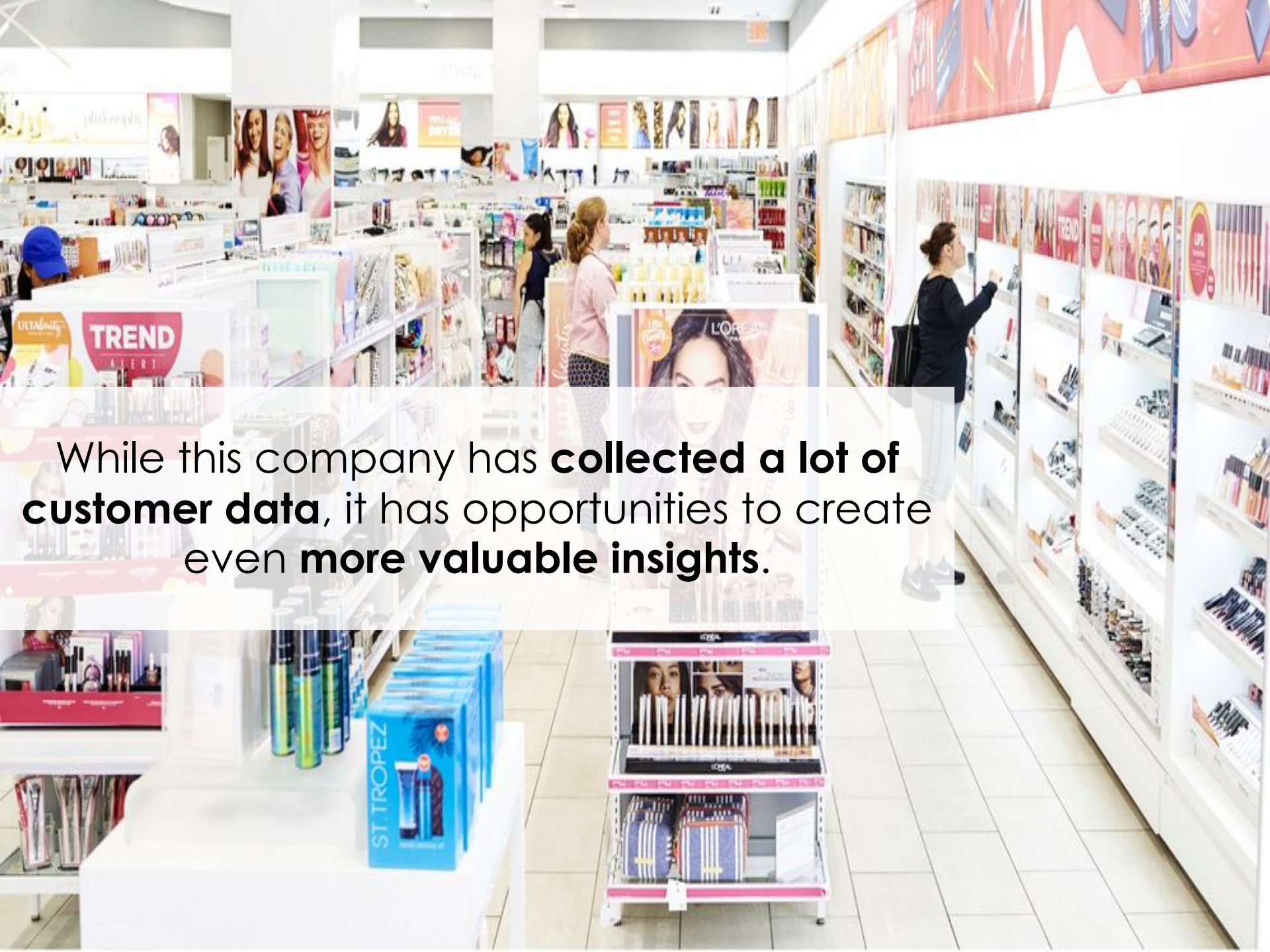
Using Data to Empower Strategic Growth

Shehryar Khursheed | Abdul Sohu | Fatih Ozer | Sardar Cheema

Barings' portfolio company is a **global retail company** that sells a variety of SKU's to a **wide customer base**.

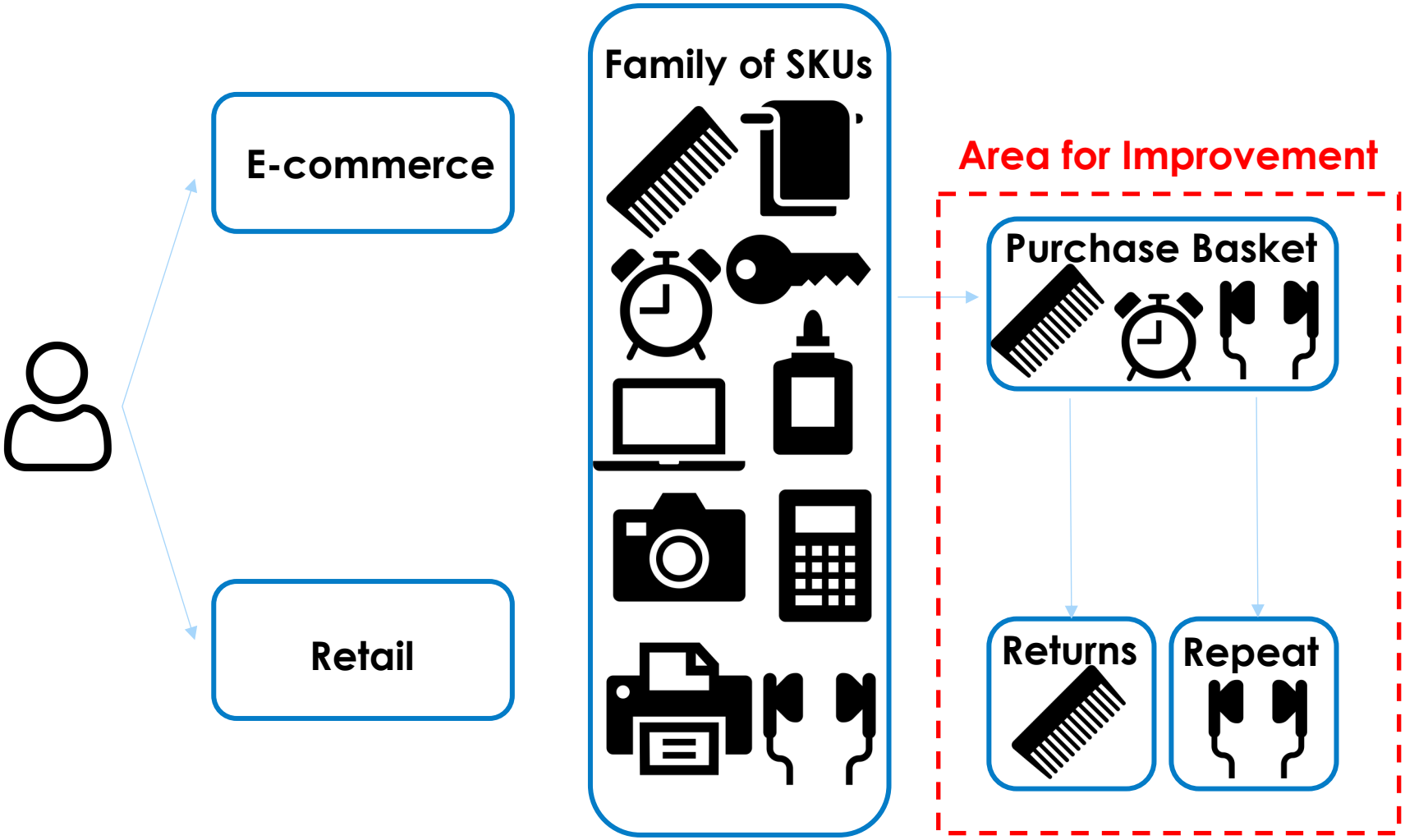


While this company has **collected a lot of customer data**, it has opportunities to create even **more valuable insights**.


We need to **leverage these insights** to position this company for **long-term strategic growth**.

The Customer Journey

There are three key areas of improvement: increasing transaction value, discouraging returns and increasing repeat shopping.

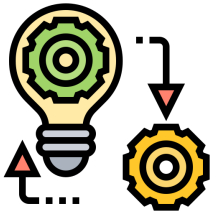


Retail Market

A successful solution should address some key challenges.

Using data to upsell and cross-sell products

Creating a customer experience that discourages returns



Identifying and focusing on the most valuable customers

How can we effectively **leverage data** to **drive growth** both offline and online, and **enhance the customer experience**?

Objectives

We aim to leverage data to segment customers, target advertising and increase basket size.

01

Drive Transaction Value

Use customer demographics to identify which products customers are most likely to purchase given their basket.

02

Discourage Returns

Figure out why people return and create proprietary scoring system for targeted promotions.

03

Encourage Repeat Shopping

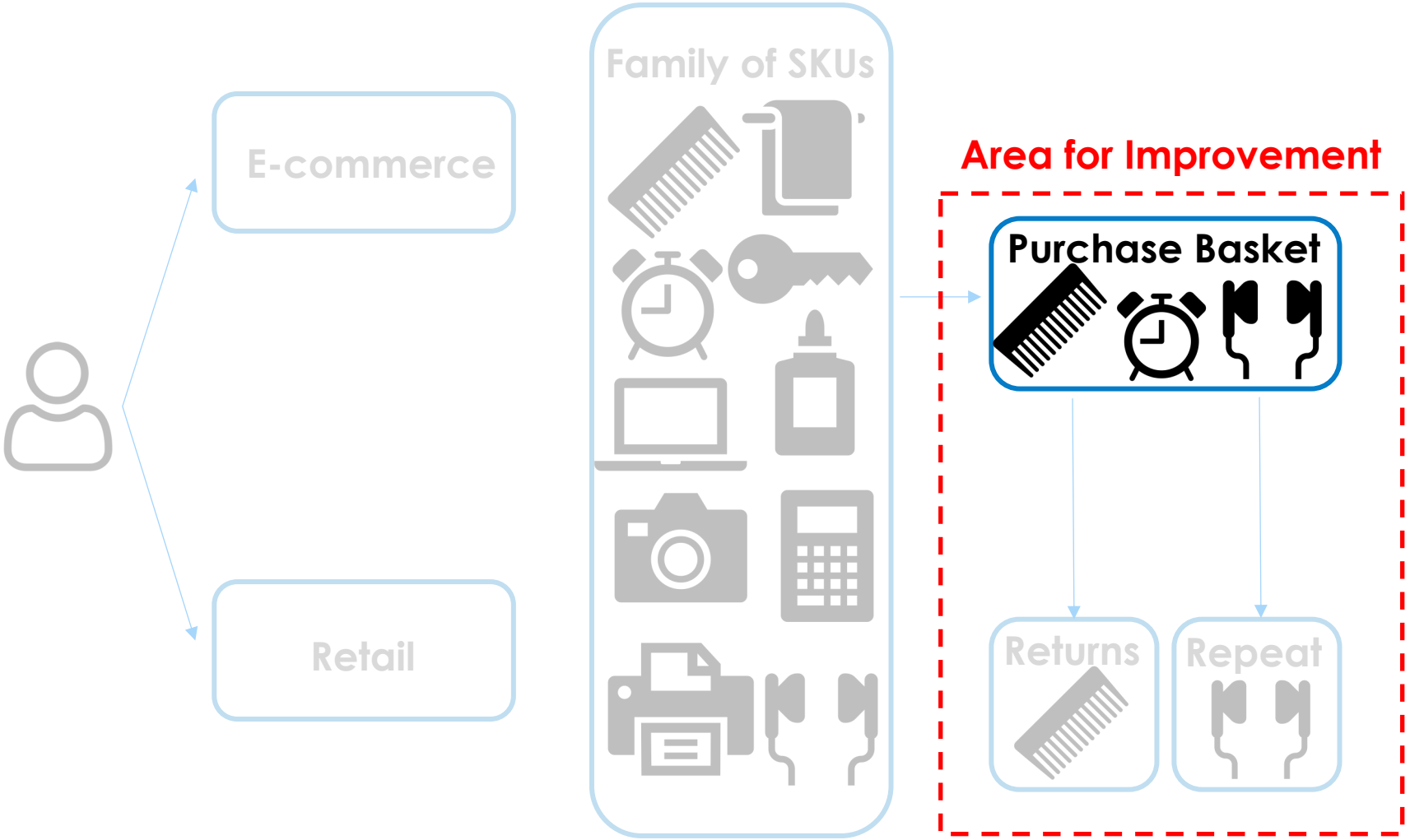
Targeting marketing to customers most likely to generate repeat business.

KPIs

- Transaction Value
- Customer Loyalty
- Return Rate

Phase 1: Enlarge the Basket

There is a significant opportunity to increase the average transaction size of customers.



Phase 1: Enlarge the Basket - Issue

Currently, there are no ways to target advertising.

Current Situation

All customers are treated similarly as one homogenous group.

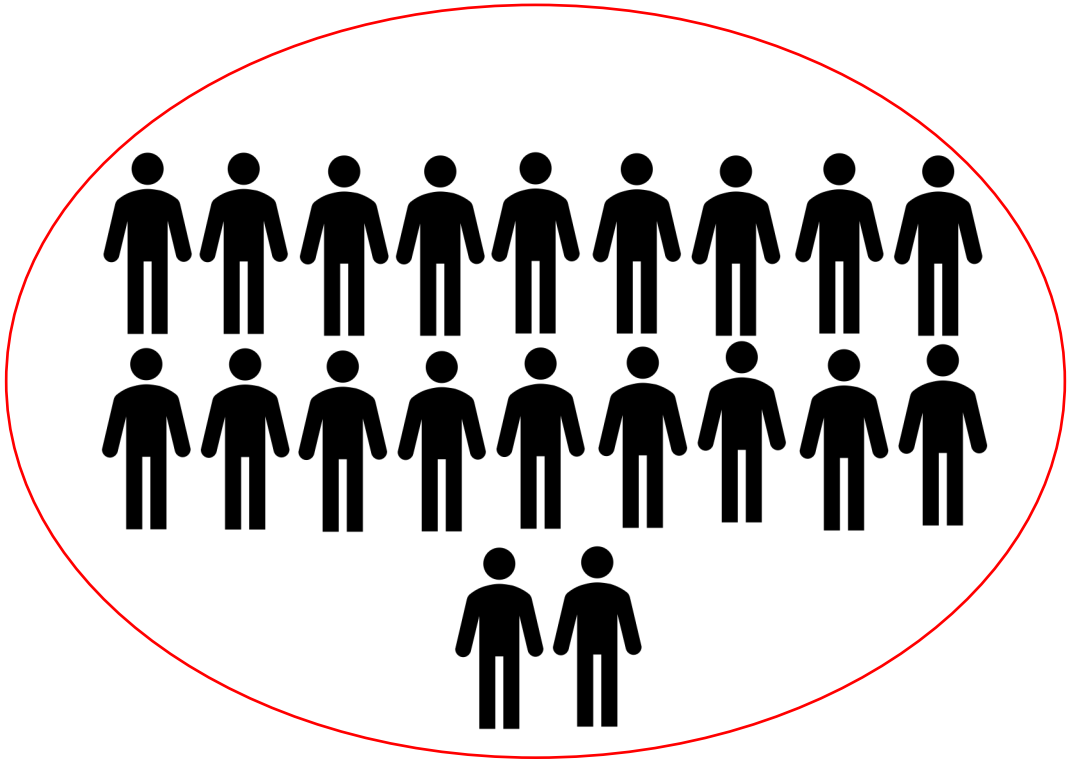
Selective marketing is not possible due to no customer segmentation.

RFM score is not a meaningful predictor of customer behavior.

As a result, marketing efforts are not as effective as they can be.

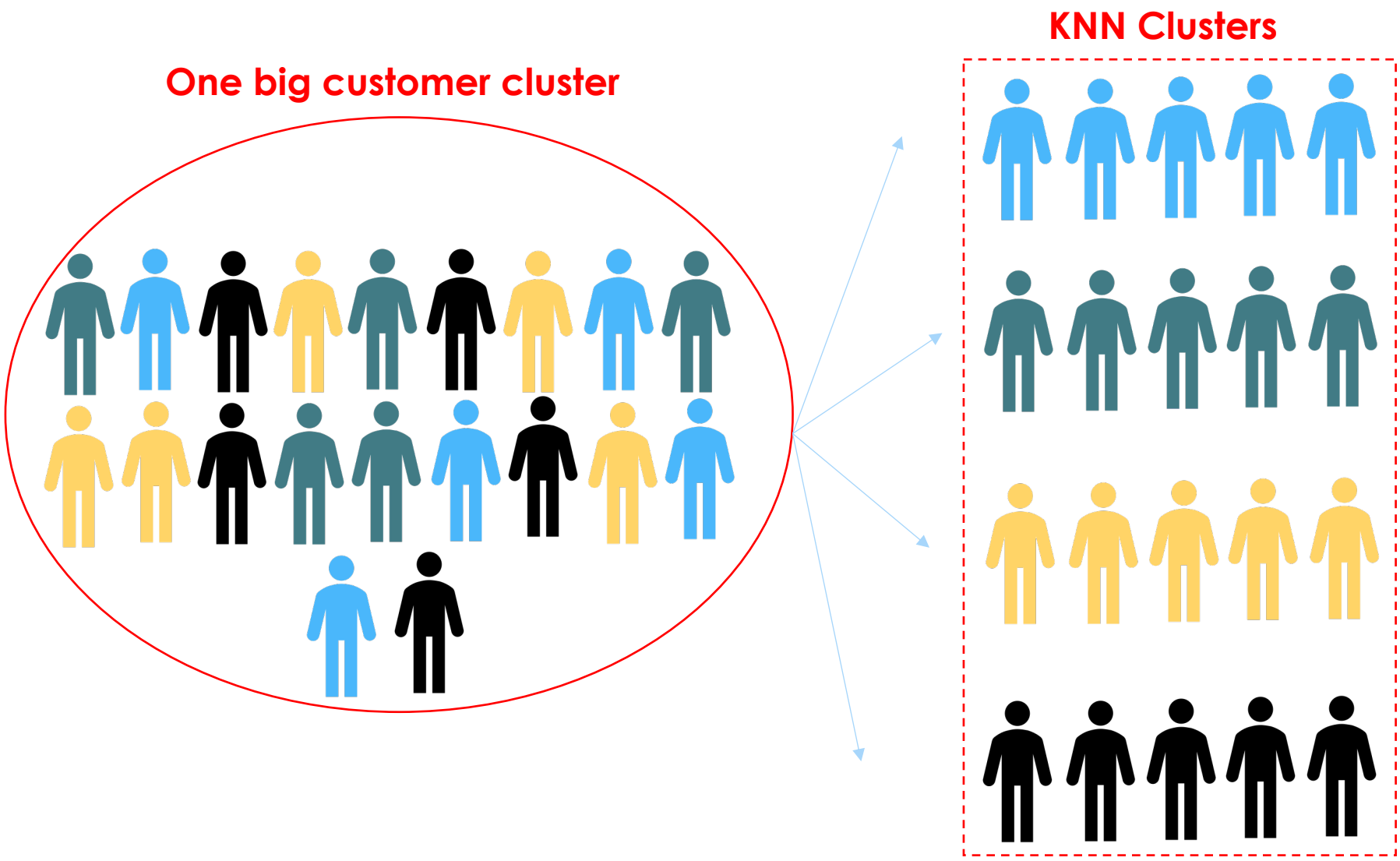
We need a way to figure out which customer is interested in **<u>what product category, and when</u>**.

One big customer cluster

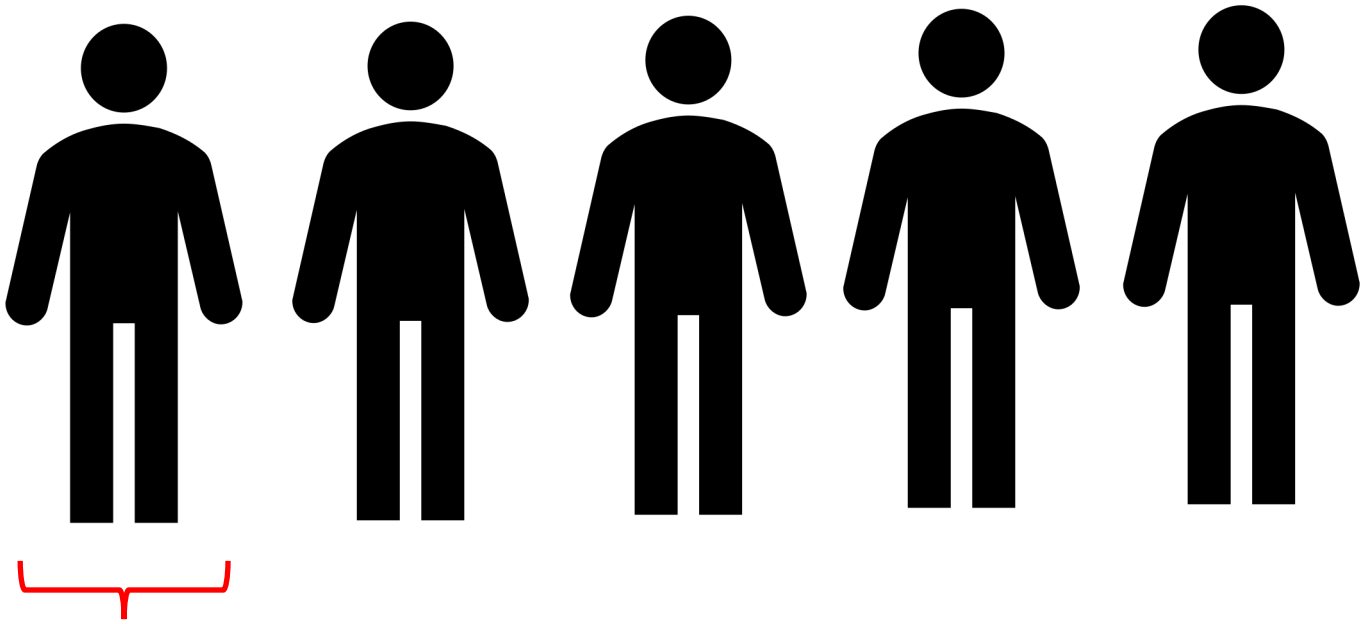


Who Prefers What?

Phase 1: Enlarge the Basket - Method



Phase 1: Enlarge the Basket - Method



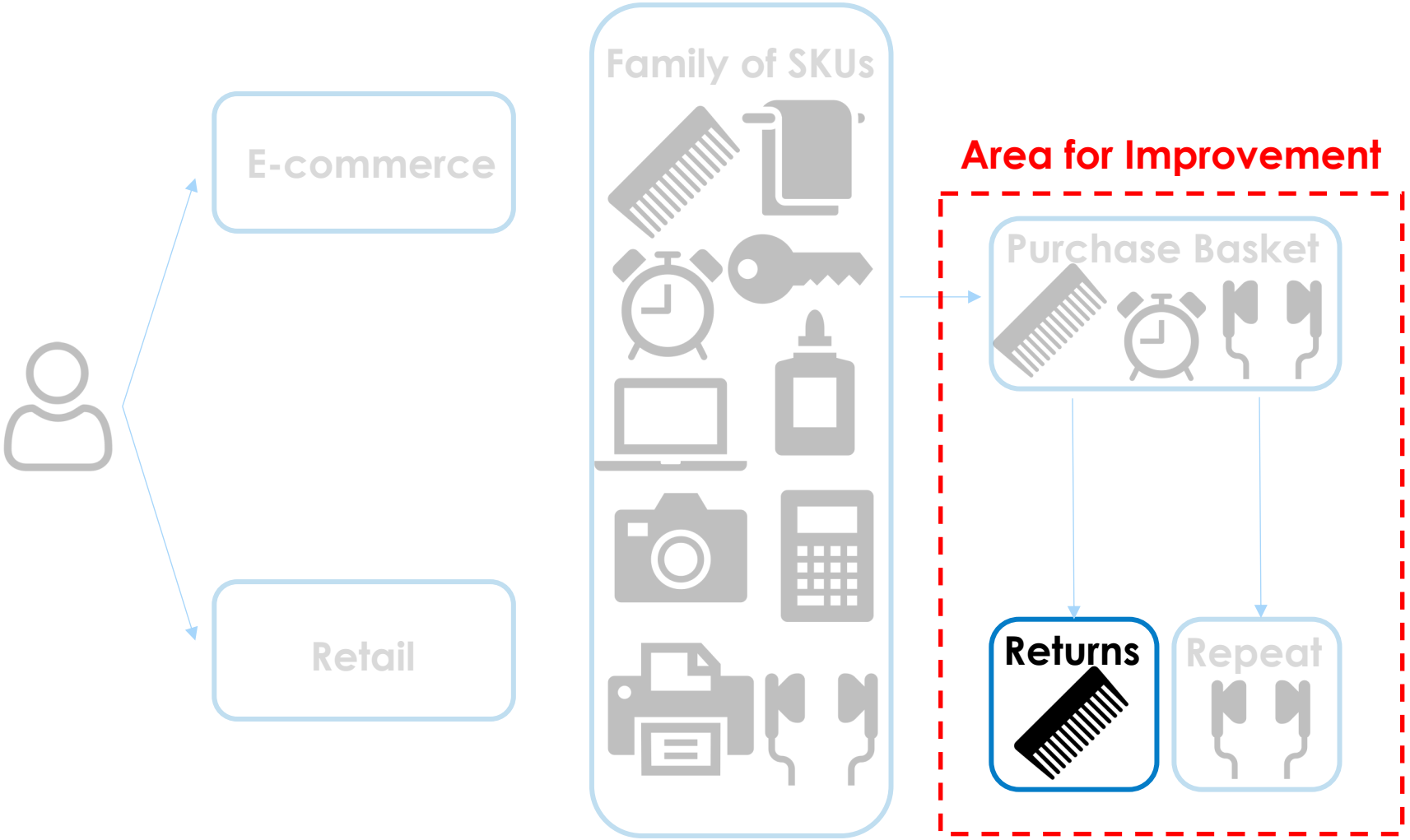
~20%* accuracy predicting what product category someone is interested in (40 categories total).

~40%* accuracy predicting what time someone would be willing to make a purchase.

*After running K-nearest neighbors on source of transaction and customer title to predict product category and time of sale.

Phase 2: Discourage Returns

We need to create a system of data generation that will allow us to use predictive analytics to discourage returns.



Phase 2: Discourage Returns- Issue

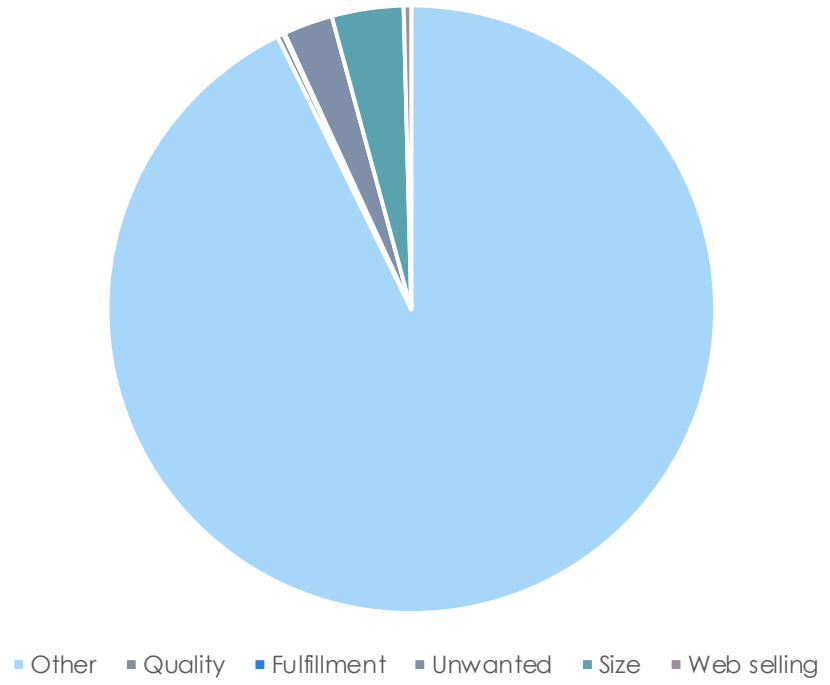
Predictive analytics are difficult if we don't know underlying return reasons in "other".

Returns in the Women's Wear Segment

Returns are concentrated in the women's wear product category.

~8% of products sold in this category were returned vs. ~2% for the second highest returned product category.

However, returns reasons were concentrated in the "other" segment, indicating a need for the company to generate more specific responses.



Big Issue: *We don't know WHY products are being returned*

Phase 2: Discourage Returns- Method

Our Proposed Solution

Illustrative Example	Data Generated

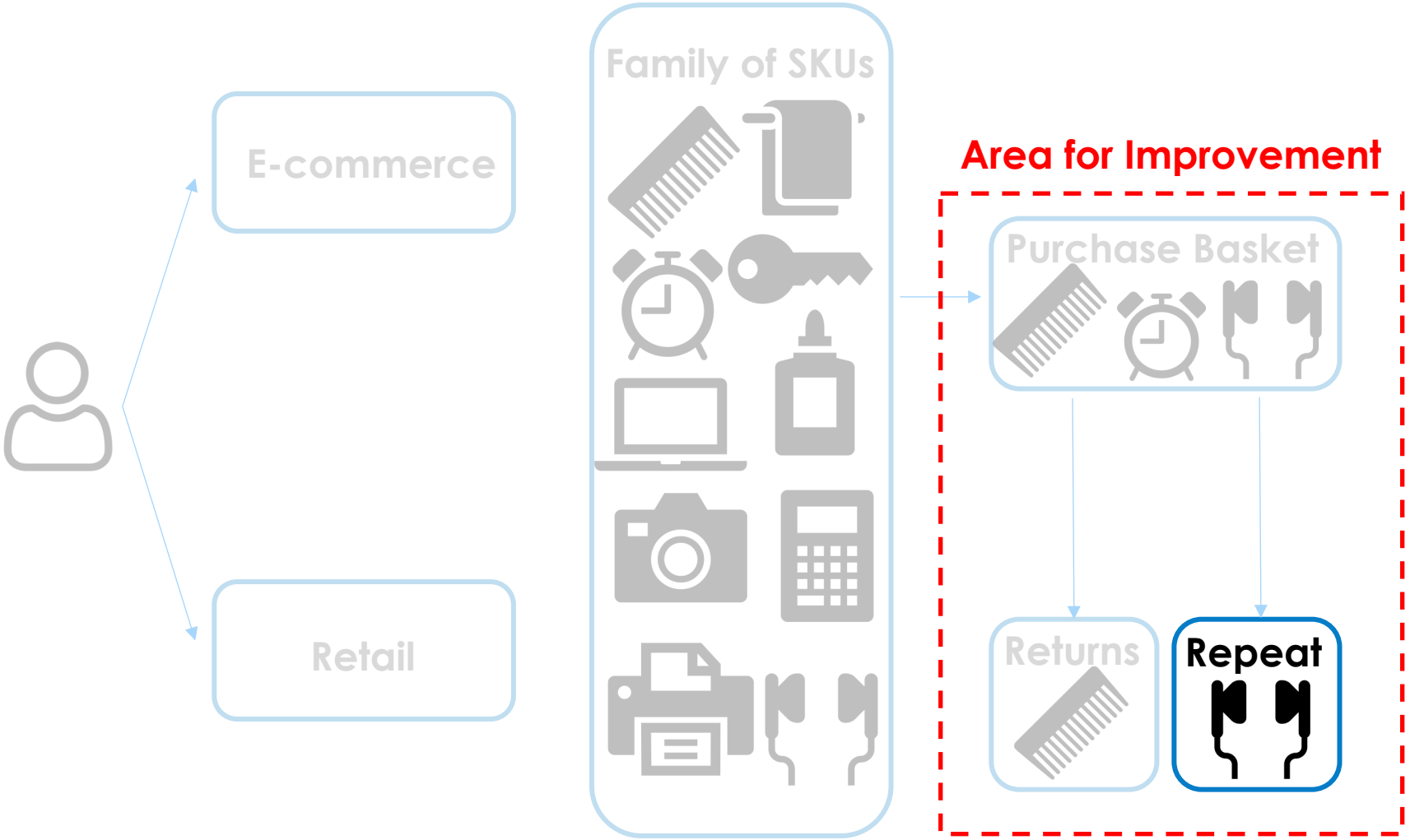
Primary Category

Secondary Category

User-generated Data

Phase 3: Encourage Repeat Purchases

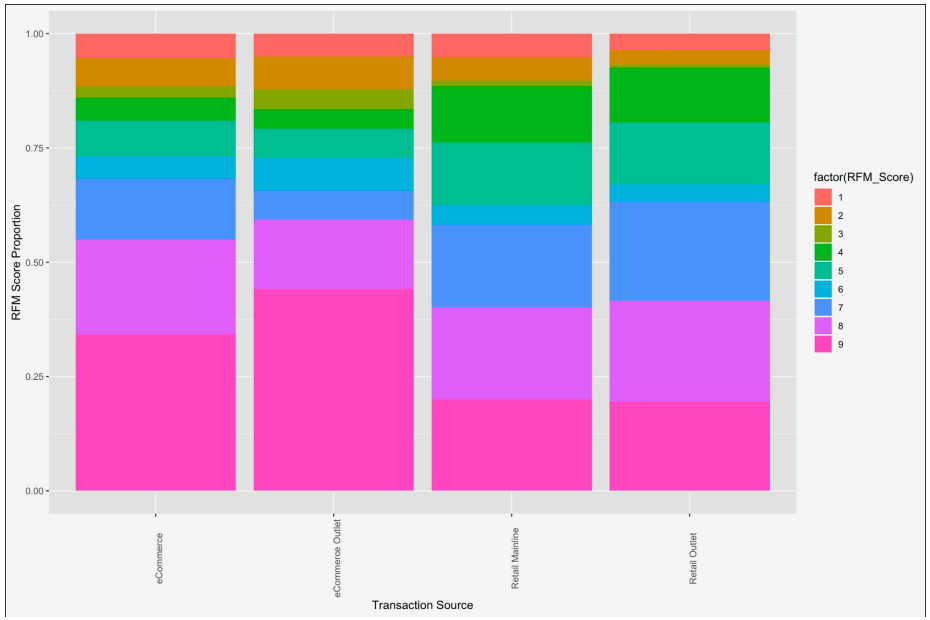
RFM does not accurately predict returns. Our ranking more accurately predicts these return purchases.

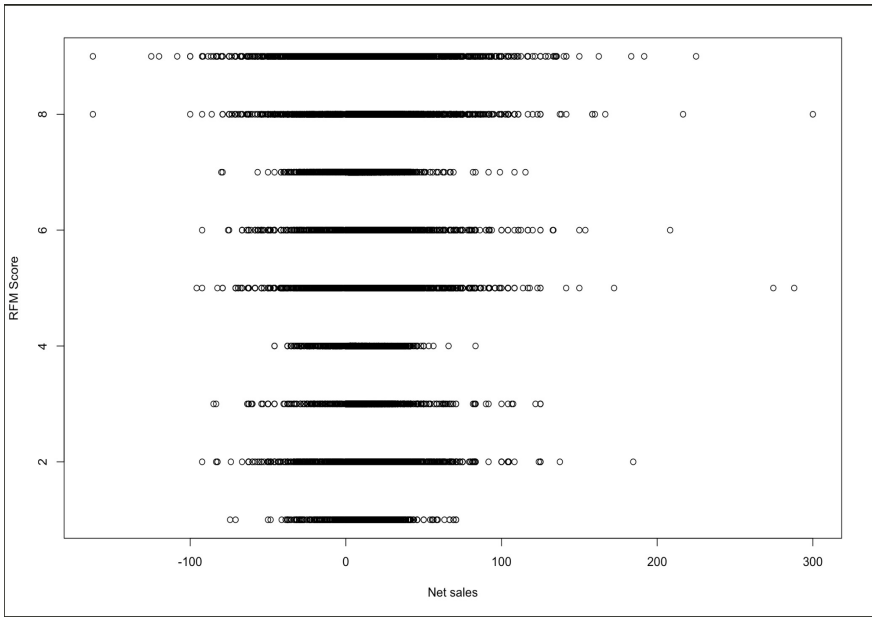


Phase 3: Encourage Repeat Shopping- Issue

RFM Score isn't the best predictor to base marketing decisions on.

Big Issue: *Current methods of customer segmentation (i.e. RFM score) have little predictive value*





1. *Highest cohort (7-9) show similar shopping patterns across channels*

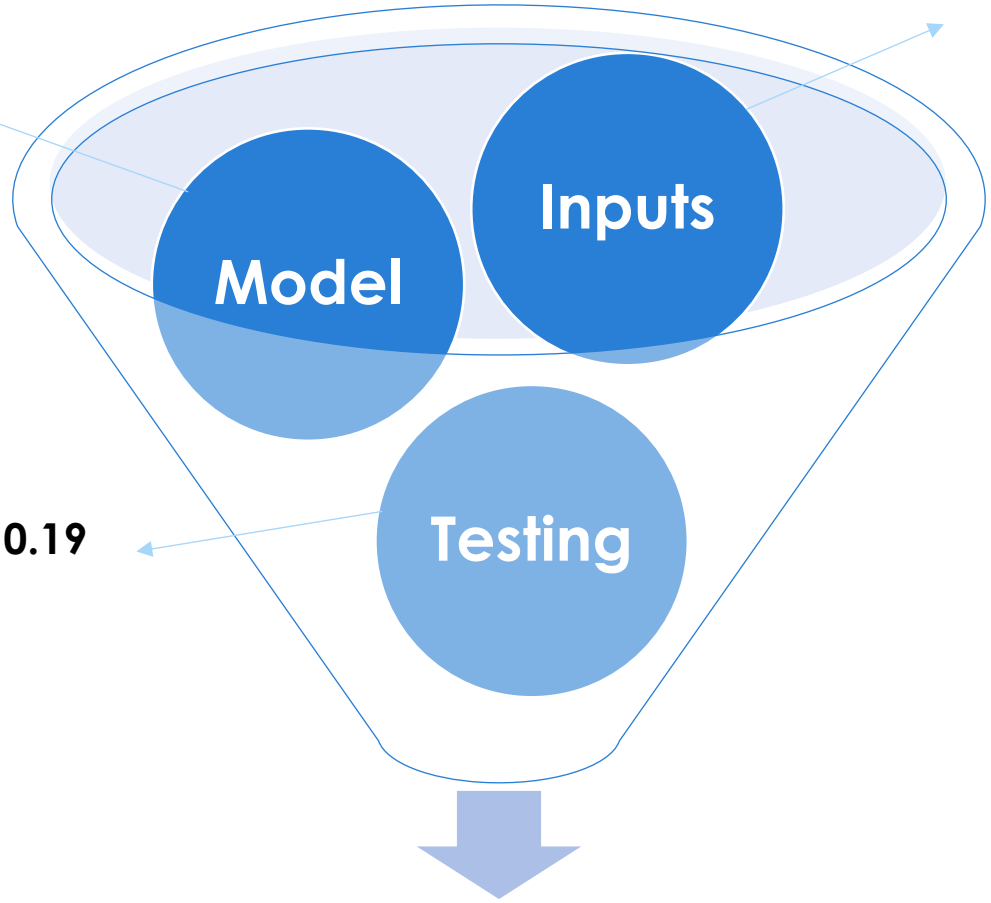
2. *No significant revenue/customer differences across RFM scores*

Phase 3: Encourage Repeat Shopping- Method

We created a new scoring system that has far greater predictive power.

1. LASSO Regression
2. Random Forest

1. Transaction Source
2. Product Category
3. Channel Name



Out-of-Basket MSE = 0.19

mtry = 6, ntree = 500

Will this customer buy an item again?

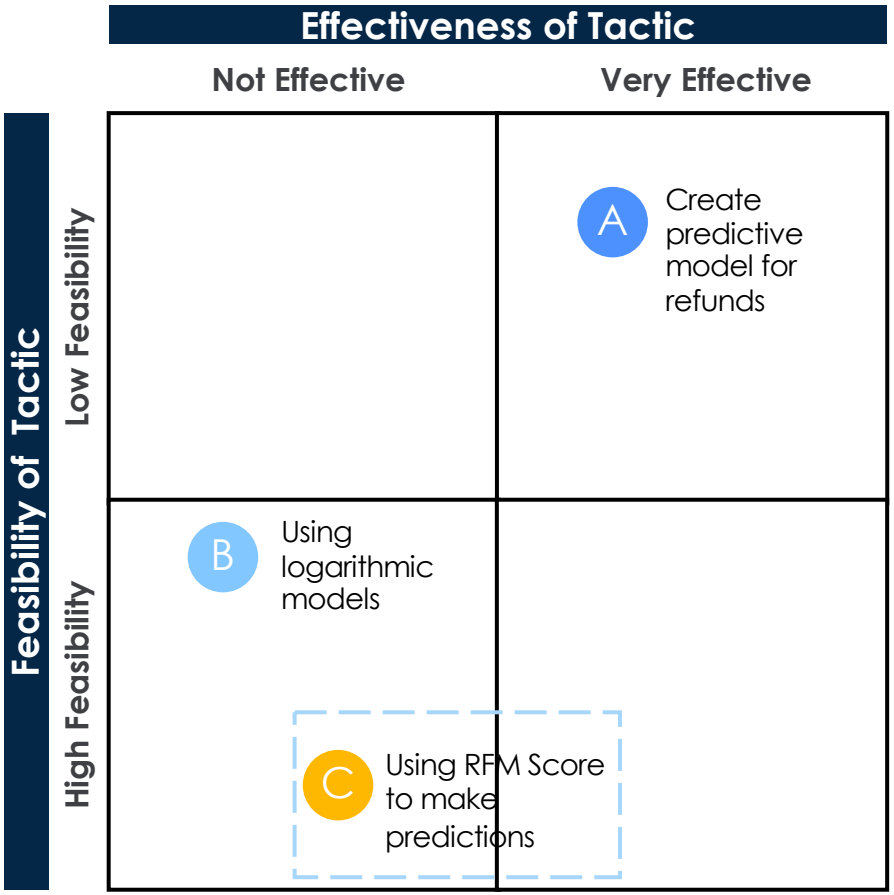
Other Recommendations

We recommend the following to drive better insights

Recommendation	Rationale
Refine customer score	Our analysis showed that RFM score is not extremely useful in generating strategies. Creating a proprietary score should drive a competitive advantage.
Modify customer data fields	To more accurately profile customers, there must be a clear distinction between numerical and categorical data. For example, for prefixes, customers can enter anything they want.
Generate unique customer accounts	A subset of ~500,000 rows from the data set show that ~40% do not have a unique customer account, which hinders ability to make predictions about repeat purchases.
Create separate database for deliveries	Helps keep actual transactions separate from "pseudo-transactions" such as free deliveries.

Alternative Strategies

We chose our strategy over these alternatives to maximize feasibility and impact.



Weakness

A Not enough information to accurately assess why a product is being returned.

B Most of the glm models our team generated had very poor ROC curves.

C There is no clear differentiating factor across RFM Score cohorts, especially in terms of mean revenue generation per customer.

Conclusion

Our models should drive increased revenues and generate more value from the same customer base.