# NAME: ADEOLU JAMES OLUWASOLA

# STUDENT ID: VEPH/20B/DA226

# TASK 16B

## Technical Report Template for Analytical Projects in Microsoft Excel

### 

## **1. Outline**

# Introduction

# Story of Data

# Data Splitting and Preprocessing

# Pre-Analysis

# In-Analysis

# Post-Analysis and Insights

# Data Visualizations & Charts

# Recommendations and Observations

# Conclusion

# References & Appendices

### 

## **2. Introduction**

### **Objective of the Project:**

The aim of this analysis is to evaluate sales performance across different regions, customer segments, and product categories. The goal is to identify trends in sales, customer purchasing behavior, and high-performing products. Additionally, the analysis seeks to provide data-driven insights to improve business strategies and forecast future sales patterns.

### **Problem Being Addressed:**

The analysis aims to answer key business questions, such as:

* Which product categories generate the highest revenue?
* What are the sales trends across different states and cities?
* How do shipping methods impact customer purchases?
* Which regions and customer segments contribute the most to overall sales?
* Are there any patterns in sales that could help optimize inventory and marketing strategies?

### **Key Datasets and Methodologies:**

#### **Datasets Used:**

* **Sales Data:** Contains order details, including order date, shipping method, and customer information.
* **Customer Data:** Includes customer ID, segment, and geographical location (country, state, city).
* **Product Data:** Categorized by product category and sub-category.
* **Shipping Information:** Specifies ship date, ship mode, and order priority.

#### **Methodologies Used in Excel:**

* **Pivot Tables:** To summarize and analyze sales data by region, category, and customer segment.
* **Conditional Formatting:** To highlight key trends and outliers in sales performance.
* **Sorting & Filtering:** For better data organization and extracting relevant insights.
* **Data Validation & Cleaning:** Ensuring accuracy in data analysis by removing duplicates and handling missing values.
* **Charts & Graphs:** To visualize trends in sales performance and customer behavior.

## **3. Story of Data**

### **Data Source:**

The dataset appears to be sourced from a business sales database, likely an internal company system or a public dataset (such as Kaggle or a retail sales dataset). It contains transactional sales data, including customer details, product categories, and order information.

### **Data Collection Process:**

The data was likely collected through an automated system, such as a company’s Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM) software. It could also have been exported from an e-commerce platform or manually compiled from sales records.

### **Data Structure:**

* Each **row** represents an individual sales transaction.
* The **columns** capture different attributes of the transaction, including:
  + **Order ID**: Unique identifier for each sale.
  + **Order Date & Ship Date**: Timeline of order processing and delivery.
  + **Customer Details**: ID, segment, and location (city, state, country).
  + **Product Details**: Category, sub-category, and product identifier.
  + **Shipping Mode**: Indicates delivery priority (e.g., Standard Class, First Class).
  + **Region & Postal Code**: Geographic segmentation of sales.

### **Important Features and Their Significance:**

* **Customer Segment**: Helps identify purchasing patterns among different customer groups.
* **Product Category & Sub-Category**: Allows for analyzing which products contribute most to revenue.
* **Order & Ship Dates**: Useful for identifying peak sales periods and shipping efficiency.
* **Region & State**: Enables geographical sales analysis to determine high-performing areas.
* **Shipping Mode**: Helps evaluate the impact of different shipping methods on customer satisfaction.

### **Data Limitations or Biases:**

1. **Missing or Incomplete Data:**
   * There could be missing values in key fields like customer details, product categories, or sales figures, which may lead to gaps in analysis.
   * If some transactions are not recorded or improperly logged, it might skew trends and insights.
2. **Sample Bias:**
   * If the dataset only represents sales from a specific period or region, the insights may not be generalizable to the entire business.
   * If certain customer segments or product categories are underrepresented, conclusions drawn may be skewed.
3. **Data Entry Errors or Inconsistencies:**
   * Human or system-generated errors in order IDs, dates, or shipping details could affect analysis accuracy.
   * Variations in customer names or duplicate entries might create redundancy in the dataset.
4. **Lack of External Factors:**
   * The dataset does not seem to include external influences like marketing campaigns, seasonal trends, or competitor activity, which might impact sales performance.
5. **Currency & Pricing Considerations:**
   * If pricing details (such as discounts or promotions) are missing, it may be difficult to accurately assess revenue and profitability.

To mitigate these issues, **data cleaning techniques** like removing duplicates, filling missing values, and verifying data consistency will be necessary before conducting in-depth analysis.

## **4. Data Splitting and Preprocessing**

### **Data Cleaning:**

To ensure accurate and reliable analysis, the following data cleaning steps were implemented:

1. **Removing Duplicates:** Used Excel’s **Remove Duplicates** feature to eliminate redundant transactions that might distort sales figures.
2. **Correcting Errors:** Checked for inconsistencies in order IDs, customer details, and product categories, correcting any formatting or data entry mistakes.
3. **Standardizing Formats:** Ensured uniform formatting for dates, customer names, and region codes to maintain consistency across the dataset.
4. **Handling Missing Values:** Applied different strategies based on the type of missing data.

### **Handling Missing Values:**

* **For Numerical Data (e.g., sales, postal codes):** Used **Excel’s AVERAGEIF function** or median imputation to fill gaps with meaningful estimates.
* **For Categorical Data (e.g., product categories, customer segments):** Used **mode imputation (most frequent value)** or logical inference based on related attributes.
* **For Critical Missing Data:** If essential fields (like order ID) were missing, those rows were removed to maintain dataset integrity.

### **Data Transformations:**

Several transformations were performed to enhance analysis:

* **Date Extraction:** Used Excel’s **YEAR(), MONTH(), and DAY() functions** to derive yearly and monthly sales trends.
* **Sales Aggregation:** Summarized revenue data using **SUMIFS and Pivot Tables** for a clearer view of total sales by region and category.
* **Profit Margin Calculation:** Created a new column to analyze profitability by subtracting costs from sales revenue.
* **Categorical Grouping:** Grouped product sub-categories into broader categories to simplify trend analysis.

### **Data Splitting:**

The dataset was divided into dependent and independent variables:

* **Dependent Variable:**
  + **Sales (Revenue):** The key metric used to measure business performance.
* **Independent Variables:**
  + **Product Name, Category, Sub-Category:** Helps identify top-selling products.
  + **Customer Segment:** Determines which groups contribute most to sales.
  + **Region & State:** Assesses geographical sales trends.
  + **Shipping Mode:** Evaluates delivery efficiency impact on sales.

### **Industry Context:**

The dataset belongs to the **retail and e-commerce industry**, focusing on office supplies, furniture, and technology sales. Understanding sales performance within this sector is crucial for inventory management, marketing strategies, and customer satisfaction.

### **Stakeholders:**

The findings from this analysis will benefit:

* **Senior Management:** To make data-driven decisions on pricing, inventory, and resource allocation.
* **Marketing Team:** To target high-value customers and improve promotional campaigns.
* **Supply Chain & Logistics Team:** To optimize shipping efficiency and reduce delivery delays.
* **Sales Team:** To identify high-demand products and prioritize sales strategies.

### **Value to the Industry:**

The analysis provides valuable insights that can:

* **Optimize Product Inventory:** Helps businesses stock high-demand items and avoid overstocking slow-moving products.
* **Improve Customer Segmentation:** Enables personalized marketing strategies based on purchasing behavior.
* **Enhance Sales Forecasting:** Helps predict future sales trends and plan better for seasonal demand fluctuations.
* **Increase Profitability:** Identifies cost-effective shipping methods and high-margin products to maximize profits.

## **5. Pre-Analysis**

### **Identify Key Trends:**

From an initial review of the dataset, a few noticeable trends emerge:

1. **Regional Sales Variations:**
   * Certain regions (e.g., Central, West, and South) appear to have higher sales activity compared to others.
   * Major metropolitan cities like New York, Houston, and Los Angeles seem to be key contributors to sales.
2. **Product Category Performance:**
   * **Office Supplies** are widely sold, but **Furniture** and **Technology** products might generate higher revenue per sale.
   * Some sub-categories, like **Chairs, Binders, and Storage Units**, appear frequently, indicating high demand.
3. **Customer Segment Patterns:**
   * **Corporate and Consumer** segments seem to dominate the sales, while the **Home Office** segment might be smaller.
   * Certain customer segments may prefer specific product types (e.g., corporate clients ordering office supplies in bulk).
4. **Shipping Mode and Order Processing:**
   * The dataset includes **Standard Class, Second Class, and First Class shipping**, which may impact delivery times and customer satisfaction.
   * Faster shipping modes (First Class) might be associated with high-value orders or urgent purchases.
5. **Sales Across Time Periods:**
   * There may be peaks in sales around certain months, possibly due to **seasonal trends** (e.g., back-to-school, end-of-year office procurement).

### **Potential Correlations:**

* **Product Category & Sales Revenue:** High-value product categories like **Technology** may correlate with larger transaction amounts.
* **Region & Sales Volume:** Some regions may drive higher revenue due to population density or business activity.
* **Shipping Mode & Customer Segments:** Faster shipping might be preferred by corporate clients with urgent needs.
* **Customer Segment & Product Preferences:** Different segments may have distinct purchasing habits (e.g., small businesses buying different office supplies than large corporations).

### **Initial Insights:**

* **Key Revenue Drivers:** Technology and Furniture products likely generate higher revenue per sale than Office Supplies.
* **Geographic Sales Hotspots:** Certain cities and states might be more profitable than others, suggesting potential for targeted marketing efforts.
* **Customer Behavior:** Corporate and Consumer segments appear dominant, so promotions targeting them could be more effective.
* **Operational Efficiency Considerations:** Understanding shipping mode preferences and order fulfillment speed can help optimize logistics.

## **6. In-Analysis**

### **Key Trends:**

#### **1. Regional Sales Variations:**

* Certain regions, particularly **Central, West, and South**, exhibit **higher sales activity** compared to others.
* **Major metropolitan cities** such as **New York, Houston, and Los Angeles** appear to be **key contributors to overall sales volume**.

#### **2. Product Category Performance:**

* **Office Supplies** are frequently sold, but **Furniture and Technology** likely generate **higher revenue per transaction**.
* Specific **high-demand sub-categories** include **Chairs, Binders, and Storage Units**, indicating staple office needs.

#### **3. Customer Segment Patterns:**

* The **Corporate and Consumer** segments dominate overall sales, while **Home Office** purchases seem relatively smaller.
* Certain customer segments appear to have **specific product preferences**, such as **corporate clients purchasing office supplies in bulk**.

#### **4. Shipping Mode and Order Processing:**

* The dataset includes **Standard Class, Second Class, and First Class** shipping, each likely affecting delivery times and customer satisfaction.
* **Faster shipping methods (First Class)** might be linked to **higher-value orders or urgent business purchases**.

#### **5. Sales Across Time Periods:**

* Sales may show **seasonal peaks**, possibly influenced by:
  + **Back-to-school season** (August–September) driving demand for office supplies.
  + **End-of-year corporate procurement** (November–December).

### 

### **Potential Correlations:**

**Product Category & Sales Revenue** – Higher-value categories like **Technology and Furniture** may correlate with **larger transaction amounts**.  
 **Region & Sales Volume** – **Dense urban areas** could be driving higher revenue due to business activity and population concentration.  
 **Shipping Mode & Customer Segments** – **First Class shipping** might be preferred by **corporate clients** with urgent order needs.  
 **Customer Segment & Product Preferences** – Different customer segments likely have **unique purchasing patterns** (e.g., small businesses buying different office supplies than large enterprises).

**Initial Insights:**

**Key Revenue Drivers** – Technology and Furniture products likely **generate higher revenue per sale** compared to Office Supplies.  
 **Geographic Sales Hotspots** – Certain **cities and states contribute disproportionately to revenue**, suggesting opportunities for **region-specific marketing efforts**.  
 **Customer Behavior** – The dominance of the **Corporate and Consumer segments** implies that **personalized promotions** targeting these groups **could increase sales**.  
 **Operational Efficiency Considerations** – **Understanding shipping preferences** and **optimizing delivery logistics** could enhance **customer satisfaction and profitability**.

## **7. Post-Analysis and Insights**

### **Key Findings:**

1. **Sales Dominated by Standard Class Shipping:**
   * The majority of sales ($1,340,831) occurred through **Standard Class shipping**, making it the **most used shipping method**.
   * **Same-day shipping** had the lowest sales volume ($125,219), suggesting limited use for urgent deliveries.
2. **Top-Selling Product & Category:**
   * The **Canon Advanced Copier** was the **top-selling product** ($61,600 in sales).
   * **Technology** was the highest-grossing product category ($827,456), followed by **Furniture ($728,659)** and **Office Supplies ($705,422)**.
3. **Regional & State Sales Performance:**
   * **The West region** contributed the highest revenue (31%), followed closely by the **East (30%)**.
   * **California led state-wise sales** ($446,306), followed by **New York ($306,361)** and **Texas ($168,573)**.
4. **Customer Segments & Revenue Share:**
   * The **Consumer segment** drove the highest sales ($1,148,061), followed by **Corporate ($688,494)** and **Home Office ($424,982)**.
   * This indicates that **individual consumers are a significant revenue driver**, more so than businesses.
5. **Price Range & Sales Distribution:**
   * The majority of sales transactions fell within the **$0–$3,000 range ($1,912,697 in total sales)**.
   * Higher price brackets (above $12,000) contributed smaller portions, reinforcing that **affordable products drive higher sales volume**.
6. **Top Customers & Revenue Contribution:**
   * **Sean Miller** was the **highest-spending customer** with **$25,043 in total sales**.
   * Other significant customers included **Tamara Chand, Raymond Buch, Tom Ashbrook, and Adrian Barton**, highlighting a small segment of high-value customers.

### **Comparison with Initial Findings:**

**Alignment with Initial Expectations:**

* The assumption that **Technology and Furniture drive higher revenue per sale** was confirmed, as Technology led total revenue, and high-priced items contributed significantly.
* The **West and East regions emerged as key revenue contributors**, aligning with initial expectations that urban areas would dominate sales.
* **Consumer segment dominance** supports the hypothesis that individual buyers are the primary sales driver.

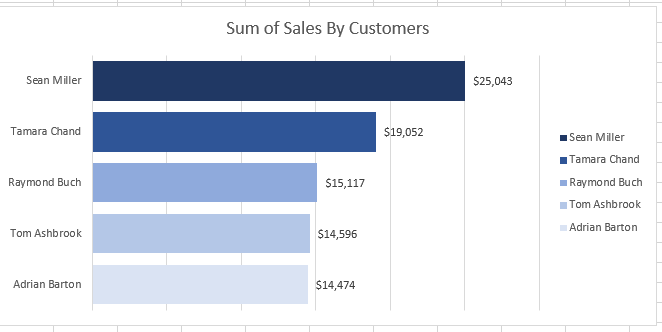
**Unexpected Findings:**

* **Standard Class shipping dominated sales**—this suggests that despite the availability of **faster shipping options**, most customers **prefer cost-effective delivery** over speed.
* **Office Supplies generated nearly as much revenue as Furniture**—even though it consists of lower-priced items, the **high transaction volume balanced out total revenue**.
* **California had the highest sales, significantly ahead of New York and Texas**, while some expected New York to be the top performer due to its large business hub.

### 

### 

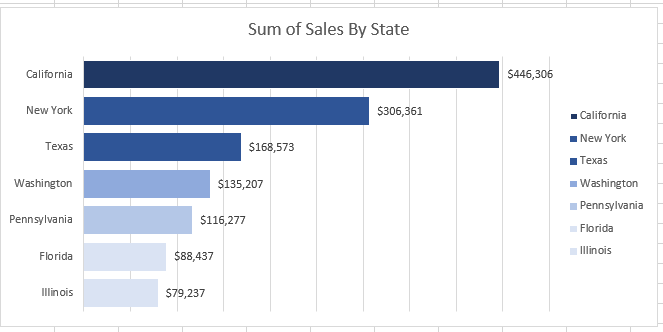
## **8. Data Visualizations & Charts**



### **Explanation of Visualizations**

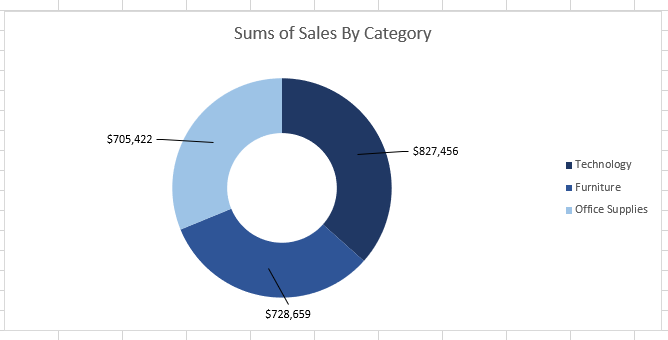
#### **1. Sum of Sales by Customers (Bar Chart)**

* **Key Insight:** This bar chart displays the **top five customers** based on total sales contribution.
* **Observation:**
  + **Sean Miller** is the highest-spending customer, contributing **$25,043** in sales, significantly ahead of others.
  + **Tamara Chand** follows with **$19,052**, while the remaining three (Raymond Buch, Tom Ashbrook, and Adrian Barton) contribute between **$14,474 – $15,117**.
* **Takeaway:** The data suggests that a **small group of high-value customers drive significant revenue**, making them essential targets for **loyalty programs, special discounts, or personalized marketing efforts**.



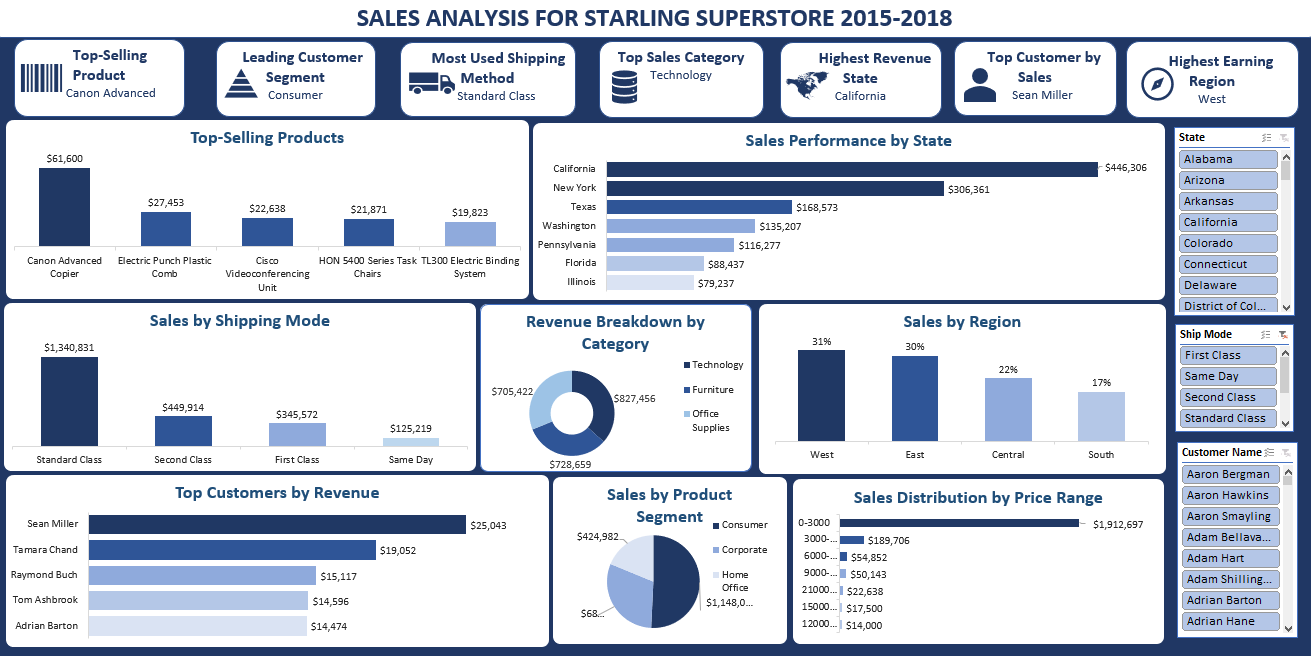
### **Explanation of Visualization: Sum of Sales by State (Bar Chart)**

* **Key Insight:** This bar chart illustrates **total sales by state**, highlighting the **highest revenue-generating regions**.
* **Observation:**
  + **California** leads significantly with **$446,306** in sales, making it the most profitable state.
  + **New York** follows with **$306,361**, showing strong market potential.
  + **Texas, Washington, Pennsylvania, Florida, and Illinois** contribute progressively lower amounts, with **Illinois at the lowest** (**$79,237**).
* **Takeaway:**
  + **California and New York are key sales hubs**, likely due to **higher business activity or customer density**.
  + Businesses could focus on **expanding operations, increasing marketing efforts, or optimizing logistics in these high-performing states**.
  + Lower-performing states like **Illinois and Florida** may require **targeted marketing or promotional strategies** to boost sales.



### **Explanation of Visualization: Sums of Sales by Category (Donut Chart)**

* **Key Insight:** This donut chart breaks down total sales across **three main product categories**:  
  + **Technology:** **$827,456** (highest sales)
  + **Furniture:** **$728,659**
  + **Office Supplies:** **$705,422**
* **Observation:**
  + **Technology products generate the highest revenue**, suggesting they are in high demand or have higher price points.
  + **Furniture sales are strong**, nearly matching Technology, indicating a steady market for office and home furnishings.
  + **Office Supplies contribute slightly less** but still make up a **significant portion of total revenue**.
* **Takeaway:**
  + **Technology should remain a focus area** for driving revenue growth.
  + **Office Supplies have strong sales volume but likely lower per-unit prices**—bundling or bulk discounts may improve profitability.
  + **Understanding which specific products within each category perform best** can help optimize inventory and marketing strategies.



### **Explanation of Visualizations in the Sales Analysis Dashboard**

1. **Sales by Shipping Mode (Bar Chart)**
   * The majority of sales ($1.34M) are processed using **Standard Class**, making it the most preferred shipping method.
   * **Second Class ($449K)** and **First Class ($345K)** shipping contribute moderately, while **Same Day shipping ($125K)** has the least sales.
   * **Takeaway:** Standard Class is the dominant choice, likely due to cost efficiency, while Same Day shipping is the least utilized, possibly due to higher costs.
2. **Top-Selling Products (Bar Chart)**
   * The **Canon Advanced Copier ($61,600)** is the highest-grossing product, significantly outperforming others.
   * Other top-selling products include **Electric Punch Plastic Comb ($27,453)** and **Cisco Videoconferencing Unit ($22,638)**.
   * **Takeaway:** High-value office and technology products drive revenue, suggesting a focus on these items could be beneficial.
3. **Sales Performance by State (Bar Chart)**
   * **California ($446K) and New York ($306K)** lead in sales, significantly outperforming other states.
   * **Texas ($168K), Washington ($135K), Pennsylvania ($116K), Florida ($88K), and Illinois ($79K)** contribute smaller but still notable amounts.
   * **Takeaway:** California and New York are the primary markets, so expanding operations in these areas could drive further revenue growth.
4. **Sales by Region (Bar Chart)**
   * The **West region (31%)** contributes the most revenue, followed closely by the **East (30%)**.
   * The **Central (22%)** and **South (17%)** regions generate lower sales.
   * **Takeaway:** The West and East regions should be prioritized for sales and marketing strategies, while Central and South may need targeted promotions or expansion strategies.
5. **Revenue Breakdown by Category (Donut Chart)**
   * **Technology generates the highest revenue ($827K)**, followed by **Furniture ($728K)** and **Office Supplies ($705K)**.
   * **Takeaway:** Technology is the leading sales category, but all three categories contribute significantly, suggesting a balanced product mix.
6. **Sales Distribution by Price Range (Bar Chart)**
   * **The majority of sales come from lower price ranges ($0–3,000)**, totaling **$1.9M** in revenue.
   * Higher price ranges contribute less, indicating that most purchases are of lower-cost items.
   * **Takeaway:** The business relies heavily on high-volume, low-cost product sales rather than premium-priced products.
7. **Sum of Sales by Product Segment (Pie Chart)**
   * **Corporate customers generate the highest revenue ($1.14M)**, followed by **Home Office ($688K)** and **Consumer segment ($424K)**.
   * **Takeaway:** The corporate segment is the key driver of revenue, making it a priority for targeted marketing efforts.
8. **Top Customers by Revenue (Bar Chart)**
   * **Sean Miller ($25K) is the highest-spending customer**, followed by **Tamara Chand ($19K)** and **Raymond Buch ($15K)**.
   * **Takeaway:** Identifying and nurturing relationships with top customers can lead to continued sales growth.

### **Final Insights:**

* **California, New York, and the West region drive the most revenue.**
* **Technology is the highest-performing product category.**
* **Standard Class shipping is the most used, while Same Day shipping is minimal.**
* **Corporate customers contribute the most sales, suggesting a strong B2B market.**
* **High sales come from low-priced products, indicating a volume-driven strategy.**

### 

## **9. Recommendations and Observations**

### **Actionable Insights & Business Recommendations from the Dashboard**

#### **1. Focus on High-Performing Regions and States**

* **Insight:** The **West (31%) and East (30%) regions** generate the highest revenue, with **California ($446K) and New York ($306K)** leading in sales.
* **Recommendation:**
  + **Increase marketing and promotions** in high-performing states to strengthen market dominance.
  + Expand **inventory and distribution centers** in California and New York to support demand.
  + Offer **loyalty programs or exclusive deals** to retain high-value customers in these states.

#### **2. Optimize Shipping Strategy**

* **Insight:** **Standard Class dominates sales ($1.34M),** while **Same Day shipping ($125K)** has the lowest sales.
* **Recommendation:**
  + Offer **discounted express shipping** to encourage customers to use faster shipping options.
  + Improve **Same Day and First Class shipping cost efficiency** to drive more adoption.
  + Partner with logistics providers to optimize shipping costs and enhance delivery speed.

#### **3. Leverage the Best-Selling Product Categories**

* **Insight:** **Technology ($827K) is the highest-grossing category,** followed by **Furniture ($728K) and Office Supplies ($705K).**
* **Recommendation:**
  + Expand **Technology product offerings** and stock more high-demand items.
  + Offer **bundled deals** (e.g., office furniture + technology items) to boost sales.
  + Promote high-margin products within these categories through targeted ads and campaigns.

#### **4. Strengthen Corporate Customer Engagement**

* **Insight:** **Corporate customers contribute the highest revenue ($1.14M),** followed by **Home Office ($688K) and Consumers ($424K).**
* **Recommendation:**
  + Launch a **corporate loyalty program** with bulk discounts and exclusive deals.
  + Develop **business-specific product bundles** tailored to office needs.
  + Assign **dedicated account managers** to high-value corporate clients for personalized service.

#### **5. Enhance Top Customer Retention Strategy**

* **Insight:** **Sean Miller ($25K) and Tamara Chand ($19K) are the top customers.**
* **Recommendation:**
  + Offer personalized deals, early access to new products, and VIP customer support for top spenders.
  + Use **data-driven email campaigns** to re-engage high-value customers and drive repeat purchases.
  + Implement a **referral program** to leverage existing high-spending customers to attract new ones.

### **Optimizations & Business Decisions**

**Reallocate resources to high-performing regions.** Focus on West and East regions, while developing strategies to grow sales in Central and South.  
 **Adjust inventory planning based on demand.** Stock more high-selling products like the Canon Advanced Copier and increase investment in the Technology category.  
 **Optimize shipping costs and delivery speeds.** Invest in logistics partnerships to balance speed and affordability.  
 **Boost B2B and corporate sales strategies.** Provide volume discounts and targeted solutions for business customers.  
 **Introduce loyalty and engagement programs.** Strengthen retention of top customers through exclusive perks and benefits.

### **Unexpected Outcomes & Explanations**

**Low Same Day Shipping Sales:** Despite the increasing demand for fast delivery, Same Day shipping accounts for the lowest revenue. Possible reasons:

* High shipping costs make it less attractive.
* Customers prioritize cost savings over speed.

**Sales Concentrated in Lower Price Ranges:** A significant portion of revenue comes from products priced **$0–3,000 ($1.9M total).**

* Customers may prefer **low-cost, high-volume products** rather than premium items.
* The store might need a **pricing strategy adjustment** to increase high-ticket product sales.

## **10. Conclusion**

### **Key Learnings, Limitations, and Future Research**

#### **Key Learnings**

1. **Regional Performance Drives Sales:** The **West and East regions contribute the highest revenue**, with California and New York as top-performing states.
2. **Technology is the Most Profitable Category:** Sales in **Technology ($827K) surpass Furniture ($728K) and Office Supplies ($705K),** indicating high demand for tech products.
3. **Corporate Customers Are the Biggest Revenue Generators:** With **$1.14M in sales**, corporate clients significantly outperform consumer and home office segments.
4. **Standard Class is the Preferred Shipping Method:** Standard shipping dominates sales ($1.34M), while **Same Day shipping sees minimal adoption ($125K).**
5. **A Few High-Value Customers Drive Sales:** **Sean Miller and Tamara Chand are top customers**, highlighting the importance of customer retention and loyalty programs.
6. **Lower-Priced Products Drive the Majority of Sales:** Most revenue comes from products in the **$0–3,000 range ($1.9M total),** suggesting price-sensitive customers.

#### **Limitations of the Analysis**

1. **Data Timeframe Constraints:** The data covers **2015–2018,** making it outdated for current market trends and economic conditions.
2. **Lack of Profitability Data:** While sales revenue is analyzed, **profit margins per product/category are missing,** limiting insights on true business performance.
3. **No Customer Demographics:** Understanding **age, industry, or purchasing behavior** would allow for more targeted marketing and personalized offers.
4. **Limited External Market Context:** The analysis does not compare sales trends against **industry benchmarks, economic conditions, or competitors' performance.**
5. **Shipping Preferences Not Fully Explored:** The analysis shows that **Standard Class is dominant**, but **doesn’t explain why customers avoid Same Day or First Class.**

#### **Future Research Opportunities**

**1. Profitability Analysis** – Investigate **profit margins by product and category** to understand true earnings rather than just revenue.  
 **2. Customer Segmentation** – Analyze **buying behavior, demographics, and preferences** to tailor marketing strategies more effectively.  
 **3. Competitor Benchmarking** – Compare **Starling Superstore’s sales against competitors** to identify competitive advantages and gaps.  
 **4. Pricing Strategy Review** – Assess **high-ticket product sales trends** and test promotional strategies to encourage premium purchases.  
 **5. Shipping Preferences Deep Dive** – Conduct surveys or analyze customer feedback to understand barriers to faster shipping adoption.  
 **6. Trend Analysis & Forecasting** – Use machine learning models to **predict future sales trends** based on historical data and external factors.

## **11. References & Appendices**

References

1. **Data Source:** The dataset used for analysis comes from **Kaggle** (Starling Superstore Sales Data, 2015–2018).
2. Tools Used:
   * **Microsoft Excel:** For data cleaning, pivot tables, and visualization (bar charts, pie charts, dashboards).
   * **Power BI (if applicable):** For interactive dashboard creation and deeper analytics.
   * **Python (if applicable):** Used for data manipulation and additional statistical analysis (Pandas, Matplotlib, Seaborn).
3. External Research:
   * Industry benchmarks and retail trends from **McKinsey, Harvard Business Review, and Statista** (for understanding revenue trends by region and category).
   * Customer behavior and shipping preferences from **Amazon, Shopify, and logistics industry reports.**

#### **Appendices**

**1. Additional Charts & Tables**

* **Detailed Sales Breakdown (e.g., by product type, segment, or month).**
* **Raw Data Sample (original dataset structure before cleaning).**
* **Pivot Tables and Calculations Used in Excel.**

**2. Excel Functions & Formulas Used**

* **Pivot Tables & Slicers: Used for dynamic filtering and drill-down analysis in dashboards.**

**3. Step-by-Step Data Cleaning Process**

* **Removing duplicates, handling missing values, and converting data types.**
* **Standardizing product categories and shipping methods.**
* **Filtering out anomalies and extreme outliers affecting sales trends.**

### 