Supervised (outcome)

1. Regression X independent / Y dependent (Numbers)
2. Classification X Independent / Y dependent (classes)

Unsupervised (outcome)

1. X Independent, no Y variable
2. Divide them to different Groups / Clusters
   1. When = Unknown patterns discovery
      1. Eg. C1, C2, C3 (three Clusters)

What

1. If Only x Values, we need to calculate the distance between datapoints to create clusters
2. If Only X values, divide them into groups
   1. If group number / no of clusters is already defined = K means Clustering
   2. If group number / no of clusters is not defined = Hierarchical Clustering

Proximity Matrix

1. Reg
   1. SSE
   2. RMSE
   3. MAE
   4. COD/R2 = 0-1 NEAREST to 1 => is better
2. Class
   1. pRecision
   2. recall
   3. AUC
   4. Accuracy
   5. Confusion Matrix
3. Clustering Matrix
   1. SC (Sillohette Quiotient) => Whether clustering techniques is good or not
      1. Values = -1 to 1
         1. Value = 1, Cluster Visibility is very clear (highly dense clustering)
         2. Value = 0, Clusters are indifferent, distance between cluster to cluster is very small / not that significant
         3. Value = -1, Completely Wrong, assigning itself wrong.
            1. Incorrect Clustering
      2. Better to have 0-1, the near to 1 again is the best HC
      3. SC score = (b-a) / max(a,b)
         1. A = in the Same Cluster, mean distance between sample and all other datapoints
         2. B = in the nearest cluster, mean distance between the sample and all other datapoints.