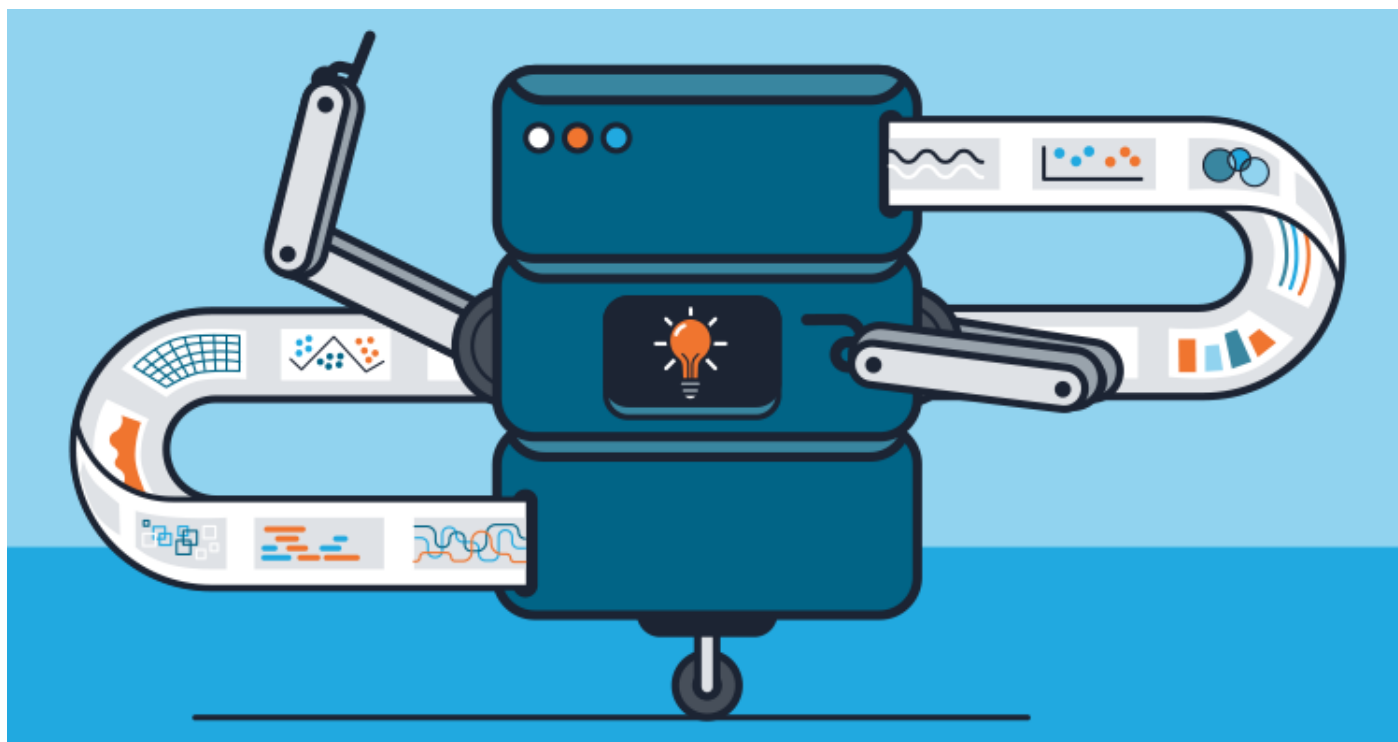# K-Means: a popular algorithm for segmentation

Zylotech™

Customer Data & Analytics Blog

Janet Wagner, on February 21, 2019 | 3 minute read

Many businesses today are using machine learning algorithms to build intelligent applications. One of the ways algorithms are used in marketing is customer segmentation- grouping customers based on specific qualities such as demographics or behavioral patterns.

In machine learning, one method for segmenting customers is [clustering](). Clustering algorithms use an [unsupervised learning]() approach which means

that the algorithms are not trained with labeled training data. Instead, the algorithms learn to identify patterns and find structures in the data on their own. Common clustering algorithms include k-means, [Gaussian mixture models](#) (GMMs), and [hierarchical clustering](#). When it comes to segmentation k-means is a popular algorithm choice.

What is K-Means?

The [k-means](#) algorithm finds clusters of data with similar attributes then creates distinct groups. Each group contains data points that are similar to each other, and the number of groups is represented by the variable K. You must input the total number of clusters (K) to be created by the algorithm. For example, if you want to group customers by annual income or spend data, the input for variable K would be the number of groups you'd like the algorithm to look for.

Each cluster has a centroid which is the location (real or imaginary) that represents the center of the cluster. The algorithm selects data points and groups them together in a way so that each data point is attached to a cluster with the shortest mean distance ("means"). The geometry or metric used by the algorithm is the distances between points. The [Euclidean distance](#) function is commonly used with k-means. The algorithm iterates until all the centroids move less than the [convergence distance](#).

A Popular Choice for Segmentation

So, why is k-means a popular choice when it comes to segmentation? K-means is one of the simplest unsupervised learning algorithms to implement and use for solving clustering problems. The

 algorithm only requires one parameter (K) to be inputted. Some of the other clustering algorithms are harder to implement than k-means because they have more parameters that must be set. And for some of the other clustering algorithms, computation tends to be more complex with greater overhead than k-means. Another reason k-means is widely used for segmentation is that the algorithm works well on large datasets; it scales well to a large number of samples. K-means is a general-purpose clustering algorithm that works well for a wide variety of verticals and use cases.

Not Without Some Drawbacks

While k-means is a fast and easy-to-use algorithm, it's not without some drawbacks. The algorithm assumes clusters are even in size and spherical. In some cases, this assumption may reduce the accuracy of the algorithm. K-means is well suited for flat geometry clustering use cases. If the clusters have complicated geometry, then k-means will not work well. An algorithm that can handle clusters with non-flat geometry would be needed.

Understand Customers at a Granular Level

Segmentation allows marketers to understand customers at a granular level and build highly personalized marketing campaigns that target audiences when they're ready to buy.

You could use an open-source machine learning framework like [scikit-learn](#) or [TensorFlow](#) to implement a clustering algorithm like k-means for customer segmentation. However, our self-learning customer data platform makes customer segmentation a whole lot easier.

Our platform does the marketing data [prep for you](#) and lets you easily separate customers into [micro-groups](#) such as LifeStage, LifeStyle, Potential, and RFV.

Interested in customer segmentation and data analytics? [Request a demo](#).

Janet Wagner is a Zylotech contributing writer.

If you liked this post, check out our recent [blog post: Common in algorithms in marketing.](#)

Subscribe to our blog newsletter for all things customer data and analytics, AI and marketing, sent monthly.

Save