# Profiling Customer Cluster From Behavioral Data

6620422012 ธนัทพงศ์ คุณากรเกษม  6620422026 วิญณา บางสร้อย

6620422015 ชานุกฤต วรรณวัฒน์  6620422027 นภทีป์ กาญจนาวิลาส

**NIDA** WISDOM for Sustainable Development
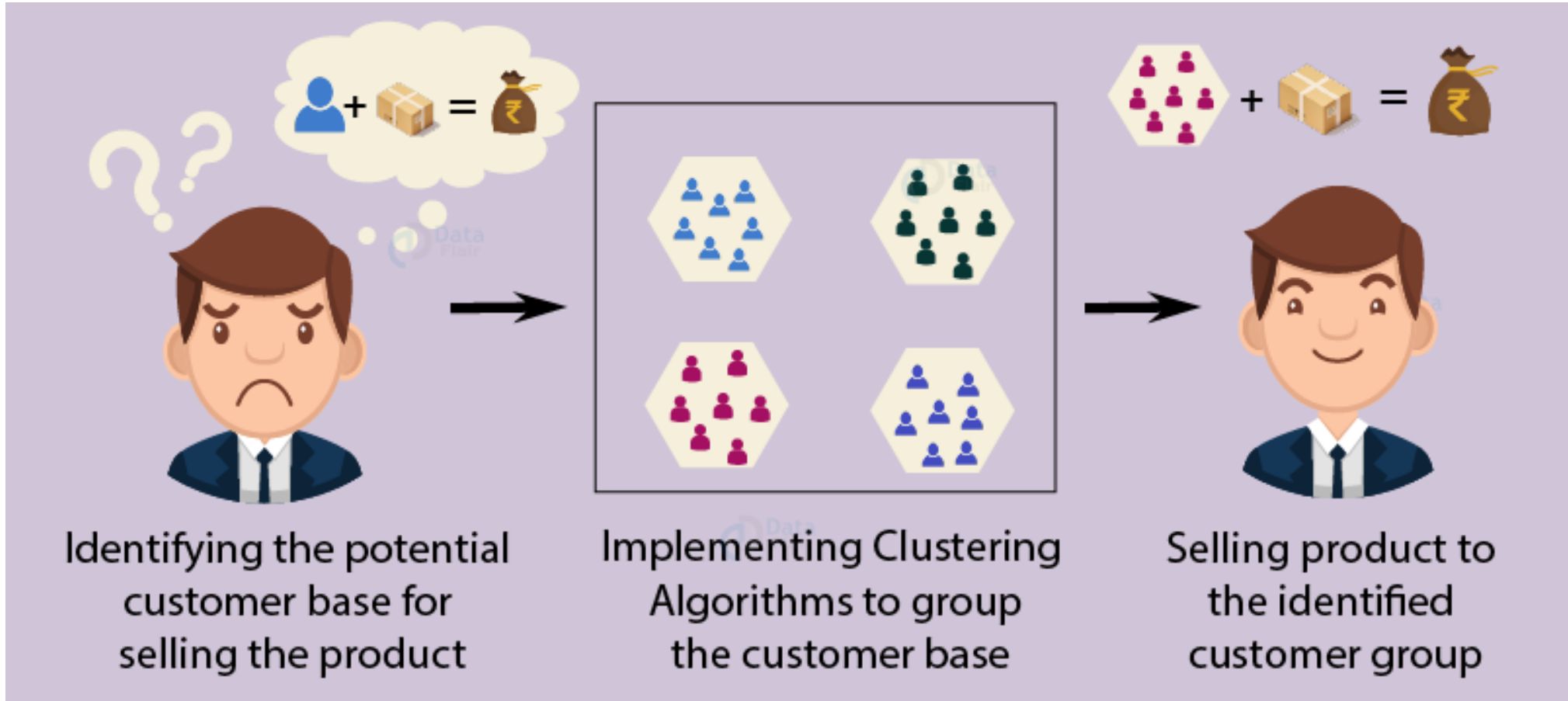
**GSAS** Graduate School of Applied Statistics

## Think You Know Your Customers? Think Again!

In today's competitive market, understanding customer behavior is crucial for businesses to succeed. This project focuses on identifying distinct customer groups based on their behaviors and demographics. By clustering customers, businesses can gain valuable insights into the factors that drive customer preferences and loyalty.

Our team has developed the following dashboards:
- **Overview of Customer Data**
- **Customer Segmentation**
- **Insight Dashboard based on Customer Segments**



Identifying the potential customer base for selling the product → Implementing Clustering Algorithms to group the customer base → Selling product to the identified customer group

## Traditional vs AI Approach

| Aspect | Traditional Approach | AI Approach |
|---|---|---|
| Basis | Predefined criteria (e.g., demographics). | Insights from algorithms. |
| Data Sources | Surveys, focus groups, basic metrics. | Large, diverse datasets (e.g., CRM, web). |
| Insights | Basic understanding. | Deep, multidimensional analysis. |
| Accuracy | Prone to bias, oversimplified. | Precise, pattern-based insights. |
| Speed | Slow due to manual processes. | Fast and automated. |
| Personalization | Limited targeting. | Enables precise personalization. |
| Cost | High manual effort and cost. | Cost-effective after setup. |

## Clustering Algorithms

### K-Mean with PCA Clustering

| | Income | Kidhome | Teenhome | Recency | MntWines | MntFruits | | PC1 | PC2 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 58,138 | 0 | 0 | 58 | 635 | 88 | 0 | 3.9061 | -1.3603 |
| 1 | 46,344 | 1 | 1 | 38 | 11 | 1 | 1 | -2.2067 | -0.6226 |
| 2 | 71,613 | 0 | 0 | 26 | 426 | 49 | 2 | 1.9702 | -0.7049 |
| 3 | 26,646 | 1 | 0 | 26 | 11 | 4 | 3 | -2.5575 | -1.6737 |

After evaluating multiple models, K-Means was selected for its superior silhouette score. To optimize the clustering process, we applied Principal Component Analysis (PCA) to reduce the 21 dimensions of customer data to 2 components (PC1 and PC2) before performing clustering with K-Means.

### The Elbow Method
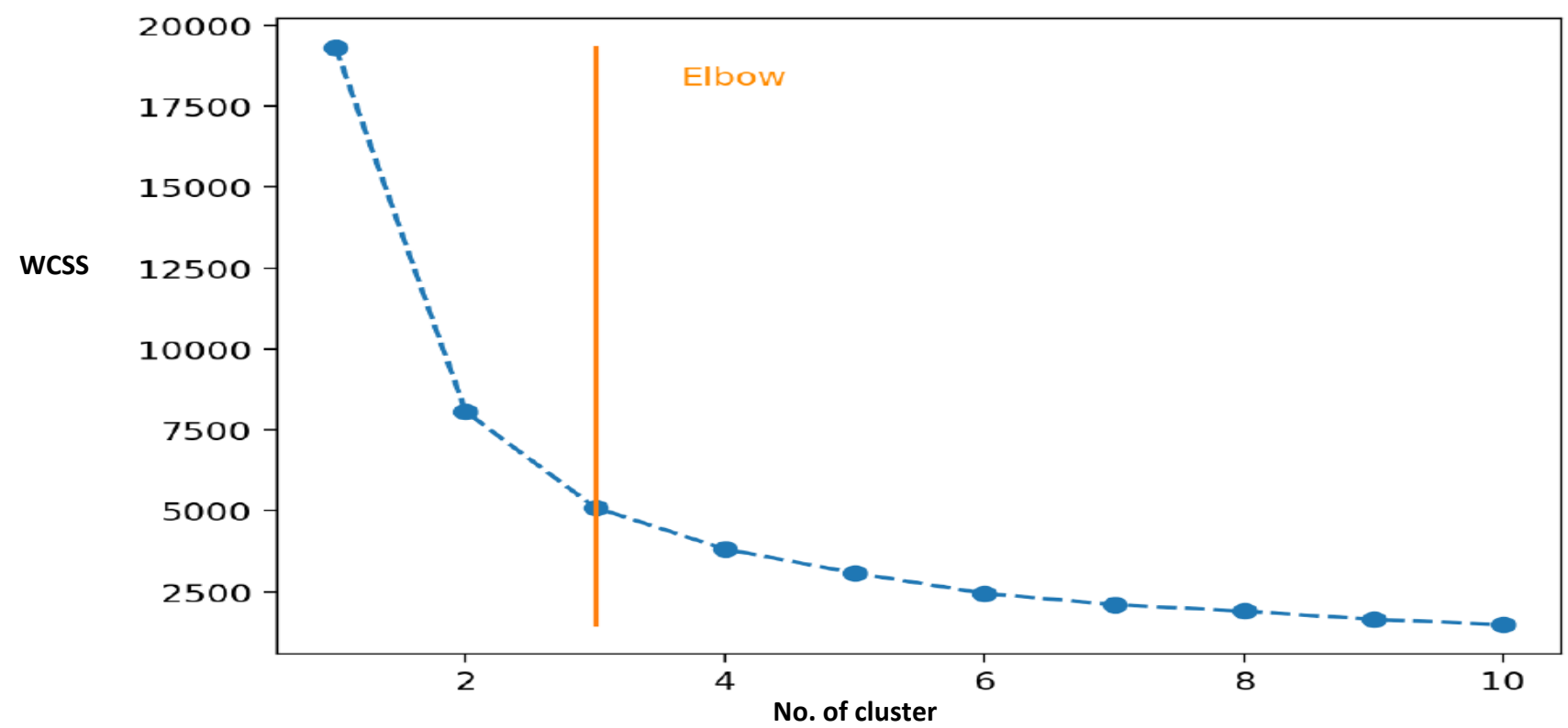


Using the Elbow Method, we tested 'k' values from 1 to 10 and identified the optimal number of clusters at the "elbow point," where the WCSS shows a significant reduction before leveling off.

### Visualize Customer Segmentation
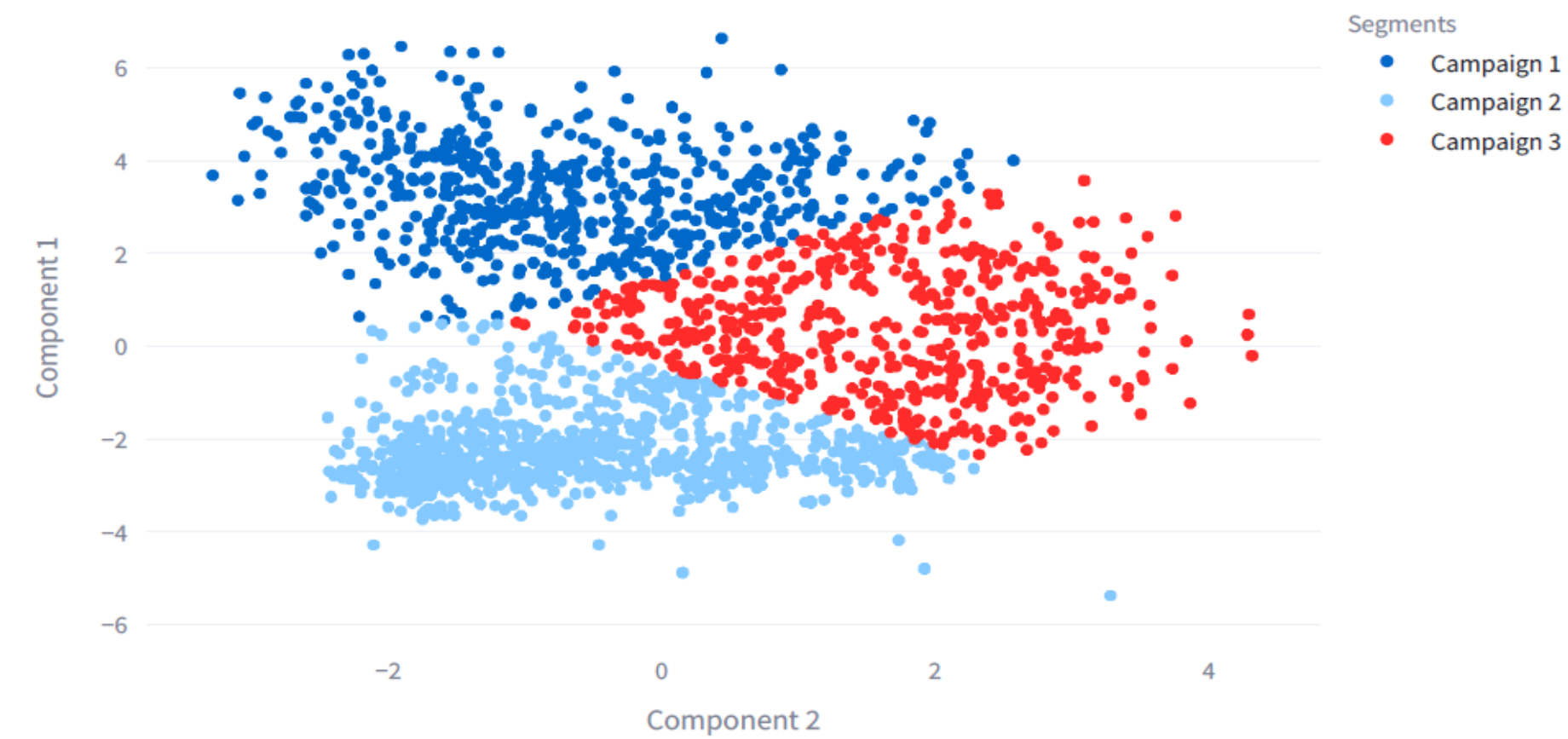


Applying the optimal clusters from the Elbow Method, we visualized customer segments in a dot plot, with points representing customers plotted on PC1 and PC2 axes, color-coded by cluster.

## Profiling Customer Cluster



After identify customer segment, we create an interactive visualization to profile each customer segment. The graph allows us to explore the unique characteristics of each cluster. Users can adjust the number of clusters dynamically to refine segmentation.

This visualization helps businesses better understand their customers' preferences, enabling the development of targeted products or marketing campaigns tailored to each segment's needs.

## Gemini Integration

🤖 **Ask Gemini-Pro about data**

Limitation only 50 samples of data given to generative AI.

Ask Gemini-Pro...

We integrated AI Gemini into the overview dashboard, allowing users to input queries and receive analyses based on customer data.

### References

https://github.com/mayonaise01/dads5001_data_product