

### At a glance

- ▲ ~30Mn rows of synthetic UK bank transactions
- ▲ 20k synthetic individuals with full details
- ▲ 20k synthetic accounts and ~30k fraud events
- ▲ ~70mn rows of synthetic text and call metadata
- ▲ 2 years of data

### Accelerate Innovation by:

- Discovering and developing methods for detection, intervention, and prevention of multiple APP fraud types and behavioural patterns
- Prototyping machine learning, rules based interventions, and large scale data processes on safe data
- Linking complex datasets from siloed environments – Banking, Identity PII, Telco, Fraud – and placing them in the hands of experts with zero regulatory and privacy concerns

### Core data attributes (see metadata for full details)

- ▲ 4 synthetic banks, 2 telco operators
- ▲ 3 banking file types: customers, accounts and transactions
- ▲ 4 telco operator file types: customers, accounts, call metadata, text metadata
- ▲ 20k customer records, with 100Mn combined rows of banking and telco data
- ▲ Name, DOB, Address, email
- ▲ Sort Code, Account Number
- ▲ Account open date
- ▲ Transaction date, ID and description
- ▲ Transaction amount, type, and balance
- ▲ Fraud reports with 6 APP typologies
- ▲ Call and Text ID, time, source, destination, length

### Key features

Synthetic current account customers, with personal details

- 6 Authorised Push Payment Fraud Typologies
- Linked Bank transactions, Account and Identity Information, Text and Call metadata.
- Attempted and Successful fraud reports, across consumers and businesses, for 20,000 data subjects, and 2 years of history
- Zero Privacy Risk
- Wide range of modelled consumer attributes and behaviours including:
  - Real employers and merchants
  - Salaries paid on Weekly and Monthly cadences
  - Mortgages and Rent
  - Household bills, grocery shopping and essential spending: insurance, loans, council tax
  - Telecommunications: Mobile phones, Internet, TV, Streaming
  - Discretionary spending: major e-commerce sites, clothing
  - Texts and Calls biased to be within 'favourites' networks

### Key benefits

- ▲ Our synthetic data is generated from simulations, and contains the same properties as real-world data, removing any privacy issues by generating the data from scratch, meaning no real people, accounts, or transaction are contained in the output
- ▲ Available immediately with no need for Privacy Assessments or other Data Governance activity
- ▲ Can be tailored to meet exact requirements – please get in touch

### Example Metadata of each file type

- accounts

Name	Type	Min	Max	# nulls	Description
customer_id	int	0	9,445	0	Unique identifier for this personal or business customer for the bank (should be the same across all this customer's accounts at the bank)
sort	String	01-01-01	01-01-01	0	The sort code of the bank account
account_number	int	1,000,000	1,009,445	0	The account number of the bank account
open_date	Date	2018-01-01 00:00:00	2018-01-01 00:00:00	0	The date the account was opened. Formatted in CSV as naive ISO8601 date YYYY-MM-DD
type	String	Current	Current	0	Account product type - e.g. current account, credit card, savings
balance	Real	-23,801,841.76	43,079,634.19	0	Balance on the account at the time the extract was written

- transactions

Name	Type	Min	Max	# nulls	Description
tid	int	1,000,000,000,000	1,000,008,642,366	0	Unique Transaction Identifier: globally unique. N.B the counterparty transaction has a different tid
date	date	2022-01-01 00:00:42	2023-12-31 23:59:46	0	The date of the transaction. Formatted in CSV as naive ISO8601 timestamp YYYY-MM-DDTHH:MM:SS, with no timezone information
sort	string	01-01-01	01-01-01	0	The SORT Code of the bank account of the first-party customer
account	string	01000000	01009445	0	The account number of the bank account of the first-party customer
counterparty_sort	string	01-01-01	04-04-04	0	The SORT Code of the bank account of the counterparty if the transaction
counterparty_account	string	01000000	04009223	0	The account number of the bank account of the counterparty to the counterparty of the transaction
narrative	string	& OTHER STORIES	ZUPLEX	0	Narrative description of the transaction (free text)
type	string	BACS	DD	0	Code of the Transaction Type, e.g. DD for Direct Debit
amount	real	-77,578.99	77,578.99	0	Size of the transaction in pounds, with 2 decimal places. Positive for credit transactions on first-party account, negative for debit transactions on first-party account
balance	real	-23,801,841.76	43,079,634.19	0	Balance on first-party account after transaction has occurred. Negative amounts represent overdraft balances

### Example Metadata of each file type

- Personal customers

Name	Type	Min	Max	# nulls	Description
customer_id	string	0000000003	0000009385	0	Unique identifier for this personal or business customer for the bank (should be same across all this customers's accounts at the bank)
citizen_id	string	BG000273A	ZZ999988A	0	An external identifier for the bank customer, typically the national insurance number
dob	date	1907-04-17 00:00:00	2014-03-20 00:00:00	0	The customer's date of birth, formatted as an ISO-8601 date (YYYY-MM-DD)
first_name	string	Aaron	Zoe	0	The forename (first name) of the customer
middle_name	string	Aaron	Zoe	3,478	The middle name/names of the customer (possibly null)
last_name	string	Abbott	Yrooks	0	The surname (last name, family name) of the customer.
title	string	Dr	Professor	147	The title (Mr, Ms etc.) to use when communicating with the customer
sex	string	F	M	0	The gender (sex) of the customer. Currently only supports M and F.
address_line1	string	0 Adams Parkway	Studio 99Y, Andrea Crossing	0	First line of address. Free text, mixed case.
address_town	string	ABERAERON	YORK	0	Second line of address. Nominally a postal town. Upper case.
address_postcode	string	A18 3AJ	ZN2 2SX	0	The post code, nominally in standard form: upper case, with a single space separator and no padding.
land_line	string	+441001300574	+446999378724	0	Land line phone number (if any). Normally starts with +44 for UK and should have no spaces, punctuation etc.
mobile_number	string	+447100513824	+447998887073	5,250	Mobile phone number (if any). Normally starts with +44 for UK and should have no spaces, punctuation etc.
personal_email	string	aaron.barker@hotmail.com	zoe.wilkinson@hotmail.co.uk	0	Email address (home)
work_email	string	aaron.bentley@matalan.co.uk	zoe.wilkinson@smartdatafoundry.com	2,081	Work email address

- Telco personal customers

Name	Type	Min	Max	# nulls	Description
customer_id	string	0000000001	0000029821	0	Unique identifier for this personal or business customer for the operator (should be same across all this customers's accounts at the operator)
citizen_id	string	BG000273A	ZZ999988A	0	An external identifier for the operator's customer, typically the national insurance number
dob	date	1907-04-17 00:00:00	2018-05-25 00:00:00	0	The customer's date of birth, formatted as an ISO-8601 date (YYYY-MM-DD)
first_name	string	Aaron	Zoe	0	The forename (first name) of the customer
middle_name	string	Aaron	Zoe	14,882	The middle name/names of the customer (possibly null)
last_name	string	Abbott	Zell	0	The surname (last name, family name) of the customer.
title	string	Dr	Professor	623	The title (Mr, Ms etc.) to use when communicating with the customer
sex	string	F	M	0	The gender (sex) of the customer. Currently only supports M and F.
address_line1	string	0 Abdul Causeway	Studio 99V, Lees Junctions	0	First line of address. Free text, mixed case.
address_town	string	ABERAERON	YSTRAD MEURIG	0	Second line of address. Nominally a postal town. Upper case.
address_postcode	string	A18 3AJ	ZL50 3RQ	0	The post code, nominally in standard form: upper case, with a single space separator and no padding.
land_line	string	+441000006814	+446999378724	0	Land line phone number (if any). Normally starts with +44 for UK and should have no spaces, punctuation etc.
mobile_number	string	+447100513824	+447999882778	22,380	Mobile phone number (if any). Normally starts with +44 for UK and should have no spaces, punctuation etc.
personal_email	string	aaron.bentley@gmail.com	zoe.wright@hotmail.co.uk	0	Email address (home)
work_email	string	aaron.bentley@matalan.co.uk	zoe.wright@calderdalemetropolitanboroughcouncil.in fo	9,011	Work email address

- Telco personal accounts

Name	Type	Min	Max	# nulls	Description
customer_id	string	0000000000	0000034542	14,912	Unique identifier for this customer for the operator (should be same across all this customers's phone accounts/numbers at the operator)
phone_number	string	+441000514060	+447999882778	0	The Phone Number for this account.
account_number	string	01000000	01034542	0	The account number of phone account
type	string	land-line	mobile-pre-paid	0	Account type – current land-line, mobile-contract or mobile-prepaid
operator_number	string	01	01	0	Unique identifiers for this operator

- Telco personal text message metadata

Name	Type	Min	Max	# nulls	Description
tmrid	int	1,000,000,000,000	1,000,013,876,528	0	Unique Identifier for call. Globally unique. N.B. the counterparty CDR has a different tmrid
time	date	2022-01-01 00:00:04	2024-01-01 00:20:44	0	The time the text message was sent. Formatted in CSV as naive ISO8601 timestamp YYYY-MM-DDTHH:MM:SS, with no timezone information
origin_number	string	+447100093721	+447999882778	0	The number from which the text message was sent. Includes +, country code and no spaces or punctuation.
destination_number	string	+447100093721	+447999882778	0	The number to which destination number the text message was received. Includes +, country code and no spaces or punctuation.
parts	int	1	8	0	Number of 160-character parts into which the text message is divided.
cost	real	0.01	0.04	7,291,628	Nominal cost of the message in pounds sterling (£). Not charged if in bundle.
bytes	int	1	1,278	0	Size of message in bytes.
MMS_attachments	int	0	0	7,291,628	Number of multimedia attachments
texts_balance	int	0	0	7,291,628	Text messages remaining in bundle, if applicable

- Telco personal call metadata

Name	Type	Min	Max	# nulls	Description
cdrid	int	1,000,000,000,000	1,000,024,675,736	0	Unique Identifier for call. Globally unique. N.B. the counterparty CDR has a different cdrid
start_time	date	2021-12-31 23:01:54	2024-01-01 00:19:01	0	The start time and date of the call. Formatted in CSV as naive ISO8601 timestamp YYYY-MM-DDTHH:MM:SS, with no timezone information
stop_time	date	2021-12-31 23:01:56	2024-01-01 00:27:29	0	The end time and date of the call. Formatted in CSV as naive ISO8601 timestamp YYYY-MM-DDTHH:MM:SS, with no timezone information
origin_number	string	+441000006814	+447999882778	0	The number from which the call was initiated. Includes +, country code and no spaces or punctuation.
destination_number	string	+441000006814	+447999882778	0	The number upon which the call was received. Includes +, country code and no spaces or punctuation.
duration	int	1	1,928	0	The number of seconds the call lasted.
minutes	int	1	33	0	The length of the call in whole-or-partial minutes
charge_basis	string	A	B	11,104,571	Indicator of charge bad. Usually a single letter or null.
cost	real	0.00	5.60	11,104,571	Nominal cost of the call in pounds sterling (£). Not charged if in bundle
minutes_balance	int	0	0	0	Number of minutes remaining on plan (if tracked)

- Fraud reporting metadata

Name	Type	Min	Max	# nulls	Description
time_stamp	date	2022-01-01 00:17:12	2023-12-31 23:48:51	0	The time and date of a fraudulent transaction took place. Formatted in CSV as naive ISO8601 timestamp YYYY-MM-DDTHH:MM:SS, with no timezone information
kind	string	Bank Impersonation Scam	Shopping Scam	0	Kind of fraudulent scam which failed and successful scams, e.g. Shopping Scam or Failed Delivery Scam
origin_number	string	+441097713082	+447999705310	4,151	The number from which the call was initiated (scammer number). Includes +, country code and no spaces or punctuation.
destination_number	string	+447100093721	+447999882778	4,151	The number upon which the call was received (victim number). Includes +, country code and no spaces or punctuation.
fraudster_id	string	ALLTAN	eCourier.co.uk	0	An external identifier for the bank customer, typically the national insurance number or the fraudulent organization
tid	int	1,000,000,000,479	4,000,008,008,734	12,988	Unique Transaction Identifier: globally unique. N.B. the counterparty transaction has a different tid
victim_account_number	string	01000004	04005033	12,739	The victim account number of bank account
victim_sort_code	string	01-01-01	04-04-04	12,739	The victim SORT Code of the bank account
fraudster_account_number	string	01000008	04005161	12,739	The fraudster account number of bank account
fraudster_sort_code	string	01-01-01	04-04-04	12,739	The fraudster SORT Code of the bank account