

R

Delivering Fresh

WHERE ANALYTICS MEETS ORGANIC
CULINARY BLISS

Team 17 @ Communitas

Rotman
Here's where it changes.



Communitas

Business Case

Who We Are:

- Communitas, a meal kit delivery company

Our Vision:

- Provide fast, easy, yet organic and healthy food at doorsteps

Our Mission:

- Predict median household income per region
 - Target customer segments with medium-high income
 - Adjust marketing strategies and boost profits



Analytics-driven / Differentiation / Organic Food / Scalable

Rotman

Preliminary Analysis - Variables

Remove

The following three variables are the same, so we only kept the first one:

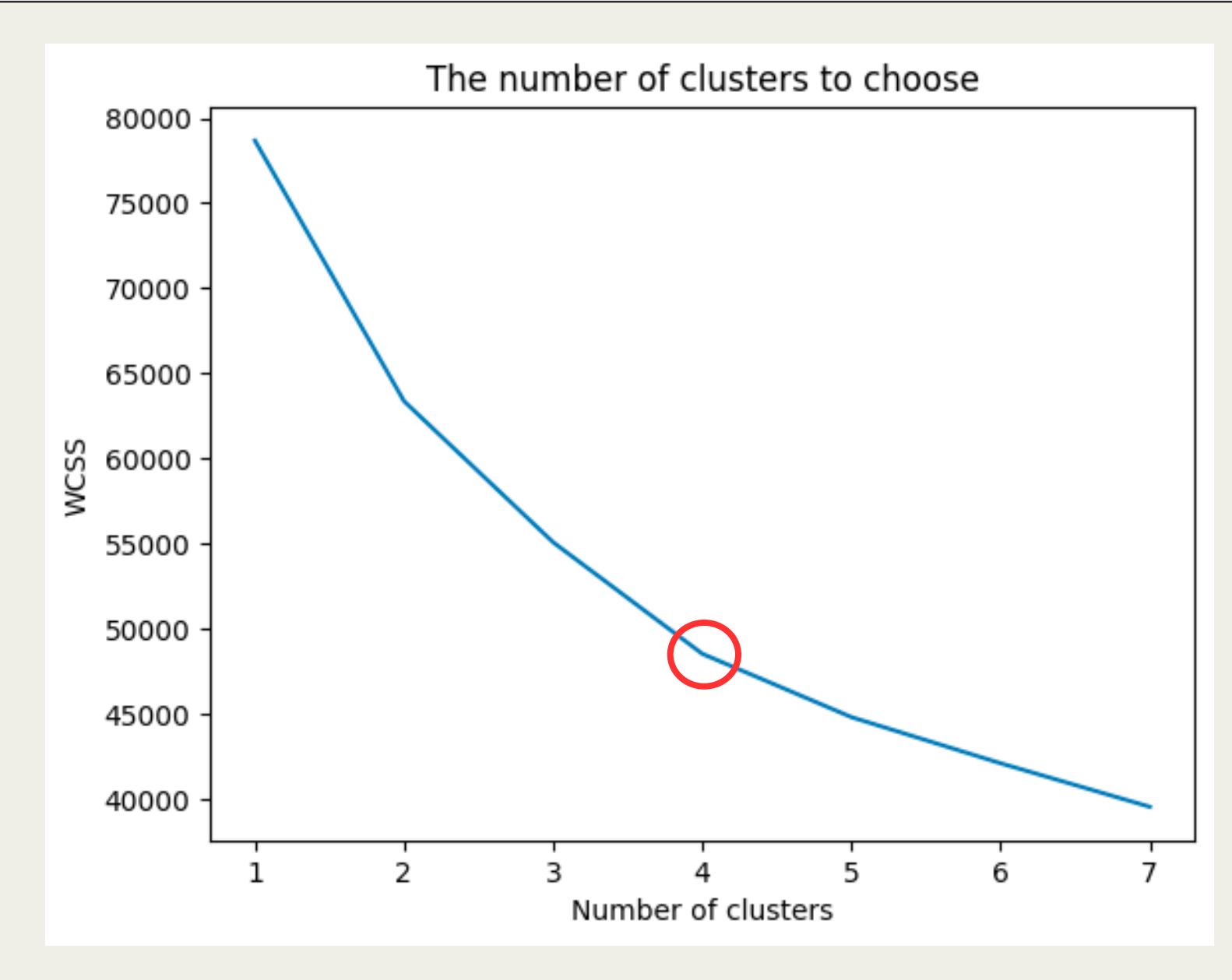
- **Basics | Total Households**
- Occupied Private Dwellings by Period of Construction | Total Households For Period Of Construction
- Occupied Private Dwellings by Tenure | Total Households for Tenure

Rename

To have a better sense of the history of households, we renamed the variables with years built as follows:

- Occupied Private Dwellings by Period of Construction | Total Households For Period Of Construction | Built Before 1961
- This is now **Total Households Aged > 63**

K and Clusters



Why Clustering?

- Efficient Customer Segmentation
- Strategic Grouping
- Business Alignment

The optimal number of clusters is 4.

K and Clusters

0

Cluster 0: Economically diverse or lower-income, densely populated areas.

1

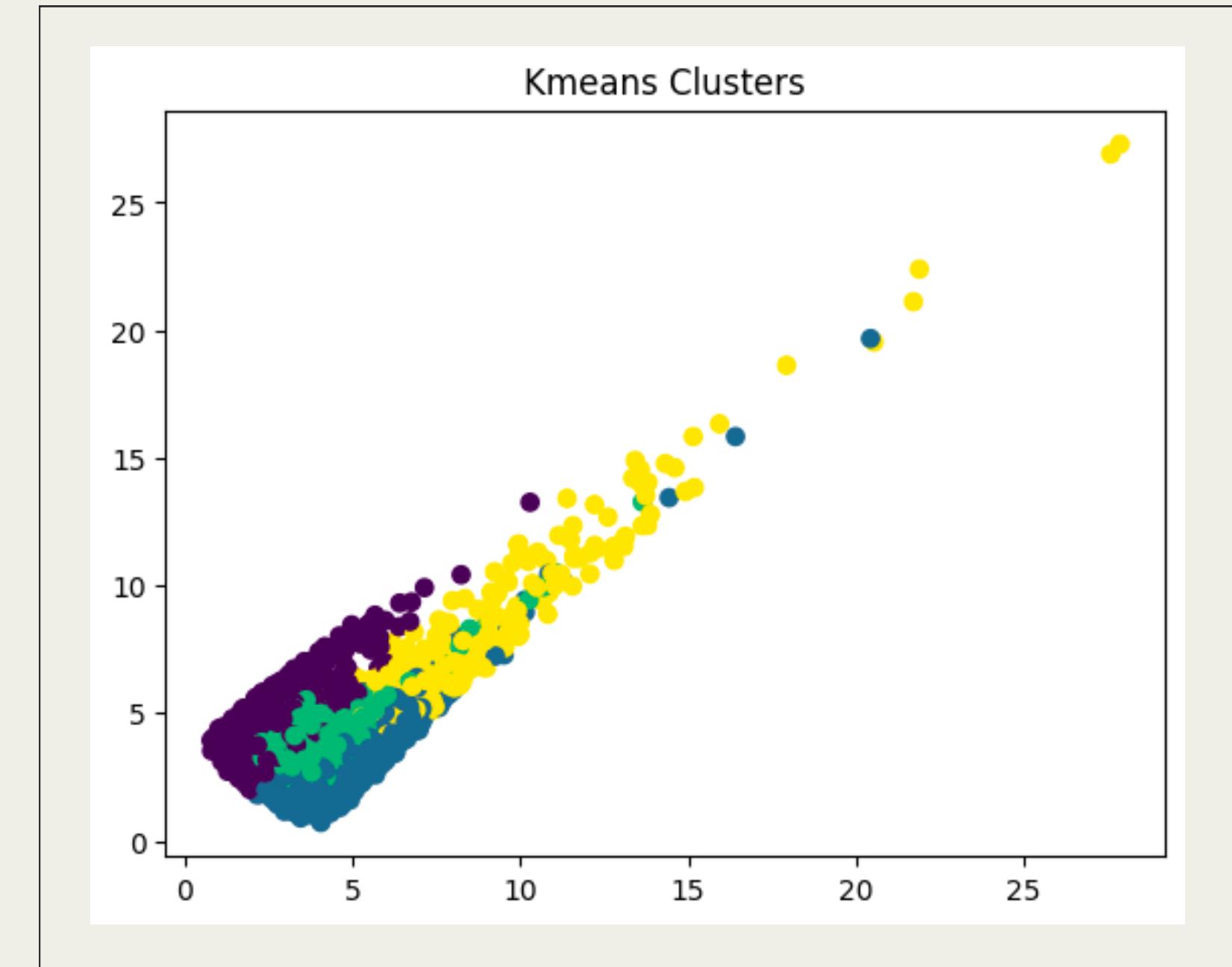
Cluster 1: Affluent, densely populated areas with higher incomes.

2

Cluster 2: Sparsely populated (suburban or rural areas) with middle-range incomes.

3

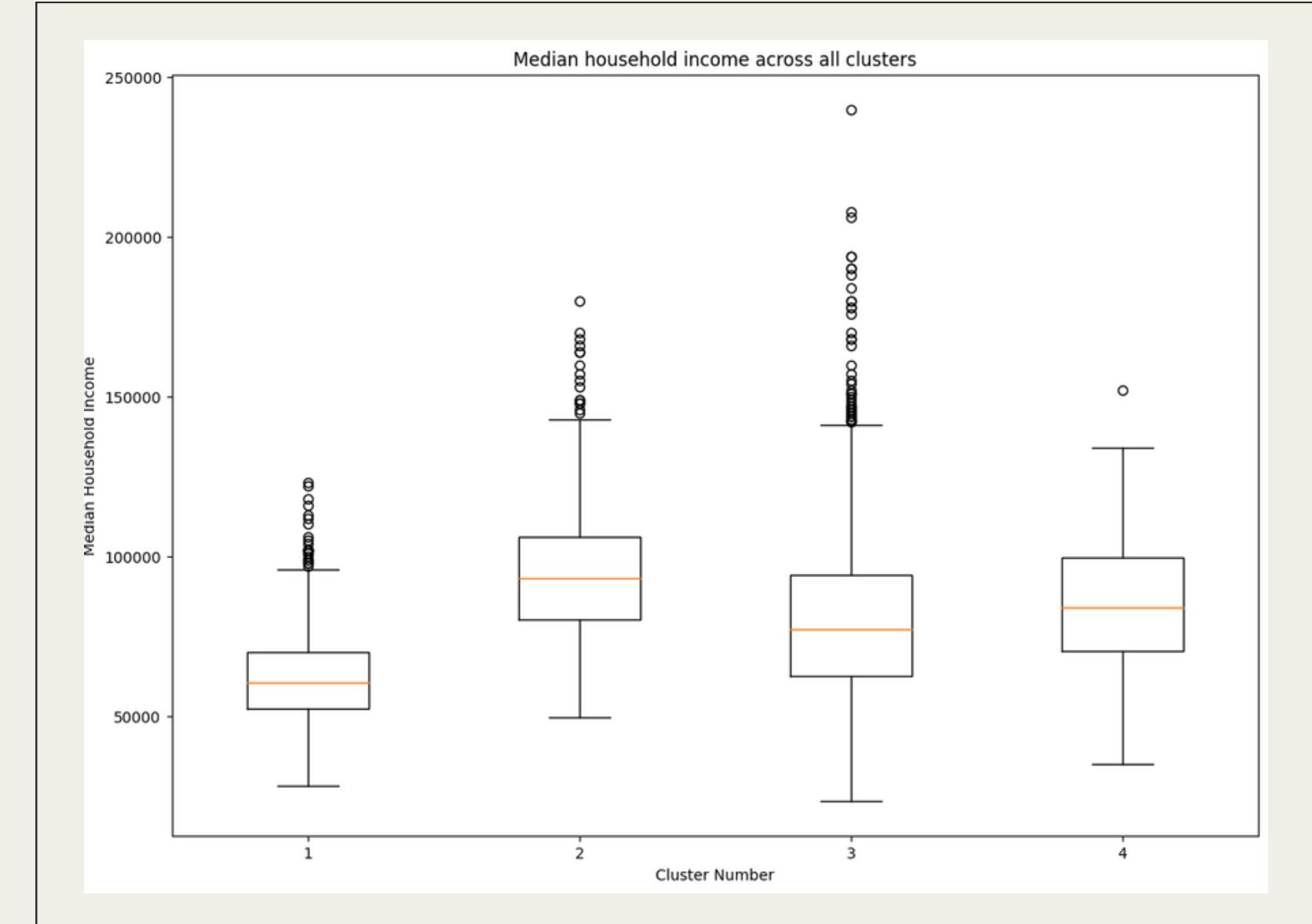
Cluster 3: Highly urbanized with the wealthy households and large population.

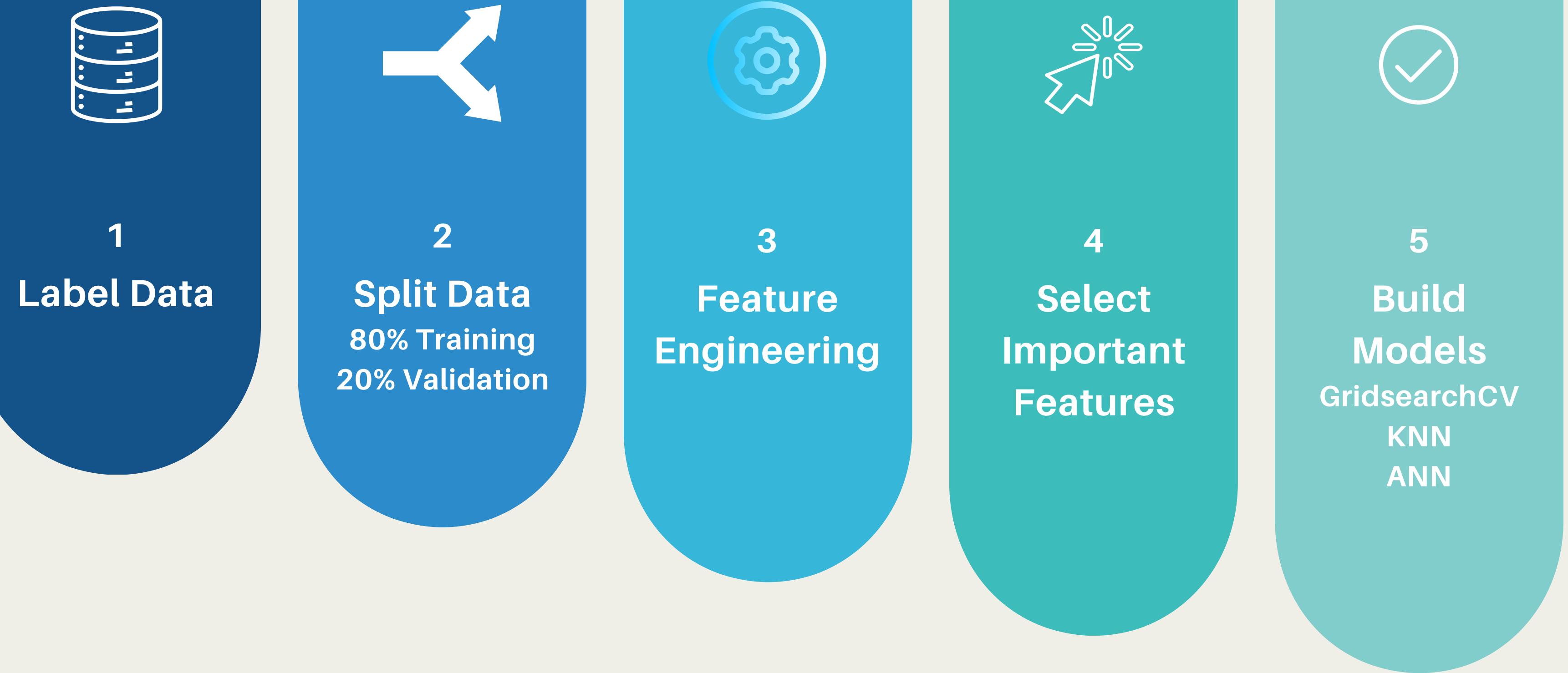


What do the groups signify?

Initial visual analysis depicts:

- Group 2 and 3 - Ideal target customers, consistent income households
- Group 4 - Prime candidate for premium line products.
- Group 1 - Broad income market, needs varied marketing to cater different income level





DATA ANALYSIS AND MODELLING PROCESS

Rotman

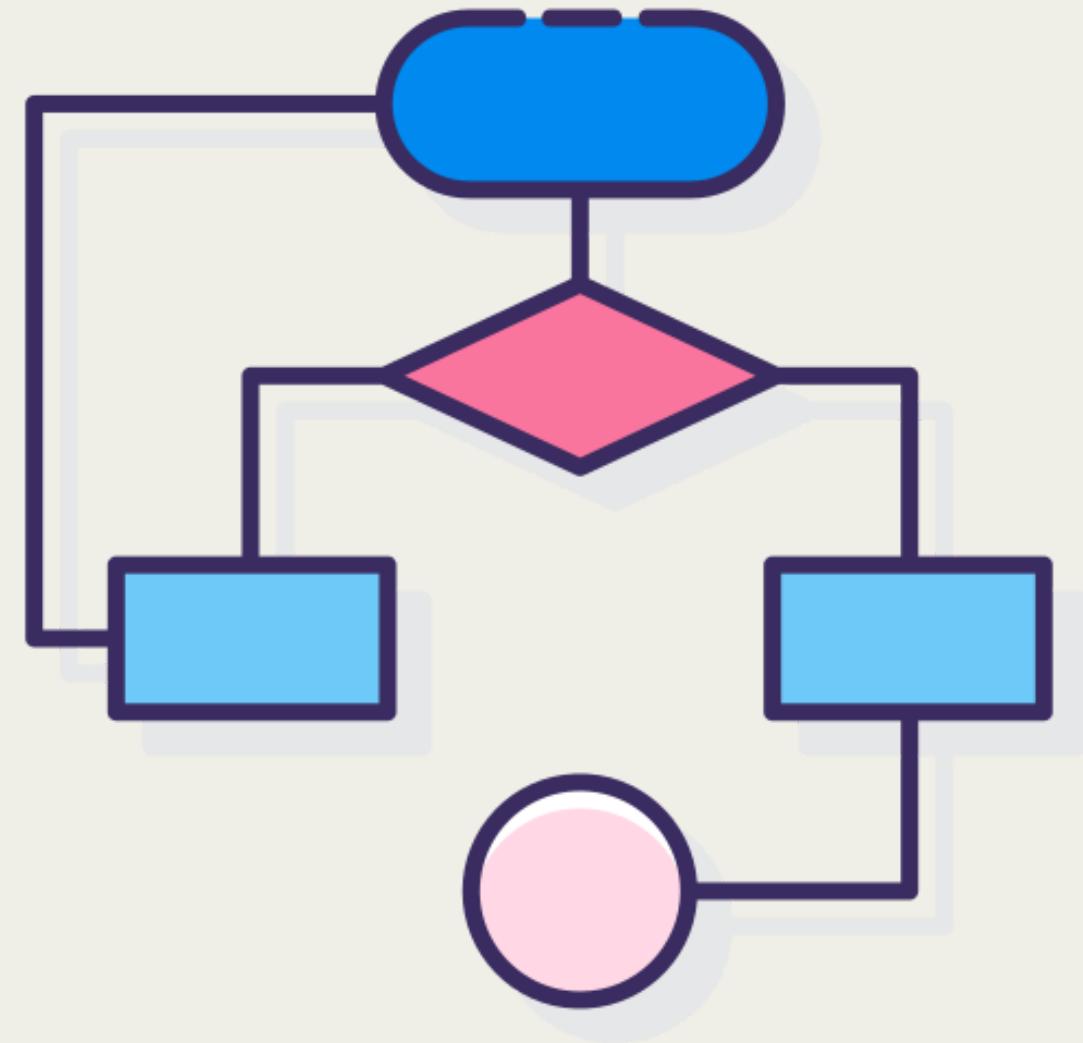
Algorithm Breakdown

- Feature Engineering

- **Decision Tree Regressor:** “maze-running” simulation; making choices at intersections
- **Lasso Regression:** “sculpting” from a raw block of marble to a furnished piece of art

- Prediction Models

- **K-Nearest Neighbors:** examining the information from the closest “neighbors”
- **Artificial Neural Network:** human brain functions; working together to solve a mystery



Evaluation Methods

- **Mean Squared Error:** throwing “darts” at a board; landing some distance from the bullseye
 - Measuring how far off the prediction result is from the actual, observed data statistic
 - Lower MSE → “darts” closer to the bullseye
- **R-squared:** a “report card” for the model
 - Prediction results vs. simple average guess
 - Closer to 100% → better the model



Model Evaluation (Clustering)

Results on validation data set

| Cluster 0 Model | Mean Square Error | R-squared | Cluster 1 Model | Mean Square Error | R-squared |
|-----------------|-------------------|-----------|-----------------|-------------------|-----------|
| KNN | 0.014 | 41.27% | KNN | 0.012 | 47.72% |
| Lasso | 0.011 | 53.84% | Lasso | 0.009 | 58.98% |
| Cluster 2 Model | Mean Square Error | R-squared | Cluster 3 Model | Mean Square Error | R-squared |
| KNN | 0.008 | 48.28% | KNN | 0.019 | 48.82 |
| Lasso | 0.007 | 56.23% | Lasso | 0.020 | 44.50% |

Model Evaluation (Without Clustering)

After Hyperparameter Tuning

| Training | Mean Square Error | R-squared |
|----------|-------------------|-----------|
| ANN | 0.0029 | 74.56% |
| KNN | 0.0033 | 70.97% |
| Testing | Mean Square Error | R-squared |
| ANN | 0.0039 | 69.56% |
| KNN | 0.0049 | 61.95% |

Limitations & Future Study

Small Dataset

01

Insufficient amount of data may negatively affect model training process.

Improvement:

- K-Fold Cross Validation
($k = 3 \rightarrow k=10$)

Model Selection

02

Add ANN model for each cluster

Pros:

- Capture complex patterns
- Better generalization

Cons:

- Computational cost

Model Accuracy

03

The use of segmentation decreases model accuracy

Improvement:

- Other clustering approach (ex. BIRCH)

Conclusions

01

02

03

04

05

*Fundamental
role*

- Intelligently segmenting our customers
- Meet expectations of customers

Segmentation

- Intelligently segmenting our customers

Strategy

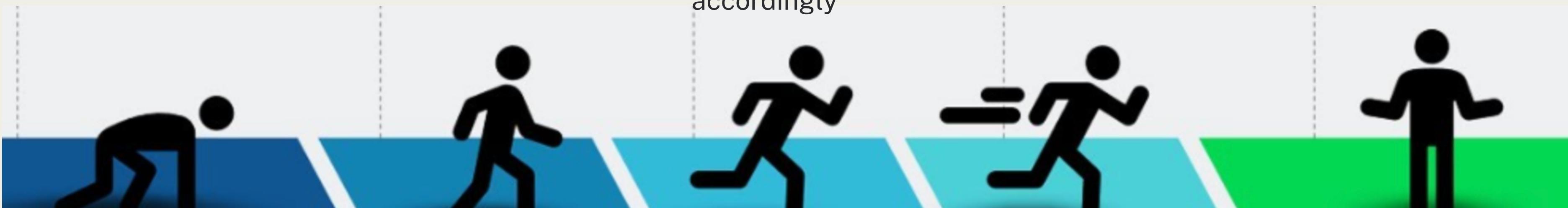
- Tailoring our services and marketing strategies accordingly

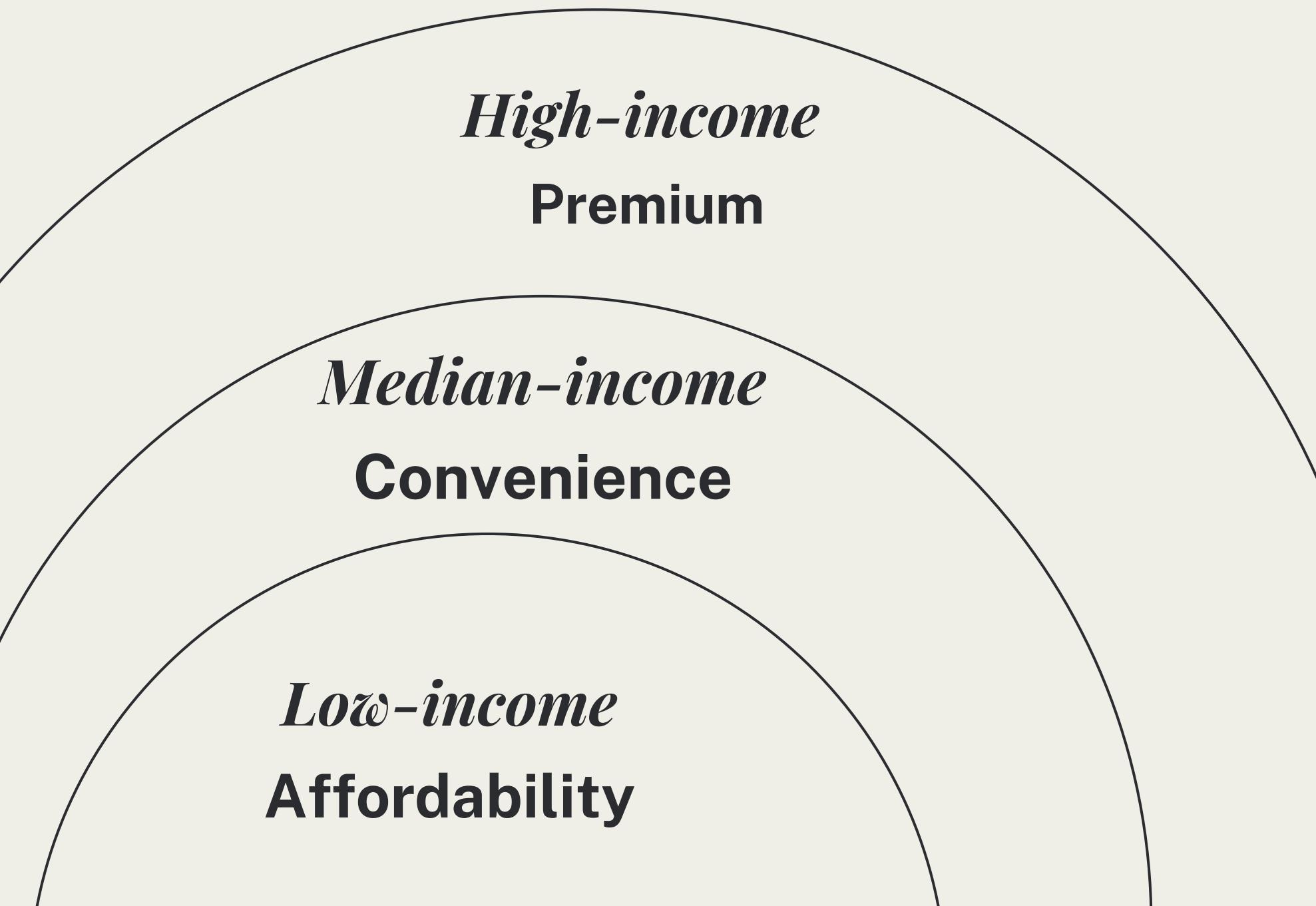
Surpass

- Exceed customers' expectation

Future

- Leadership in the healthy food delivery market





Dynamic Customer Preferences

Customer preferences change rapidly.

- Keep up with these trends and adapt to the segmentation strategy accordingly

Income-Based Segmentation Risks

Income level doesn't directly correlate with spending preferences or priorities.

Customer Retention

Recognize that customers from different income brackets have varied expectations and preferences.

- Maintain open channels of communication



R

Thank you!

FAST, EASY, AND ORGANIC - AT YOUR DOORSTEP

Phyllis, Jack, Jiaming, Lead Data Scientists

Rishi, Dorothy, Senior Consultants

Sydney, Outreach Intern

@ Communitas

November 28, 2023

Communitas 