Alternative Credit Scoring

Mohan Jayaraman
MD Decision Analytics, Business Information
Experian Asia Pacific

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Experian – An introduction

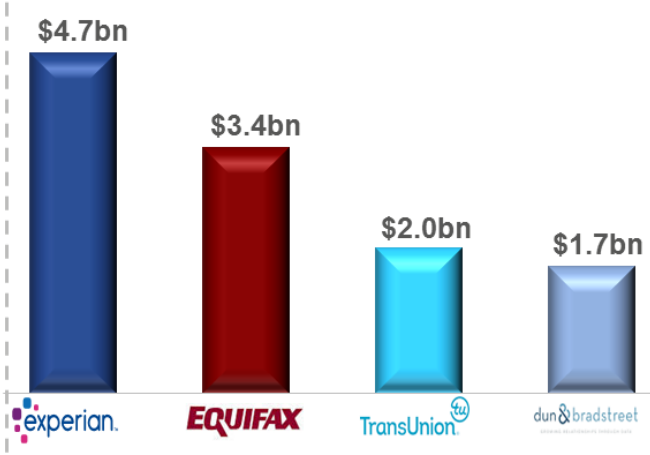
Our Business

- Credit Services
- Decisioning & Analytics
- Consumer Services
- Data Quality
- Audience Targeting

11 Business Credit Bureaus Globally

18 Consumer Credit Bureaus Globally

Market leader with unparalleled reach globally



5 YEARS RUNNING
Experian named one of *Forbes* top 100 most innovative companies

Powering opportunities globally

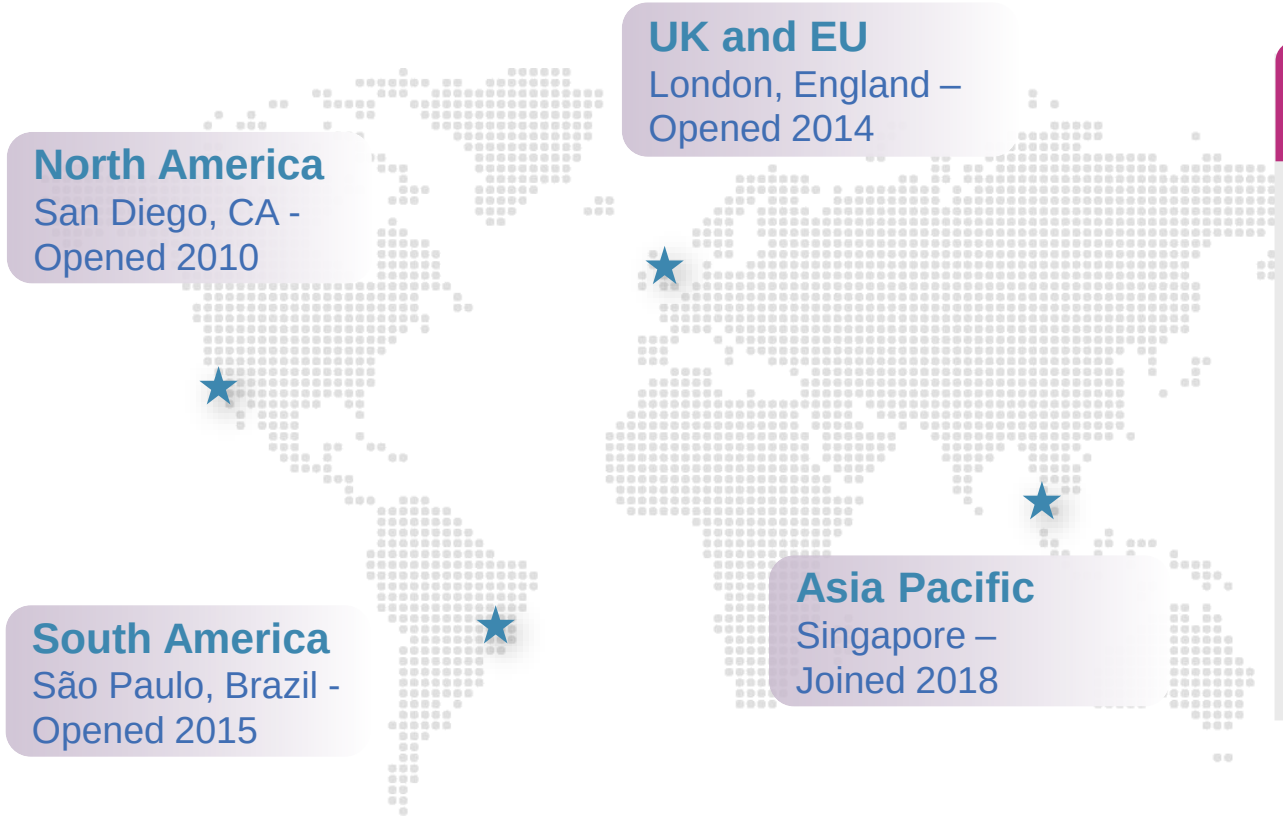
- **17,000** employees in **37 countries**
- **12,000+ clients** around the globe
- Diverse mix of expertise in **40+ countries**
- One of **Forbes' "World's Most Innovative Companies" 5th year**
- More than **300+** issued and pending patents
- **Data Labs dedicated scientists and machine-learning experts**
- **7.8 million** conversations with people, including education on how debt affects credit scores
- Supported **45,000** fraud victims in the past year
- Social innovation products have helped **2.7 million** people access essential services

Argentina	Costa Rica	Italy	Philippines	Turkey
Australia	Denmark	Japan	Poland	U.A.E
Austria	France	Malaysia	Russia	UK
Brazil	Germany	Mexico	Singapore	United States
Bulgaria	Greece	Monaco	South Africa	Venezuela
Canada	Hong Kong	Netherlands	South Korea	Vietnam
Chile	India	New Zealand	Spain	
China	Indonesia	Norway	Taiwan	
Colombia	Ireland	Peru	Thailand	

Experian DataLabs

Research and development, innovation and incubation

~100 data scientists, technologists and industry experts focused on innovation



Purpose

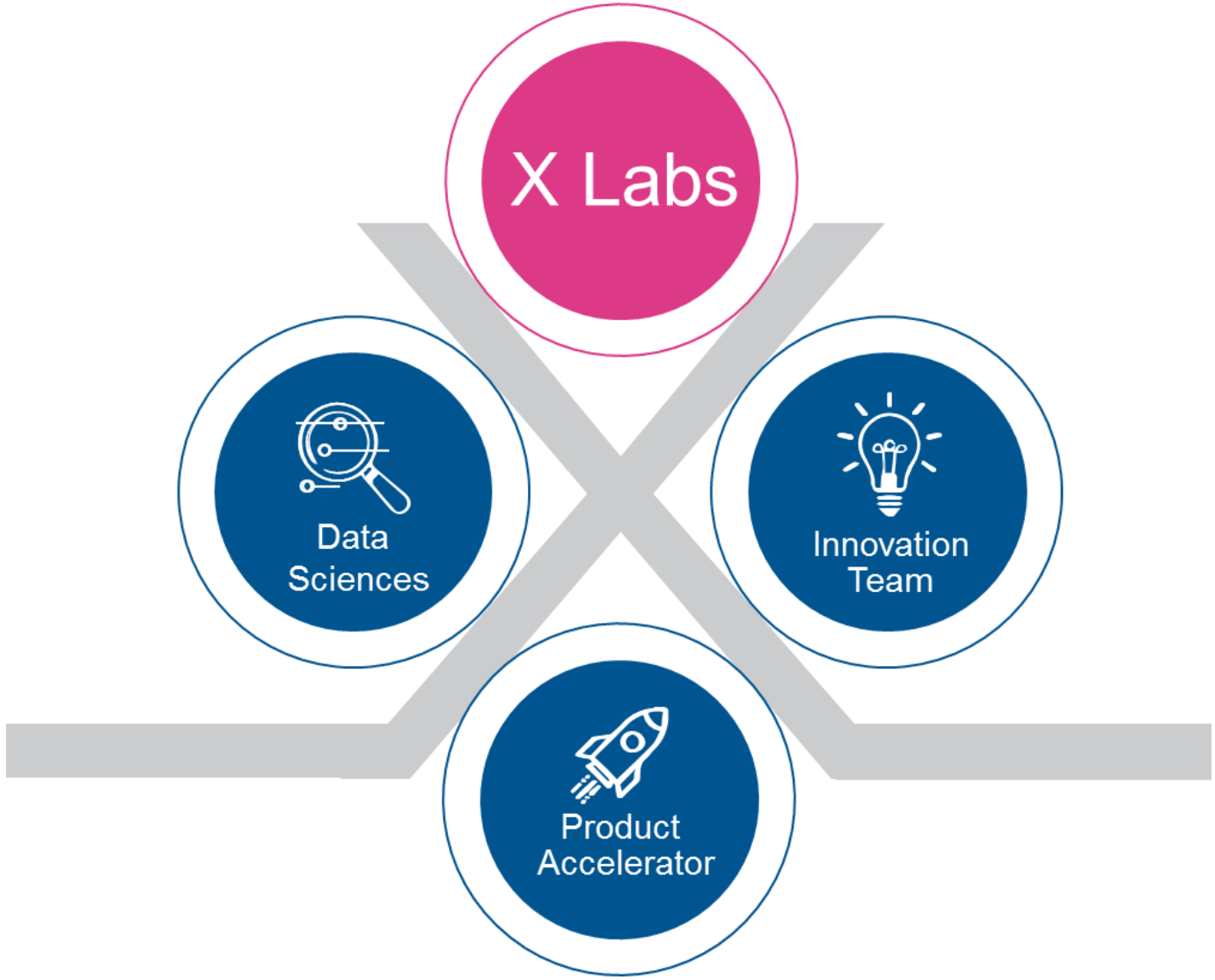
- Innovation with new data sources, advanced analytics and new technologies
- Emphasis on financial services, retail, telecom and healthcare
- Safe and secure environment to innovate with Experian and partner data assets

Approach

- Hypothesis driven problem solving
- Direct client engagement
- Tackling new, currently unsolved industry and client issues

Experian DataLabs

Team composition



Experian DataLabs

Team competencies

Artificial Intelligence/ Machine Learning

Explainable Artificial Intelligence

Advanced data science/ non-linear modeling

New technologies / tech stack and architecture

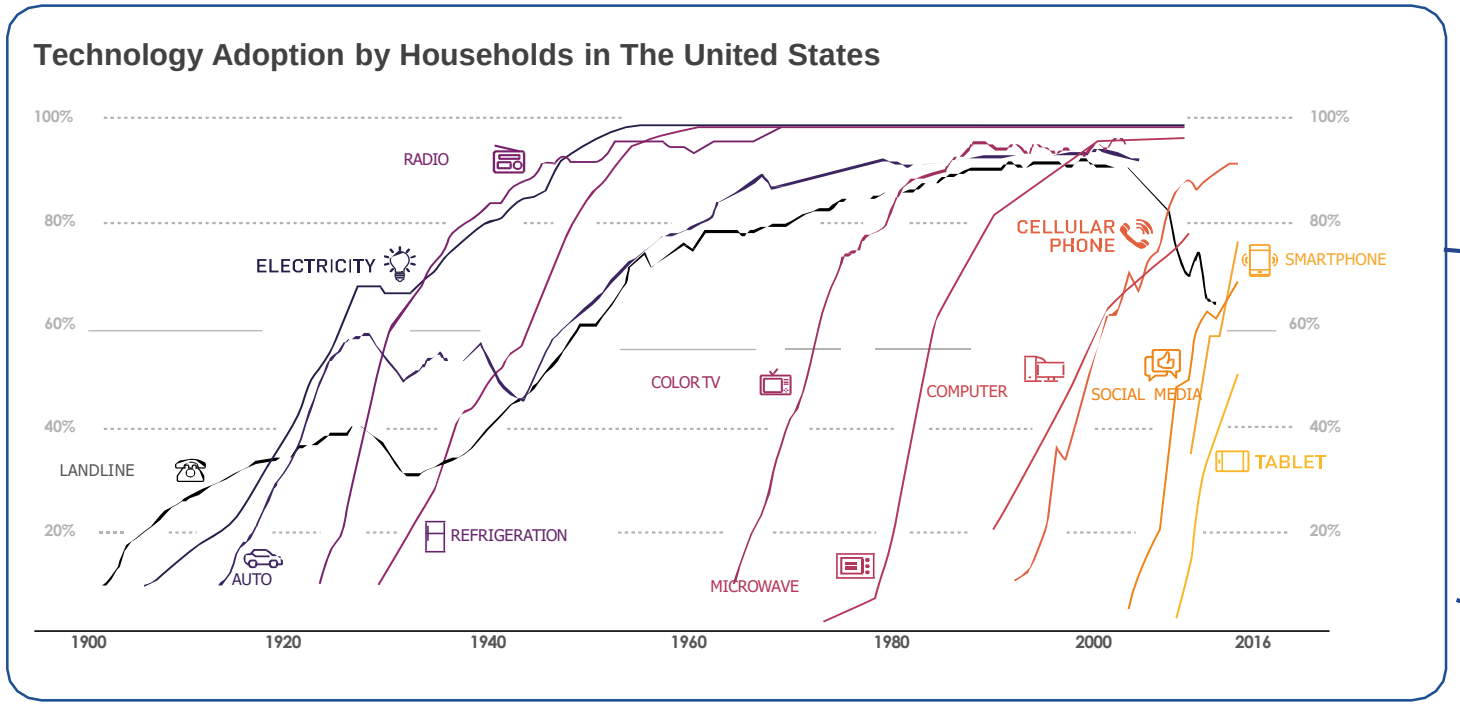
New data source integration

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Consumer preferences are changing

Faster adoption cycles seem the rule – Digital natives are here to stay



Consumer preferences changing

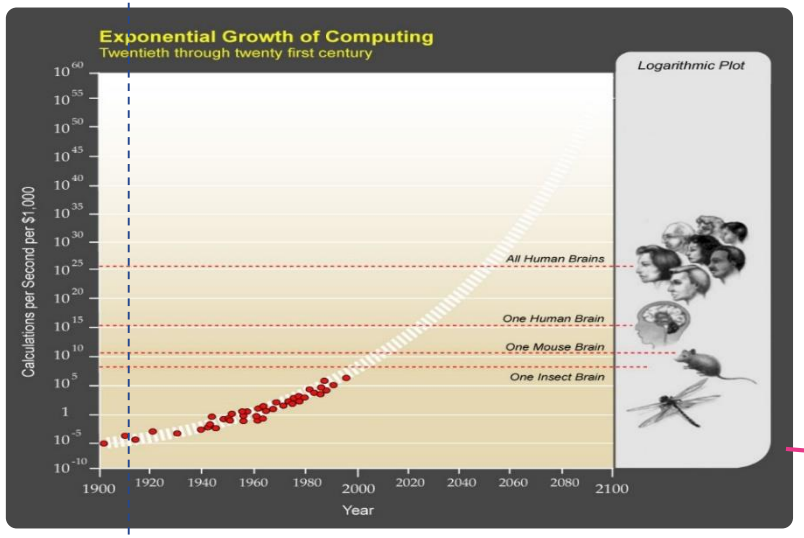
…leading to increased tech adoption rate

Aided by the growth in technology capability

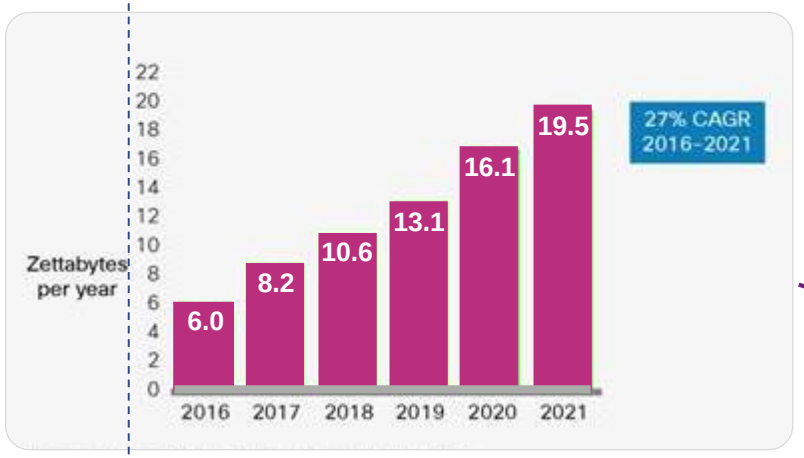
“2.5 quintillion bytes of data created each day…” - IBM

The IBM Model 350 disk file a storage space of 5MB from **1956**

Sandisk Ultra micro SD storage space of 128GB from **2019**



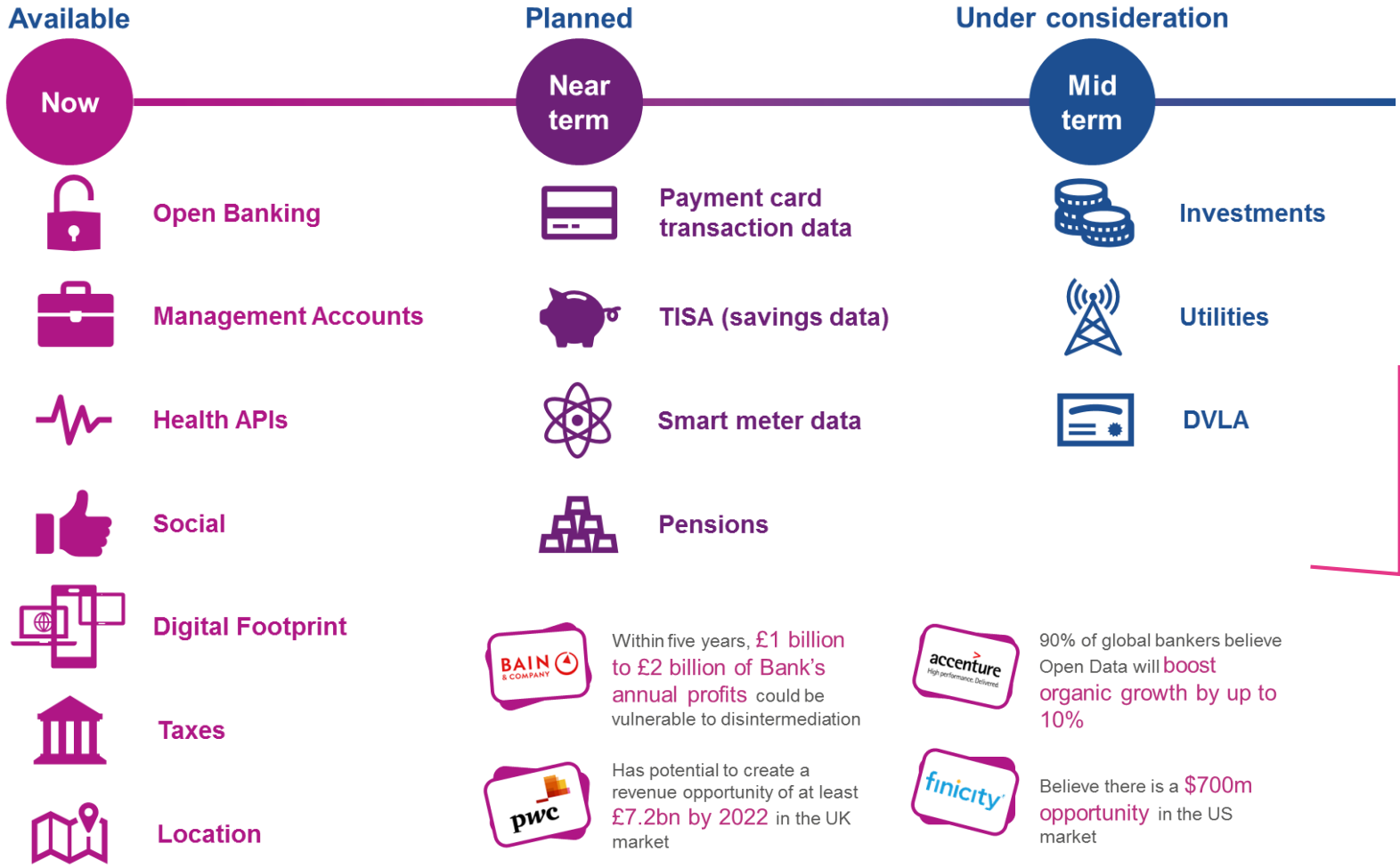
Exponential growth of computing power



…and cloud processing capacity

And Open data regulation is adding to the data explosion

Open data will profoundly impact our core markets



Consumer data management amongst the top regulatory focus areas

"Open Banking: A Wave Bigger Than The Internet And Mobile" Forbes

"Why the future of Data is Open' " WALL STREET JOURNAL

"An earthquake in European banking" The Economist

"Open banking a 'revolution' to challenge banks' dominance' " BBC NEWS

BAIN & COMPANY — Within five years, £1 billion to £2 billion of Bank's annual profits could be vulnerable to disintermediation

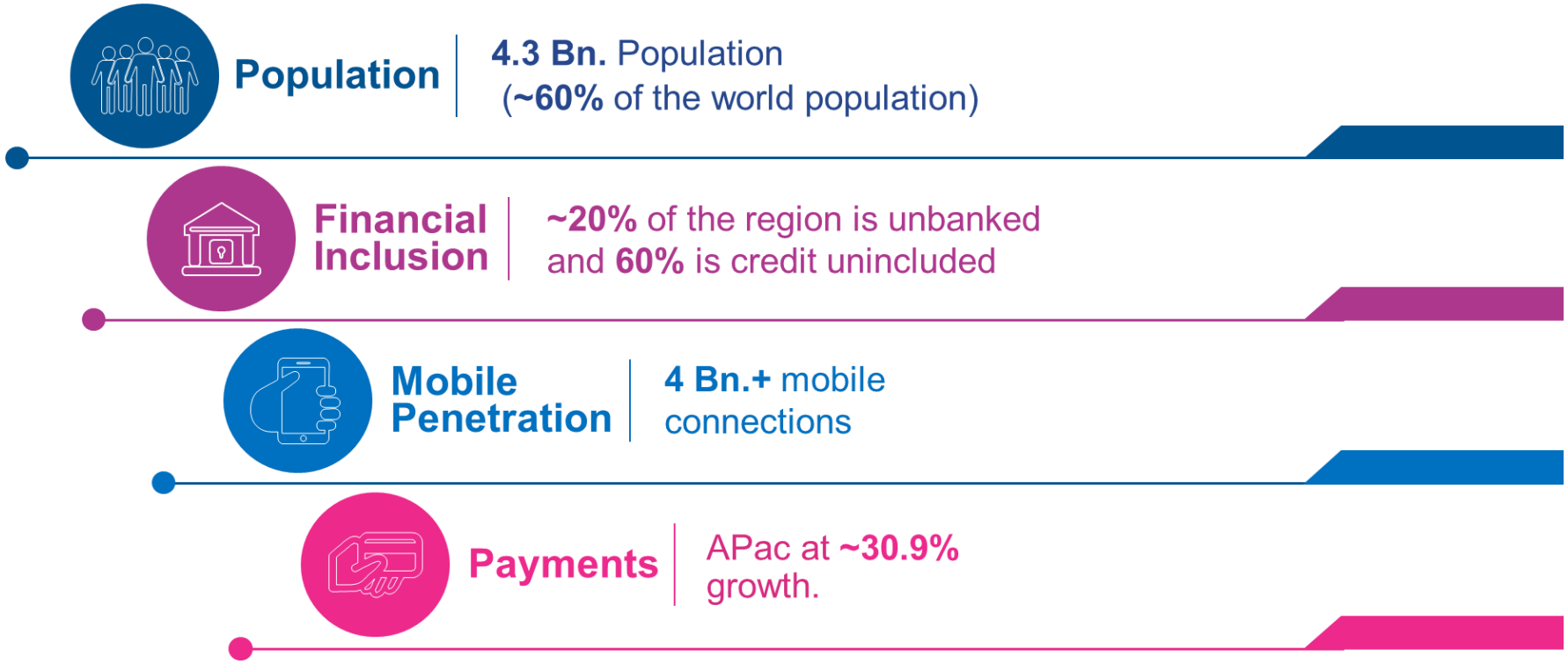
pwc — Has potential to create a revenue opportunity of at least £7.2bn by 2022 in the UK market

accenture — 90% of global bankers believe Open Data will boost organic growth by up to 10%

finicity — Believe there is a $700m opportunity in the US market

All of this is even more relevant in APac

Most markets in APac are moving with consumer preferences and leap frogging tech generations



…Alternate data in APac is actually data…

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Multiple components needed to bring alt data models to life

Consumer consent and collaboration based models will be the de-facto standard in the new world

Clients need to be able to make faster, more reliable decisions and with deeper insight than ever before

Data

New data in structured and unstructured form needs to be constantly evaluated to keep updating the organizational view of

Analytics techniques

Advanced analytics and machine learning combined with a wide range of accessible data can aid with problem solving as well as improving products, services and customer insight.

Platforms

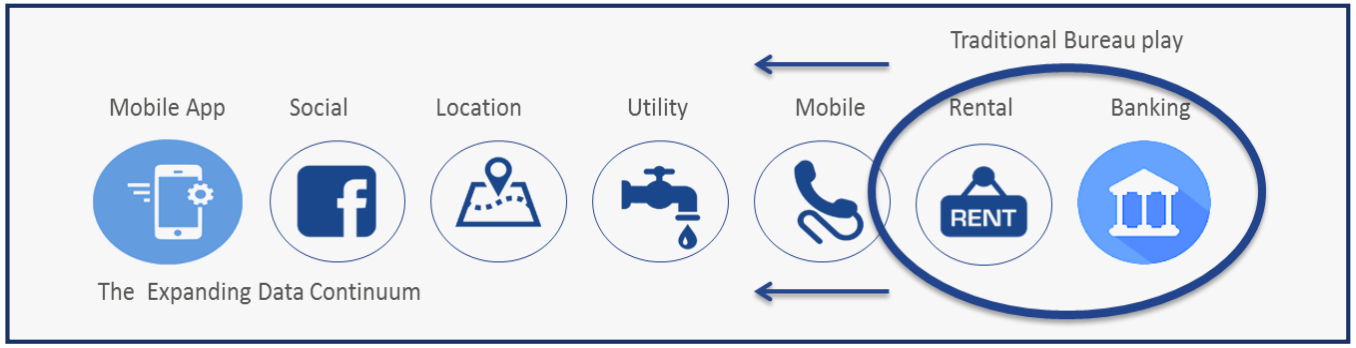
Many businesses are increasing their investment for machine learning and advances analytics however many are still struggling to transform it into insight.

Collaborative business models

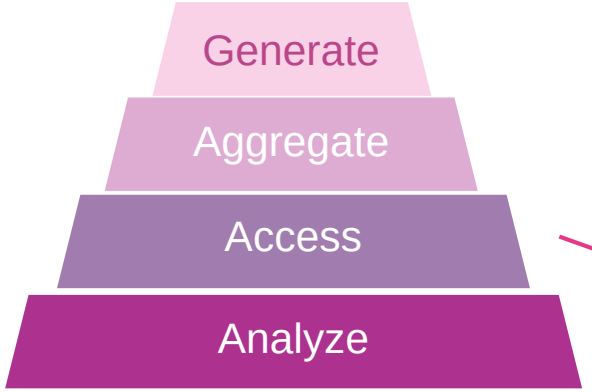
Clients who understand how competitors are using data can begin to see where and why they are taking the competitive edge. This then allows them to respond accordingly.

Traditional data usage models focused on ownership dying out..

Consumer consent and collaboration based models will be the de-facto standard in the new world



Data ownership is getting disaggregated



And Access to data is becoming more important than ownership

The alternate data universe is growing

Telco data

- Mobile usage
- Data usage
- Contact information
- Top up
- Roaming information
- GPS location
- ID check

Travel data

- Travel pattern
- Income proxy
- Demographic profile
- Customer segment
- Contact information
- ID check

Payments data

- Travel pattern
- Spends
- Income proxy
- Demographic profile
- Customer segment
- Contact information
- ID check

Ecommerce data

- Spend pattern
- Income proxy
- Demographic profile
- Contact information
- ID check

Non Bank

- Payment behaviour
- Debt consolidation
- Asset holdings
- Contact information
- ID check

PF/Tax payments

- Spend pattern
- Income proxy
- Demographic profile
- Contact information
- ID check

Market place

- Leda generation
- Credit assessment
- Id and verification

Insurance data

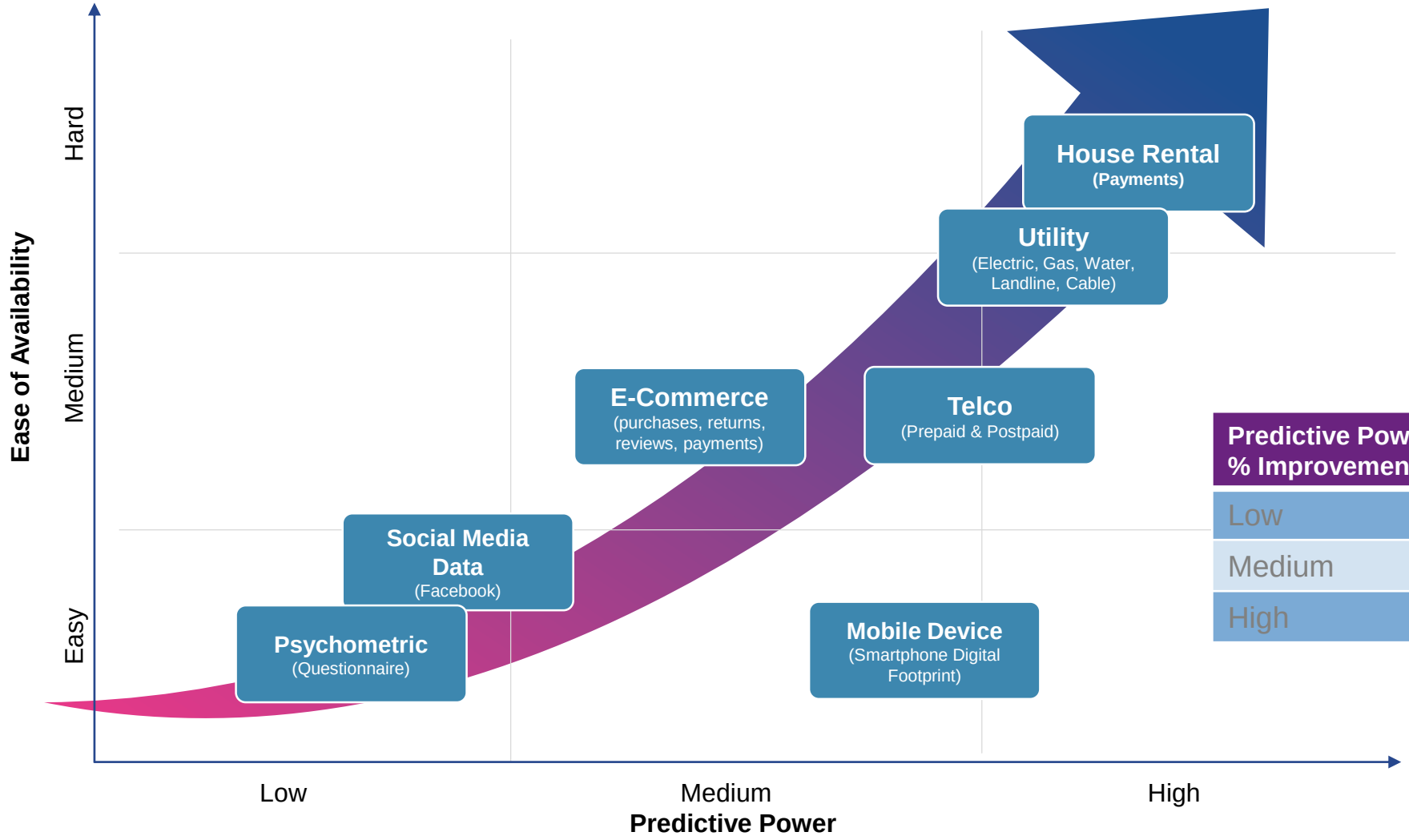
- Asset holdings
- Premium payment behaviour
- Contact information
- ID check

Other utilities

- Income proxy
- Monthly usage
- Contact information
- ID check

But not all data is born equal

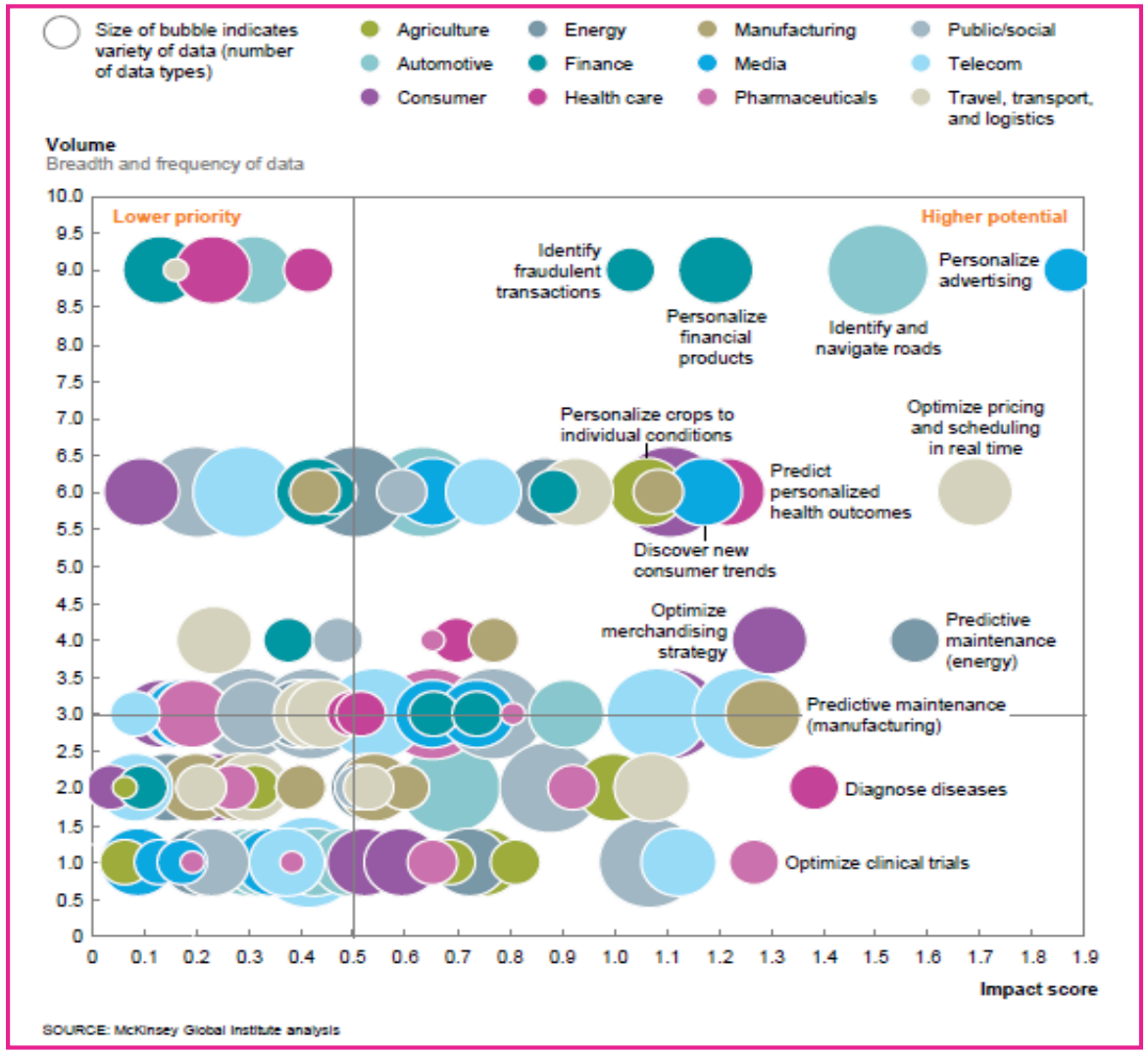
The right data can provide life of upto 40%+ in Gini



Predictive Power: % Improvement in Gini for Thin File Cases	
Low	Up to 20%
Medium	20% - 40%
High	40% +

Use cases of Scoring are growing in the digital world

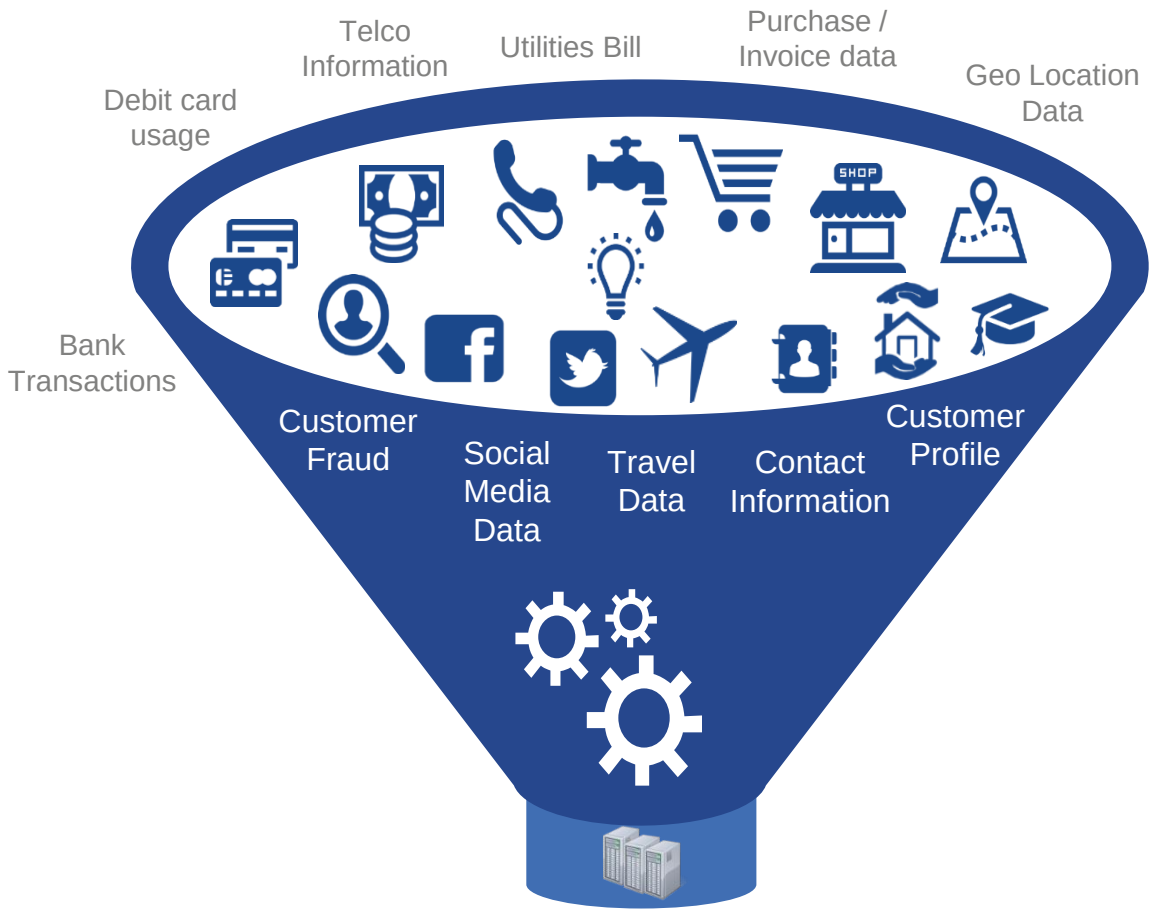
Multiple ways of viewing the consumer necessary given new revenue models



Credit segregation is not the only problem to solve in the current world

Other use cases for alternate scoring

Consumer 360 profiles….

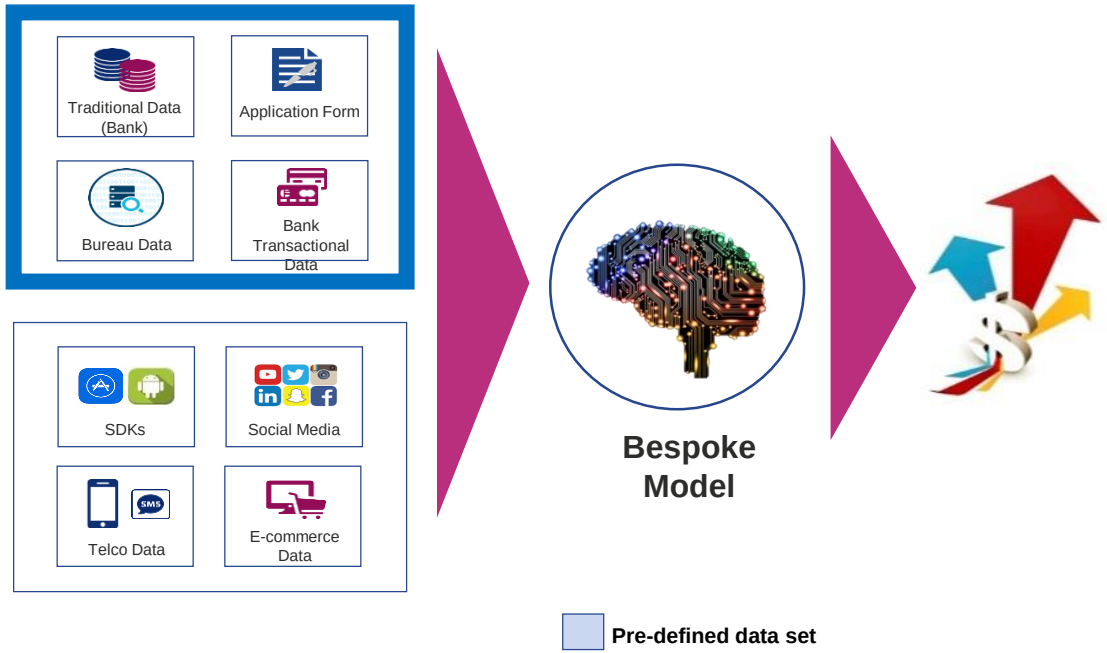


- **Unified risk scoring for credit**
- **Income assessment**
- **Socio-economic Profiling**
- **Identity verification**
- **Fraud detection**
- **And many more use cases**

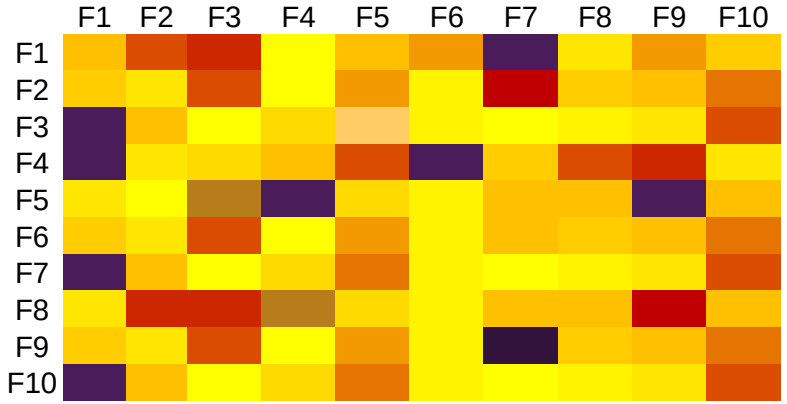
Modelling with ML for lift

Apply Machine Learning

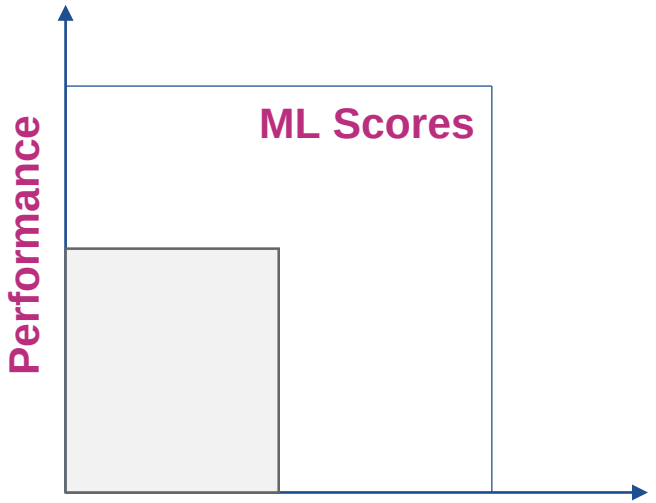
- **ML scores have stronger predictive power**, allowing for lower risk and/or higher acceptance
- ML scores allow to use non standard data sources which can further raise predictive power



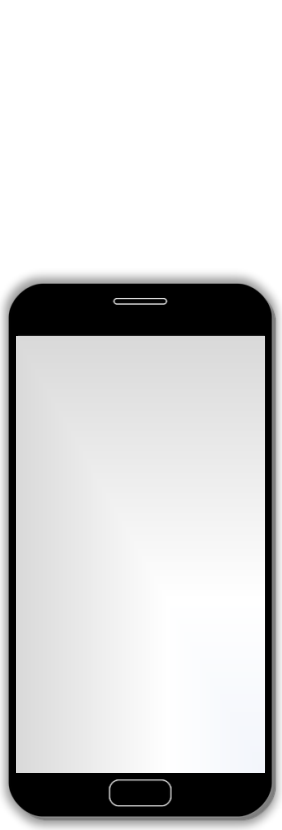
Machine Learning – Micro Segments



Similar colored cell indicates similar population bucket



Modelling with ML for micro segmentation



ML to model variables to pick micro patterns

Demographics

Calling Behavior

Data Usage

Geo Location

Social Network

Data Usage

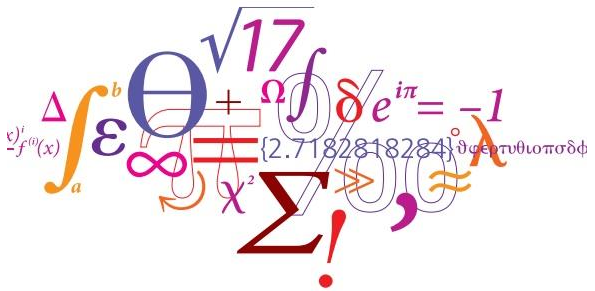
SMS

Apps Usage

Top-Ups

VAS

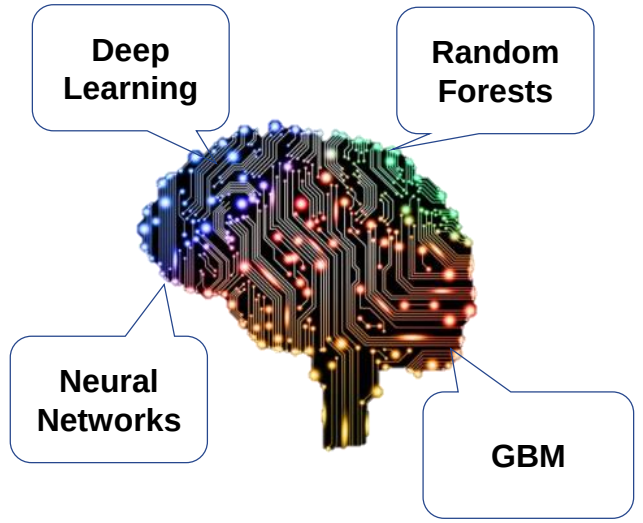
Feature Engineering



High dimension Feature Space

1500+ Transformed Features

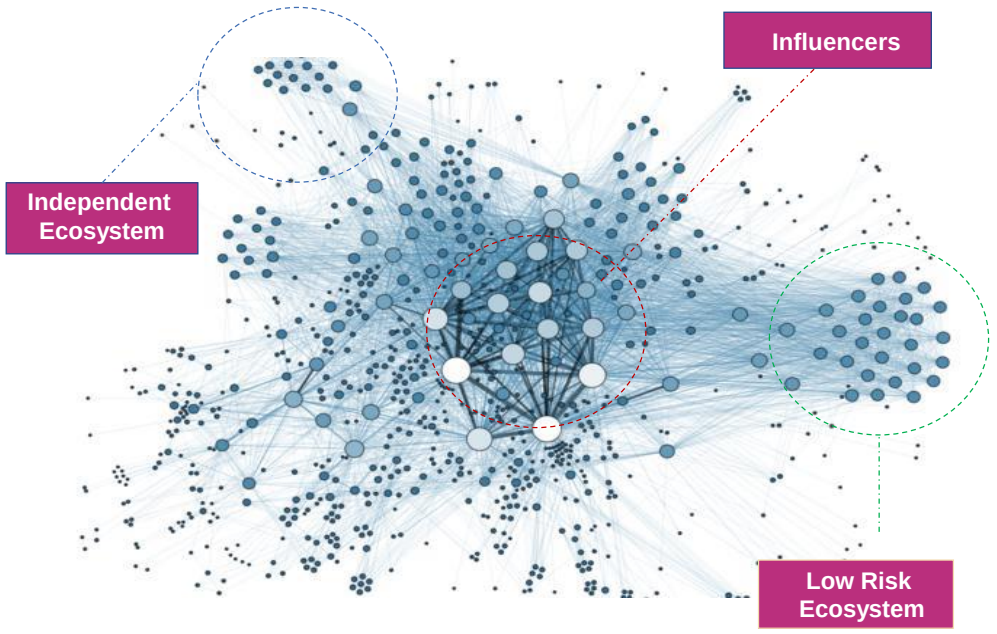
Machine Learning Models



- Boost conversion with ML and pre-approved leads
- Optimized decision system

Network analysis based models

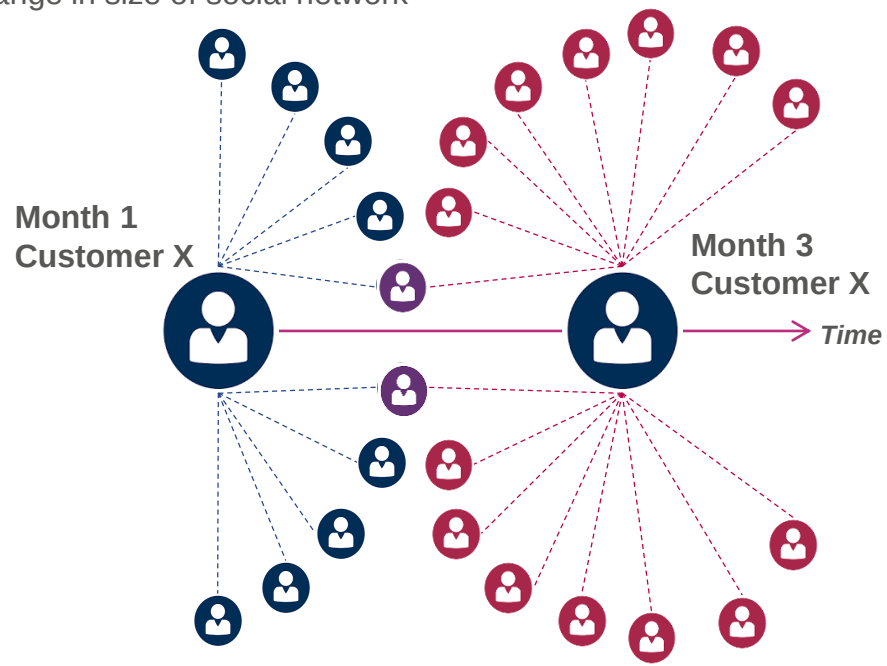
Social Network



Illustrative Image.
Source : Network World :

Change in Social Network

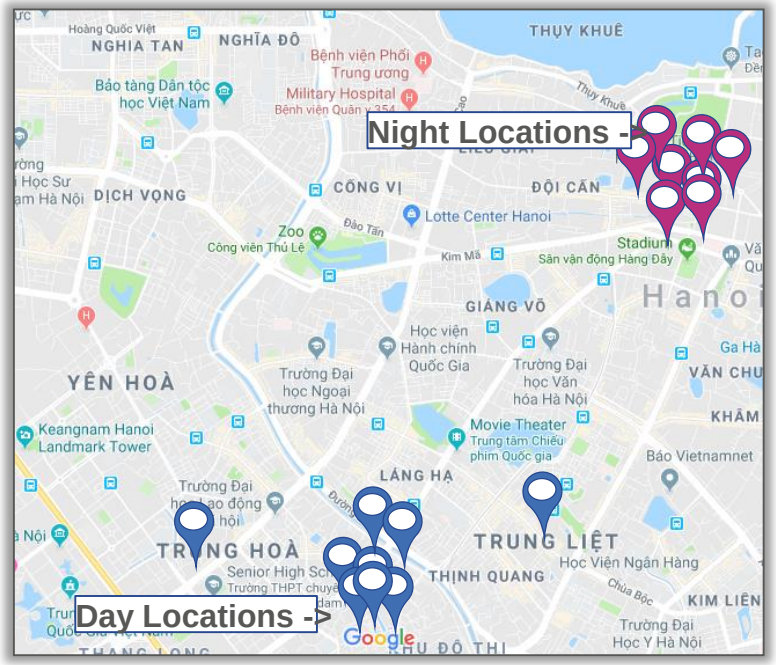
Overlap between social network in Month 1 as compared to Month 3
Change in size of social network



Using Sensitive information by finding data patterns

Customer Mobility

The mobility of a customer across weekdays / weekends indicate the type of employment and vehicle ownership

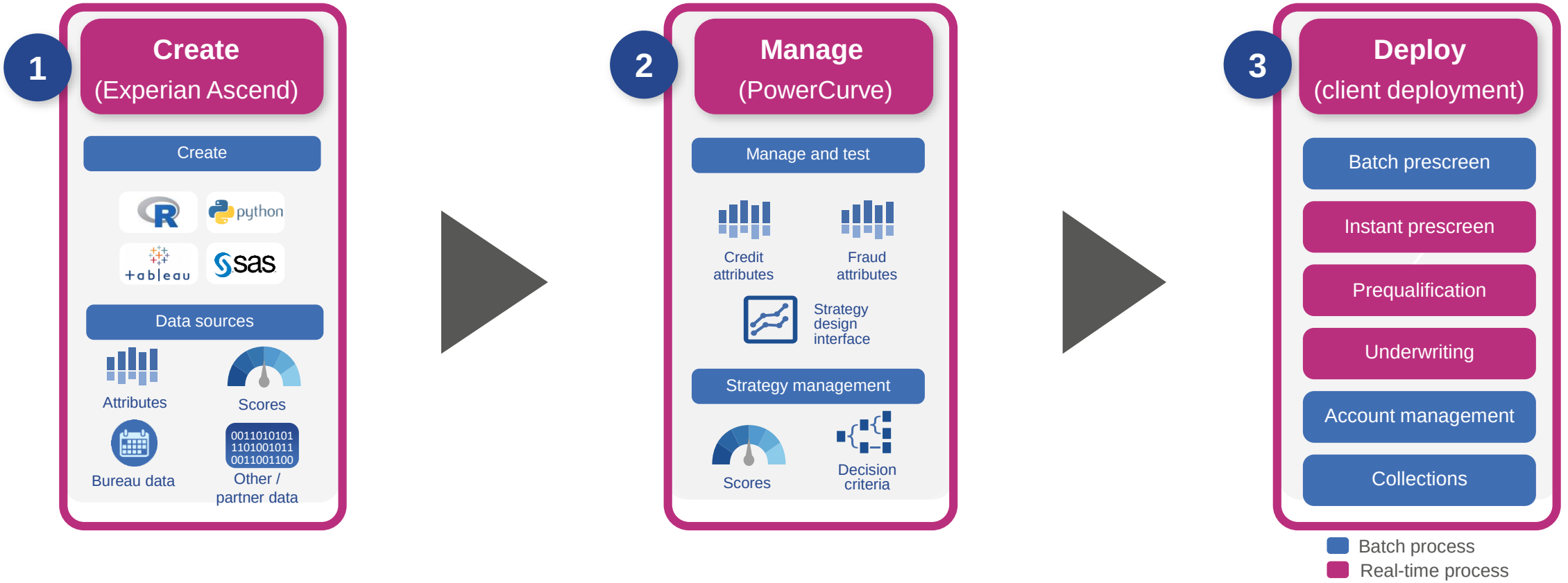


Illustrative Image.

Illustrative Image.

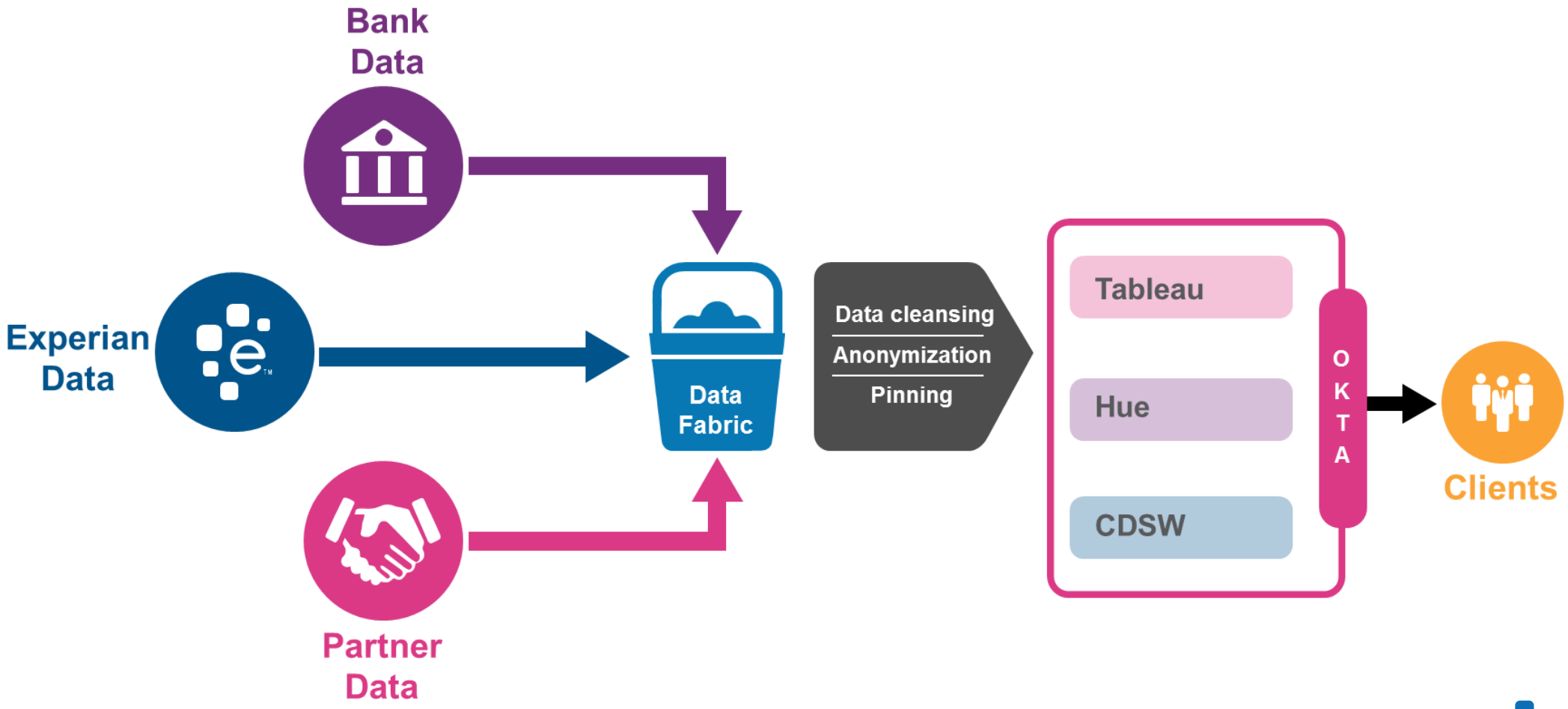
Platform capabilities needed to build, manage and deploy

Combining realtime and batch capabilities a requirement



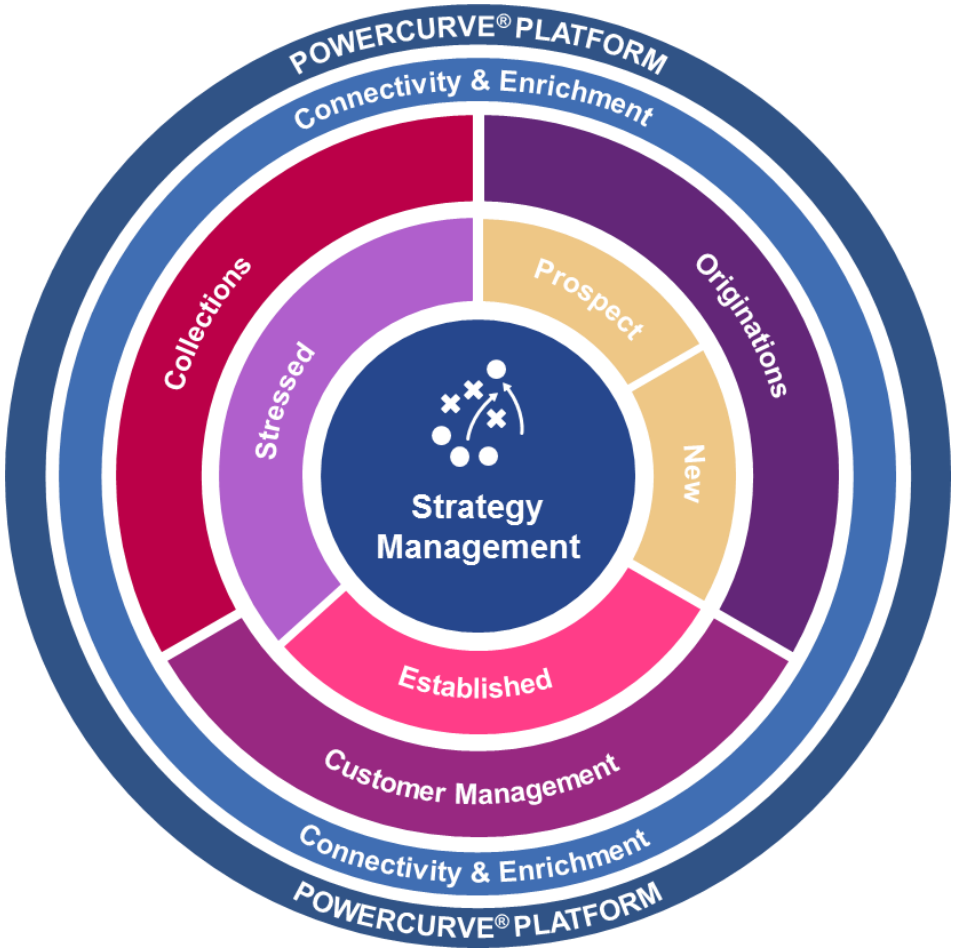
Platform for model creation…

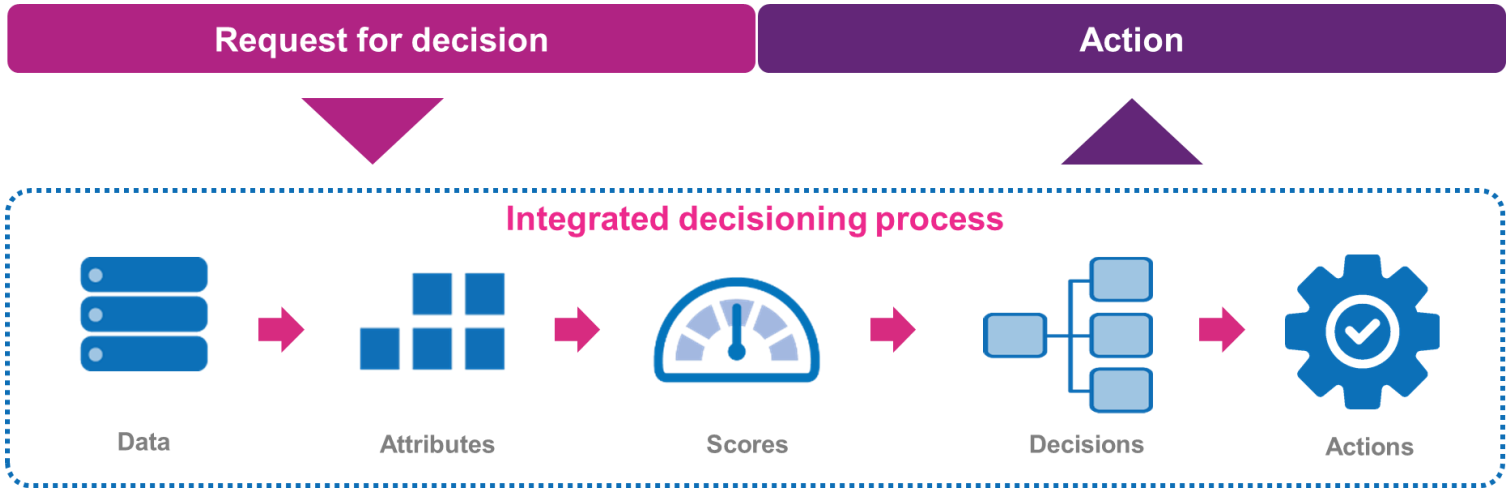
Data across sources with multiple tools



Platform for model maintenance and deployment

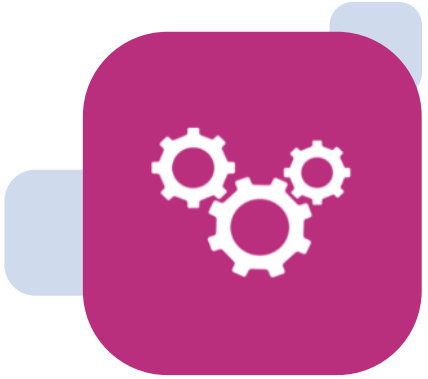
Ability to manage and deploy multiple models across the lifecycle





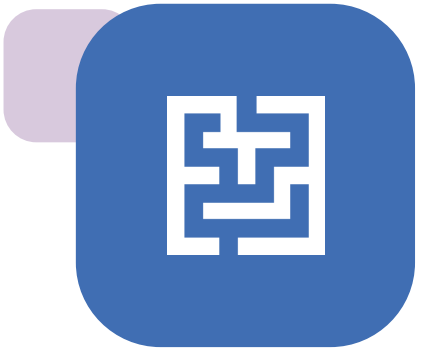
Within Experian - PowerCurve® allows organisations to manage and deploy models with ease

Key platform requirements



Unified platform

Integrated design studio



Decision execution

Data access

Monitoring and reporting

Modular framework

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Telco Scores

Consumer:

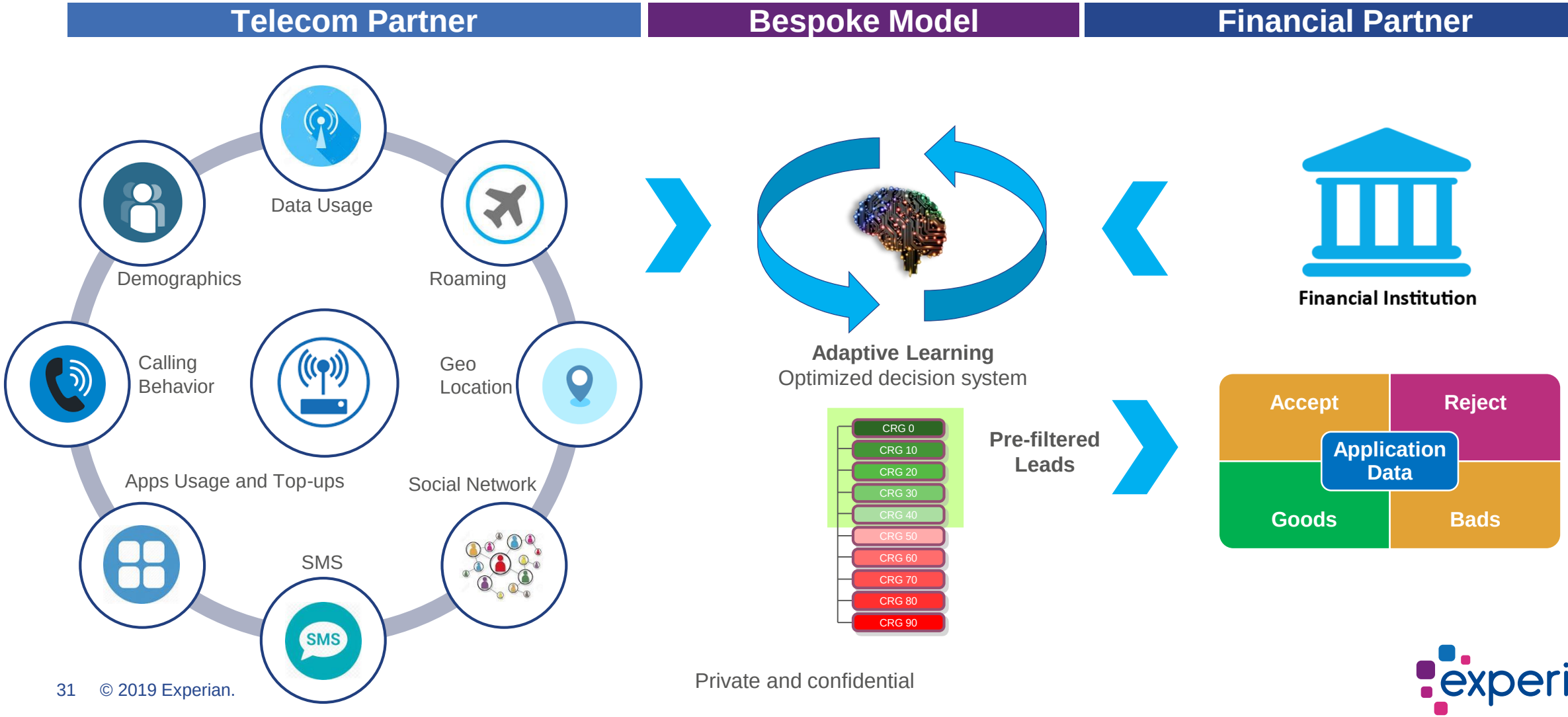
For individuals' who do not have bank account / financial records, Telco data is useful in predicting their risk level.

Experian:

Make use of Telco data and machine learning to boost predictive power and increase acceptance for small and micro loans.

How does Telco data scoring for cash loans work

Entry point to credit inclusion



Attributes collected from Operators for Risk Models

Illustrative list of variables

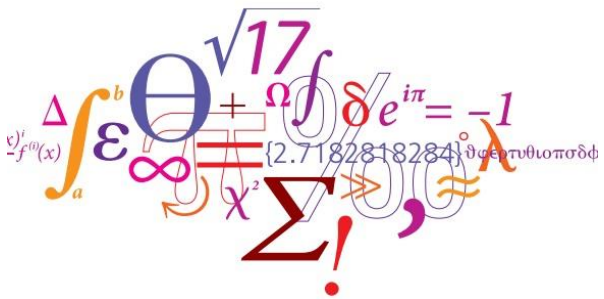
Pre Paid/Post Paid	• Information of Activation of Sim card or Tenure of customer along side other details mentioned in the application form • Postpaid defaults, credit, churn and payment information. Availability of mobile wallets
Data Usage	• Data used, revenue generated from data, hourly usage of data, data related value added services • Additional details on applications used, websites browsed over day and night
Geo-Location	• Mobility from CDR data, day and night presence, density of location • Prominent location with attributes from census, publically available data
Top Up History	• Top-up information, type, size and frequency of top-up along side channel of top-up • Channel of bill payment, invoiced amount, payment terms, and mode of payment – Bank, Credit, Wallet information
Calling/SMS Patterns	• Statistics on Call duration/count, calling phone numbers, towers, SMS sent and received • Time Of Day calling, weekday/weekend, inactivity, calling consistency
Demographics	• Demographics information recorded at the time of customer filing for sim-card • Inferred demographic details

ML techniques to generate telco scores

Transformation of telco data to 1500+ feature space and application of ML to pick up micro patterns

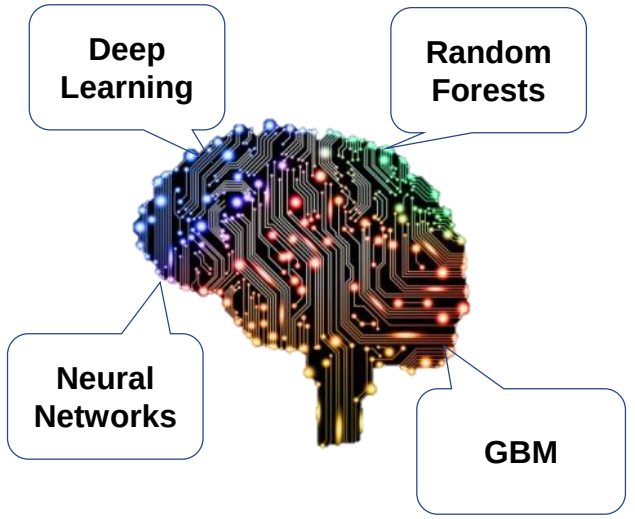
Demographics	Calling Behavior
Data Usage	Geo Location
Social Network	Data Usage
SMS	Apps Usage
Top-Ups	VAS

Feature Engineering



High dimension Feature Space
1500+ Transformed Features

Machine Learning Models



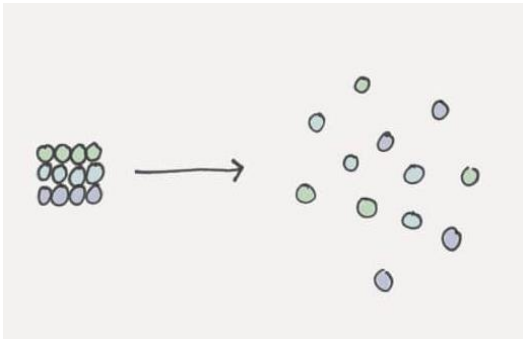
- Boost conversion with ML and pre-approved leads
- Optimized decision system

Some out of the box attribute creation

Interaction- Ratios

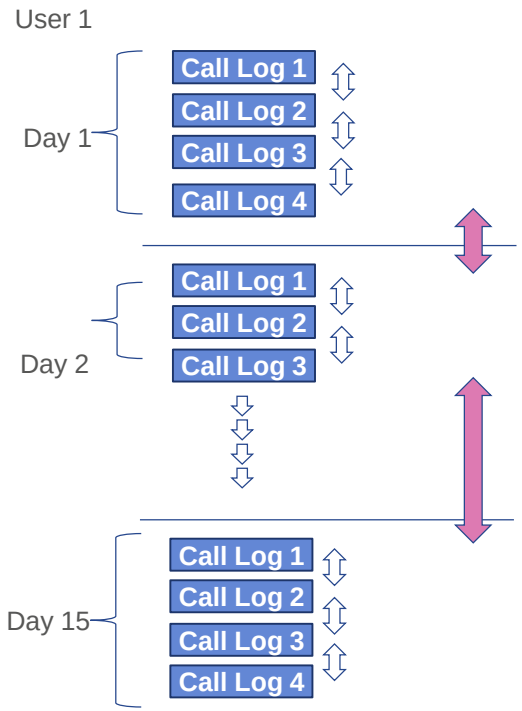
- SMS vs Voice
- Weekday/Weekend
- Day vs Night
- Office vs non-office

Entropy



- Geo-location (cell ID)
- Communication
- Time of Day entropy

Inter Call time gap/User Hibernation



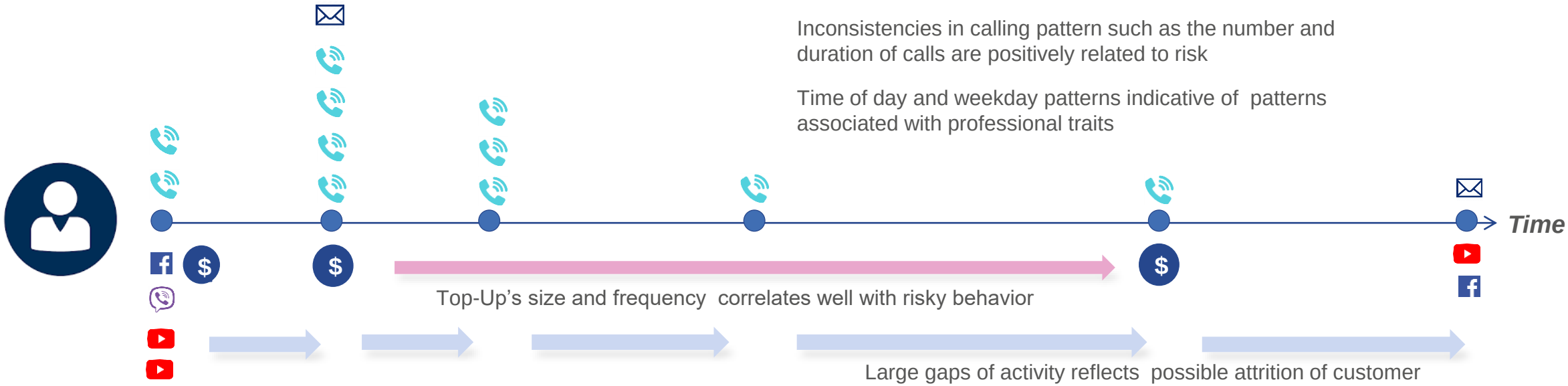
Velocity Based Variables

Month 1 **Month 2** **Month 3**

Call revenue/Data usage decreasing

Calling behavior patterns correlates to risk

We transform the raw CDR (Call detail records) into behavioral patterns to correlate with risk

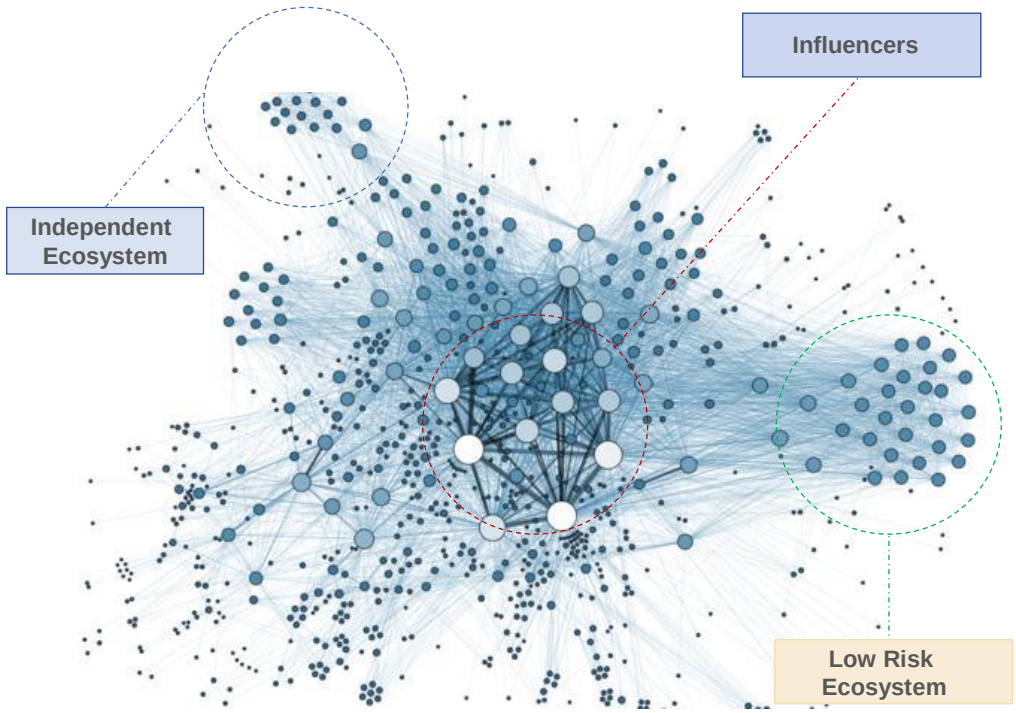


High ghost-hour activity and randomness in calling / sms and data consumption are risky

Network Analysis

Depending on the network of the individual we track the change in trends over months

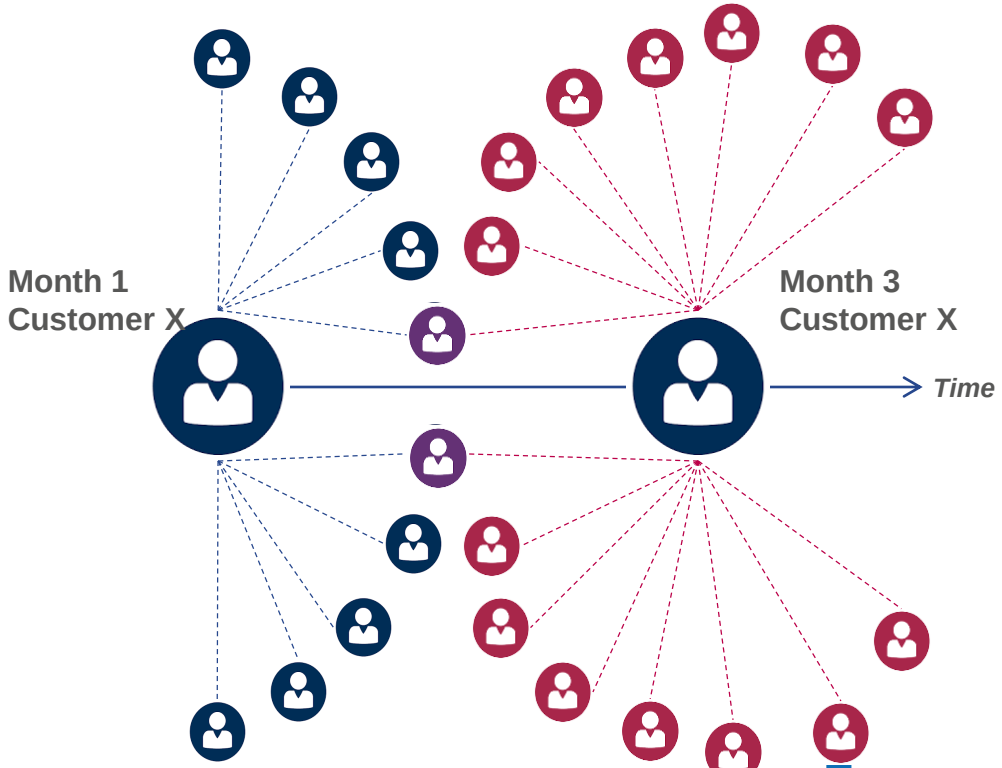
Social Network



Illustrative Image.
Source : Network World :

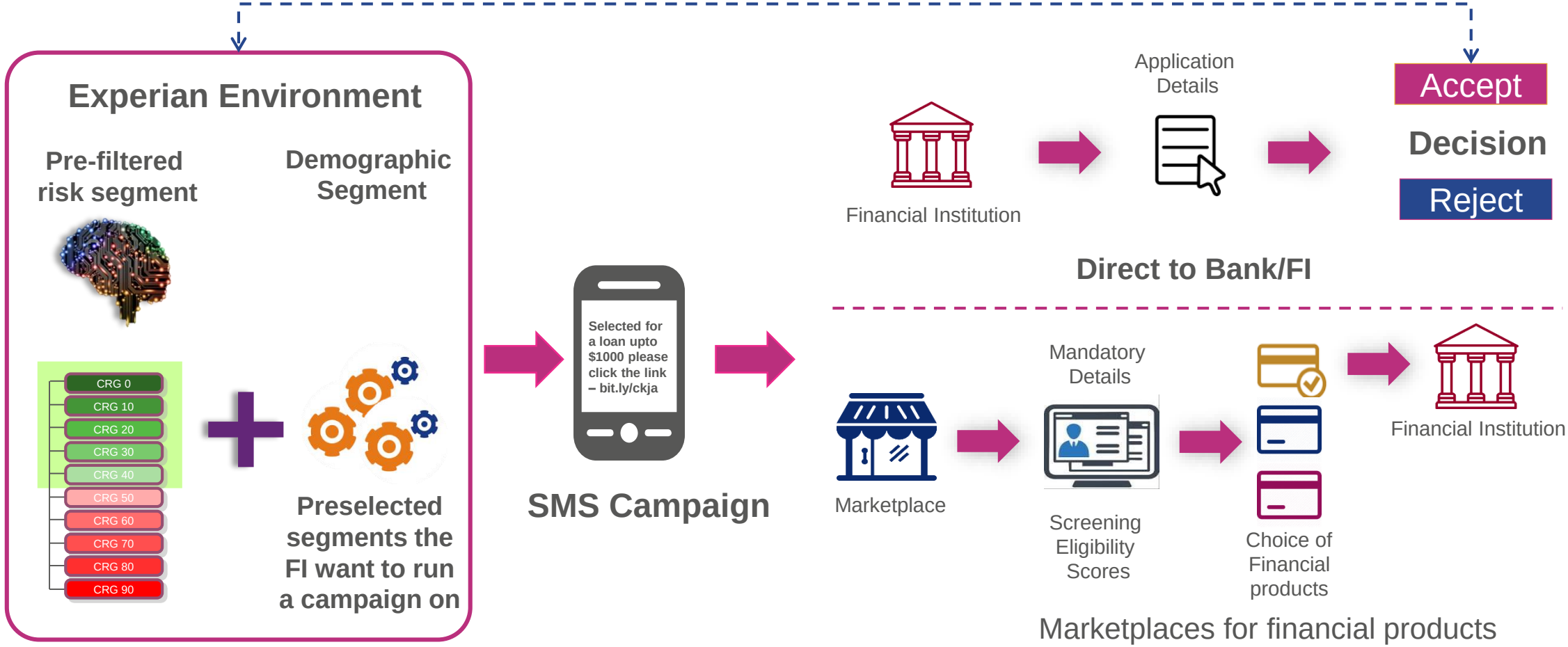
Change in Social Network

Overlap between social network in Month 1 as compared to Month 3
Change in size of social network



Customer Journey

Lead Generation through Telco Scores



Device Score

Consumer:

For individuals' who do not have bank account / financial records, Mobile data is useful in predicting their risk level.

Experian:

Make use of Mobile data and machine learning to boost predictive power and increase acceptance for small and micro loans.

Device Score

Potential to credit profile more than 5BN consumers globally

Harnessing the power of 'ready-to-use' Alternative Data

Device Score can assess the creditworthiness of applicants who are unable to obtain credit via the use of traditional data:

- Taps into one of the richest sources of alternative data to predict credit risk.
- Quick **ROI** as solution is scalable, available globally and can be implemented within days.
- Seamless customer experience through automation. No manual intervention.
- Stringent processes to ensure maximum customer privacy and data security.

Device Score enriches the credit decisioning process and provides reliable credit scores to a wider base of applicants.

1. Customer gives consent on the use of Device Score.

2. Device data is captured, encrypted and sent to the Experian Cloud.

3. Device data is stored, transformed and processed using sophisticated analytical models in Experian Cloud, generating a credit score.

4. Score of a customer is made available (via API or a query portal) to enrich the credit decisioning process.

5. Models are fine-tuned based on close monitoring of the performance and feedbacks received.

Predictability of Device Scores

Use of ML over rich attributes from device prove to be extremely powerful to predict risk

Challenges

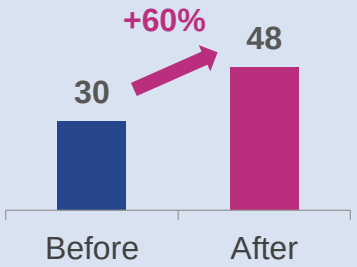
- Increase small ticket size lending for both thin file and thick file customers by:
  - Improving the accuracy of lending decisions made at less risk.
  - Maintain or lower NPLs and bad debt rates.

Approach

- Alternate data (e.g., SMS data, demographic data) was used t:o
  - Enrich the data points available and;
  - Improve the predictability of the existing models using machine learning frameworks.

Results

Thin File Segment



- ✓ Additional 60% improvement in GINI score for the thin file segment.

Only Device Data

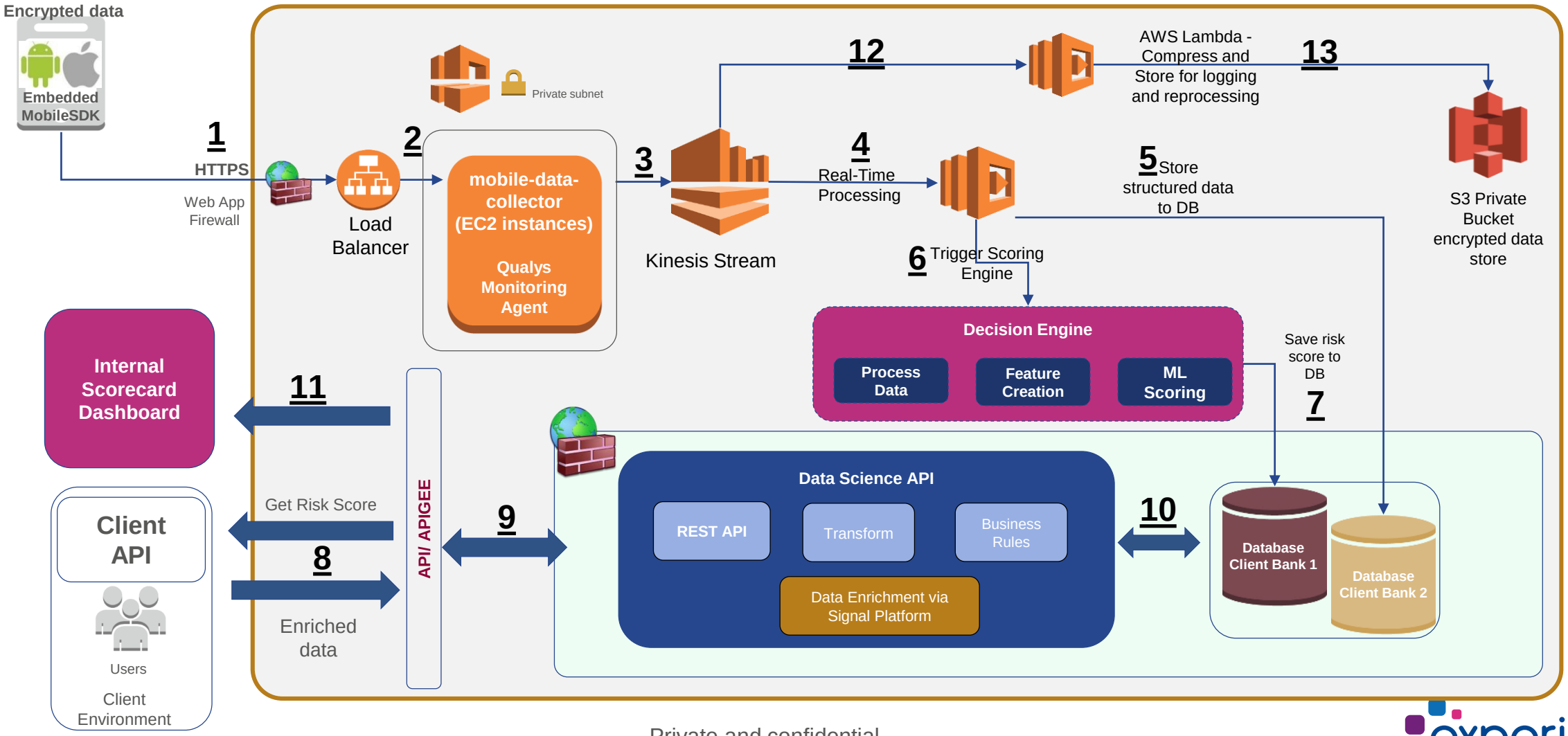
41.8 GINI

- ✓ Standalone predictive power of device data better than some bureau segments

Increase in overall lending volumes as credits are extended to more customers at a lower risk

Architecture for Scoring Device Data

Server less architecture on cloud allows to generate scores in realtime and handle scale



X Score
Alternate Credit Score for lending

Consumer:

For individuals' who do not have bank account / financial records, Telco data is useful in predicting their risk level.

Experian:

Make use of Telco data and machine learning to boost predictive power and increase acceptance for small and micro loans.

Problem Statement

BUREAU

TELCO

DEVICE

MARKET PLACE

PAYMENT / WALLET

E-COMM

OTHER DATA SOURCES

Bureau covers **less than 20%** of the population.

90%
Access to mobile services

80%+
Prepaid subscribers

20-40%
Access to banking services

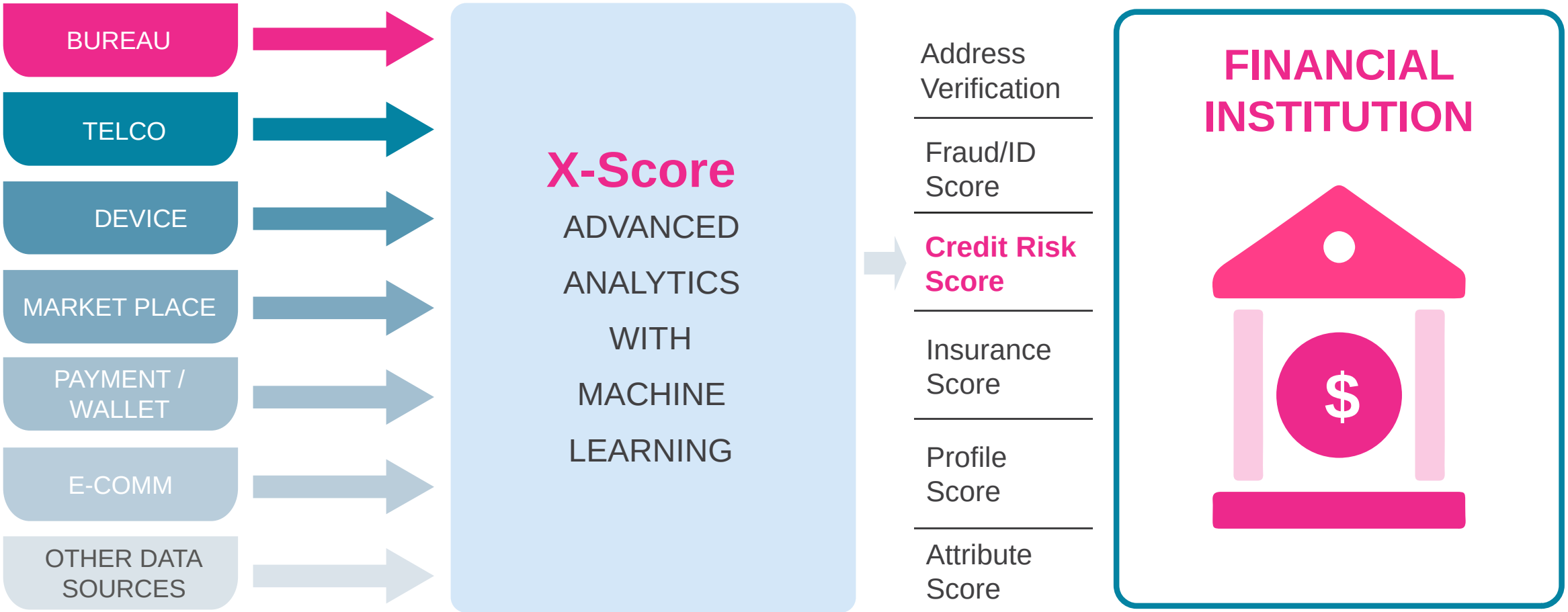
<5%
Access to credit services

APAC non-mature market reality

FINANCIAL INSTITUTION

Solution

X-Score enables credible information access to banks for credit decisioning



Score Pooling

Challenge in Ensemble methods

- Data Sharing restrictions from clients where they only share a score
- Distribute revenue across multiple partners fairly basis prediction confidence and accuracy of model

Solution : Consensus Approach

Final Score to be dependant on :

- **Goodness of Model** – Accuracy , AUC, KS ,etc
- **Level of Uncertainty** – if probability is close to 0 or 1 then Uncertainty in prediction is low

1. Determining the Local uncertainty $U_{ij}=-\Sigma p_i \log(p_i)$
2. Determining the Global uncertainty $c_{ij}=fn(A_j-A_i)$
3. Weight matrix Initialization – Transient state $wij=1/(U_{ij}^{-2}*\Sigma_j U_{ij}^{-2})$
4. Weight matrix – Steady state
5. Weight vector calculation
6. Consensus driven X score

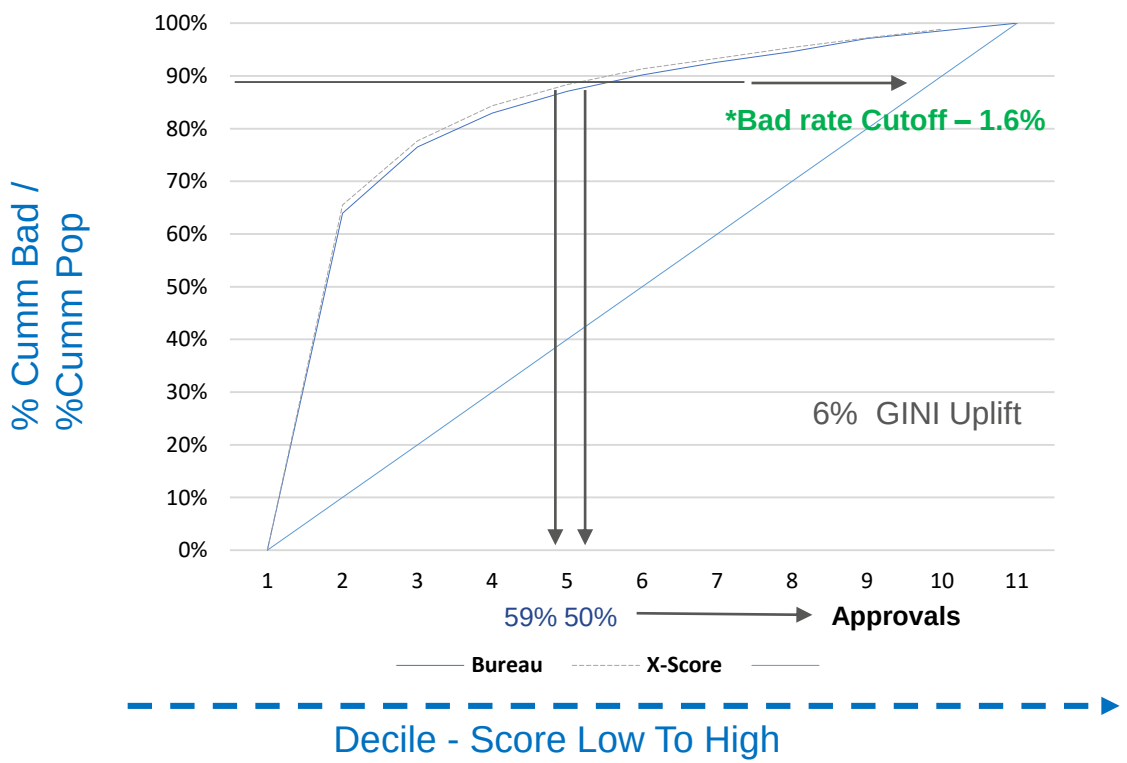
Weight vector can be directly calculated without multiple iterations - it is equivalent to left eigen vector of initial weight matrix "W", corresponding to eigen vector "1"

These parameters are final weightage to combine multiple scores to X-Score

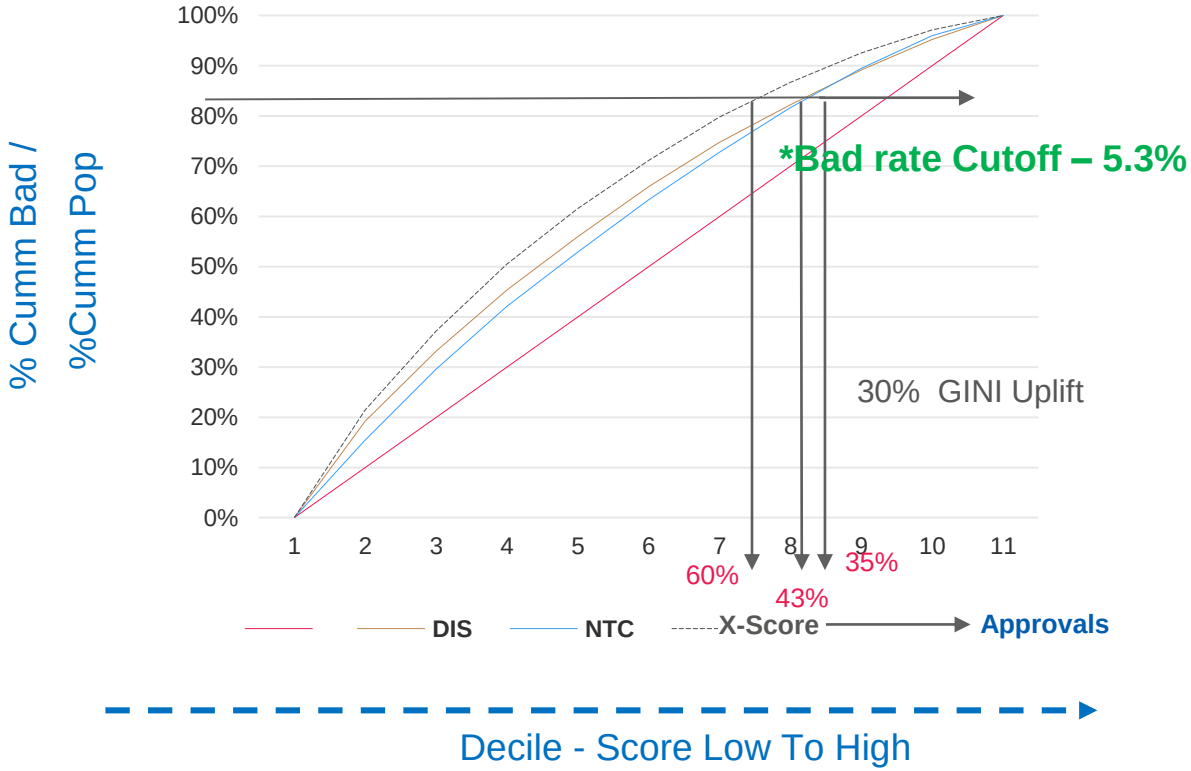
Private and confidential

Impact by combining alternate data scores (6% GINI Uplift on existing Bureau Score)

Thick file - The Consensus X Score improve approvals by 9%* (with same risk level)



Thin File – (NTC and Payment Score) - The Consensus X Score improve approvals by 17%* (with same risk level)



Recommended Usage of X-Score

For banks to use our model scores – we provide strategies for score cutoffs

Independent Score (along with risk policies & procedures)

Risk Band	Model Score	Cum Size	Strategy
1	[620, 640)	5%	Accept
2	[600, 620)	10%	
3	[580, 600)	15%	
4	[560, 580)	21%	
5	[540, 560)	24%	
6	[520, 540)	29%	
7	[500, 520)	35%	
8	[480, 500)	41%	
9	[460, 480)	45%	Low Limit
10	[440, 460)	49%	
11	[420, 440)	55%	
12	[400, 420)	60%	Refer
13	[380, 400)	65%	
14	[360, 380)	70%	
15	[340, 360)	75%	Reject
16	[320, 340)	80%	
17	[300, 320)	85%	
18	[280, 300)	90%	
19	[260, 280)	95%	
20	[240, 260)	100%	

Cut-off

Adjust the cut-off strategy by fine-tuning the expert model based on performance and risk appetite

- Experian X Score can be used along with existing credit policies in making lending decisions.
- Use case for assessing no hit & thin file customers who are 'new to bureau' with very limited credit data.

Dual Scoring (along with risk policies & procedures) Experian X Score

Existing A-Score

- X Score can be used in conjunction with existing scoring mechanism(s) to strengthen the lending decisions across the customer base.
- Telco and device data contains powerful behavioural information that can improve the predictiveness.
- Use case for Telco and device data through an existing partner to maximize coverage, even for thick file customers.

Low-side Override/ Swap-In (along with risk policies & procedures)

Risk Band	Model Score	Cum Size	Strategy
1	[620, 640)	5%	Accept
2	[600, 620)	10%	
3	[580, 600)	15%	
4	[560, 580)	21%	
5	[540, 560)	24%	
6	[520, 540)	29%	
7	[500, 520)	35%	
8	[480, 500)	41%	
9	[460, 480)	45%	
10	[440, 460)	49%	Low Limit
11	[420, 440)	55%	
12	[400, 420)	60%	Refer
13	[380, 400)	65%	
14	[360, 380)	70%	
15	[340, 360)	75%	
16	[320, 340)	80%	Reject
17	[300, 320)	85%	
18	[280, 300)	90%	
19	[260, 280)	95%	
20	[240, 260)	100%	

Cut-off

Identify and accept good potential customers below the cut-off score

- Use X Score to specifically focus on customers who are below the cut-off line.
- Use case for reducing type II errors of rejecting potentially good customers. Once accepted, these customers can be closely monitored using more complete data collected.
- Existing predictive models can be improved.

Alternate Data Projects in APac

Proven capability in risk scoring various verticals in emerging markets like SEA and India

	Device Scores	Telco Scores	Commerce Scores
Machine Learning	NLP, GBM XG Boost Real-time Solution	Random Forests, GBM Network Analysis Geolocation Analytics	Logistic Regression XG Boost
Performance Indicators	SMS and App Usage Data : **GINI 41** Version2 SDK Data: **Expected GINI 30+**	Behavior Data : **GINI 33** Lifestyle Profile Data: **GINI 25**	Wallet Data: **GINI 24** Merchant POS Data: **GINI 20**
Clients On boarded	Live with clients - 60% Approval Rates 4+ Clients in Pipeline	Data on boarded – Live in Q3 FY20 3+ Telco in Pipeline 200M customers in Indonesia Potential 300M customers in India	One client previously live and 30% approval rate Live in India

Credit Risk Grade

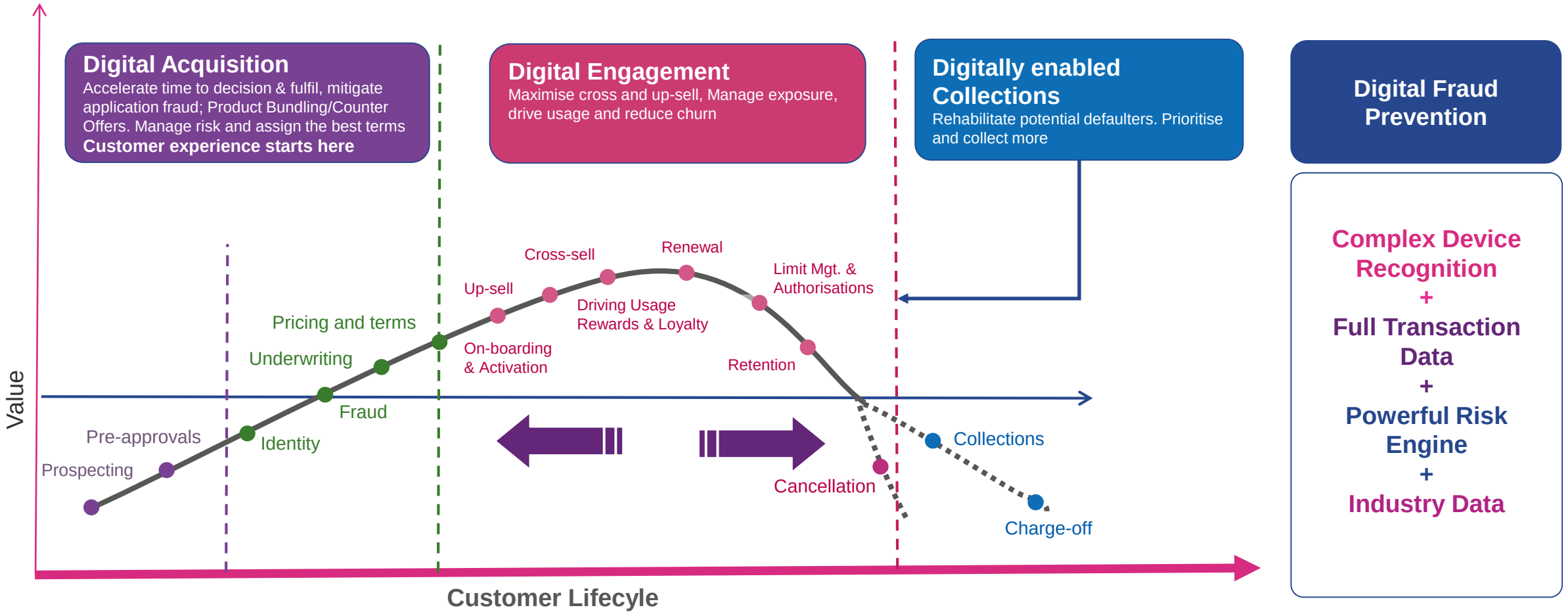
- CRG 0
- CRG 10
- CRG 20
- CRG 30
- CRG 40
- CRG 50
- CRG 60
- CRG 70
- CRG 80
- CRG 90

GINI – Measures the rank ordering or performance of a model to predict risk

Agenda

- Experian and the X-labs
- Industry evolution and the APac Context
- Alternate Credit Scoring - building blocks
- Implementations – Telco, Wallets & Open data
- Summary and considerations

Decisioning lifecycle for a digitally enabled bank



Decisions across the customer lifecycle

A better customer journey

Business impact can be tremendous

Originations

15%

Increase in sales through improved accept rates

5%

Reduction in bad debt through better exposure management

80%

Reduction in processing time through automated decisions

Customer Management

10%

Increase in acceptance rate for cross-sell/up-sell

10%

Increase in profitability through risk-based pricing

14%

Increase in profit with better limit setting strategies

Debt Management

26%

Improvement in delinquency rate

25%

Increase in efficiency with an improved collections process

Thank You !

experian™