Bellevue University

K-Means Clustering

Ashley Mayo

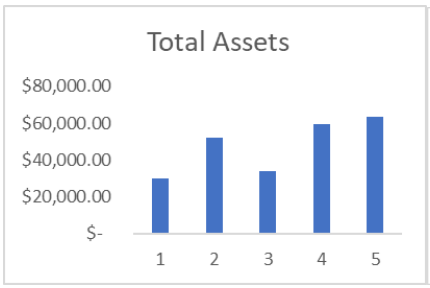
DSC 500: Introduction to Data Science

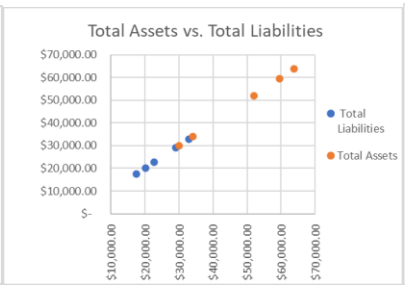
An unsupervised learning algorithm, called K-Means Clustering, allows data points to be identified for each of the k clusters (Bento, 2018). This algorithm creates a group of clusters through the data points, separated into those that have similar variance. The within-cluster sum of squares will be like each other, while data points in other clusters have lower similarity levels (Bento, 2018). An important aspect of this algorithm to address is, you will need to specify the parameter K, which represents the number of clusters (Bento, 2018).

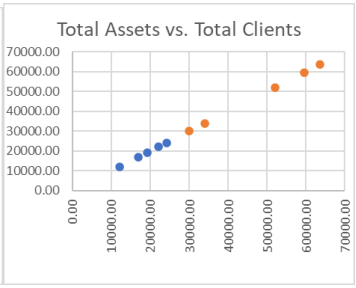
To demonstrate how K-Means Clustering is utilized the graph below will be used to provide an example. The data below shows the assets, liabilities, and equity of the Pawfect Care Company, and the total clients for the year.

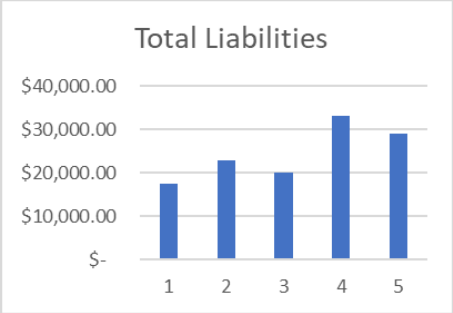
Year	Total Assets	Total Liabilities	Total Clients
2018	$30,000	$17,500	12,150
2019	$52,000	$22,750	17,005
2020	$34,000	$20,125	19,175
2021	$59,500	$32,990	22,225
2022	$63,700	$29,075	24,230

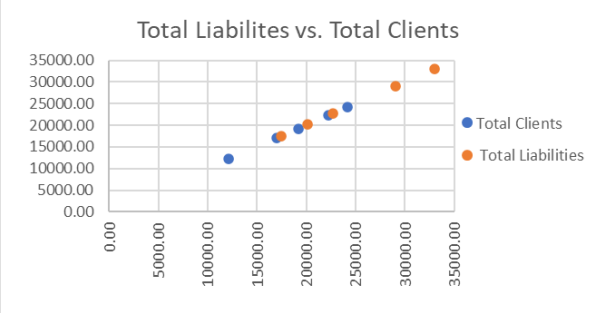
The benefit of applying k-means clustering to this data set is that it allows the company to potentially improve their performance. It will allow them to focus on the years with higher performance and evaluate what was done differently compared to the other years.











The graphs above were formulated from the data set in the table. K-means clustering would be to help uncover what groups to cluster the data into. In this situation where it may not be as obvious for the clusters, it is important to find the K value. Once the distortion is found through plotting the datasets together, the K value is clear. Additional graphs can be formulated after the k value is found and then analysis can be used to provide insight into the company's performance numbers.

References

Bento, C. (2018, Dec. 3). *K-means in real life: Clustering workout sessions*. Medium. Retrieved on July 13, 2023, from https://towardsdatascience.com/k-means-in-real-life-clustering-workout-sessions-119946f9e8dd