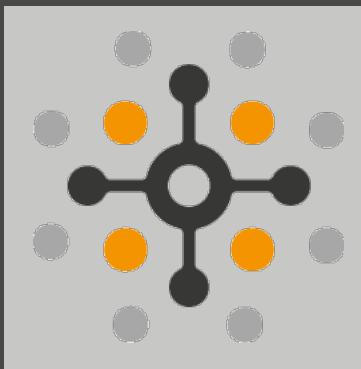


WHY YOU SHOULD CARE ABOUT **SYNTHETIC DATA**

Presented by Real Impact Analytics

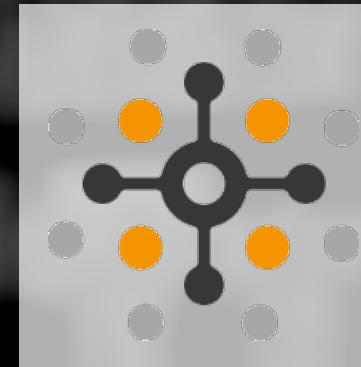


QUESTIONS?

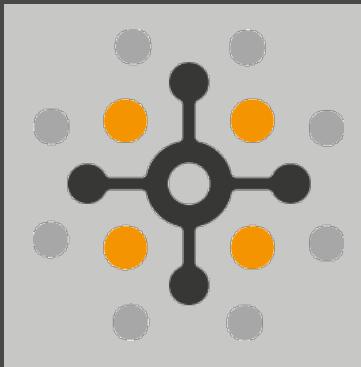


#SYNTHETICRIA

SYNTHETIC DATA OVERVIEW

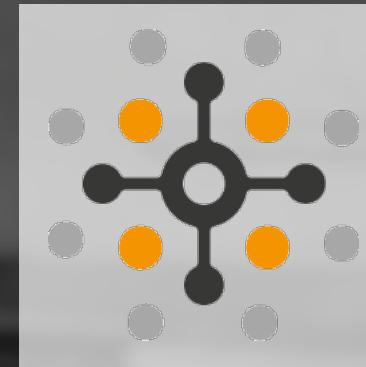


- 03 **What** is synthetic data?
- 08 **Why** use it?
- 13 **How** to create it?
- 25 **Who** creates it?
- 27 **Conclusion**

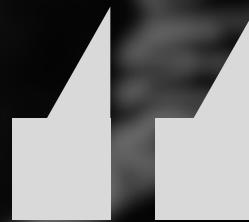


WHAT IS SYNTHETIC DATA

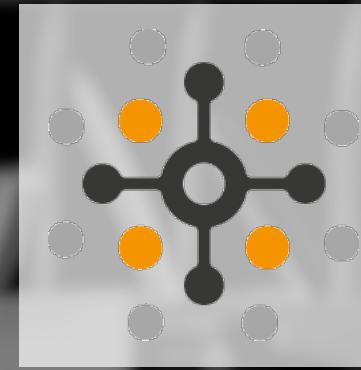
WHAT IS SYNTHETIC DATA?



“Generic and artificial data used to mimic real-world data sets.



WHAT IS SYNTHETIC DATA?



“

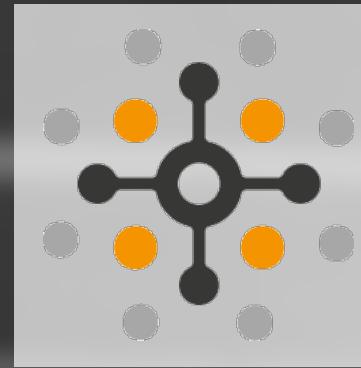
Generic and artificial data used to mimic real-world data sets.



Protect people's privacy

substitutes real data that contains personal information

WHAT IS SYNTHETIC DATA?



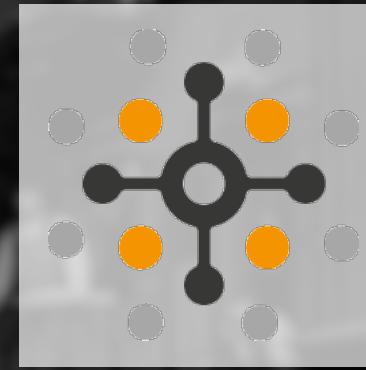
Generic and **artificial** data used to **mimic** real-world data sets.

Test robustness and accuracy



during software development

WHAT IS SYNTHETIC DATA?



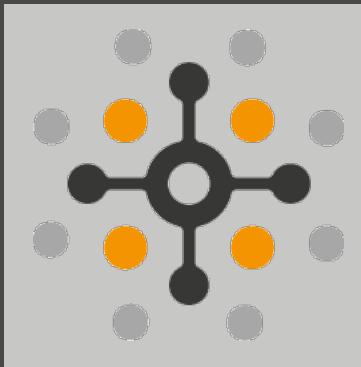
“

Generic and artificial data used to **mimic** real-world data sets.

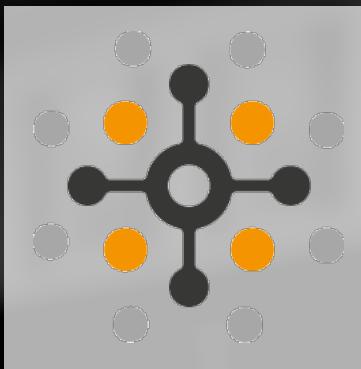


.... Create artificial base

with similar features of real data sets

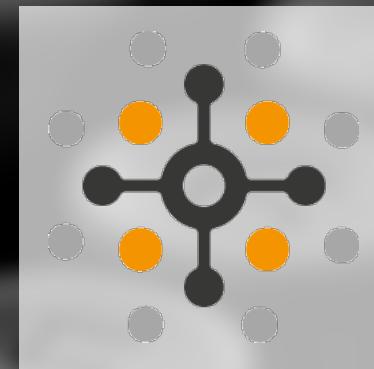


WHY USE IT?

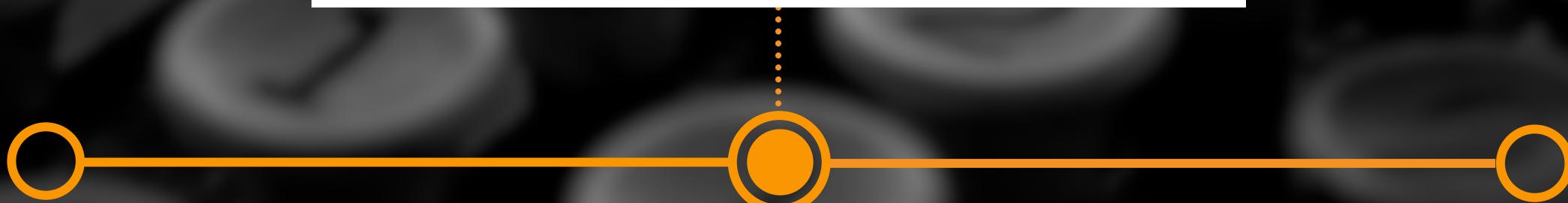


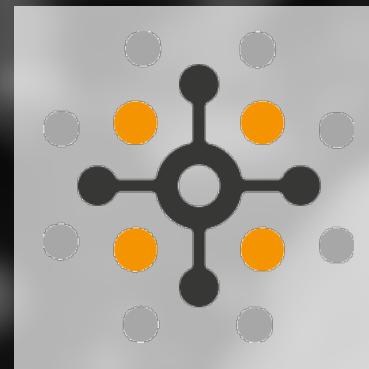
Use of actual data sets is no longer allowed, to protect everyone's right to privacy.





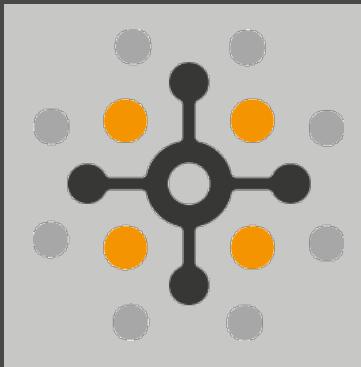
To develop big data tools, we need realistic data sets for testing algorithms and easy data visualization.





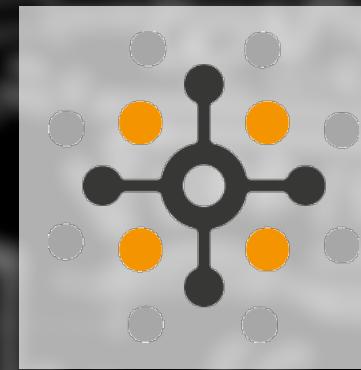
Synthetic data - similar to real data sets & shareable to public - acts as a substitute without invading anyone's privacy.





HOW
TO CREATE IT?

HOW TO CREATE IT?

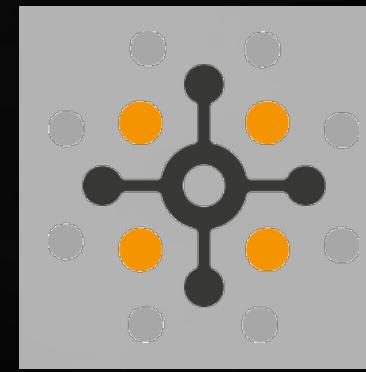


1 DRAWING
NUMBERS

OR

2 AGENT-BASED
MODELLING

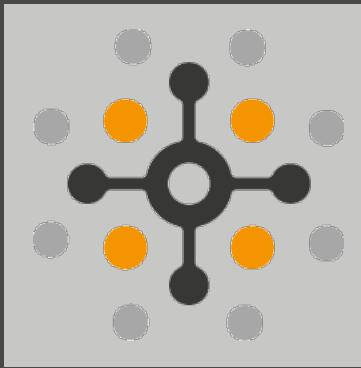
HOW TO CREATE IT?



1

DRAWING NUMBERS

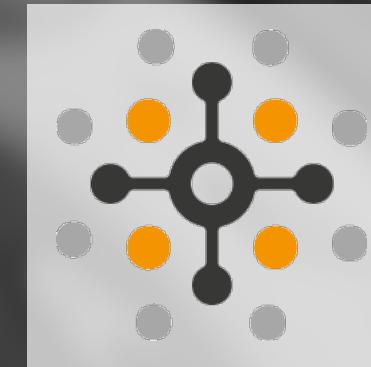
Observe real-world statistic distributions from original data to reproduce artificial bases by drawing simple numbers.



EXAMPLE TELECOM DATA

DRAWING NUMBERS

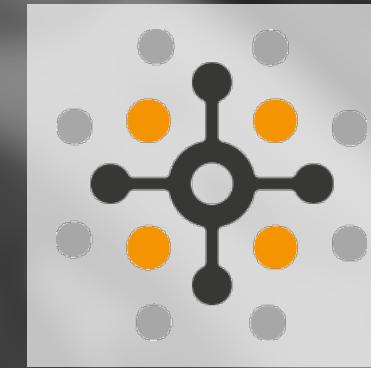
DRAWING NUMBERS



1 / 3

Observe the real
temporal distributions of texts
and phone calls from CDR data
(call detail records).

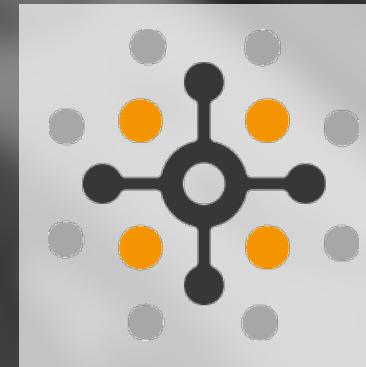
DRAWING NUMBERS



Create an artificial base of customers.

2
/3

DRAWING NUMBERS

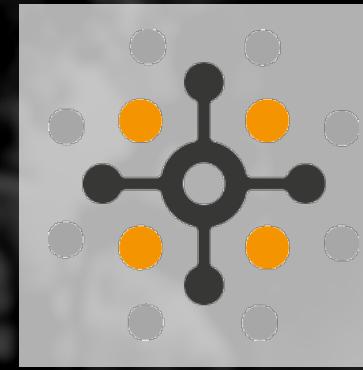


3 / 3

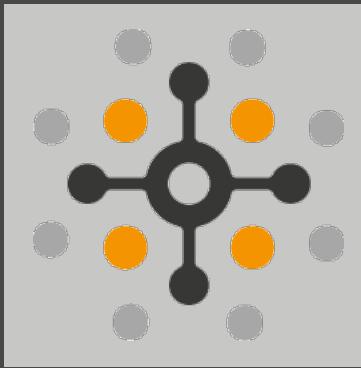
Simulate texts and phone calls with time stamps following the distributions. The goal is to simulate CDR so they follow the same distribution as real CDRs.

2 AGENT-BASED MODELLING

HOW TO CREATE IT?



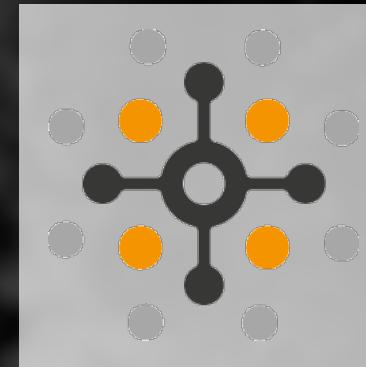
Create physical models to explain observed behaviour to generate generic, random data using this model.



EXAMPLE TELECOM DATA

AGENT-BASED MODELLING

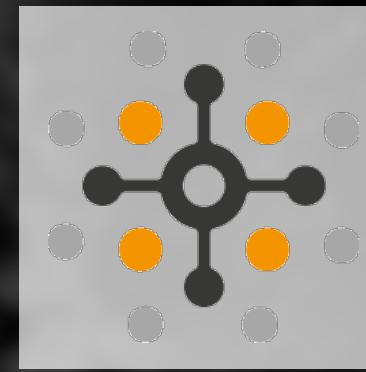
AGENT-BASED MODELLING



1 / 3

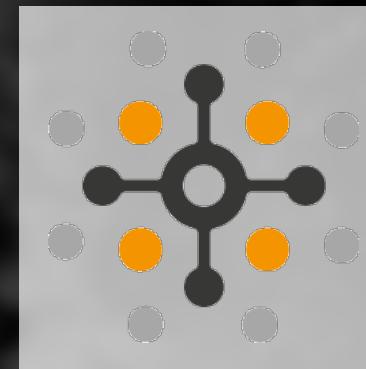
Analyze real data from texts and phone calls, identifying temporal and behavioural patterns.

AGENT-BASED MODELLING

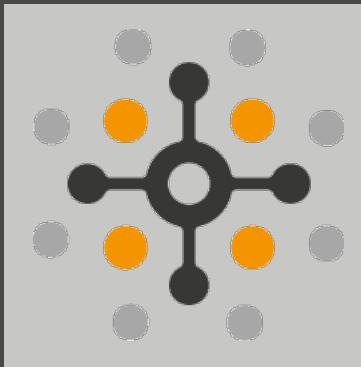


Create a physical model based
on those observations and
evolutions over time.

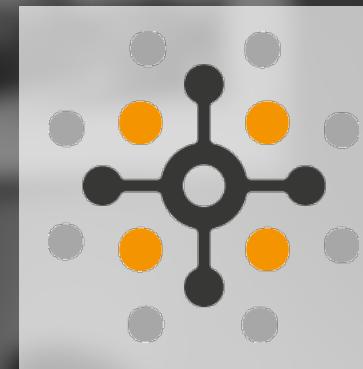
AGENT-BASED MODELLING



This model simulates texts and phone calls over time as they would occur in real life.



WHO
CREATES IT?



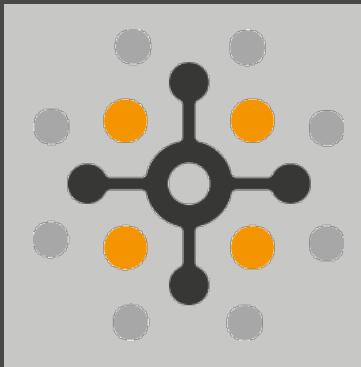
WHO CREATE(S) IT?

DEPENDING ON THE COMPLEXITY OF THE DATA SET

IN-HOUSE
DEVELOPMENT

OR

AD-HOC
DEVELOPMENT

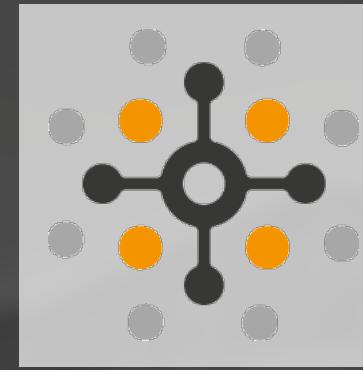


CONCLUSION

SYNTHETIC DATA

CONCLUSION

SYNTHETIC DATA



Your ability to generate realistic synthetic data is essential to developing algorithms and software that will maximize the value of your big data tools, without transgressing privacy laws.

About Us

Real Impact Analytics (RIA) taps into rich telecom data to capture its value. The data is turned into action with big data apps that ease our clients' day-to-day work.

RIA provides guided and predictive analytics through proprietary software. Five of the top ten global telecom operators trust us to enhance customer experience through Customer Value Management, and optimize daily operations with our Commercial Excellence apps.

To learn how Real Impact Analytics can create the same value for you, visit realimpactanalytics.com.



info@realimpactanalytics.com



realimpactanalytics.com



@RIAnalytics



@ReallImpactAnalytics



Real Impact Analytics