



How to use Machine Learning to reach your goals

Wind Energy Denmark – October 1, 2019

The Data Analytics Value Proposition:

1

What happened?
Descriptive analytics

BI

2

Why did it happen?
Causal analytics

**Econometrics
and statistics**

3

What will happen?
Predictive analytics

**Machine
learning etc.**

4

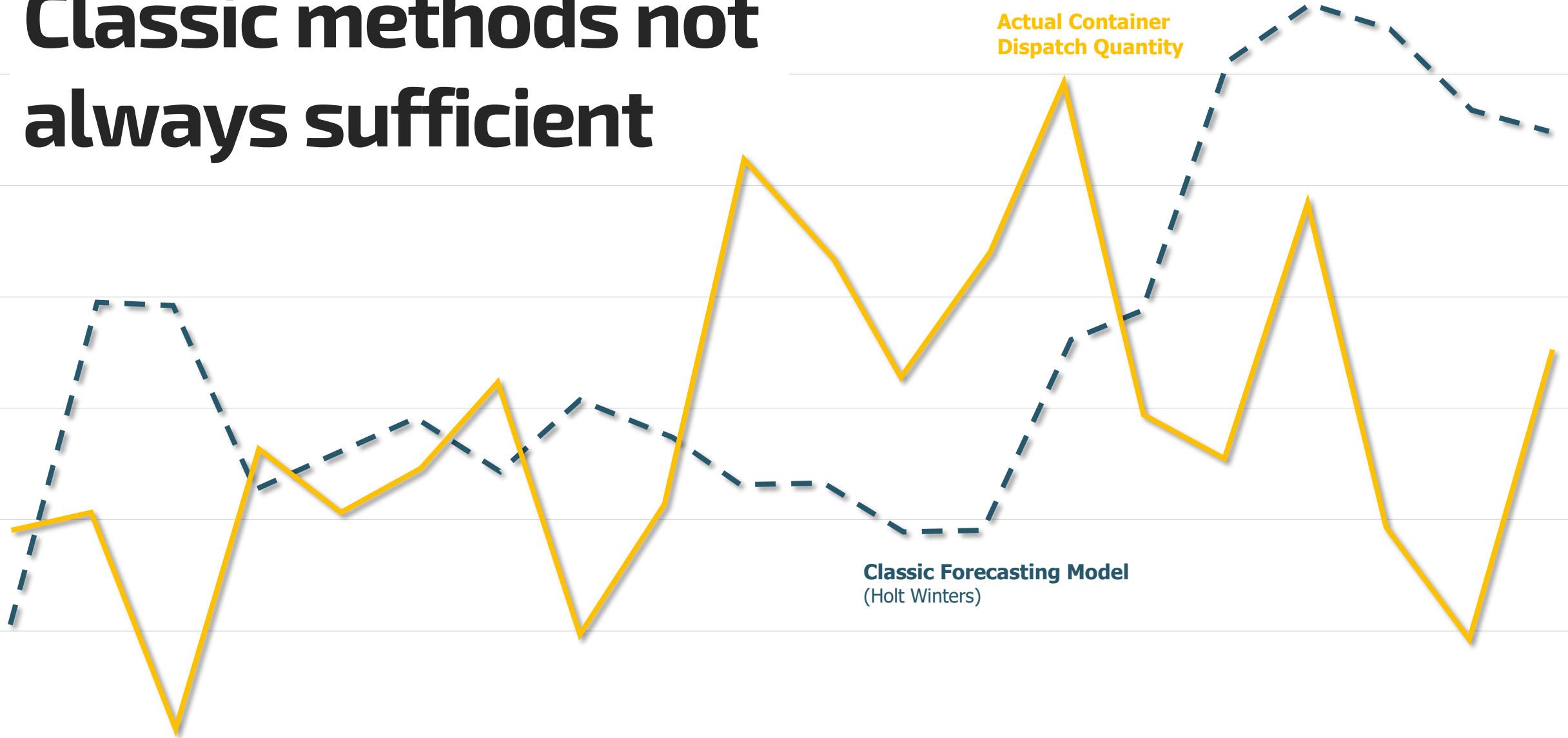
How can we make it
happen?

Prescriptive analytics

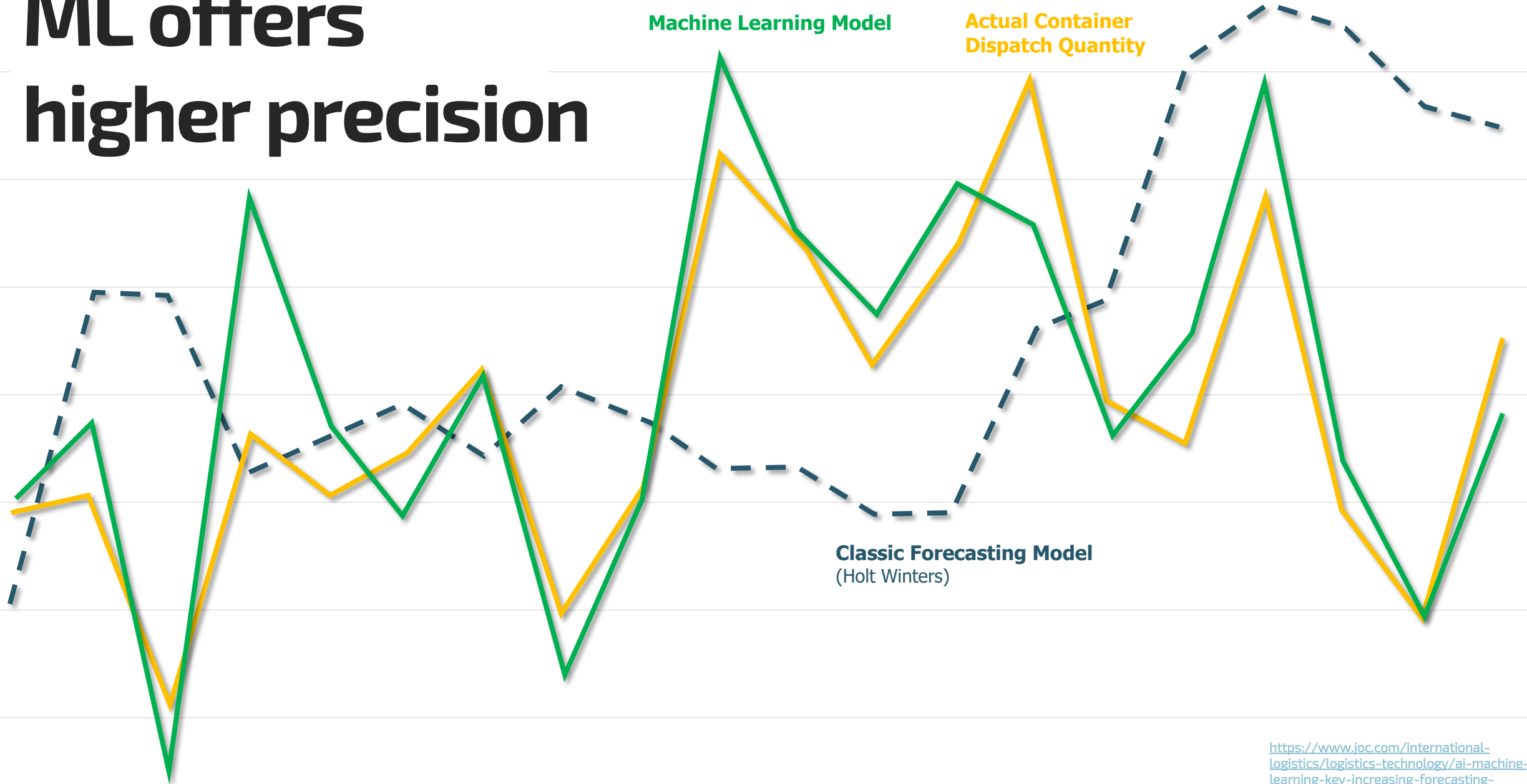
Data driven organization

Innovation
happens here!

Classic methods not always sufficient



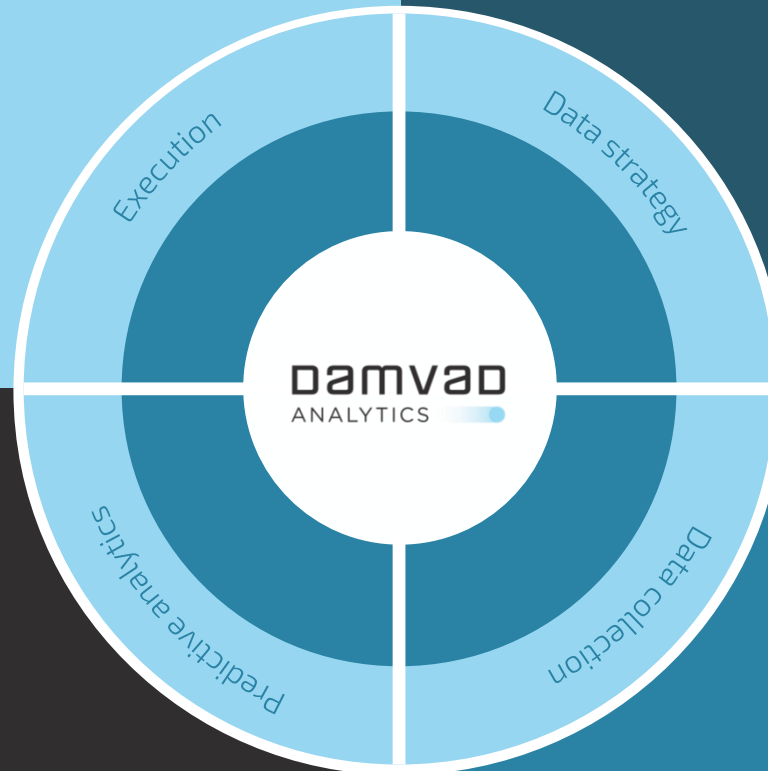
ML offers higher precision



Focus on value creation

Convert to value!

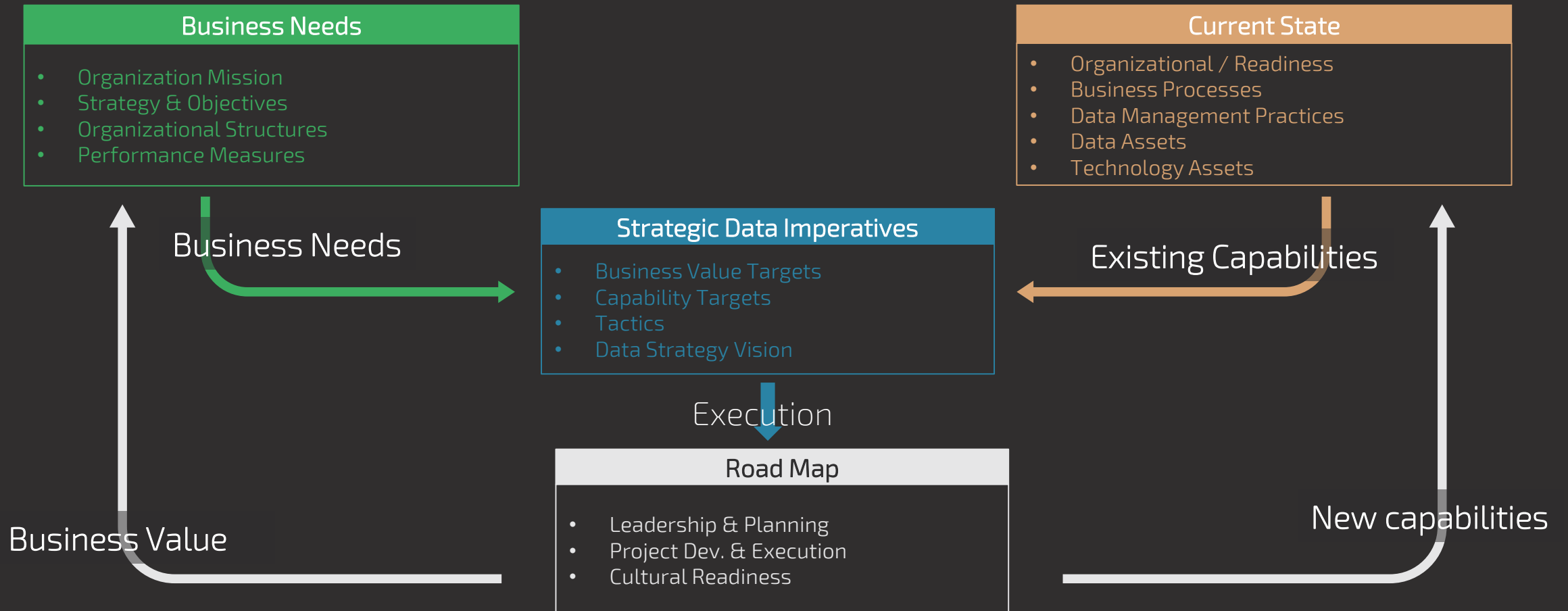
Identify and process value!



Build applicable and scalable models!

Structure and cultivate!

Let the road map reflect the strategy



Avoid large standalone projects

One big project – that failed



Jan

Feb

Mar

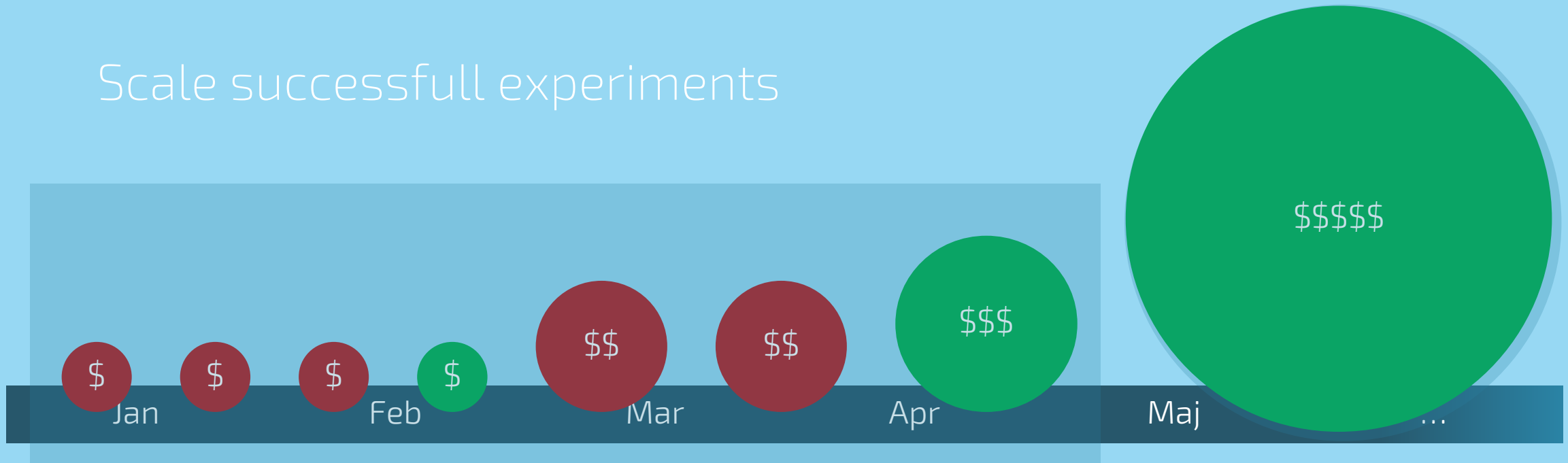
Apr

Jun

...

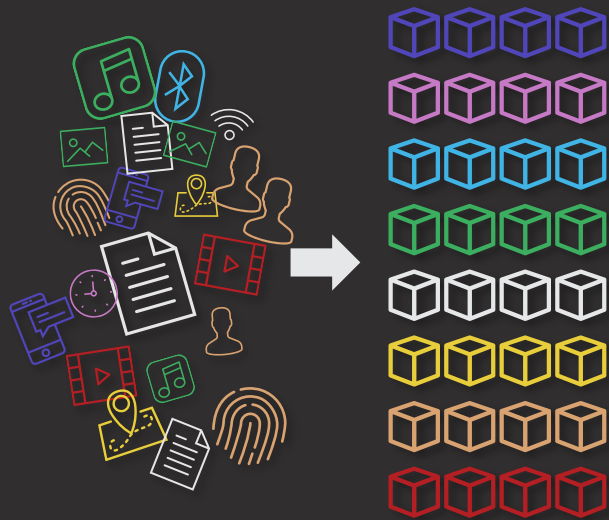
Trial and error

Scale successful experiments



Projekt proces

Data preparation



Get an overview of data

- Access to data
- Structure data
- Descriptive statistics

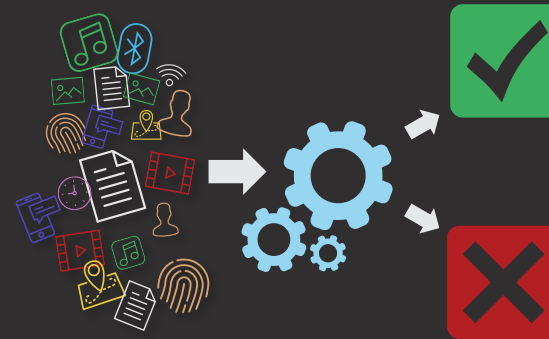
Data exploration



Analysis insights

- Cluster Analysis
- Dimension Reduction
- Feature Engineering

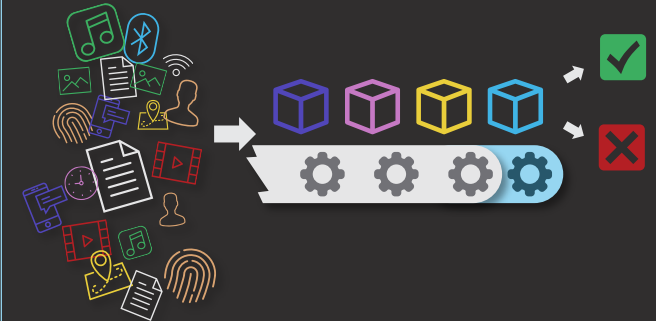
Modelling



Build predictive model

- Train ML models
- Evaluate performance
- Choose final model

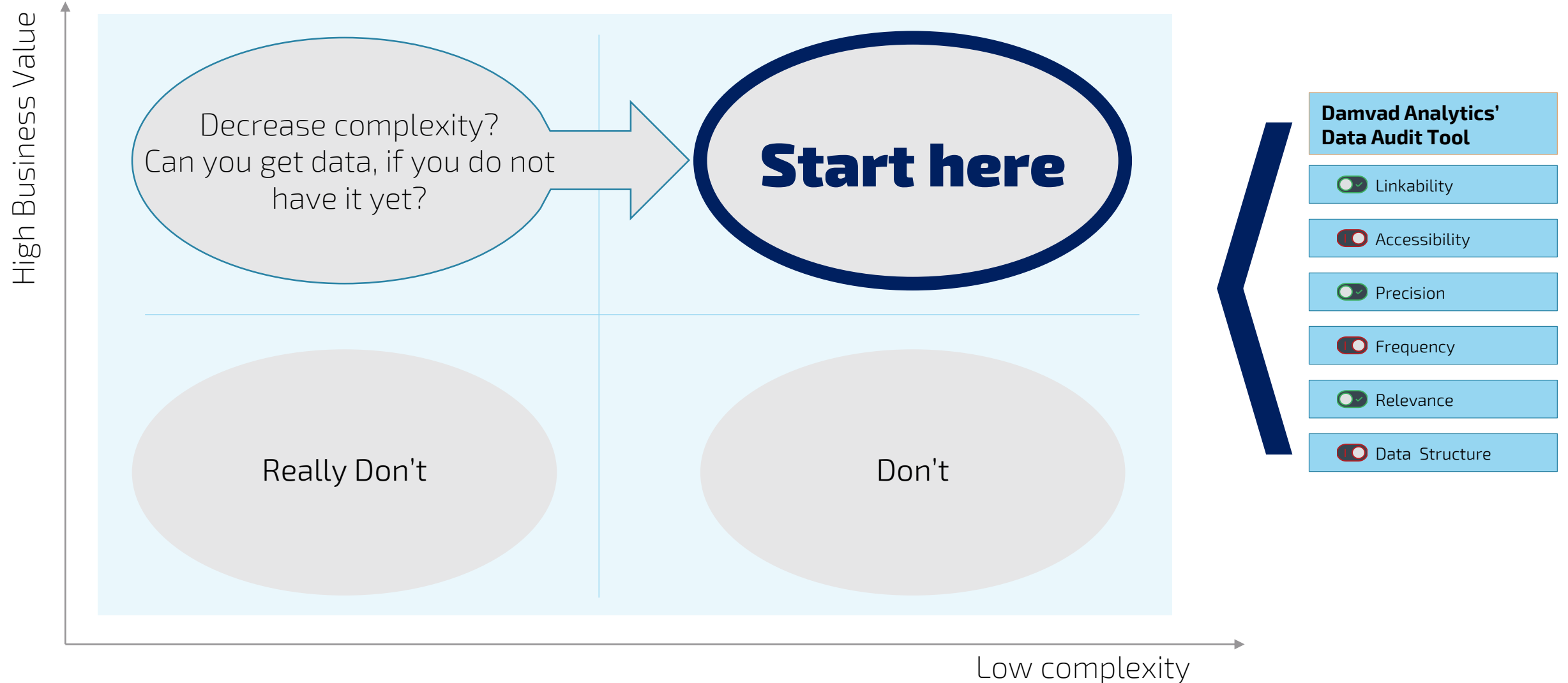
Production



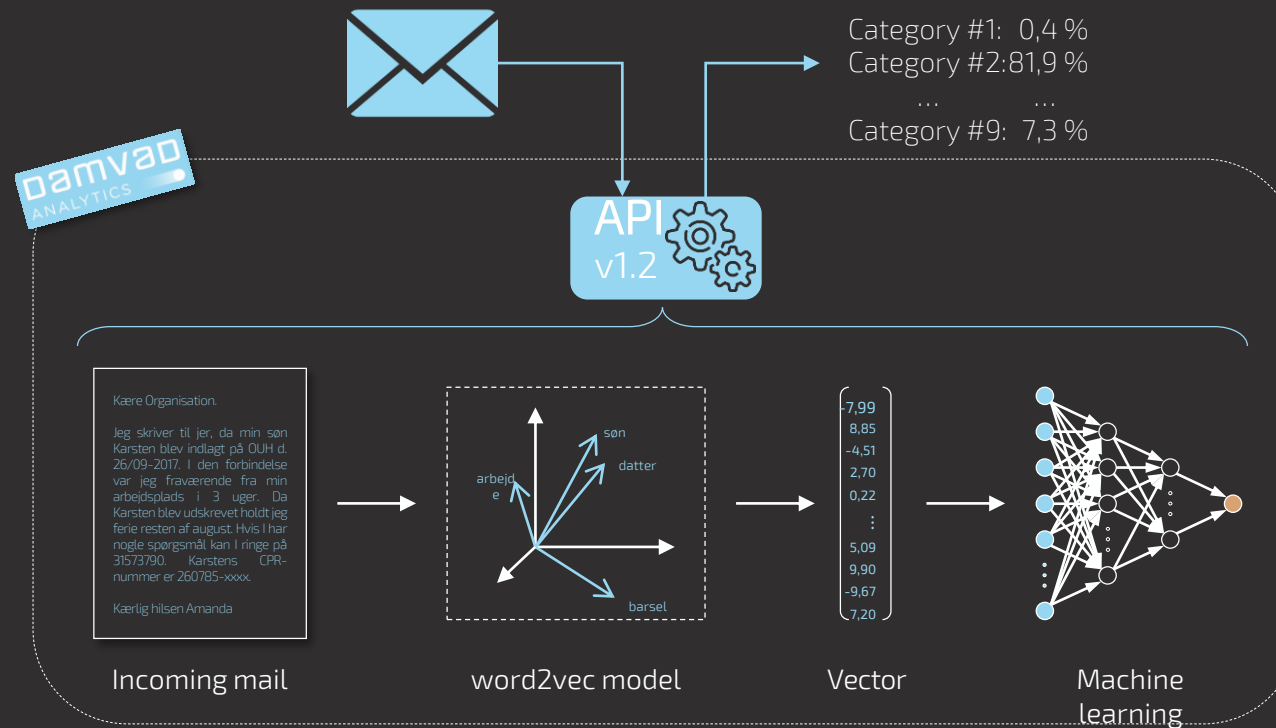
Implement model

- API
- Docker container

Focus on high-value and low-complexity data



Product wrapping in containers and API's



Open Source is very much ready for use



ULM-FiT

ELMO

BERT

?

AnonymizeAPI

Kære Peter.

Jeg hedder Ulla og skriver til dig, da min mand Gunnar og jeg ønsker at lade os skille. Vi har boet på Ulykkevej 42 i Sorgby i 24 år og vi ønsker at dele boet. Hvis du får brug for det er mit CPR-nr. 241279-2048 og du kan fange mig på enten 24759248 eller ulla@familienflid.dk.

DIT SYSTEM DIT SETUP DINE DATA

Kære #navn.

Jeg hedder #navn og skriver til dig, da min mand #navn og jeg ønsker at lade os skille. Vi har boet på #adresse i #by i 24 år og vi ønsker at dele boet. Hvis du får brug for det er mit CPR-nr. #cprnr og du kan fange mig på enten #telefonnummer eller #email.

AnonymizeAPI

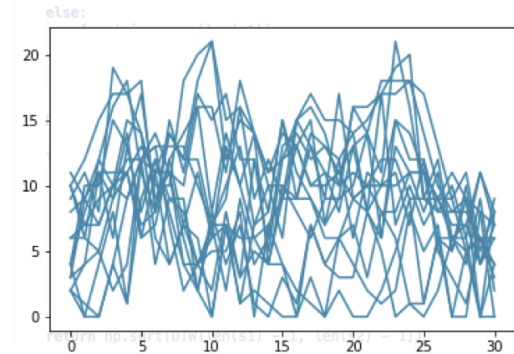


Case: Cluster Analysis

```
def DTWDistance(self, s1, s2, w=None):
    """
    Calculates dynamic time warping Euclidean distance between two
    sequences. Option to enforce locality constraint for window w.
    """
    DTW = {}

    if w:
        w = max(w, abs(len(s1) - len(s2)))

        for i in range(-1, len(s1)):
            for j in range(-1, len(s2)):
                DTW[(i, j)] = float('inf')
```

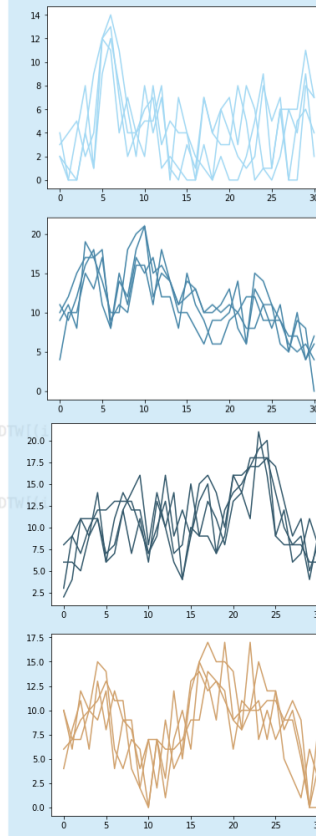


```
def LB_Keogh(self, s1, s2, r):
    """
    Calculates LB_Keogh lower bound to dynamic time warping. Linear
    complexity compared to quadratic complexity of dtw.
    """
    LB_sum = 0
    for ind, i in enumerate(s1):

        lower_bound = min(s2[(ind - r if ind - r >= 0 else 0):(ind + r)])
        upper_bound = max(s2[(ind - r if ind - r >= 0 else 0):(ind + r)])

        if i > upper_bound:
            LB_sum = LB_sum + (i - upper_bound) ** 2
        elif i < lower_bound:
            LB_sum = LB_sum + (i - lower_bound) ** 2
```

DTW clustering

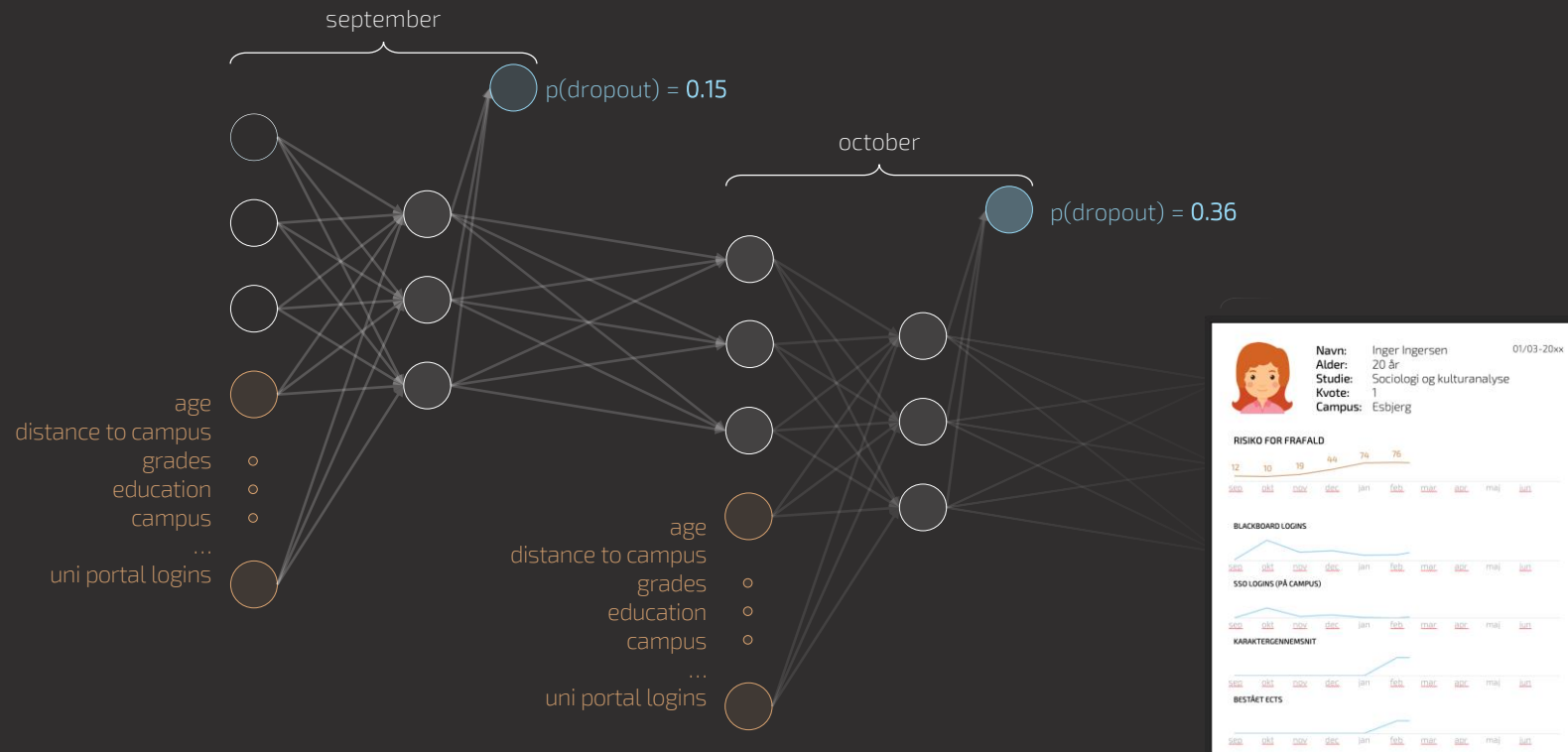


```
def k_means_clust(self, data, num_iter, w, progress=False):
    """
    k-means clustering algorithm for time series data. dynamic time warping Euclidean
    distance used as default similarity measure.
    """
    self.centroids = random.sample(data, self.num_clust)

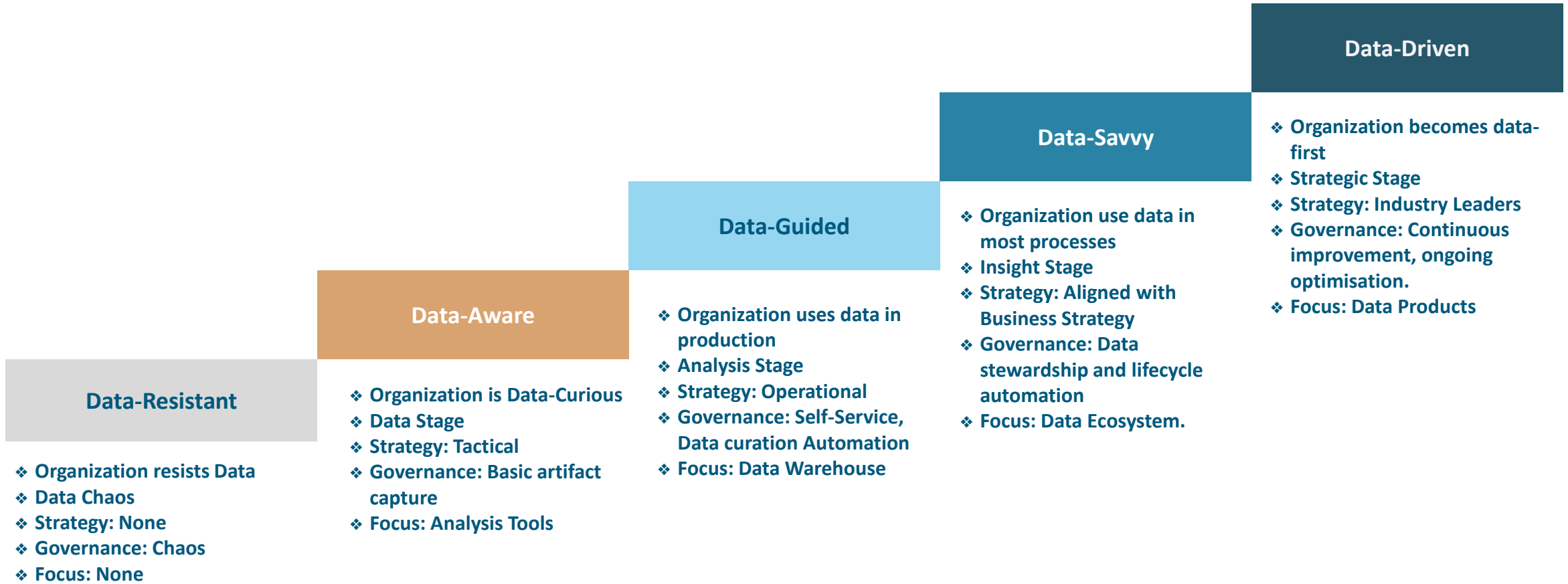
    for n in range(num_iter):
        if progress:
            print('iteration ' + str(n + 1))
        # assign data points to clusters
        self.assignments = {}
        for ind, i in enumerate(data):
            min_dist = float('inf')
            closest_centroid = None
            for k in range(self.num_clust):
                dist = DTWDistance(self, i, self.centroids[k], w)
                if dist < min_dist:
                    min_dist = dist
                    closest_centroid = k
            self.assignments[ind] = closest_centroid

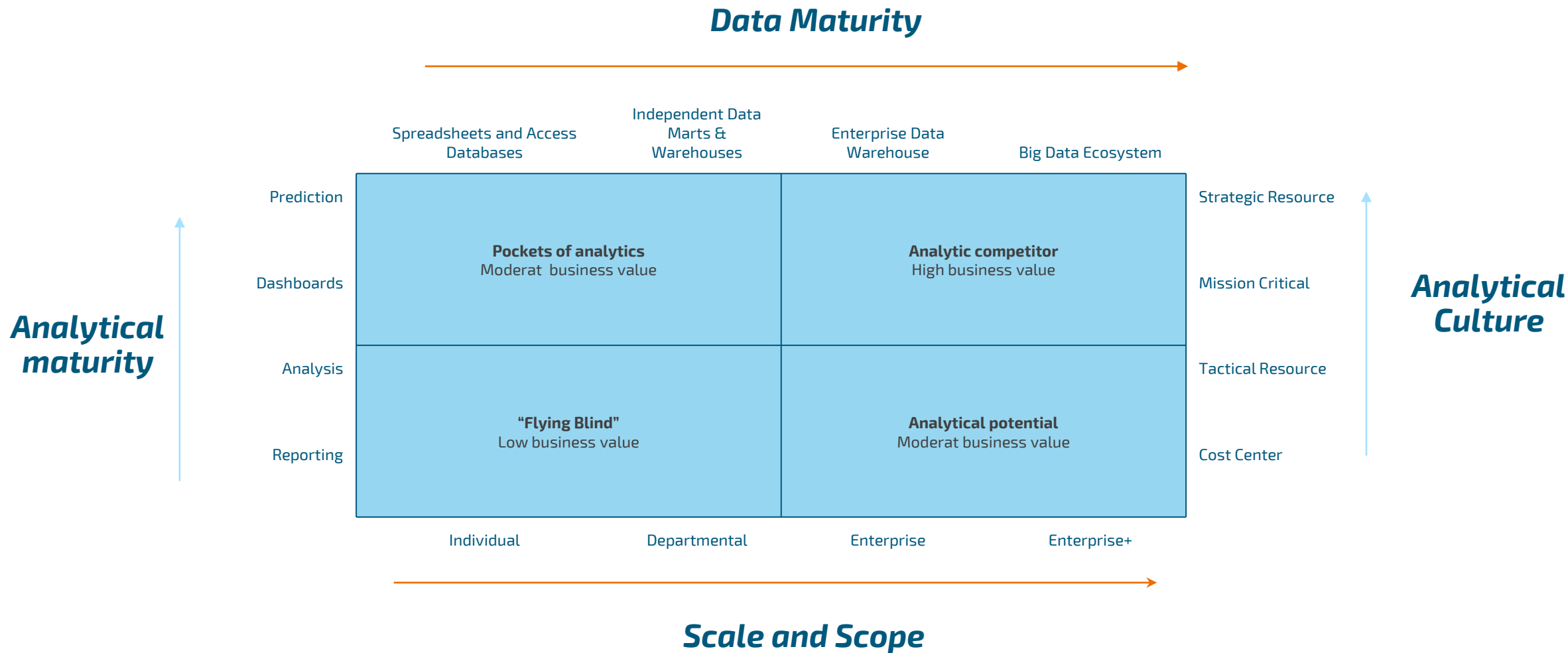
        # recalculate centroids
        for key in self.assignments:
            for k in range(self.num_clust):
                self.centroids[k] = [sum(len(s1) * assignments[k] for s1 in clust
```

Case: Churn Prediction



Is your organization data driven?





DAMVAD Analytics
Havnegade 39,
DK-1058 Copenhagen K

