

CUSTOMER PROFILING USING K-MEANS CLUSTERING FOR RETAIL PRODUCT INSIGHTS



Group Members (Group 3)

- 6722040174
- 6722040224
- 6722040158
- 6722040950

Aye Moh Moh Htwe

Thaw Thaw Hnin Yee Aye

Ma Thi Mai Dinh

U.A.D.Nirosha Lakmali

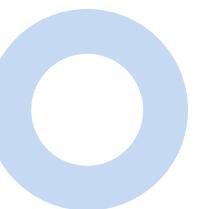


Table of Contents

- Introduction
- Problem Statement
- Objectives
- Data Collection & Data Preprocessing
- Process Flow
- Model Development
- Customer Segmentation
- Managerial Insights & Marketing Strategies
- Conclusion



Introduction



E-retailers struggle to understand varied and evolving customer behaviors



Traditional segmentation ignores real shopping behavior



Inefficient targeting leads to wasted marketing spend and missed sales



K-Means clustering groups customers by actual purchase behavior



Enables data-driven marketing and smarter business decisions

Problem Statement



Poor Customer Insight

Difficulty identifying meaningful differences in behaviour



Low Marketing ROI

Generic campaigns fail to engage key segments



Off-Target Suggestions

Irrelevant product recommendations reduce conversions



Stock Mismatch

Overstock and stockouts due to unclear demand patterns



Objectives

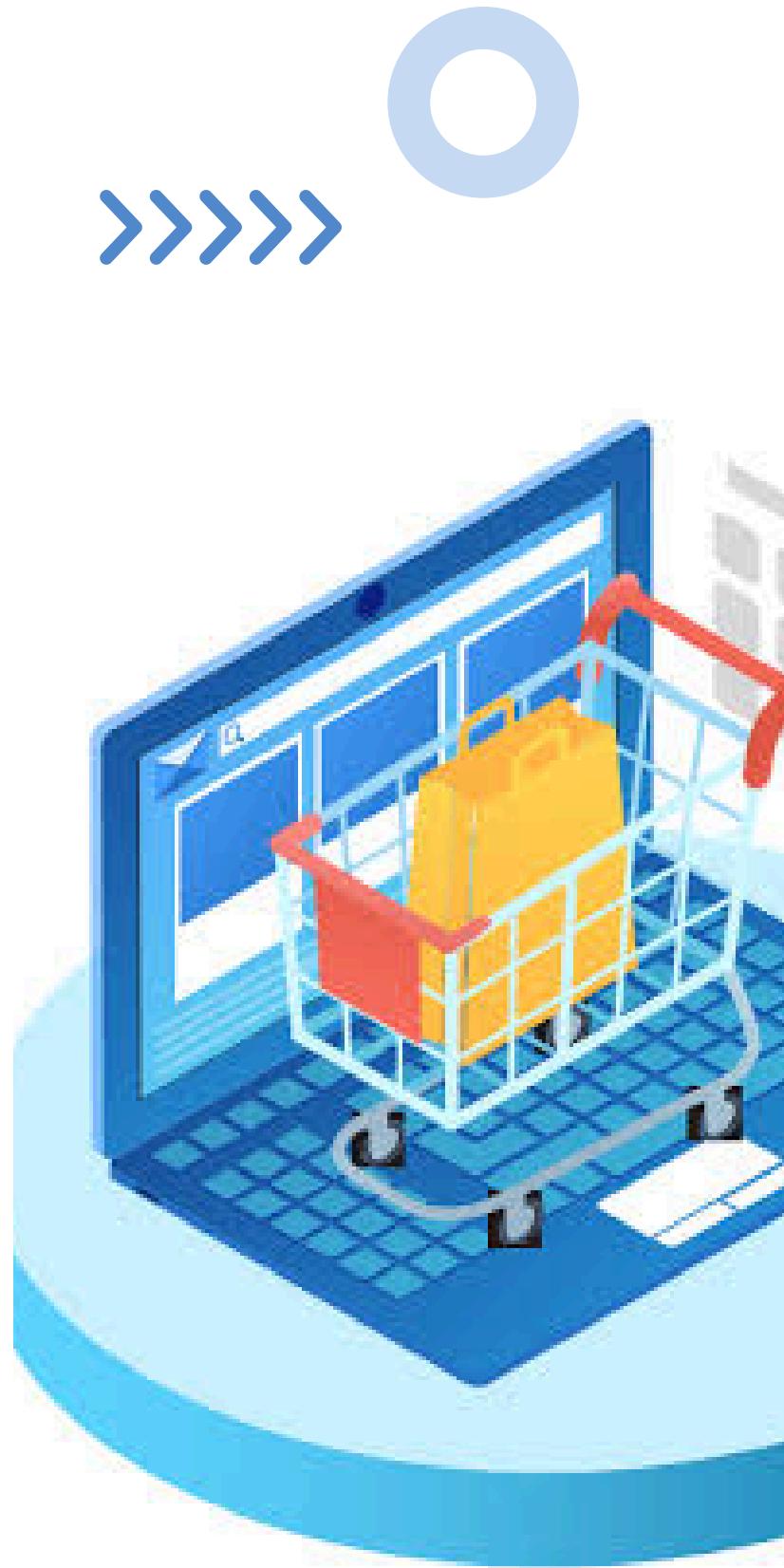
Segment customers by demographics and behavior



Targeted Marketing
Promote the right products to the right groups



Personalized Recommendations
Suggest items based on cluster preferences



Data Collection

 Dataset Source: Kaggle / Business Transaction Records

 Data that we included in the model

Category name

Price

Product name

Payment method

Quantity

Review score

Gender

Age



Data that we excluded in the model

Customer ID

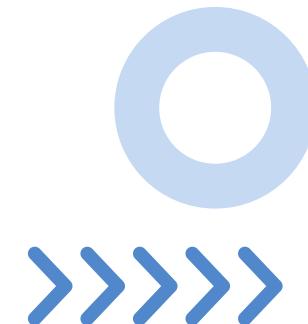
Order date

Product ID

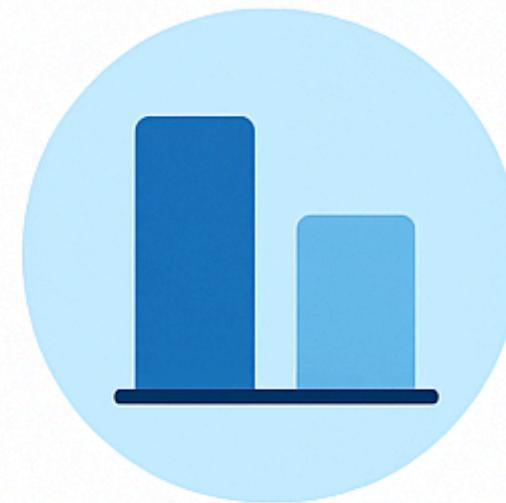
Category ID



Data Preprocessing



Removed
irrelevant fields
(Customer ID,
Order Date,
Product ID,
Category ID)



Split data:
80% training,
20% testing

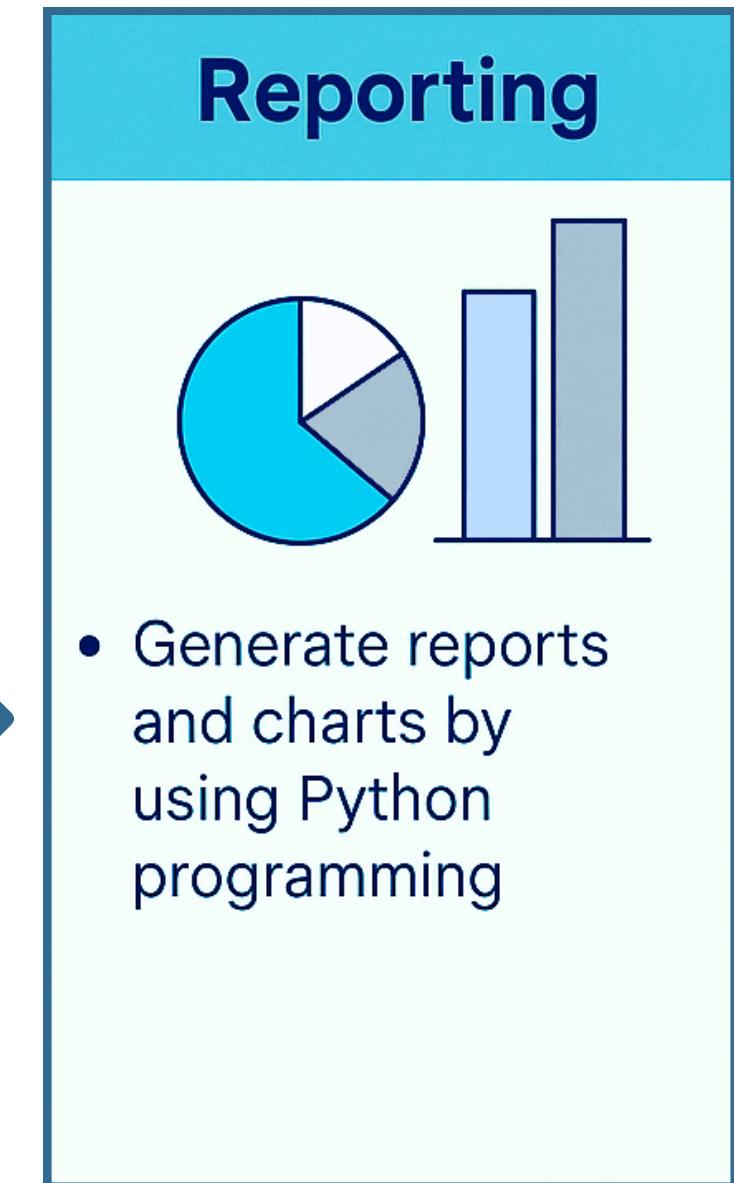
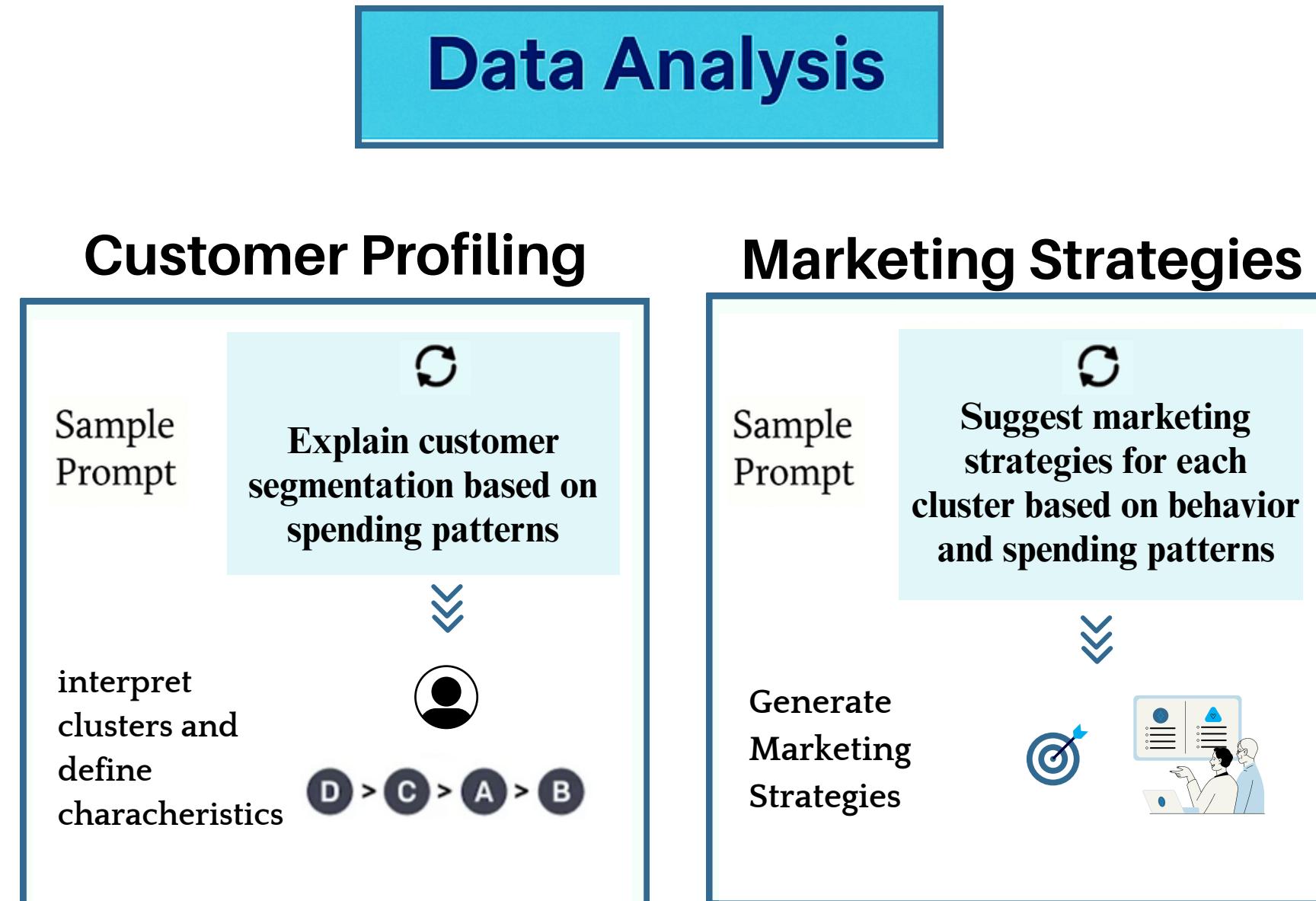
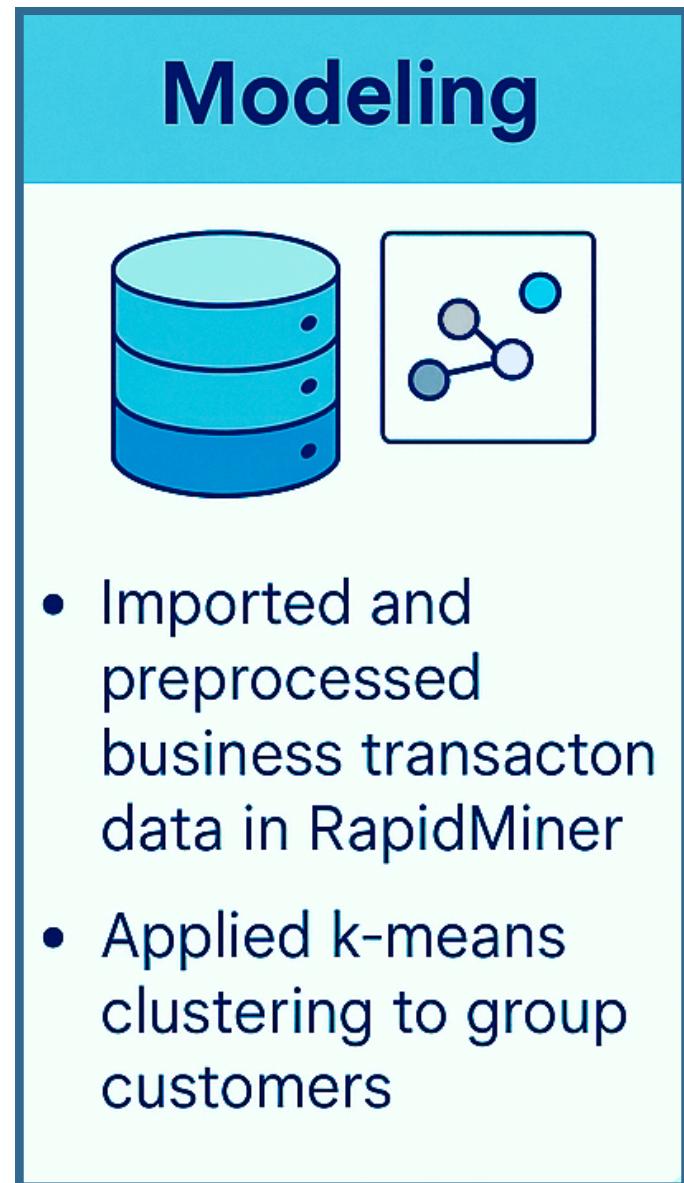


Filled missing
values
(Gender,
Review Score)
(Gender, Payment
Method, Product
Name)

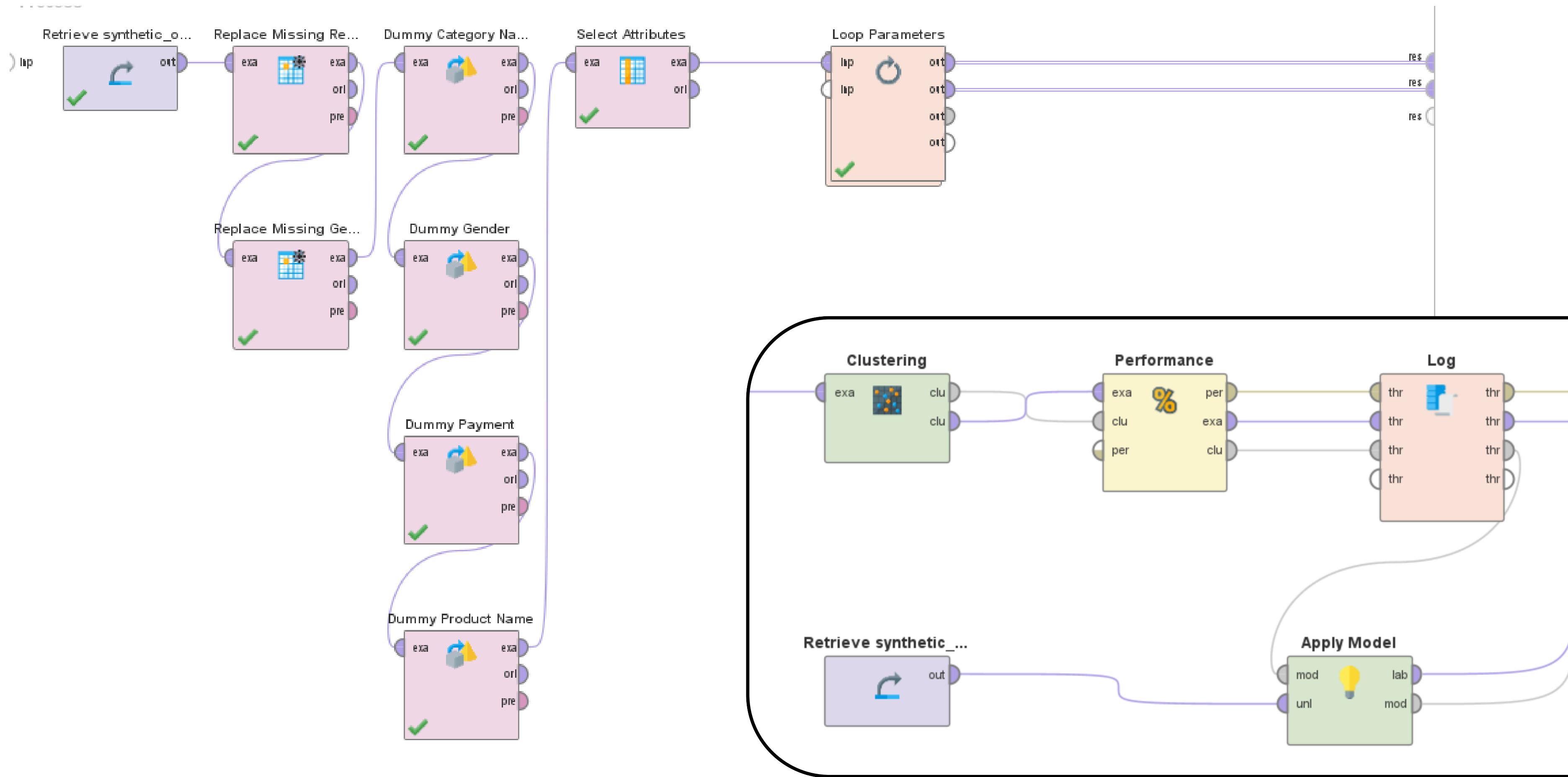


Encoded
categorical
variables
(Gender, Payment
Method, Product
Name)

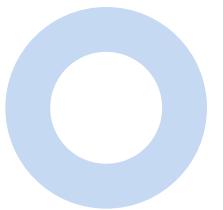
Process Flow



Model Development



Elbow Point



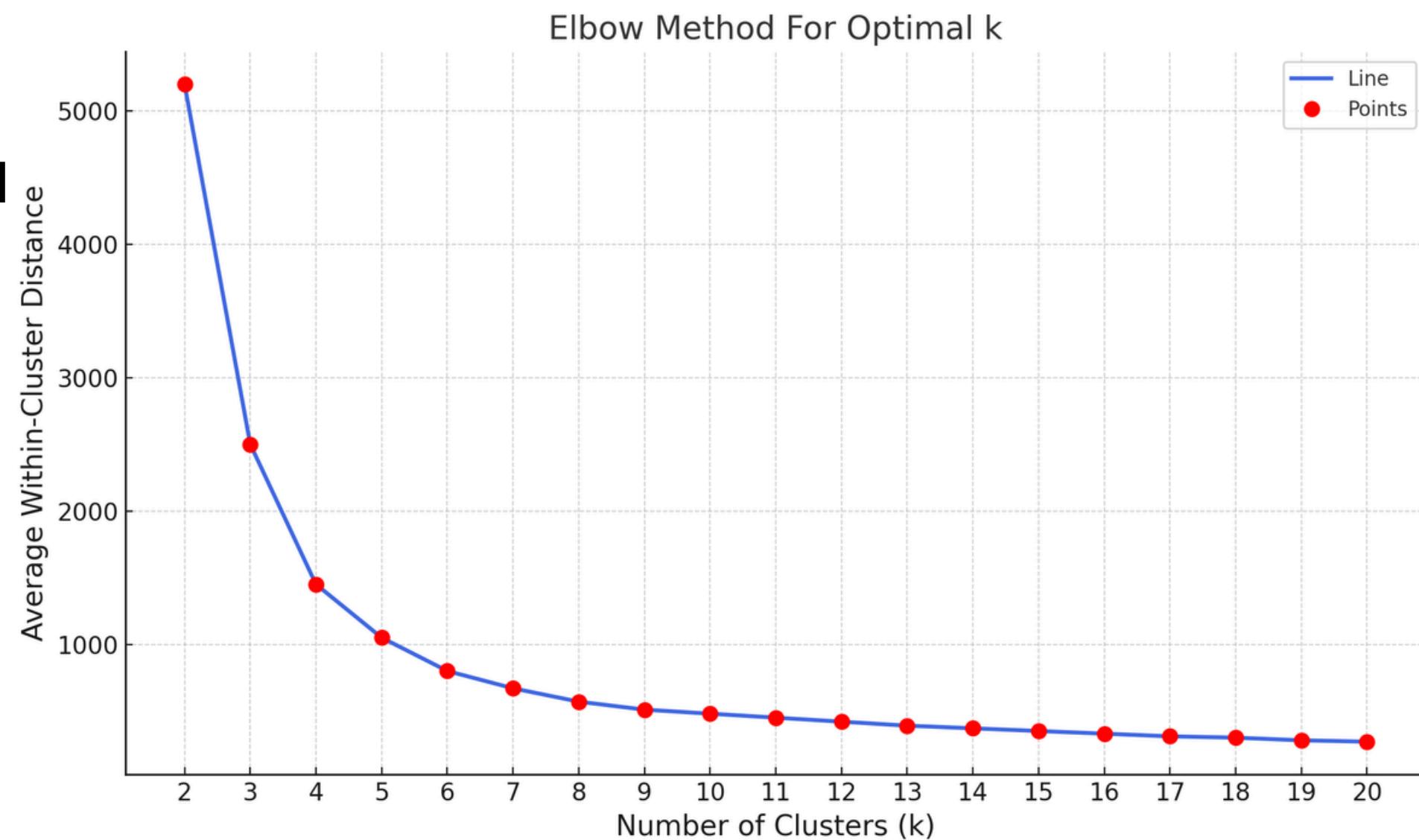
Accuracy
Clusters are meaningfully separated



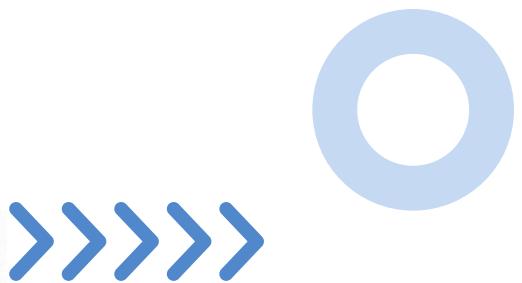
Simplicity
The model remains easy to interpret

Elbow point at $k = 5$

This is the optimal value because it
offers a balance between Accuracy
and Simplicity



Customer Segmentation

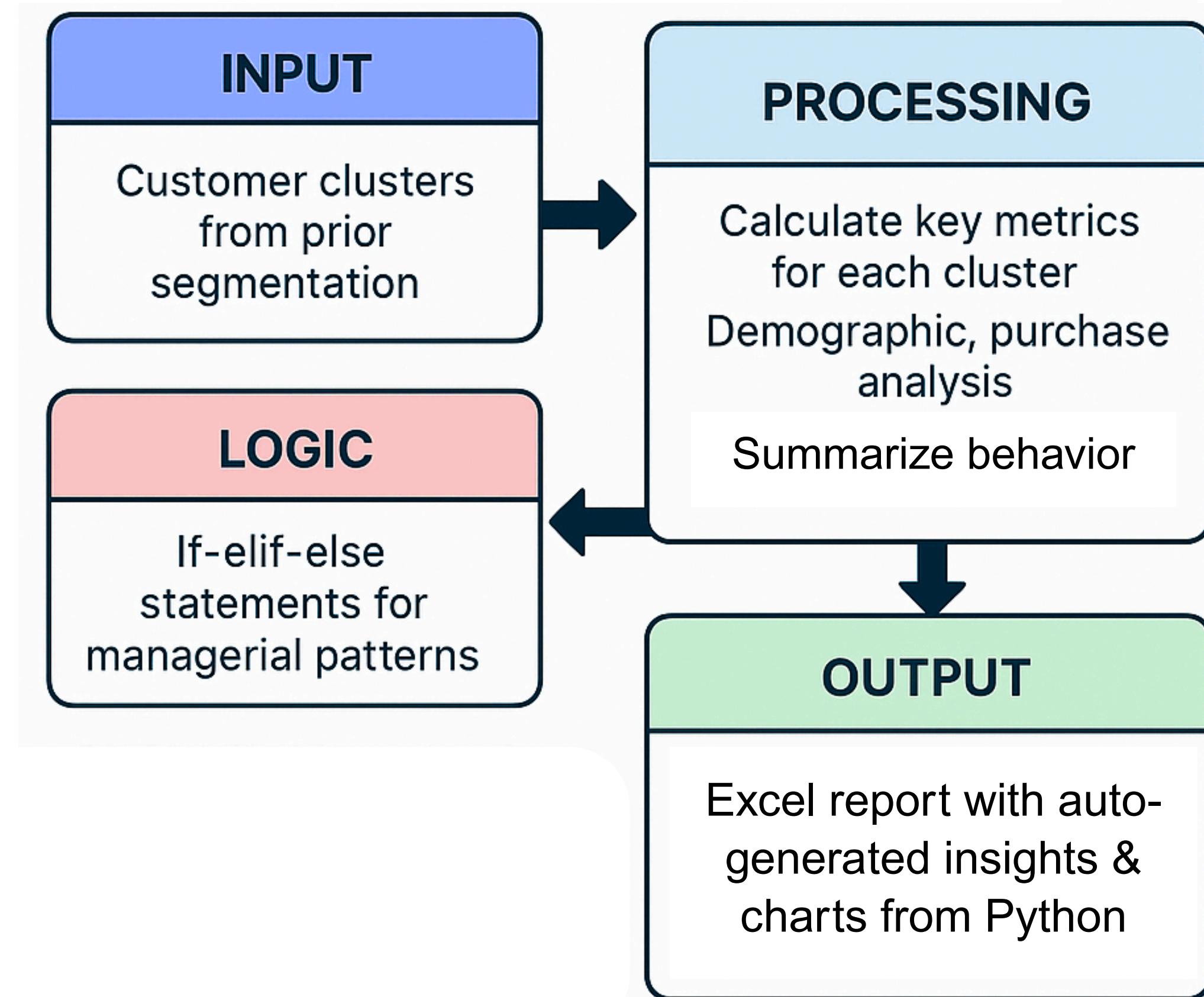


Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gadget Lovers	Outdoor Enthusiasts	Luxury Buyers	Bookish Buyers	Fashionistas
22.4% electronics	16.8% Sports & outdoors	20.4% Home & living	20.4% Book & Stationery	20.1% Fashion
180	134	163	163	160
Total Spend	Total Spend:	Total Spend	Total Spend	Total Spend
\$ 123222 AVG \$674	\$ 123833 AVG \$924	\$ 102853 AVG \$631	\$ 117503 AVG \$721	\$ 130349 AVG \$815

Automated Reporting and Visualization from Clustered Data using Python

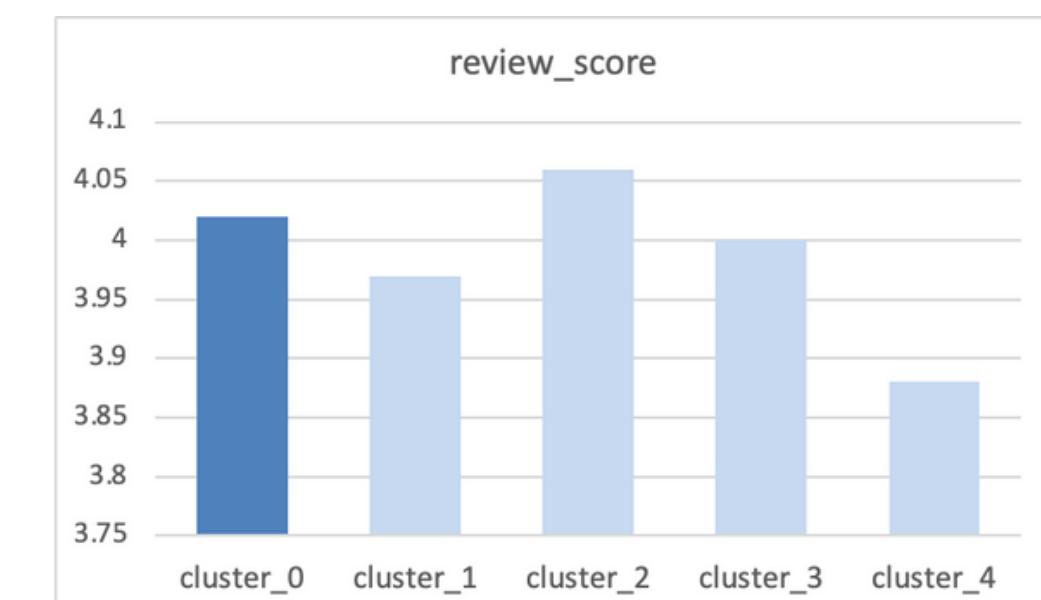
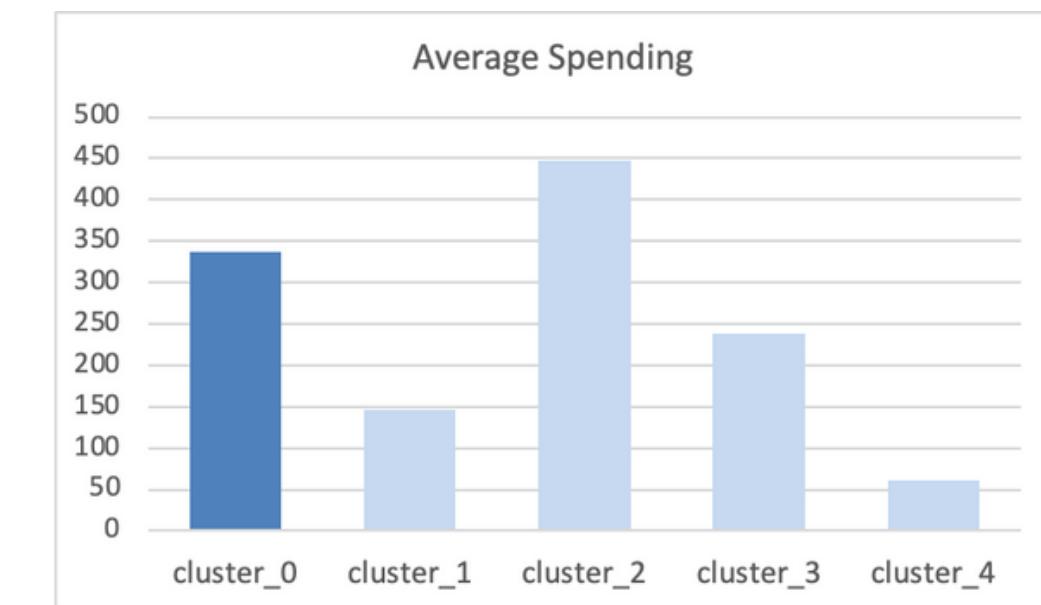
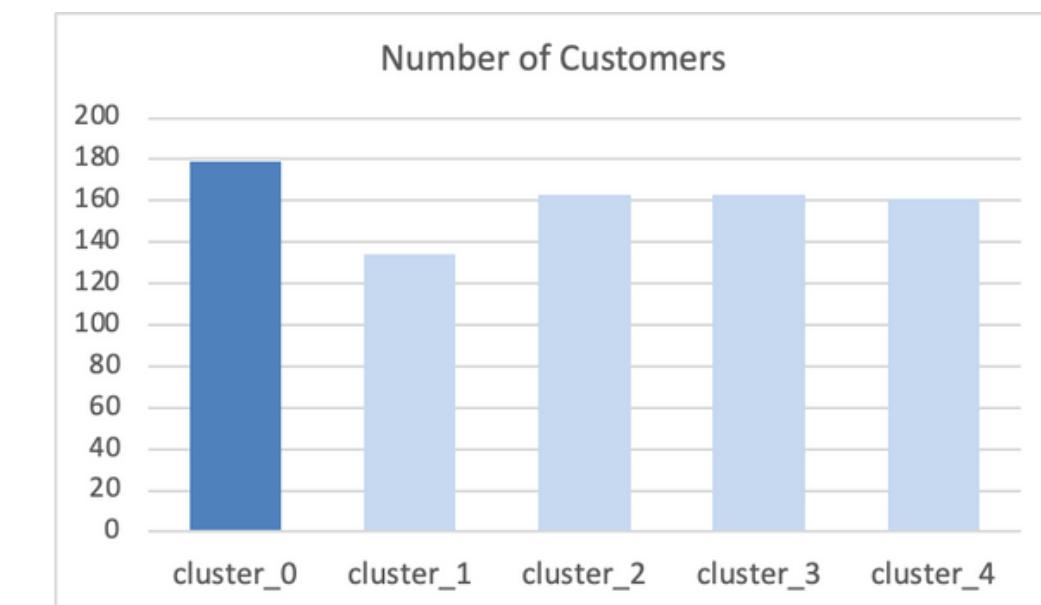


>>>>

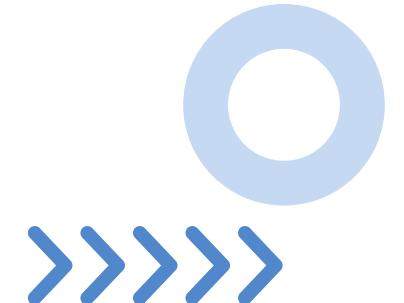


Sample of Automated Reports and Recommendation Generated from Python

Cluster cluster_0 consists of **179 customers** with an **average age of 44.23**. They typically purchase **3 units per order** at an **average price of \$337.44 per item**. Their average **review score is 4.02 out of 5**. The pricing behavior of this group is moderate compared to others. Their satisfaction level is average compared to other segments. It is also the largest cluster, making it a key target for broader marketing efforts. **Frequently bought products include: Yoga Mat, Soccer Ball, Eraser.**



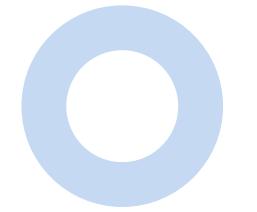
Recommended actions: Use mass marketing and wide-reaching campaigns. Ensure popular items are well-stocked: Yoga Mat, Soccer Ball, Eraser.



Managerial Insights and Marketing Strategies

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
				
Gadget Lovers	Outdoor Enthusiasts	Luxury Buyers	Bookish Buyers	Fashionistas
electronics	Sport and Outdoors	Home and living	Book & Stationery	Fashion
 22.4%	 16.8%	 20.4%	 20.4%	 20.1%
 \$337	 \$146	 \$447	 \$237	 \$60
 4.02	 3.97	 4.06	 4	 3.88
Use Mass Marketing & Loyalty Programs	Re-engage with offers and combos	Focus on premium upselling	Try simple offers & Keep monitoring this group	Introduce Discounts to elevate satisfaction

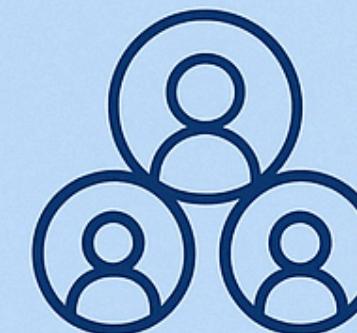
Conclusion



Applied K-Means in RapidMiner to segment customers by age, spending, and behavior



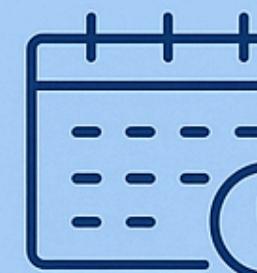
Used the Elbow Method to find the optimal number of clusters



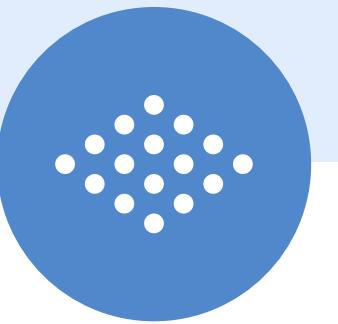
Identified distinct customer groups with unique preferences



Enabled personalized recommendations for better decisions



Future work:
Incorporate seasonal trends and time-based analysis



**THANK
YOU!**




```
# Recommendations
recommendation = "\nRecommended actions: "
if row['price'] == max_price:
    recommendation += "Focus on premium upselling, exclusive bundles, and early access programs."
elif row['price'] == min_price:
    recommendation += "Introduce discounts, value packs, and loyalty rewards."
if row['review_score'] == min_review:
    recommendation += "Improve service or product quality based on feedback."
if row['count'] == max_size:
    recommendation += "Use mass marketing and wide-reaching campaigns."
if row['count'] == min_size:
    recommendation += "Apply personalized outreach to grow or re-engage this group."

# Fallback
if recommendation.strip() == "Recommended actions:":
    recommendation += "Keep monitoring this group and try simple offers to see what works best."

# Add inventory recommendation
recommendation += f" Ensure popular items are well-stocked: {', '.join(top_items)}."

# Combine
insights.append(note + recommendation)
```