SWIRE COCA-COLA

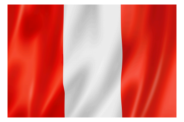
DELIVERING GROWTH THROUGH DATA INSIGHTS

GROUP 8

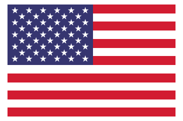
MEET OUR TEAM

WAYNE PARK

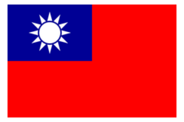
ESTEFANY ALVARADO



NICK ACOSTA



JOCELYN CHANG



ANAIS CORRAL

TABLE OF CONTENT

BUSINESS PROBLEM

Optimizing logistics by **identifying high-potential customers** currently on

white truck (ARTM) delivery → *red truck (direct delivery)*

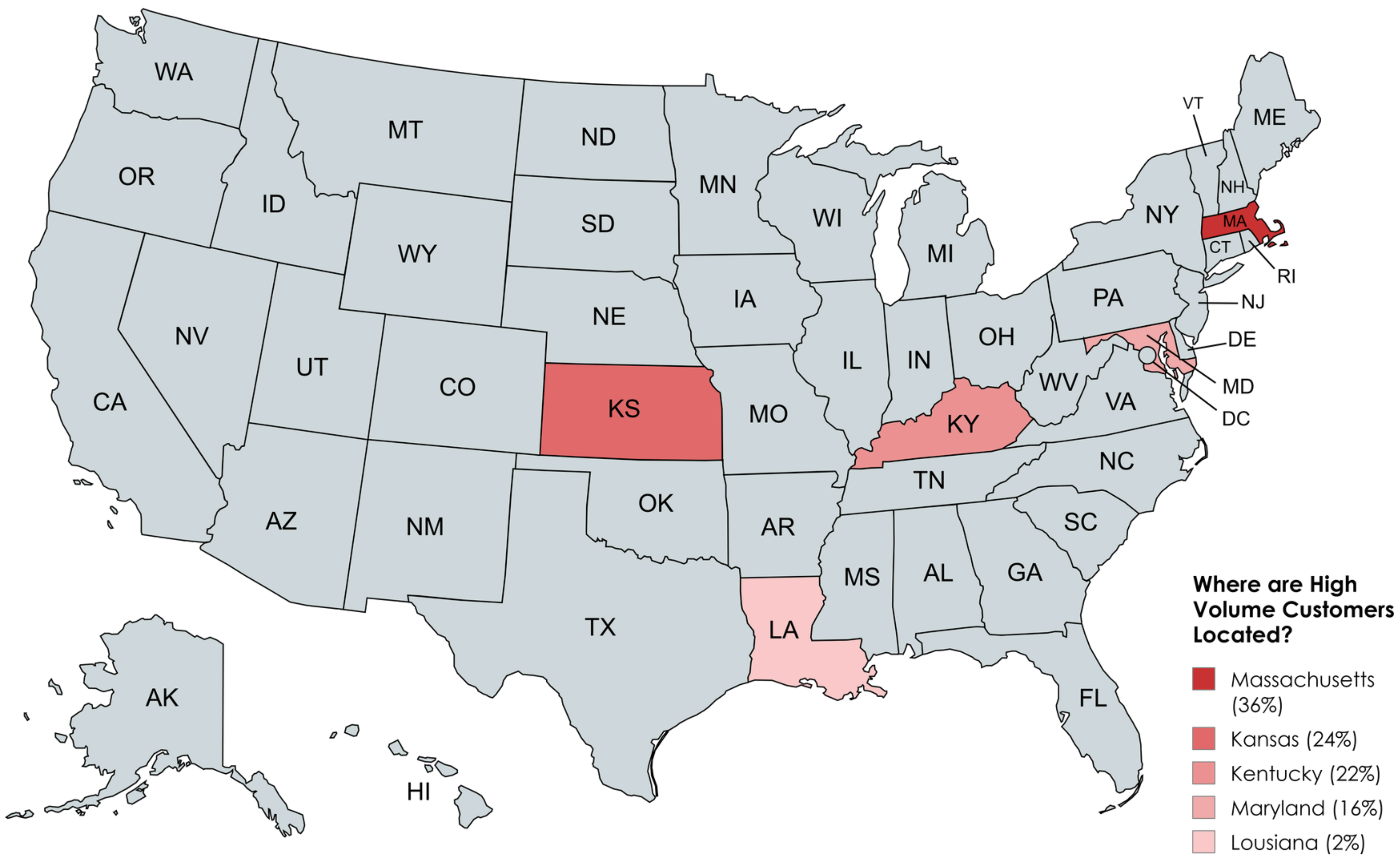
to support future growth while maintaining cost efficiency.

OUR APPROACH

- We used modeling approaches to predict high-growth customer accuracy.
- Evaluated high volume costumers by state and delivery costs to discover relationships between the two.
- Performed customer segmentation through XGBoost with Classification Trees as a cross-validation to determine the characteristics of high-growth customers.

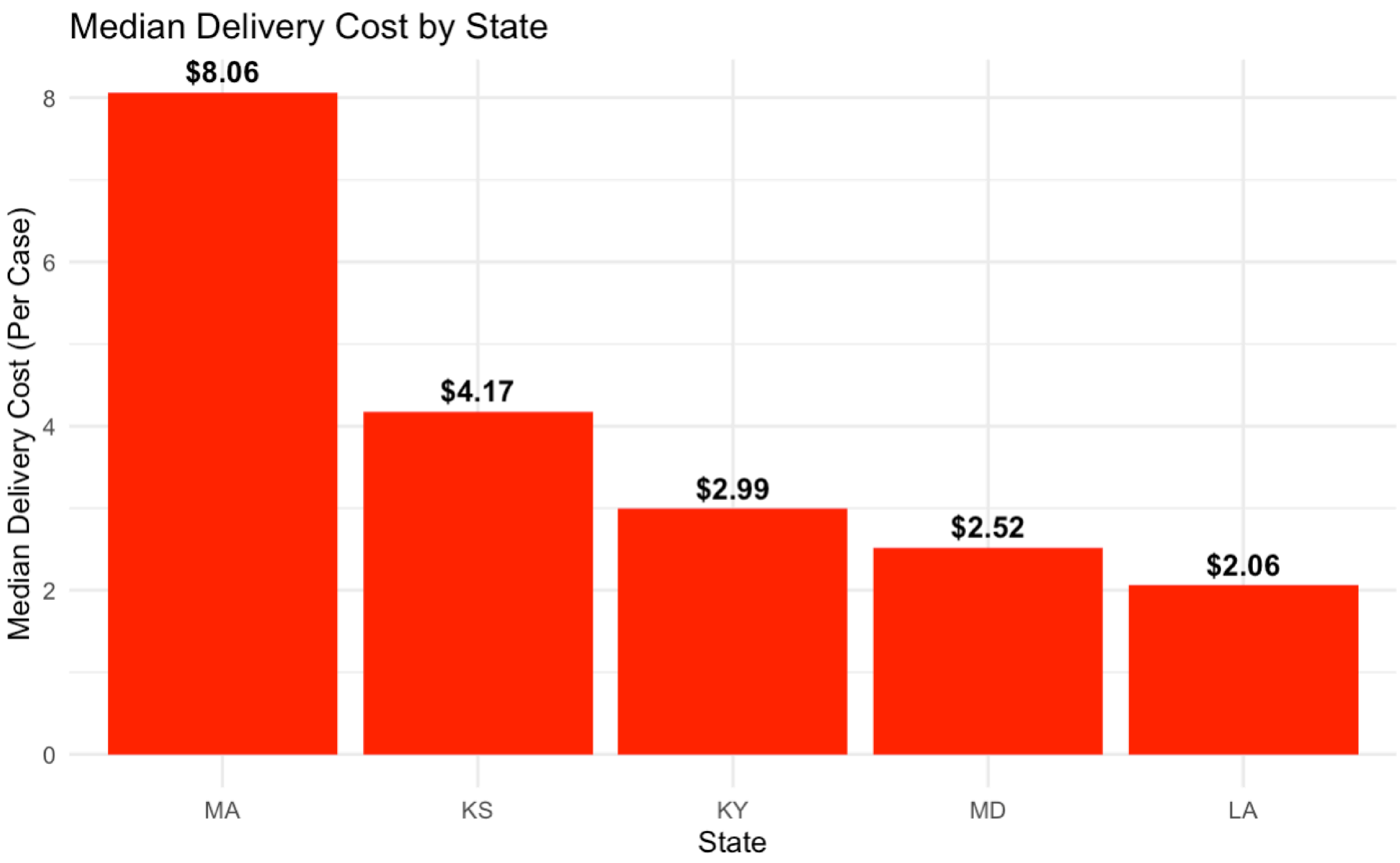
HIGH VOLUME CUSTOMERS BY STATE

*Cases & Gallons



STATE BY DELIVERY COST

*Cases & Gallons



WHY STATE-LEVEL DELIVERY COST ANALYSIS MATTERS?

- Geographic Disparities in Distribution Costs
- Direct Impact on Profit Margins
- Data-Driven Resource Optimization

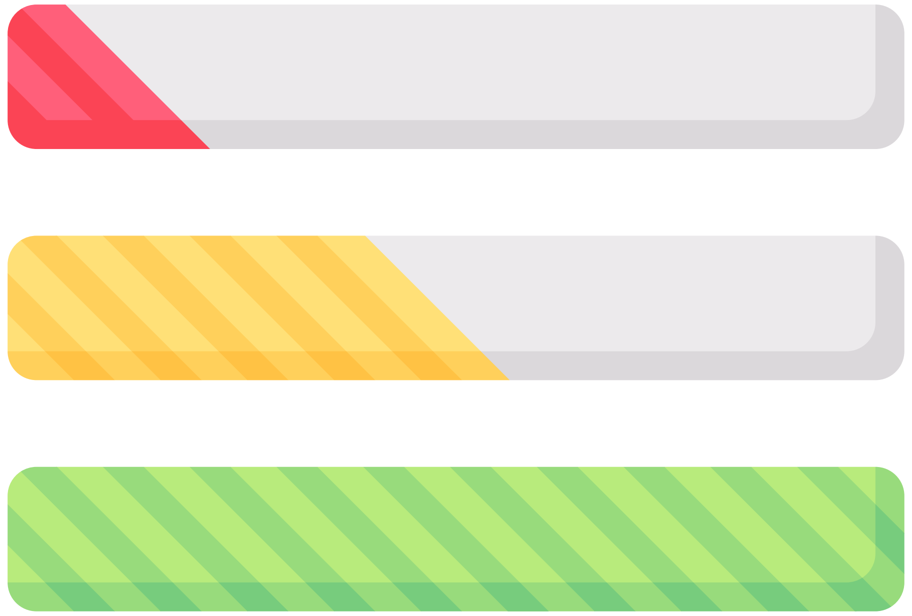
SEGMENTATION

XGBOOST MODEL → PREDICT CUSTOMER GROWTH RATES

- HISTORICAL INCREASES IN <u>DELIVERED CASES</u>

AFTER SEGMENTATION → BETTER PERFORMANCE

SEGMENTATION



LOW GROWTH

MEDIUM GROWTH

HIGH GROWTH

HIGH GROWTH CUSTOMERS

TOP CHANNEL: DINING

TOP TRADE CHANNEL: FAST CASUAL DINING

TOP FREQUENT ORDER TYPE: SALES REP

TOP STATE: MASSACHUSETTS

AVG. DELIVERY COST: $2.88

RECOMMENDATIONS & NEXT STEPS

1 Transition High-Growth Customers to Red Truck Delivery

- Target “High Growth” customers identified by XGBoost.
- Prioritize ZIPs 67647 (KS), 42033 (KY) & 2298 (MA).
- Focus on fast-casual & comprehensive dining segments (hotels, fast food chains, etc).
- Strong fit: Customers using Sales Rep channel with rising volume.

RECOMMENDATIONS & NEXT STEPS

2 Optimize Number of Orders to Reduce Delivery Costs

- Current avg. cost: **$2.88** per delivery.
- Avg. Orders per Customer per Year: **232.08**
- Frequent but small orders → Bundle orders

IMPLEMENTATION STRATEGY

Offer bulk order discounts to improve delivery efficiency.

Ex) Reduce to 100 Orders/Year

- 📉 % Reduction: ~**56.6%** (Saving)

RECOMMENDATIONS & NEXT STEPS

3 Strengthen Sales Rep Channel Engagement

- Sales Reps are the primary order channel for growing customers
- Empower reps with tools for account growth & personalized offers
- Drive loyalty and conversion to direct delivery

IMPLEMENTATION STRATEGY

Performance-based incentives, offering rewards (bonuses and discounts) for meeting sales and customer engagement targets.

RECOMMENDATIONS & NEXT STEPS

3 Estimated Strategy Impact:

- Nearly **20,000 customers** currently order through Sales Reps
  - **5% of customers** → 996
- Average order value: $250
- Average orders per year: 527.5
- Profit margin: 30%

If incentivized Sales Reps convert just 5% of their customers, they can generate:

→ **+$130 M in new revenue**

→ **$39 M in profit *annually***

THANK YOU!

Q&A

RECOMMENDATIONS & NEXT STEPS

3 Estimated Impact: **explanation**

- There are almost 20,000 customers (19,929 to be exact) who regularly order through the Sales Rep channel.
- If performance-based incentives help Sales Reps influence just 5% of their customers to switch to direct delivery, that would be around 996 customers.
- High-growth customers place frequent orders — about 527 times a year — and typically spend around $250 per order.
- Based on this, converting those 996 customers could generate:
  - **over $130 million in new revenue**, or
  - roughly **$39 million in profit annually** (assuming a 30% profit margin.)

RECOMMENDATIONS & NEXT STEPS

3 *$250 avg order value explanation

- We don't have exact pricing per case in the dataset, but we do know how many cases are typically delivered per order.
- From our analysis, high-growth customers receive around 16–17 cases per order on average.
- To estimate the value of an order, we assumed a conservative price of $15 per case, which is in line with wholesale beverage pricing.
- So: **17 cases × $15 = $255, which we round to $250 per order as a reasonable estimate.**

Revenue = Number of Customers × Avg Order Value × Orders per Year Revenue

996 × 250 × 527.5 = **$131,887,500**

Profit = Revenue × Profit Margin

131,887,500 × 0.30 = **$39,566,250**

RECOMMENDATIONS & NEXT STEPS

3 *30% profit margin explanation

We assumed a 30% profit margin based on industry standards and Swire's own cost data. Since the average delivery cost per case is $2.88 and we estimated $15 per case in revenue, this leaves plenty of room for other operating costs — making 30% a conservative and realistic estimate.

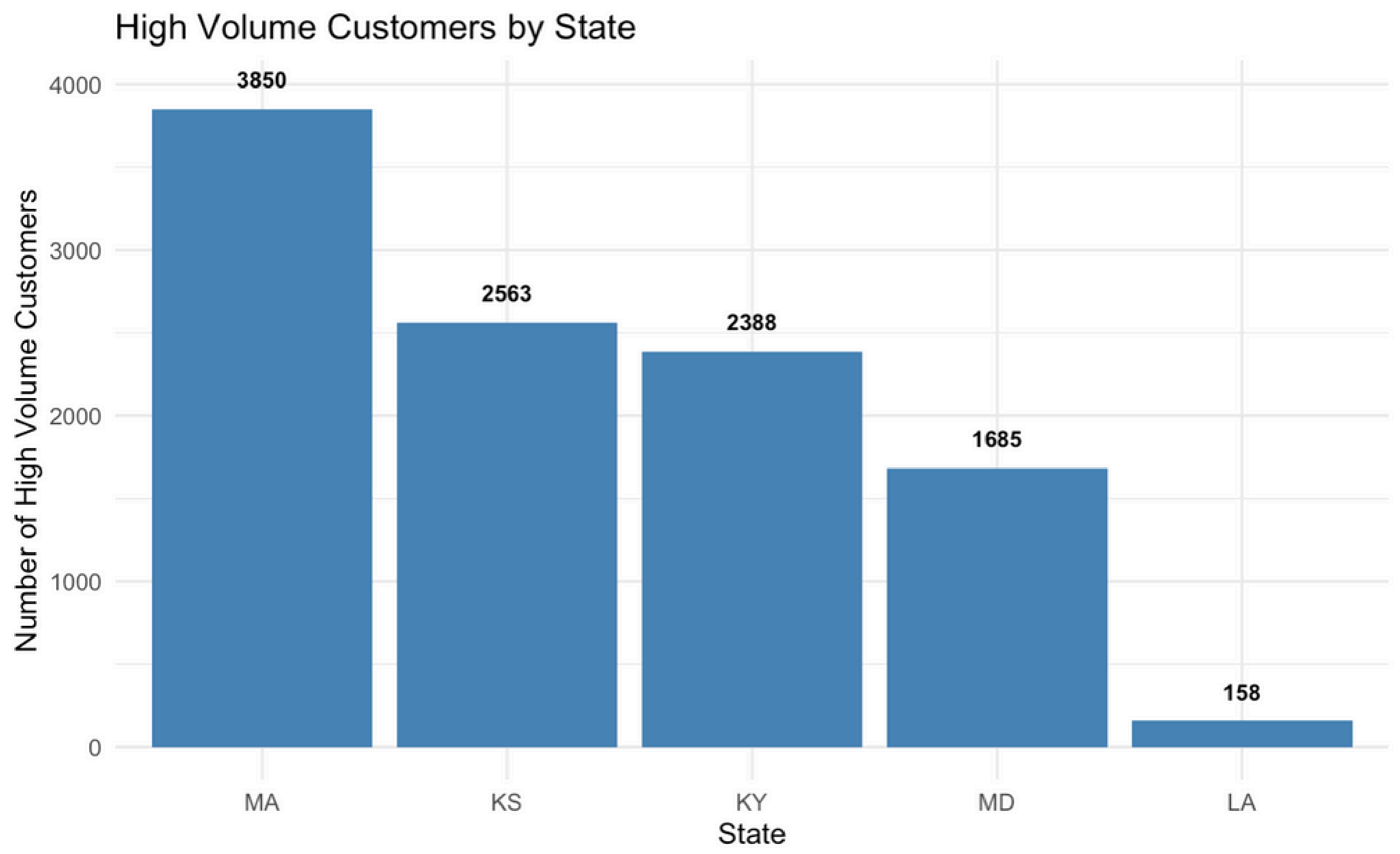
Assumed a 30% profit margin based on:

- Industry norms for companies like Swire Coca-Cola (distribution/logistics-heavy but still with strong margins),
- And data showing delivery costs are relatively low compared to estimated revenue:
  - Average delivery cost per case: $2.88
  - Estimated price per case: $15
  - That's a gross margin of ~81% before accounting for other costs — so 30% net profit is conservative and realistic.

Volume Tier Classification and Business Justification

Tier	Volume.Range..Gallons.	Reason
Low Volume	≤ 200	Slightly above median (153 gallons), includes the lower ~55% of customers.
Medium Volume	201 – 1000	Captures customers between ~50th and 90th percentile
High Volume	> 1000	Above 90th percentile. High-value customers

STATE BY DELIVERY COST



High Volume Customers Distribution by State

state	n_customers	pct_customers
MA	3850	36.2
KS	2563	24.1
KY	2388	22.4
MD	1685	15.8
LA	158	1.5

State <chr>	Median_Delivery_Cost <dbl>
KS	4.17
KY	2.99
LA	2.06
MA	8.06
MD	2.52

Subsetting the dataset:

The top_state_data contains only the rows corresponding to the states MA, KS, KY, MD, and LA.

Summing the Median_Delivery_Cost:

The aggregate function groups the data by State and then sums the Median_Delivery_Cost within each state.

Output:

The resulting state_total_cost will give you the total sum of the Median_Delivery_Cost for each of the five states.

Why State-Level Delivery Cost Analysis Matters?

- **Geographic Disparities in Distribution Costs**

→ Delivery costs are influenced by logistics infrastructure, warehousing proximity, fuel prices, and demand density across states.

→ Understanding these variations enables cost modeling at a granular level.

- **Direct Impact on Profit Margins**

→ Elevated delivery costs in high-volume markets (e.g., MA) significantly reduce contribution margins per case.

→ Identifying cost outliers is essential for improving operational profitability.

- **Data-Driven Resource Optimization**

→ State-level delivery cost insights allow for precision in allocating fleet capacity, warehouse coverage, and last-mile routing.

→ Informs capital investment decisions (e.g., adding a fulfillment node or rerouting carriers)

POSSIBLE QUESTIONS

- What threshold have you used to determine the customer segmentation (low, medium, and high)? /How did you define "high-growth" customers?

- Are there other interesting characteristics in the other 2 segments? why not showing them?
- Exactly what customer groups does Swire use from our deliverables?

- Why did you choose XGBoost and Classification Trees for segmentation?
- Did you validate your model's predictions? If so, how?
- What distinguishes high-growth customers from low and medium growth ones?
- What would be your first step if Swire implements your recommendations tomorrow?
- How did you ensure that the segmentation model accurately captures meaningful and actionable differences between customer segments, and what steps did you take to validate the effectiveness of the segmentation in driving targeted strategies?
- *How did you identify which regions had both high customer demand and low delivery costs?*

CLASSIFIED TREES SEGMENTATION RESULTS:

Top Node Section

Segment 28

number of customers : 397751

avg. cases: 31.58501

avg. gallons: 7.90808

avg. cost: 2.870263

top channel: DINING

top trade channel: FAST CASUAL DINING

Determining the growth for customers?

Customer growth within the Classified Trees were based on actual delivered cases

Inactive customers == 0

Low = < 100

Medium = < 500

High = > 500

XGBOOST

High Growth	Low Growth	Medium Growth
8484145	4926171	497201

High Growth	Low Growth	Medium Growth
61.004024	35.420924	3.575052

Threshold-based segmentation:

- If a customer's growth rate is **above 10** → High Growth
- **Between 1 and 10** → Medium Growth
- And **0 or below** → Low Growth

Steps:

- First, we calculated each customer's growth rate based on their delivered cases → model performed better with delivered cases rather than gallons + segmented them into High, Medium, and Low Growth groups.

- Then, we trained an XGBoost model to help us identify what factors or behaviors are most predictive of being in the High Growth segment.

How we categorize "High Growth" customers:

- We assessed each customer's growth in delivered cases over time. → **historical increases in delivered cases.**
- If a customer's delivered cases increased consistently over the observation period, they were classified as "High Growth."
- We did not apply an arbitrary percentage threshold or top % method. Instead, we directly observed their growth trend in delivered cases.

- Order Type
- Frequent Order Type
- Cold Drink Channel
- Trade Channel, Zip Code
- Median Delivery Cost
- Customer Decision

**using the group_by() and count() functions in the dataset

- **<u>EXAMPLE</u>**:

After segmentation, RMSE dropped from <u>17.11</u> to <u>11.51.</u>

This result show that segmenting customers significantly improved prediction accuracy.

XGBOOST

High Growth	Low Growth	Medium Growth
8484145	4926171	497201

High Growth	Low Growth	Medium Growth
61.004024	35.420924	3.575052

**using the group_by() and count() functions in the dataset

Steps:

- First, we calculated each customer's growth rate based on their delivered cases → model performed better with delivered cases rather than gallons + segmented them into High, Medium, and Low Growth groups.

- Then, we trained an XGBoost model to help us identify what factors or behaviors are most predictive of being in the High Growth segment.

- **EXAMPLE**:

After segmentation, RMSE dropped from 17.11 to 11.51.

This result show that segmenting customers significantly improved prediction accuracy.

🧾 **CURRENT STATE:**

- AVG. ORDERS PER CUSTOMER PER YEAR: **232.08**
- AVG. DELIVERY COST PER ORDER: **$2.88**

🧮 CURRENT ANNUAL DELIVERY COST PER CUSTOMER:
232.08 × $2.88 = $668.39

REDUCE TO 100 ORDERS/YEAR

- DELIVERY COST = 100 × $2.88 = $288.00
- 💰 SAVINGS: $668.39 - $288.00 = $380.39
- 📉 **%** REDUCTION: (1 - 100 / 232.08) × 100 = **~56.6%**

SWIRE COCA-COLA