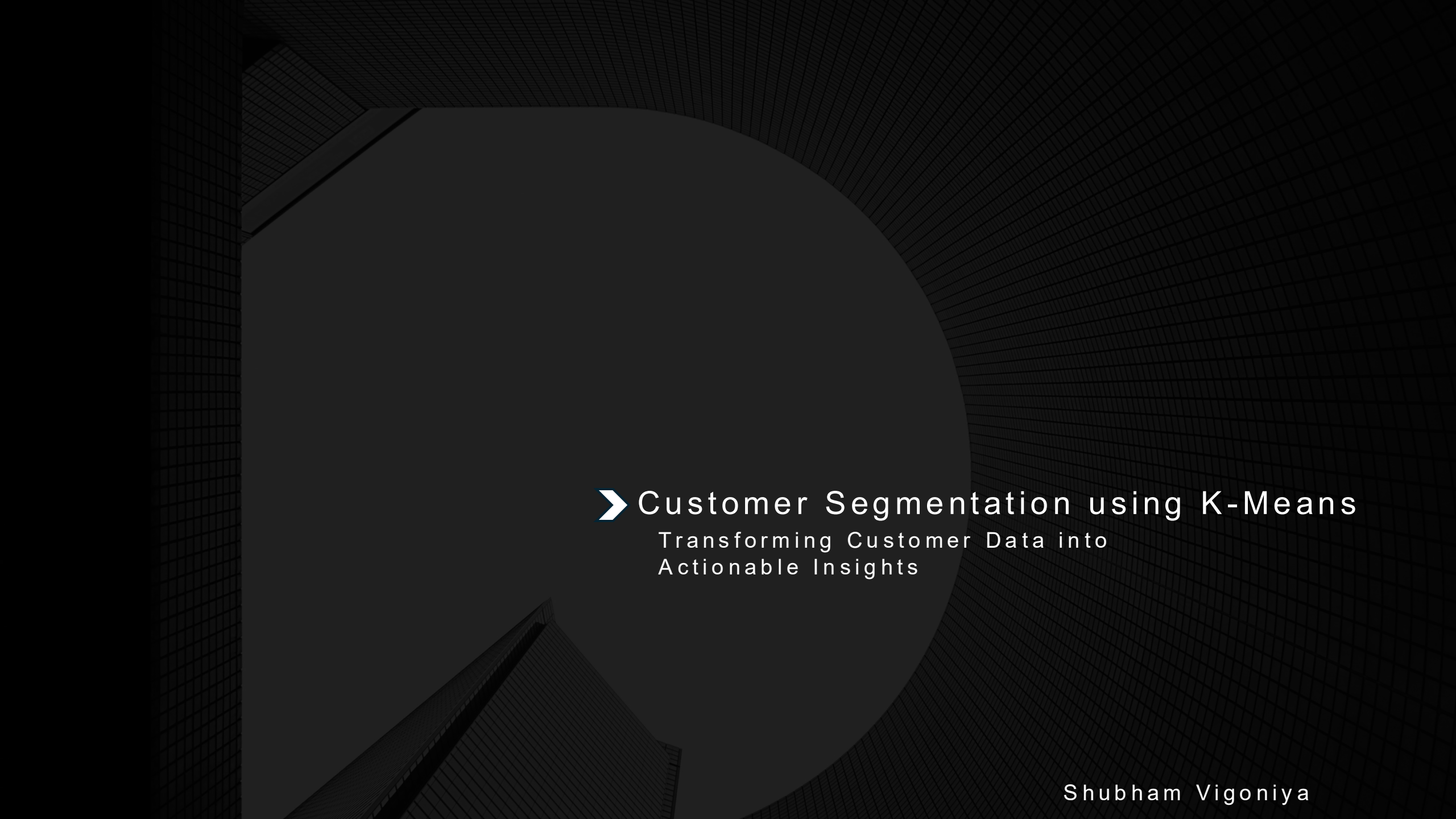


# Customer Segmentation using K-Means

Transforming Customer Data into Actionable Insights

Shubham Vigoniya

# CONTENTS

# THE BUSINESS PROBLEM & PROJECT GOAL

**Problem:** Need to move beyond generic marketing and personalize customer engagement.

**Goal:** Group customers based on purchasing behaviour and demographics to define targetable segments.

**Expected Value:** Higher campaign ROI, improved customer retention, and tailored product strategy.

# DATA OVERVIEW AND FEATURES

Source: customer_data.csv (2240 records, 29 features)

Key Data Categories:

1. Demographics: Income, Age, Education, Marital_Status.

2. Product Spend: MntWines, MntMeatProducts, Total_Spent.

3. Engagement: Recency (Days since last purchase), Campaign Acceptance (AcceptedCmp1-5).

# DATA CLEANING AND FEATURE ENGINEERING

**New Features Created:**

1. **Age** (Calculated from Year_Birth).
2. **Total_Spent** (Sum of all product spend columns).
3. **Total_Children** (Sum of Kidhome and Teenhome). –

**Handling Missing Data:** Imputed 24 missing Income values using the median to maintain distribution integrity.

**Outlier Management:** Noted and addressed outliers in Year_Birth (e.g., min 1893) and extreme Income values.

# FEATURE TRANSFORMATION & SCALING

1. **Marital Status Consolidation:** Simplified rare/inconsistent categories (Absurd, YOLO, Alone) into Single for cleaner segmentation.

2. **Categorical Encoding:** Used LabelEncoder for remaining categorical features (Education, Marital_Status).

3. **Feature Scaling (Standardization):** Applied StandardScaler to all numerical features to prevent features with larger scales (like Income) from dominating the K-Means algorithm.

# CLUSTERING METHODOLOGY: K-MEANS

**Model Choice:** Unsupervised learning via **K-Means Clustering**.

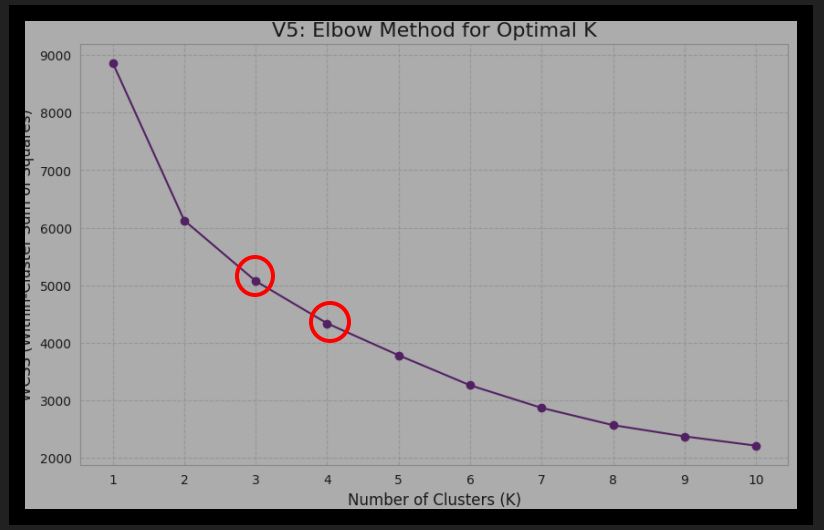
**Mechanism:** Iteratively groups data points by minimizing the **Within-Cluster Sum of Squares (WCSS)**. - **Input:** The final set of scaled, cleaned, and engineered features.

# FINDING THE OPTIMAL K: THE ELBOW METHOD

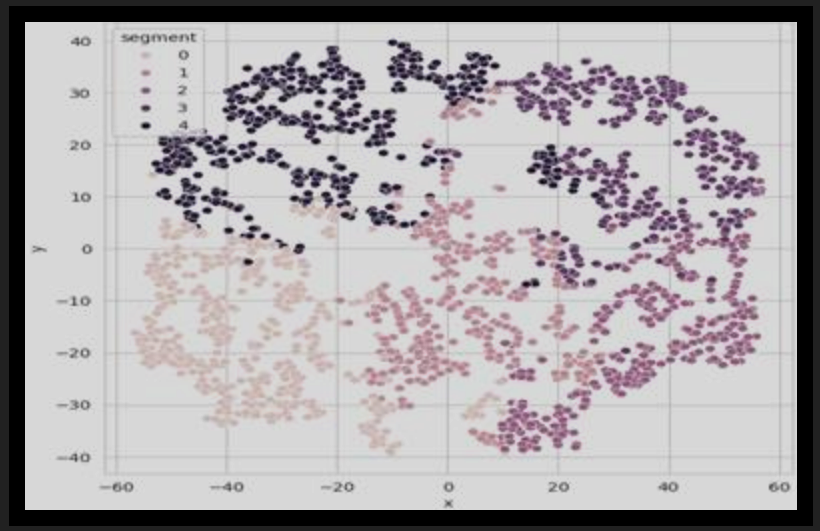
**Analysis:** Performed the Elbow Method to evaluate WCSS for K values from 1 to 10.

**Result:** The elbow point indicates that **K=3** or **K=4** are the optimal number of clusters.

**Decision: K=4** was chosen for a slightly finer, more actionable segmentation.




Optimal value of K




Cluster k = 4 color label map

# SEGMENTS

**Segment 0:** The Affluent Spenders (High-Value):

- Size: Smallest segment (e.g., 16.1% of the total base).
- Traits: Highest Income (~88k), Highest Total_Spent (~$1,257).
- Behaviour: Low Recency, zero children at home, high success rate in major campaigns (especially Cmp5).
- Focus: Premium products, personalized luxury offers, exclusive loyalty status.

**Segment 2**: The Established Spenders (Loyal Mid-High):

- Size: Second smallest segment (e.g., 14.9%).
- Traits: High Income (~68k), High Total_Spent (~$903).
- Behaviour: Highly engaged (low Recency), high rate of Catalog Purchases.
- Focus: Cross-selling, retaining loyalty, targeted campaigns to increase overall spend frequency.

Segment 3: The Average/Mid-Market.

Size: Mid-sized segment (e.g., 23.6%).

Traits: Average Income (~51k), Mid-range Total_Spent (~$401).

Behaviour: Moderate engagement across all channels, lower campaign response rate than the top two groups.

Focus: Promotional deals, increasing basket size, conversion campaigns for high-margin categories.

Segment 1: The Budget-Conscious Family.

Size: Largest segment (e.g., 45.3%). - Traits: Lowest Income (~29k), Lowest Total_Spent (~$46), highest Kidhome.

Behaviour: High number of web visits (deal hunting), lowest purchase frequency.

Focus: Price-sensitive offers, family bundles, driving in-store traffic with discounts.

# STRATEGIC RECOMMENDATIONS FOR MARKETING

**Targeted Budget Allocation:** Focus the majority of high-cost, personalized campaigns on Segments 0 and 2 (High-Value).

**Campaign Design:** Segment 1 needs value/necessity-focused messaging; Segment 0 needs exclusivity/premium messaging.

**Product Strategy:** Develop product lines or bundles specifically to capture more wallet share from Segments 1 and 3.

# CONCLUSION

Based on the Elbow Method analysis, the optimal number of clusters for segmenting this customer data is likely K=3 or K=4.

K=3 provides a strong, simple segmentation into groups that may represent Low-Value, Mid-Value, and High-Value customers.

K=4 offers a slightly finer segmentation, potentially separating High-Value customers into "New" (low recency) and "Loyal" (high recency) groups.


THANK YOU