Fourcasters

Rakamin Data Science
Batch 35

WOMart Sales Analysis
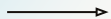
01

Background

Dataset Overview

Contains **sales data** for 18 months from 365 WOMart stores. This data includes Store Type, Location Type, Region Code, Holiday Status, Discounts offered daily, the Number of Orders each day.

Background

Problem
Store types S2 and S3 have sales that are consistently lower than the average total sales, indicating below–average performance.

Metrics
Sales

Goal
Improving the sales performance of underperforming store types to achieve or surpass the average total sales

Objectives
1. Analyze the factors contributing to underperformance in sales for specific store types
2. Develop actionable business recommendations to rectify the identified sales underperformance factors

Meet our Team

And their roles

Atthoriq Putra Pangestu
Data Scientist

Alvida Dwiki C
Business Analyst

Emir Akbar
Data Analyst

Reny Rafiqah
Business Intelligence Analyst

Darell Timothy Tarigan
Data Analyst

Nur Baiti Listyaningrum
Data Analyst

02

Data Exploration

Statistical Summary

	ID	Store_id	Store_Type	Location_Type	Region_Code	Date	Holiday	Discount	N_order	Sales
count	188340	188340	188340	188340	188340	188340	188340.000000	188340	188340.000000	188340.000000
unique	188340	365	4	5	4	516	NaN	2	NaN	NaN
top	T1000001	1	S1	L1	R1	2018-01-01	NaN	No	NaN	NaN
freq	1	516	88752	85140	63984	365	NaN	104051	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	NaN	0.131783	NaN	68.205692	42784.327982
std	NaN	NaN	NaN	NaN	NaN	NaN	0.338256	NaN	30.467415	18456.708302
min	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	0.000000	0.000000
25%	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	48.000000	30426.000000
50%	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	63.000000	39678.000000
75%	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	82.000000	51909.000000
max	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	371.000000	247215.000000

# Statistical Summary

Categorical variables:

- There's **no duplicates in ID**, meaning that the dataset doesn't have any duplicate records because each ID represents one observation.
- Store with **Store Type S1** has the most frequency in the dataset.
- Store with **Location Type L1** has the most frequency in the dataset.
- Store with **Region Code R1** has the most frequency in the dataset.
- Most of the store sales record **doesn't have discount.**

Numerical variables:

- The typical order quantity from each store is around **68 orders** per day.
- `N_order` has a **slightly right skewed** data, meaning that several stores sold a higher quantity of product than other stores in a day.
- Sales has a **slightly right skewed** data, meaning that several stores hit a higher sales than other stores in a day.
- Based on Holiday, most days are **not holidays.**
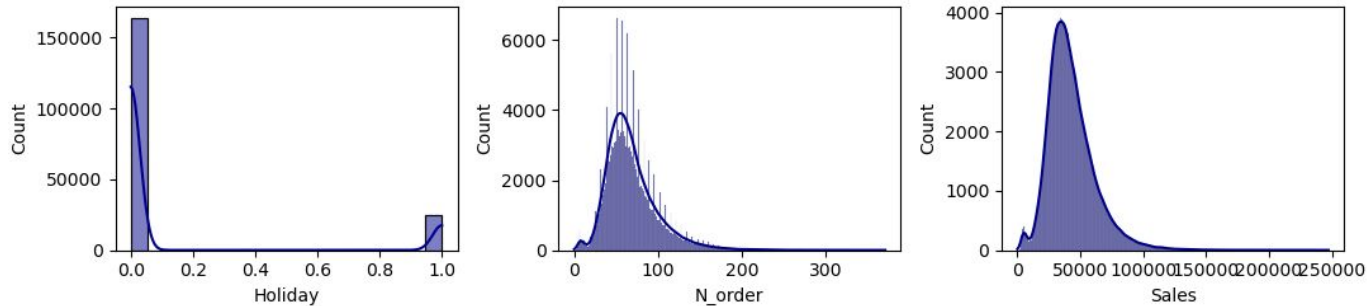
Univariate: Boxplot

Numerical Variable Distributions: Boxplot
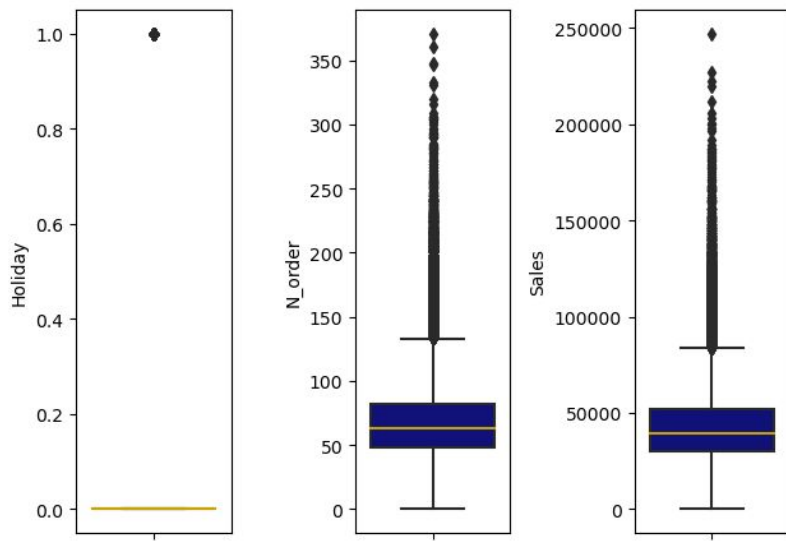
- Holidays, marked with a value of 1 in the box plot, are **rare occurrences** in the dataset.
- On an average day, stores typically sell between **48 and 82 orders**, but some stores sometimes exceed **140 orders** in a single day.
- Daily sales typically range from **$30,000 to $50,000** per store, but certain stores sometimes achieve sales **exceeding $80,000** in a day.
- Notably, there is **one extreme outlier** in the Sales variable, with a value of approximately $250,000 in a single day.
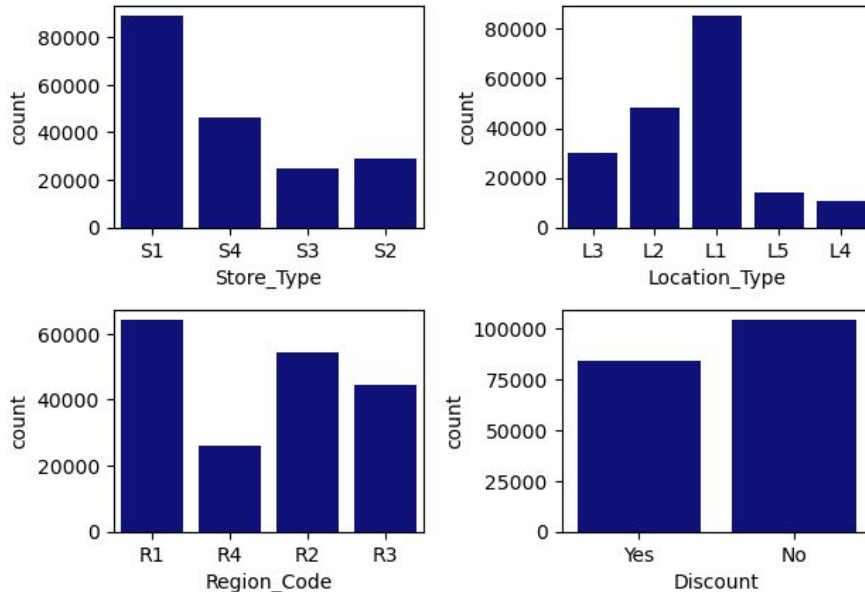
Univariate: Histogram

Numerical Variable Distributions: Histogram

– It's evident that **holidays are infrequent** in the dataset.
– N_order shows a **slight right–skewed** distribution.
– Sales also display a **slight right–skewed** distribution.
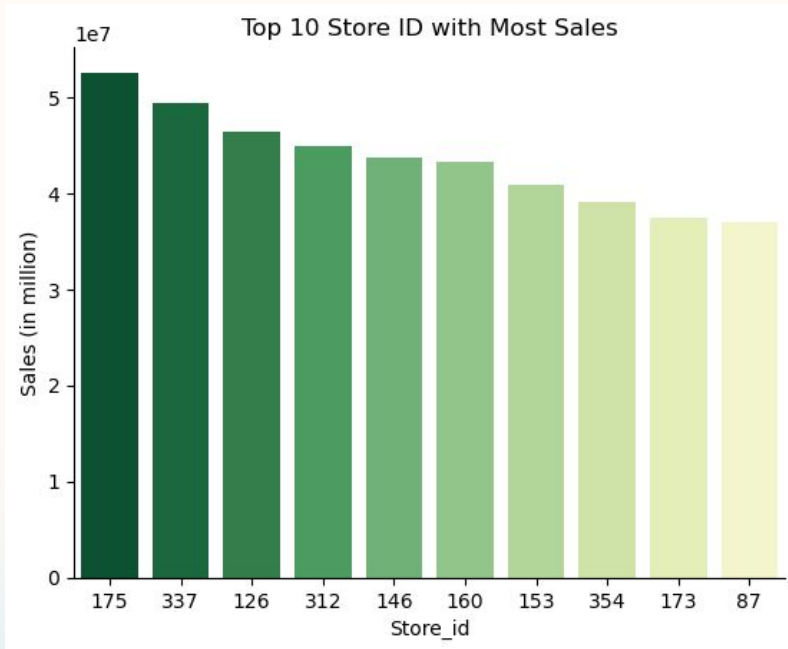– There are days when the stores **don't make any sales** and orders.

Univariate: **Boxplot**


Categorical Variable Distributions

- **S1** is the most common Store_Type.
- **L1** is the most common Location_Type.
- **R1** is the most common Region_Code.
- Discounts are **less commonly applied** to orders per store.

Bivariate: Top 10 Stores with Most Sales

Top 10 Store ID with Most Sales

– Store ID 175 has the most sales in the last 18 months with **over $52 million** in total.

# Bivariate: Store Type Sales

Total Sales by Store Type
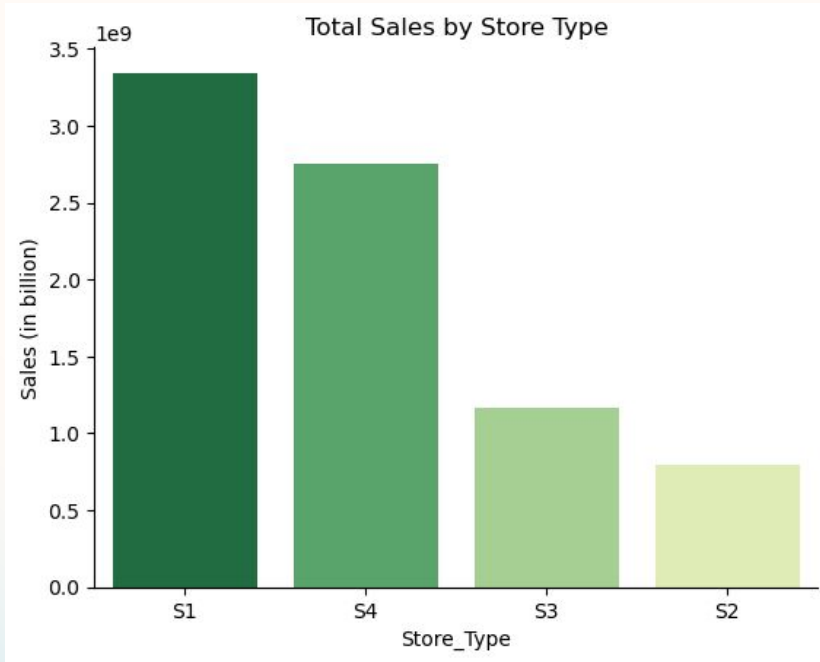
- **Store Type S1** has the **most** total sales in the last 18 months with around **$3.3 billion** of sales.
- **Store Type S2** has the **least** total sales in the last 18 months with around **$795 million** of sales.

Bivariate: Location Type Sales


Total Sales by Location Type
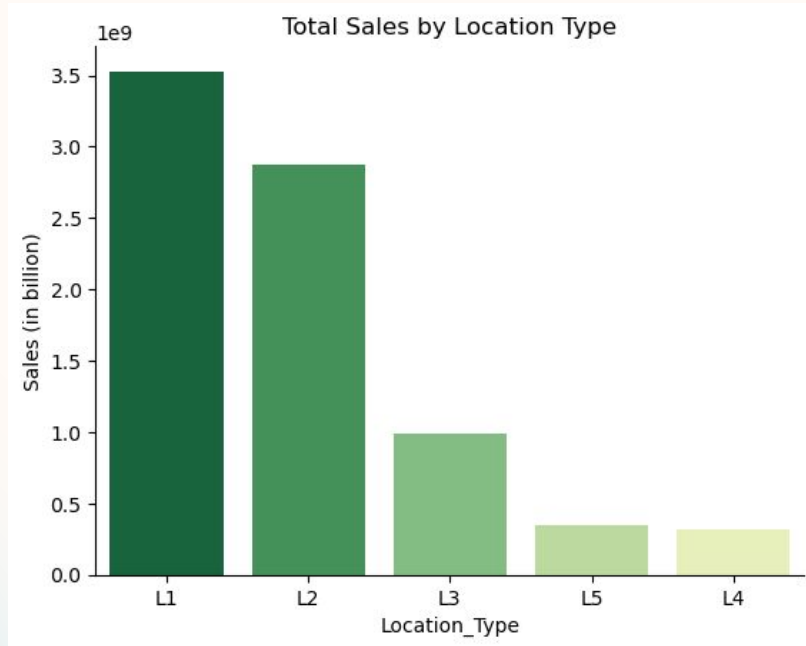
- – **Location Type L1** has the **most** total sales of around **$3,5 billion** in the last 18 month.
- – **Location Type L5 and L4** has the **least** total sales of around **$315 – $351 million** in the last 18 month.

Bivariate: Regional Sales

Total Sales by Region Code
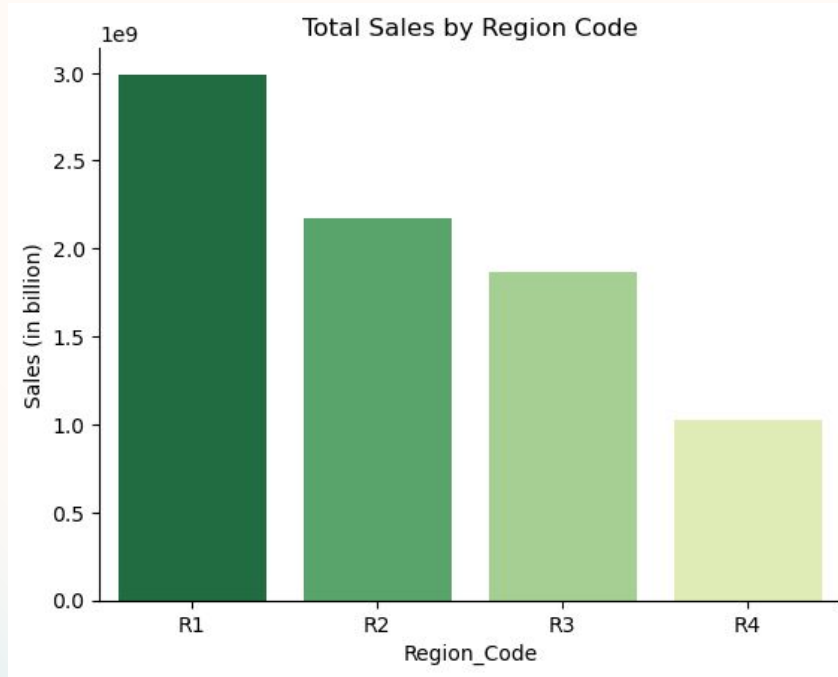
- Region Code R1 has the most total sales of around $2,9 billion in the last 18 months
- Region Code R4 has the least total sales of around $1 billion in the last 18 months

Monthly Total Sales

Monthly Total Sales
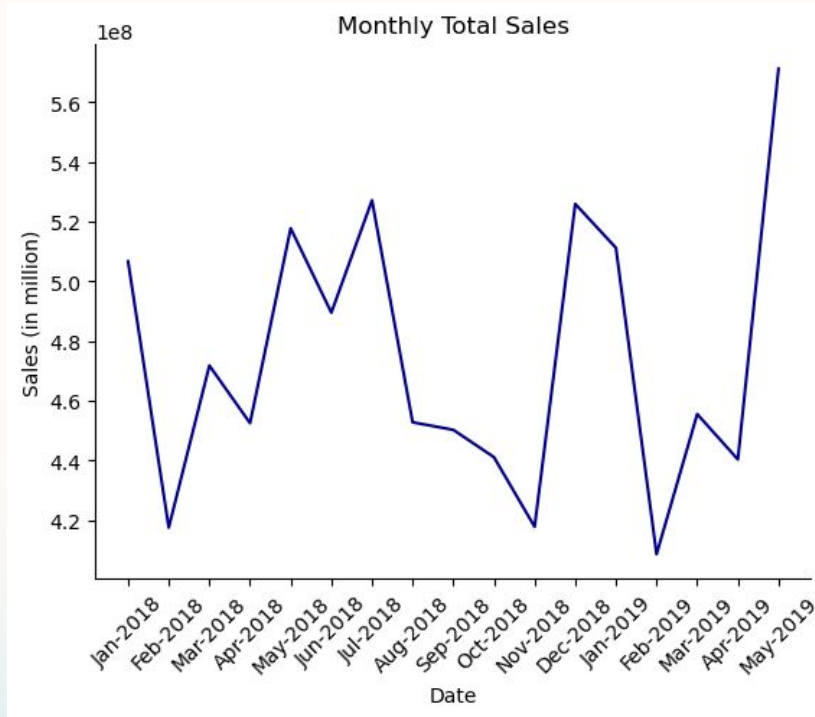
– In 2018, sales showed a **consistent uptrend** from **February to July**, followed by a **decline toward November**, which might be attributed to **post–holiday spending fatigue**. However, **sales surged in December**, likely due to **holiday–related shopping**.

– In the first four months of 2019, there was a **sharp decline in sales** from January to February, possibly reflecting **post–holiday spending fatigue**. Subsequently, there was a **significant and rapid uptrend** leading into May.

– Looks like the **uptrend of February to May** in 2019 is **much higher** than 2018. This could indicate **business growth** or **improved sales strategies** compared to 2018.

Monthly Total Sales by Holiday

Monthly Total Sales by Holiday Status
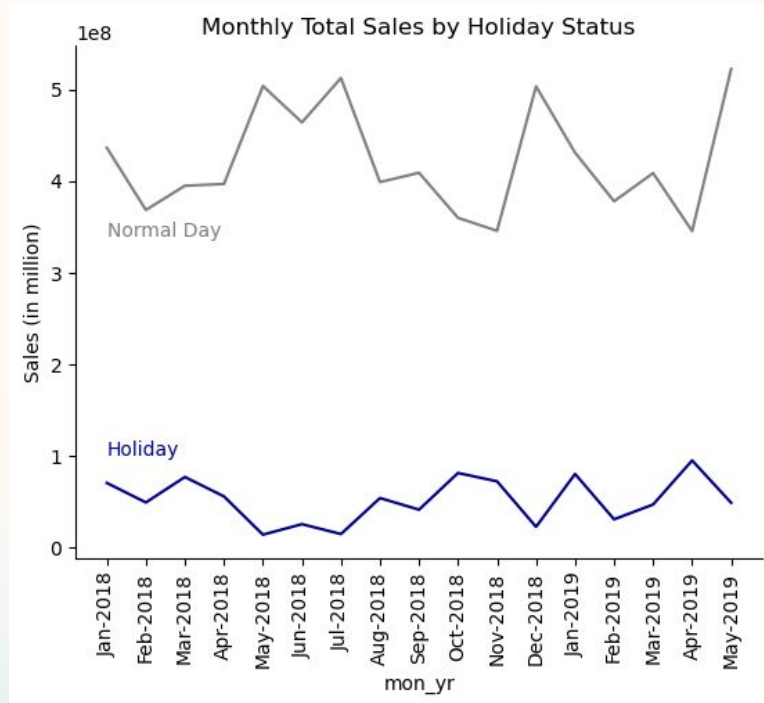
- There's a **big difference** in total sales between **regular days and holidays** over the past 18 months. **Regular days** have been **much better for sales**.

- Interestingly, the 'holiday' label in the dataset **doesn't match** the usual holiday pattern. For example, during November to December, when we'd expect higher sales due to big holidays like Christmas and New Year's Eve, the 'holiday' label shows a decrease instead.

- We could consider this sales gap between holidays and regular days as an **issue**. However, **we need more information**, especially a more **precise description** of what the 'holiday' label means from the dataset owner.

Multivariate: Correlation Heatmap

- **More orders mean more sales**, and more sales mean more orders, they're **strongly connected** (obviously!!).

- Sales and the number of discounts have a relationship; **more discounts can mean more sales**, but it's **not a very strong** connection.

- The number of orders and the number of discounts are related too. **More discounts can lead to more orders**, but it's **not a very strong** link either.

So, What's the Problem?

Some store types, locations, and regions are not doing as well as others (their sales are below average)

Let's see the sales based on store types, locations, and regions.

# Lower than average location type sales

Sales by Location Type
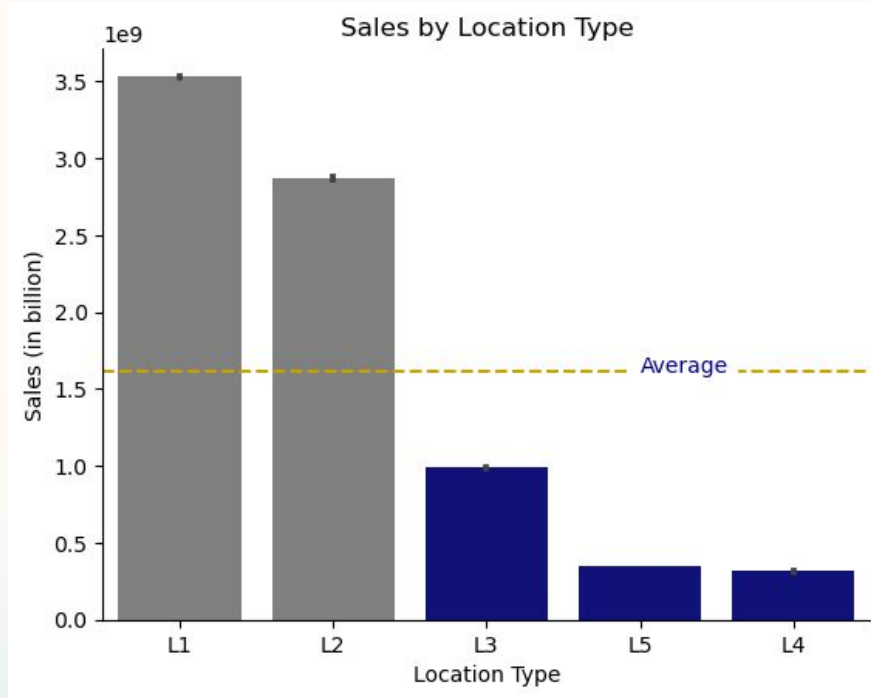
– Location Types **L3, L4, and L5** have much **lower sales than the average** of all location types. This suggests that these location types might **not be the best for stores**. Possible reasons could include **not understanding the customers well** in those locations, the **store layout** not being set up in the best way, and other things like that.

Lower than average regional sales
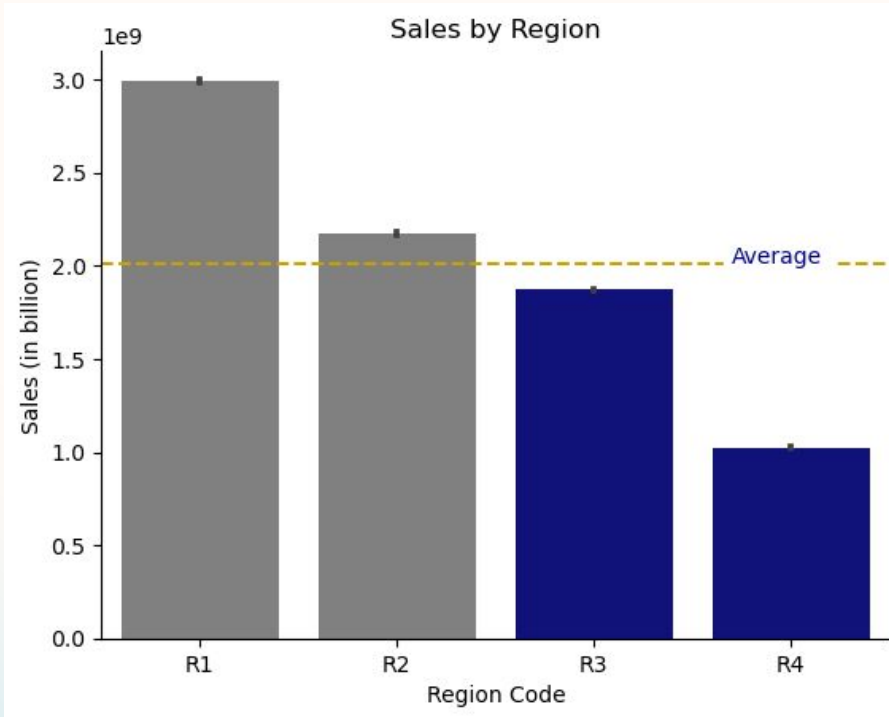
– **Region Codes R3** have sales slightly **below average**, but **R4** is **significantly lower** than average. This could be because the strategies they're using to attract customers in that region **might not be very effective**, or perhaps **there aren't many customers** in that area who fit their **target market**.

Lower than average store type sales
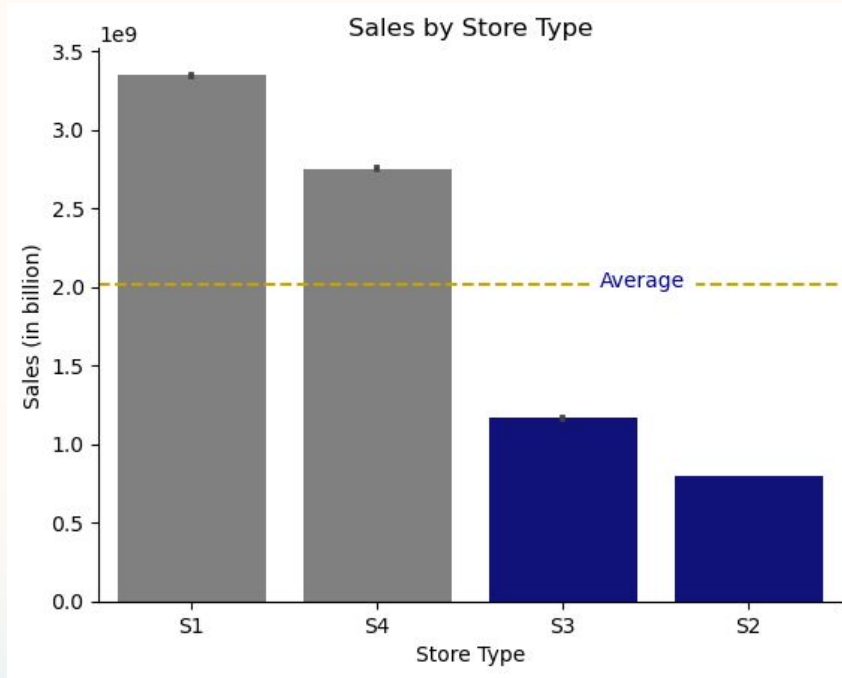
Sales by Store Type

– **Store types S3 and S2** have sales much **lower than the average** of all stores. This might be because they **don't have a wide variety** of products, the products they have **might not be what the local customers want**, or other reasons.

Key Takeaways

The following store type, location, and region is underperforming in sales:

- – Store Type: **S2 and S3**
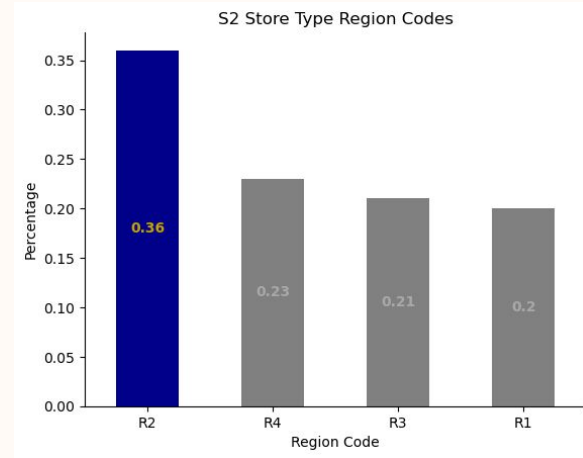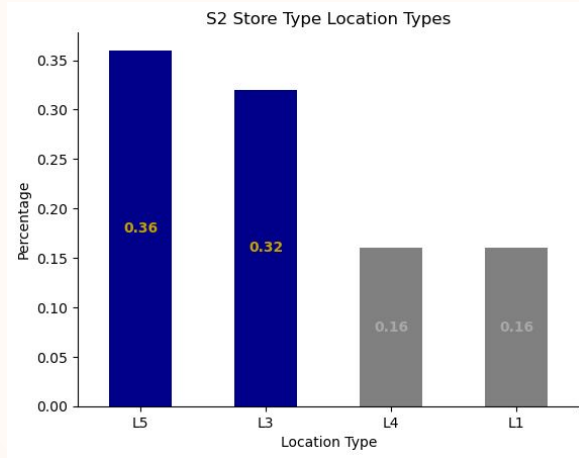- – Location Type: **L3, L4, L5**
- – Region: **R4**

We'll focus on one problem first

S2 and S3 Store Type Investigation

Factor analysis that could cause sales to go lower than average

S2 Store Location Types and Region

- – Most S2 stores are in areas with **location types L5 and L3**, where **sales are lower** than average.
- – S2 stores are mainly in **region R2**, which has **slightly higher sales** than the average.

  The issue might be with the **location types**, as L5 and L3 locations have lower sales performance.

S3 Store Location Types and Region

S3 Store Type Location Types



S3 Store Type Region Codes
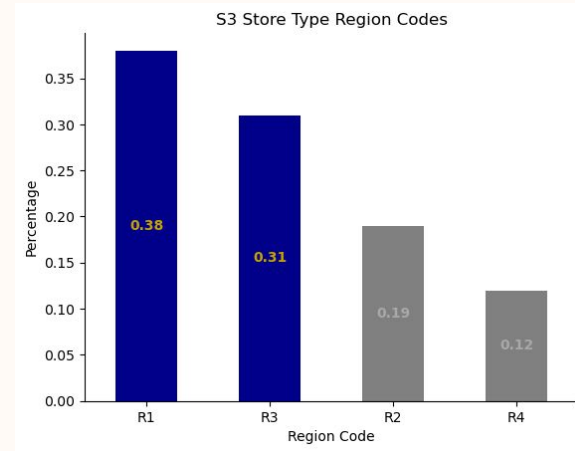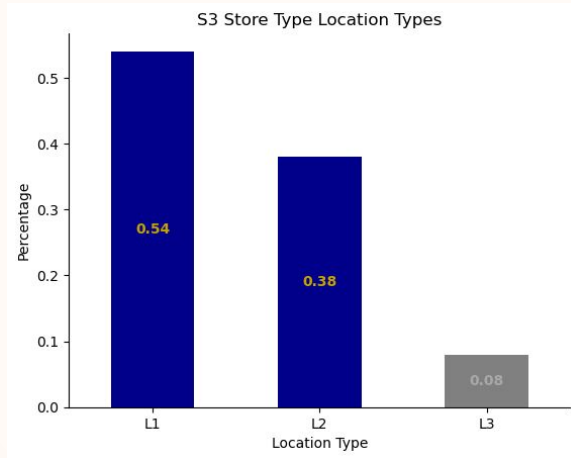
– Many S3 stores are found in **excellent spots**, mostly **L1 and L2**, which have sales above average.
– S3 stores are mostly in **R1 and R3 regions**. R1 shows good sales, but S3 stores in **R3** might need some attention as their sales are slightly **below average**.

The issue might be in the **region code**, particularly **R3**, where sales are below average.

S3 Store Location Types and Region

Comparison of Monthly Sales Across all Store Types
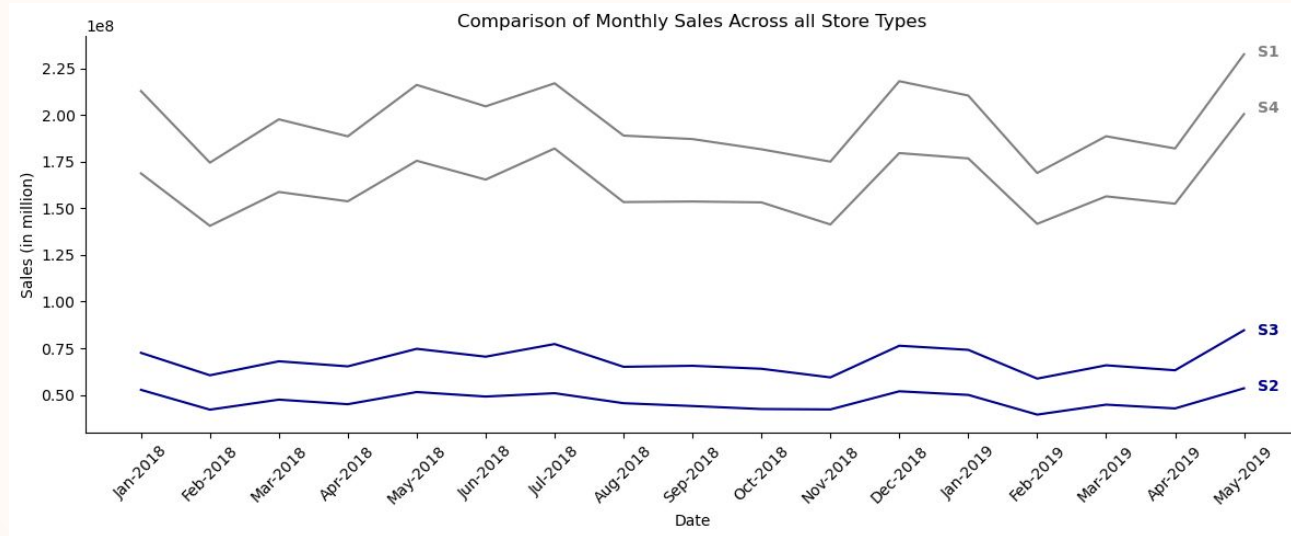
–   Looking at the monthly sales, S2 and S3 stores show a **big gap** compared to the rest. Also, the sales for S2 and S3 appear **flat**, indicating they might **not change much** with **seasons or holidays** compared to other store types.

Potential Factors

Lower than Average Sales in S2 and S3 Store Types → Less responsive to seasonal and holiday events
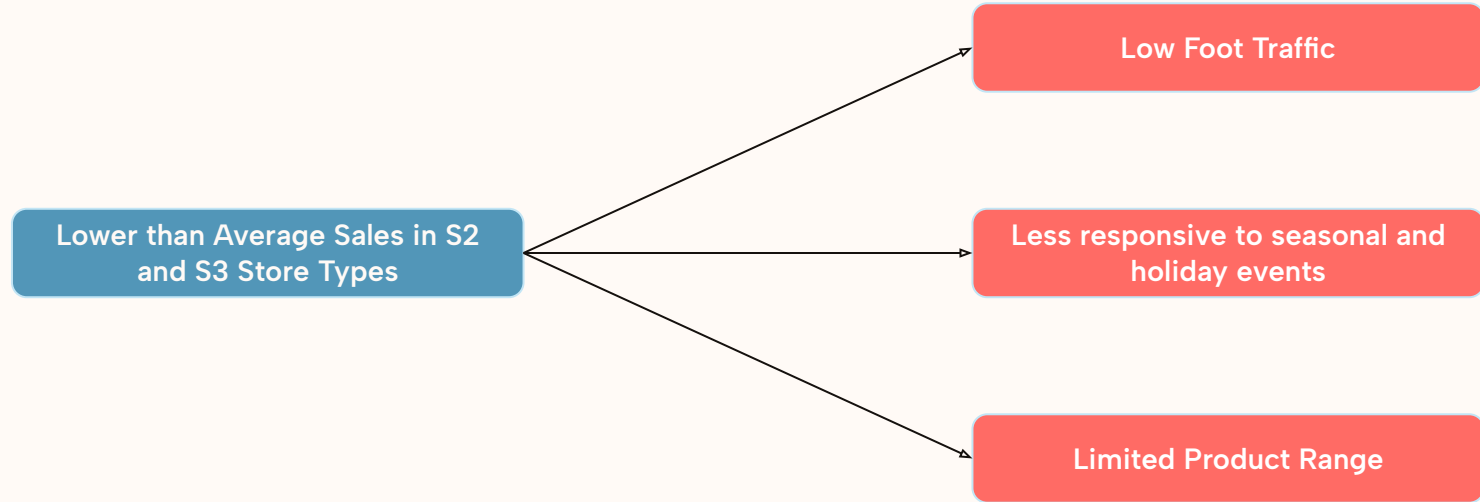
Average Order Value across all store types

Average Order Value Across all Store Types

- Store Type **S2** has **high AOV**, despite **poor sales performance** based on the sales figure below average total sales
- Store Type **S3** has **average AOV,** despite **poor sales performance** based on the sales figure below average total sales

- Store type S2 and S3 may have **lower foot traffic** or fewer customer compared to other types. This could result in fewer overall sales despite a high AOV, as there are fewer transactions.
- S2 and S3 may have **limited product range** or variety, which could discourage customers from making multiple purchases during the same visit

Potential Factors

Average Order Value across all location types
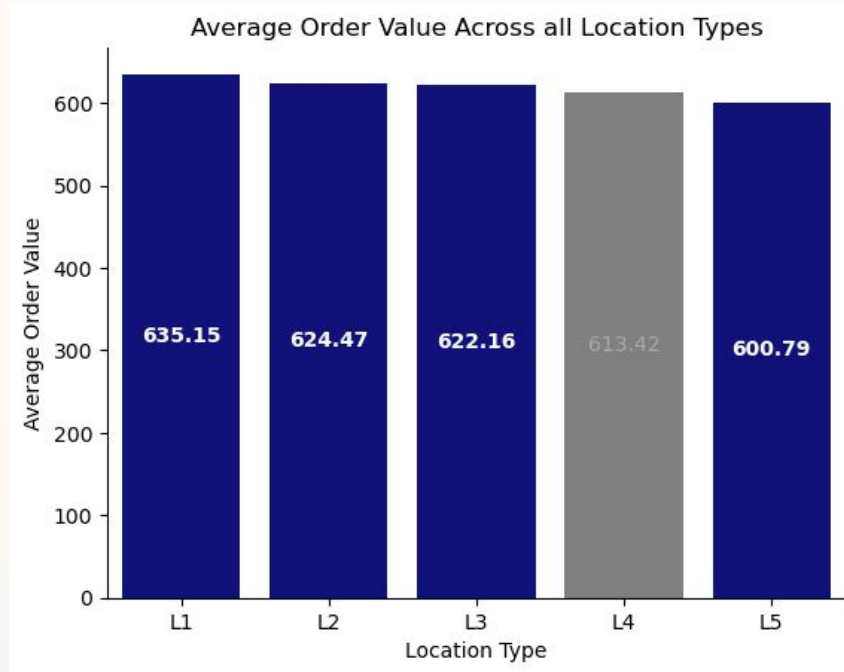
AOV for S2 ($680) and S3 ($639) **exceeds their location's AOV**. But, sales figures for both are still below average. *Why?*

– S2 and S3 may **not match local preferences**, leading to lower foot traffic despite higher spending per customer.
– S2 and S3 face **competition from nearby stores**, requiring a detailed look at competitors' offerings and prices.

Potential Factors

Lower than Average Sales in S2 and S3 Store Types

Product Misalign with Local Preferences

Low Foot Traffic

Less responsive to seasonal and holiday events

Limited Product Range

Competitive Landscape

Average Order Value across all regions
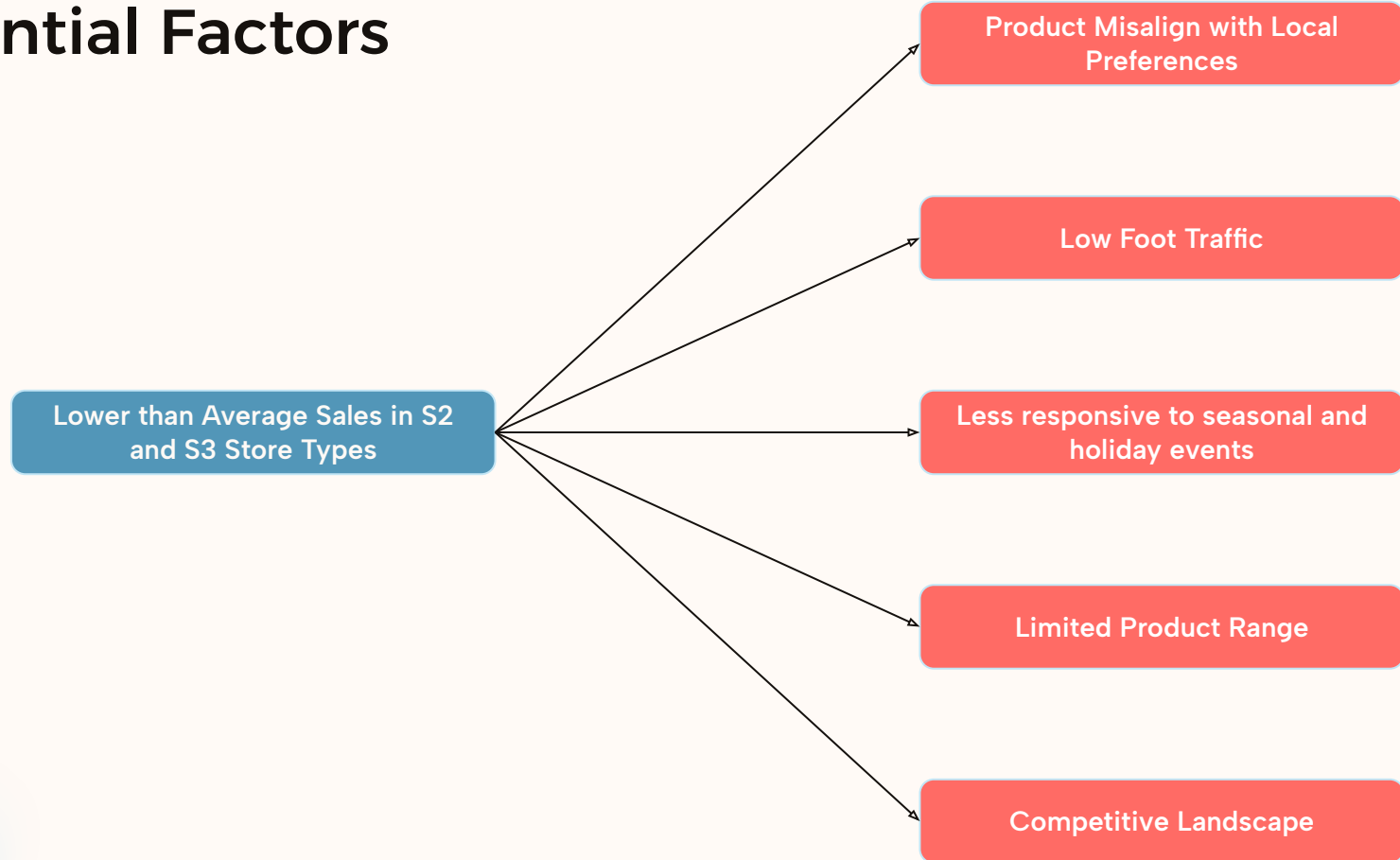
Average Order Value Across all Region Codes

Store types S2 and S3 have **higher AOV than their region**, but lower sales than the average. *Why?*

– S2 and S3 **attract high–AOV customers**, but have a **smaller overall customer base**, potentially leading to lower sales.

Potential Factors

Lower than Average Sales in S2 and S3 Store Types

- Limited Market Reach
- Product Misalign with Local Preferences
- Low Customer Traffic
- Less responsive to seasonal and holiday events
- Limited Product Range
- Competitive Landscape
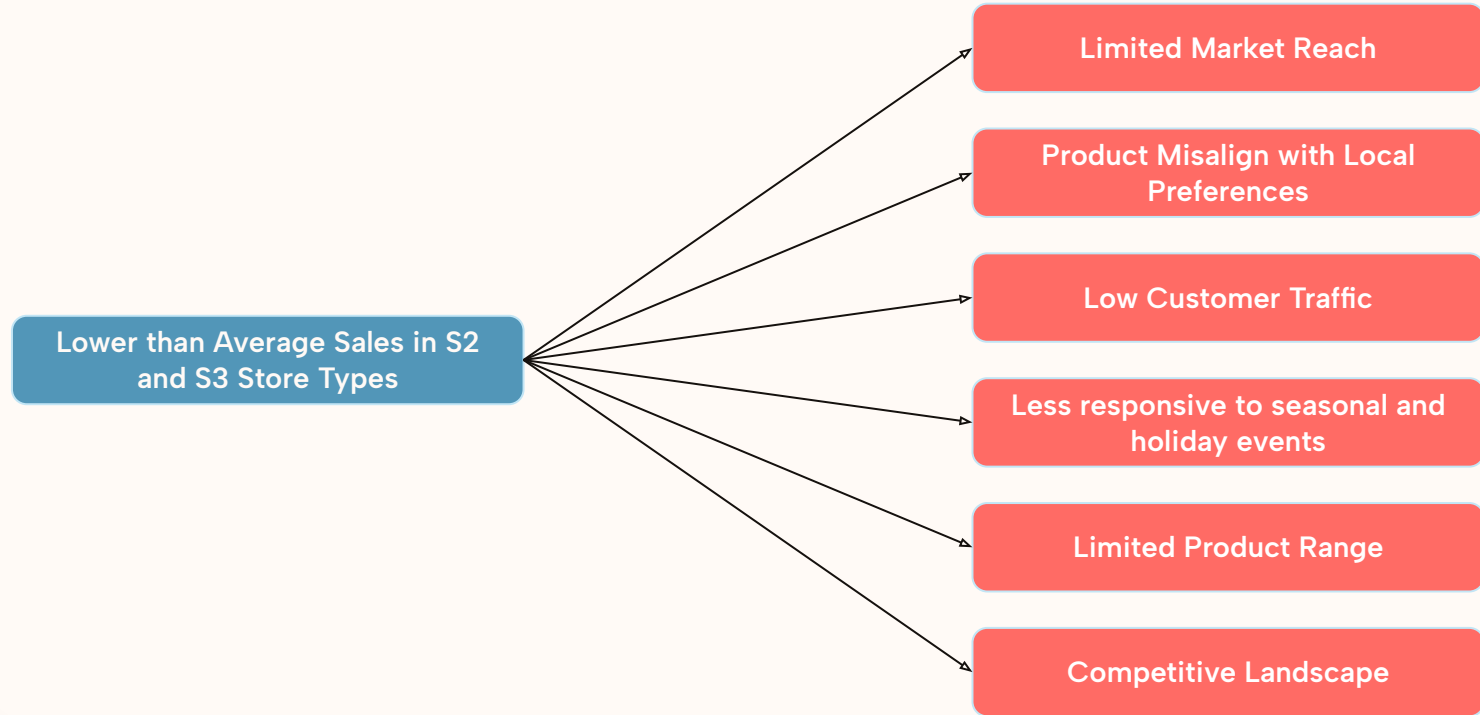
So, **What can we do**?

Business Recommendations

Factor: Limited Market Reach

Online Visibility

Make S2 and S3 stores more visible online. Use local SEO to ensure they show up in local search results.

Also, use social media ads (like on Facebook and Instagram) to reach people in the areas where they are.

Referral Program

Start programs where happy customers can tell their friends and family about S2 and S3 stores.

Give them rewards for bringing in new customers. This way, more people will know about the stores through word-of-mouth.

Business Recommendations

Factor: Product Misalign with Local Preferences

Market Research

Research local preferences in L1, L2, L3, and L5 to align store products better.

May include popular product categories, brands, and features that resonate with the local population to better align the stores' products.

Local Partnerships

Collaborate with local suppliers, artisans, or producers to source products that are unique to the region.

Highlighting locally-made items can attract customers looking for authentic and distinctive offerings

Business Recommendations

Local Marketing

Run local marketing campaigns in S2 and S3 communities, like community–focused social media contests with prizes.

For example, run local social media contests with prizes for S2 and S3 communities.

Special Events

Host special events or promotions that encourage foot traffic, such as weekend sales events, new product launches, or store anniversaries.

We can collaborate with local influencers or organizations to drive traffic

Business Recommendations

Factor: Less Responsive to Seasonal and Holiday Events

Seasonal Merchandising

Create seasonal plans for S2 and S3 stores. Select products, promotions, and ads for each season and holiday.

Make sure these plans match what local customers want.

Holiday-Specific Promotions

Offer holiday promotions like '12 Days of Christmas Discounts.'

Each day, a different product gets a bigger discount as we approach Christmas.

Extended Holiday Hours

Extend store hours during holidays, accommodating customers who prefer to shop outside regular business hours.

For example, open the store from normal opening hours to midnight at Halloween Day

Business Recommendations

Factor: Limited Product Range

Collaborations

Collaborate with local food producers, artisans, or suppliers for unique S2 and S3 products.

Customers can sample and purchase locally-made food and beverages

Customer Feedback

Gain feedback from customers of S2 and S3 stores about the products they'd like to see in stores through the use of feedback stations or online surveys with gifts.

Carefully review and align the product accordingly to the store plan

Business Recommendations

Factor: Competitive Landscape

Competitor Assessment

Research local preferences in L1, L2, L3, and L5 to align store products better.

This may include popular product categories, brands, and features that resonate with the local population to better align the stores' products.

Pricing Strategy

Adjust pricing strategies as needed to remain competitive while maintaining healthy profit margins.

Consider running promotions or discounts on select products to attract price-sensitive customers

Thanks!