



Super Store Sales Data

Data Source

This data is accessed from
[https://www.kaggle.com/datasets/aditisaxena20/superstore-sales-dataset?
resource=download](https://www.kaggle.com/datasets/aditisaxena20/superstore-sales-dataset?resource=download)

Introduction

What kind of data we have?

order_id	order_date	ship_date	ship_mode	customer_name	segment	country	region	category	sub_category	product_name
AG-2011-2040	Jan 1, 2011	Jun 1, 2011	Standard Class	Toby Braunhardt	Consumer	Algeria	Africa	Office Supplies	Storage	Tenex Lockers, Blue
IN-2011-47883	Jan 1, 2011	Aug 1, 2011	Standard Class	Joseph Holt	Consumer	Australia	Oceania	Office Supplies	Supplies	Acme Trimmer, High Speed
HU-2011-1220	Jan 1, 2011	May 1, 2011	Second Class	Annie Thurman	Consumer	Hungary	EMEA	Office Supplies	Storage	Tenex Box, Single Width

sales	quantity	discount	profit	shipping_cost
408	2	0	106.14	35.46
120	3	0.1	36.04	9.72
66	4	0	29.64	8.17

Introduction

Column That We have:

- 1. order_id
- 2. order_date
- 3. ship_date
- 4. ship_mode
- 5. customer_name
- 6. segment
- 7. state
- 8. country
- 9. market
- 10. region
- 11. product_id
- 12. category
- 13. sub_category
- 14. product_name
- 15. sales
- 16. quantity
- 17. discount
- 18. profit
- 19. shipping_cost
- 20. order_priority
- 21. year

Introduction

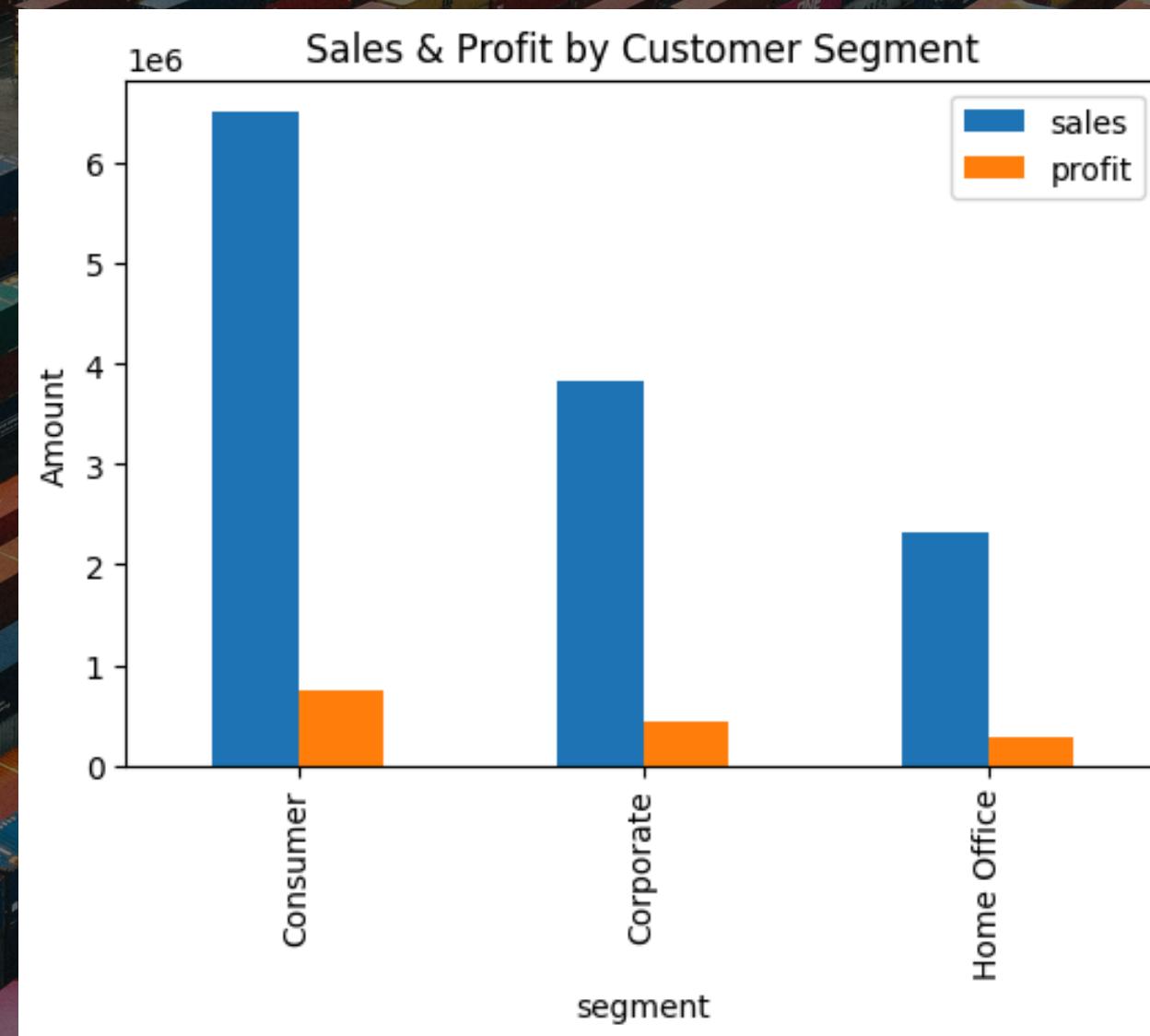
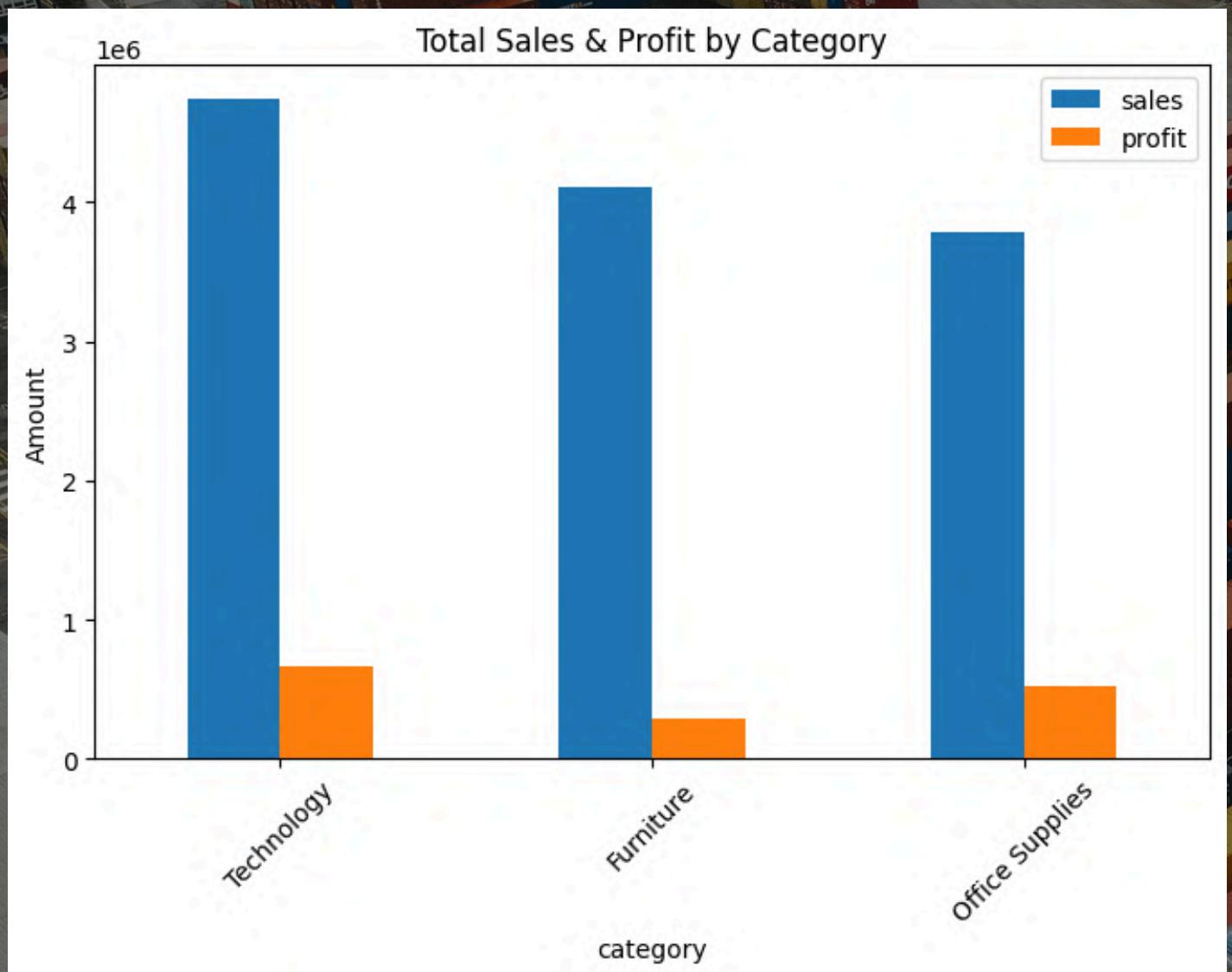
Column That We have:

- 1. order_id
- 2. order_date
- 3. ship_date
- 4. ship_mode
- 5. customer_name
- 6. segment
- 7. state
- 8. country
- 9. market
- 10. region
- 11. product_id
- 12. category
- 13. sub_category
- 14. product_name
- 15. sales**
- 16. quantity
- 17. discount
- 18. profit**
- 19. shipping_cost
- 20. order_priority
- 21. year

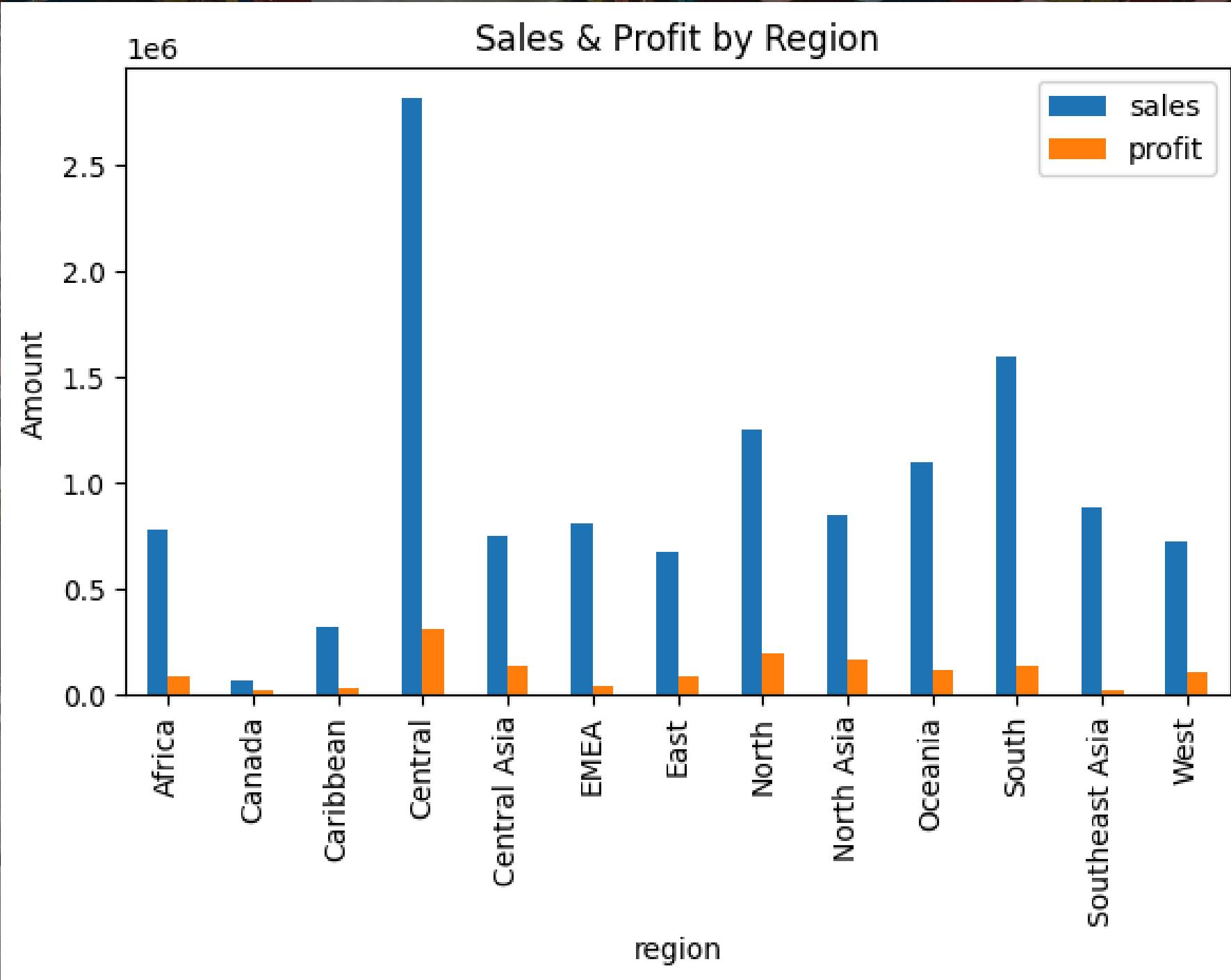
Of Course!

The main importance here is sales and profit

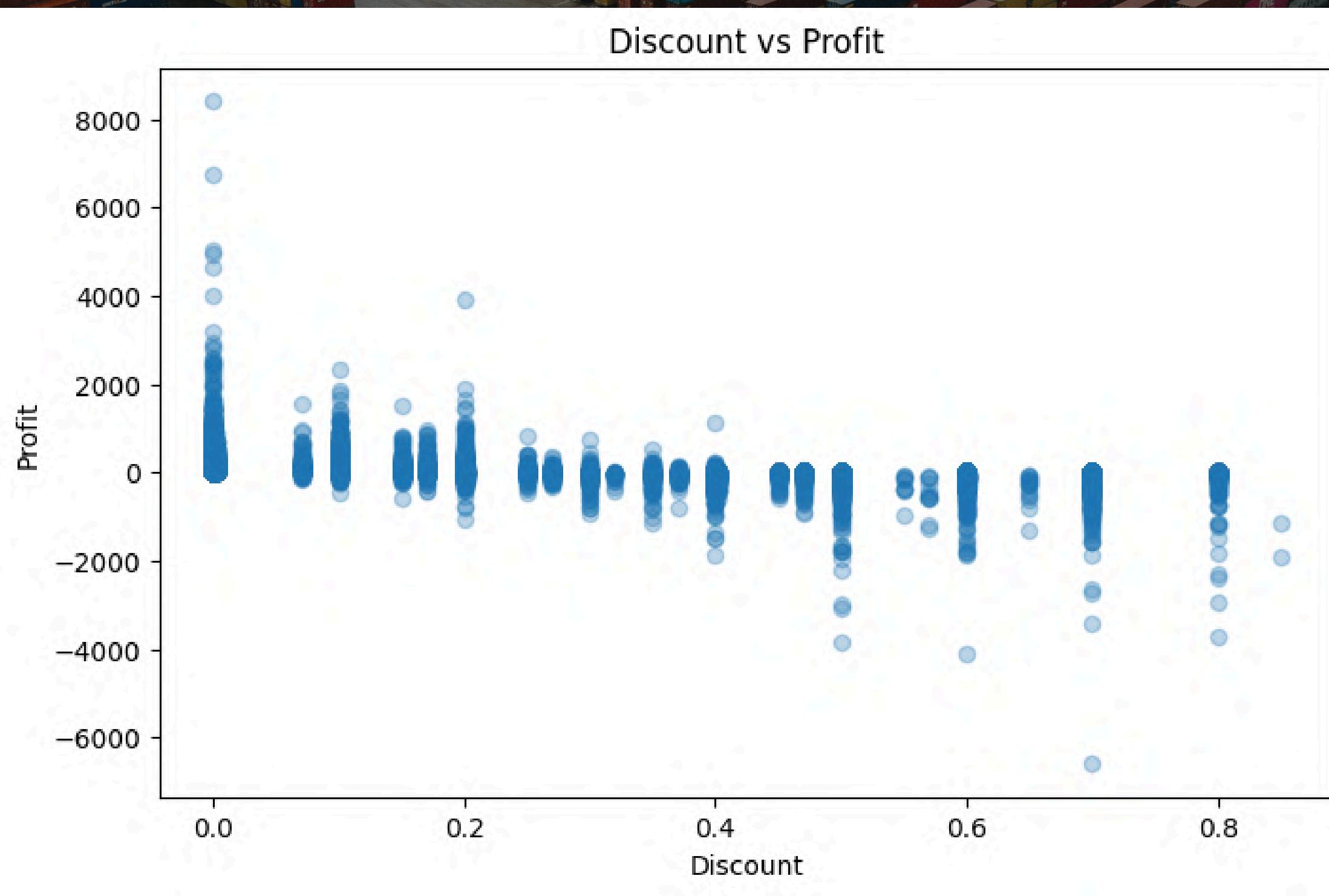
Raw Graph Introduction



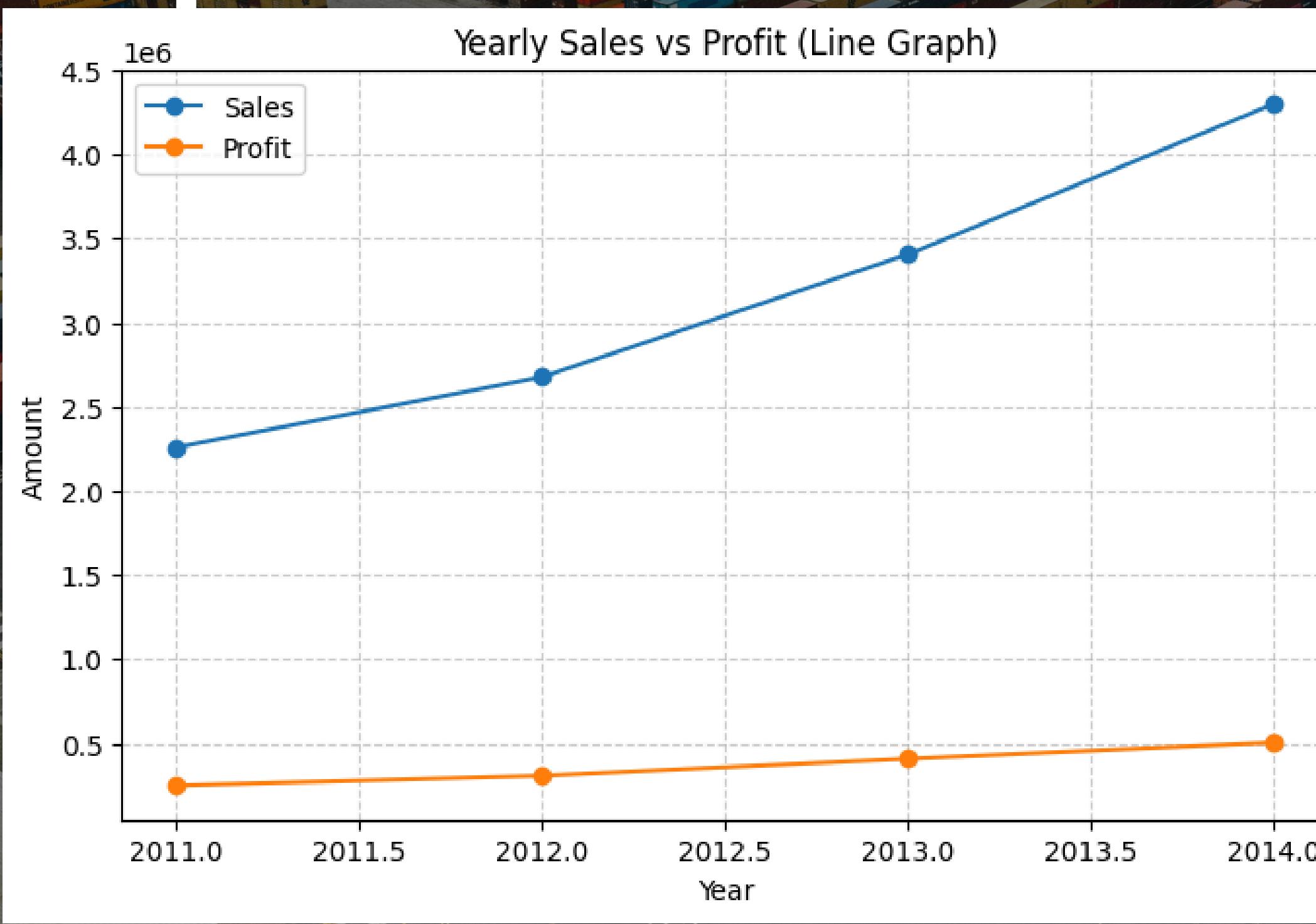
Raw Graph Introduction



Raw Graph Introduction



Raw Graph Introduction

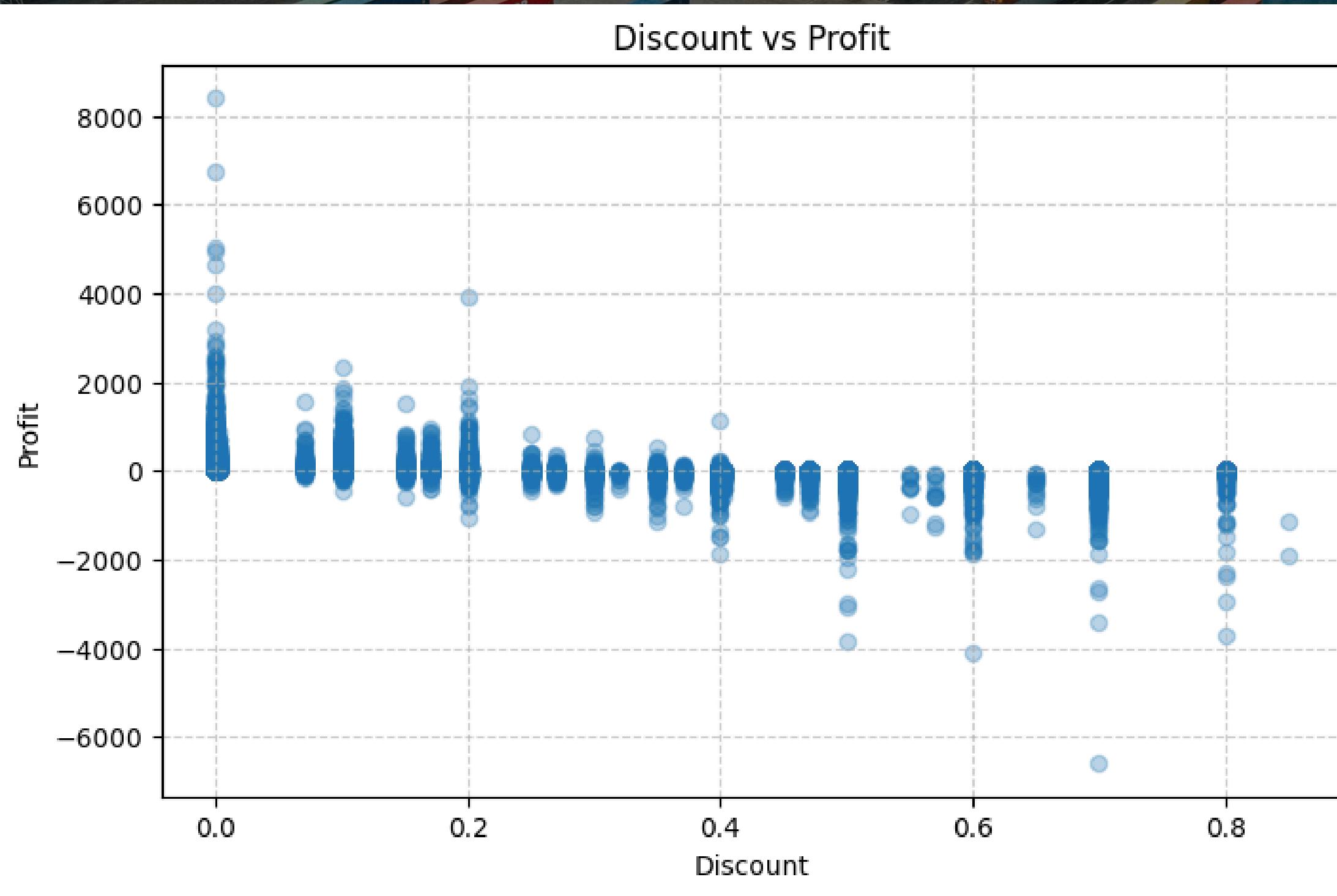


Hypothesis based on the data

- **Discounts** → maybe too aggressive, eroding margins.
- **Product category mix** → some categories (e.g., Furniture) bring sales but weak profit.
- **Regional costs** → certain markets (e.g., Southeast Asia) have high sales but very low margins, likely from shipping or local pricing.

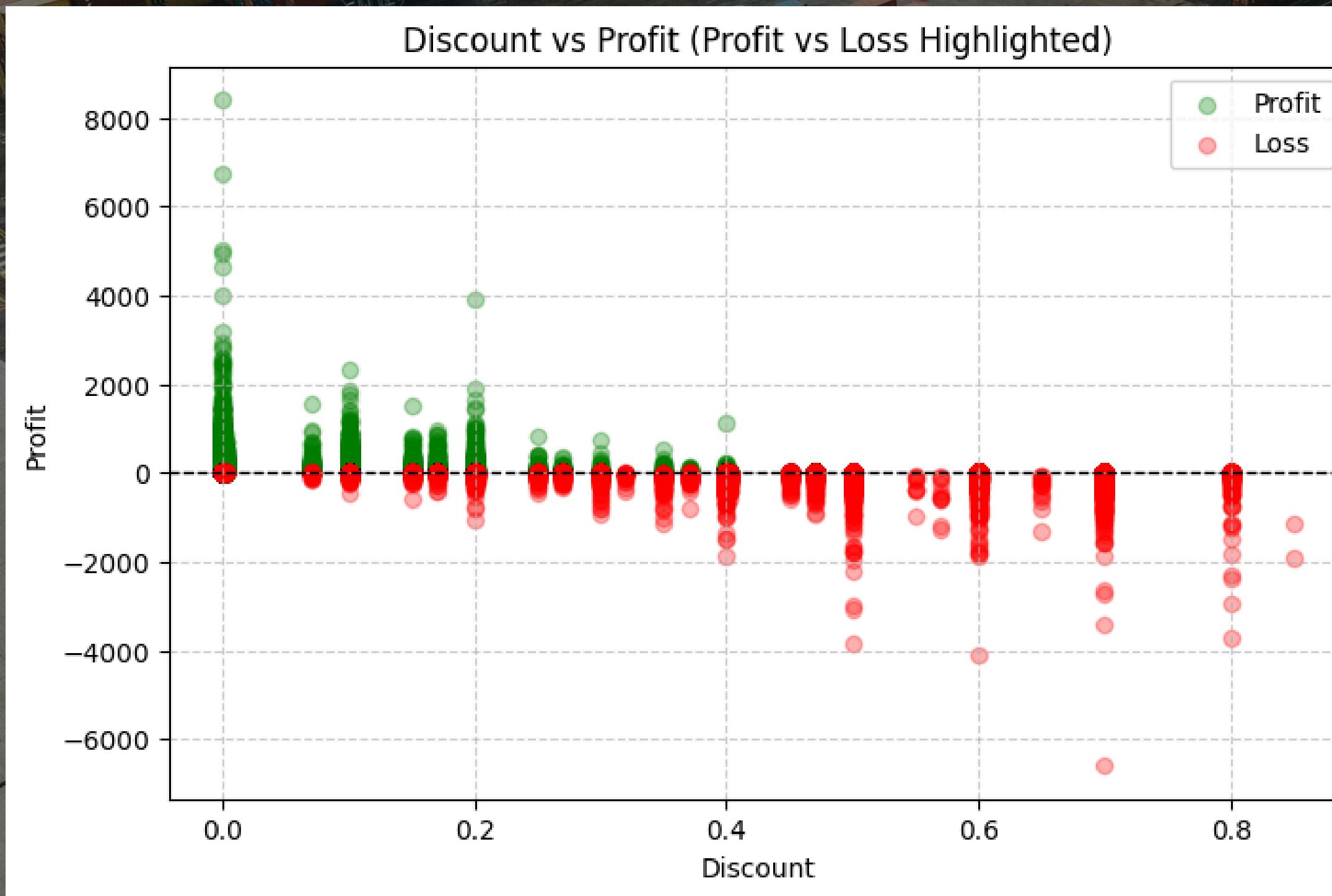
Let's start from discount

Anything weird?



Let's start from discount

How about now?



discounts lead to profit vs loss:

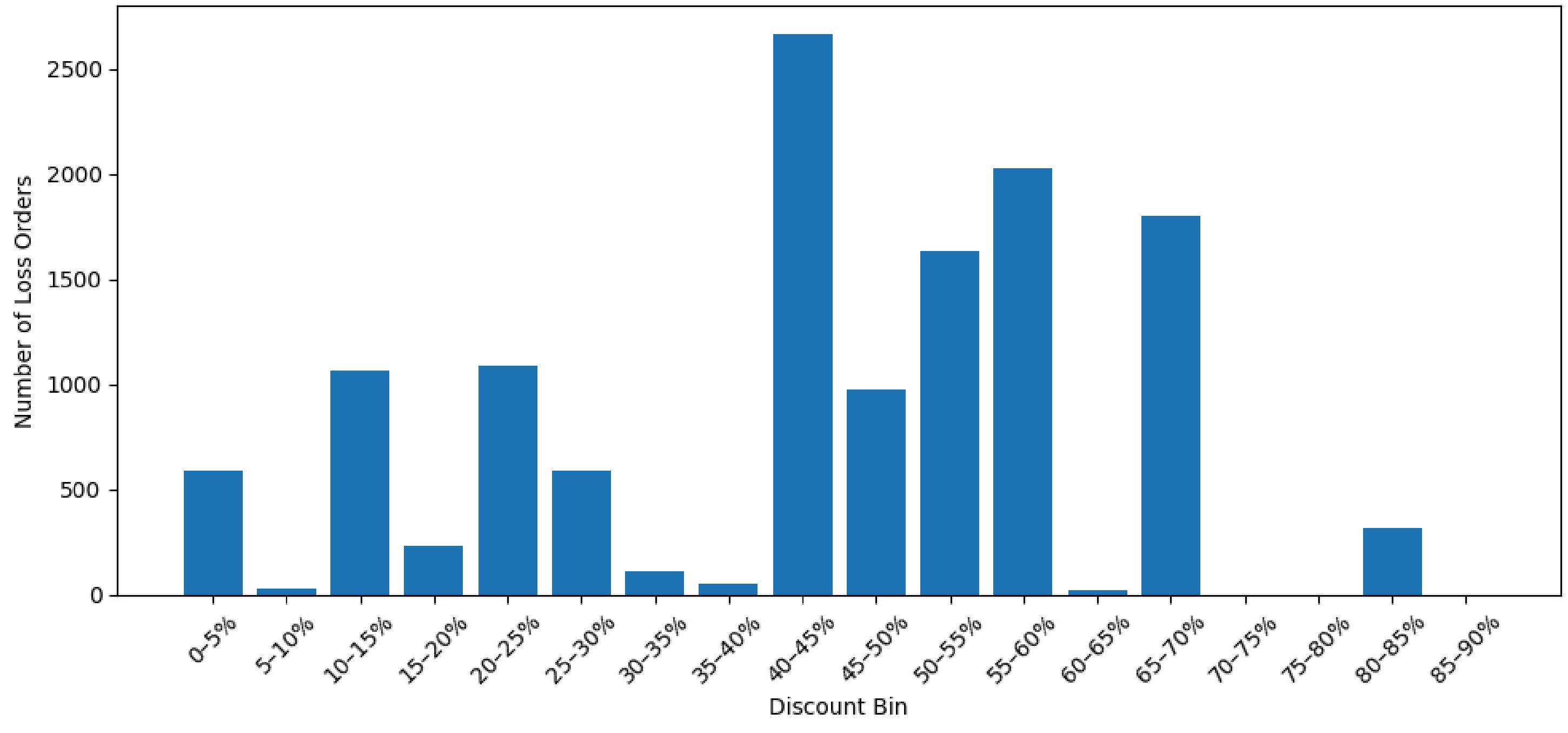
- Green points = orders with **profit ($\approx 38,079$ orders)**.
- Red points = orders with **losses ($\approx 13,211$ orders)**.

Key Numbers

- About **25.8% of all orders are loss-making**.
- Losses are heavily concentrated at higher discount levels.
- Profitable orders cluster mostly in the **0–20% discount range**. (is this statement correct?)

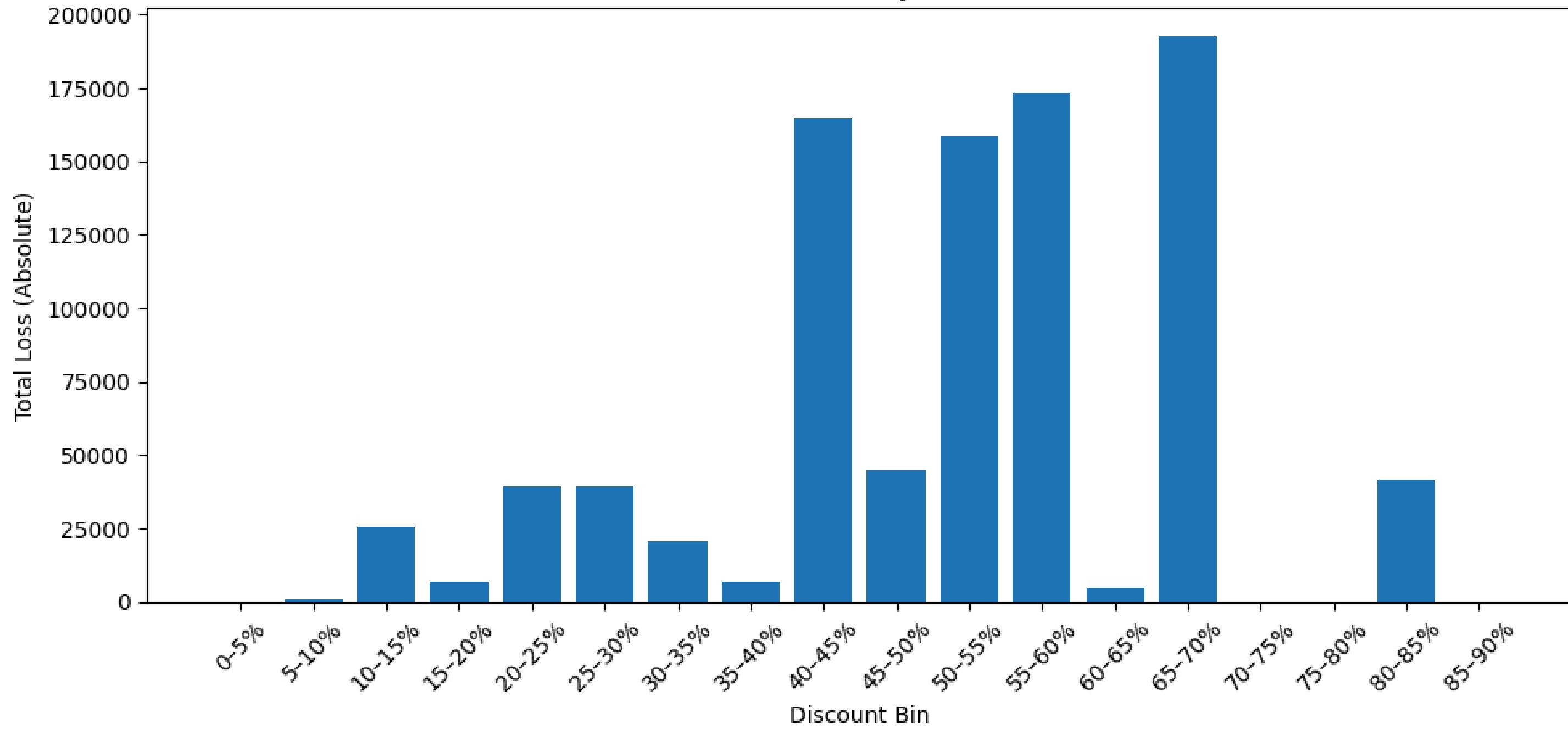
Discount Bin

Count of Loss-Making Orders by Discount Bin



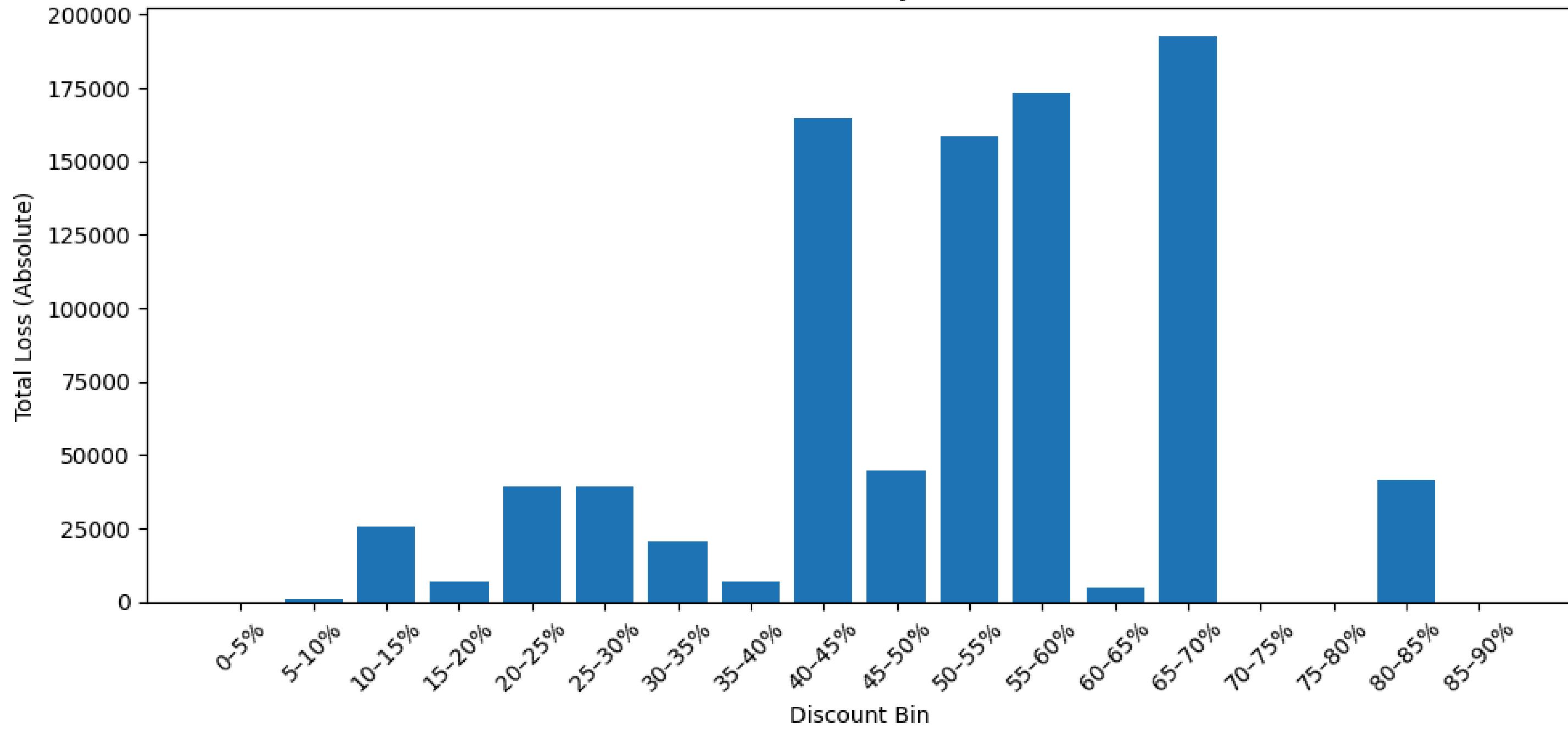
Discount Bin

Total Loss Amount by Discount Bin



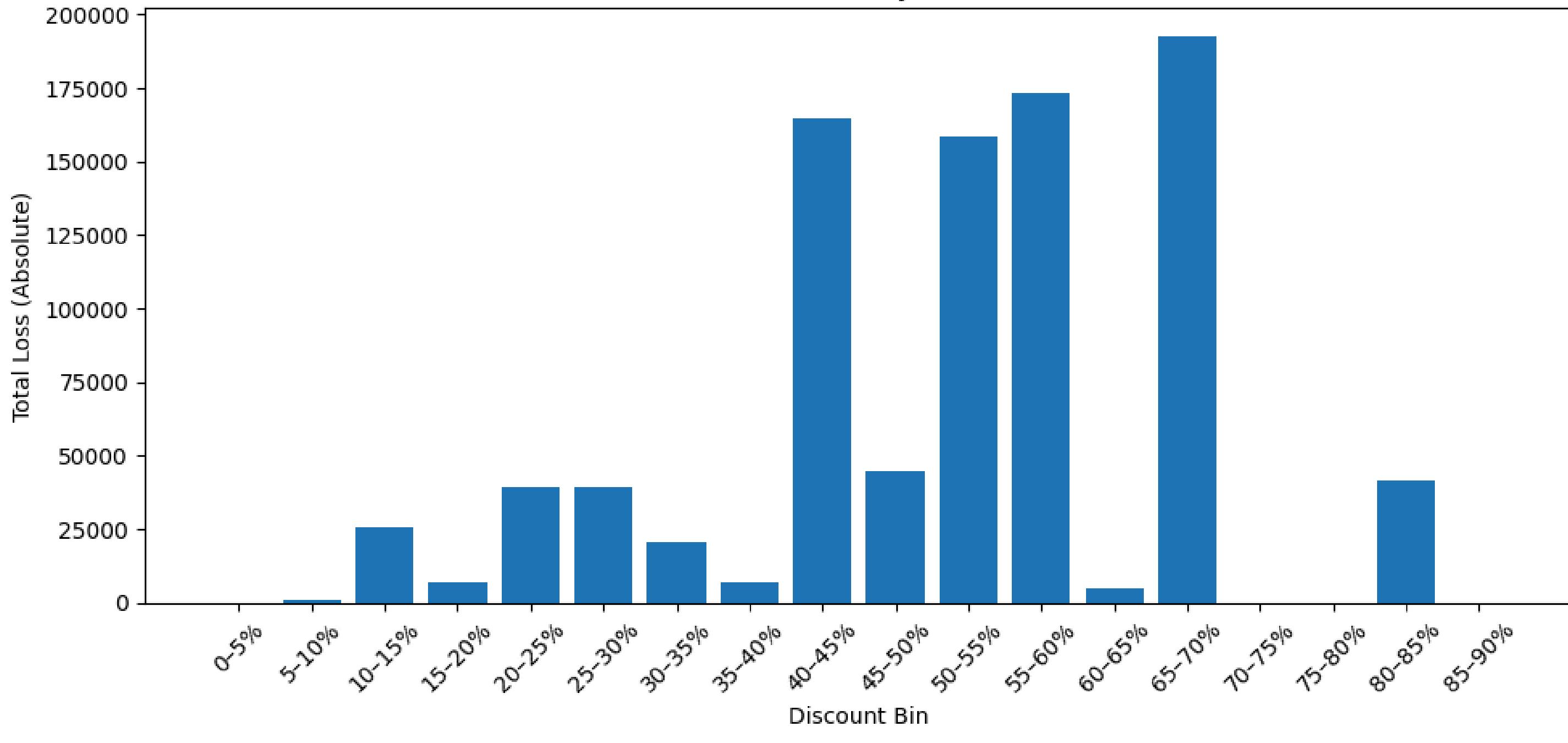
Discount Bin

Total Loss Amount by Discount Bin



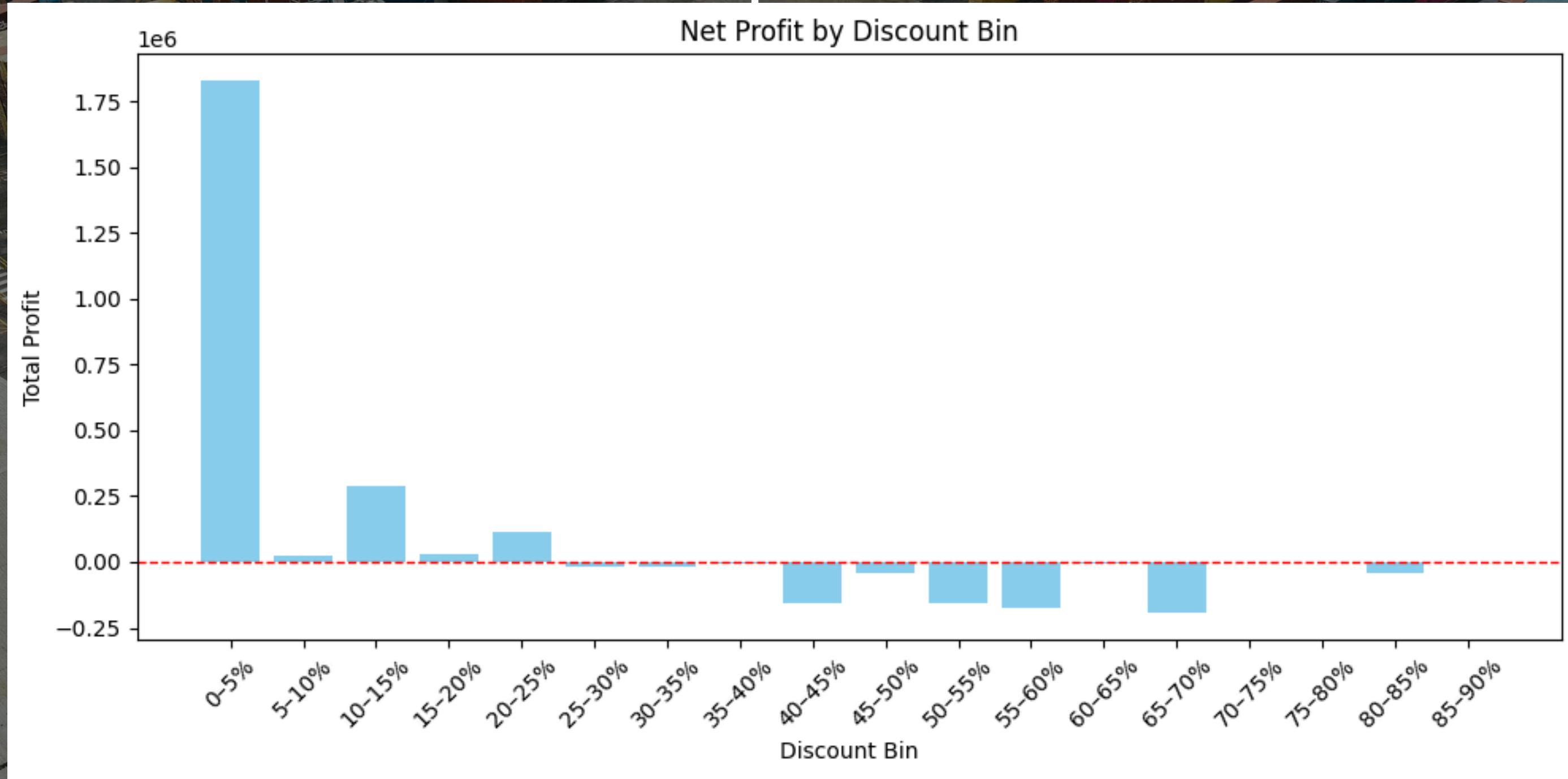
Discount Bin

Total Loss Amount by Discount Bin



Discount Bin

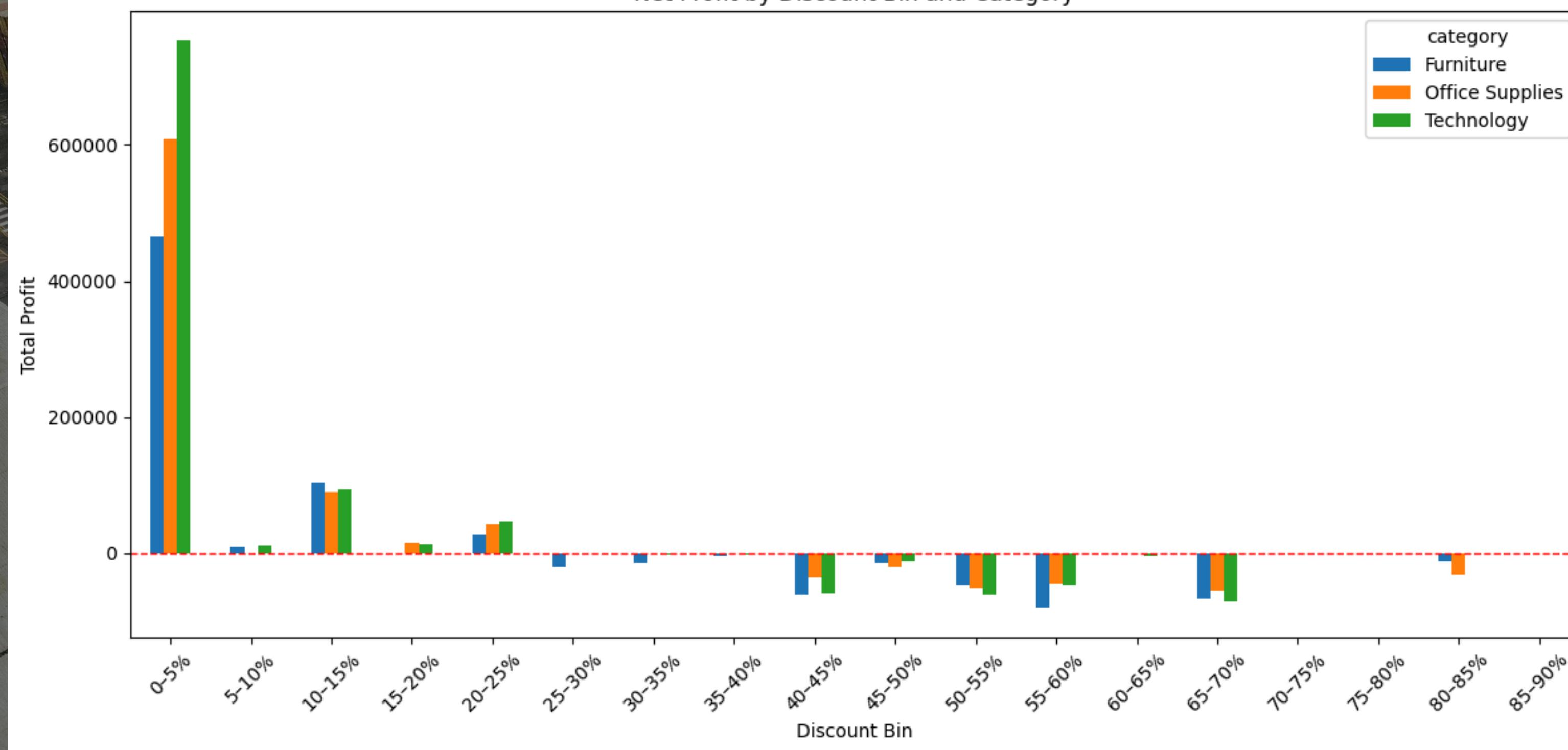
Before we have loss, but how about profit?



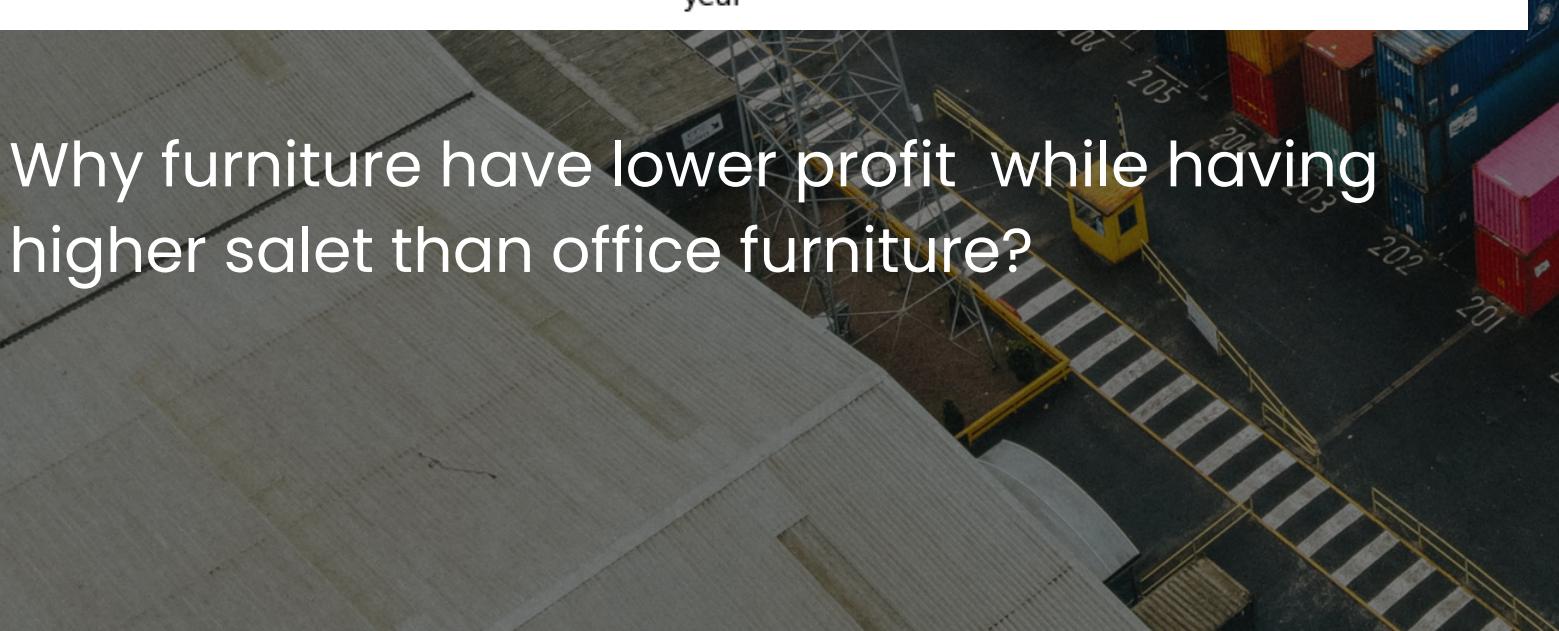
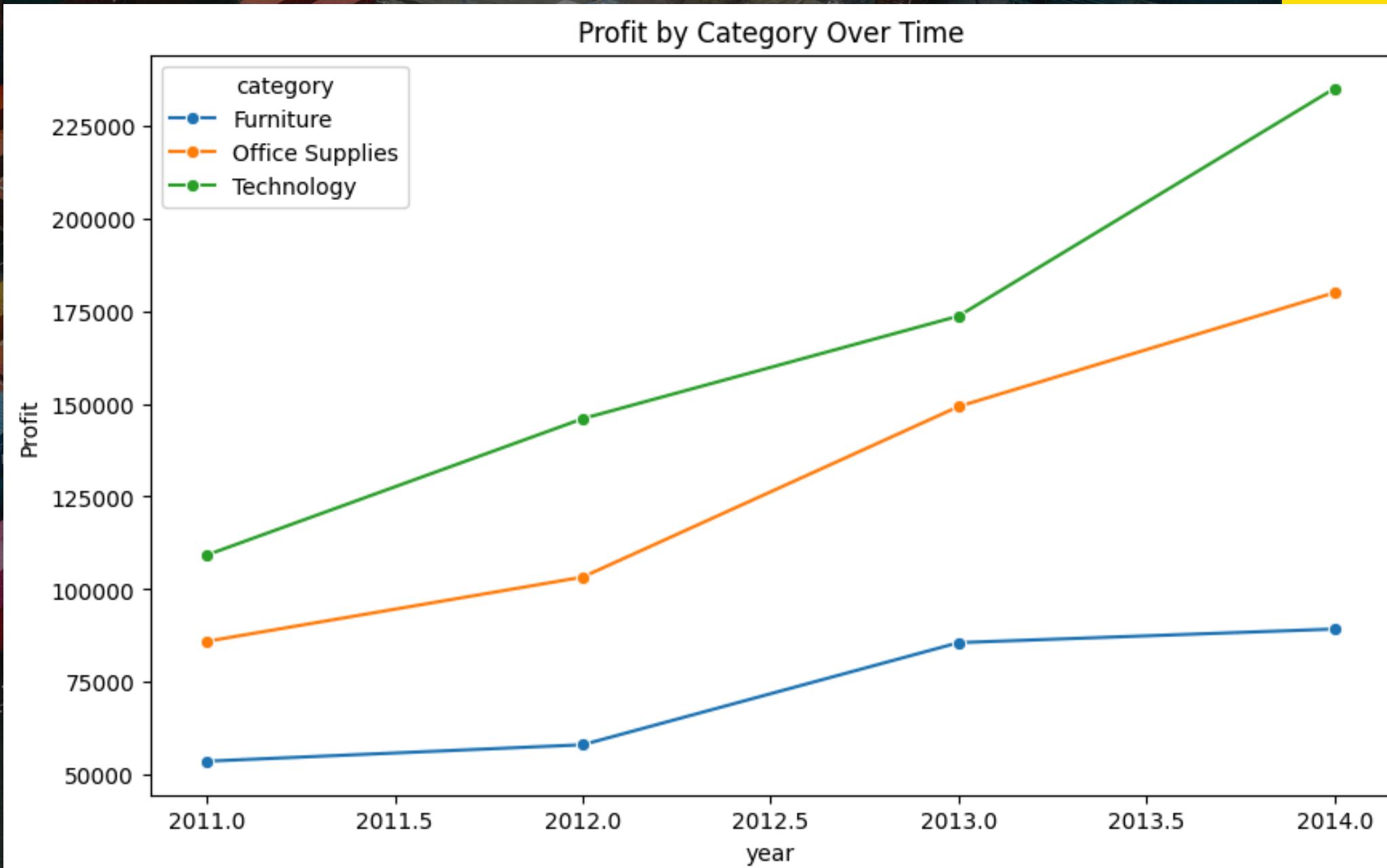
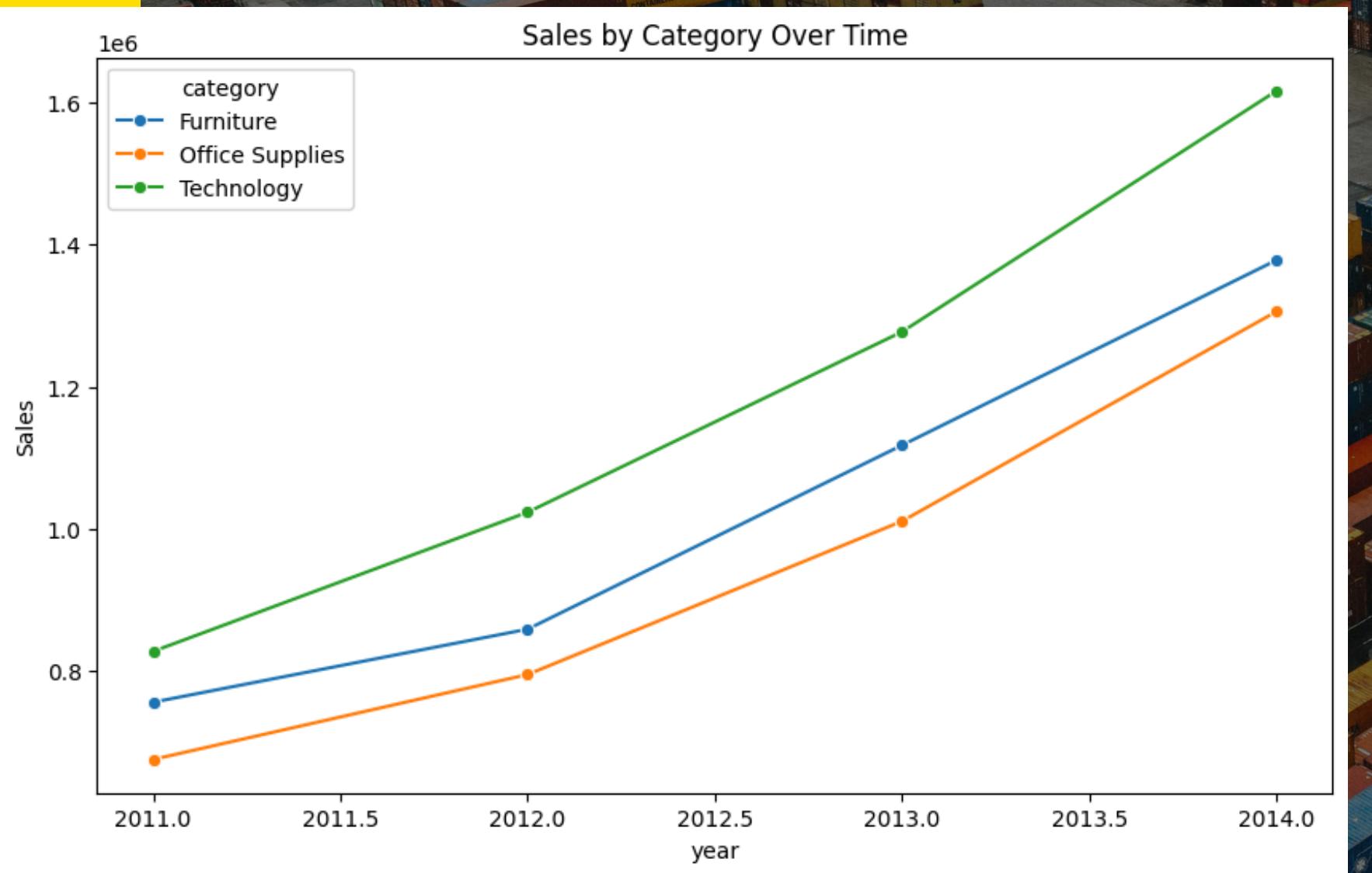
Discount Bin

What if it has connection with the category?

Net Profit by Discount Bin and Category

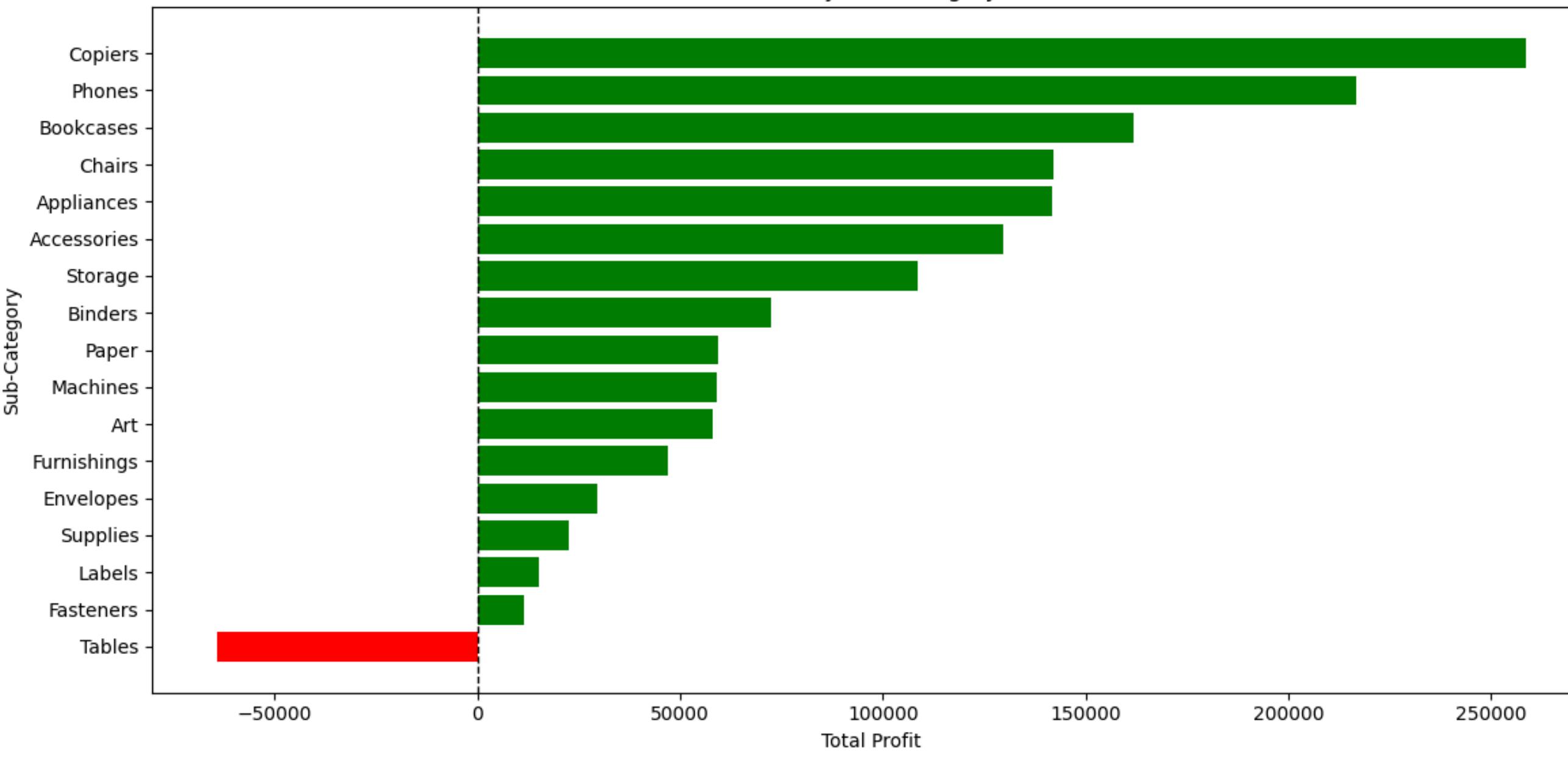


Let's dive into category

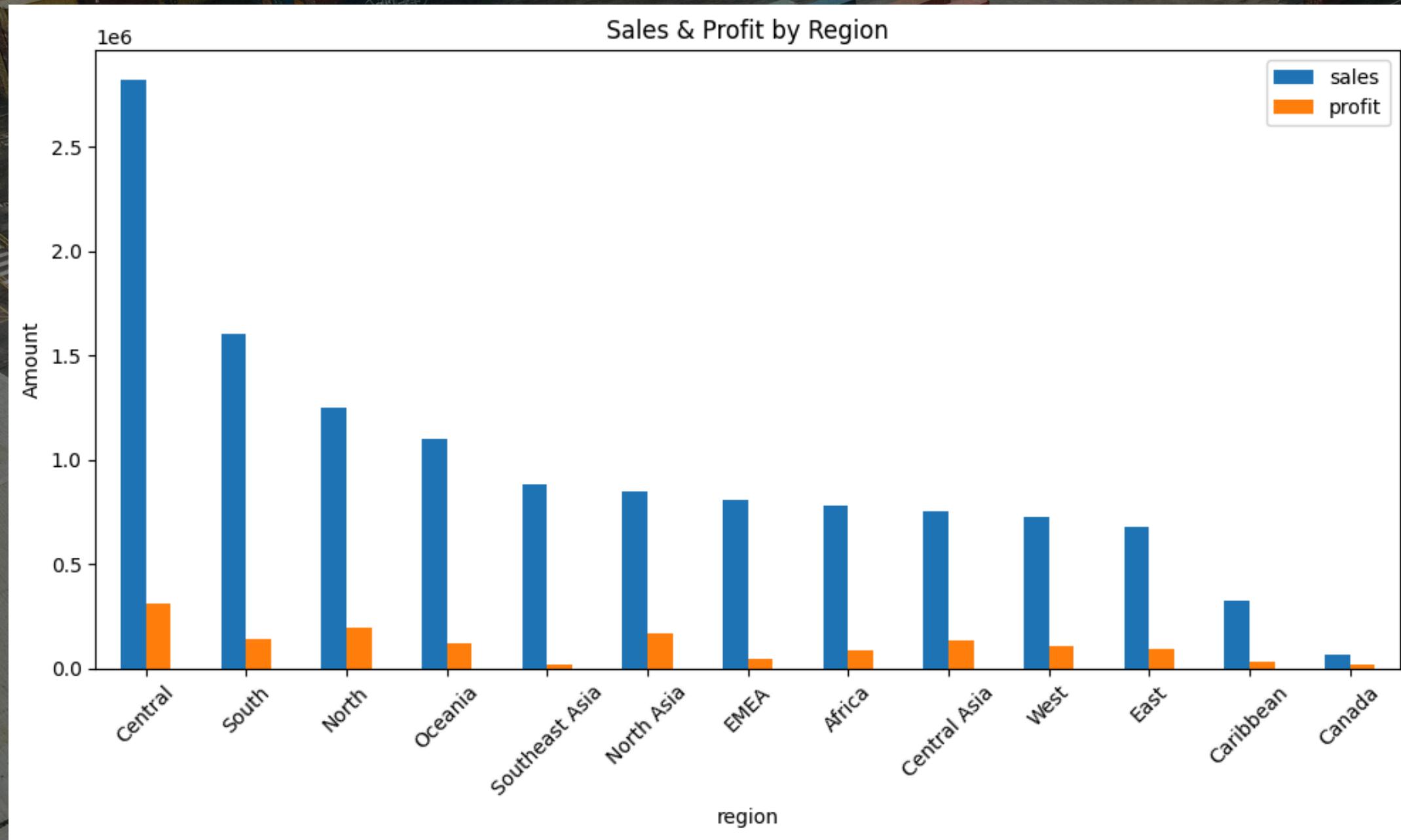


Sub Category Profit!

Profit by Sub-Category



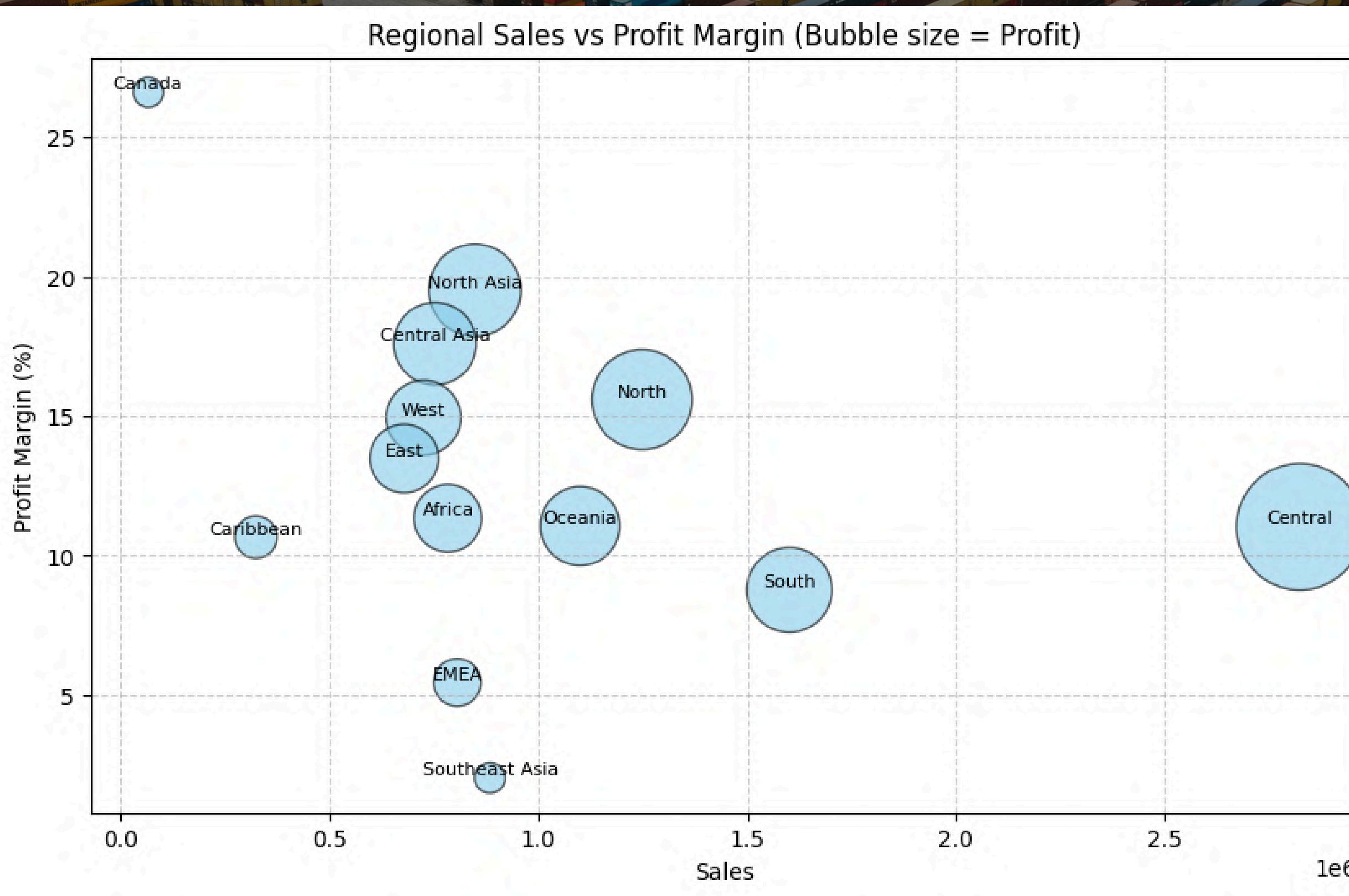
What about the connection with region?



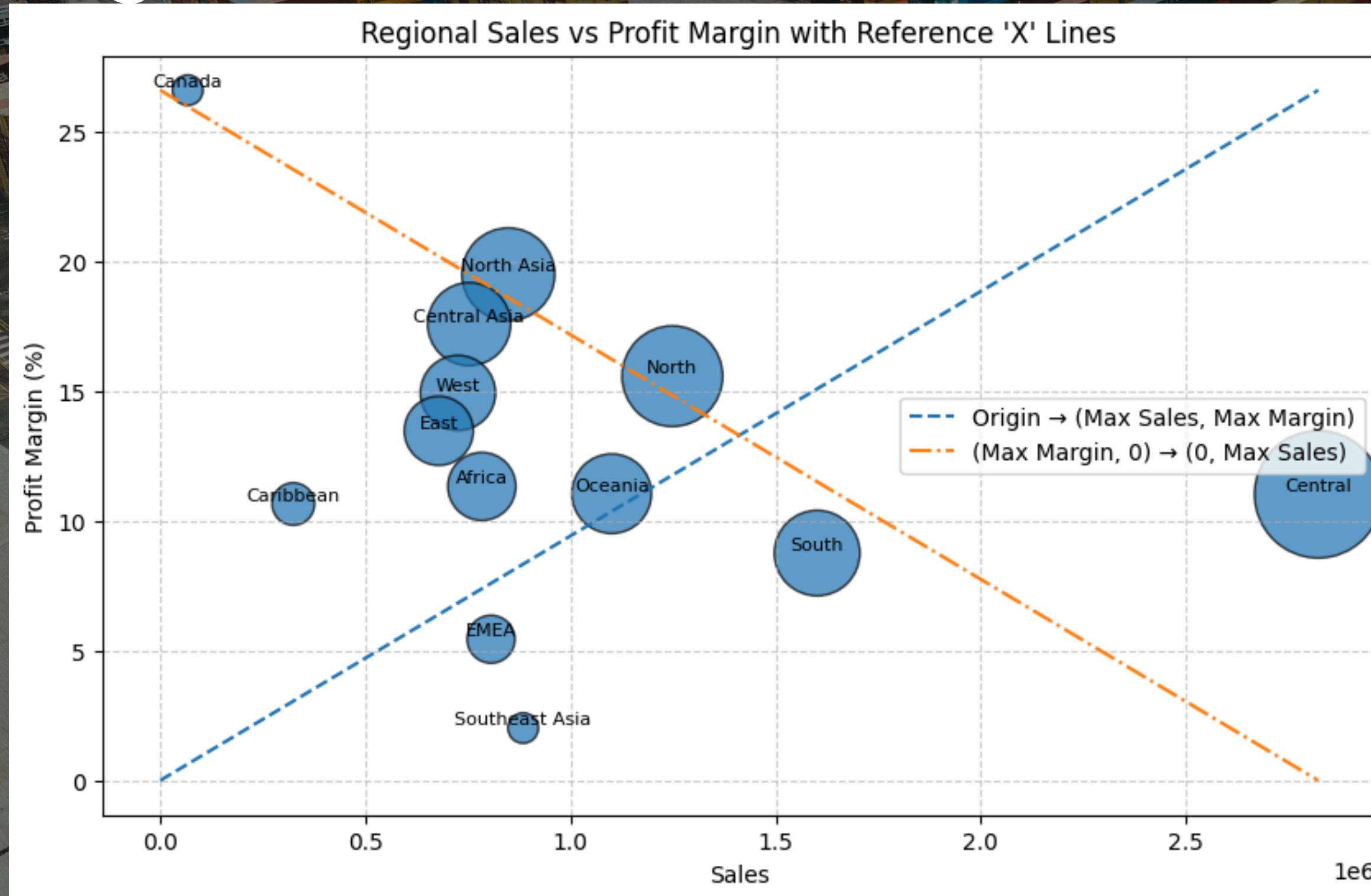
Where profit margin = profit / sales * 100

region	sales	profit	profit_margin_%
Central	2822399	311403.98164	11.033308
South	1600960	140355.76618	8.766975
North	1248192	194597.95252	15.590386
Oceania	1100207	121666.64200	11.058523
Southeast Asia	884438	17852.32900	2.018494
North Asia	848349	165578.42100	19.517725
EMEA	806184	43897.97100	5.445155
Africa	783776	88871.63100	11.338907
Central Asia	752839	132480.18700	17.597413
West	725514	108418.44890	14.943674
East	678834	91522.78000	13.482351
Caribbean	324281	34571.32104	10.660915
Canada	66932	17817.39000	26.620137

What about the connection with region?



What about the connection with region?

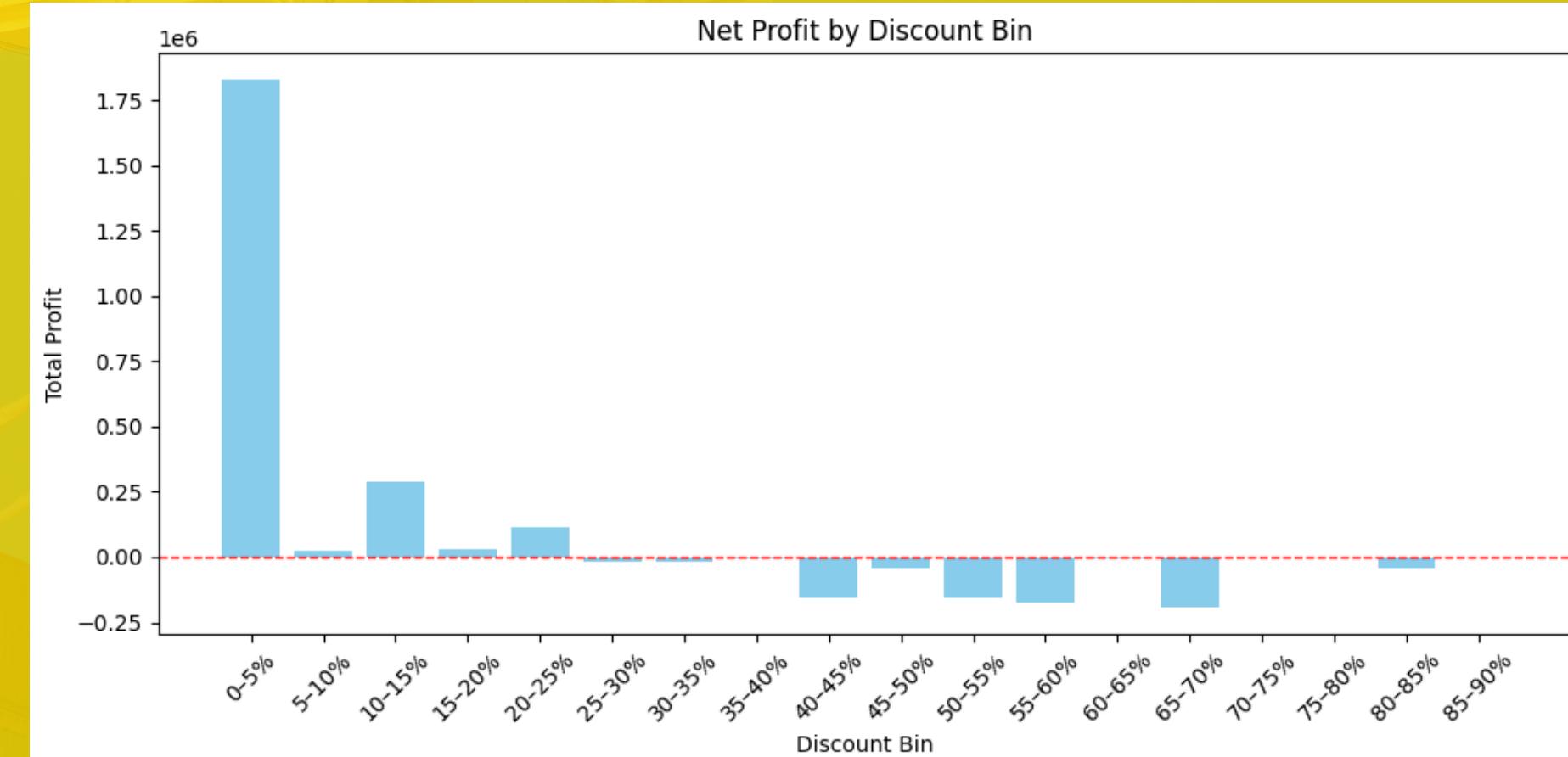


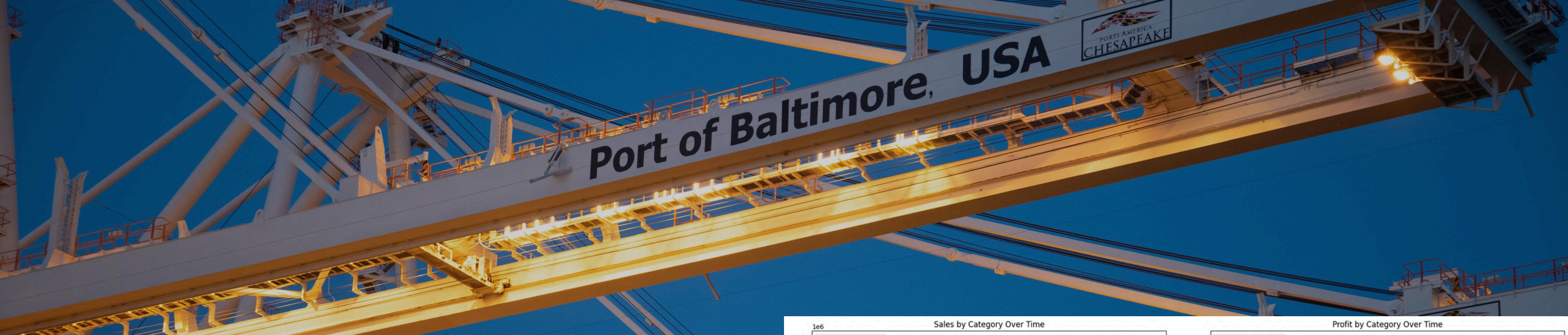
- Central → high sales, moderate margin (~11%), large bubble = **big absolute profit**.
- North Asia & Central Asia → smaller sales but very high margins (~18-20%) → **efficient and profitable; good for expansion**.
- Southeast Asia → medium sales but terrible margin (~2%) → big **underperformer**.
- Canada → tiny sales but extremely high margin (~27%) → **niche but very efficient**.
- North & Oceania → **solid balance, good mix of sales and margin**.



Hypothesis one: Discounts

- Evidence:
 - Scatter plot + bar charts showed that discounts above 20–25% flip orders into losses.
 - Net Profit by Discount Bin: Profitable up to 20%, but losses in the 40–45% range.
- Conclusion:
 - Discounts are a major driver of profit leakage.
 - Company should cap discounts at ~20%, and treat higher discounts only as exceptions (e.g., clearance).





Hypothesis two: Product Category Mix

- Evidence:
 - Furniture sales are high but profit is weak .
 - Sub-category breakdown shows Tables are loss-making, dragging Furniture down.
 - In contrast, Phones, Copiers, Chairs, Bookcases, Appliances are strong profit generators.
- Conclusion:
 - Profit problems are not across all Furniture, just in certain sub-categories.
 - Company should fix or drop unprofitable sub-categories (Tables) while doubling down on profitable ones.

