**QTW-Week 5 Presession Submission-Clustering**

**This week was all about Clustering**

**Clustering:** Clustering is an unsupervised machine learning algorithm in which no target variable is given. We need to find the relationships between the variables and structure within the data.

**Clustering Algorithms:**

* MiniBatchKMeans
* Affinity Propagation
* DBSCAN
* Spectral
* Agglomerative Clustering
* Birch
* MeanShift
* Ward
* OPTICS
* Gaussian Mixture

**Distance:** Shortest distance between 2 data points in a straight line.

Euclidean distance is considered as Standard distance.

Manhattan is considered as non-standard distance.

Text

Description automatically generated with medium confidence

The above formula is called Minkoskwi distance formula where,

P=1 is **Manhattan** distance

P=2 is **Euclidean** distance

P=infinity is **Chebyshev** distance

**Cosine Distance:** Cosine similarity concept is used to measure the similarity of two vectors moving in same direction.

It is measured by the cosine of the angle between the two vectors and determines whether the vectors are pointing ion roughly the same direction. It is often used on document similarity in text analysis.

**Question:** What approach we must use, if there is no clue of how many cluster will be formed?