Clustering

→ unsupervised machine learning
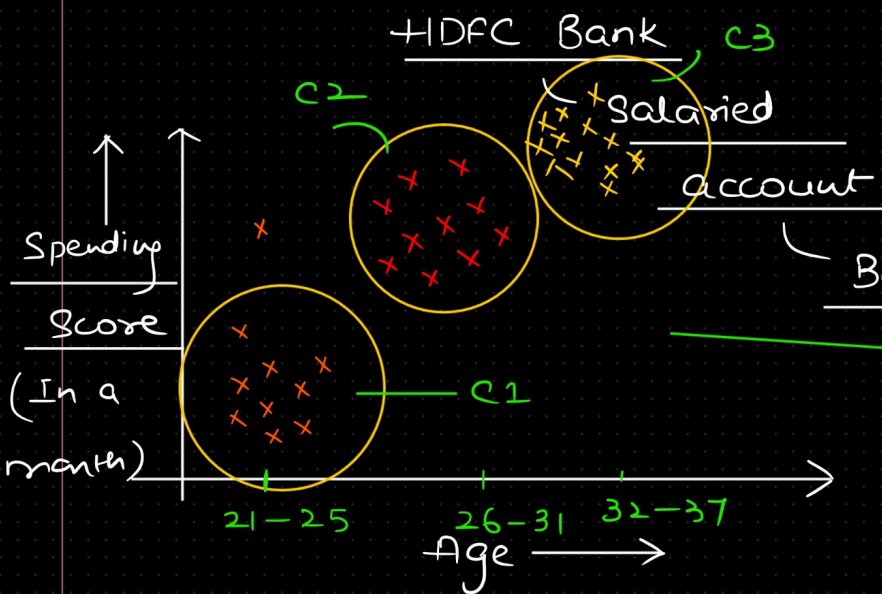
↳ labelled data not available

HDFC Bank, C3

C2

Salaried account

Bengaluru

→ Customer Segmentation

Spending Score (In a month)

C1

21-25   26-31 32-37
Age →

Clusters

K-Means Clustering

↳ Elbow Method
↓
Optimal value of k (num of cluster)

① Euclidean
② Manhattan
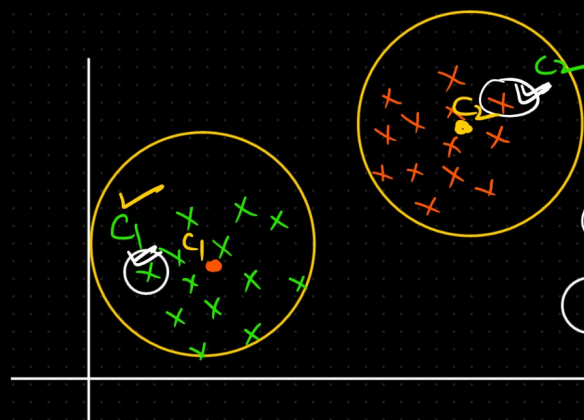
C1 (Distance of all data points from cluster C1 & C2)
C2

K = 2

Euclidean Distance

•————————•
$P_1(x_1, y_1)$ $P_2(x_2, y_2)$

$$\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}$$
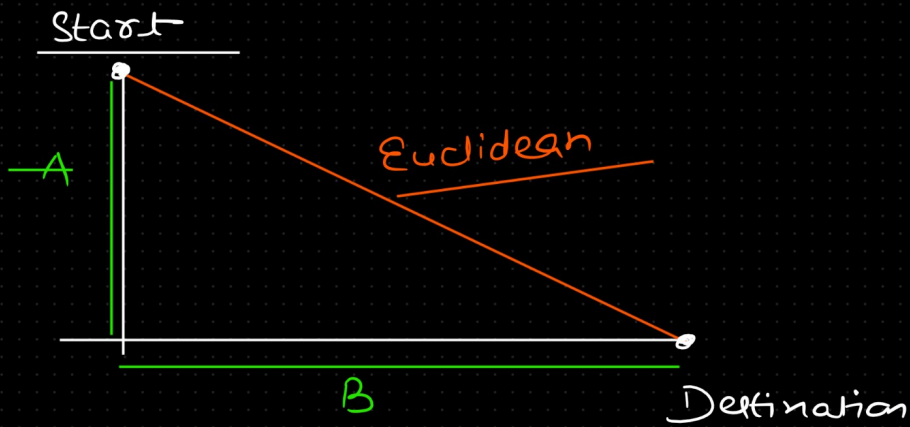
Manhattan Distance

P_1 ———————— P_2
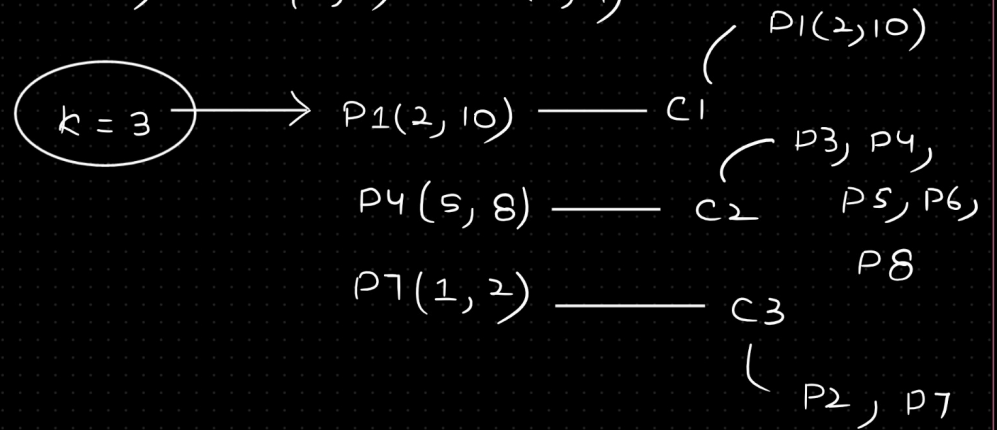(x_1, y_1) (x_2, y_2)

$$|y_2 - y_1| + |x_2 - x_1|$$

P1 (2, 10)   P2 (2, 5)   P3 (8, 4)   P4 (5, 8)   P5 (7, 5)

P6 (6, 4)   P7 (1, 2)   P8 (4, 9)

P1 (2, 10)

$k = 3$ $\longrightarrow$ P1 (2, 10) —— C1

P3, P4,

P4 (5, 8) —— C2   P5, P6,

P8

P7 (1, 2) —— C3

P2, P7

Start

A

Euclidean

B

Destination

A + B = manhattan

Distance