



# REGIONAL SALES EDA

By Amy Smith

# PROBLEM STATEMENT:

Sales teams often lack a clear, data-driven understanding of regional performance, making it difficult to identify growth opportunities and optimize resources. This project aims to analyse and visualize regional sales data to uncover trends, evaluate profitability, and support strategic decision-making.

## **Business Questions:**

Inconsistent revenue and profit performance across U.S. regions

Lack of visibility into seasonal swings, top SKUs, and channel profitability

**Goal:** Leverage 5 years of historical data to pinpoint growth levers and optimize strategy

# APPROACH AND PROJECT WORKFLOW

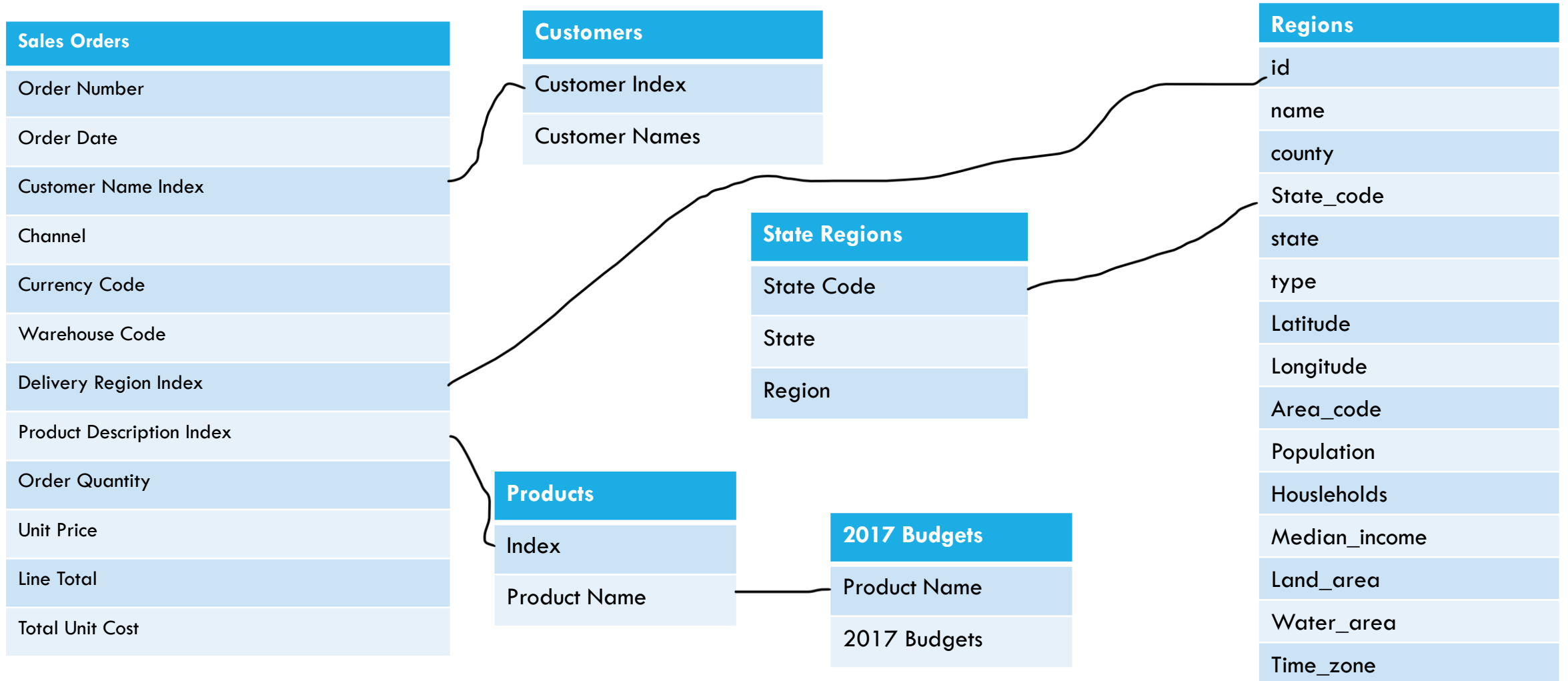
The general approach with this EDA is to first import the data into Google Colab, then read in the individual worksheets as individual data frames.

From there we can conduct data understanding and potential processing/cleansing by checking for null entries and merge lists together based on the presence of unique identifying keys between tables. We can also derive certain fields like profit.

We then conduct an EDA using several Python libraries to explore the data and extract insights centred around the primary business questions identified before summarising potential recommendations.

# DATA OVERVIEW

In this case of this data-set, our data has been spread across 6 datasets with the following features:



# EXPLORATORY DATA ANALYSIS

We have a count of 64,104 records

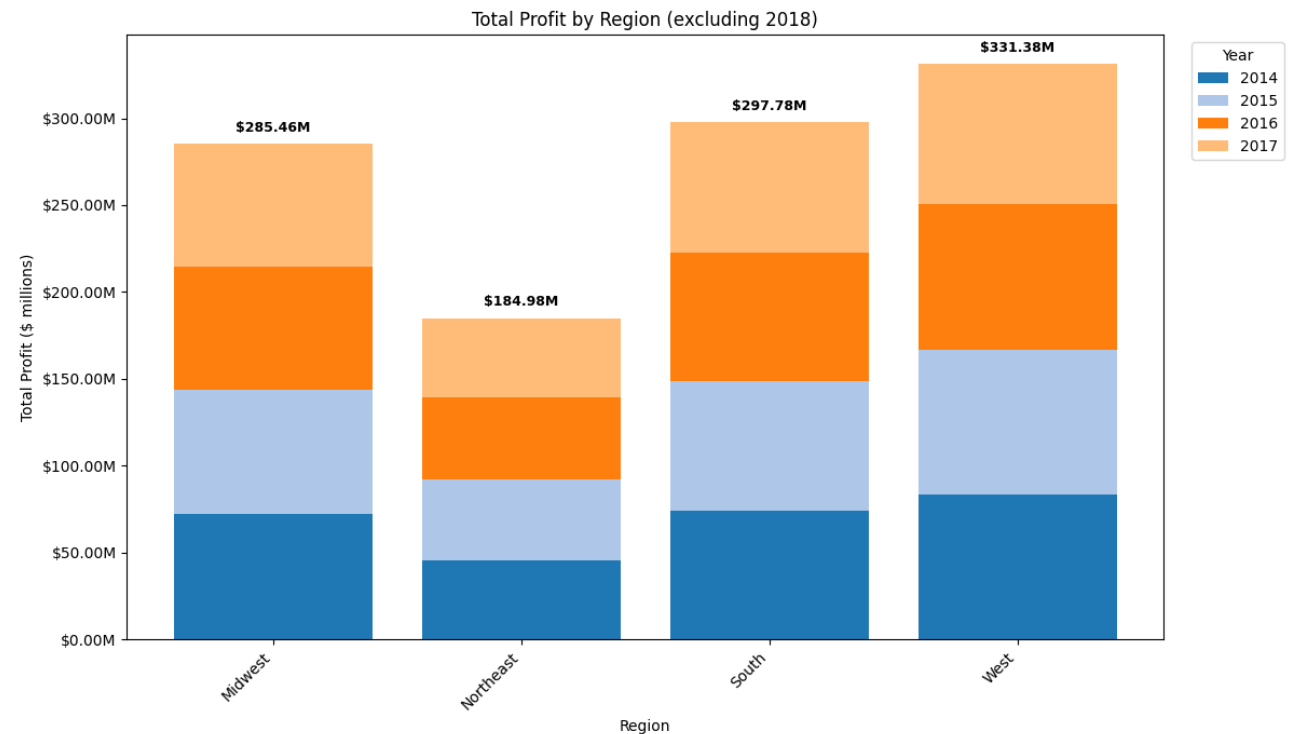
Here we can see some overall information of the numeric features of the data table

	count	mean	min	25%	50%	75%	max	std
OrderDate	64104	2016-01-29 01:28:20.935979008	2014-01-01 00:00:00	2015-01-13 00:00:00	2016-01-27 00:00:00	2017-02-13 00:00:00	2018-02-28 00:00:00	NaN
Order Quantity	64,104.00	8.44	5.00	6.00	8.00	10.00	12.00	2.28
Unit Price	64,104.00	2,284.38	167.50	1,031.80	1,855.90	3,606.28	6,566.00	1,663.60
Line Total	64,104.00	19,280.68	837.50	8,019.90	14,023.10	27,416.40	78,711.60	15,429.60
Total Unit Cost	64,104.00	1,432.08	68.68	606.22	1,084.50	2,046.93	5,498.56	1,107.71
No. Households	64,104.00	57,025.20	0.00	22,639.00	29,518.00	45,409.00	3,113,535.00	143,773.29
Median Income	64,104.00	59,396.94	0.00	43,163.00	53,587.50	71,897.00	149,837.00	21,556.19
Index	64,104.00	14.91	1.00	6.00	15.00	24.00	30.00	8.79
Profit	64,104.00	17,848.60	695.12	7,326.45	12,967.45	25,144.60	75,801.52	14,493.98
Margin	64,104.00	0.92	0.83	0.90	0.93	0.94	0.97	0.03
Unit Cost	64,104.00	183.83	5.72	70.57	136.00	256.64	1,028.96	157.69
year	64,104.00	2,015.59	2,014.00	2,015.00	2,016.00	2,017.00	2,018.00	1.20
month	64,104.00	6.35	1.00	3.00	6.00	9.00	12.00	3.52

# EXPLORATORY DATA ANALYSIS

Based on these, we can see that the West and South region make up the highest proportion of profits, while the Northeast comes up as the lowest by far.

Depending on the intent of the company, you could either look to run experimental targeted campaigns in the Northeast to see if this would increase their overall profit (as they may be an underserved region) or you could look to capitalise on already higher profit regions such as the West to increase the already higher proportion.

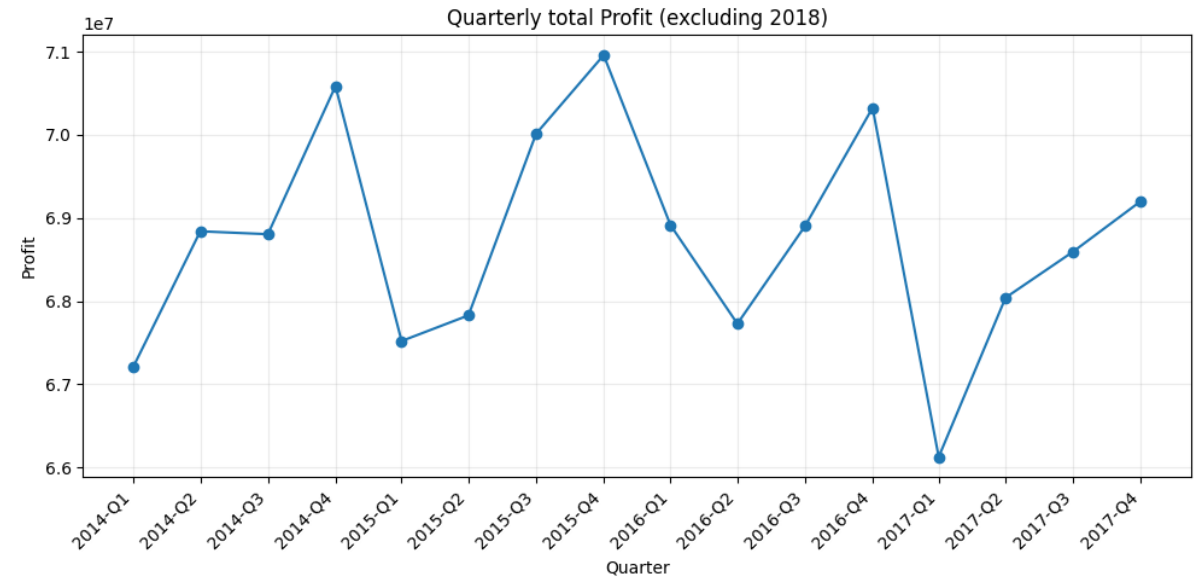


# EXPLORATORY DATA ANALYSIS

We can see here that there are universally large dips in profits around the first quarter of the year, while there are major uplifts in the last quarters.

Depending on the nature of the product, this could suggest a natural seasonality (like holiday items) or otherwise a further analysis could be done to explore why the first quarter is a regular dipping point.

If we assume that this is not a naturally seasonal item, we could consider running experimental campaigns in these lower periods to see what affect they may have on first quarter performance.

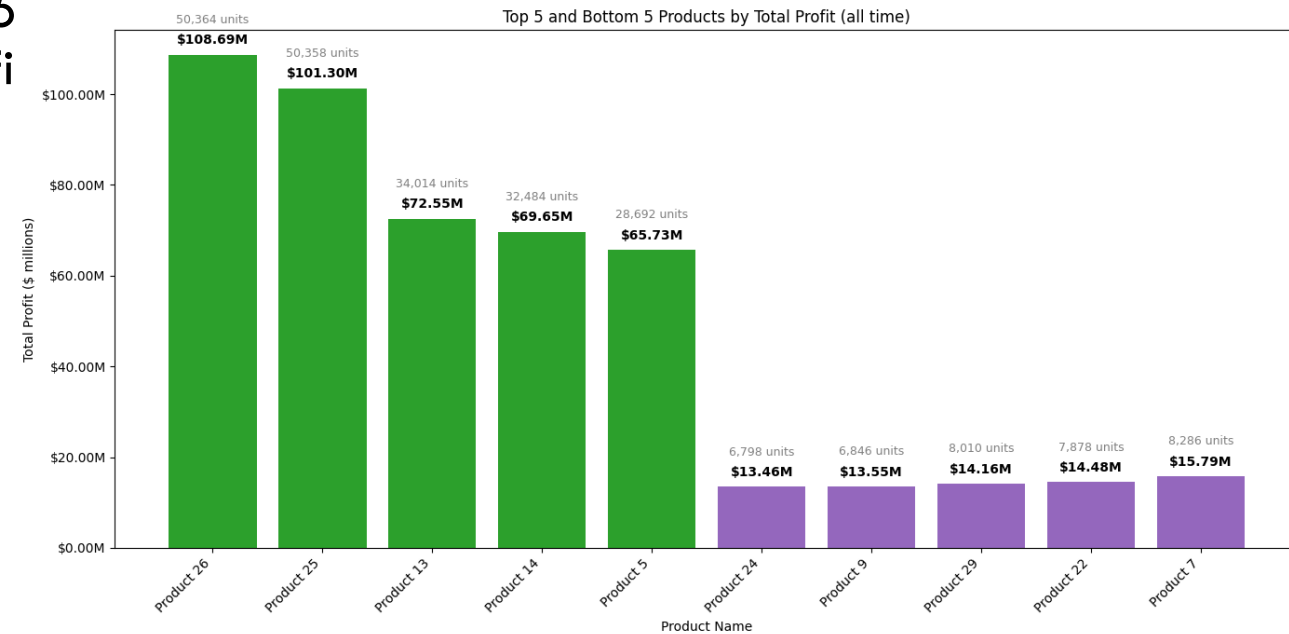


# EXPLORATORY DATA ANALYSIS

From this we can get a clear picture of our 5 best and worst performing products by profit. Product 26 and 25 are our top performers both in units sold and total profit (which makes sense given they are each generally around 16,000 units sold higher than Product 13).

The worst performers were Products 24 and Product 9 and were also the lowest selling units overall.

Given this information, you could consider whether you still need to stock the lowest profit products or if you could cut their budgets. You may also want to put more resource behind the top 5, though be aware this could make you a bit top-heavy if a competitor came out with a similar product to them at potentially a lower price.



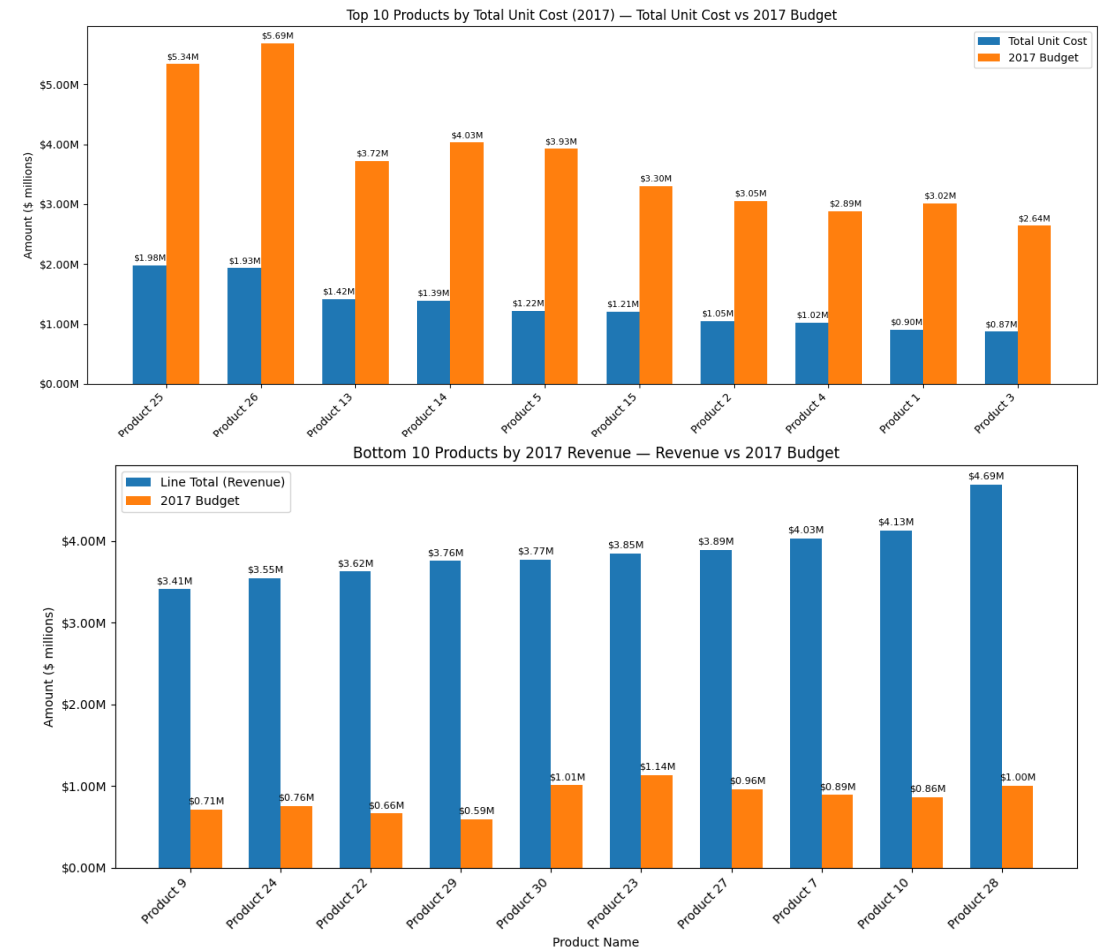


# EXPLORATORY DATA ANALYSIS

This chart shows that all products cost below their given budgets to produce and shows a weak correlation between the budget offered and the amount of units solve. It does show that the budgets for 25 and 26 are also quite a lot higher than the other budgets.

You could potentially experiment on reallocating some of the budget from 25 and 26 to other mid performing products to see if any rise in their sales and profits counteracts any potential loss of 25 and 26.

Here we can see that the lowest revenue products also have some of the lowest budgets, which could mean that if we reallocate some funds from the higher budgets to here that they may increase, though we would have to double check their margins before making that choice.

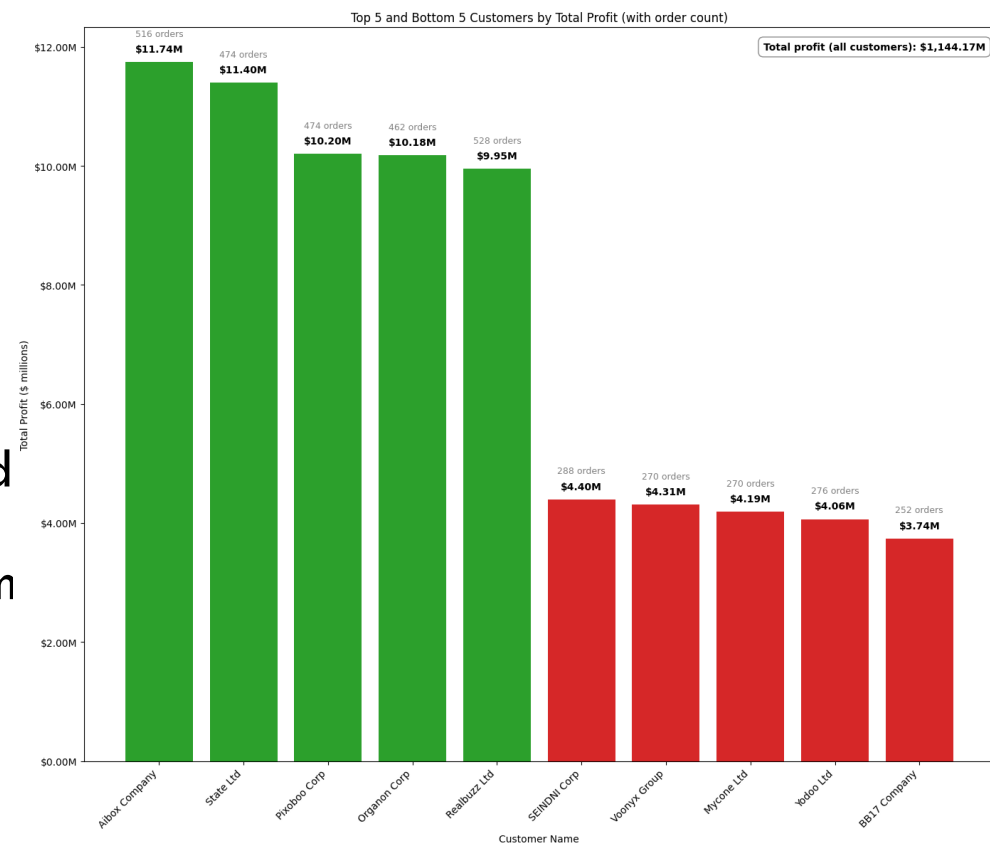


# EXPLORATORY DATA ANALYSIS

This chart shows that we have some very high value customers who are carrying our total profits, which suggests a heavy dependence on these accounts. They also have the highest order counts, which means they buy a lot and frequently.

The bottom 5 meanwhile still contribute meaningfully to the total and should not be disregarded.

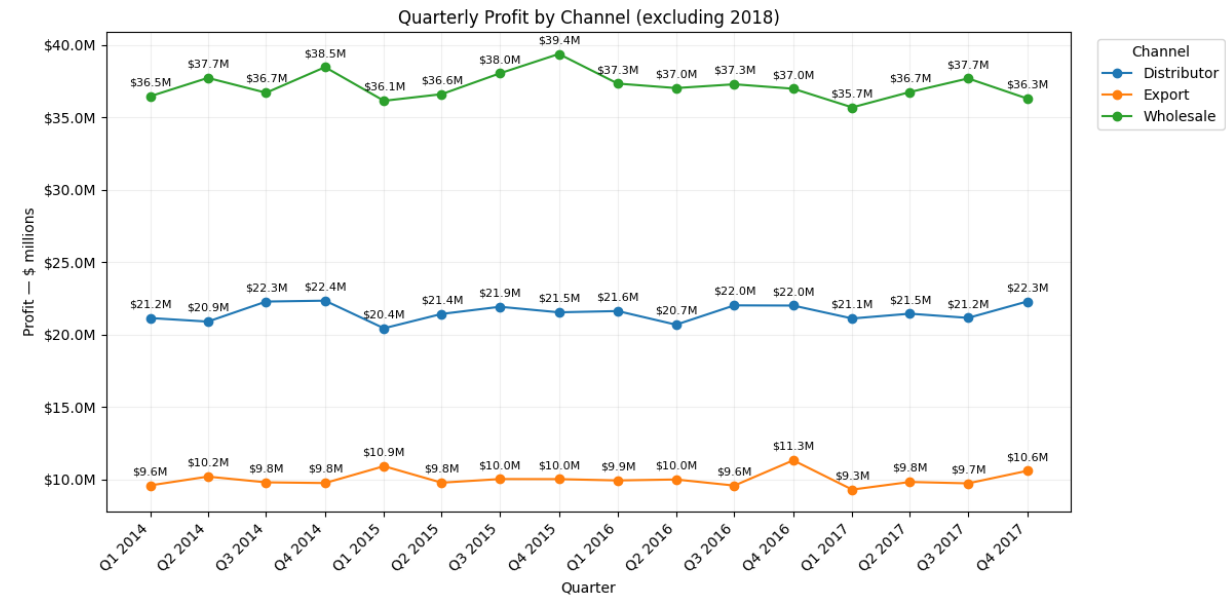
Given the risk of losing any one of these top 5 and their impact on the overall profit, it would be a good move to mitigate risk by ensuring their custom with assigned account owners or similar. You could also look into cross-sell opportunities for mid level companies to spread risk.



# EXPLORATORY DATA ANALYSIS

From here we can see that Wholesale is the overall dominant channel and has the largest profit contribution. Distributor sits in the middle and is relatively stable quarter on quarter while export is the lowest but is less volatile than wholesale.

Targeted investment in Distributor and Export channels could help spread risk and increase general profit, while also performing some measures to protect wholesale like confirming health of contracts and their operational efficiency.



# SUMMARY OF RECOMMENDATIONS

Protect currently high performing customers and marketing channels by auditing their contract health and assigning account owners.

Look into diversifying risk across other mid-tier accounts to increase profit and mitigate top-heavy risks to profile

Revisit product budgets, raising budgets for underfunded products to realise greater revenues and run controlled experiments shifting a small portion of promotional/budget from 25/26 to mid performers and measure net lift

Run experiments in Q1 (promos, bundles, targeted ads) and measure incremental orders and profit

Test Northeast: run targeted, low-cost demand tests to assess addressable upside in under-penetrated regions