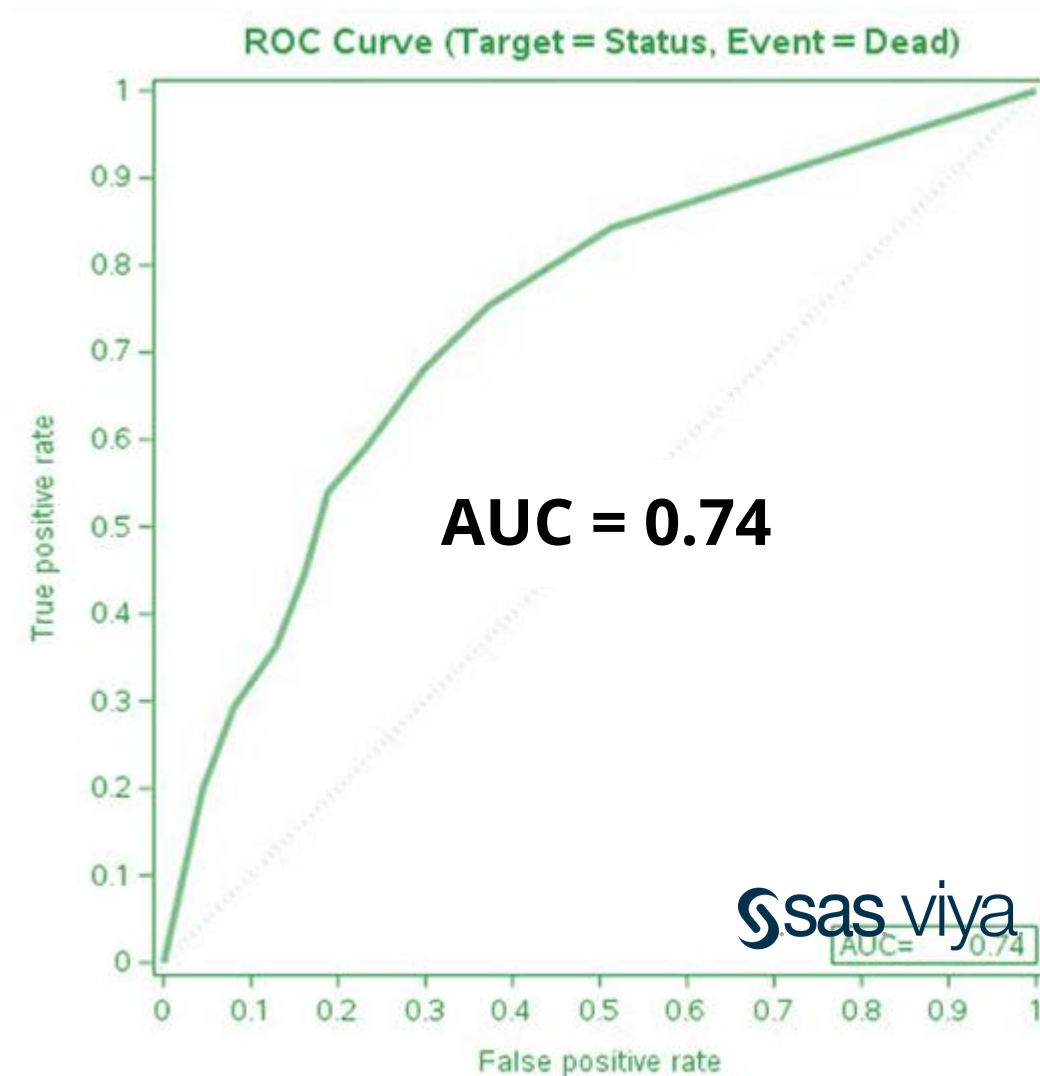
Synthetic data

sas viya

Case Study: Predicting deterioration and mortality as part of cancer research on synthetic data



Case Study: Predicting deterioration and mortality as part of cancer research on synthetic data



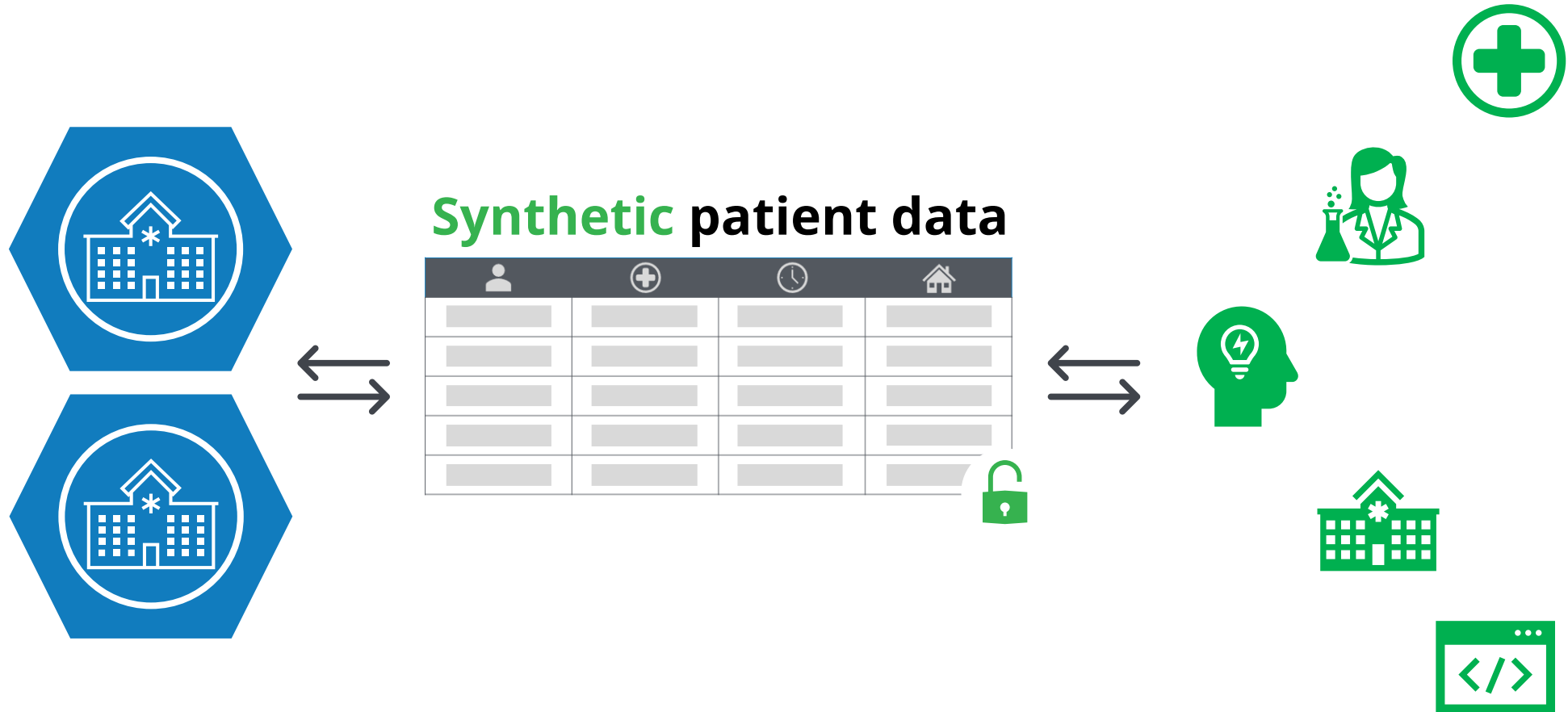
Case Study: Predicting deterioration and mortality as part of cancer research on synthetic data

Conclusion

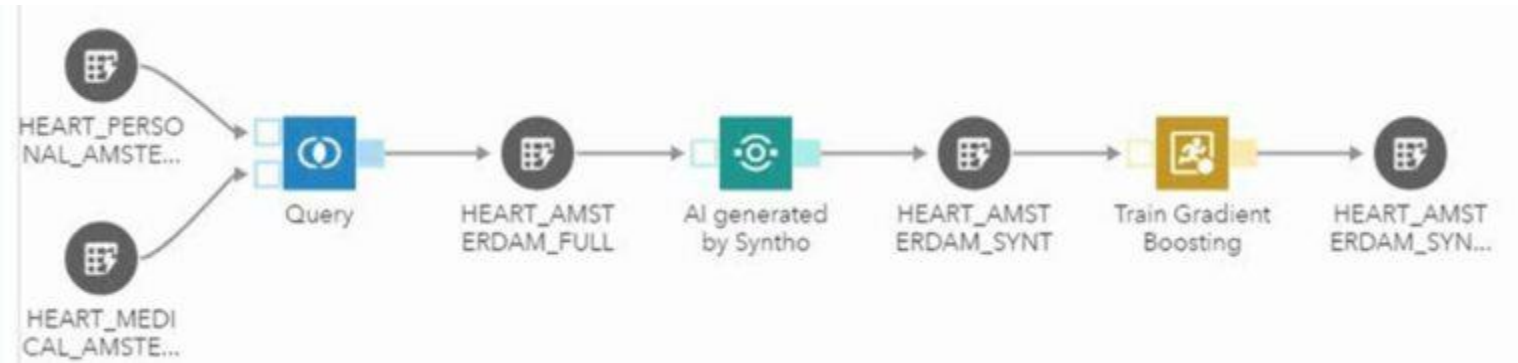
Privacy-sensitive life science data can be unlocked with synthetic data within the hospital to realize data-driven insights

- ✓ Less risk
- ✓ More data
- ✓ Faster data access

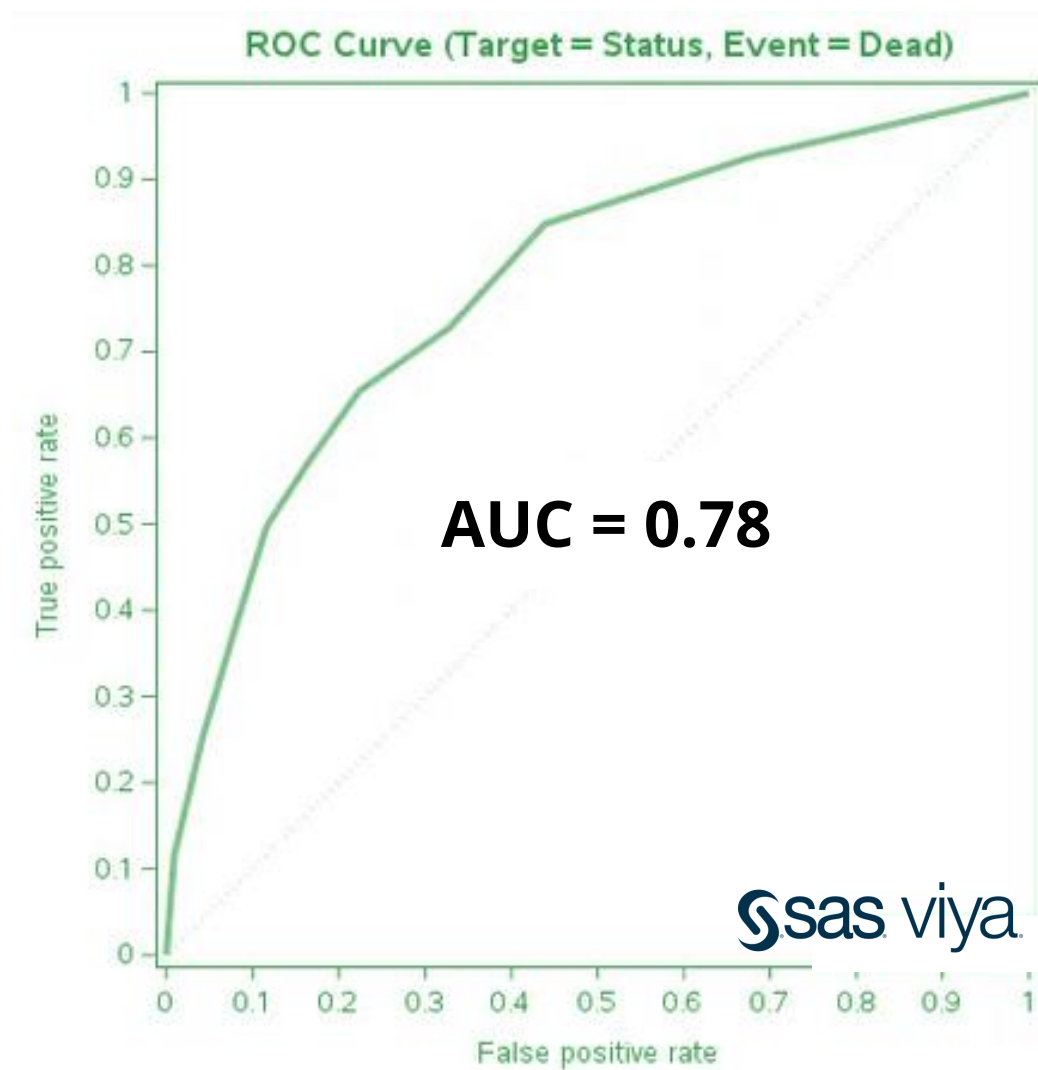
Case Study: Combine synthetic data from *two* hospitals



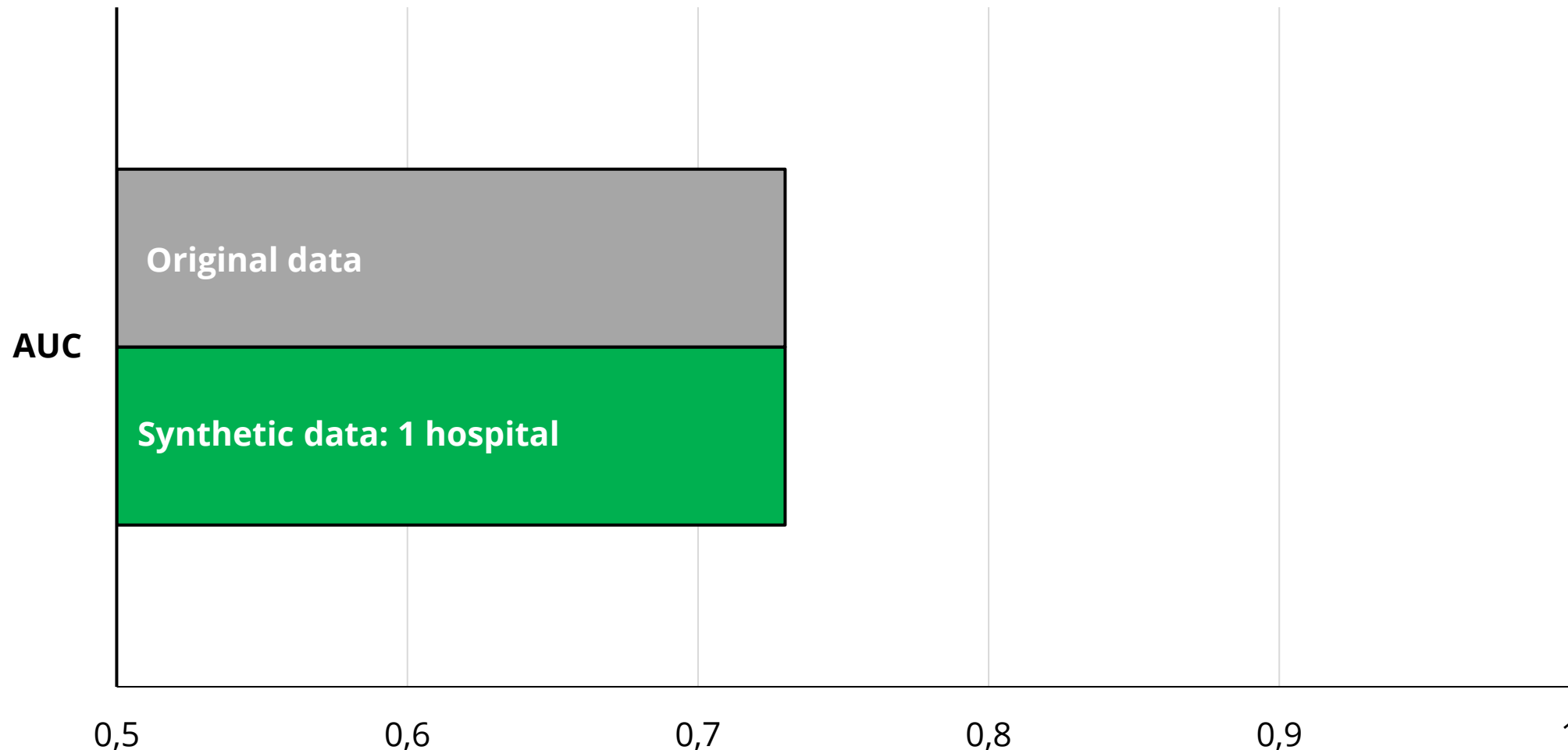
Setup



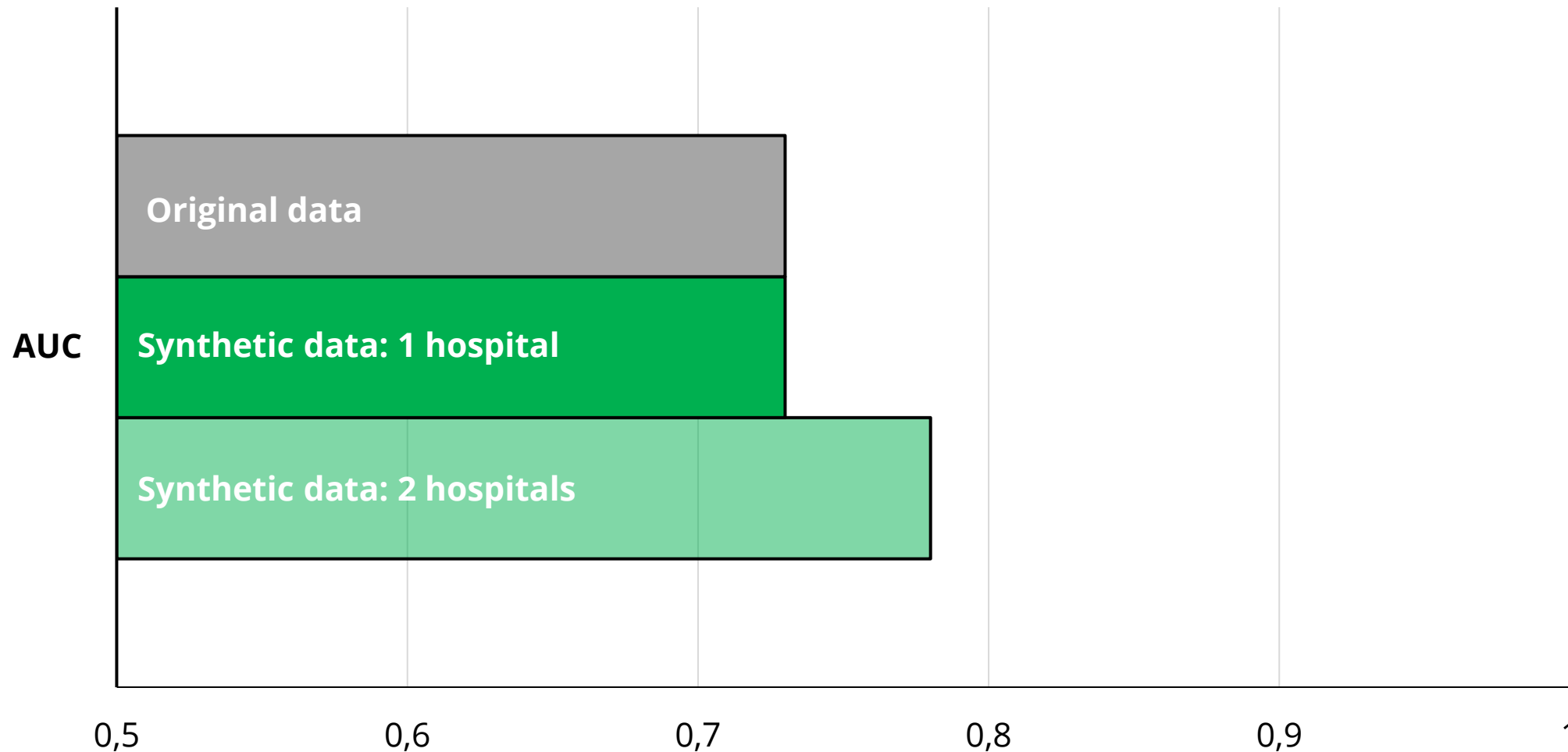
Case Study: Combine synthetic data from *two* hospitals



Case Study: Synthetic data is as-good-as-real and can be used for clinical research as alternative for sensitive real data



Case Study: Synthetic data can easily be shared. Access to more data could improve the accuracy of predictive models.



50%

Syntho unlocks data

\$4T

Realize data opportunities

Turn data access into a **competitive advantage**

70%

An increase in industry collaborations is expected with the **use of privacy tools**

30%

More revenue for organizations that **earn and maintain digital trust** with customers and patients

50%

Of data will be **unlocked by privacy-enhancing techniques**

Synthetic data – Real people!



Wim Kees Janssen

Founder & CEO

Contact: **kees@syntho.ai**