📈 XYZ Co. Regional Sales Data Analysis Plan (2014-2018)

📌 Problem Statement & Objectives

XYZ Co. aims to analyze its 2014-2018 sales data to identify core **revenue and profit drivers** across products, channels, and regions. The goal is to uncover **seasonal trends** and outliers, align performance against the **2017 budgets**, and use these insights to optimize pricing, promotions, and market expansion for sustainable growth while mitigating concentration risk.

---

1. Data Wrangling and Preparation Summary 🖌️

DataFrame	Description	Original Shape	Status
df_sales	Core sales transactions	(64104, 12)	Merged
df_products	Customer details	(175, 2)	Merged
df_customers	Product details	(30, 2)	Merged
df_regions	Detailed city/region info	(994, 15)	Merged
df_state_reg	State-to-Region mapping	(49, 3)	**Needs Cleaning/Merge**
df_budgets	2017 Product Budgets	(30, 2)	**Needs Merge**

Export to Sheets

Data Cleaning Checks

- **Null Values:** The core df_sales table has **no missing values** (df_sales.isnull().sum()), indicating a high-quality transaction log.
- **Consistency:** All column names were converted to **lowercase** (df.columns = df.columns.str.lower()) to ensure consistent indexing.

🔗 Merging Steps Completed

The df_sales data has been successfully enriched with descriptive information from three tables using **left merges**:

1. **Customer Names:** Joined on customer name index.
2. **Product Names:** Joined on product description index.
3. **Geographic Detail:** Joined on delivery region index.

Visual Confirmation of Merge Integrity:

---

2. Feature Engineering and Final Merges 🛠️

To make the data ready for strategic analysis, we create key performance indicators (KPIs) and complete the final merges.

A. Core Metric Calculation (Profitability)

The most crucial step is deriving gross profit, a direct profit driver.

New Column	Formula	Purpose
gross_profit	line total - (total unit cost * order quantity)	Determines actual profit per transaction.
profit_margin	gross_profit / line total	Standardized metric for comparing profitability across products/channels.

Export to Sheets

Python

```
# Code Snippet: KPI Calculation

df['gross_profit'] = df['line total'] - (df['total unit cost'] * df['order quantity'])

df['profit_margin'] = df['gross_profit'] / df['line total']
```

B. Final Data Integration

The remaining dimensional tables must be merged to enable regional and budget comparisons.

1. **Clean and Merge State Regions:** This step connects low-level geography (state_code) to higher-level business **Region** (South, West, etc.) for high-level analysis.

Python

```
# Code Snippet: Example Cleanup & Merge (assuming initial header fix)

# The merged dataframe should now have a 'region' column.

df = df.merge(df_state_reg, how='left', left_on='state_code', right_on='Column1')
```

2. **Merge 2017 Budgets:** This is critical for assessing the 2017 goal alignment.

Python

```
# Code Snippet: Merge Budgets
```

```
df = df.merge(
  df_budgets,
  how='left',
  left_on='product name_x', # Use the proper product name column
  right_on='Product Name'
)
```

---

3. Strategic Visualization Plan 📊

The following interactive visualizations directly address the objectives outlined in the problem statement.

A. Identifying Key Revenue and Profit Drivers

Objective	Visualization Type	Key Metrics/Dimensions	Strategic Insight
Revenue Drivers (Product & Channel)	**Interactive Treemap**	Size: line total; Hierarchy: product name channel	Optimizing the top products and their most effective sales channels (e.g., increased promotion in the most successful channel).
Profit Drivers (Product Profitability)	**Box Plot**	X: product name; Y: profit_margin	Comparing margin distribution across products to identify high-variance or consistently low-margin items that need pricing review.

Export to Sheets

Visualization Example:

B. Seasonal Trends and Performance Alignment

Objective	Visualization Type	Key Metrics/Dimensions	Strategic Insight
Seasonal Trends	**Time Series Line Plot**	X: orderdate (resampled Monthly/Quarterly); Y: line total	Planning inventory and marketing campaigns around predictable peak seasons and analyzing growth trends year-over-year.
Performance vs. Budget	**Dual-Axis Bar Chart**	X: product name; Y1: 2017 Budgets; Y2: 2017 Sales	Immediate identification of product lines that require investment (exceeded target) or corrective action (missed target).

Export to Sheets

Visualization Example:

C. Mitigating Concentration Risk (Geographic Analysis)

Objective	Visualization Type	Key Metrics/Dimensions	Strategic Insight
Concentration Risk	**Choropleth Map (State Level)**	State Color: Total Revenue; Label: State/Region Name	Highlights excessive revenue reliance on a few geographic areas. Guides strategy for safe expansion into low-contributing states/regions.
Region-Level Deep Dive	**Bar Chart of Median Income vs. Sales**	X: region; Y1: total sales; Y2: median_income	Understanding the demographic link to sales success, useful for informing future market penetration strategies.

Export to Sheets