McKinsey & Company

Data and analytics innovations in Credit Portfolio Management

Discussion Document

March, 2022

IACPM

McKinsey & Company

McKinsey and IACPM together completed a survey on new developments in data and analytics for credit portfolio management

The survey had 3 main objectives

Survey range of industry practice on:

- Emergence of alternative data sources for credit risk identification, assessment, and monitoring
- Related evolution of analytics tools

Understand degree to which different data types and analytical approaches are in use/under consideration across institutions, geographies, subsectors, and CRE/C&I portfolio segments

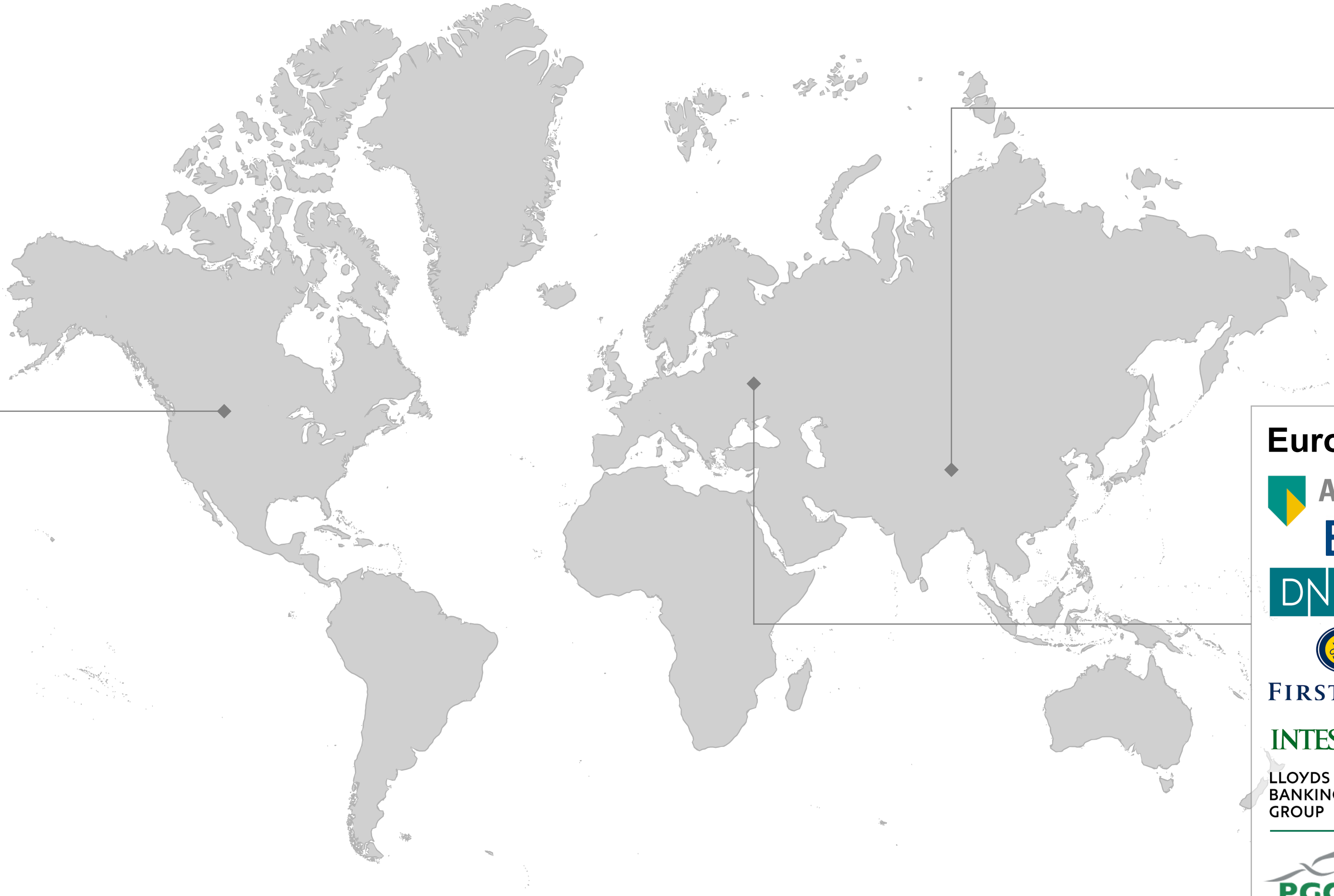
Develop insights on current state and path forward for participants to incorporate next generation data and analytics

IACPM McKinsey & Company

44 financial institutions participated in the survey from across Americas, APAC, and EMEA

Americas

AIG
Bank of America
ATB Financial
Bancolombia
BMO
HTLF Strength. Insight. Growth.
fcc Farm Credit Canada
National Bank Financial Markets
KeyBank
PNC
M&T Bank
Prudential
Goldman Sachs
RBC
Regions
usbank



Asia Pacific

ANZ
CommonwealthBank
DBS
MUFG
Maybank
OCBC Bank
nab
UOB
SBI
SMBC

Europe Middle East and Africa

ABN·AMRO
absa
BBVA
BARCLAYS
DNB
IDB
Credit Suisse
FirstRand
standard chartered
UBS
Intesa Sanpaolo
UniCredit
Lloyds Banking Group
Munich RE
NATIXIS BEYOND BANKING
PGGM
Swiss Re

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

Content

Summary of survey results	**20 mins**
Perspectives on selected topics	30 mins
Questions	10 mins

Summary of results

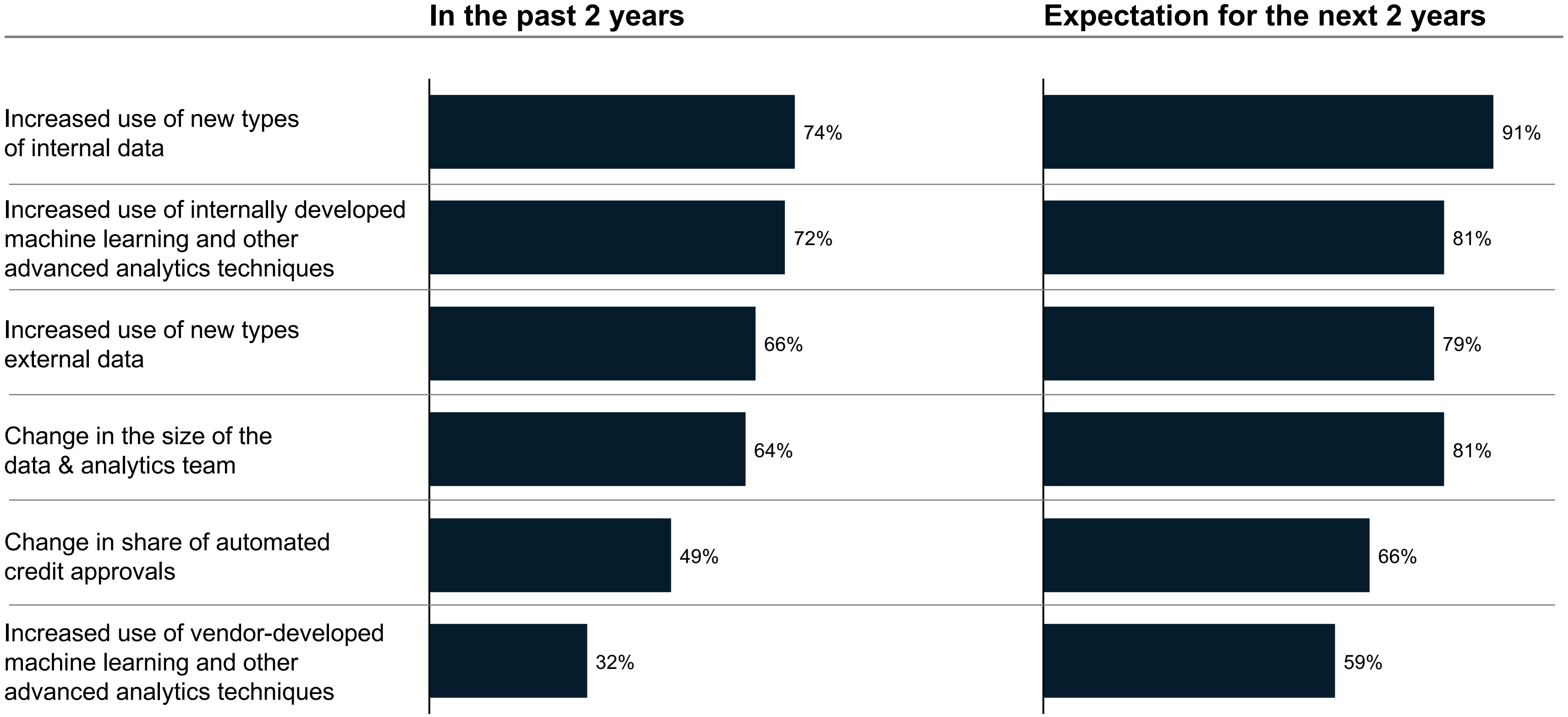
A	**Trends**	Over the next 2 years, larger number of participants expect significant increase in the use of internally developed advanced techniques and new types of external data
B	**Challenges**	Data quality assessment and talent management are the top challenges for use of both advanced analytics and innovative data solution
C	**Climate**	Majority of the participants believe that Climate and ESG is the next biggest challenge for credit assessment
D	**Investment and Strategic Goals**	Top investment areas have been data tech and data acquisition. Over the next 2 years, participants expect a greater role of innovative data and advanced analytics in improving Credit Strategy and Client Experience
E	**Use cases**	Machine learning models are primarily gaining traction for Risk Scoring of SMEs and Early Warning across the board Innovative external data sources are more used for Corporate segment while SME segment uses more innovative internal data sources
F	**Impact**	Use of innovative data and/or advanced analytics improves model accuracy , turn-around-time, automated decisioning and time spend on analysis, with higher benefit observed in the SME segment

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

A: Over the next 2 years, a vast majority of participants expect significant increase in the use of new types of external data and internally developed advanced techniques

How has your firm's data/analytics for credit decisions changed in the past 2 years and how do you expect it to change in the coming 2 years?

% participants see increase in trends



In the past 2 years, over 60% of the participants have seen an increase in the:

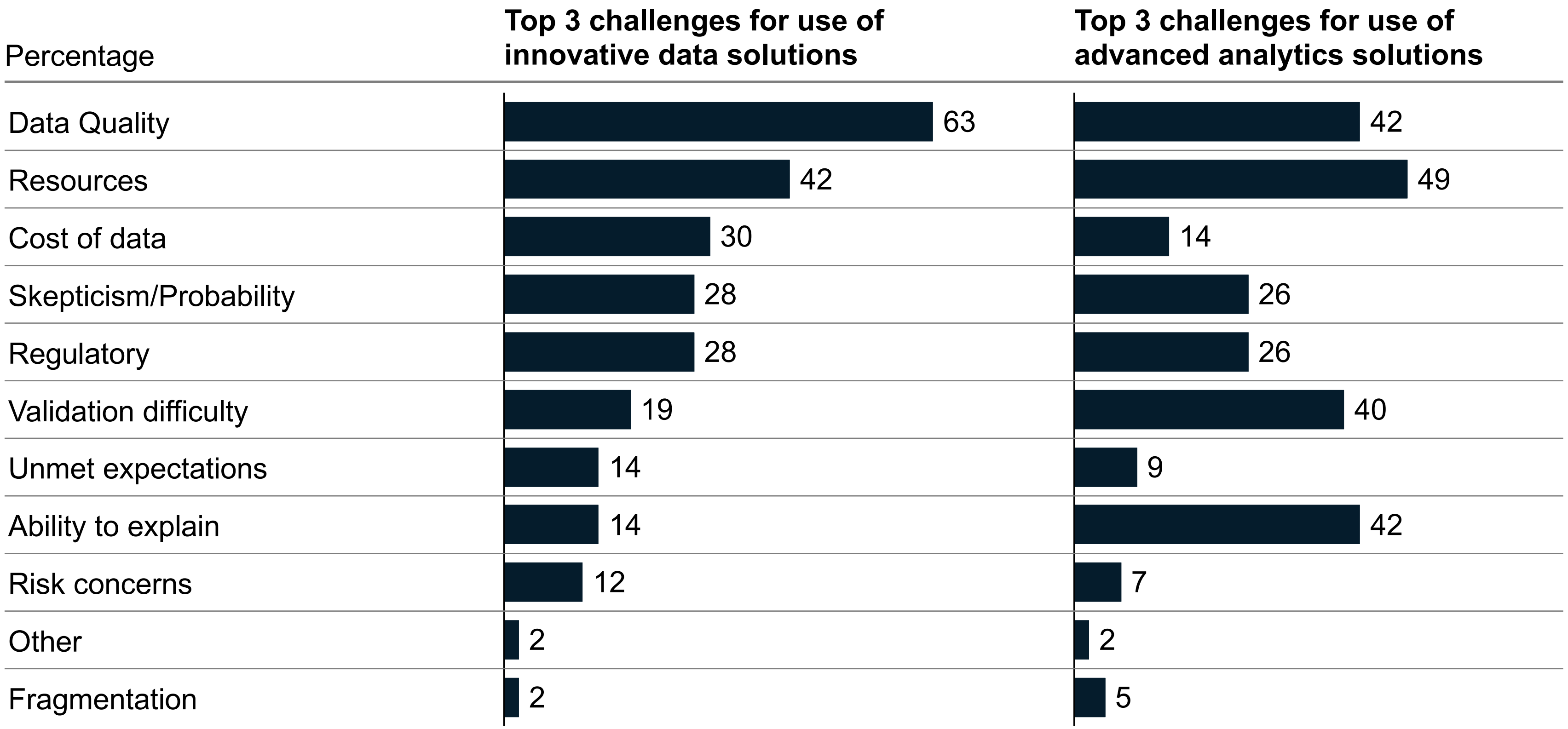
- Use of new types of internal and external data
- Use of internally developed advanced techniques
- Size of data and analytics team

Over the next 2 years, even larger % of participants expect this trend to continue

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

B: Data quality assessment and talent management are the top challenges for use of both advanced analytics and innovative data solution

Currently, what are the major challenges faced by your firm that constrain the use of innovative data or advanced analytics (e.g., machine learning and AI)?



Major challenges for use of advanced analytics solutions are:

- Attract, retain and develop resources
- Ability to explain
- Data quality
- Validation

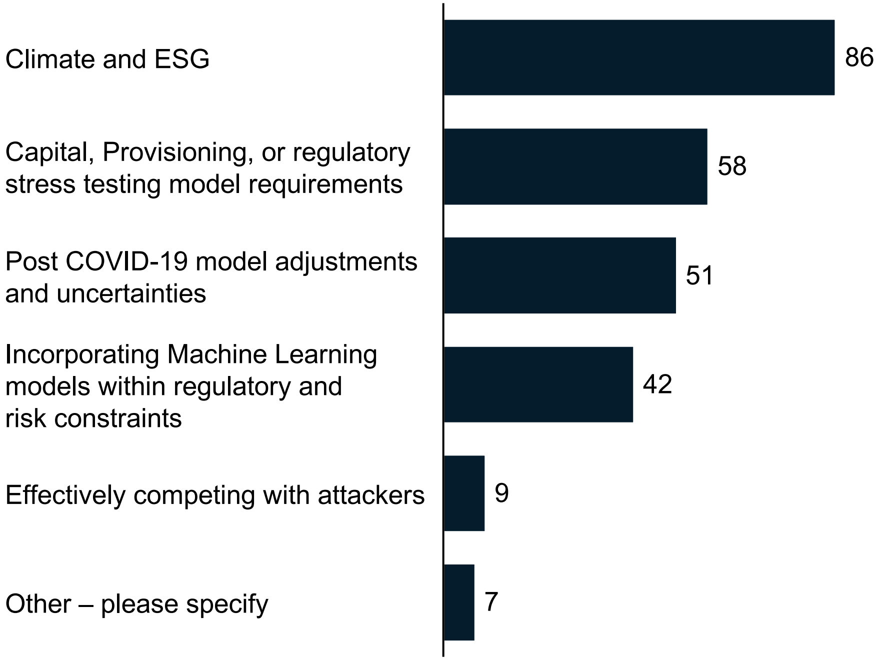
While for using innovative data, key challenge in data quality assessment and talent

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

C: 86% of the participants believe that Climate and ESG is the next biggest challenge for credit assessment

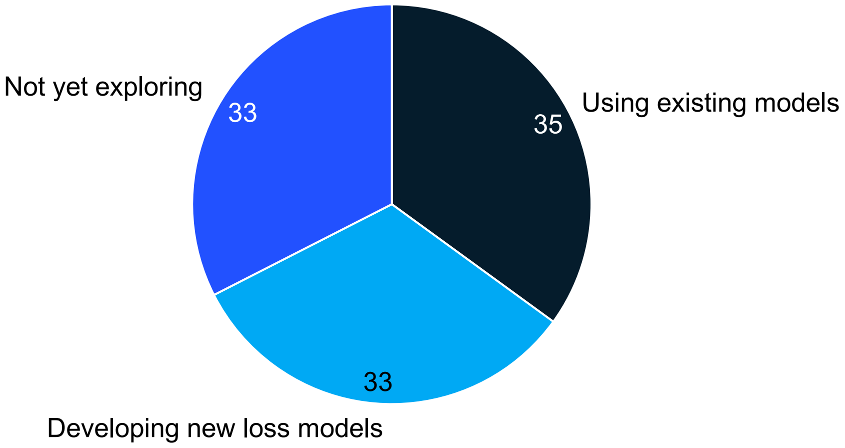
27: What are the biggest challenges facing credit risk and credit portfolio management analytics in the next 2-3 years?

Percentage



25: For incorporating the impact of climate risk ,are you using existing loss models with climate shock applied to input variables? Or are you developing new loss models to assess it?

Percentage



86% of the participants believe climate and ESG is the next biggest challenge for credit assessment, followed by COVID-19 adjustments, capital and regulatory exercise, and using machine learning models

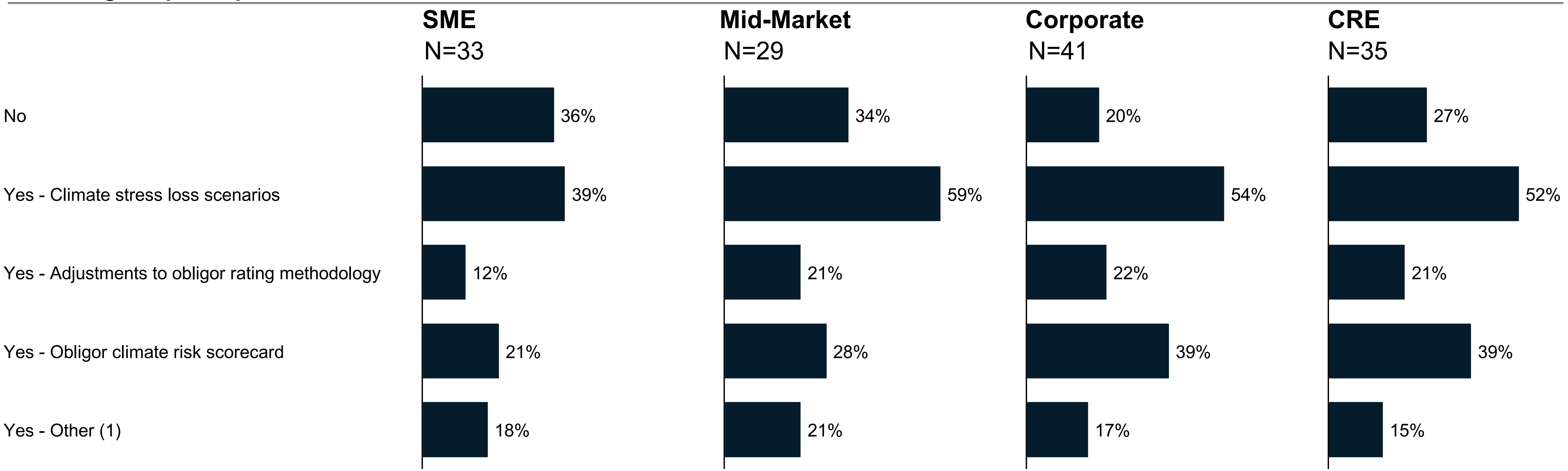
One third of the participants plan to use existing credit models to translate the climate impact to credit risk and another one third of the participants plan to develop new loss models for climate assessment

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

C: > 50% of participants have implemented/or are planning to implement climate stress loss analyses in mid-market, corporate, or CRE space

Have you implemented or are planning to implement in the next 12 months any changes to the credit assessment/ adjudication and monitoring models to capture the impact of climate change? (transition and physical risks)

Percentage of participants



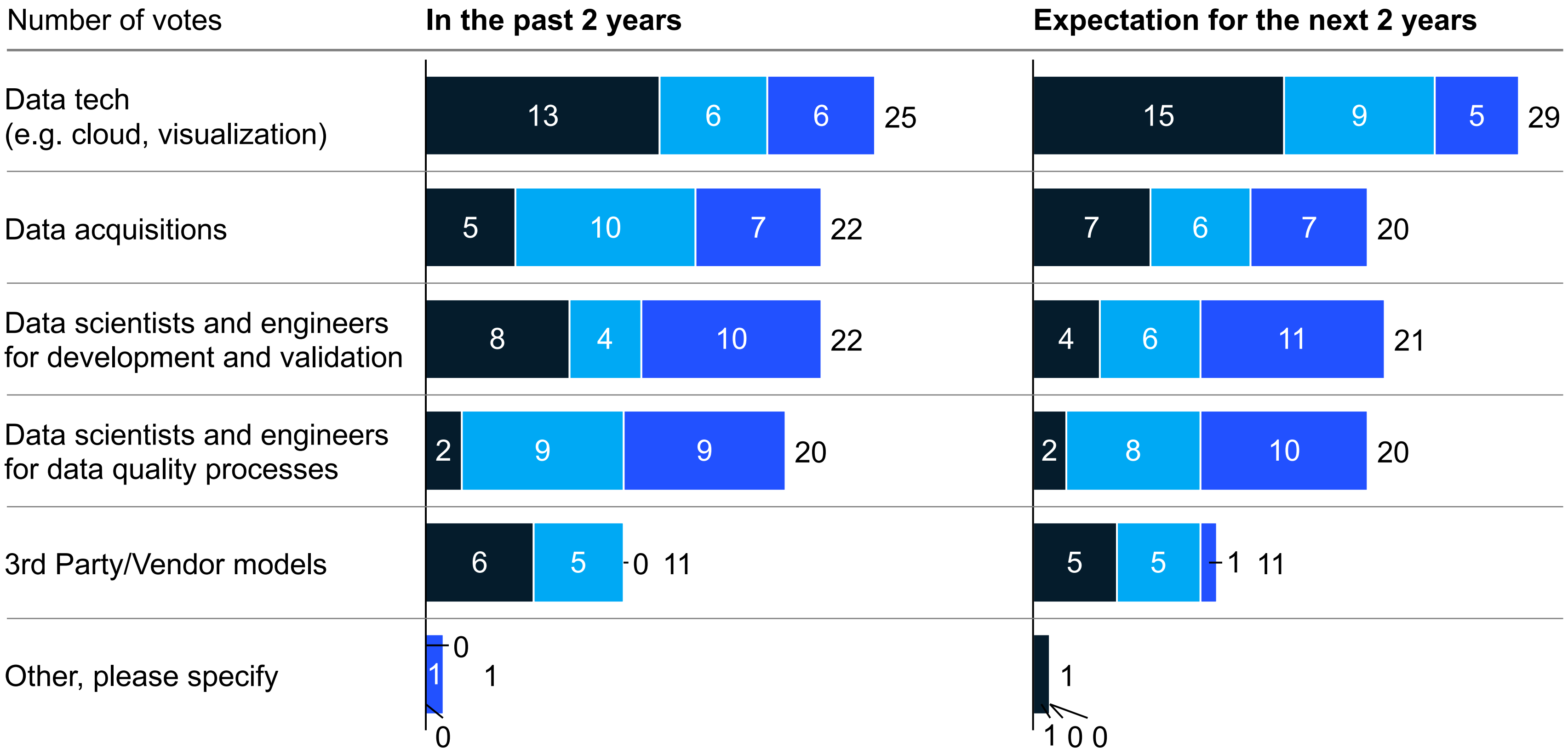
1: E.g., Adjustments to obligor rating methodology and climate stress loss scenarios, but beyond 12 months. Bucketing of risks (geography, industry, property type segments).

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

D: In the past 2 years, top investment areas have been data tech and data acquisition – this trend is expected to continue over the next 2 years

Largest 2nd largest 3rd largest

Where have you made the most investments in the past 2 years and where do you expect to invest the most in the coming 2 years?



In the past 2 years, the top investment areas for participants were data tech and data acquisition

And this trend in expected to continue over the next 2 years with higher expected investment

Other top investment areas include talent for both development/ validation and data processing

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

E: Machine learning models are primarily gaining traction for Risk Scoring of SMEs and Early Warning across the board

Innovative data sources

What methodologies are being "used in production", "validated" or "in pilot" for each of the listed use cases for the Corporate and SME portfolio

Percentage		Expert based qualitative models	Statistical approaches (e.g., linear or logistic regression or CHAID)	Simulation-based models (e.g., monte-carlo approaches to economic capital)	Machine learning tree algorithms (e.g. random forest, gradient boosting)	Other machine learning models (e.g. support vector, NLP, neural network)
Corpo-rate	Risk score/ Rating/ PD model	44%	68%	10%	2%	0%
	Early warning indicators	39%	37%	5%	12%	5%
	Credit pricing	44%	24%	5%	0%	0%
	IFRS 9/ CECL	39%	56%	7%	0%	0%
	Stress Testing	32%	46%	29%	0%	0%
SME	Risk score/ Rating/ PD model	41%	72%	6%	13%	3%
	Early warning indicators	44%	56%	9%	16%	3%
	Credit pricing	47%	38%	3%	6%	0%
	IFRS 9/ CECL	44%	66%	6%	3%	0%
	Stress Testing	38%	53%	28%	0%	0%

Expert based and statistical models are most widely applied approaches across the spectrum of use cases

Simulation based models are more widely used for stress testing

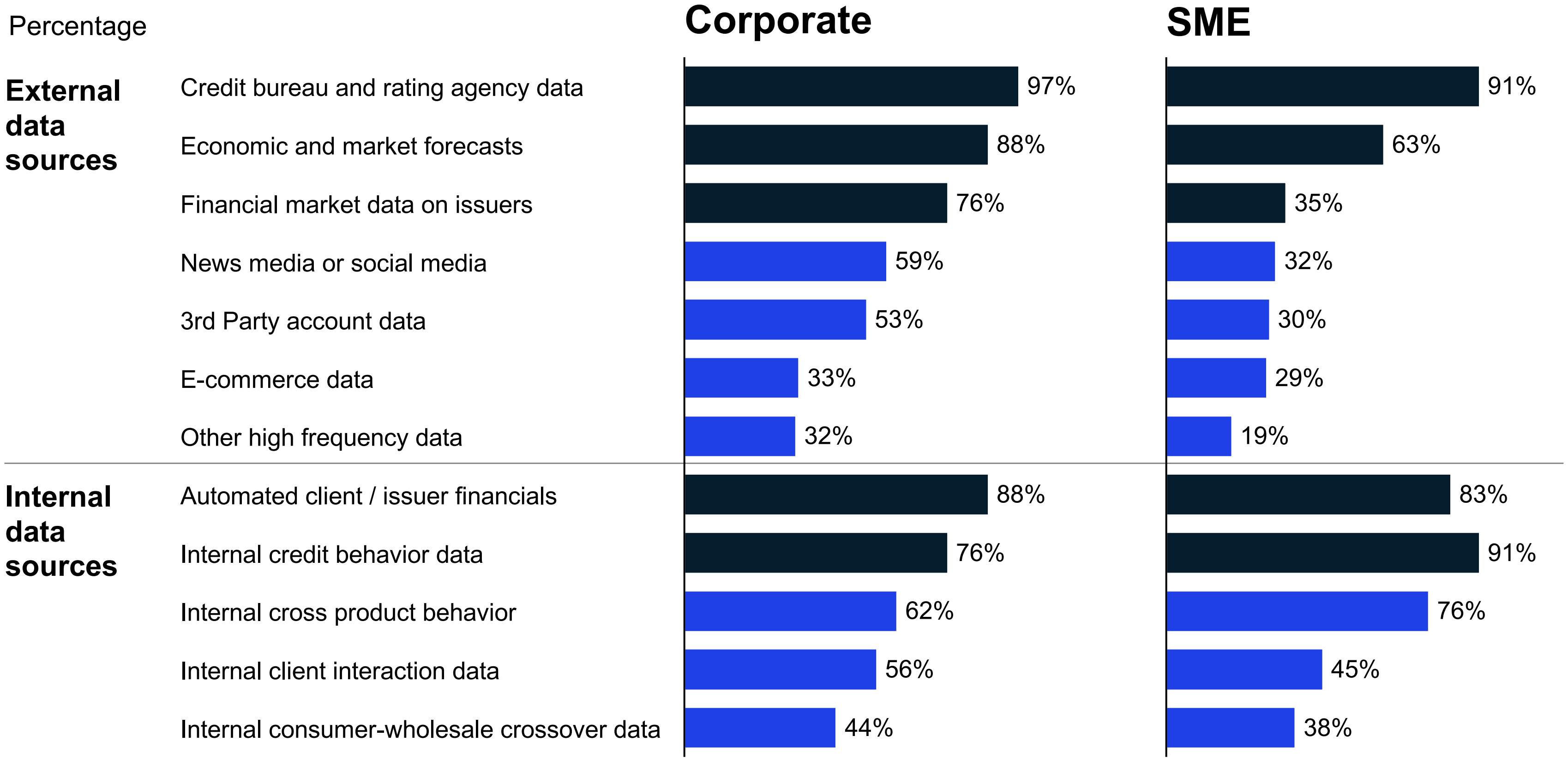
Machine learning models are getting traction for Risk Scoring, Early Warning, and Pricing

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

E: Innovative external data sources are more used for Corporate segment while SME segment uses more innovative internal data sources

Innovative data sources

Which of the following categories of data are being used in production, under pilot or under consideration for credit risk management use cases within the Corporate portfolio?



External data sources:

For Corporates, over 50% of the participants are using, piloting or considering New media or social media and 3[rd] Party account data, a higher proportion than for SMEs

Both segments use E-commerce data at similar rates

Internal data sources:

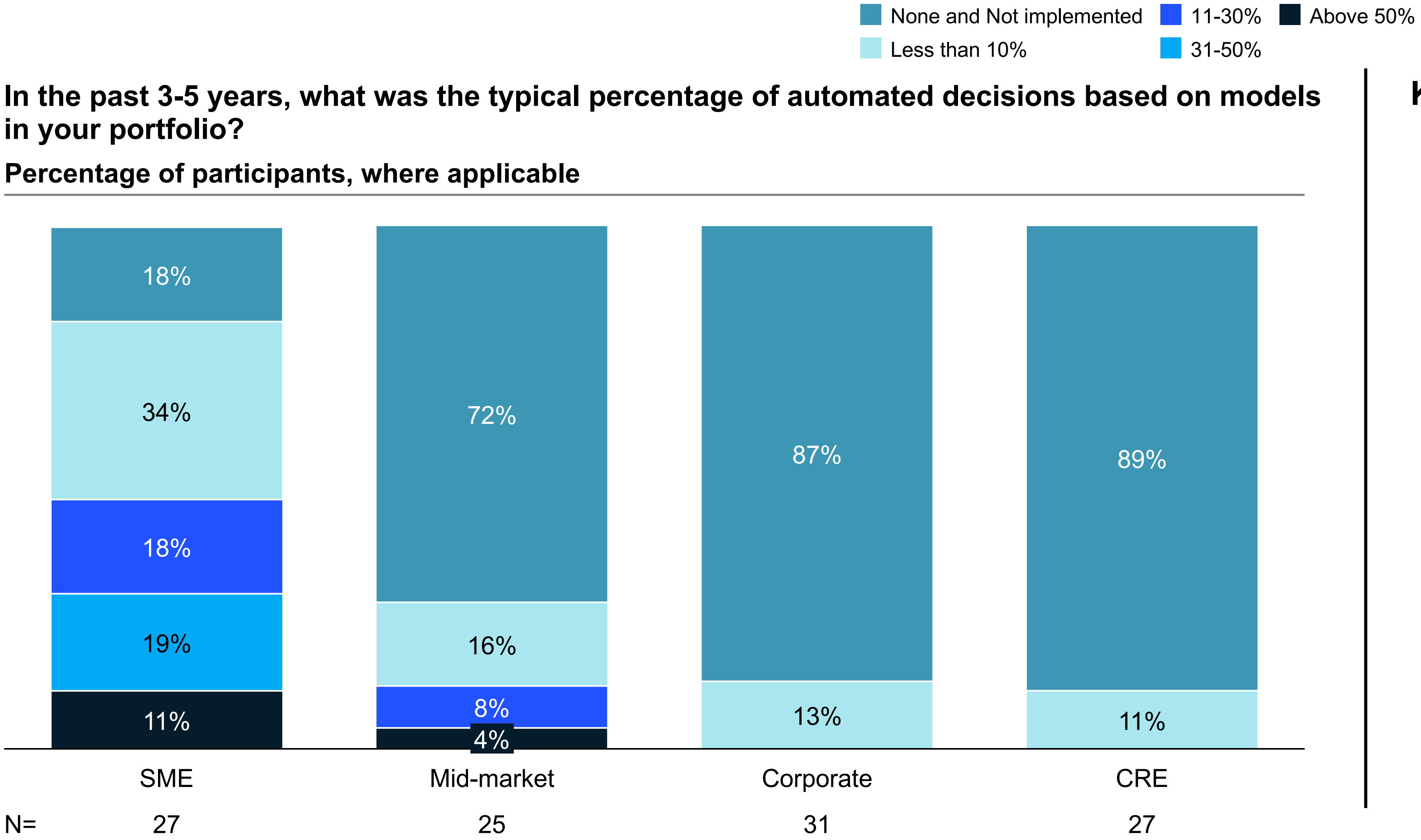
For Corporates, over 70% of the participants are using, piloting or considering Automated client/ issuer financials and internal credit behavior data, while for the SME segment in addition to above 2, internal cross product data also has large share

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

F. Automated decisions are largely a feature of SME portfolios

In the past 3-5 years, what was the typical percentage of automated decisions based on models in your portfolio?

Percentage of participants, where applicable



Key insights

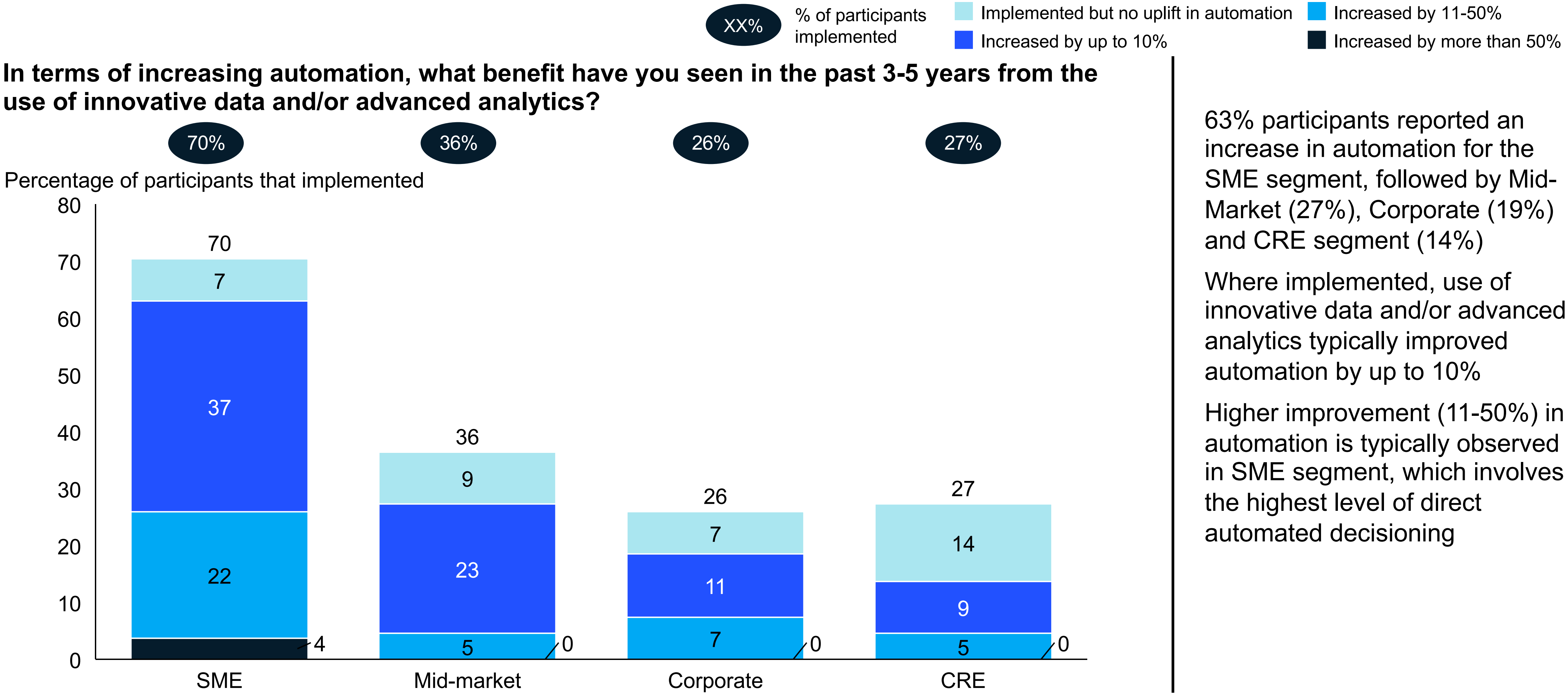
Fully automated decisions for a material portion of the portfolio is almost exclusively a feature of SME portfolios

However, there are pockets of portfolio with full automation, even for mid-market and others

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

F: Where implemented, use of innovative data and/or advanced analytics typically achieved increased automation

In terms of increasing automation, what benefit have you seen in the past 3-5 years from the use of innovative data and/or advanced analytics?



63% participants reported an increase in automation for the SME segment, followed by Mid-Market (27%), Corporate (19%) and CRE segment (14%)

Where implemented, use of innovative data and/or advanced analytics typically improved automation by up to 10%

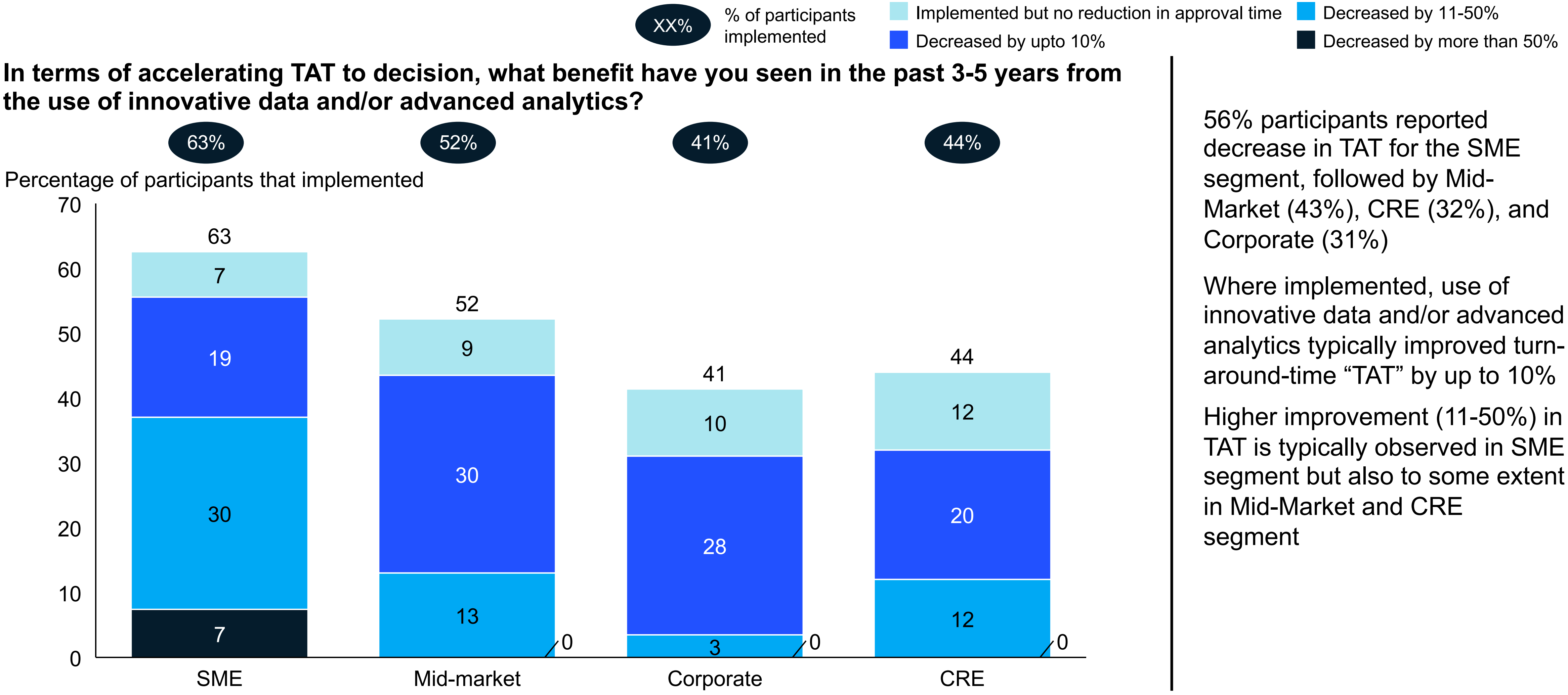
Higher improvement (11-50%) in automation is typically observed in SME segment, which involves the highest level of direct automated decisioning

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

F: Where implemented, use of innovative data and/or advanced analytics significantly improved turn-around-time "TAT" for SMEs

XX% — % of participants implemented

In terms of accelerating TAT to decision, what benefit have you seen in the past 3-5 years from the use of innovative data and/or advanced analytics?



56% participants reported decrease in TAT for the SME segment, followed by Mid-Market (43%), CRE (32%), and Corporate (31%)

Where implemented, use of innovative data and/or advanced analytics typically improved turn-around-time "TAT" by up to 10%

Higher improvement (11-50%) in TAT is typically observed in SME segment but also to some extent in Mid-Market and CRE segment

Source: McKinsey/IACPM Survey on data and analytics innovations in Credit Portfolio Management – October 2021

Agenda

Summary of survey results	20 mins
Perspectives on selected topics	**30 mins**
Questions	10 mins

Perspectives on specific topics : Climate Risk, Next generation data and analytics, and Scenario planning and forecasting

Climate Risk

Setting up a modular, portfolio specific methodology for scenario analyses is critical, prioritization will depend on exposure to high-risk sectors

Climate risk impact on portfolio requires inter-disciplinary skills and mobilization across credit, front-line, and model risk management

Next generation data and analytics

Using cross-product data with help from AI/ML can drive both revenue growth and automated credit decision

For larger obligors, availability of analytics and accessible data is key for turnaround time reduction, however, doing so requires treating data quality as more than a ‘regulator-required’ effort

Scenario Planning and Forecasting

Building a flexible infrastructure to forecast and optimize portfolio is more critical than ever

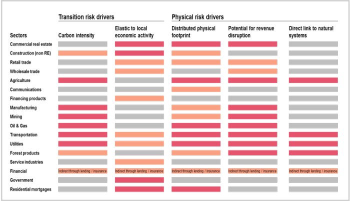
Rather than waiting for a full-scale solution, banks would want to establish analytics and organizational capabilities that enable rapid ‘what-if’ analyses

Key learnings on the materiality of Climate Risk on Credit, based on our extensive work with banks

Risk is concentrated in “pockets” across the portfolio; banks need to take a targeted approach

Both physical and transition risk lie in very targeted areas of the portfolio: for example, for a large global bank we identified that approximately ~15% of their loan book was materially exposed to climate risk

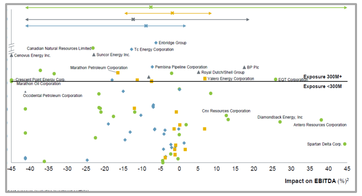
Banks need to perform heatmapping to focus their efforts on the high risk portfolios and risks: even within a CRE portfolio for a large US bank, we found that majority of the credit impact came from 10% of the portfolio



The “average” impact is moderate in the near-term, but there is high degree of counterparty-level variability

We found that even for high risk industries, the average impact is moderate: For example, in a portfolio of upstream O&G companies, the impact by 2025 under below 2C scenario was ~7% median impact on EBITDA

However, the difference between winners and losers is stark: in the upstream O&G example, we saw several counterparties with up to ~40% impact on EBITDA, while there were other companies that saw a positive EBITDA impact

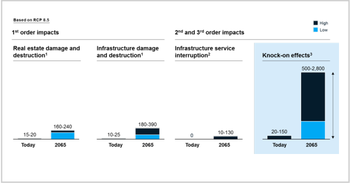


The majority of the risk is in knock-on impacts that most banks do not model

Direct damages are immaterial on credit; Knock-on effects can dwarf direct impacts, e.g., 4.5x the impact of direct 1st order impact for a Muni flood example

Material risk drivers include community deterioration, geographic transition risk, and broader ecosystem impacts (e.g., insurance cost and availability)

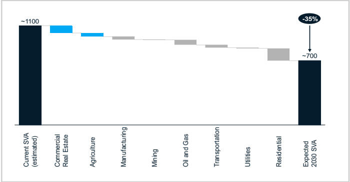
Real estate losses are driven by asset pricing (property values and cap rates), not physical damage



Climate is not a “capital” problem; however it can have real impact on returns / economic profit

A ‘CCAR mindset’ of focusing on capital risks will underestimate the business value risk and miss the opportunity to steer the business

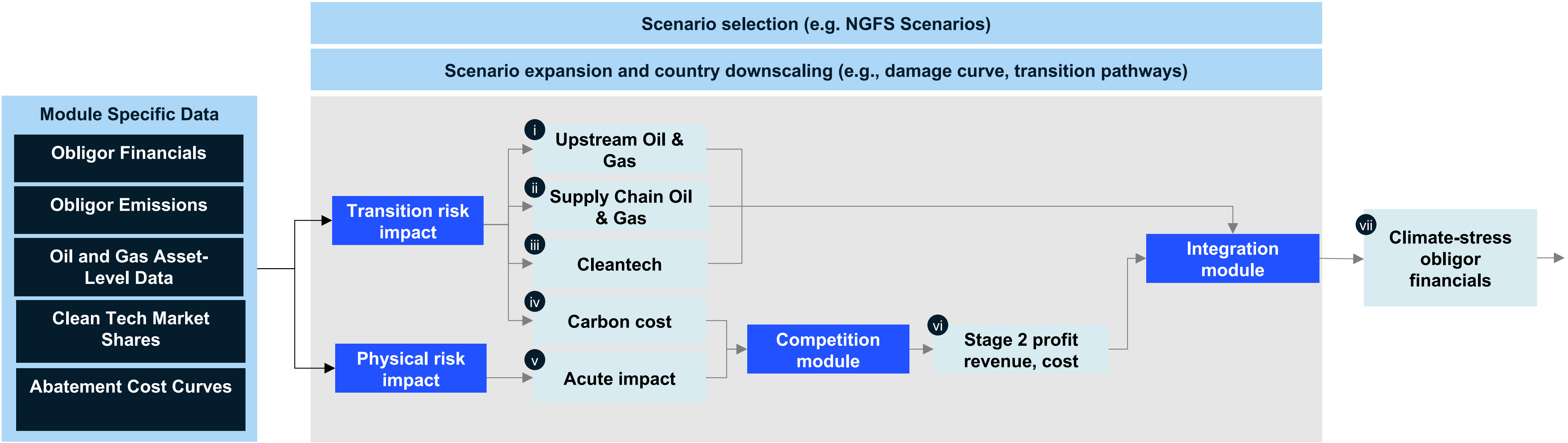
For a North American bank, we identified that 35% of economic profits could erode by 2030 without taking action on key pockets of climate risk exposures



Scenario analyses for corporate and mid-market obligors needs to be sector-specific: oil and gas example

Methodology isolates individual aspects of scenario analysis through modules that calculate Oil and Gas obligor impacts from:

- Oil and Gas demand changes across Upstream and Downstream Operations
- Clean Technology demand changes
- Carbon Costs
- Acute Physical Hazard damage costs

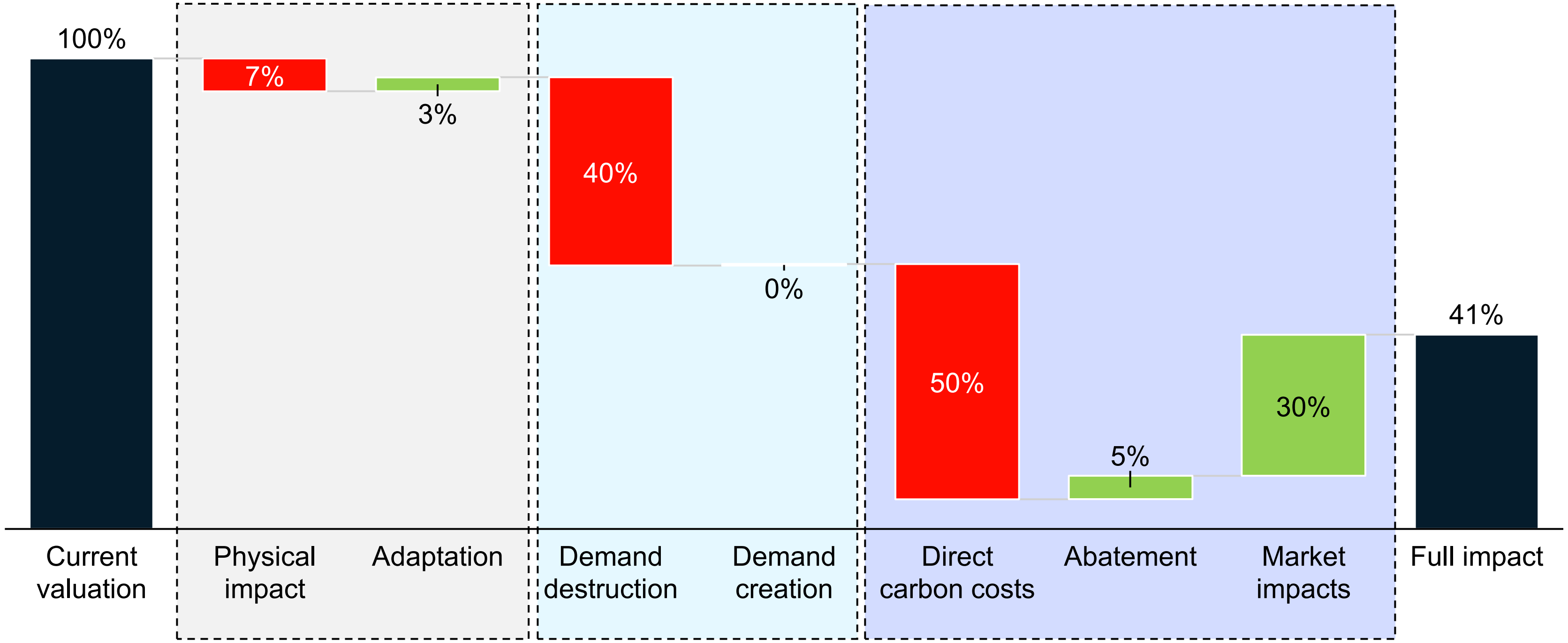


Sanitized example : Demand destruction, direct carbon cost, and market impacts for a specific obligor

Delayed transition scenario – 2050

ASSUMING NO MANAGEMENT/BANK ACTION

Valuation impact waterfall – average impacts weighted by exposure



Physical risk impact

Revenue impact

Driven by current business composition, breakeven cost, and oil and gas price

Carbon cost impact

Driven by carbon emissions which drives carbon costs, abatement and ability to pass cost to customers/ market impacts

Takeaway

- **Demand destruction, direct carbon cost and market impacts** are the top 3 drivers of the climate impact to the O&G portfolio
  - Demand destruction is related to the business composition and breakeven cost
  - Direct carbon cost is related to the carbon price and the amount of output that can be sustained
  - Market impact is related to the ability to pass cost to customers and gain market share from other players

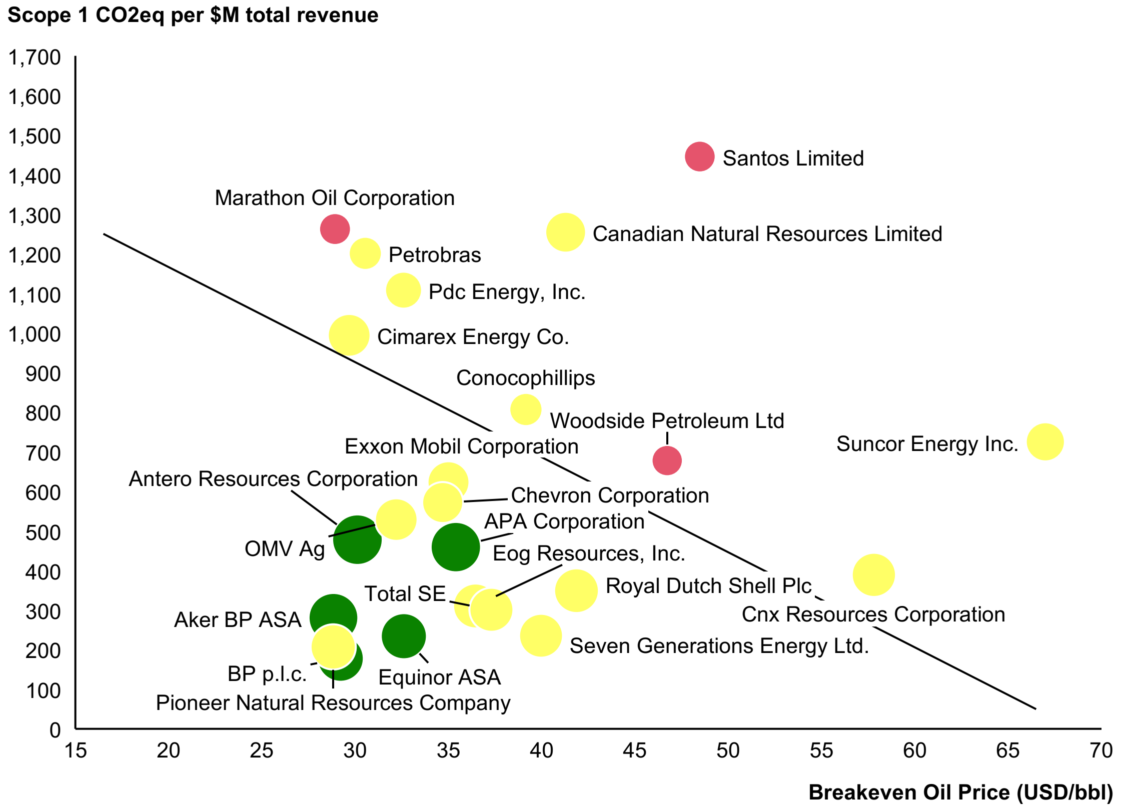
Sanitized example: Portfolio level impact for a sample of public oil and gas obligors

Impact of each driver on earnings for upstream companies

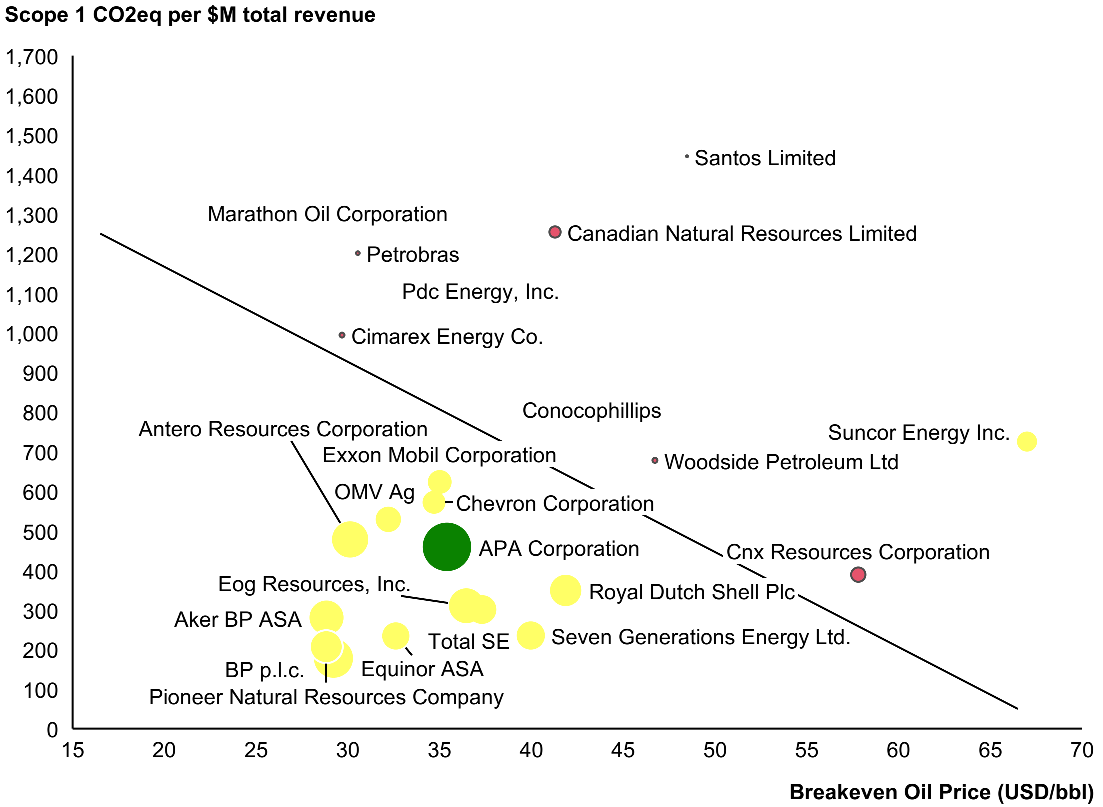
Bubble size: **Market cap**

- 100% of current market cap
- Earnings increase
- Earnings decrease 51-100%
- Earnings decrease <=50%

A 2030: Immediate transition scenario



B 2050: Delayed transition scenario



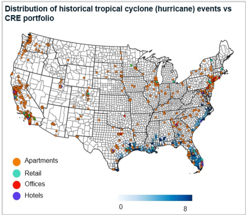
Note: Scenario analysis assumes no mitigation action by bank or borrowers; Source: McKinsey Energy Insights, Rystad

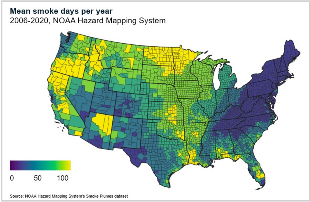
Case study: Using a phased framework to develop bespoke scenario impact assessment tool for CRE portfolios (1/2)

Identify risks and specify scenarios

Identify hazards

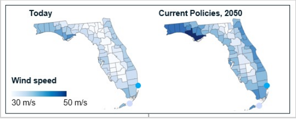
- Flood, hurricane and wildfire most important, given the bank's footprint
- Transition risk leading to GVA impact in counties with high fossil-fuel dependency

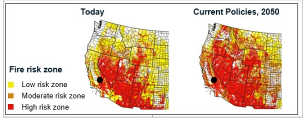




Downscale scenarios

- Hurricane wind severity in every mile across areas where storms have been observed
- Probability of flood and inundation levels

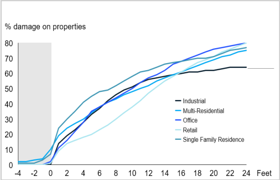




Assess impact on obligor's financials

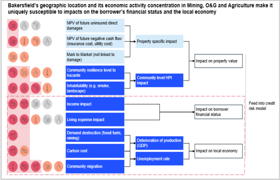
Impact to Net Operating Income

- **Uninsured damage** from physical risk event
- Short term revenue decrease due to **business interruption**
- **Increase in insurance cost** due to increase in frequency and severity of physical risk



Impact to Property Value

- Decline in property value due to change in expected impact of climate events



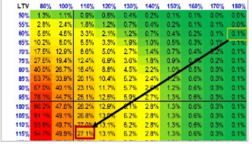
Additional drivers of credit worthiness

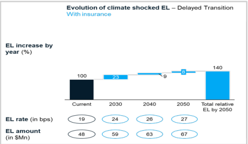
- Community and infrastructure resilience or vulnerability
- Additional sponsor risk
- Risk mitigants like levees

Quantify portfolio impact

Mechanism to translate financials to expected loss

- NOI and property valuation impact Loan-to-value and DSCR of properties, translated into PD, LGD and EL
- Detailed sensitivity testing, e.g.:
  - — Change in cap rate
  - — Insurance unavailability for specific hazard like fire, or in a specific state



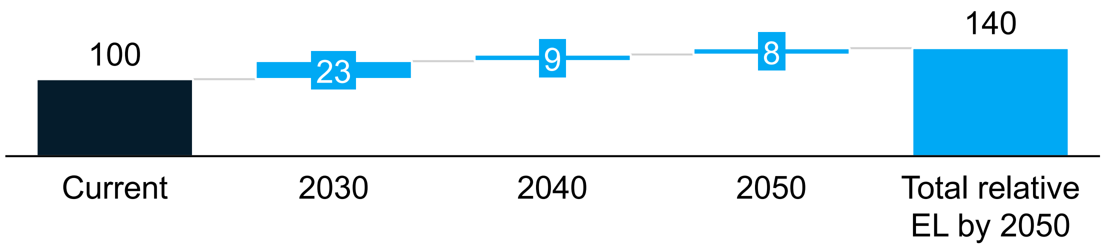


Case study: Using a phased framework to develop bespoke scenario impact assessment tool for CRE portfolios (2/2)

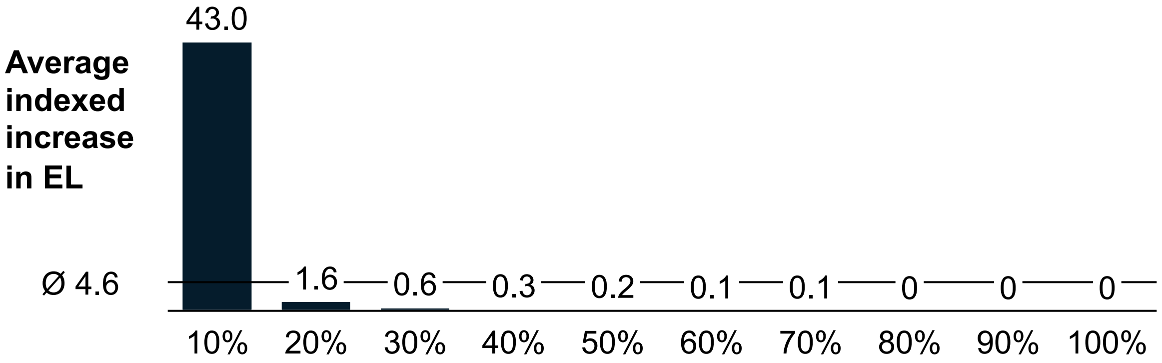
Sanitized example from US based CRE portfolio of a large bank #1

Evolution of climate shocked Expected Loss
Scenario : Delayed Transition

With insurance

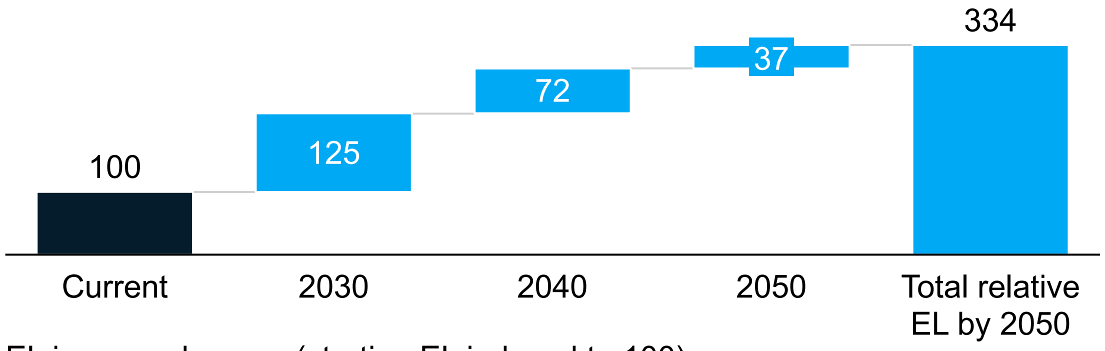


Climate impact to EL arranged by deciles



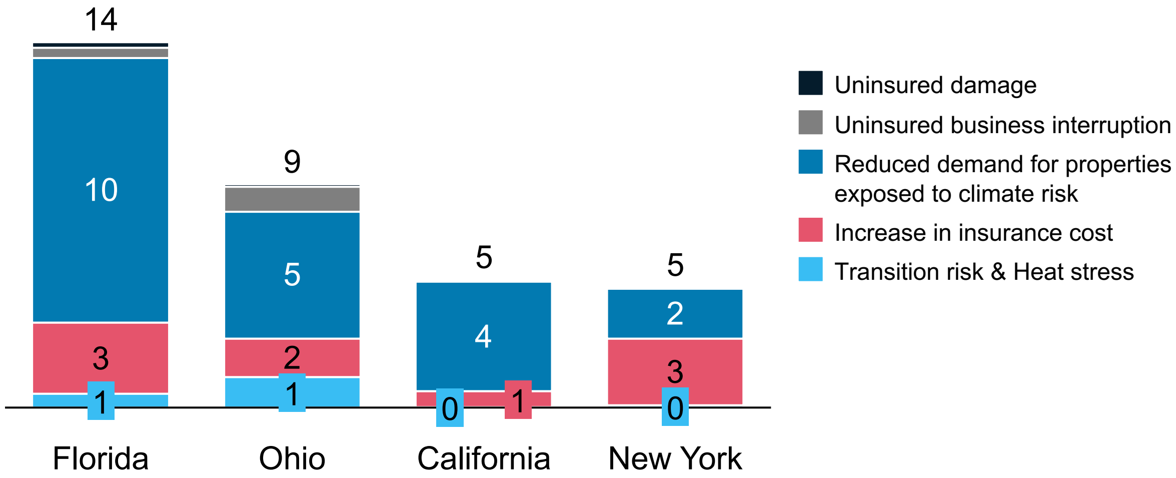
Evolution of climate shocked Expected Loss
Scenario : Delayed Transition

Without insurance



EL increase by year (starting EL indexed to 100)

Net Cash Flow decline by transmission channel



Deposits data can be used to create an up-to-the-minute estimate of SMB's financials

Context

Deposits transactions can be analyzed using a transaction classifier to derive an estimate of

- **SMB's financials (e.g. revenues, revenue growth, and profits)**

Using a reinforcement learning and natural language processing, deposits transactions are analyzed to provide structure to unstructured data

In a client application, this method yield **95% transaction classification accuracy** and over 300 risk signals and competitive indicators, resulting in substantial improvement in predictive power of credit risk models

Risk signals extracted from transactions

SME financial position	Estimated revenues and profits Business seasonality Sectoral dependencies
Competitive information	Multi-banking clients Loans with other lenders Interest rate changes by other lenders Fintech relationships (e.g. Paypal)

Strategic implications

Competition to become the primary bank will intensify

In turn, **banks that host the deposit account where the payroll is deposited have a competitive advantage**

Deposit accounts are already increasingly being bundled with other financial products

Fees for deposit accounts are expected to reduce further

Case example: Leveraging transaction data to transform credit decisioning

Illustrative; Client example

Traditional lending process	Reimagined credit decisioning
Customer completes credit application form	Customer completes online application (pre-populated, leveraging API enabled data sources)
Customer provides bank account and provides simple historic financial statements	ETB: Transactional data from own system, used to build synthetic financial statements NTB: Customer provides permission[1] to access bank transactions data to build synthetic financial statements
RM manually reviews data and performs credit assessment (scoring model, uses credit bureau, and historic/ narrow financial data)	10101 01010 10101 Algorithm integrates data and performs credit assessment (e.g., leveraging transaction classier to build financial statements, link to risk signals)
Manual decision communicated to customer	Decision communicated to customer (instant time yes)

More predictive power

Up to **40%** additional predictive power in some subsegments

More customer insights

95% Accuracy of financial inflows and outflows from classifier

More sophisticated understanding of risks

>300 Predictive risk signals – linked to transaction classifications

An AI-driven asset deployment driven approach can accelerate data quality improvement significantly

1 AI toolkit for data

Detection: Ready to implement package to assess data quality

Correction: Relationship discovery and anomaly detection to find errors

Repair: AI driven correction through an open-loop process

2 Deployment accelerators

Deployable as pipelines that can be stand-alone for immediate results and integrated into Data Platforms to continuously monitor and improve data quality (e.g., Apache Airflow integration with Collibra), platform agnostic deployment

3 Training modules

A new way of working, including roles, talent, an a fast-paced Agile operating model

Co-development of solutions through build-operate-transfer to sustain the impact

When **implemented together** these 3 components significantly accelerate data quality capabilities

Data Quality can be addressed through modularized tools in three core areas

	Detection	Correction	Repair
Objectives	**Measure the quality of each record** on a scale of Low, Medium, High data quality to help prioritize remediation	**Recommend corrections** to data quality errors **prioritized by business impact** and measure the confidence (e.g., 95%)	**Validate recommendations** with experts and incorporate changes into underlying data systems feeding reports
How machine learning helps	Automatic generation of reports on the profile of data, inferred relationships between tables and anomalies, root cause identification to prioritize upstream interventions	Rule-mining and clustering algorithms recommend corrections to data quality errors, quantify confidence, and help estimate business impact	Validate corrections with experts (open-loop) confirming only lower confidence recommendations manually and automated approval of validated changes
Illustration	Detect top 1th percentile interest rate, unusually late maturity date as a potential error	Correct loan type with attribution, country of origination, and address with 95% confidence	Flag and automatically fix reporting dates that were corrected by experts repeated
Ready-to-deploy tools	Data relationship discovery Entity disconnect identification Root cause identification Anomality detection	Automated DQ rule generation Free-form text corrections Comparison to 3rd party data Corrections that integrate attribute, record, and database signals	Collaborative workflows Programmatic data validation

High impact examples AI being used to improve data quality

Root cause detection and correction to errors in 5M customer accounts

Corrected 5M free-form occupations to accelerate AML Customer Risk Rating using a neural language model and traditional fuzzy matching

Accelerating an enterprise data transformation with AI

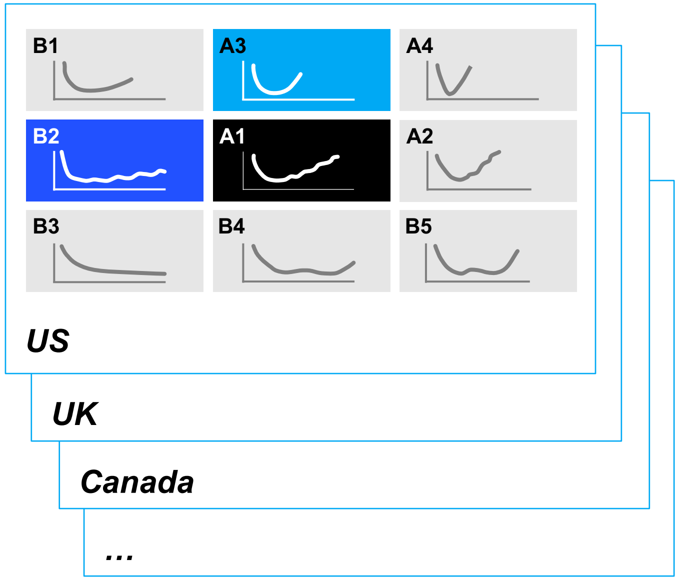
Automated data quality detection and correction with AI (human-in-the-loop) to reduce data transformation timeline by 30-40 percent

Automated identification of errors in CRE loan data

Automated detection of data quality errors in over 100,000 loans using time series anomaly detection to detect issues in real time and prioritize remediation

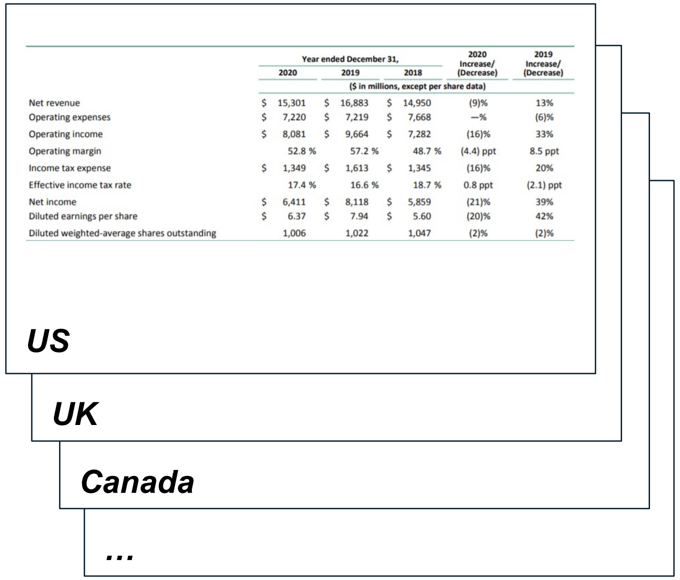
Scenario planning forecasting approach needs to be tailored to bank's footprint flexible to incorporate different driving factors

Scenarios from underlying reasons



Develop inflation scenarios based on set of underlying reasons (e.g., supply chain bottleneck leading to 'cost-push' inflation) and consider incorporation of important macro-linkages (e.g., currency fluctuation risk)

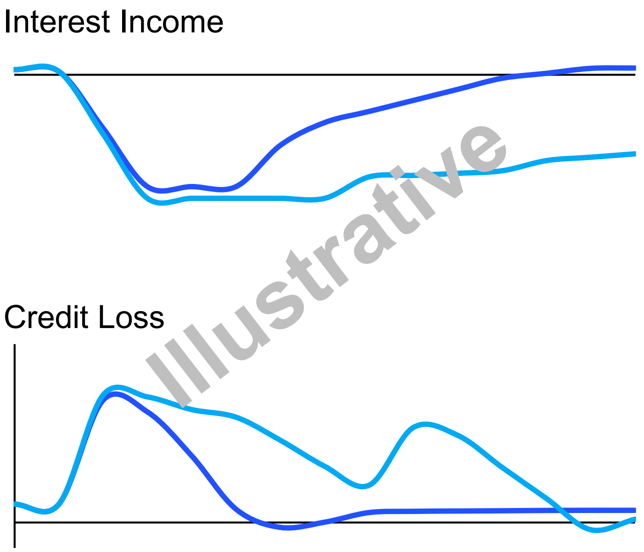
Business drivers from scenarios



	Year ended December 31, 2020	2019	2018	2020 Increase/ (Decrease)	2019 Increase/ (Decrease)
	($ in millions, except per share data)				
Net revenue	$ 15,301	$ 16,883	$ 14,950	(9)%	13%
Operating expenses	$ 7,220	$ 7,219	$ 7,668	—%	(6)%
Operating income	$ 8,081	$ 9,664	$ 7,282	(16)%	33%
Operating margin	52.8 %	57.2 %	48.7 %	(4.4) ppt	8.5 ppt
Income tax expense	$ 1,349	$ 1,613	$ 1,345	(16)%	20%
Effective income tax rate	17.4 %	16.6 %	18.7 %	0.8 ppt	(2.1) ppt
Net income	$ 6,411	$ 8,118	$ 5,859	(21)%	39%
Diluted earnings per share	$ 6.37	$ 7.94	$ 5.60	(20)%	42%
Diluted weighted-average shares outstanding	1,006	1,022	1,047	(2)%	(2)%

Identify drivers that are likely to be inflation-sensitive for business portfolios / geographies, and direction of impact – e.g., inflation may drive up transaction volume in the short term; however, inflation may also reduce demand in the long term

Full estimates from business drivers and optimize across BUs



Develop analytics to project the underlying drivers and the business portfolio financials

Based on projection results, **synthesize implications** for **strategic decision-making**

Uncertainty in economic path require a forecasting approach that is nimble, flexible, and responsive to emerging risks like inflation

Scenario Generation and Refinement

Customized scenarios for portfolio/segments
- Can be macro-economic or event driven
- Recent macro-trends (e.g., EPOP ratio) core to design

↑

Business driver forecast through models

Volume, Revenue, and Expense Forecast Models tied to business drivers (e.g., line utilization) and scenario conditioned

↑

Implementation Engine

One-click solution for aggregation, reporting and visualization

Versatility
- Flexibility to expand current internal scenarios with new variables
- Ability to add macro and event driven overlays
- Ability to refresh scenarios weekly vs. monthly

Coherence
- Model and analytics developed for relevant drivers and separately for relevant geography
- Models capture time trend, macroeconomic factors, seasonality and latent portfolio factors in a transparent manner

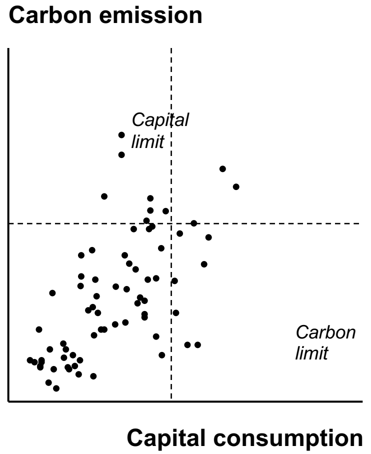
Flexibility
- Designed to capture, test and visualize business actions under different macro-conditions
- Structured to generate report on portfolio and sub-segment level with different ‘what-if’

In addition to standard capital or liquidity constraints, exploring emissions related constraints can help in portfolio alignment

Preliminary list of constraints incorporated in the approach (to be refined based on observations during design phase)

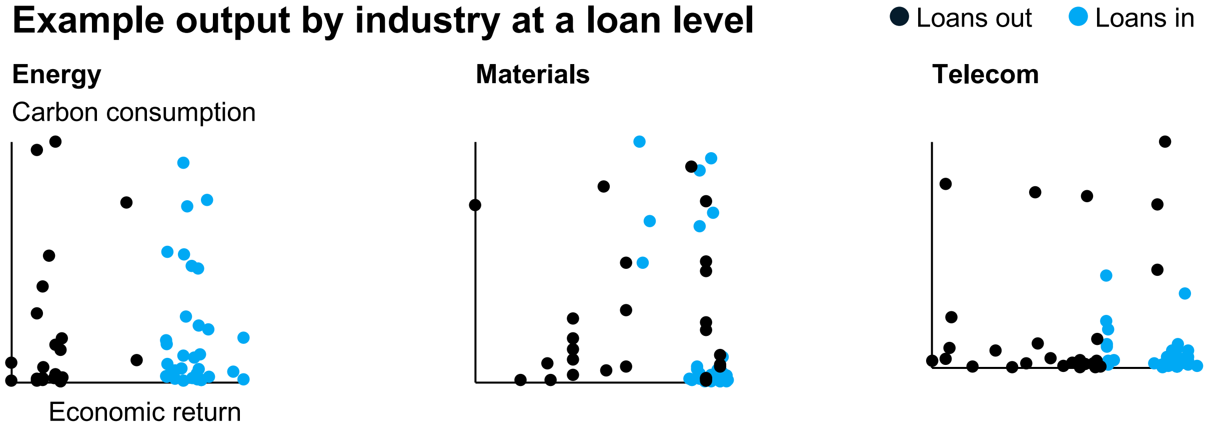
	Product	Industry	BU
Capital	✓	✓	✓
Risk Weighted Assets	✓	✓	✓
Expected Loss	✓	✓	✗
Origination/balance growth	✓	✓	✓
Liquidity coverage ratio	✗	✗	✗
Emissions	✗	✓	✓

Approach to incorporate emissions related constraints



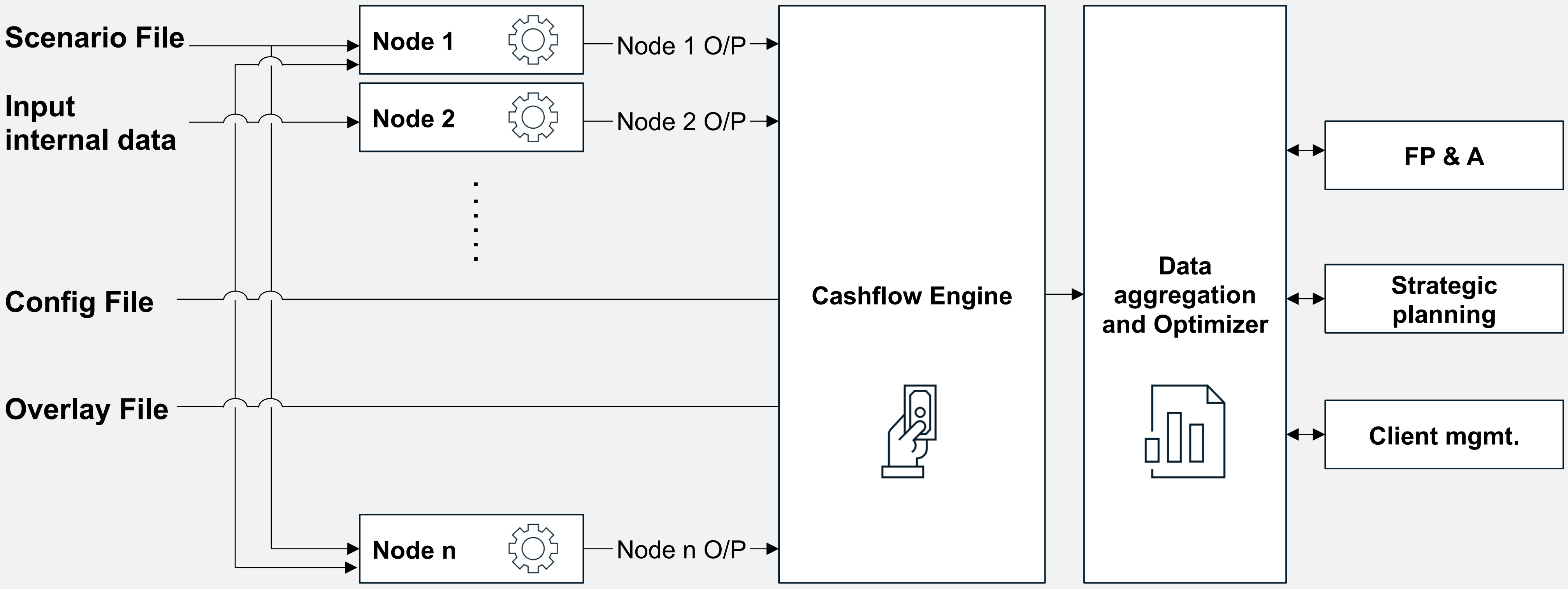
- Include carbon limit based on benchmark scenario at North America with relevant downscaling to Canada/Alberta
- Add constraint for net zero target of total portfolio emissions
- Develop functionality to add sector specific targets and connect with potential sector-specific carbon intensity metric

Example output by industry at a loan level



Microsegment based forecast models are coded in engine with a Balance Sheet Optimizer added for ongoing scenario planning

Example forecasting engine architecture for volume and revenue estimation



Input

Prepare input files and data

Pre-processing (e.g., Python)

Run pre-processing code in Python e.g., overlays, variable transformation

Cashflow calculator (e.g., Spark)

Run microsegment level calculations

Aggregation

Aggregate predictions, conduct sensitivity

Add-on modules

Feed output to additional modules per specific use case

Architecture tailored based on level of manual intervention desired, and size of input data

Content

Summary of survey results	20 mins
Perspectives on selected topics	30 mins
Questions	**10 mins**

Questions