



UNSUPERVISED LEARNING DIABETES DATASET

Akash Kadia

A decorative geometric pattern on the left side of the slide, featuring a grid of squares with various patterns: a blue square with white concentric circles, a purple square with white concentric circles, a blue square with white concentric circles, a purple square with white concentric circles, a blue square with white concentric circles, and a purple square with white concentric circles.

FLOW STRUCTURE

- ❑ Project Goal
- ❑ EDA - Exploratory Data Analysis
- ❑ K-Means Clustering
- ❑ Hierarchical Clustering
- ❑ PCA
- ❑ Conclusion



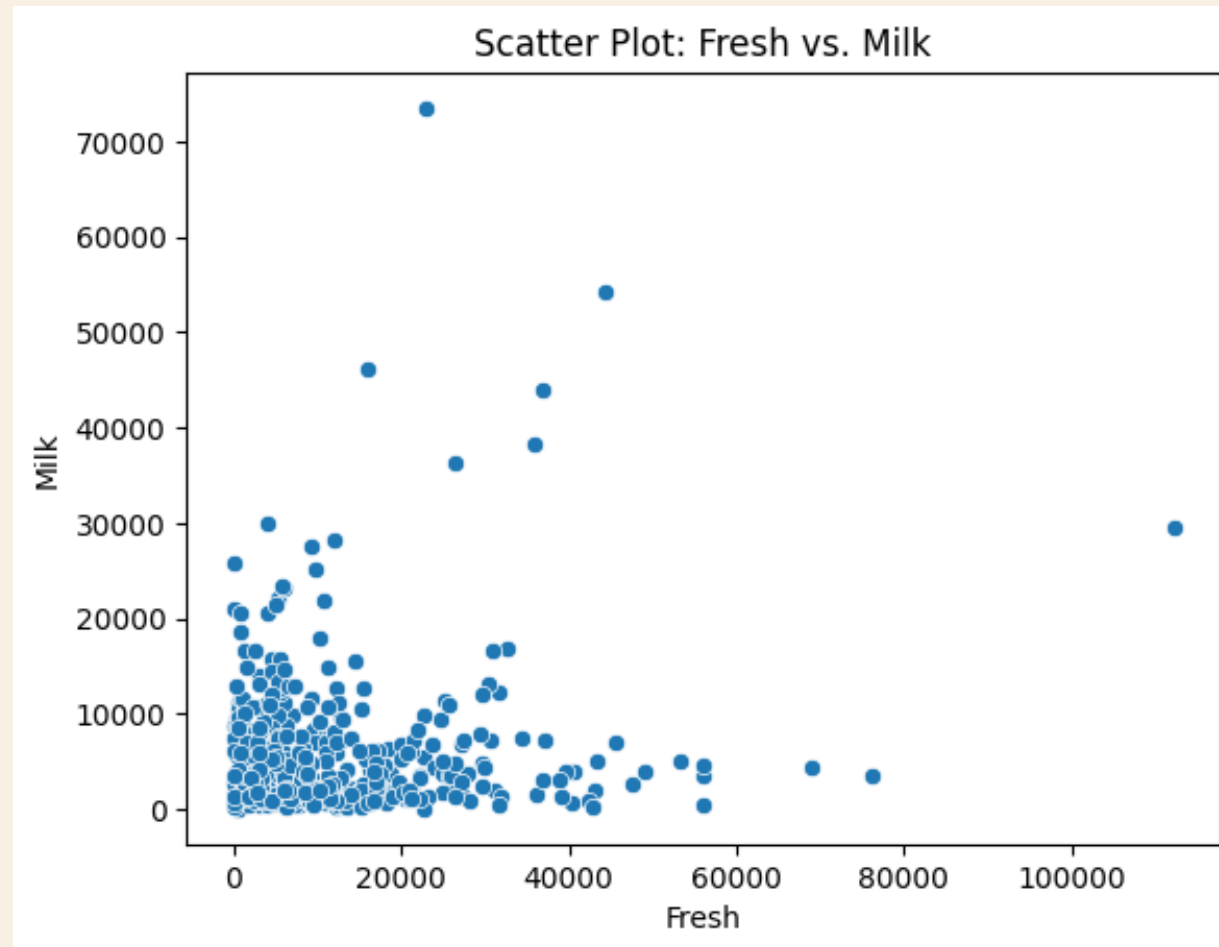
PROJECT GOAL

- ❑ Unsupervised Learning Techniques
- ❑ gain insights from the data sets to make informed decisions

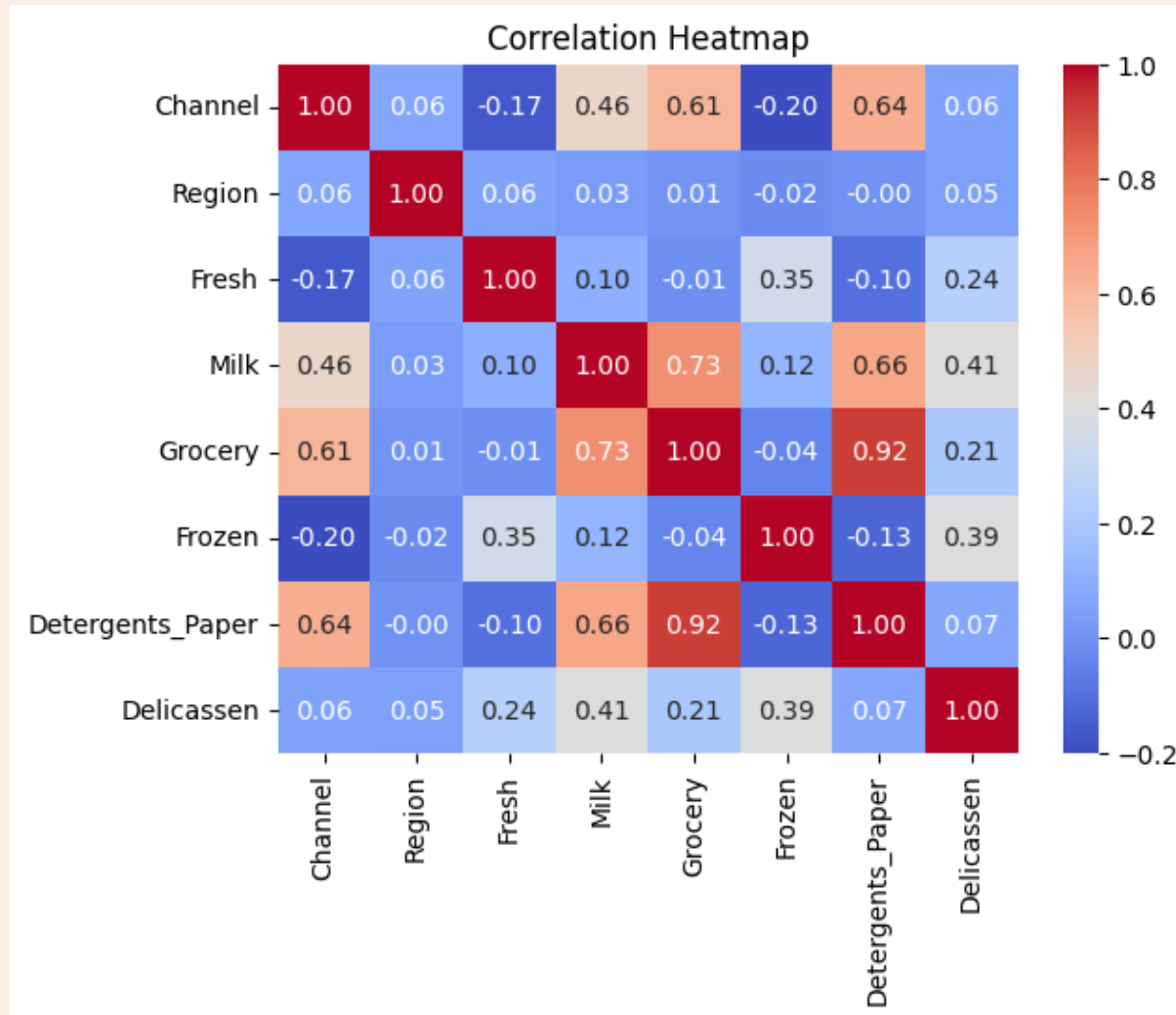


EXPLORATORY DATA ANALYSIS

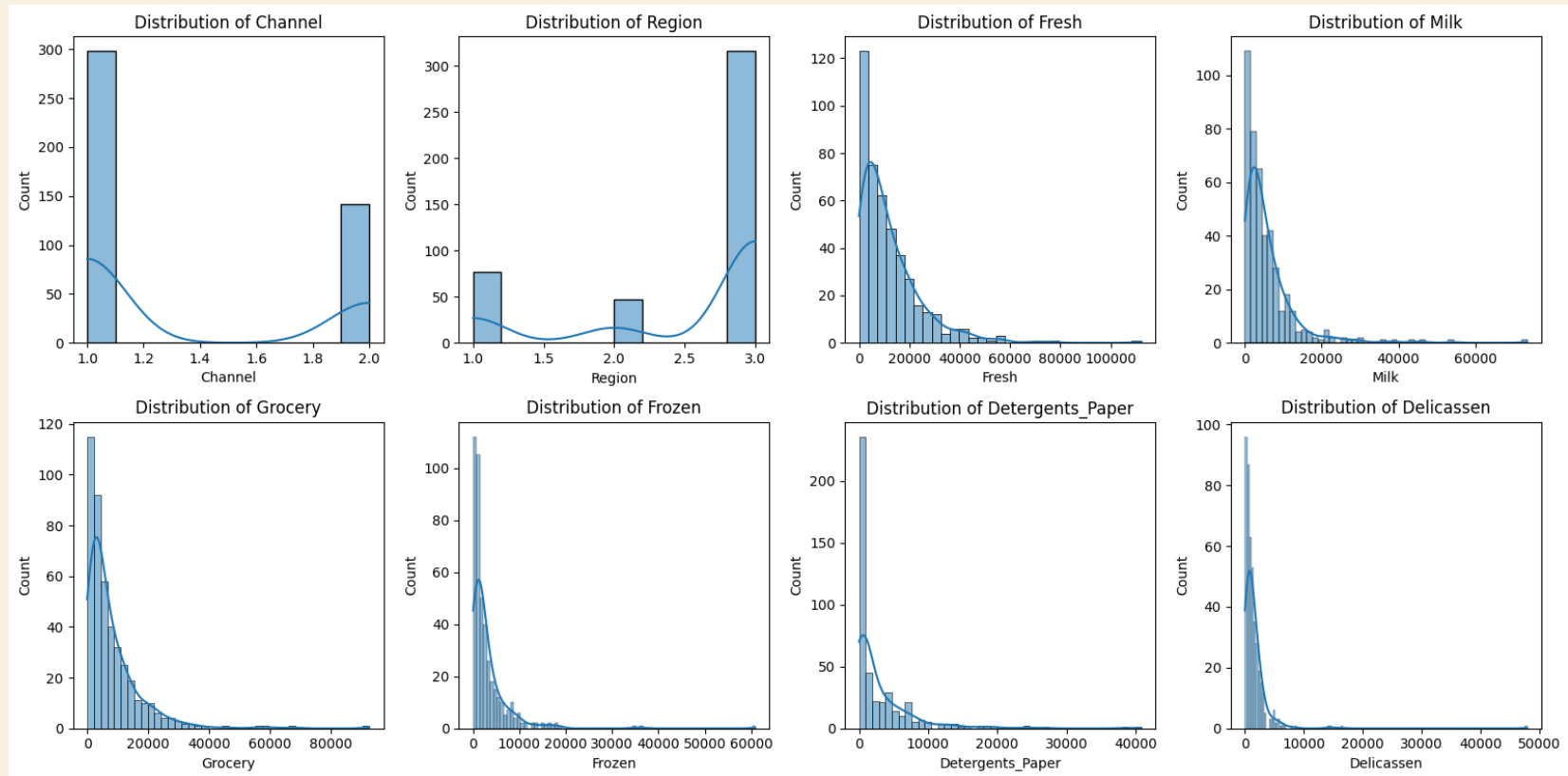
FRESH VS MILK



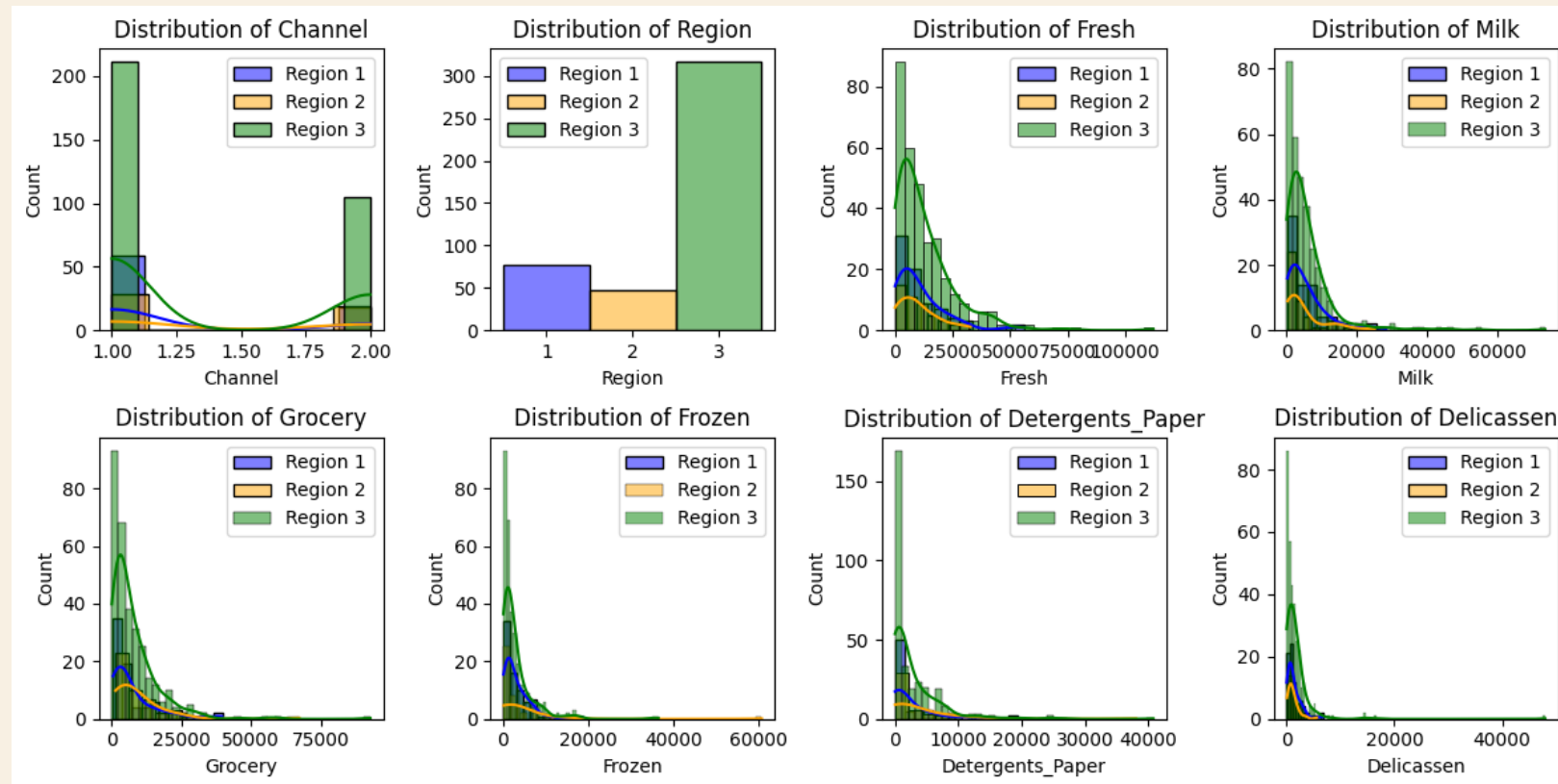
CORRELATION HEATMAP



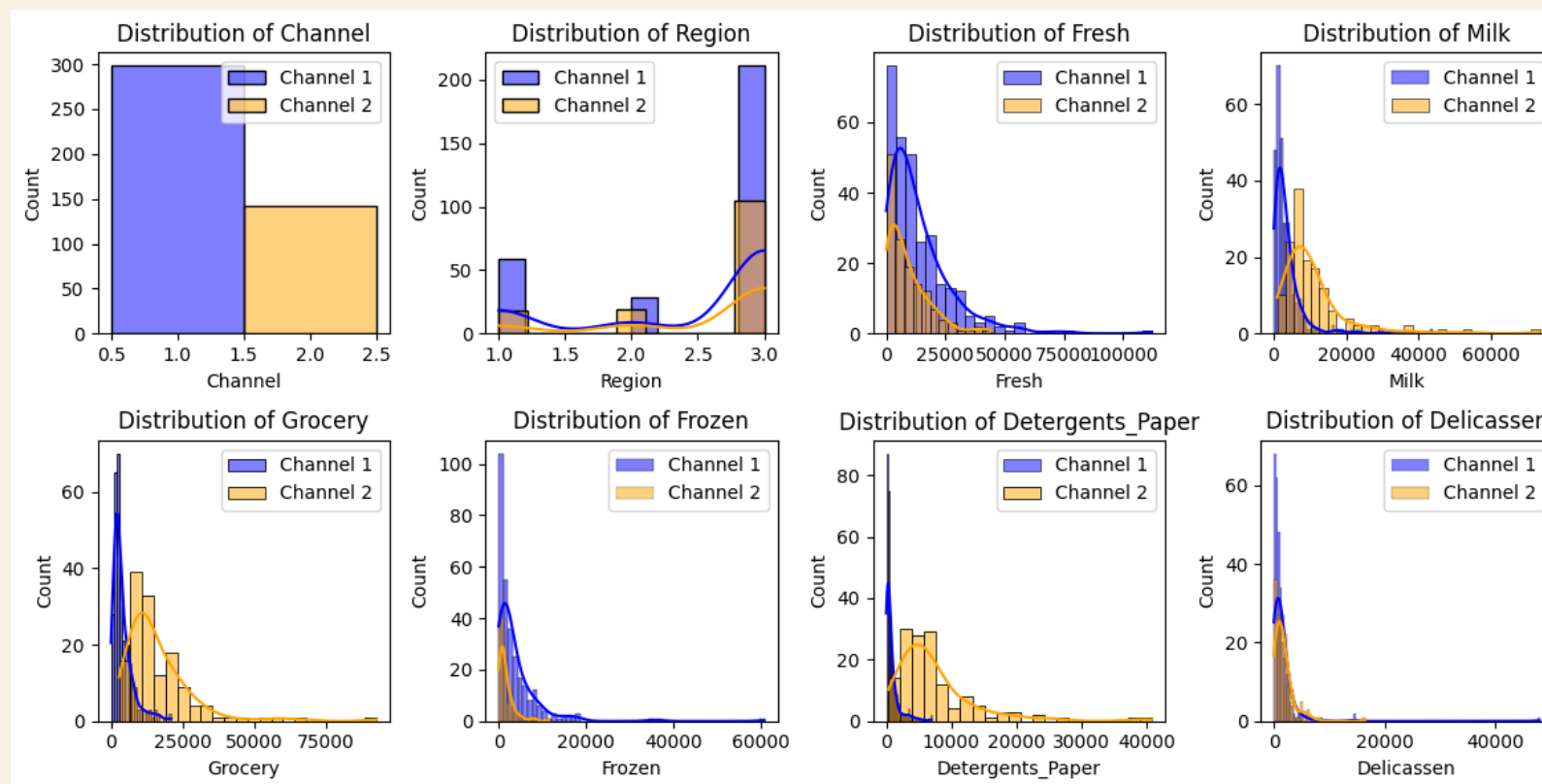
DISTRIBUTION OF EACH VARIABLE



DISTRIBUTION OF EACH VARIABLE WITH REGIONS



DISTRIBUTION OF EACH VARIABLE WITH CHANNELS





K-MEANS CLUSTERING

HYPOTHESIS

Different customer segments can be identified based on their purchasing behavior in the 'Fresh', 'Milk', 'Grocery', 'Frozen', 'Detergents_Paper', and 'Delicassen' categories.

PROCESS

- Use Elbow Method to determine K value
- Perform K-Means clustering
- $SSE = 1074854303.35$

SSE is high so scale the data for better result

SCALING

- For better SSE can use two scaling methods
 1. Standardization
 2. Normalization

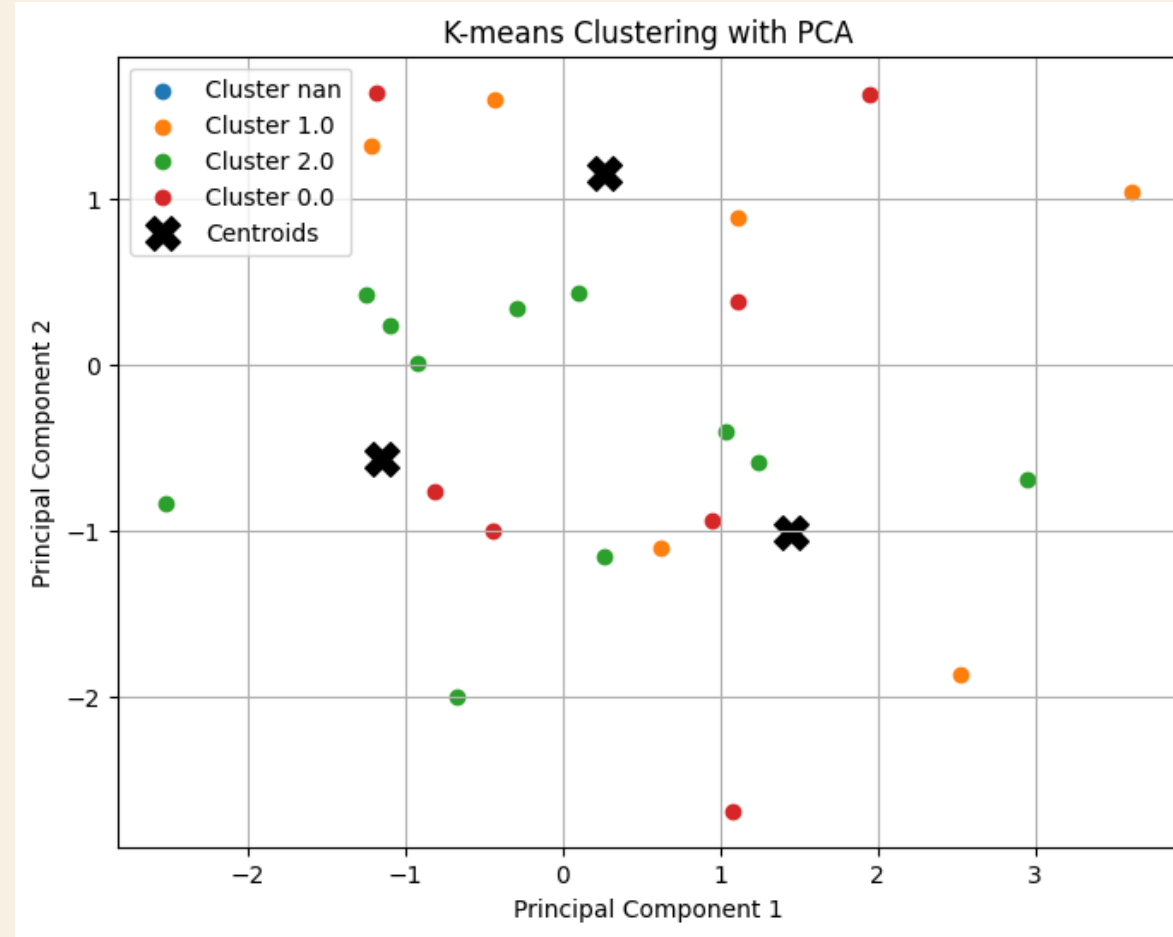
Used Standardization

SCALED DATA

- Use Elbow Method to determine K value
- Perform K-Means clustering
- $SSE = 490.07$

well-defined clusters

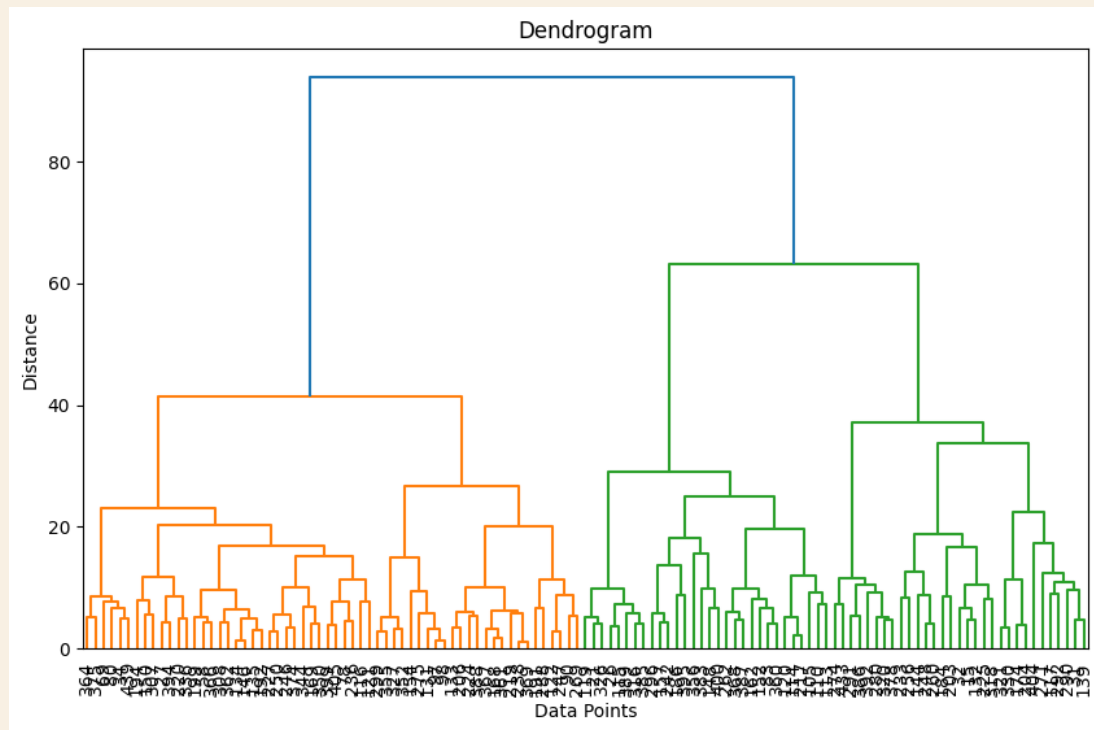
K-MEANS CLUSTERS WITH PCA



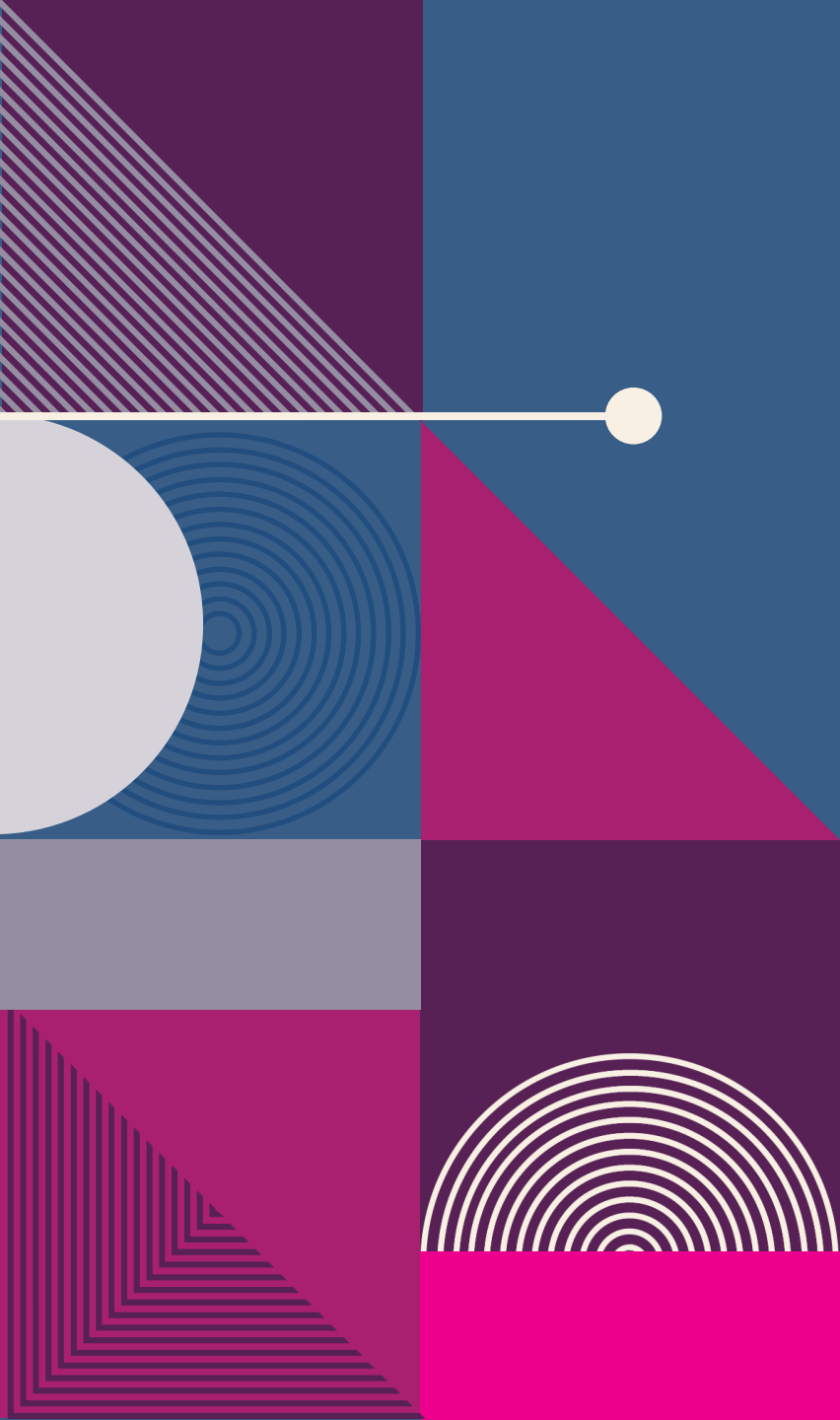
The top-left corner of the slide features a complex arrangement of geometric shapes and patterns. It includes a dark purple square, a blue square with concentric circles, a magenta triangle, a light purple circle, a grey rectangle, a magenta square with concentric circles, and a magenta square with a series of parallel lines. A white line with a dot extends from the top-left towards the center of the slide.

HIERARCHICAL CLUSTERING

DENDROGRAM

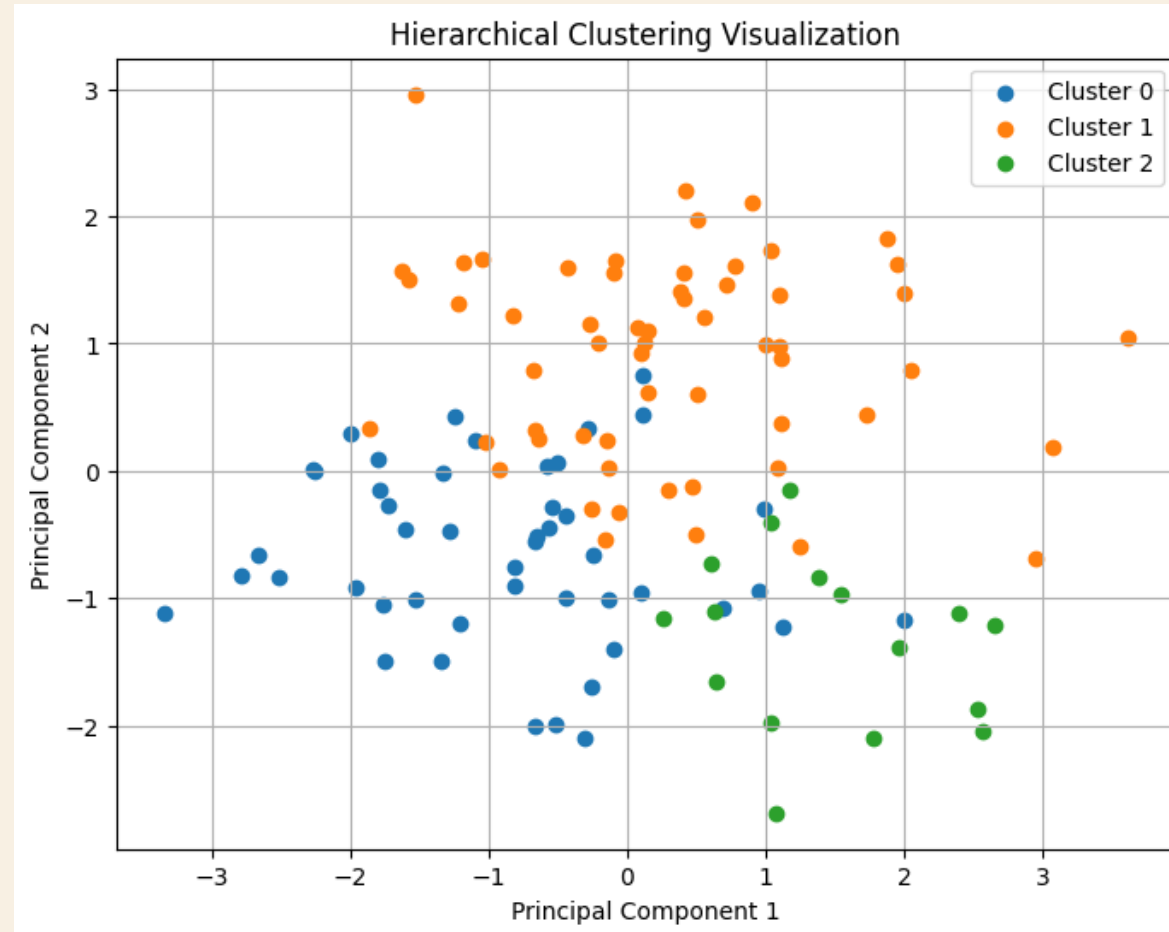


- Number of clusters = 3
- Created cluster assignment for PCA and visualization
- Can use PCA or t-SNE



PCA

HIERARCHICAL CLUSTERING





CONCLUSION

- The scatter plot obtained from K-means clustering shows more scattered and disrupted data points, indicating that the clusters may not be well-defined or separated in the reduced space.
- The hierarchical clustering diagram exhibits well-defined clusters, suggesting that the algorithm has effectively grouped similar data points together based on the similarity measure, resulting in more cohesive and distinct clusters.

An abstract geometric design on the left side of the slide. It features a dark blue background with various geometric shapes and patterns. A white circle is positioned near the top left. Below it, a light blue semi-circle is visible. To the right of the semi-circle, there is a pink triangle with diagonal lines. Further down, there is a pink square with a pattern of concentric lines. At the bottom, there is a pink triangle with a pattern of concentric lines. The overall design is modern and minimalist.

THANK YOU