2018 Predictive Analytics Symposium

Session 06: Customer Segmentation and Profitability Analysis

SOA Antitrust Compliance Guidelines
SOA Presentation Disclaimer

Customer segmentation and profitability analysis

Moderator: Michael Niemerg

Presenter: Talex Diede

SOCIETY OF ACTUARIES®

SOA Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- -**Do not** discuss prices for services or products or anything else that might affect prices
- -**Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- -**Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- -**Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- -**Do** alert SOA staff and/or legal counsel to any concerning discussions
- -**Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

Who is Talex?

- MS in Computational Finance & Risk Management - University of Washington
- Been with Milliman for 5 years, focused on data science and predictive analytics
  - Seattle Life Practice
  - Primarily modeling policyholder behavior for annuities
- Favorite programming language: R


DO WE HAVE ANY ACTIONABLE ANALYTICS FROM OUR BIG DATA IN THE CLOUD?
YES, THE DATA SHOWS THAT MY PRODUCTIVITY PLUNGES WHENEVER YOU LEARN NEW JARGON.
MAYBE IN-MEMORY COMPUTING WILL ACCEL-ERATE YOUR APPLICA-TIONS.
PLUNGE, PLUNGE, PLUNGE.
Dilbert.com DilbertCartoonist@gmail.com
1-9-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick

Agenda

- Motivation
- Business applications: Profitability analysis
- Segmentation: the technical side
- Questions

Why does this matter?

SOCIETY OF ACTUARIES®

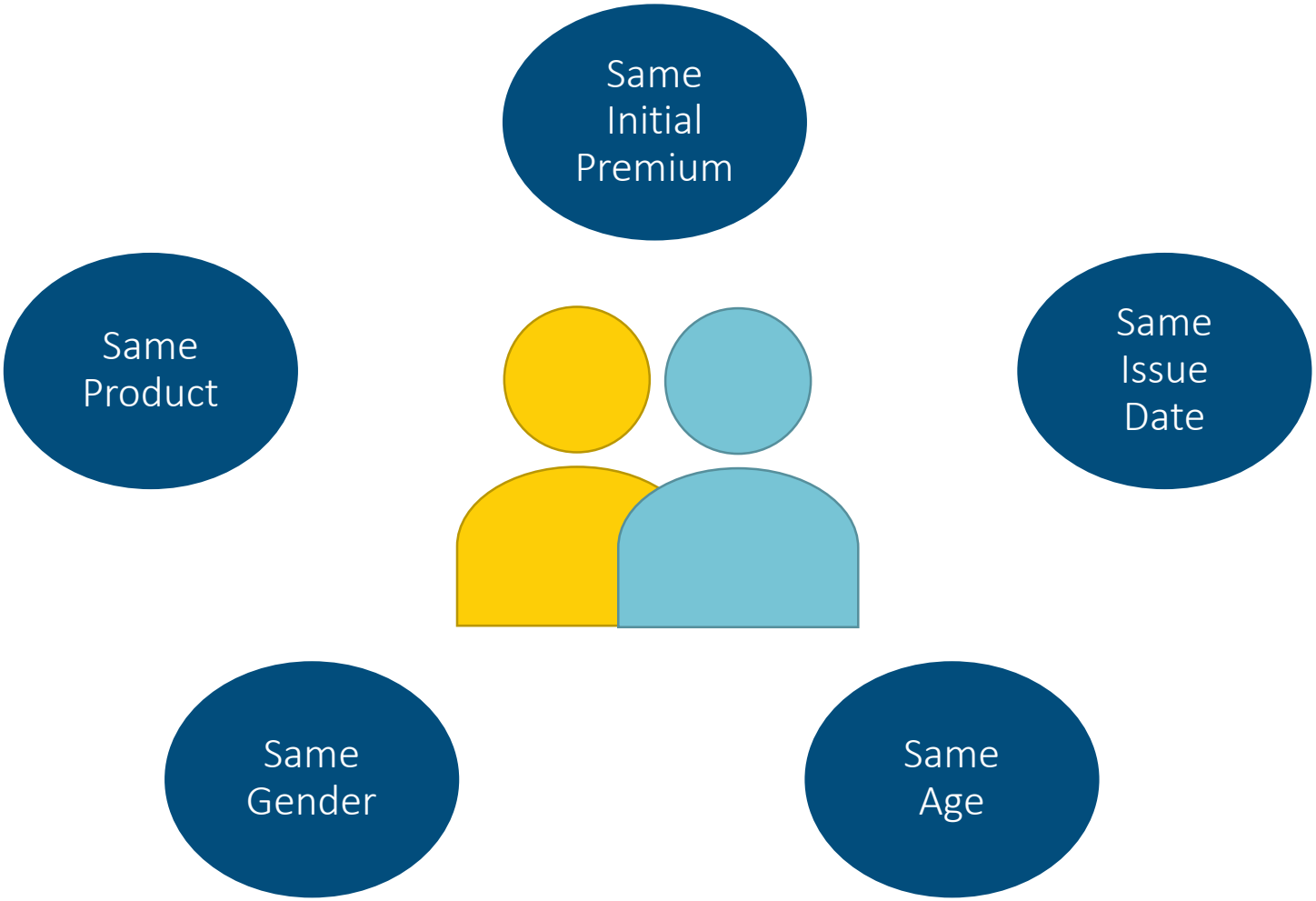
Customer segmentation

- Understand the needs of the customer
- Build tailored products
- Efficiently market existing products

Profitability analysis

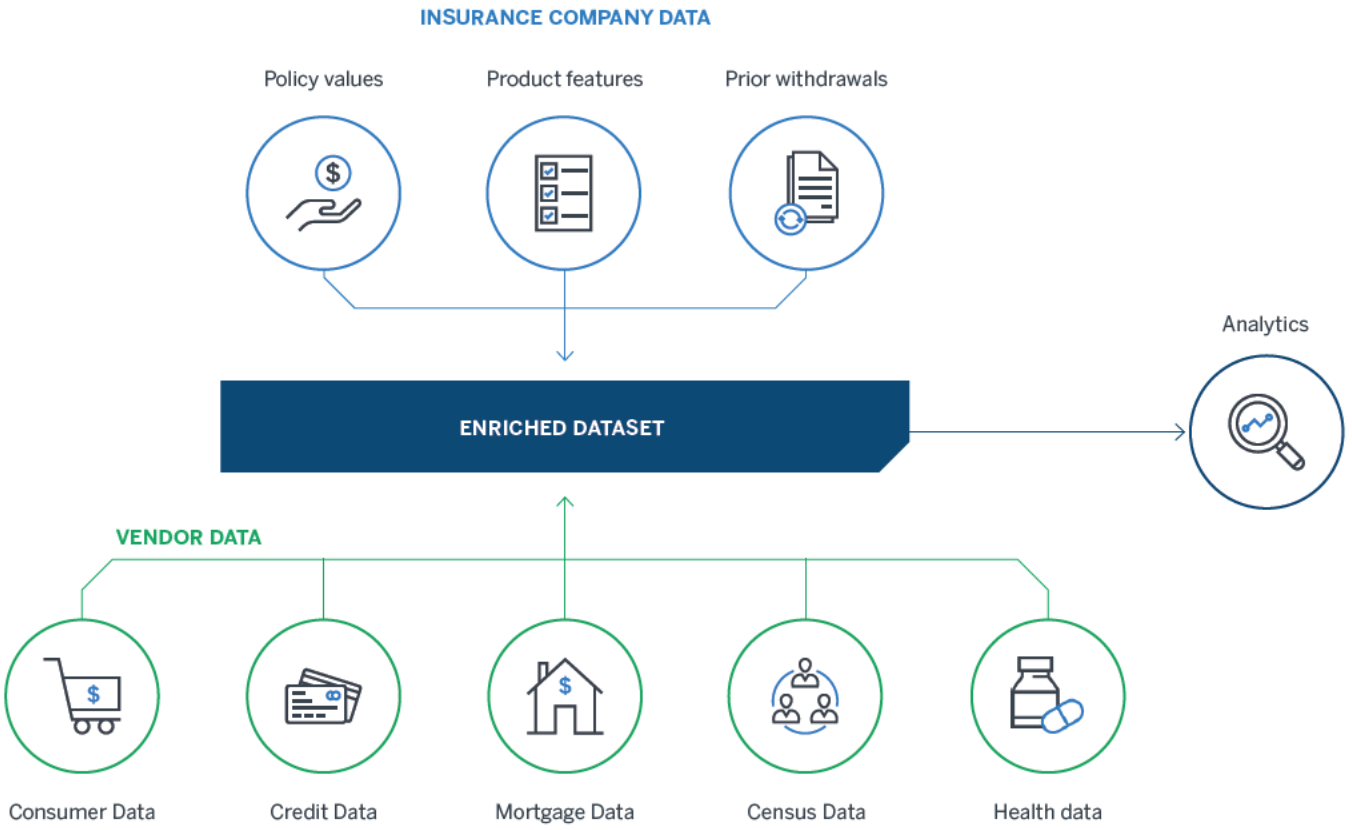
SOCIETY OF ACTUARIES®

Current paradigm

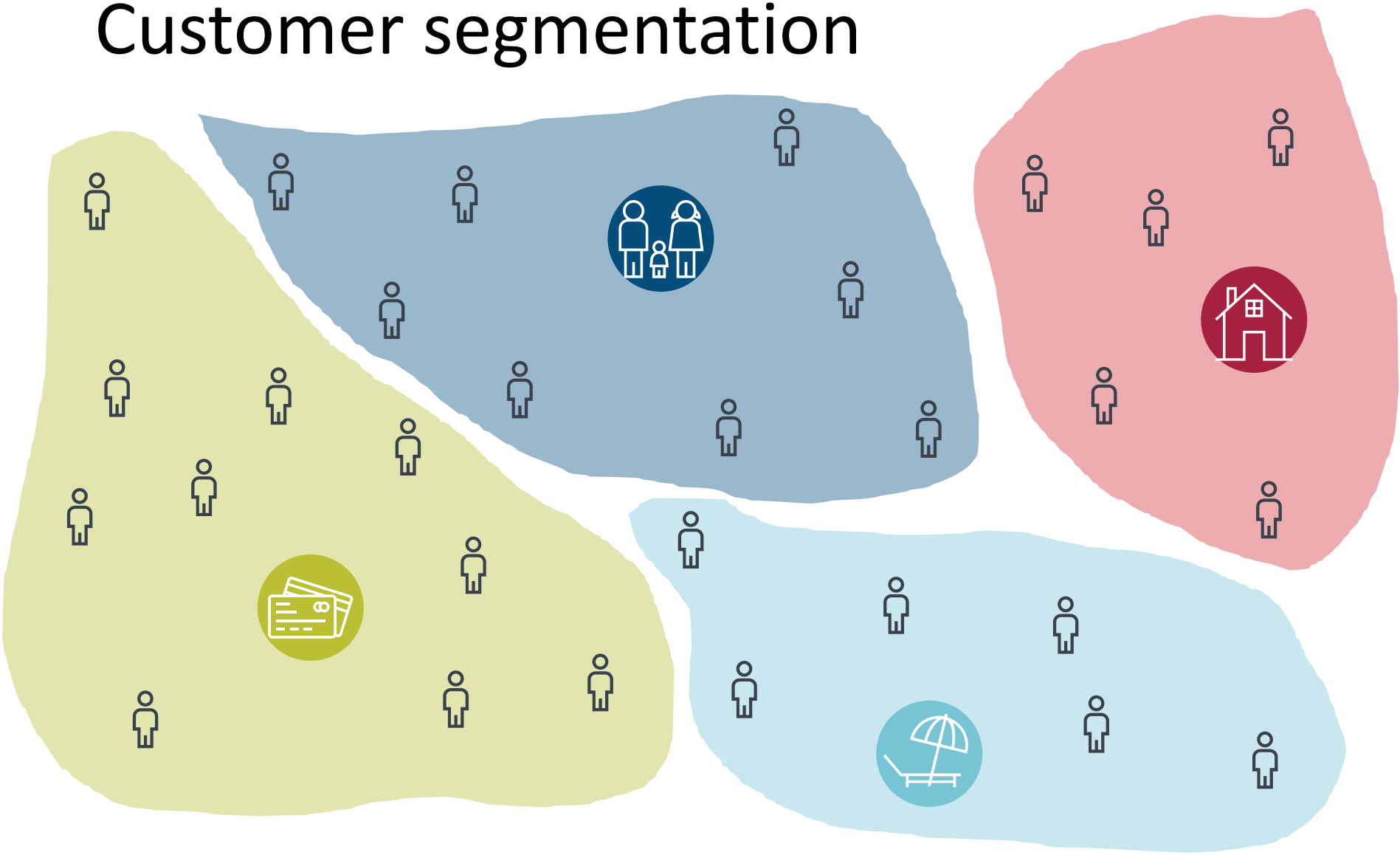


Understanding your customers

Data enrichment

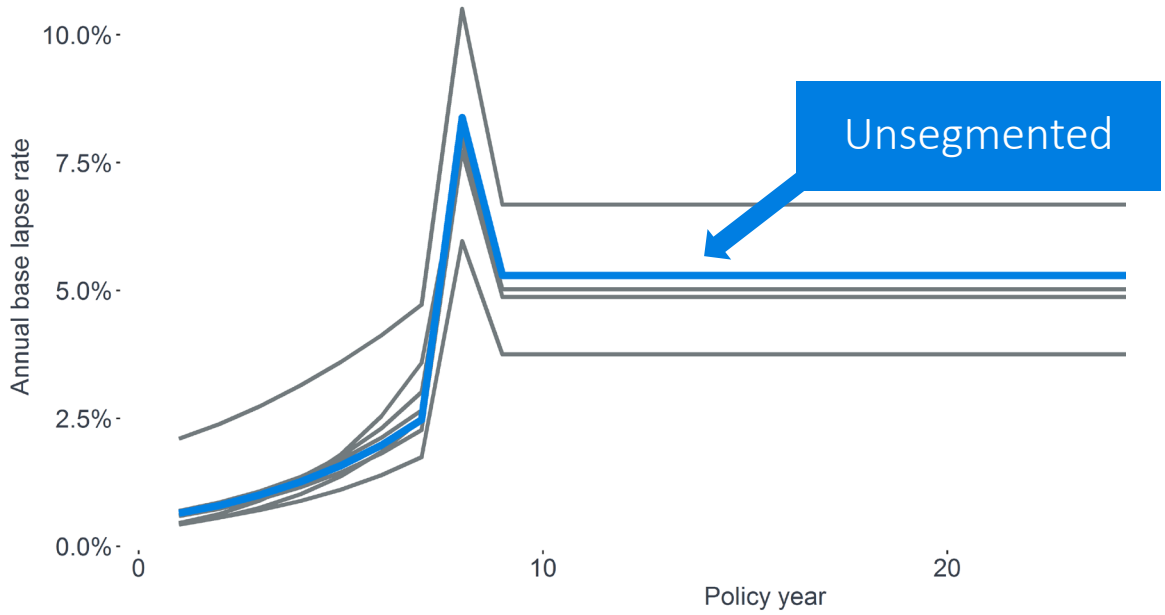


Customer segmentation



Segmented assumptions

- Fit behavior models to each customer segment revealing how people use their insurance differently



Project profitability

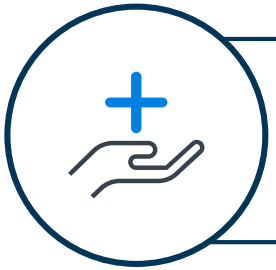
- Calculate “profitability” measure at seriatim level

Customer Profile

Cash Flow Projection Model

Customer Level Profitability (CLP)

Join to other available information

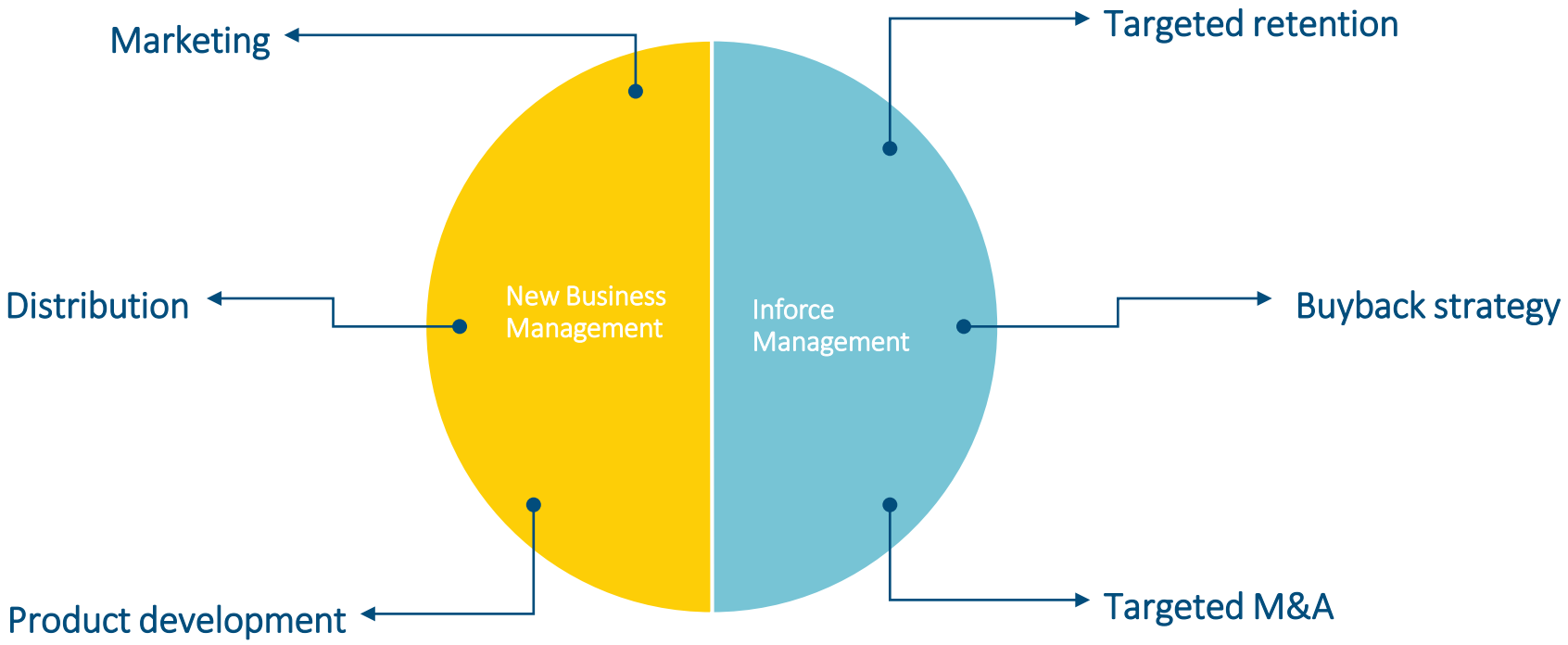


Propensity to buy

Cost of marketing

Financial advisor

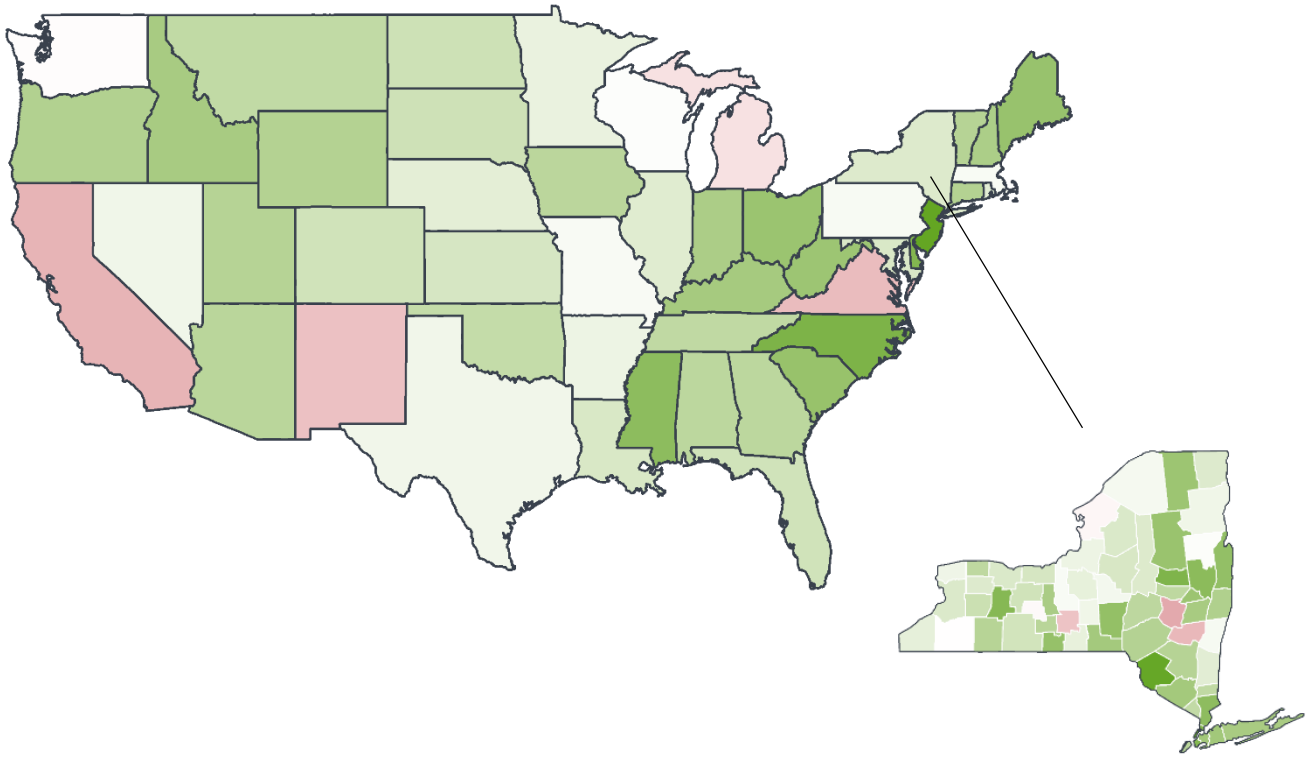
Business applications



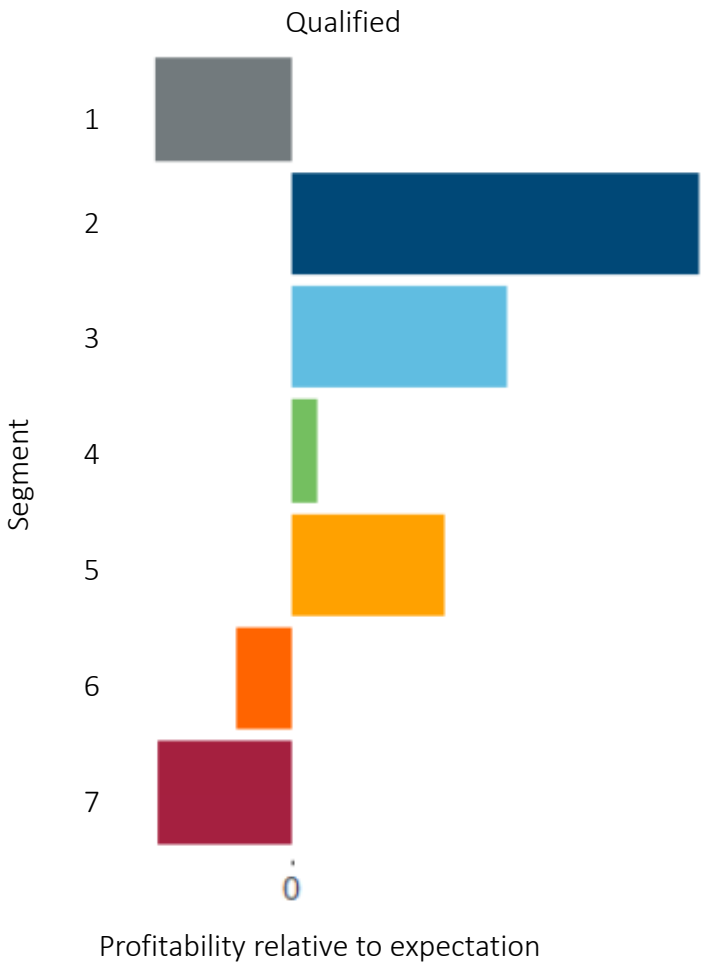
Marketing

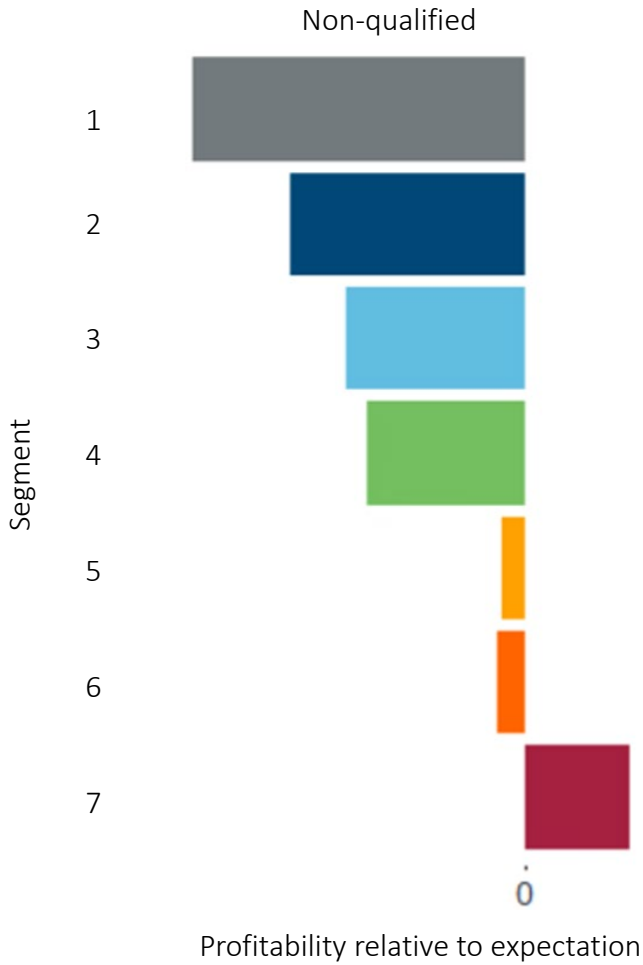
Higher Profitability

Lower Profitability

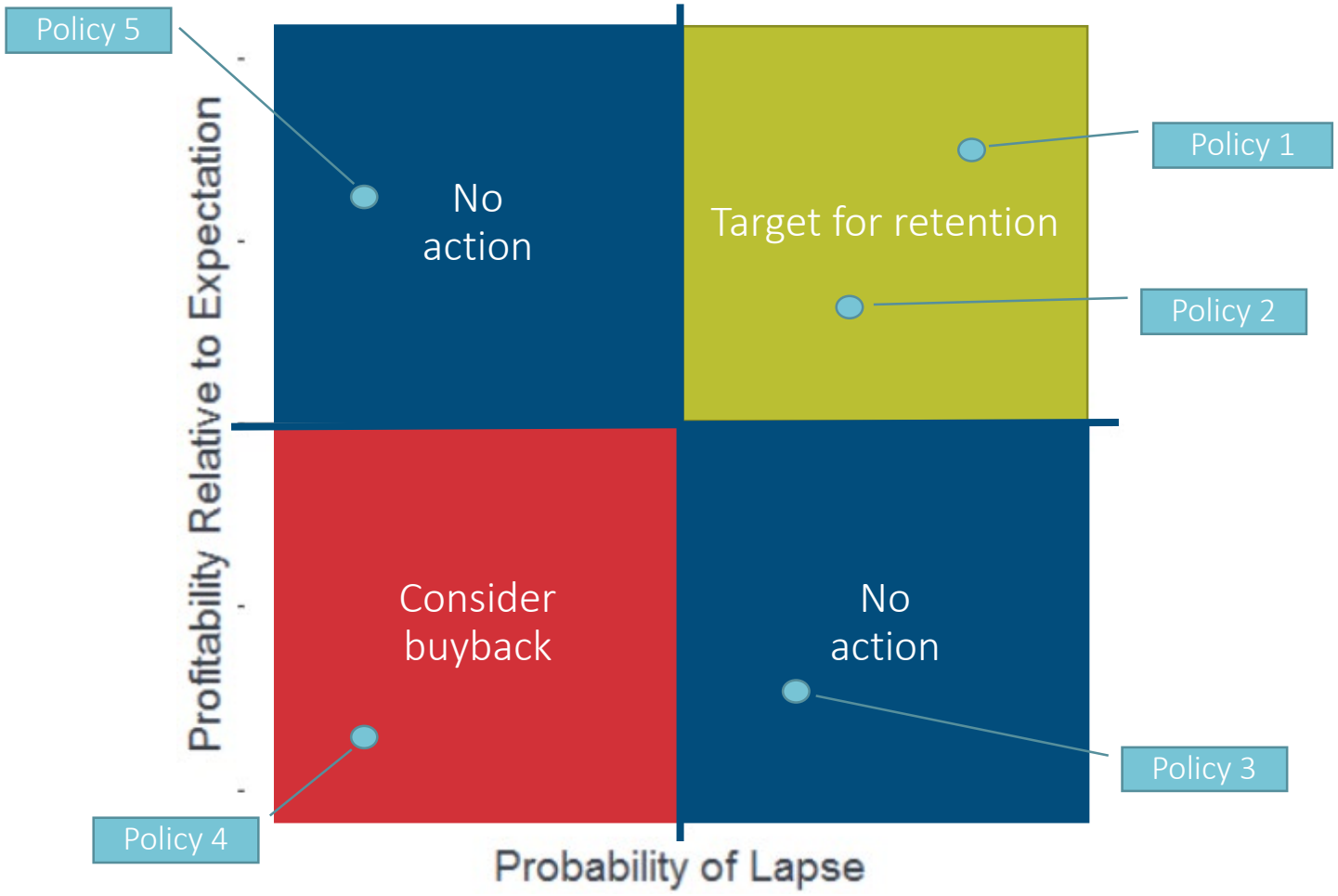


Product development

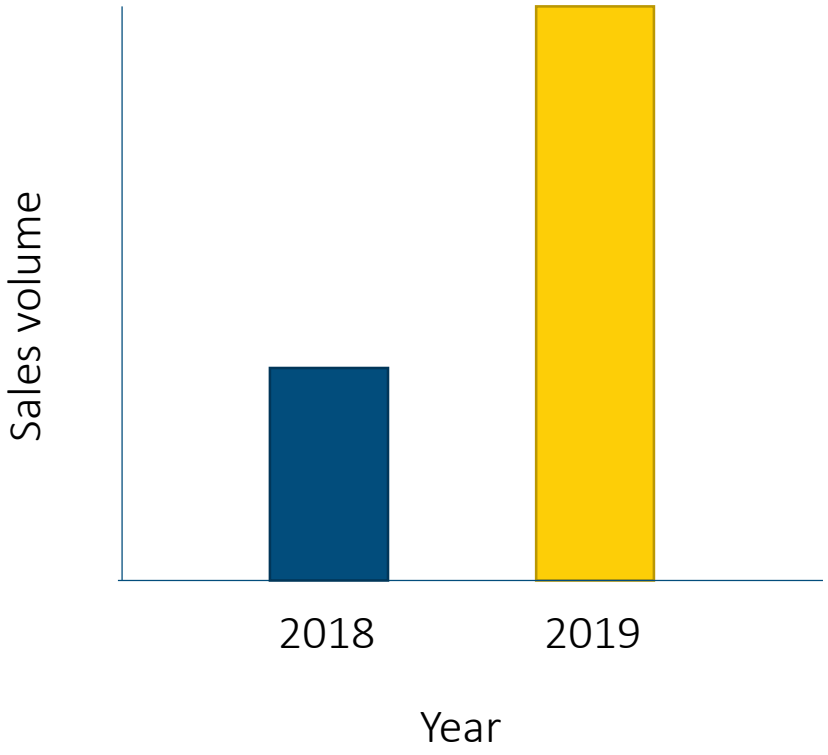




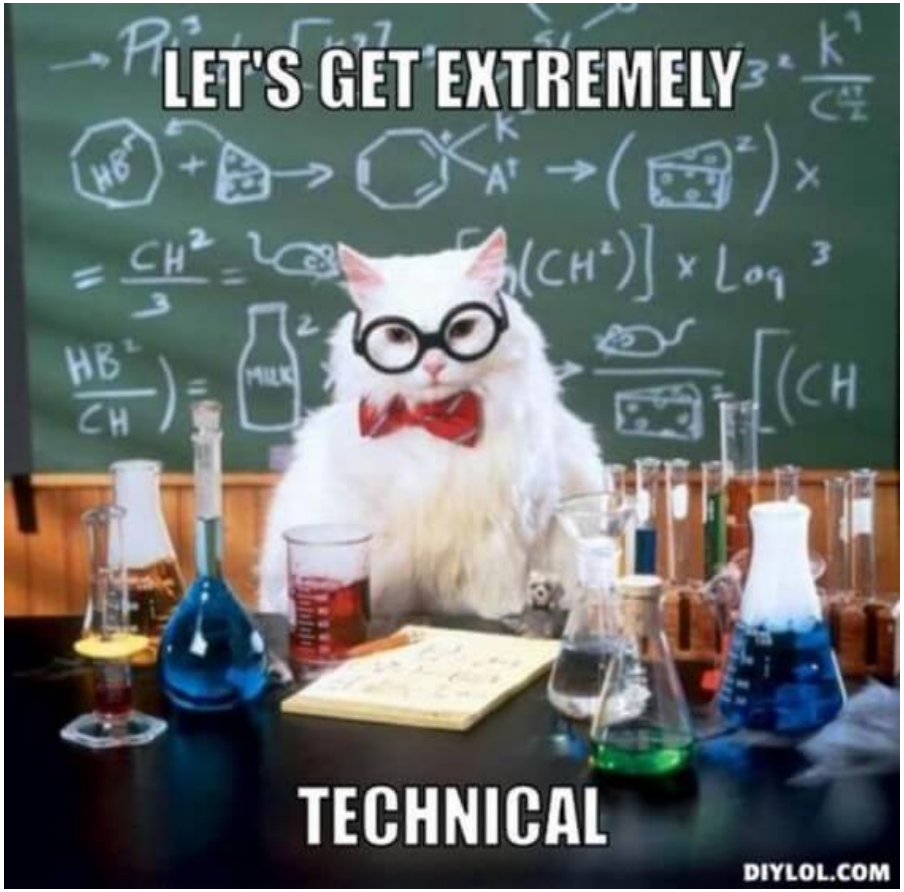
Targeted retention



Track impact of actions taken



But wait, how do I determine customer segments?



Questions before we move on?

SOCIETY OF ACTUARIES

Segmentation:
A technical look

SOCIETY OF ACTUARIES®

Types of clustering

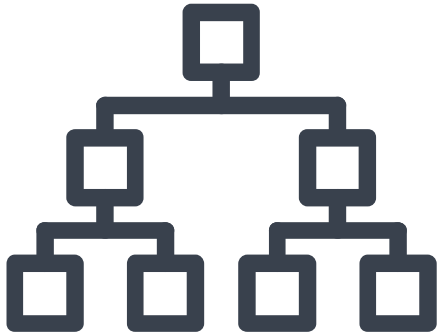
- Connectivity-based clustering
- Centroid-based clustering
- Distribution-based clustering
- Density-based clustering

Types of clustering

- Connectivity-based clustering
- Centroid-based clustering
- Distribution-based clustering
- Density-based clustering

Connectivity-based clustering

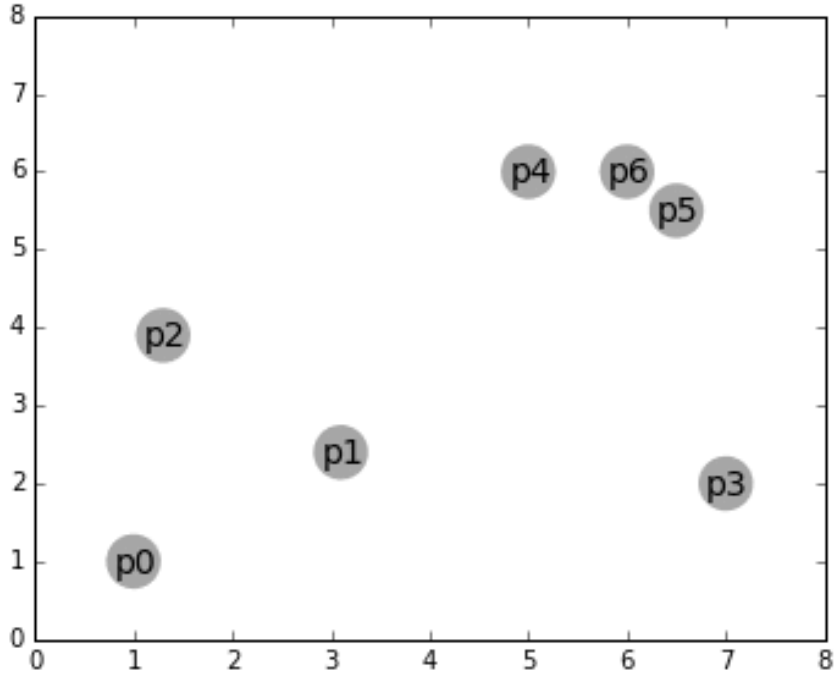
- **Defining principal:** data points are more related to nearby data points than to data points far away
- **Algorithm:** Hierarchical clustering
  - Agglomerative or Divisive
  - *https://www.soa.org/Library/Newsletters/Predictive-Analytics-and-Futurism/2018/april/2018-predictive-analytics-iss17.pdf*
- **Pros:**
  - Easy to interpret
  - Can choose # of clusters after
- **Cons:**
  - Not scalable

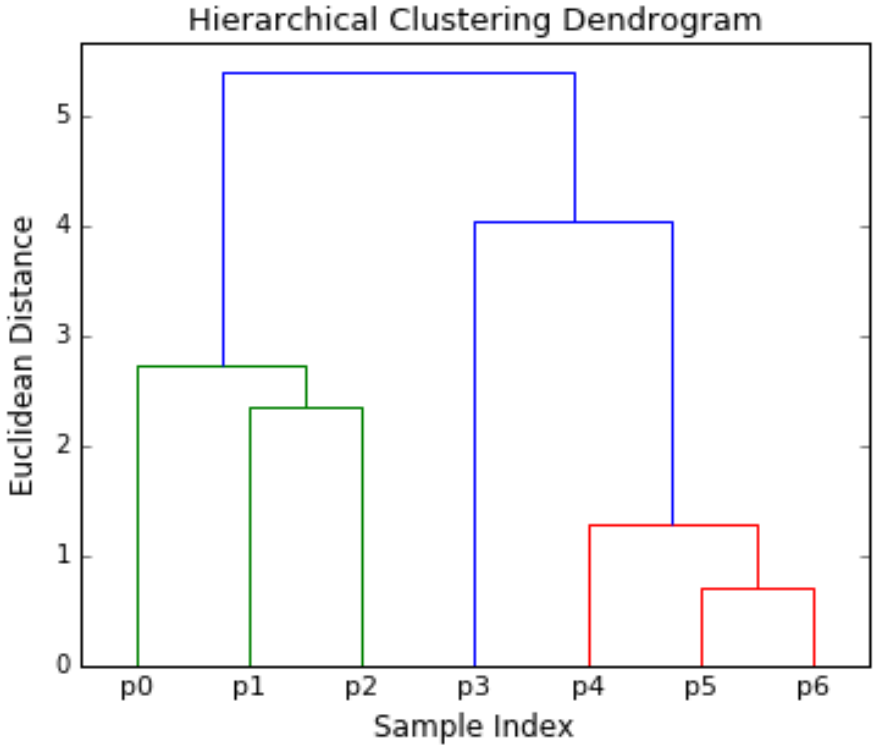


Process

- Standardize variables
- Treat each data point as its own cluster
- Select distance metric
- Combine two closest clusters into one
- Select and cut to number of clusters

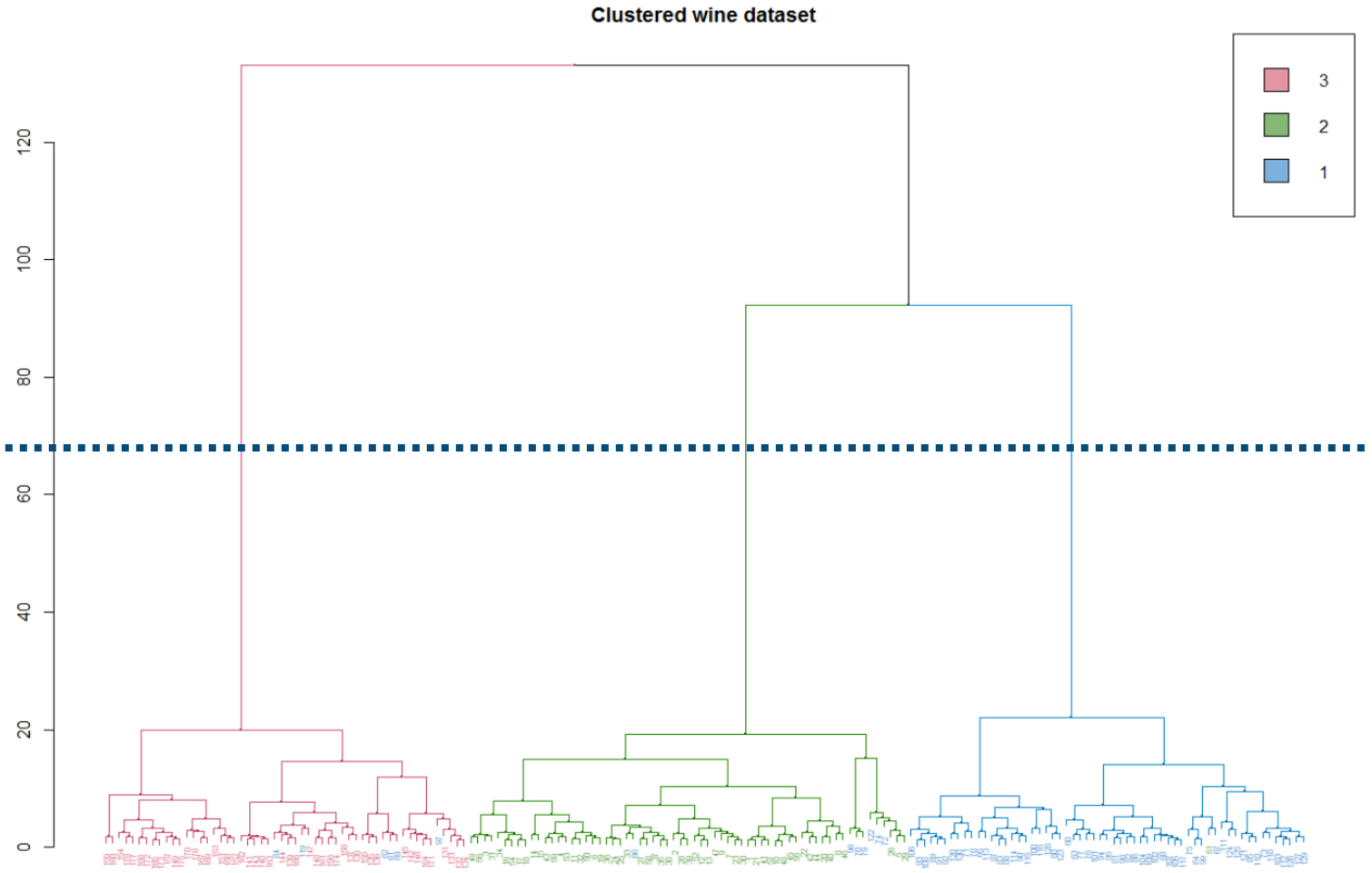
Process





*https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

Dendrogram: *tree diagram*

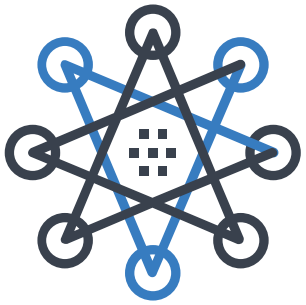


Types of clustering

- Connectivity-based clustering
- Centroid-based clustering
- Distribution-based clustering
- Density-based clustering

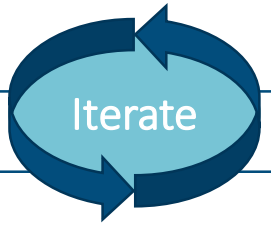
Centroid-based clustering

- **Defining principal:** data points are defined by their closeness to the centroid of the clusters
- **Algorithm:** K-means, K-medians
- **Pros:**
  - Computationally efficient
- **Cons:**
  - Spherical
  - Must choose # of clusters in advance

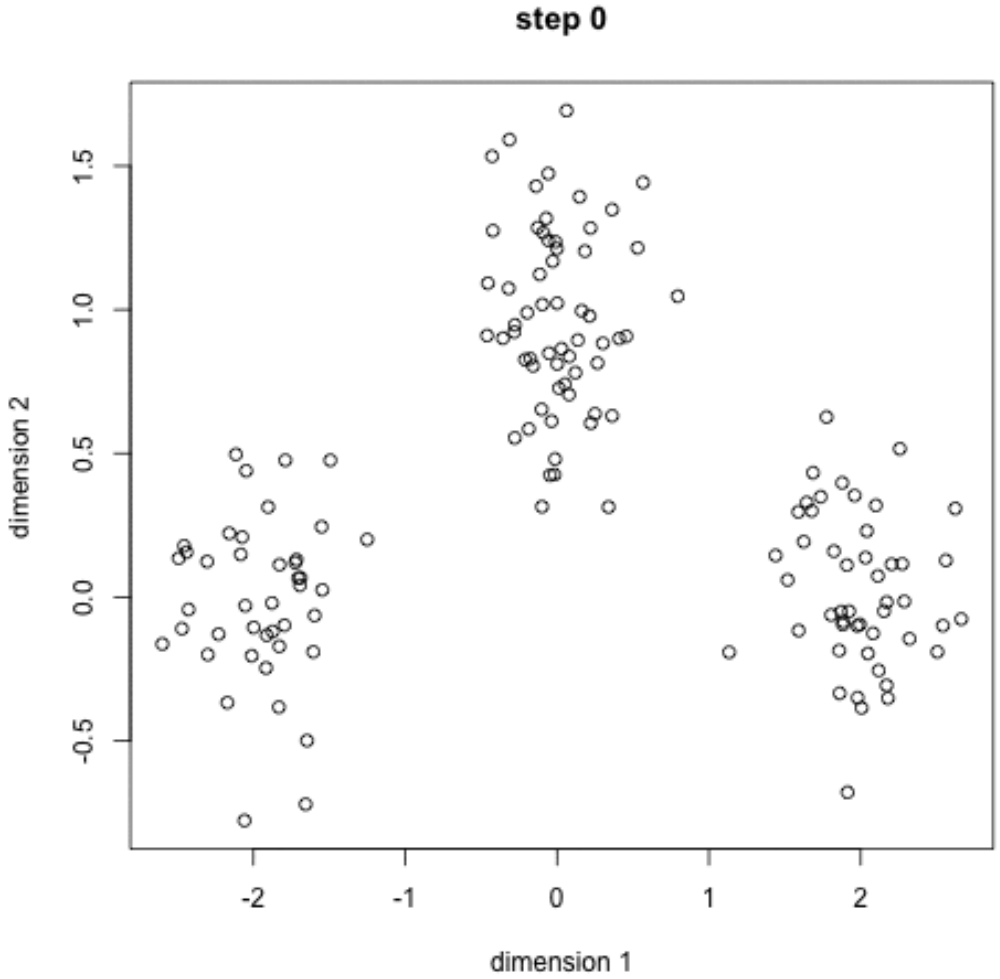


Process

- Select number of clusters (k)
- Randomly select k points (starting cluster centroids)
- Classify every point relative to closest centroid
- Re-compute cluster centroids
- Repeat for other values of k

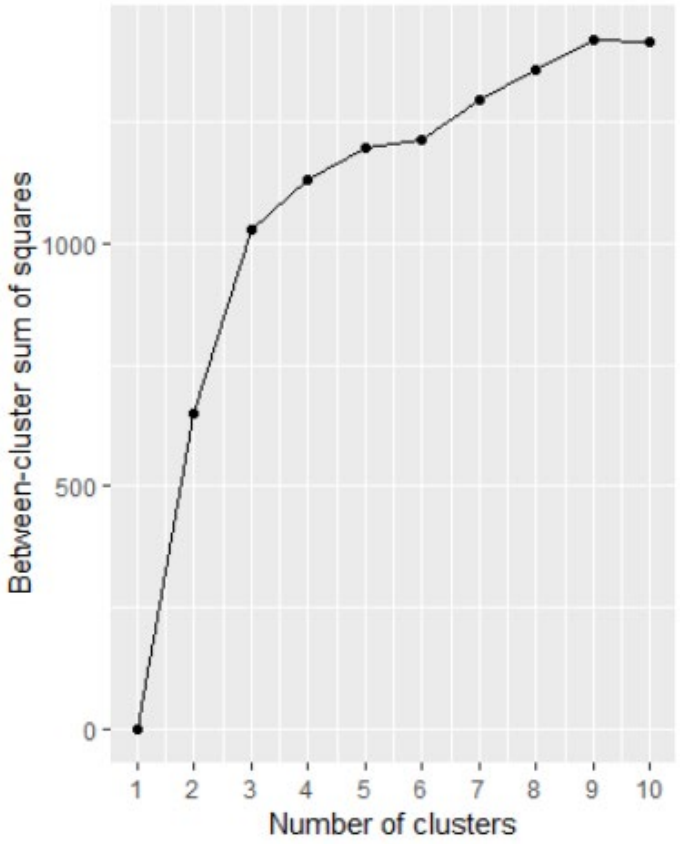


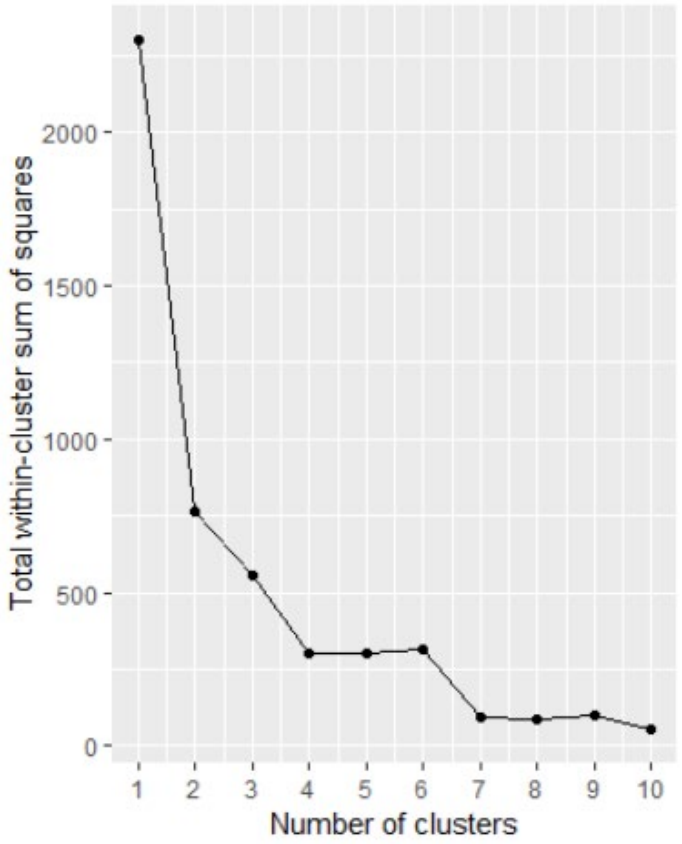
Process



*https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

Choosing k


Between-cluster sum of squares
0
500
1000
1
2
3
4
5
6
7
8
9
10
Number of clusters


Total within-cluster sum of squares
0
500
1000
1500
2000
1
2
3
4
5
6
7
8
9
10
Number of clusters

Types of clustering

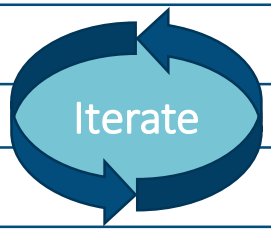
- Connectivity-based clustering
- Centroid-based clustering
- Distribution-based clustering
- Density-based clustering

Distribution-based clustering

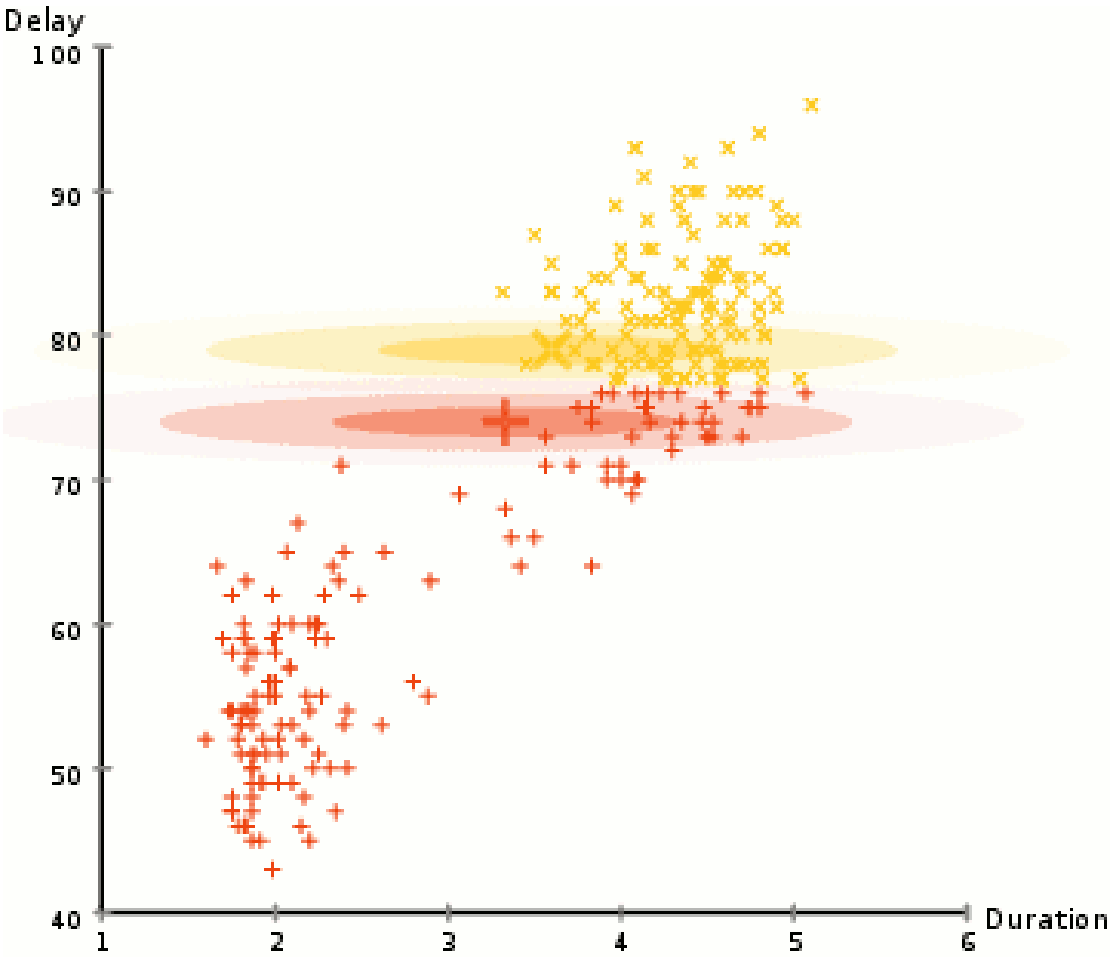
- **Defining principal:** clusters are defined as objects belonging to the same distribution
- **Algorithm:** Expectation-maximization
  - Gaussian mixture models
- **Pros:**
  - Can capture correlation and dependence between attributes
  - Can have multiple clusters per data point (mixed membership)
- **Cons:**
  - Suffer from overfitting
  - Must choose # of clusters in advance

Process

- Select number of clusters (k)
- Randomly initialize distribution parameters for each cluster
- Compute probability each data point belongs to particular cluster
- Compute new parameters to maximize probabilities



Process



*https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

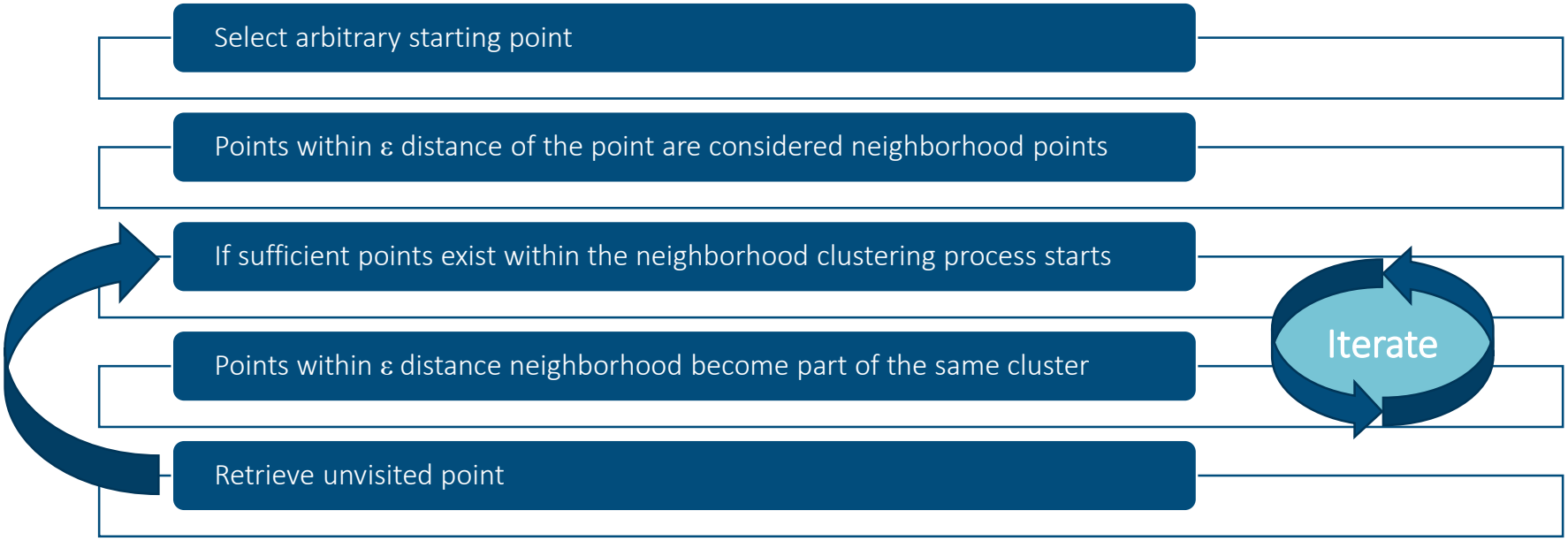
Types of clustering

- Connectivity-based clustering
- Centroid-based clustering
- Distribution-based clustering
- Density-based clustering

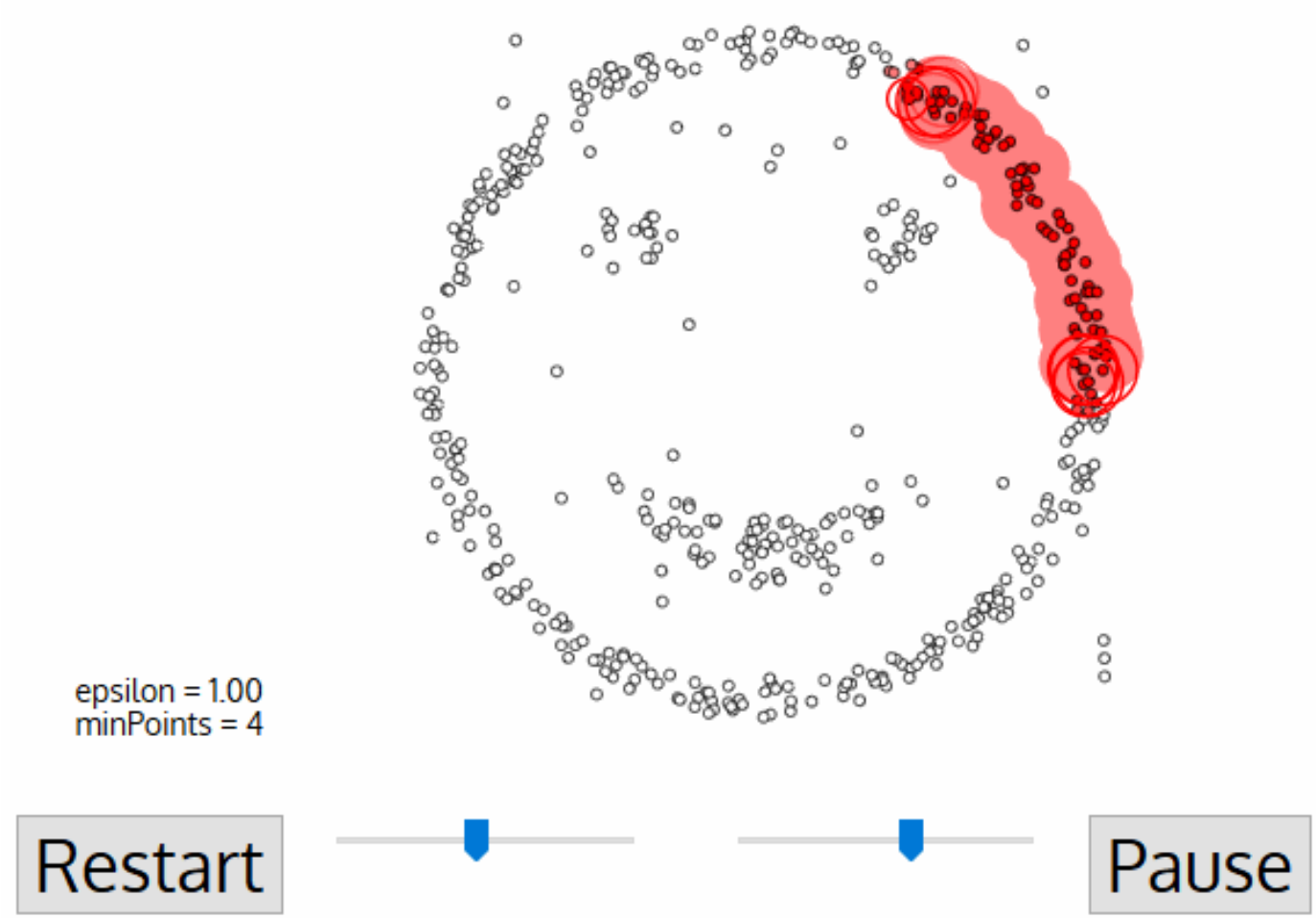
Density-based clustering

- **Defining principal:** clusters are defined as areas of higher density than the remainder of the data set
- **Algorithm:** Density-based spatial clustering of applications with noise (DBSCAN)
- **Pros:**
  - No pre-set number of clusters needed
  - Arbitrarily sized and shaped clusters
- **Cons:**
  - Struggles with clusters of varying density

Process

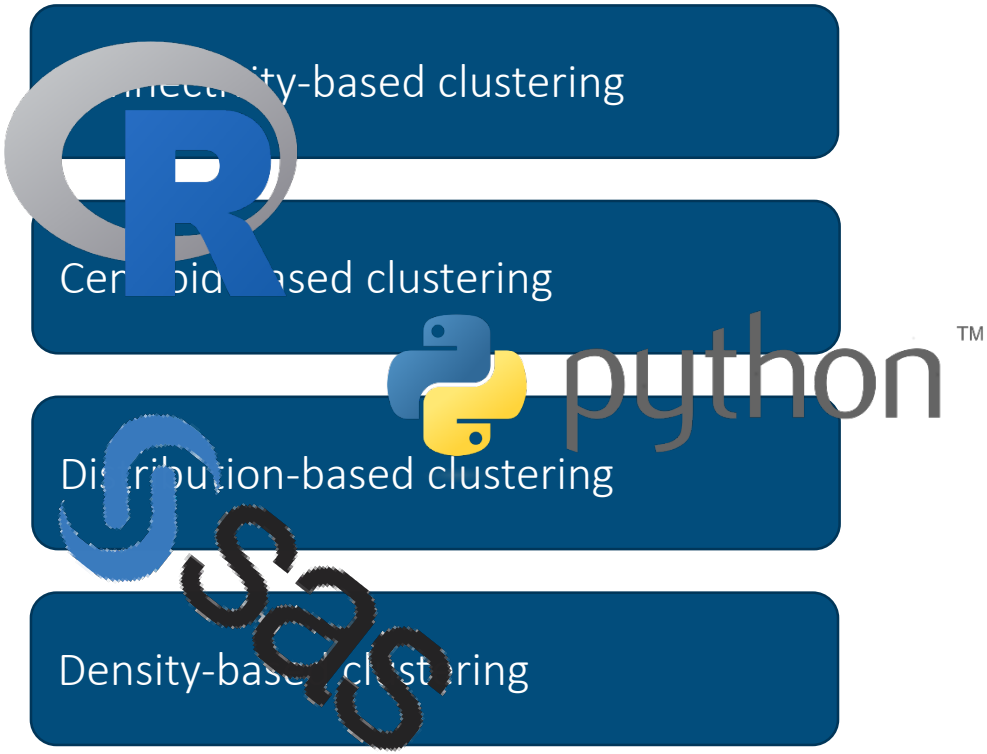

Select arbitrary starting point
Points within ε distance of the point are considered neighborhood points
If sufficient points exist within the neighborhood clustering process starts
Points within ε distance neighborhood become part of the same cluster
Retrieve unvisited point
Iterate

Process



*http://arogozhnikov.github.io/2017/07/10/opera-clustering.html

Clustering toolbox



Questions?

Talex Diede
talex.diede@milliman.com

SOCIETY OF ACTUARIES