TEAM CELONIS

TUM Data Innovation Lab

Sebastian Roßner

Master in Computational Science and Engineering

Olha Tupko

Master in Mathematical Finance and Actuarial Science

Keesiu Wong

Master in Management

Margarita Ageeva

Master in Data Engineering and Analytics

Ahmed Ayadi

Master in Electrical and Computer Engineering
Master in Management

AGENDA

STRATEGIC GOAL

PHASE I: DATA ANALYSIS

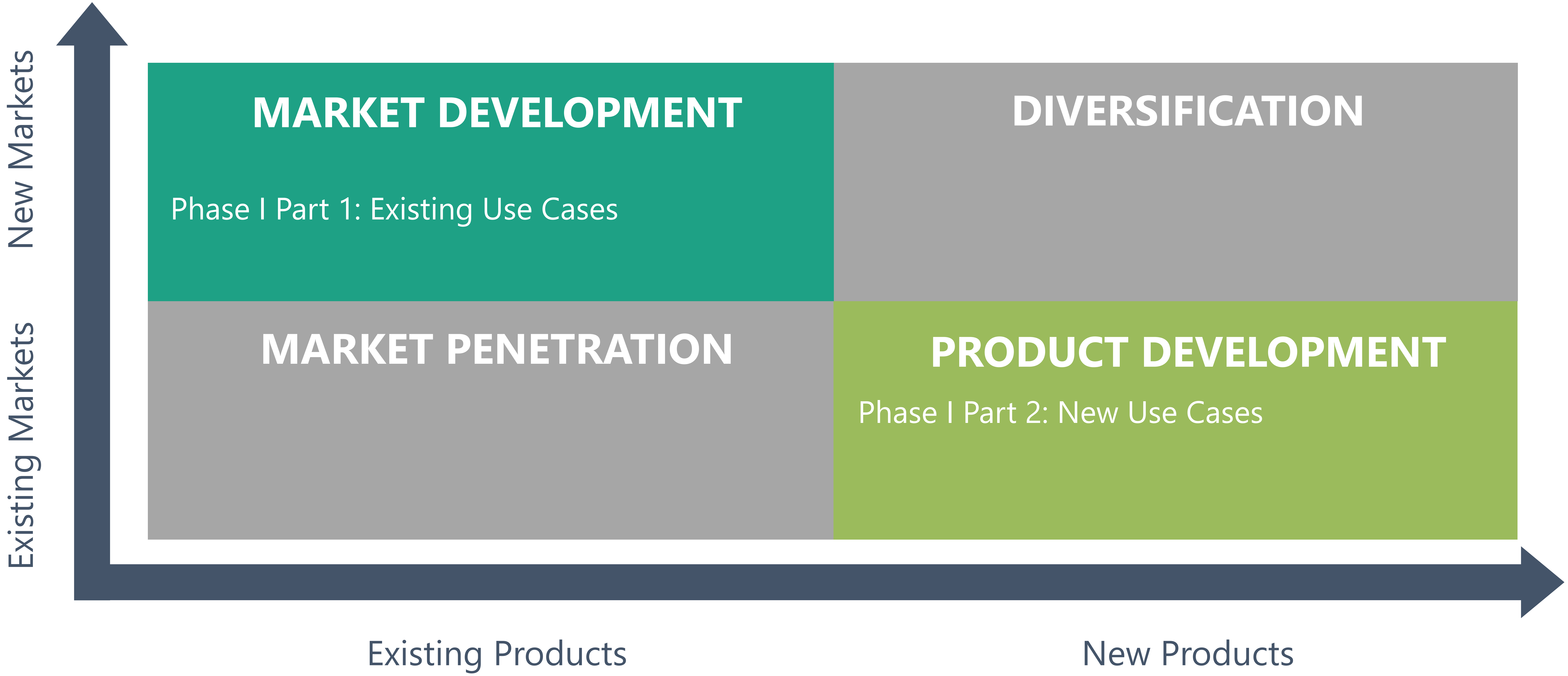
PART 1: EXISTING USE CASES

PART 2: NEW USE CASES

PHASE II: MACHINE LEARNING

STRATEGIC GOAL

Strategic positioning of the project within the Ansoff Matrix



AGENDA

- STRATEGIC GOAL
- PHASE I: DATA ANALYSIS
  - PART 1: EXISTING USE CASES
  - PART 2: NEW USE CASES
- PHASE II: MACHINE LEARNING

PHASE I: DATA ANALYSIS

PART I: EXISTING USE CASES

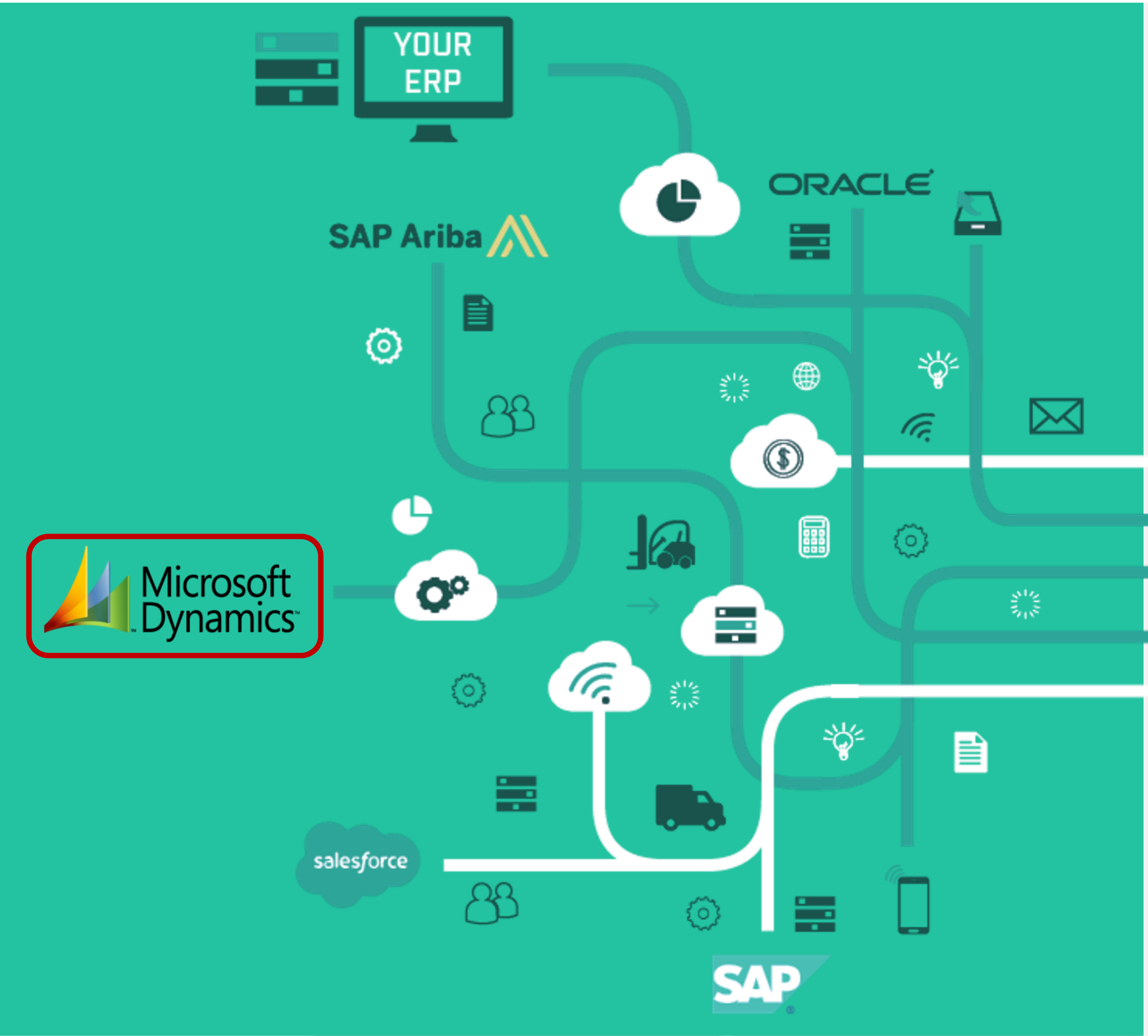
PART II: NEW USE CASES

PHASE I: DATA ANALYSIS

PART I: EXISTING USE CASES

EXISTING USE CASES

11 new Analyses in the Content Store



VISUALIZATION OF THE ACTUAL PROCESSES

AI-POWERED ROOT CAUSE ANALYSIS & IMPROVEMENT

EVENT LOG

2016-12-01 CREATE PURCHASE ORDER #1234
2016-06-23 START PRODUCTION #5678
2016-07-14 RECEIVE PAYMENT #1234
2016-07-14 SEND EMAIL #9012

PHASE I: DATA ANALYSIS

PART I: EXISTING USE CASES

PART II: NEW USE CASES

PHASE I: DATA ANALYSIS

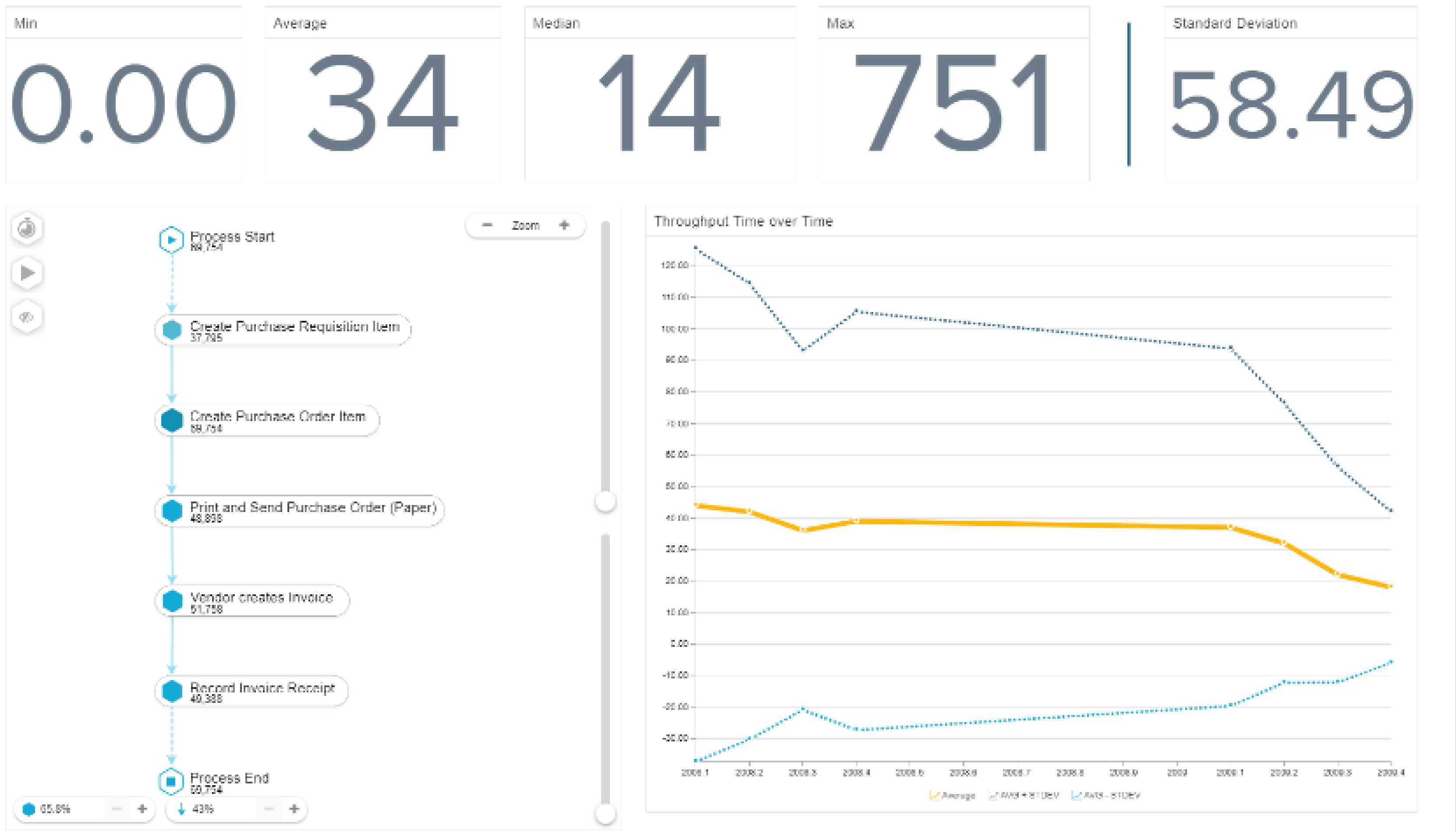
PART II: NEW USE CASES

THREE NEW USE CASES

for any process, for any ERP platform

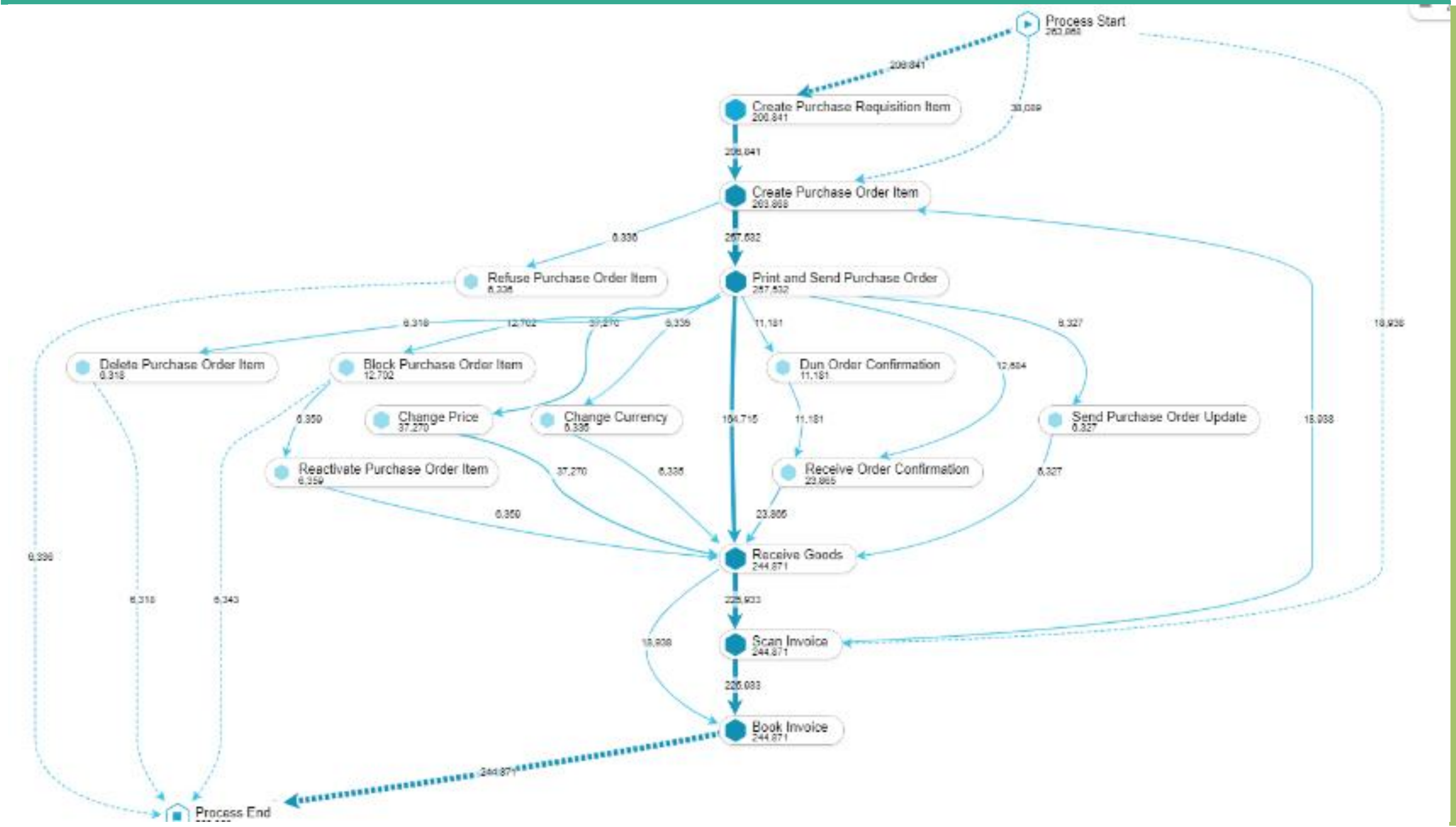
COMPLEXITY ANALYSIS

Get a feeling for the complexity for your processes – *quantified*.



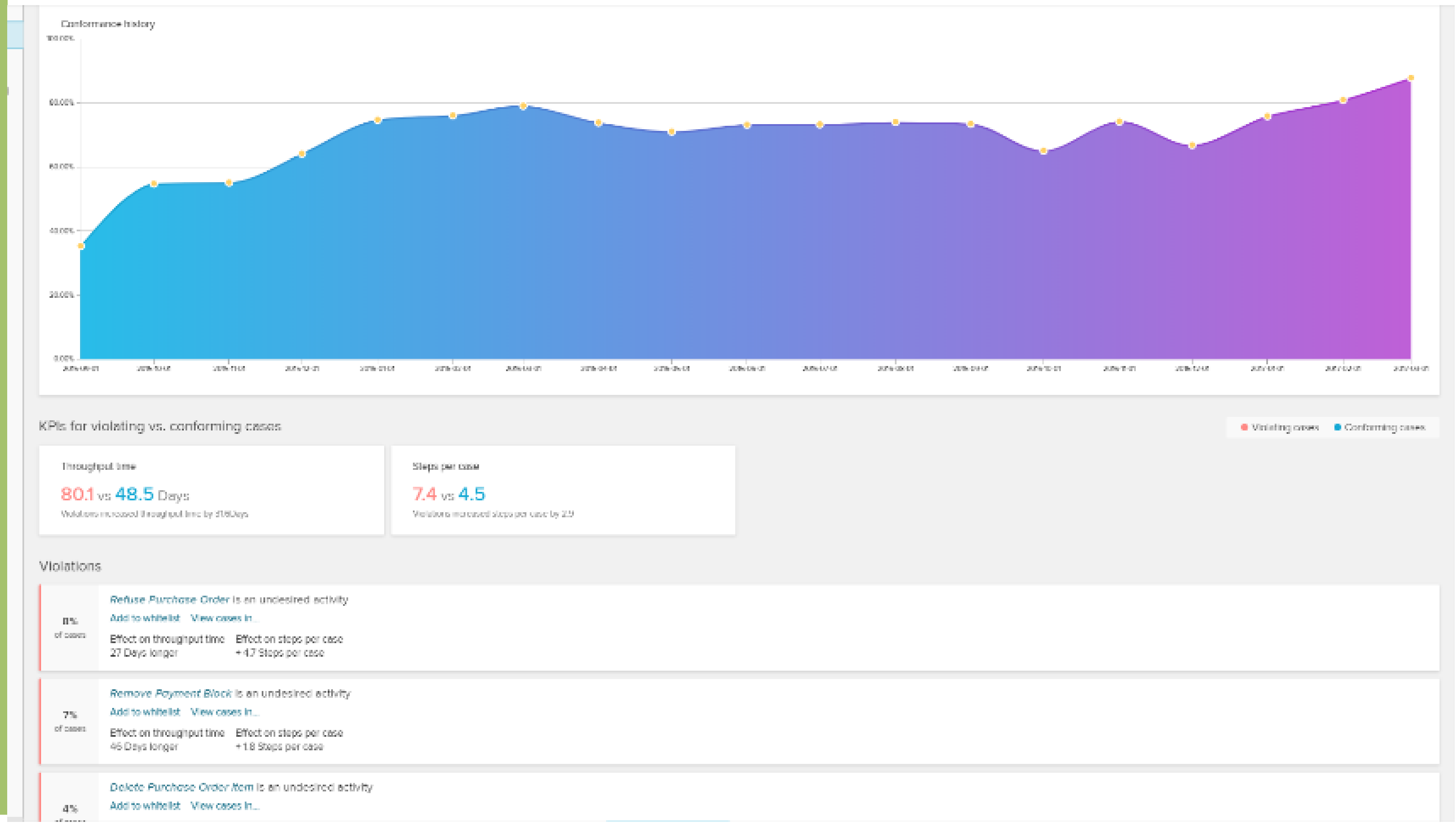
UNDESIRED ACTIVITIES

Analyze where undesired activities happen – *and why*.



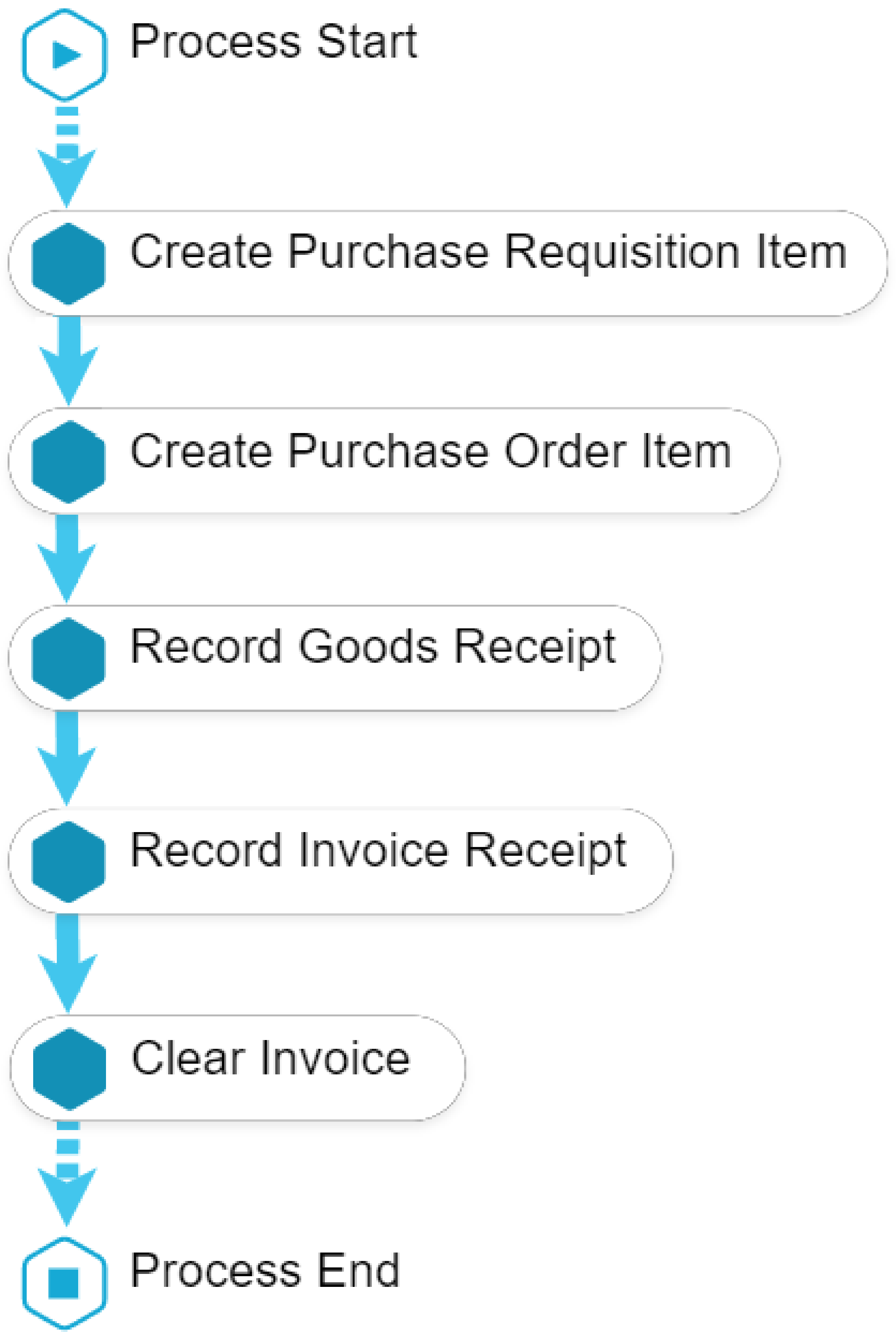
EFFICIENCY ANALYSIS

See exactly which process steps can be improved – *for sure*.



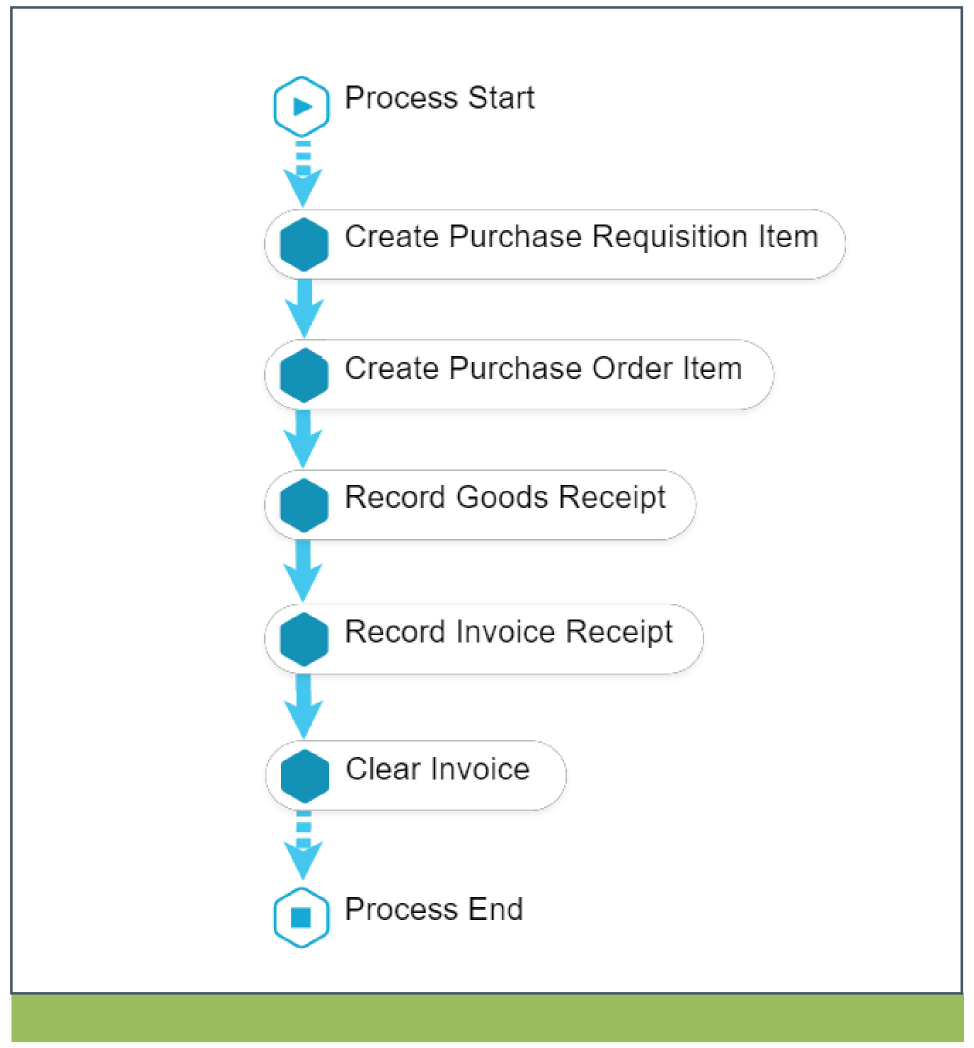
COMPLEXITY ANALYSIS

Get a feeling for the complexity of your processes – quantified.



COMPLEXITY ANALYSIS

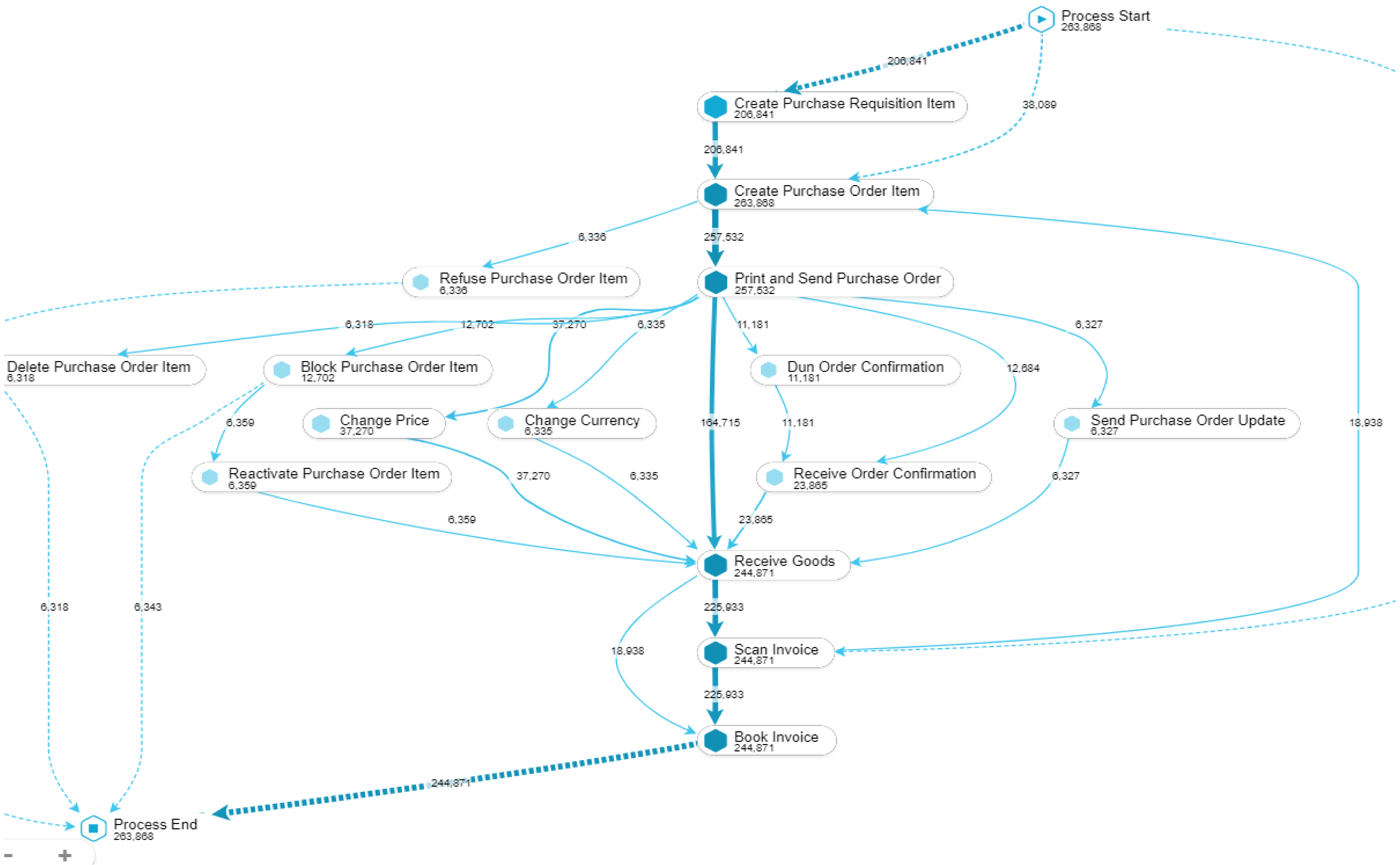
Get a feeling for the complexity of your processes – quantified.



Perfect world

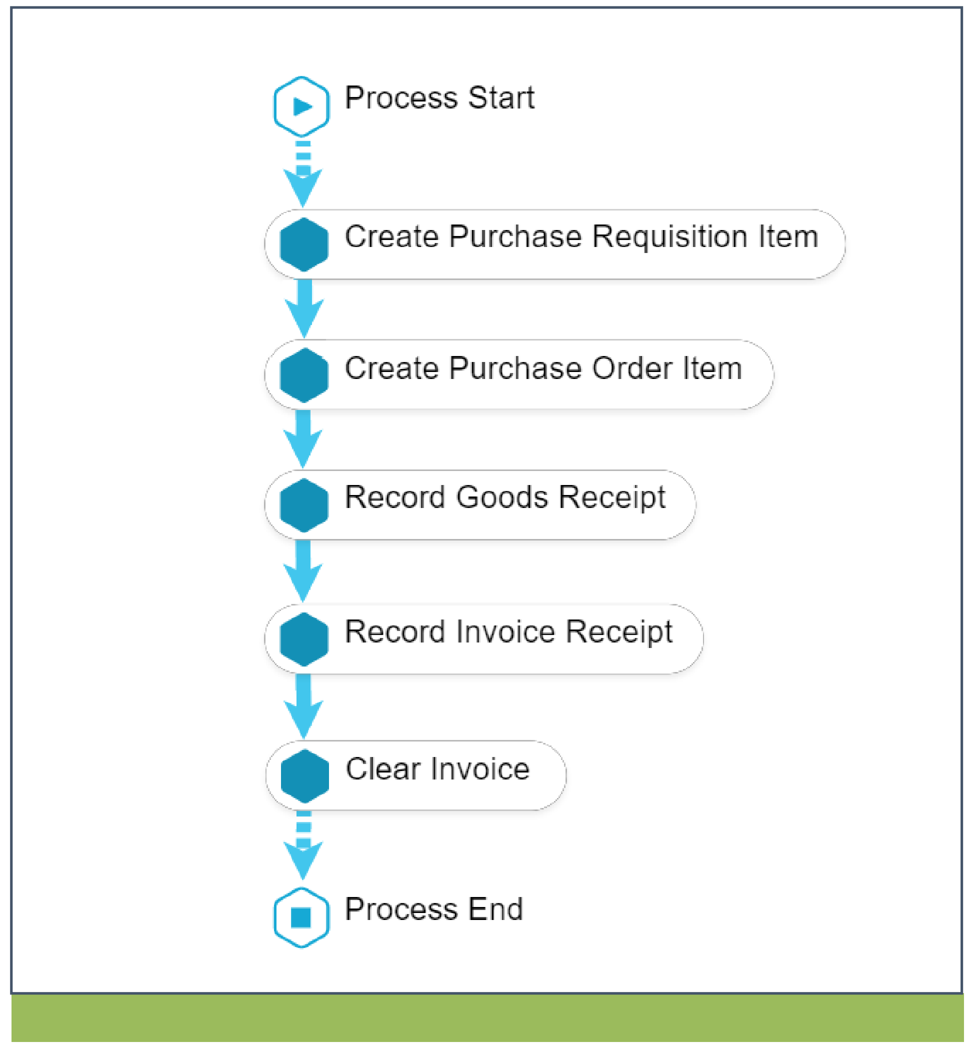
COMPLEXITY ANALYSIS

Get a feeling for the complexity of your processes – quantified.

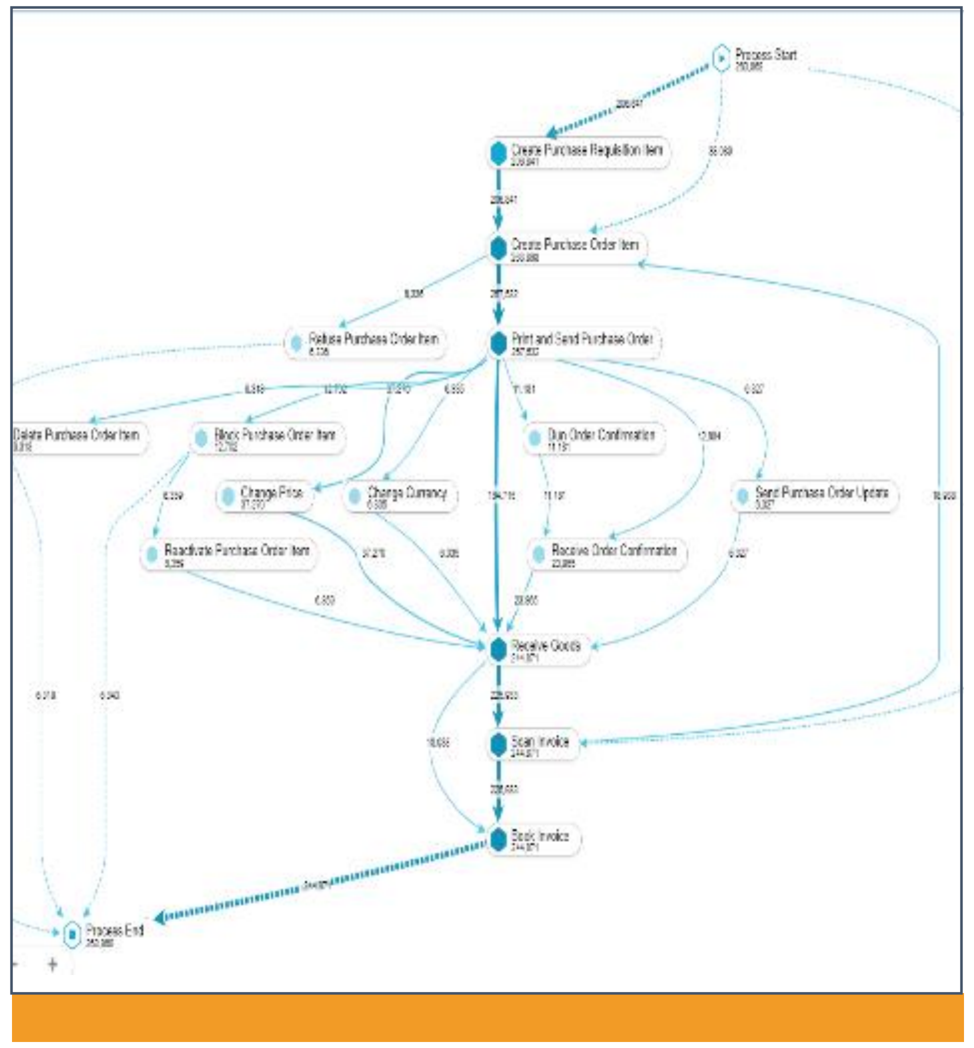


COMPLEXITY ANALYSIS

Get a feeling for the complexity of your processes – quantified.



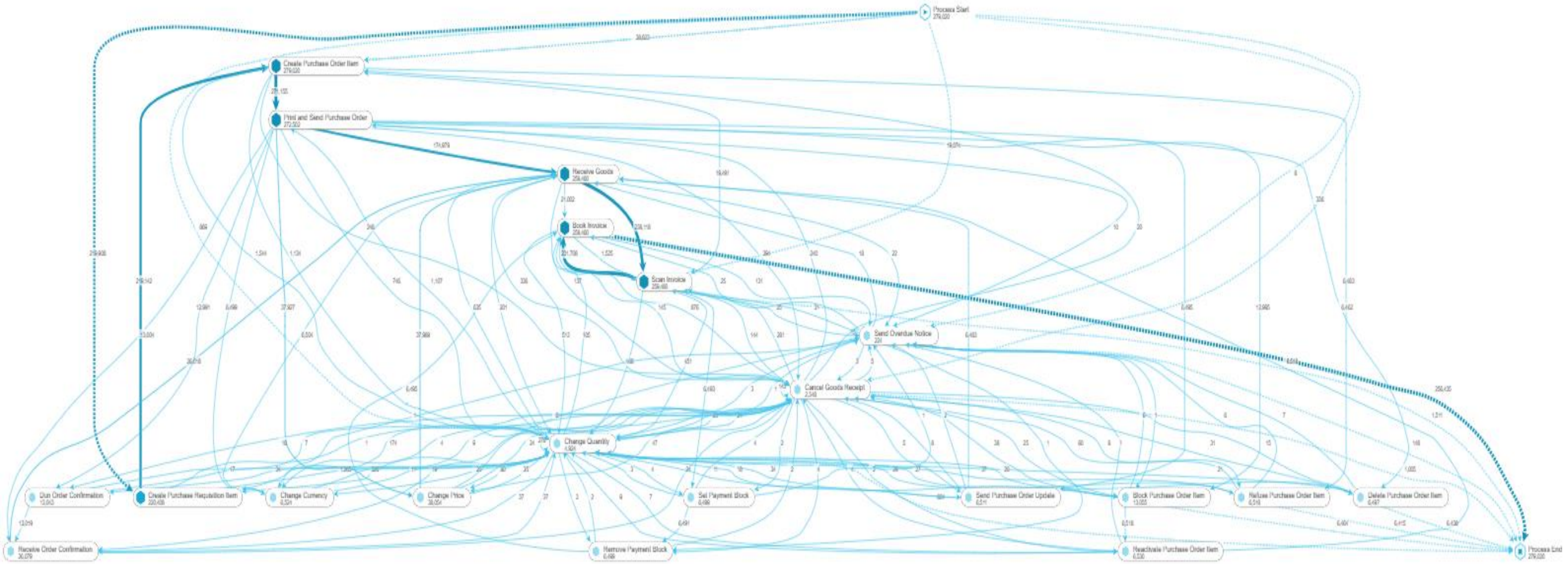
Perfect world



Expectation

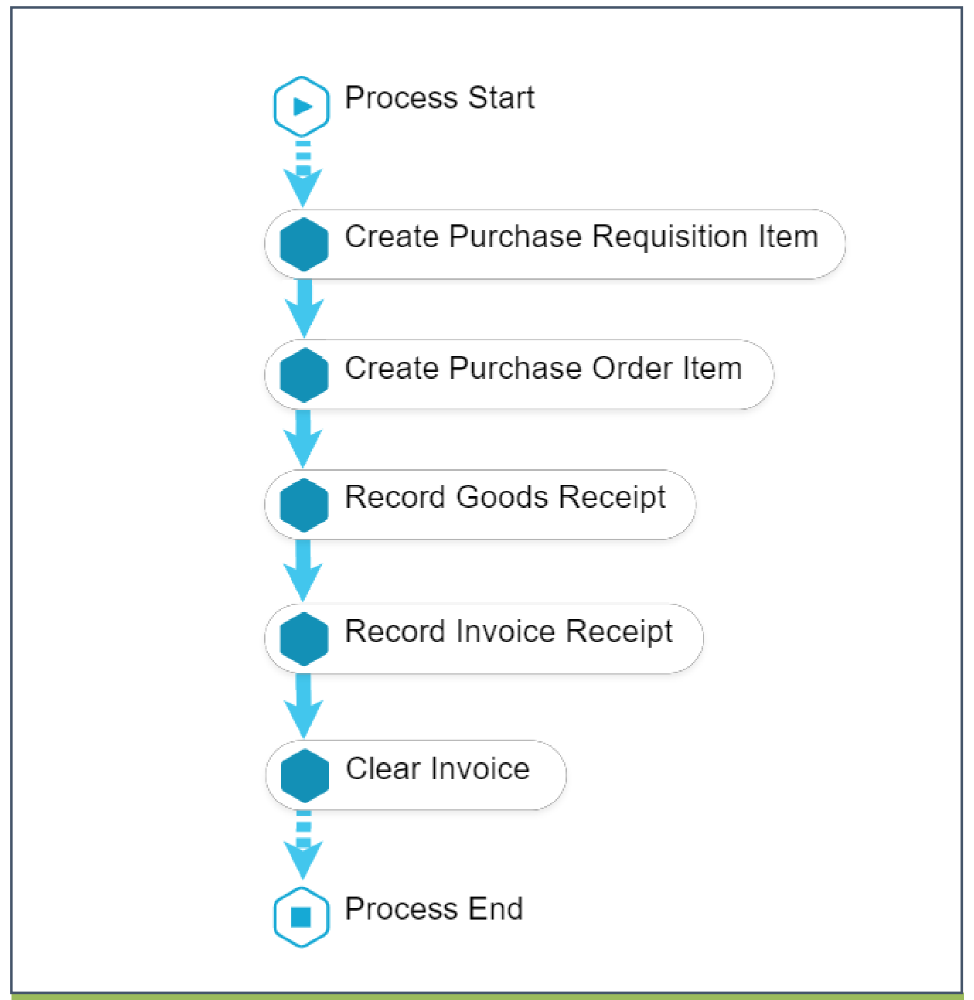
COMPLEXITY ANALYSIS

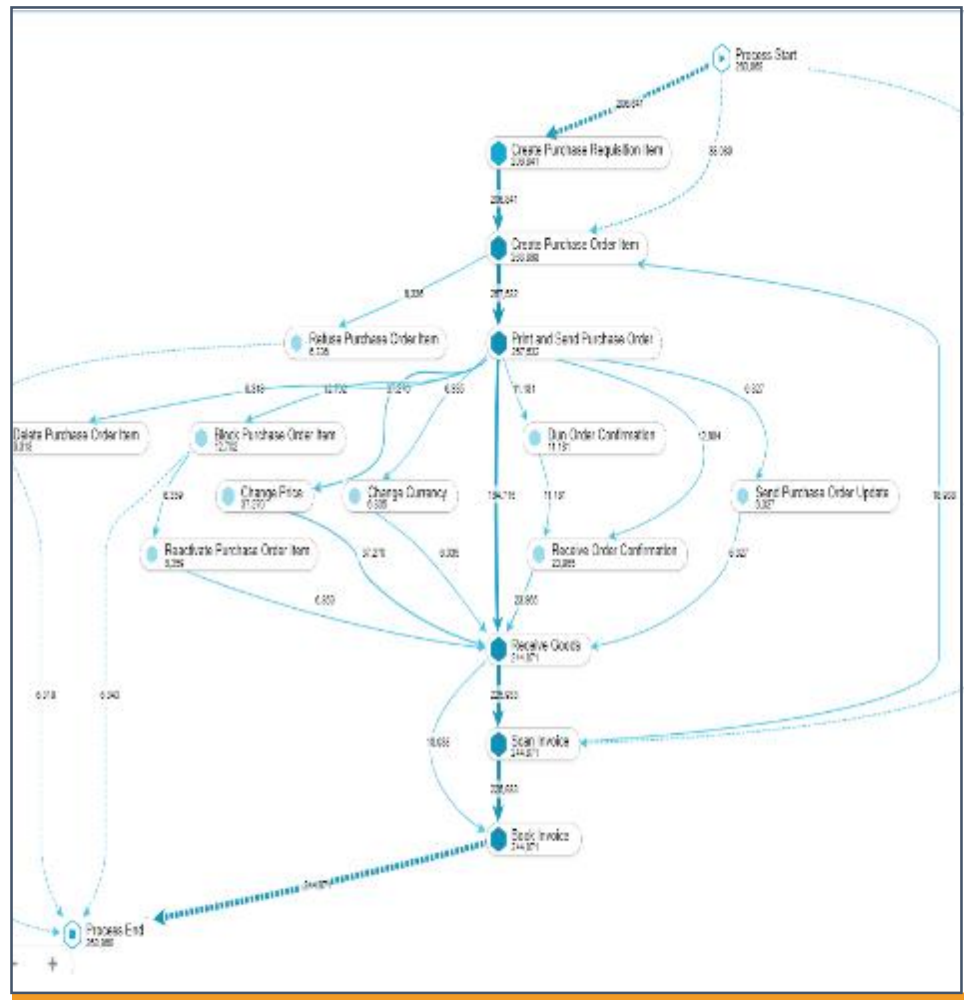
Get a feeling for the complexity of your processes – quantified.

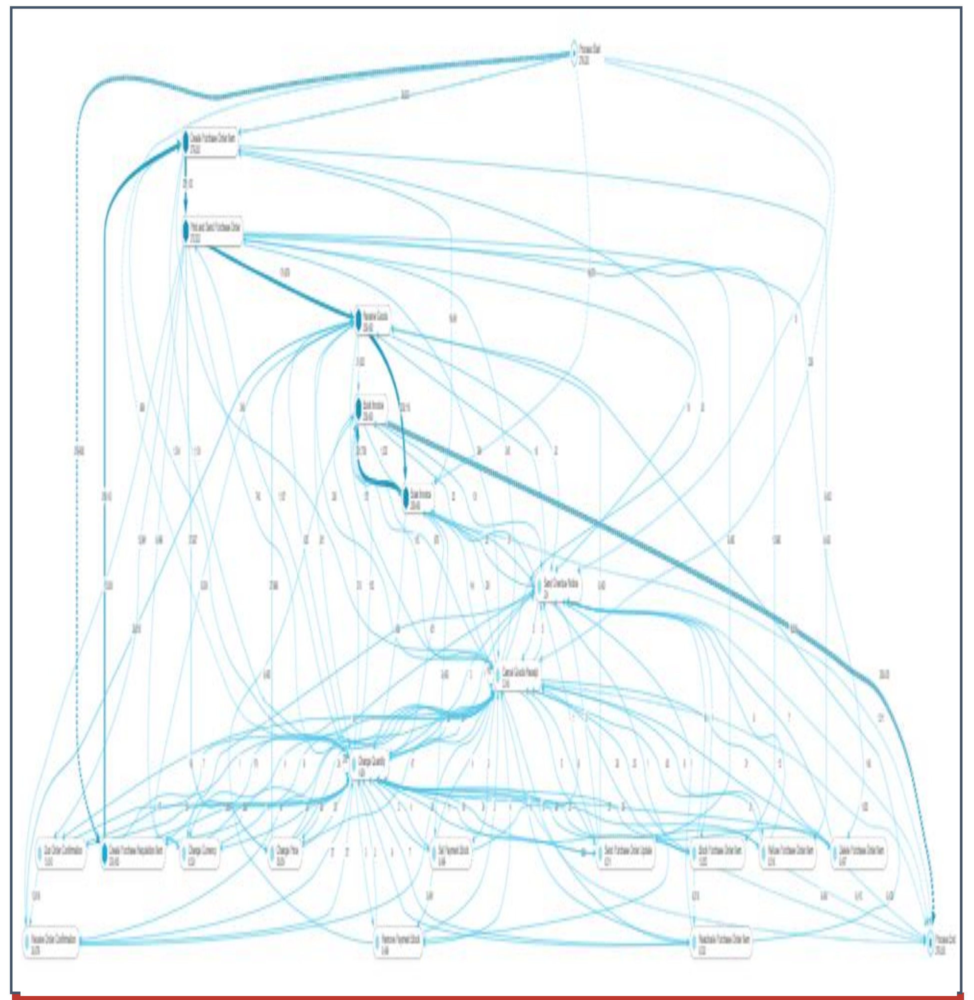


COMPLEXITY ANALYSIS

Get a feeling for the complexity of your processes – quantified.







Perfect world **Expectation** **Reality**

COMPLEXITY ANALYSIS

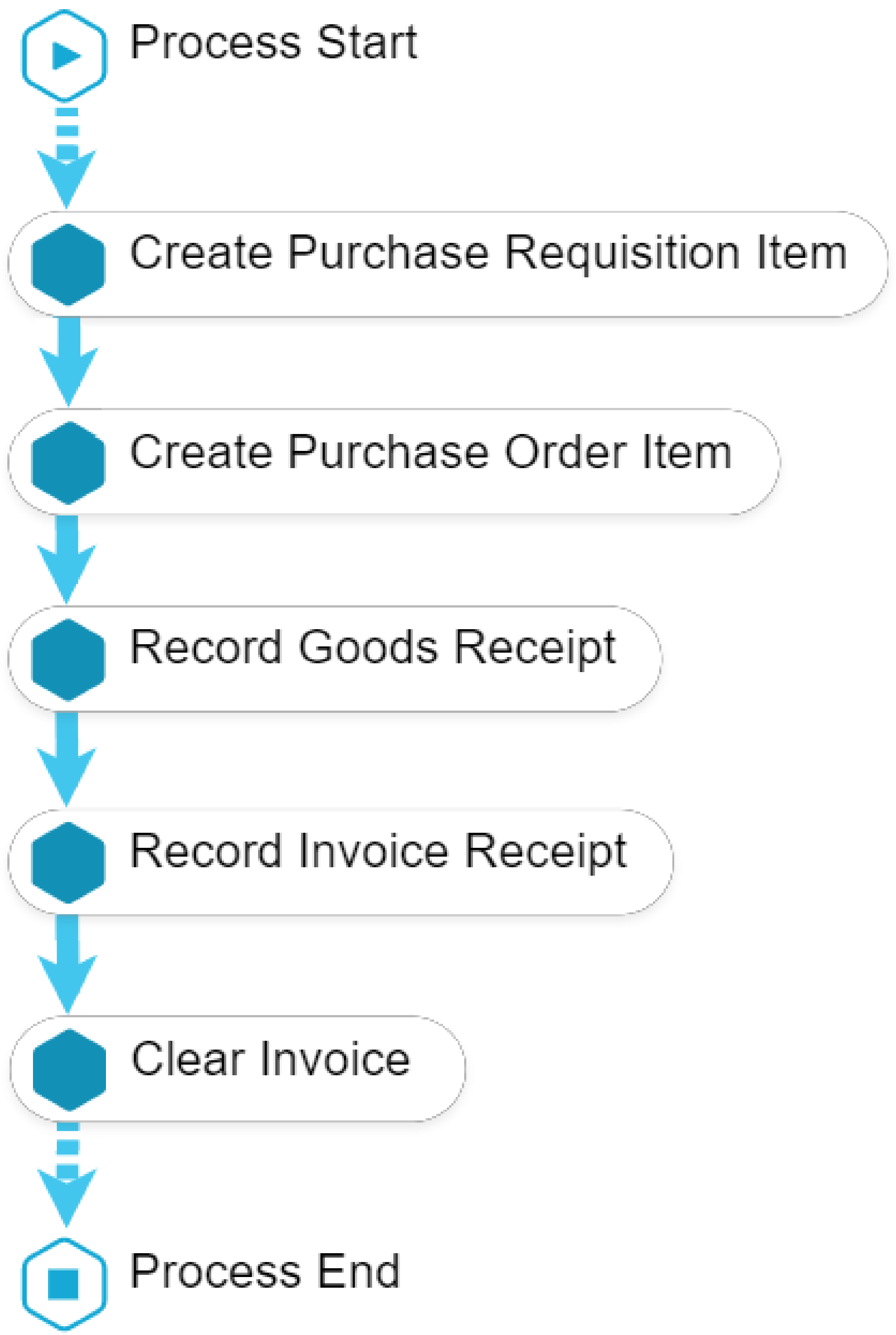
Get a feeling for the complexity of your processes – quantified.

AGENDA

- STRATEGIC GOAL
- PHASE I: DATA ANALYSIS
  - PART 1: EXISTING USE CASES
  - PART 2: NEW USE CASES
- PHASE II: MACHINE LEARNING

ROBLEM

where machine learning empowers process mining



ARIMA

Classical Statistics

ARD Regression

Bayesian Trade-off

LSTM

Advanced Machine Learning

ARIMA

Auto**R**egressive **I**ntegrated **M**oving **A**verage

$$(1-\sum_{i=1}^{p} L^{i})(1-L^{d})y_t = (1+\sum_{i=1}^{q} \theta_i L^{i})\varepsilon_t$$

Advantages

- Time series specific
- Removes polynomial trend
- Easy to implement
- Computationally fast

Limitations

- Works with stationary time series only
- Cannot deal with trend other then polynomial

ARD

Automatic Relevance Determination regression

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{X} + \varepsilon$$

$$p(\mathbf{w}|\lambda) \sim N(\mathbf{0}, \Lambda^{-1}), \Lambda = diag(\lambda_1, \ldots, \lambda_N)$$

$$p(\varepsilon_i|\alpha) \sim N(0, \alpha^{-1}), \{\varepsilon_i\}_{i \in \overline{1,N}} - i.i.d.$$

Advantages

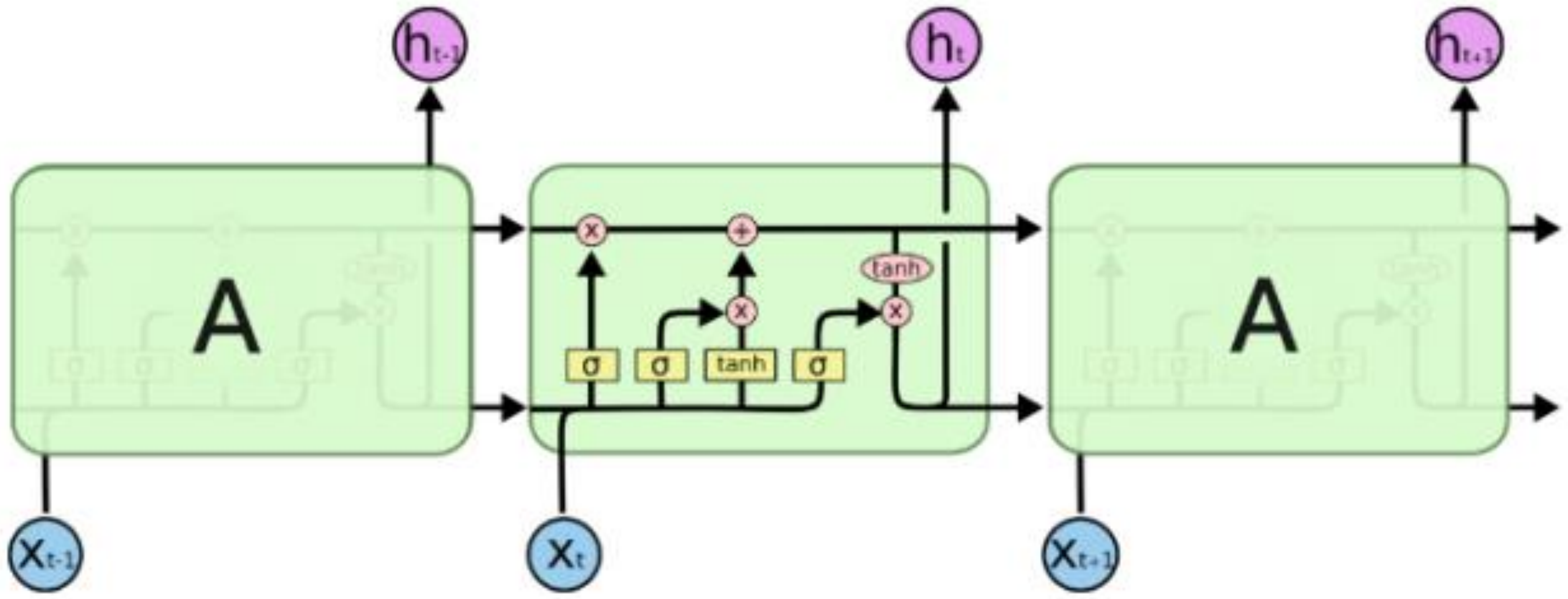
- Automatic feature selection
- Confidence intervals
- Easy to implement
- Computationally fast

Limitations

- ~~Not times series specific~~
- ~~Not robust to distributional changes~~

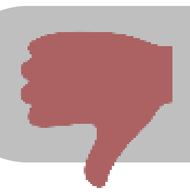
LSTM

Long **S**hort-**T**erm **M**emory recurrent neural network



Advantages

- Sequential data specific
- Captures long and short-term dependencies
- Can handle complex data structures

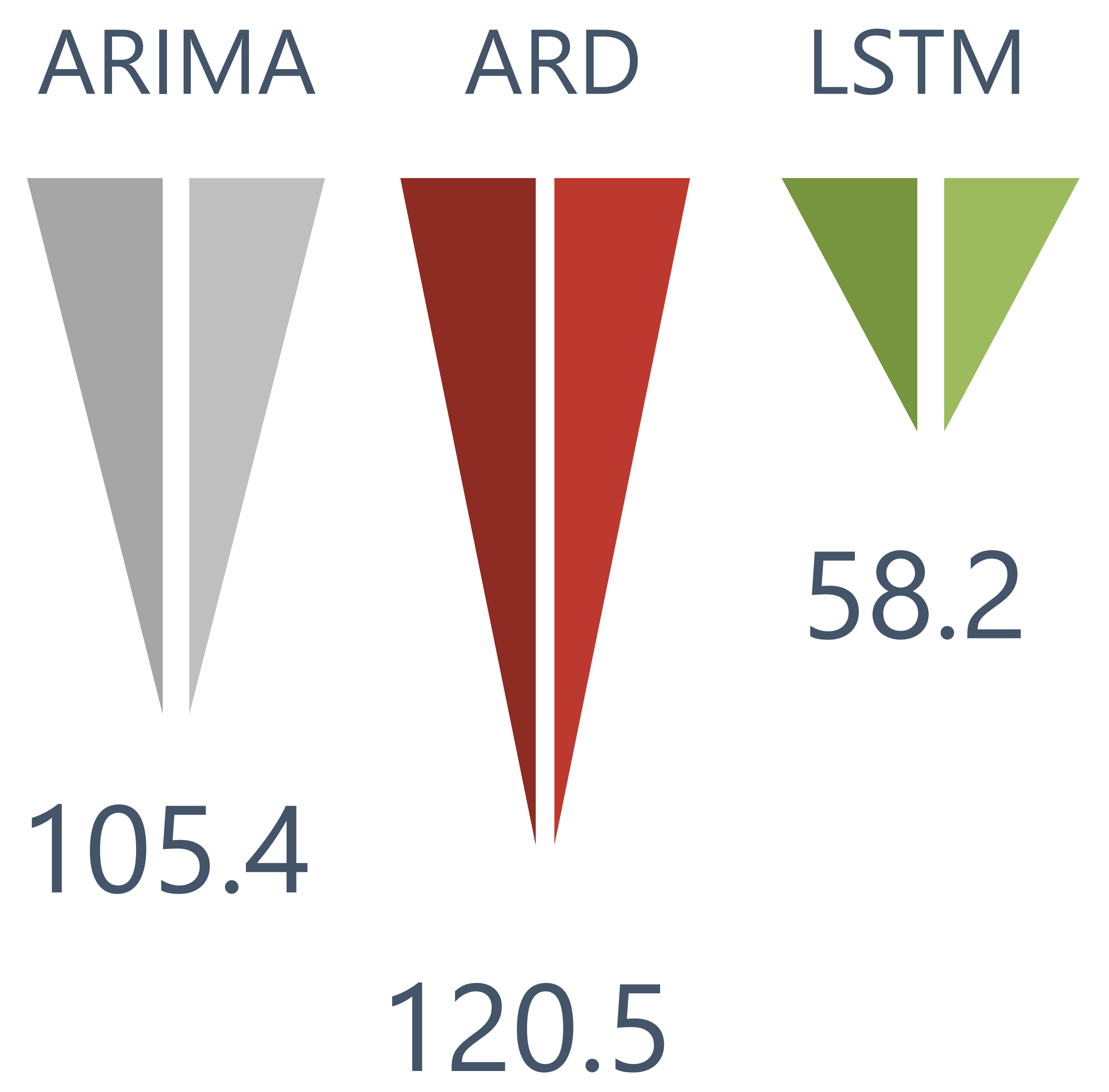


Limitations

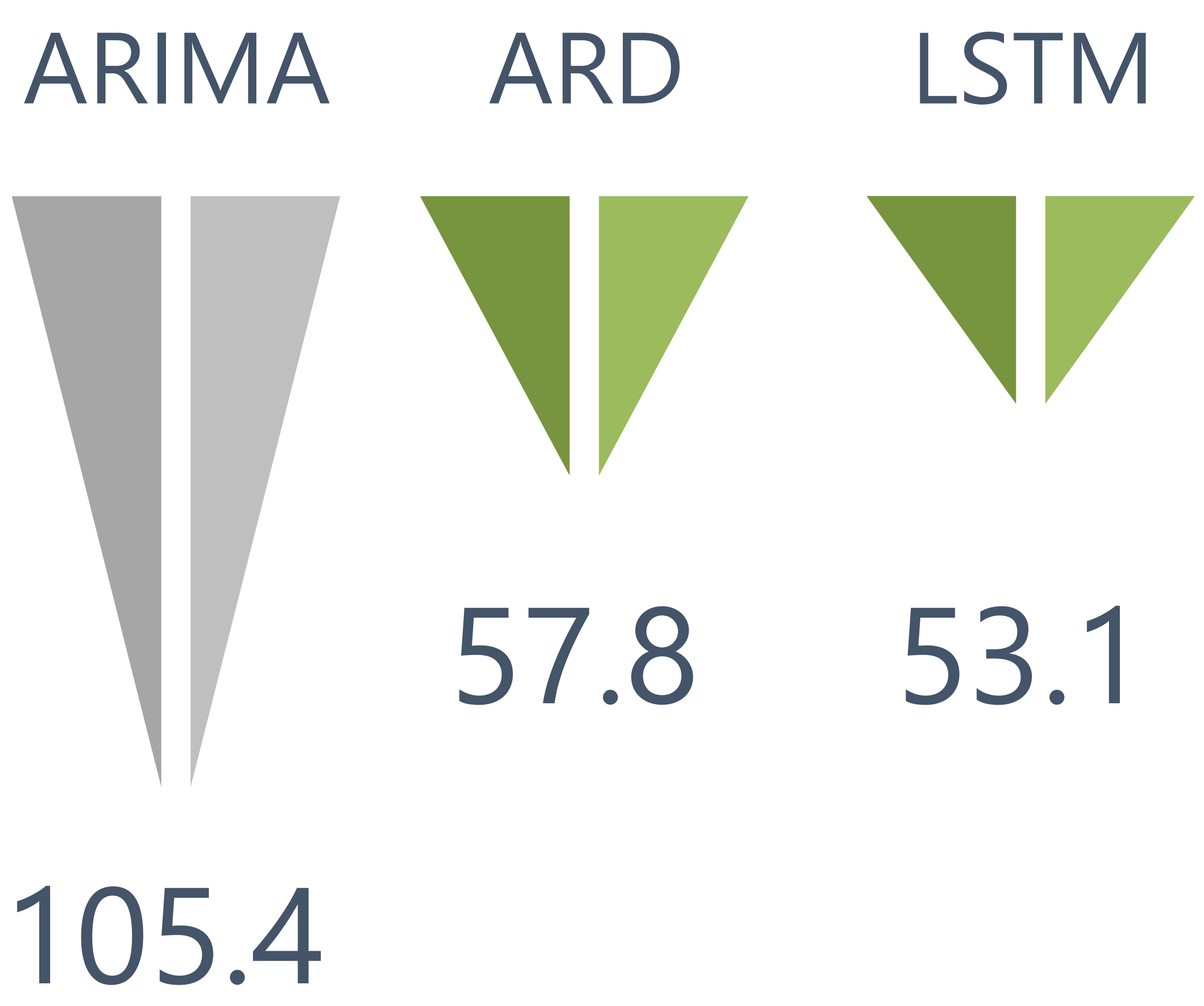
- Gives point estimates
- Requires large datasets
- Computationally expensive

RESULTS

in terms of Mean Absolute Error



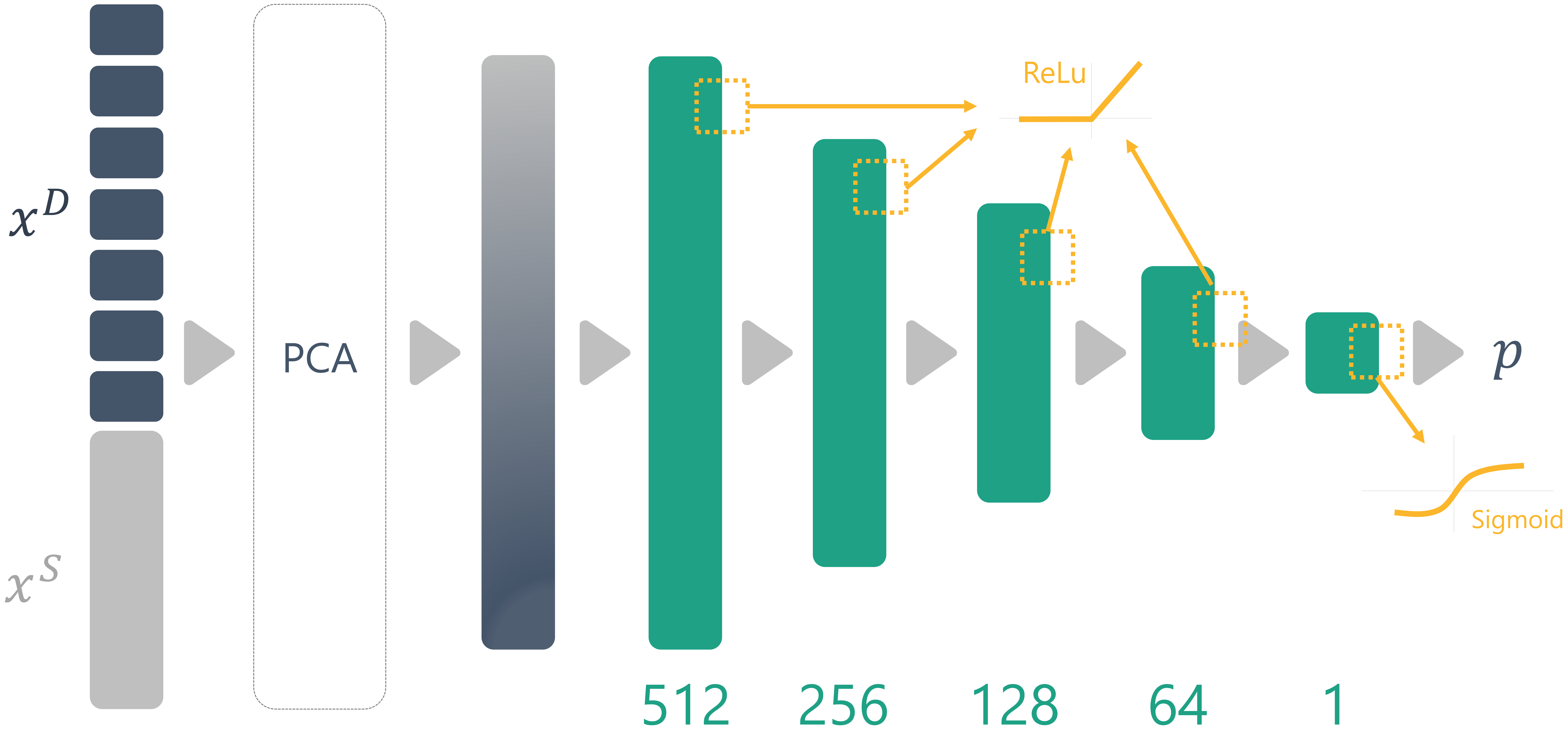
Before Truncation



After Truncation

MACHINE LEARNING

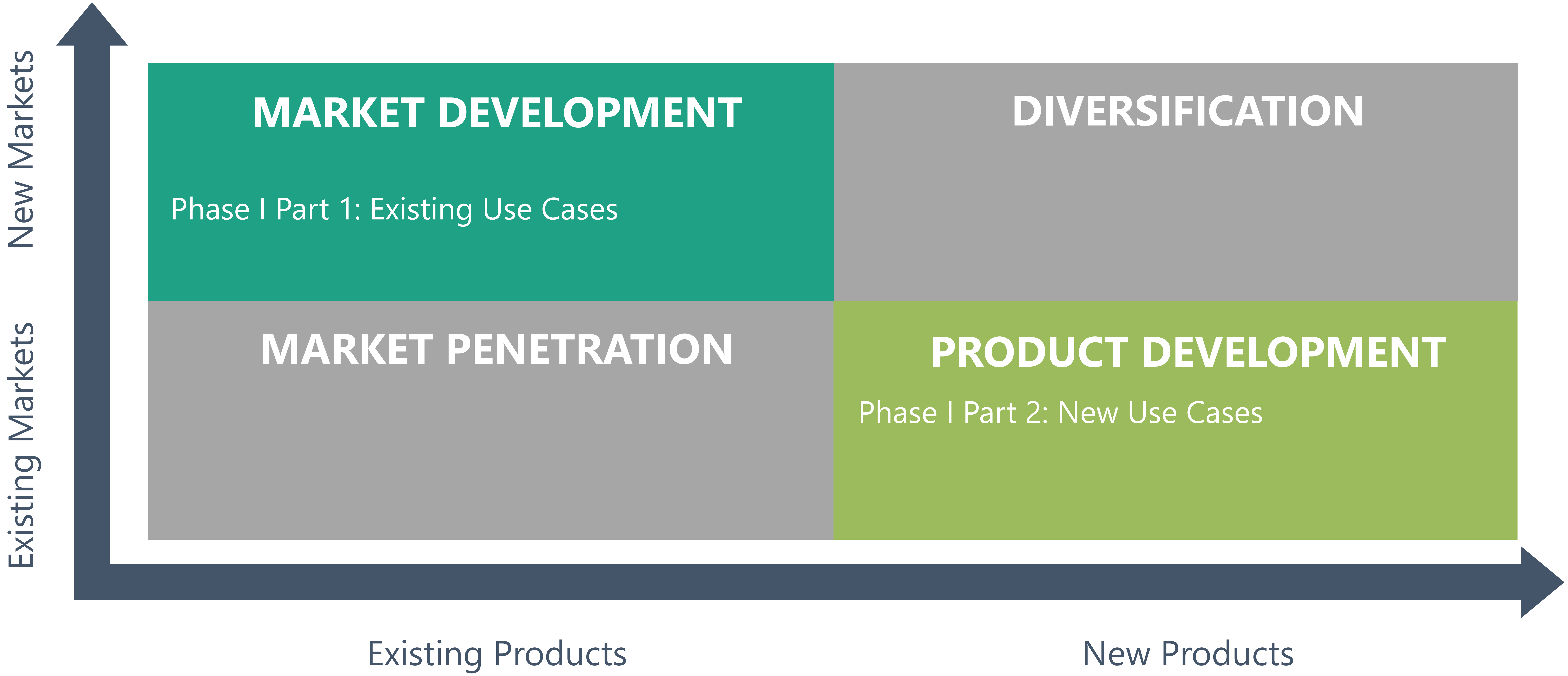
1. Naïve Bayes
2. Logistic Regression
3. K-Nearest Neighbors
4. Decision Trees
5. Random Forests
6. Feedforward Neural Networks


x^D
x^S
PCA
ReLu
p
Sigmoid
512
256
128
64
1

	Naïve Bayes	Logistic Regression	KNN	Decision Trees	Random Forests	Neural Networks
1	38	14	59	60	63	85
2	40	37	64	70	74	88
3	42	41	69	75	79	89
4	46	50	75	81	85	85
5	46	63	83	85	89	92
6	47	60	77	79	86	93
7	55	68	81	79	84	90
8	52	72	80	80	86	89
9	50	66	78	79	81	84
10	59	74	78	81	85	83
Overall	43	45	65	74	78	88

	Naïve Bayes	Logistic Regression	KNN	Decision Trees	Random Forests	Neural Networks
1	38	14	59	60	63	**85**
2	40	37	64	70	74	**88**
3	42	41	69	75	79	**89**
4	46	50	75	81	**85**	**85**
5	46	63	83	85	89	**92**
6	47	60	77	79	86	**93**
7	55	68	81	79	84	**90**
8	52	72	80	80	86	**89**
9	50	66	78	79	81	**84**
10	59	74	78	81	**85**	83
Overall	43	45	65	74	78	**88**

CONCLUSION


New Markets
Existing Markets
MARKET DEVELOPMENT
Phase I Part 1: Existing Use Cases
DIVERSIFICATION
MARKET PENETRATION
PRODUCT DEVELOPMENT
Phase I Part 2: New Use Cases
Existing Products
New Products


thank you