EMERGE

THE SOLUTIONS
TO THE WORLD'S
BIGGEST PROBLEMS
ARE IN THE DATA

There are mountains of data but...

"Only 0.5% of all data is analysed or used"

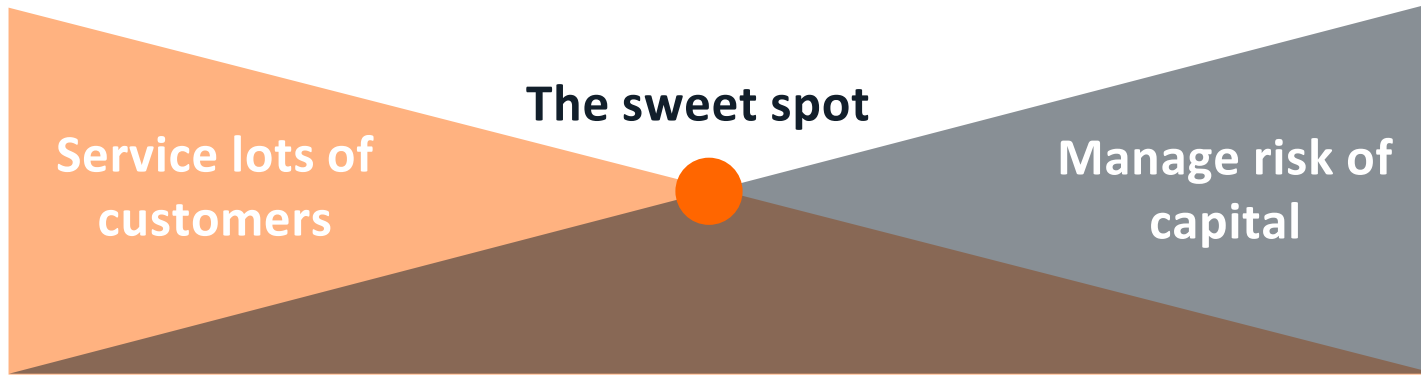Bernard Marr, Forbes.com

The data landscape

Purchased data

Internal data

Low availability

High availability

Web browsing

Consumer bureau

Geolocation

Merchant acquiring

Business bureau

Weather

Government

Economic

Facebook

Crime data

Twitter

Election results

Demographic

Other products

Transaction flows

Application

Process data

System logs

HR data

Connected devices

ERP

Purchases

Smartphone / feature phone and app usage

Digital behaviour

Public data

Data from partners

How do banks traditionally approach this?

The rules-based approach where everyone is treated the same

1 Put in rule-based controls to manage various risks

2 Controls result in a frustrating customer experience

4 Reactively try save the good clients

5 It is too late to keep the frustrated good clients

6 Find new good clients in the market

7 Client is put through risk averse entry processes

3 Good paying customers get frustrated and go elsewhere

4 We are left with the worse paying customers

5 The book's overall performance gets worse
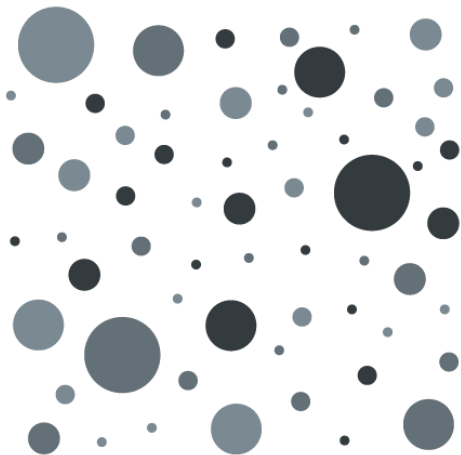
The new way of managing risk...
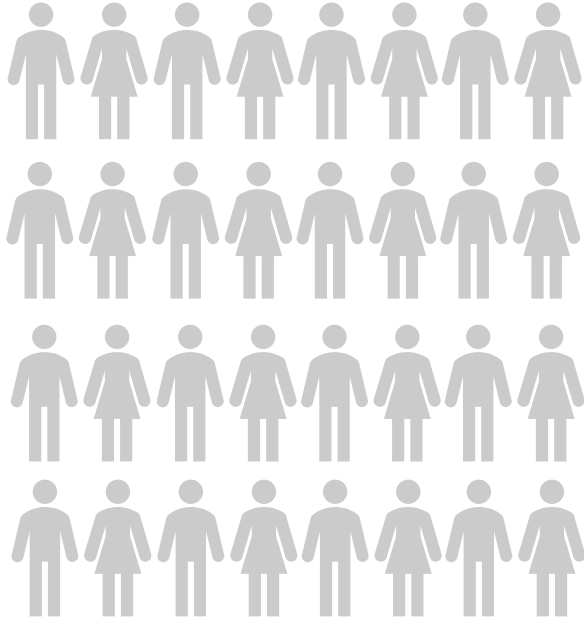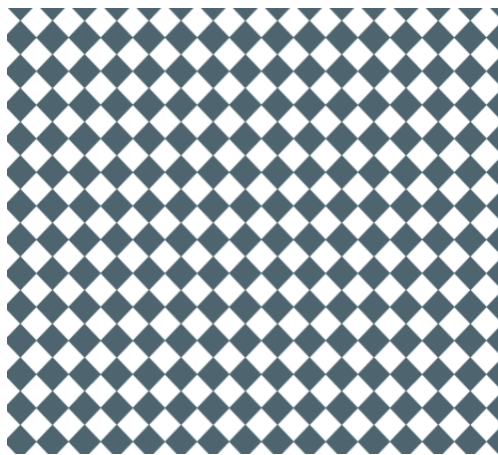
No more generic rule-based segmentation

Use machine learning to enable scalable personalised interactions

How is this done?

Provide **your data** in whatever amount and format you have it - don't spend time cleaning it first

Find unique patterns to identify pockets of value using **Emerge's proprietary machine-learning** software
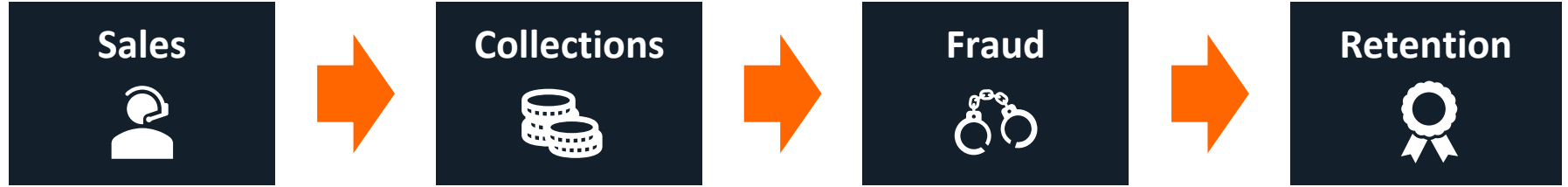

Run the model in your operational systems through an **API or batch process in** the Azure cloud

There is opportunity to personalise the full customer lifecycle

For example, let's look at a few case studies

Sales → **Collections** → **Fraud** → **Retention**

CROSS SELL

Aim
Identify the propensity for a lead to take up a lending product

Input data
Data from previous campaigns and new potential leads

25 000
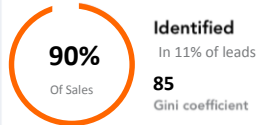Leads were implemented into the campaign
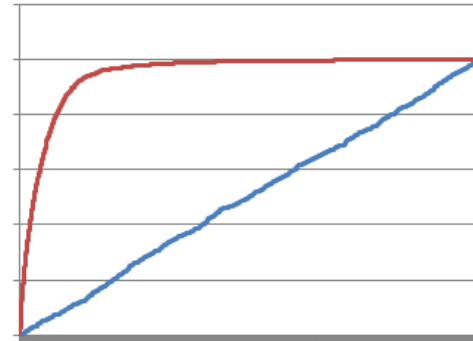
4 hours
To run the model

7
data variables

- Demographics

Model results

90%
Of Sales

Identified
In 11% of leads

85
Gini coefficient

Achieved 8x better productivity - Imagine finishing the week's dialing on Monday afternoon at tea time

COLLECTIONS

Aim

To determine which credit card clients entering collections would cure on their own within 5 days

Input data

Each account that entered collections in a year was included

1.6 million
Accounts that entered collections

9 hours
To run the model

64
data variables

- Account descriptors
- Collections data
- Demographics
- Client descriptors
- Bureau
- Affordability
- Contact details
NOTE: no transactions were used

Model results

Predict	Result	
	0	**1**
0	581 815	149 851
1	173 331	545 003

73 Gini coefficient

Perfect
(Gini of 100)

Emerge
(Gini of 73)

Random
(Gini of 0)

The steeper the curve,
the more predictive the model

**The model accurately predicted
80% of clients that cured in 5 days and
76% of clients that didn't cure in 5 days
without any transactional data**
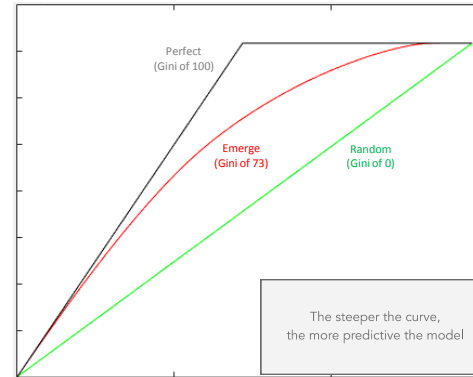
FRAUD

Aim

Identify which suspicious credit-card transactions were actually fraudulent

Input data

All data relating to transactions flagged as suspicious (based on a rule based assessment)

225 000

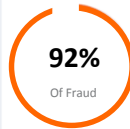suspicious transactions reviewed of a total 72 000 000 total transactions

6 hours

To run the model

120
data variables

- Account details
- Demographic details
- Credit-bureau details
- Transaction features

Model results

92%

Of Fraud

Identified
In 33% of fraud flags which is 0.1% of total transactions

Gini coefficient of **89**

Perfect
(Gini of 100)

Emerge
(Gini of 89)

Random
(Gini of 0)

The steeper the curve, the more predictive the model
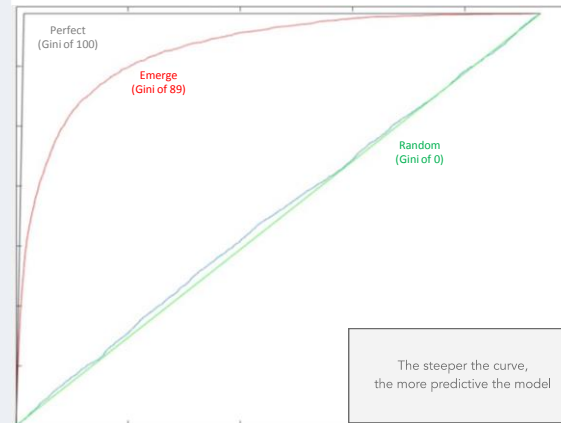
Avoided need to hire 20 extra fraud-assessment agents

RETENTION

Aim
Identify transactional accounts that will be cancelled in the next three months

Input data
All transactional accounts with a particular product type
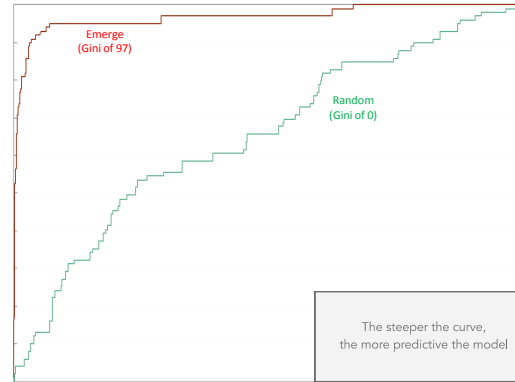
2 million
Transactional accounts

8 hours
To run the model

Data variables
- Account holder details
- Account descriptors
- Transaction details

Model results

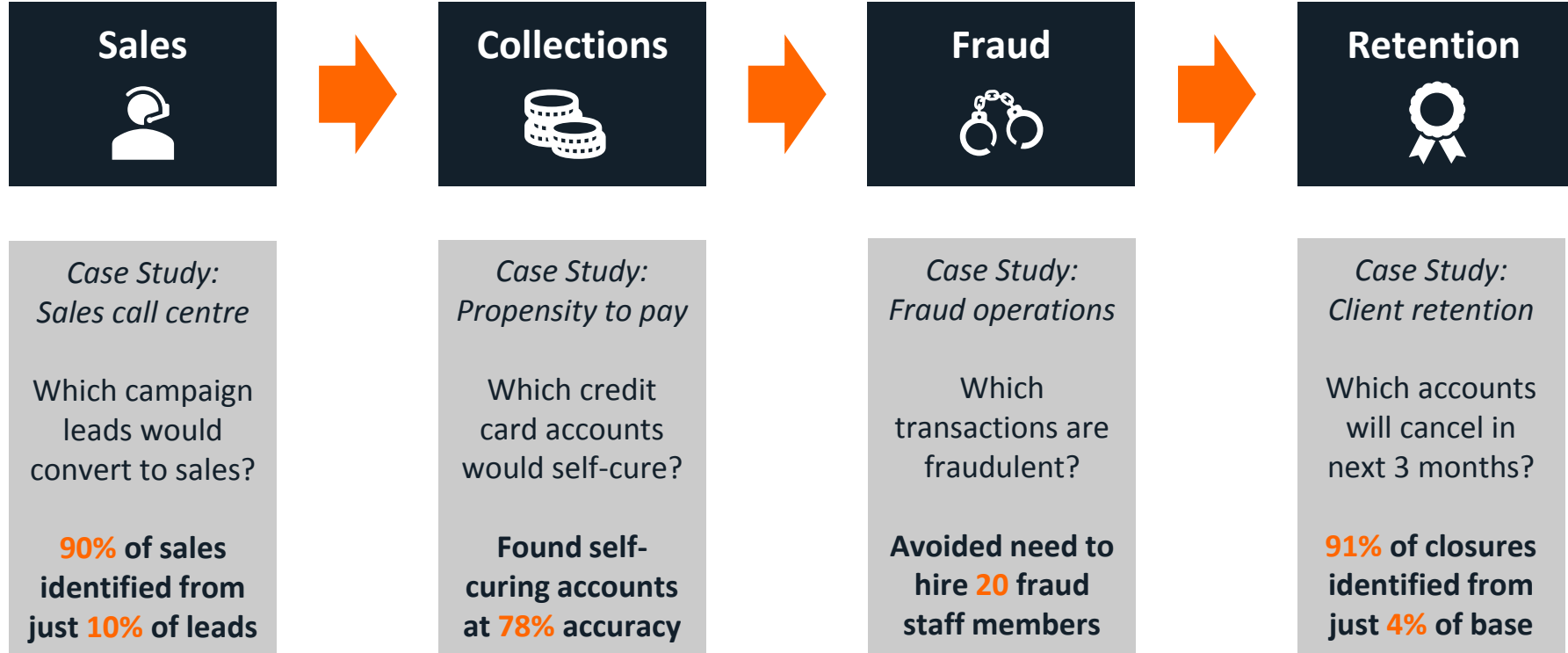**91%**
Of cancellations

Identified
In 4% of base

97
Gini coefficient
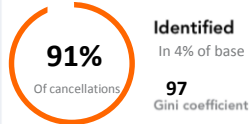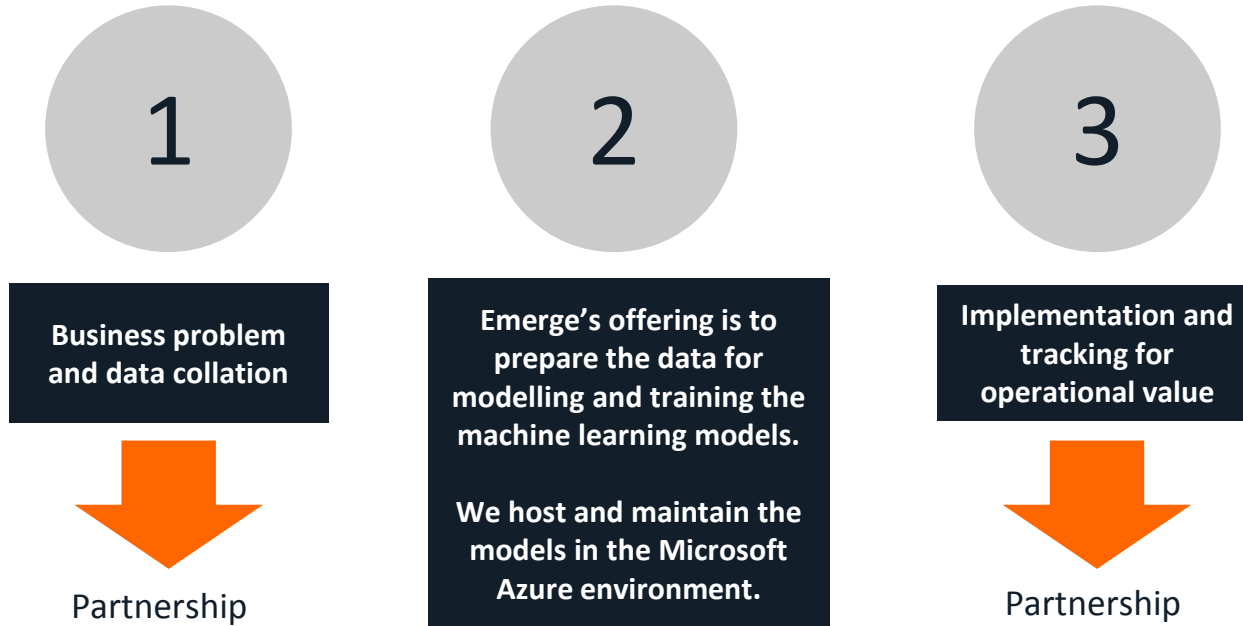
Once you know who will leave you can contact those clients and solve the issues that might have resulted in a cancellation

Emerge
(Gini of 97)

Random
(Gini of 0)

The steeper the curve,
the more predictive the model

There is opportunity to personalise the full customer lifecycle

For example, let's look at a few case studies

Sales	**Collections**	**Fraud**	**Retention**

Case Study:
Sales call centre

Which campaign leads would convert to sales?

90% **of sales identified from just** **10%** **of leads**

Case Study:
Propensity to pay

Which credit card accounts would self-cure?

Found self-curing accounts at **78%** **accuracy**

Case Study:
Fraud operations

Which transactions are fraudulent?

Avoided need to hire **20** **fraud staff members**

Case Study:
Client retention

Which accounts will cancel in next 3 months?

91% **of closures identified from just** **4%** **of base**

The way we envisage a partnership with you…

1

Business problem and data collation

Partnership

2

Emerge's offering is to prepare the data for modelling and training the machine learning models.

We host and maintain the models in the Microsoft Azure environment.

3

Implementation and tracking for operational value

Partnership

Microsoft

EMERGE

info@emergeml.com

www.emergeml.com