Regional Sales Summary Project

Marian Pattugalan

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."

Approach 1

PHASE 1: Understanding the Data and Defining the Objectives

- Gather context: What do each column in each sheet represent? How do the sheets relate to one another?
- Define the objectives (what questions am I trying to answer?)
    - Look into the distribution of revenue across the different states in USA
    - Identify which Products sell the most, which Channel is the most profitable, etc.
    - Identify the peaks and troughs of sales throughout the year

PHASE 2: Set Up

- Load the required libraries (Numpy, Pandas, Matplotlib, Seaborn)
- Import the excel file and upload each sheet as a separate dataframe
- Identify the relationships between each table and merge the data into one big df ('merged_df')
- Get an overview of the data structures, using .**head()**, .**tail()**, .**info()**, .**describe()**, and **.shape**, to identify how the data needs to be cleaned.
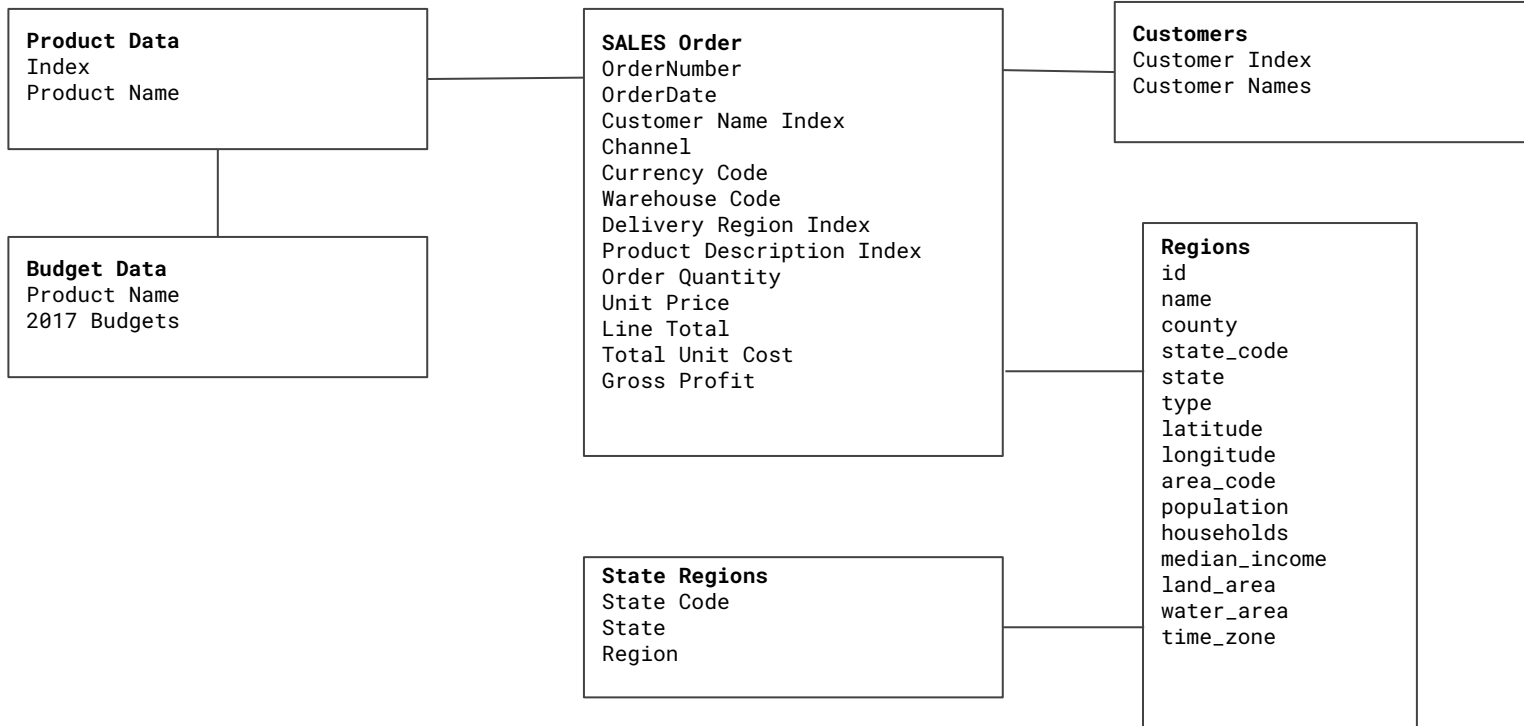
Approach 2

PHASE 3: Data Cleaning

- Checked data types, ensuring that each field was correctly assigned text, numeric, or date
- Checked data quality, double-checking for any null values, errors, duplicates, whitespaces, irrelevant columns.
- Make everything lowercase for easier recall
- Feature Engineering: split date (mm/dd/yyyy) into two groups of columns: 'Order Month Name' and 'Year' for clearer analysis; created bins for median income
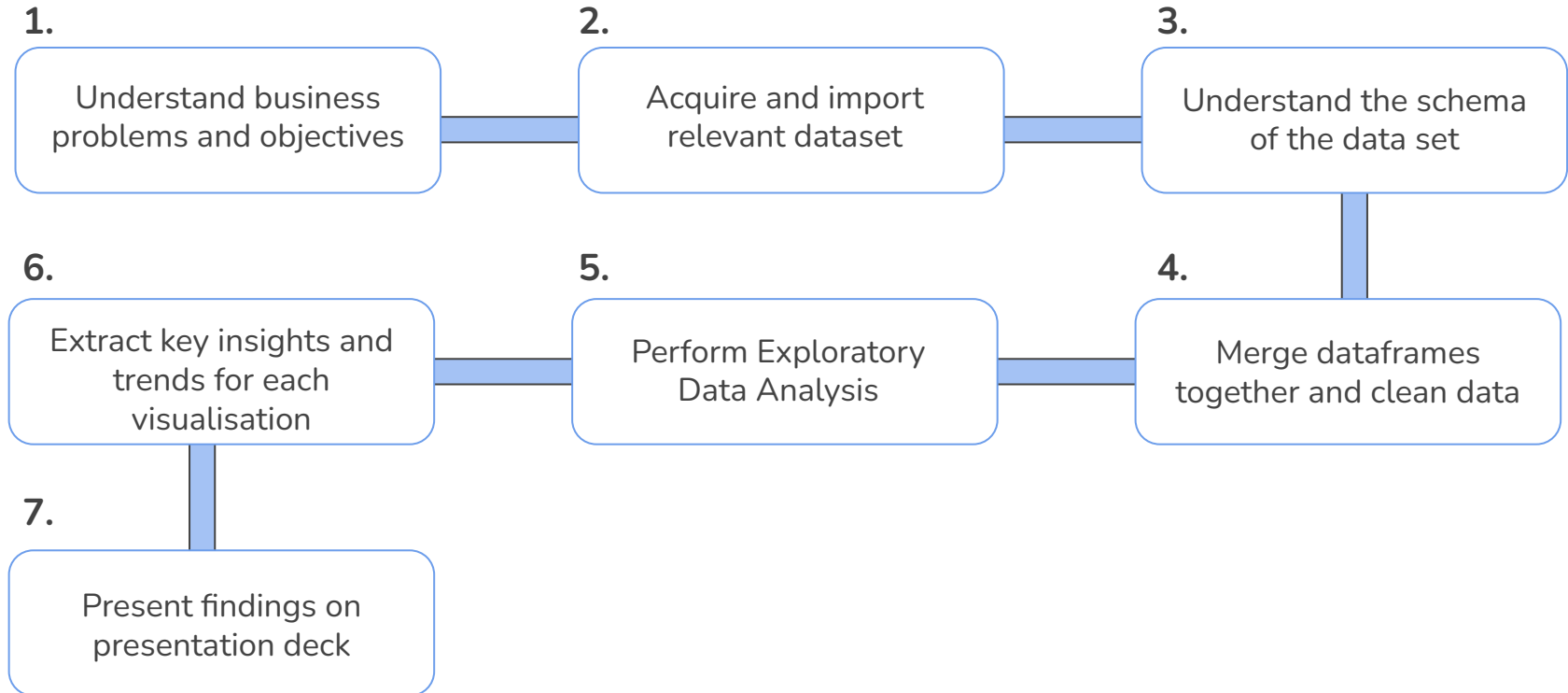
PHASE 4: Exploratory Data Analysis

- Used EDA to uncover patterns, trends, and relationships through univariate and bivariate analyses.
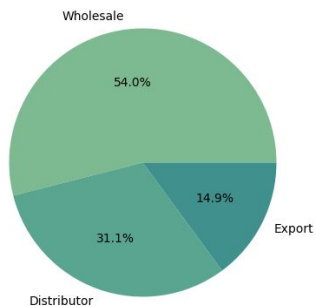
Data Overview

Product Data
Index
Product Name

Budget Data
Product Name
2017 Budgets

SALES Order
OrderNumber
OrderDate
Customer Name Index
Channel
Currency Code
Warehouse Code
Delivery Region Index
Product Description Index
Order Quantity
Unit Price
Line Total
Total Unit Cost
Gross Profit

Customers
Customer Index
Customer Names

Regions
id
name
county
state_code
state
type
latitude
longitude
area_code
population
households
median_income
land_area
water_area
time_zone

State Regions
State Code
State
Region

Project Workflow

1.
Understand business problems and objectives

2.
Acquire and import relevant dataset

3.
Understand the schema of the data set

6.
Extract key insights and trends for each visualisation

5.
Perform Exploratory Data Analysis

4.
Merge dataframes together and clean data

7.
Present findings on presentation deck

Exploratory Data Analysis (EDA)

Sales Channel Distribution

Yearly Gross Profit Composition by Channel (2014-2017)

Popularity of Products across 2014-beginning of 2018
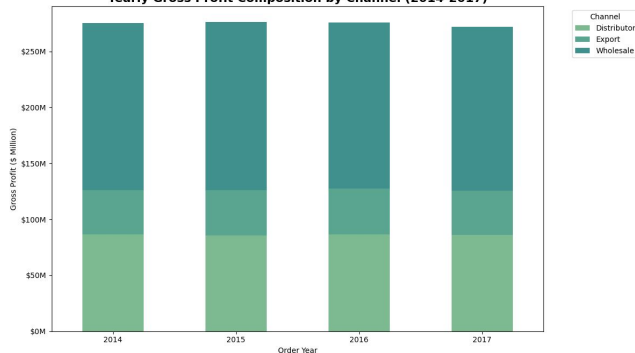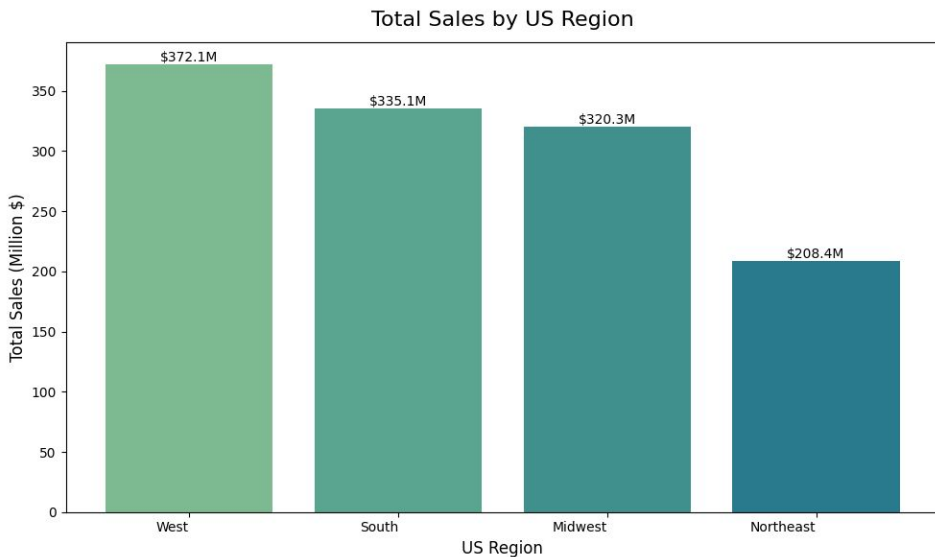

- Wholesale is the dominant channel for sales, making up majority of the gross profit.
- Exports is the smallest channel contributor.
- Most sold products: Product 25, Product 26, Product 13, Product 14, Product 15.
- The graph is heavily skewed to the right, with most sales occurring with the top 5 products.

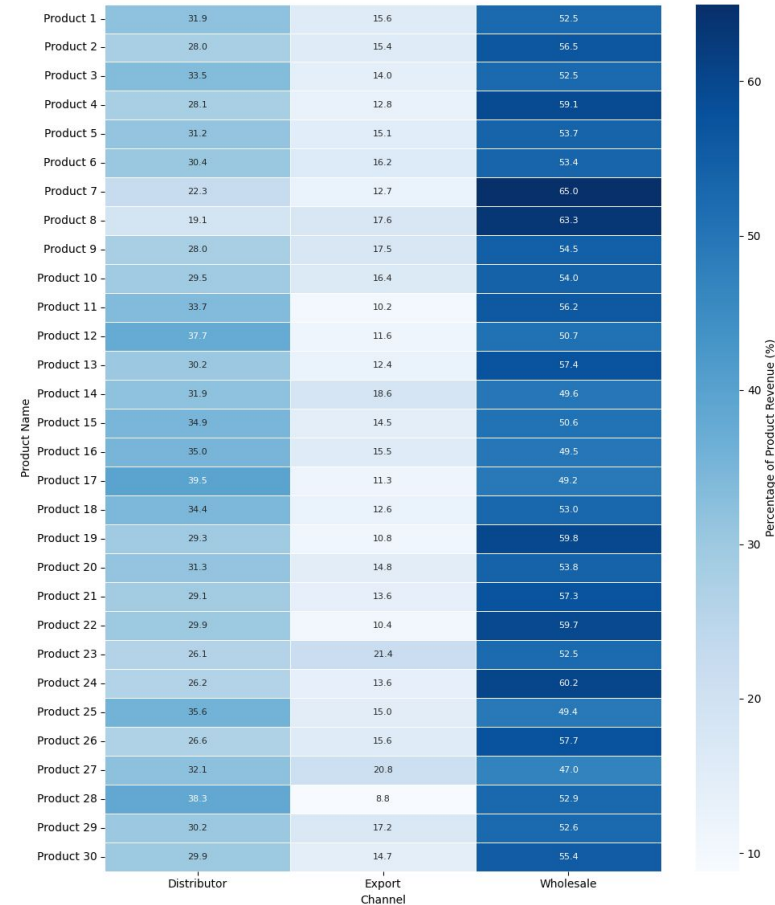# Exploratory Data Analysis (EDA)
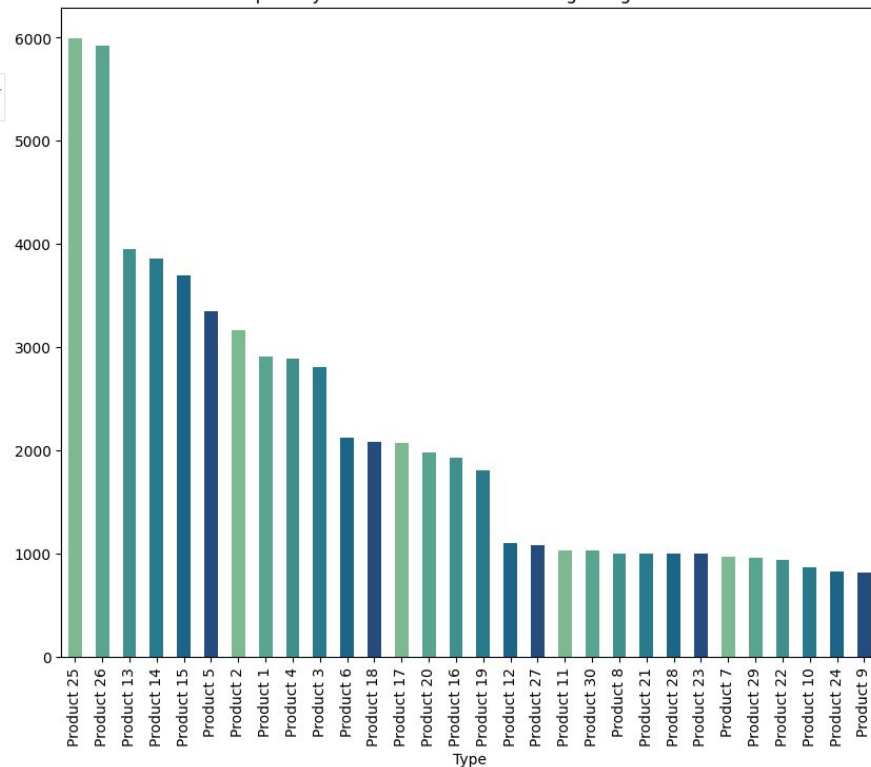
Total Sales by US Region

- The West is the most dominant region, having the highest total sales ($372M) and accounting for roughly 35% of the total sales.
- The South and Midwest both show robust sales but are lower than the West.
- The Northeast has the lowest amount of sales, approximately ~40% less than the West.
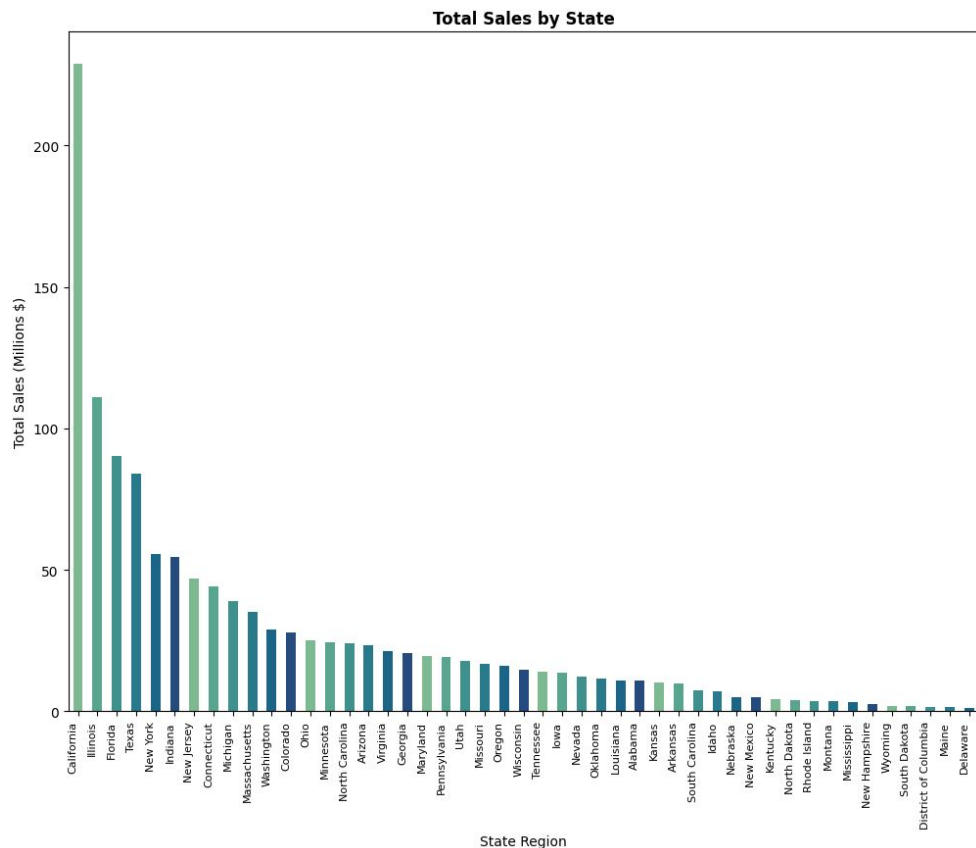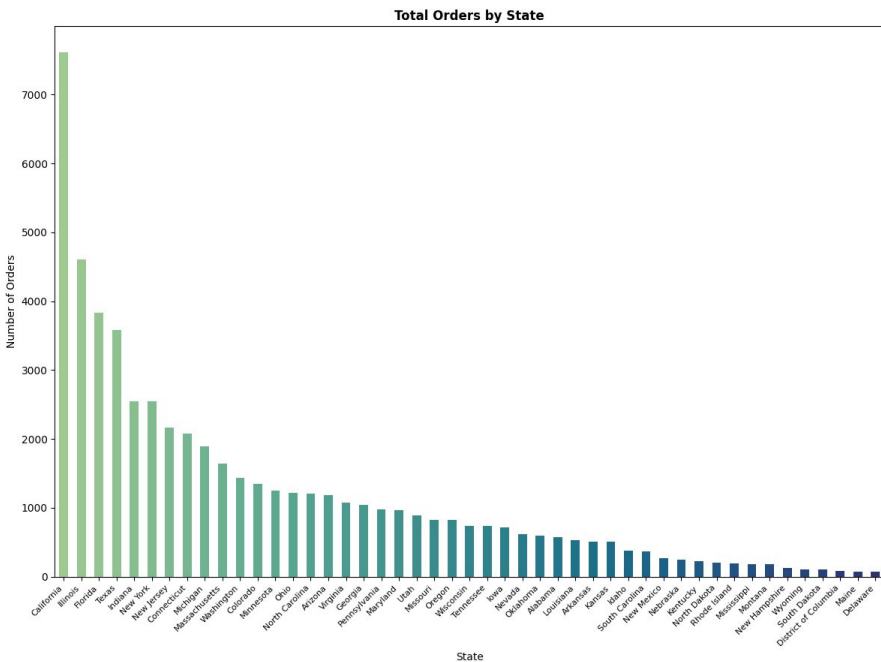
Exploratory Data Analysis (EDA)
Bivariate Analysis

- Wholesale is the primary/dominant channel for every product sold (ranging from ~47-65%), generating the highest revenue percentage for all products.
- Distributor is the second significant channel for all products (ranging from ~19-39%), with Product 12 (37.7%), Product 17 (39.5%), Product 25 (35.6%), and Product 28 (38.3%) showing higher reliance on this channel.
- Export channel is the smallest contributor for all of the products (ranging from ~8-21%), with Product 23 (21.4%) and Product 27 (20.8%) deriving a significantly larger share of their revenue from exports compared to others like Product 28 (8.8%).

### Channel Revenue Distribution for Each Product (%)

Product Name	Distributor	Export Channel	Wholesale
Product 1	31.9	15.6	52.5
Product 2	28.0	15.4	56.5
Product 3	33.5	14.0	52.5
Product 4	28.1	12.8	59.1
Product 5	31.2	15.1	53.7
Product 6	30.4	16.2	53.4
Product 7	22.3	12.7	65.0
Product 8	19.1	17.6	63.3
Product 9	28.0	17.5	54.5
Product 10	29.5	16.4	54.0
Product 11	33.7	10.2	56.2
Product 12	37.7	11.6	50.7
Product 13	30.2	12.4	57.4
Product 14	31.9	18.6	49.6
Product 15	34.9	14.5	50.6
Product 16	35.0	15.5	49.5
Product 17	39.5	11.3	49.2
Product 18	34.4	12.6	53.0
Product 19	29.3	10.8	59.8
Product 20	31.3	14.8	53.8
Product 21	29.1	13.6	57.3
Product 22	29.9	10.4	59.7
Product 23	26.1	21.4	52.5
Product 24	26.2	13.6	60.2
Product 25	35.6	15.0	49.4
Product 26	26.6	15.6	57.7
Product 27	32.1	20.8	47.0
Product 28	38.3	8.8	52.9
Product 29	30.2	17.2	52.6
Product 30	29.9	14.7	55.4

Exploratory Data Analysis (EDA)

Total Orders by State

Total Sales by State

- California is the top-performing state, with the highest total revenue, exceeding $200M.
- Illinois shows robust revenue but is significantly lower than California
- Delaware, District of Columbia, Maine's sales <= $10M, indicating low sales activity in these areas.
- Sales follow the same trend

Exploratory Data Analysis (EDA)

Sales Count by Product and Year
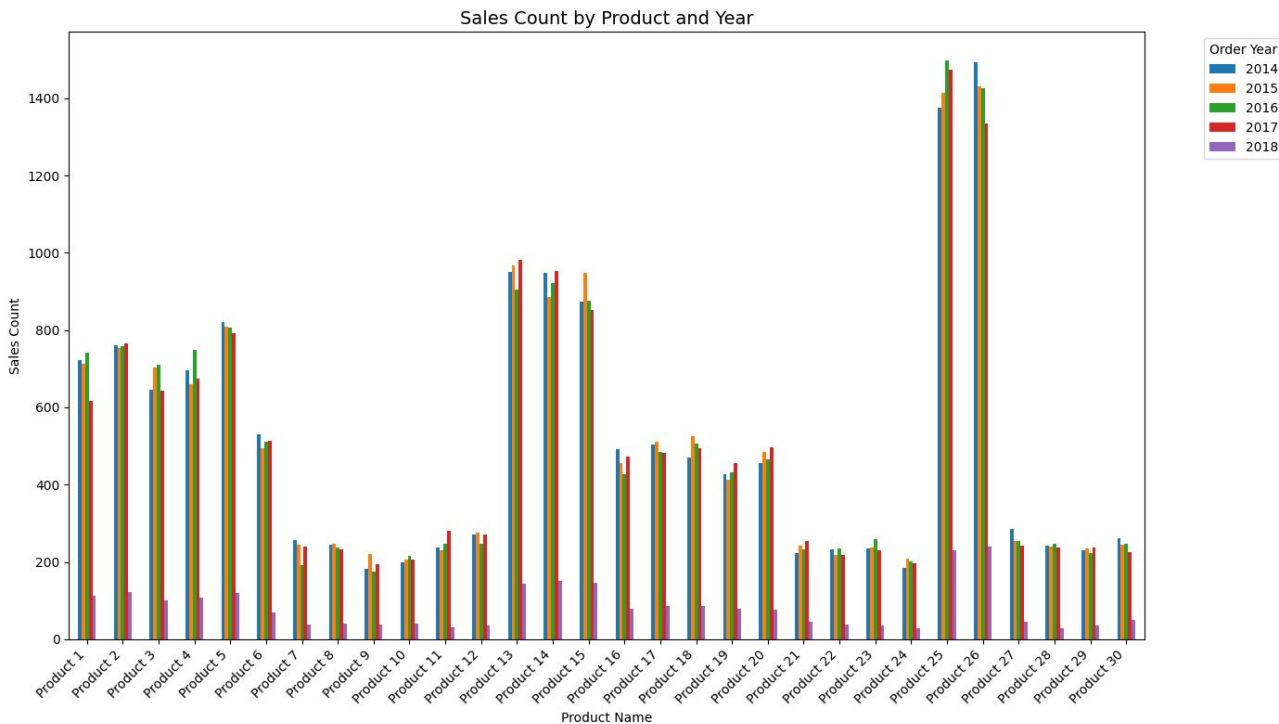
- Top 5 performing products from 2014-2017 (Product 25, Product 26, Product 13, Product 14, Product 15) account for ~36% (23,424) of total sales.
- Lowest performing products (Product 9, Product 10, Product 22, Product 24, Product 29) sold less than 300 units per year.
- Product 26 shows a significant decline (10%) in sales from 1,494 to 1,336 units) between 2014-2017
- Product 11 shows the highest growth (17%) between 2014-2017.
- Uneven distribution, with a small number of products making a large percentage of the sales volume.

Exploratory Data Analysis (EDA)


Total Revenue by Month (Year-over-Year Comparison)
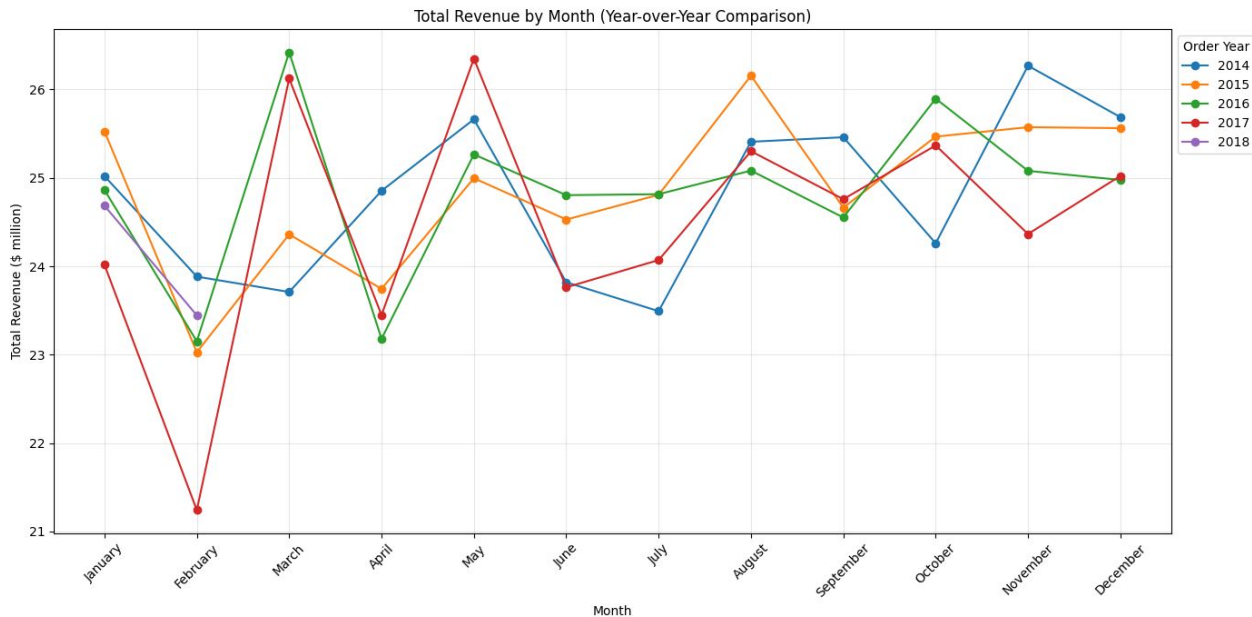
- The business is highly seasonal, showing trends with the peaks and troughs throughout the year.
- Sales stay consistent between $24-26M throughout the months per year, but there is a dramatic drop in Feb(2017) which is an outlier.
- There is a plateaus in revenue between October-December, making Q4 the highest revenue quarter in sales in every year.
- Consistent decline in January-February indicates a post-holiday trough into the new year.
- There are spike increases in April and May.
- There is an upward trajectory per year post 2014, indicating a strong year-on-year growth.
- There is no clear upward or downward trend, showing a stable seasonal pattern.
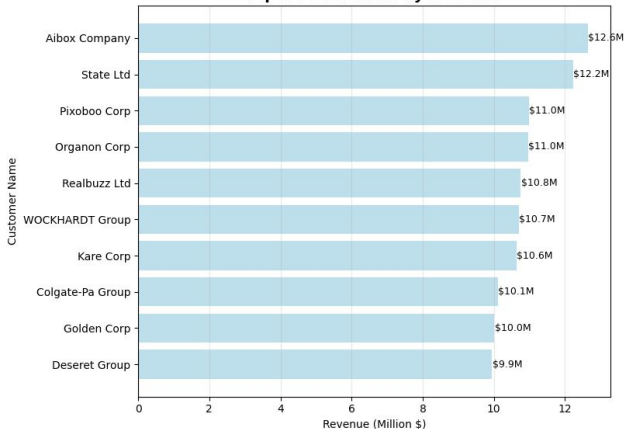
Exploratory Data Analysis (EDA)

Top 10 Selling Products (Millions $)
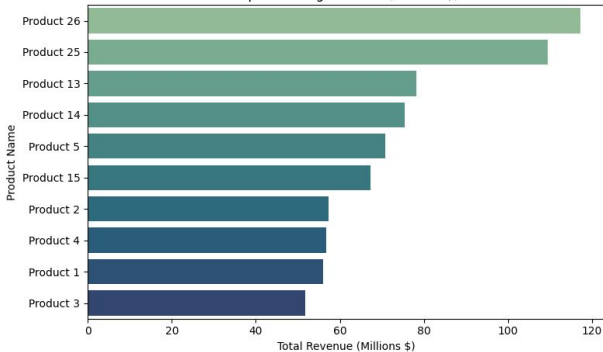
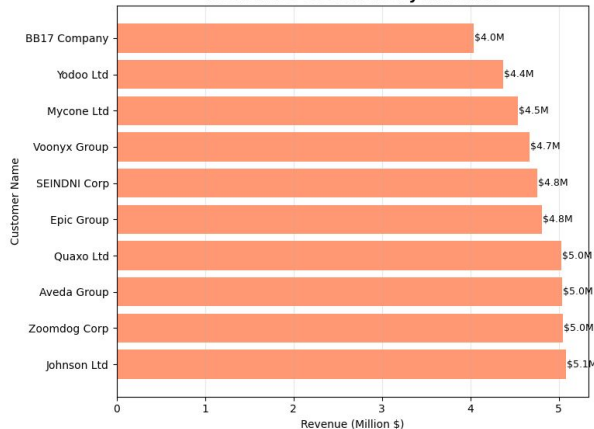Lowest 10 Selling Products (Millions $)
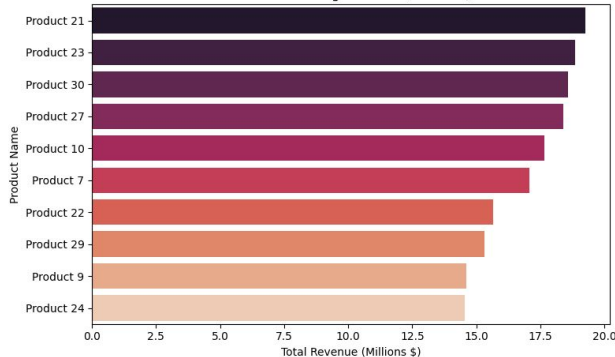
Top 10 Customers by Revenue

Bottom 10 Customers by Revenue

- Product 26 and 25 are the leading products in revenue (with revenues more than $100M).
- Product 24 is the lowest selling product, with a revenue just under $15M.
- Aibox Company is the highest revenue customer ($12.6M), closely followed by State Ltd ($12.2M).
- There is a steep 12.7% decrease between Aibox Company and the third highest revenue company.
- The top 10 bottom customers by revenue are below $5M, which is around 60% lower than the highest revenue companies.
- This shows that majority of the revenue are from the top customers.

Exploratory Data Analysis (EDA)
Bivariate Analysis on 2017

2017 Budget vs Actual Total Unit Cost by Product (Top 10)
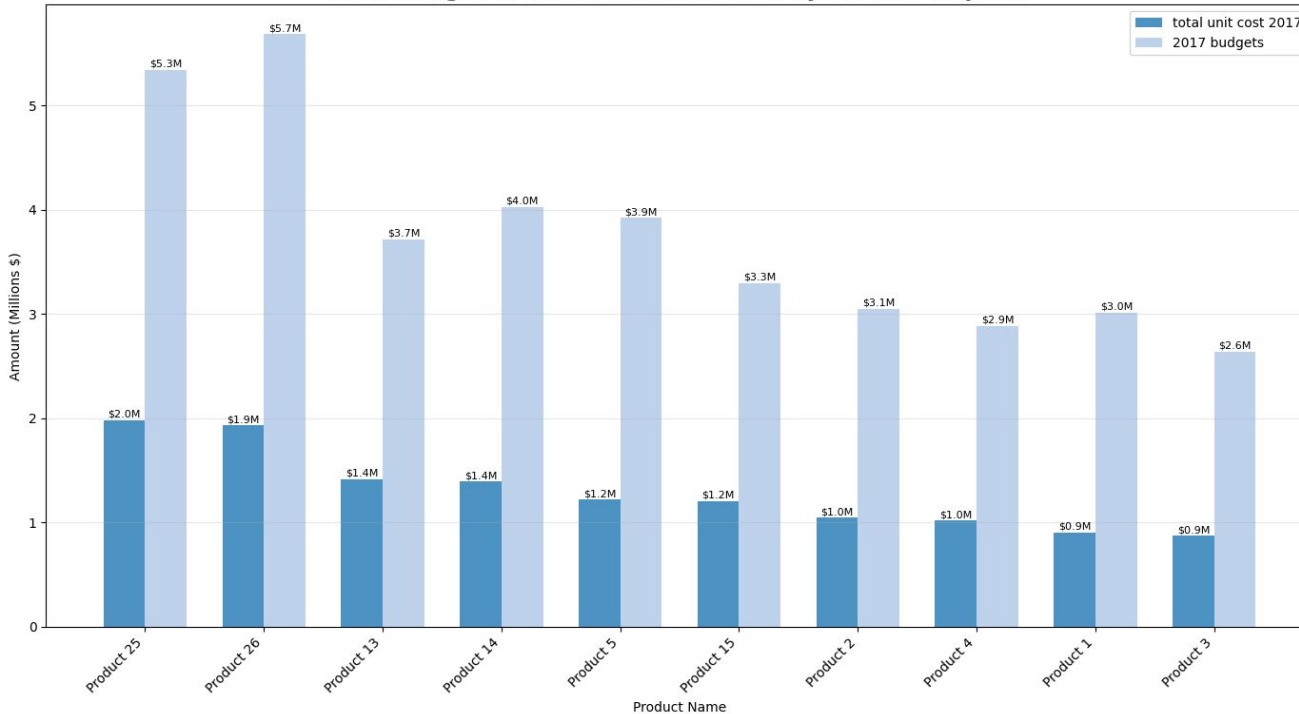
- For every product, the allocated budget significantly exceeds the actual total unit cost. Shows consistent budget overestimation in terms of raw material prices or labour.
- The cost only account for around ~35% (range: 30.8-37.8%).
- Product 25 and Product 26 have the biggest variances ($3.3M and $3.8M, probably indicating ambitious planning for sales.
- Lower actual costs could mean higher gross profits.

Key Insights

- Monthly revenue is consistent throughout the years, with consistent spikes in May and April and a strong Q4(Oct-Dec). A dip in revenue can be seen in January-February, indicating a post-holiday trough.
- The Western region (particularly California with $230M revenue and around 7500 orders) dominates in sales performance. Meanwhile, the Northeastern region (e.g. Delaware with than $5M and less than 1000 orders) has the weakest regional performance.
- Distribution of channels relies largely on wholesale: Wholesale (54%), Distributors (31%) and Exports (15%)
- Product 26 and Product 25 have the highest revenues and sales
- Product 24 and Product 9 have the lowest revenues and sales
- There is gross miscalculation on the budget of each Product.
- Based on patterns, expect 2018 to follow similar seasonality (unless disrupted by external factors). Avg annual revenue/year is around $290-300M.

Recommendations

- Northeast region's lower sales indicate that there is a big opportunity for growth here via targeted marketing, improved channel distribution, and tailored promotions.
- Provide additional product promotions, deals, or rebranding for the underperforming products
- Prioritize relationship management with top companies to sustain high revenue contribution, but also explore upselling opportunities with lower-tier customers to boost their revenue.
- Recurring February dip could inform that more marketing should be conducted in Q1
- Conduct additional promotions for Product 26 as its revenue dropped significantly in 2017.
- Utilise the growth of Product 11 to further increase its revenue for 2018.
- Recalibrate the budgets for each Product and allocate more budget for advertisements/promotion into Products which are underperforming