**Investment Banking - Data mining to help determine cross-sell opportunities.**

A MNC dealing in financial services showed interest in mining complex customer behavioural data. By deploying analytics services and series of ML and data science initiatives. Using advanced techniques like K-Means clustering, the company was able to dig into its structured data that it has been gathering for years and correlated it with an army of unstructured data to enhance cross-selling.

K-means clustering is a type of Non-hierarchical clustering that uses partitioning methods and within cluster variance as a measure to form homogenous groups.

The objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.

You’ll define a target number k, which refers to the number of centroids you need in the dataset. A centroid is the imaginary location representing the centre of the cluster.

Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.

In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid. K-means is not distance based rather based on within cluster variation (squared distance from cluster centre). This algorithm aims at segmenting data such that within cluster variation is reduced.

K-means is superior to hierarchical clustering as it is less impacted by outliers and computationally it’s faster.