

Report of Clustering on Customer Data

Arnab Roy

January 2025

1 Feature Engineering

Clustering on customer data has been performed using **Customer Details, Product Details and Transaction data**. Following features have been extracted to represent a single customer -

- Region where the customer belongs to
- Product category wise total spending
- Product category wise total products purchased

2 Algorithm Used

K-Means clustering is used for customer segmentation. From the elbow plot (WCSS vs. Clusters) the value of K is chosen as 5. The DB index after creating 5 clusters is 1.5. The data have been transformed to 2D using TSNE for visualisation.

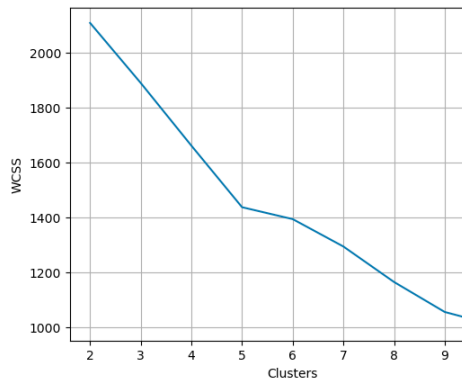


Figure 1: First image caption.

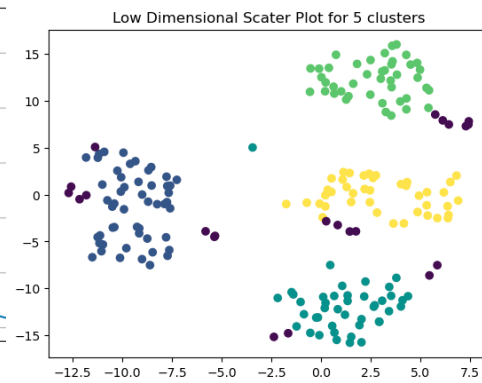


Figure 2: Second image caption.