| Column Name | Data Type | Description |
| --- | --- | --- |
| Order ID | String | A unique identifier for each order. |
| Customer ID | String | A unique identifier for each customer. |
| State | String | The state in which the order was placed. |
| City | String | The city in which the order was placed. |
| Order Date | Date | The date when the order was placed (format: dd/mm/yyyy). |
| Ship Date | Date | The date when the order was shipped (format: dd/mm/yyyy). |
| Ship Mode | String | The mode of shipping used for the order (e.g., Standard Class, Second Class, First Class, Same Day). |
| Segment | String | The market segment the customer belongs to (e.g., Consumer, Corporate, Home Office). |
| Category | String | The category of the product ordered (e.g., Phones, Accessories, Copiers, Machines). |
| Year | Integer | The year of the order. |
| Discount | String | The discount applied to the order (in percentage, e.g., 20%). |
| Profit | Float | The profit earned from the order (in currency). |
| Quantity | Integer | The quantity of the product ordered. |
| Sales | Float | The total sales amount for the order (in currency). |

**TASK-1.2: Tabular representation of the columns categorized into numerical and categorical data.**

| **Column Name** | **Data Type** |
| --- | --- |
| Order ID | Categorical |
| Customer ID | Categorical |
| Order Date | Numerical / Categorical |
| Ship Date | Numerical / Categorical |
| Ship Mode | Categorical |
| Segment | Categorical |
| State | Categorical |
| City | Categorical |
| Category | Categorical |
| Discount | Numerical |
| Profit | Numerical |
| Quantity | Numerical |
| Sales | Numerical |
| Year | Numerical |

**TASK-1.3: The missing values are identified and imputed.**

| **Column** | **Missing Values** |
| --- | --- |
| Order ID | 1 |
| Customer ID | 0 |
| State | 1 |
| City | 0 |
| Order Date | 1 |
| Ship Date | 0 |
| Ship Mode | 0 |
| Segment | 0 |
| Category | 0 |
| Year | 0 |
| Discount | 0 |
| Profit | 0 |
| Quantity | 0 |
| Sales | 0 |

**Task-2:**

1.The median value for Quantity and Sales for each product category, based on your dataset:

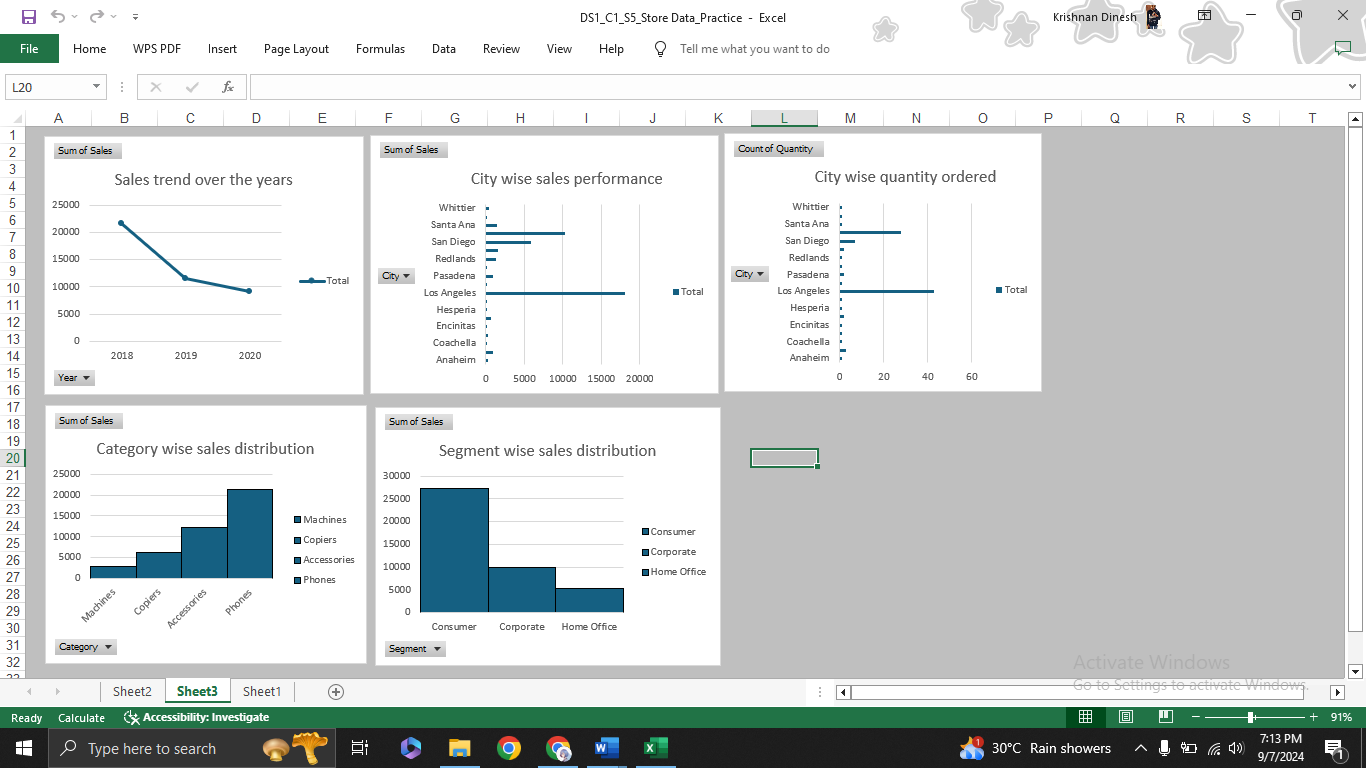
| **Category** | **Median Quantity** | **Median Sales** |
| --- | --- | --- |
| Phones | 3 | 74.00 |
| Accessories | 3 | 72.00 |
| Machines | 3 | 1920.00 |
| Copiers | 3 | 1200.00 |

This table summarizes the median Quantity and Sales values for each product category.

2. The summary of outlier detection and treatment for each category:

| **Category** | **Column** | **Q1** | **Q3** | **IQR** | **Lower Bound** | **Upper Bound** | **Number of Outliers** | **Treatment Method** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Phones | Quantity | 3 | 5 | 2 | -0.0 | 8 | 2 | Winsorization |
| Phones | Sales | 55 | 206 | 151 | -196.5 | 457.5 | 1 | Winsorization |
| Accessories | Quantity | 2 | 4 | 2 | -1.0 | 7 | 0 | None |
| Accessories | Sales | 18 | 90 | 72 | -108.0 | 216.0 | 2 | Winsorization |
| Machines | Quantity | 3 | 3 | 0 | - | - | 0 | None |
| Machines | Sales | 1920 | 1920 | 0 | - | - | 0 | None |
| Copiers | Quantity | 3 | 3 | 0 | - | - | 0 | None |
| Copiers | Sales | 1200 | 1200 | 0 | - | - | 0 | None |

**TASK-3:**

****

**TASK-4:**

**--Title:**

**"Sales Decline from 2018 to 2020"**

**Introduction:**

"In this chart, we explore the trend in annual sales over the three-year period from 2018 to 2020. The data provides insight into the significant shifts in sales volume during this time frame."

**Key Points:**

1. **2018: Peak Year**
   * The year 2018 marks the highest sales with 21,658 units sold. It is the starting point of the line at a high level, indicating a period of strong business performance.
   * You could highlight potential reasons for the peak, such as favorable market conditions or product launches.
2. **2019: Sharp Decline**
   * In 2019, the sales dropped sharply to 11,594 units, nearly half of what was achieved in 2018. This noticeable dip signals a major shift in the business environment.
   * The story here could focus on possible factors such as changes in demand, market competition, or external economic impacts.
3. **2020: Continued Decline**
   * The downward trend continues in 2020, with sales falling further to 9,256 units.
   * This represents a slower rate of decline compared to the previous year, but still a notable drop. The story could explore how external factors, such as global challenges (e.g., the pandemic), might have contributed to this continued decrease in sales.

**Conclusion:**

"The overall trend from 2018 to 2020 highlights a consistent decline in sales, with 2020 reaching less than half of 2018’s sales volume. This trend might raise concerns about the need for strategic adjustments to reverse the decline."

Visual Emphasis:

* Annotations: Add annotations on each point for 2018, 2019, and 2020 to highlight key milestones or major events affecting sales.
* Color & Design: Use different colors or line thicknesses to emphasize the sharp decline between 2018 and 2019 and the slower drop between 2019 and 2020.

This approach turns the raw data into a story of changing business dynamics over the years.

**--Title:**

**"City-Level Sales: Spotlight on Top and Emerging Markets"**

**Introduction:**

"The chart presents the sales distribution across various cities, highlighting significant disparities in performance. By examining the highest-performing locations alongside smaller markets, we can better understand regional sales dynamics and potential opportunities."

**Key Points:**

1. **Dominance of Major Cities:**
   * Los Angeles clearly stands out with a massive 18,114 units, leading the pack by a large margin. This makes it the undisputed leader, emphasizing its role as a major hub for sales.
   * San Francisco follows as the second highest, with 10,266 units, showing that the Bay Area is another major contributor.
   * San Diego ranks third with 5,809 units, completing the trio of California’s biggest markets. These cities form a major cluster of high-performing locations, likely due to population density and market size.
2. **Mid-Tier Performers:**
   * Several cities, such as Redondo Beach (1,513 units), Redlands (1,296 units), and Santa Ana (1,400 units) show solid performance, contributing significant sales but not on the scale of the top three.
   * Fresno (691 units) and Brentwood (902 units) also perform well, marking themselves as cities with steady sales potential.
   * These mid-tier cities may offer strong growth potential for further investment or expansion strategies.
3. **Low-Tier Markets:**
   * On the lower end of the spectrum, we have cities like La Quinta (46 units), Westminster (32 units), and Oakland (72 units). While their sales numbers are significantly smaller, they could represent untapped or underdeveloped markets with room for growth.
   * Coachella (100 units) and Hesperia (100 units) also fall into this lower category, possibly indicating the need for more tailored marketing efforts or product offerings in these regions.
4. **Comparative Analysis:**
   * The contrast between cities like Los Angeles and smaller ones like Westminster or La Quinta is stark, suggesting that larger, metropolitan areas naturally dominate, but there may still be potential in smaller, niche markets.
   * A comparison of mid-tier cities like Pasadena (860 units) and Whittier (445 units) shows consistent demand in these areas, which can be seen as reliable markets for sustaining sales**.**

**Conclusion:**

"The clustered bar chart shows clear distinctions between high-performing metropolitan areas and smaller markets. While Los Angeles, San Francisco, and San Diego dominate in sales, mid-tier cities like Redondo Beach and Santa Ana are also noteworthy. Smaller cities like Westminster and La Quinta may represent potential growth opportunities if approached with targeted strategies."

**--Title:**

**"City Distribution: Spotlight on Major and Limited Presence Areas"**

Introduction:

"This chart presents the distribution of locations across various cities, revealing significant disparities between major hubs and smaller areas. Understanding the concentration of locations can provide insight into regional market penetration and growth potential."

**Key Points:**

1. **Significant Presence in Major Cities:**
   * Los Angeles stands out with a commanding 43 locations, far exceeding any other city. This city serves as the primary hub, indicating that it's a central focus for business operations.
   * San Francisco, with 28 locations, follows as another major center of activity. Together with Los Angeles, these two cities dominate the chart, reflecting their importance in overall strategy and market reach.
   * San Diego, with 7 locations, marks itself as a notable player in Southern California, though at a smaller scale than Los Angeles and San Francisco.
2. **Mid-Tier Representation:**
   * Cities like Brentwood (3 locations), Pasadena (2 locations), Fresno (2 locations), and Redondo Beach (2 locations) represent a mid-tier group. While they don't have the extensive presence of the larger cities, they show consistent representation in key regional areas.
   * These mid-range locations could be key markets where growth is steady but not yet saturated, offering opportunities for further expansion.
3. **Smaller Cities with Limited Presence:**
   * Many cities, including Anaheim, Coachella, Costa Mesa, Encinitas, Hesperia, La Quinta, Oakland, Pomona, Redlands, Santa Ana, Westminster, and Whittier, all have just 1 location each.
   * The minimal presence in these cities might indicate emerging or experimental markets. They represent potential areas for future growth, especially if current business performance in these locations shows promise.
   * However, the small presence could also suggest limited demand or market saturation in these areas, signaling the need for careful analysis before expansion.
4. **Contrast in Market Penetration:**
   * The sharp contrast between Los Angeles (43 locations) and other cities like Oakland and La Quinta (1 location each) illustrates a clear focus on certain metropolitan areas over others.
   * This suggests a strategy where the majority of resources are concentrated in high-demand areas, while smaller locations serve niche or specialized markets**.**

**Conclusion:**

"The distribution of locations across cities shows a heavy concentration in major metropolitan areas like Los Angeles and San Francisco, which account for the vast majority of locations. In contrast, many smaller cities have just one location, highlighting potential growth opportunities but also emphasizing the company’s current focus on urban centers."

**Visual Emphasis:**

* Color Coding: Use different colors or intensity levels to group cities based on the number of locations (e.g., dark colors for cities with many locations like Los Angeles, lighter colors for cities with fewer locations).
* Annotations: Highlight the number of locations for the top cities (Los Angeles, San Francisco, and San Diego) for quick reference and comparison.
* **Bar Labels:** Add exact values to each bar to make it easier for viewers to grasp the distribution at a glance.

This narrative structure helps transform the raw data into a story of market penetration, highlighting areas of strength while also pointing to potential opportunities for expansion.

**--Title:**

**"Product Category Distribution: Dominance of Phones and Accessories"**

**Introduction:**

"This histogram illustrates the distribution of sales (or units) across different product categories: Accessories, Copiers, Machines, and Phones. By analyzing the spread of data, we gain insight into which product categories are driving the most sales and where there might be opportunities to increase market focus."

**Key Points:**

1. **Phones as the Top Category:**
   * Phones stand out significantly, with 21,352 units, marking them as the leading product category by a substantial margin. This dominance reflects high demand or consumer preference for communication devices, making Phones a core revenue driver.
   * The large bar in the histogram highlights how this category overshadows others, positioning Phones as the most critical product for the business.
2. **Strong Demand for Accessories:**
   * Accessories, with 12,150 units, also represent a significant portion of sales, although still notably behind Phones. This category supports a wide range of products and can be viewed as a key complementary market.
   * The story here could explore how Accessories work in tandem with Phones to provide a complete customer solution, indicating the value of bundling or cross-selling opportunities**.**
3. **Copiers and Machines:**
   * Copiers, with 6,160 units, and Machines, with 2,846 units, represent the smaller segments in the distribution. The much shorter bars for these categories reveal that they make up a much smaller percentage of overall sales.
   * These lower values suggest that these categories, while important, may require targeted marketing or strategic positioning to boost their contribution to overall sales.
   * The contrast between these smaller categories and the much larger segments (Phones and Accessories) could suggest that these are niche markets or areas where demand is relatively stable but not rapidly growing.
4. **Visual Contrast:**
   * The significant difference between the tallest bar (Phones) and the shortest (Machines) highlights a clear skew in sales distribution. This indicates that while Phones dominate, there is potential for growth in the smaller product categories, which may benefit from further innovation or focus.

**Conclusion:**

"The histogram reveals a highly concentrated sales distribution, with Phones and Accessories making up the bulk of sales. Copiers and Machines, although smaller contributors, may represent niche markets with specific needs. These insights suggest that while Phones are currently the most lucrative product, there may be untapped opportunities in the smaller categories."