Project Summary: Sales Analysis for 2019

Objective

The goal of this project was to analyze sales data from 2019 to uncover trends, patterns, and actionable insights. By processing and analyzing the data, the project aimed to answer key business questions such as the best-performing months, cities with the highest sales, optimal advertising times, frequently purchased product combinations, and top-selling products.

Steps Taken

1. **Data Collection and Integration**:
    - Sales data from multiple CSV files (each representing a month) was consolidated into a single dataset for comprehensive analysis.
    - The combined dataset was saved into a new file, `Combined_Sales_2019.csv`, for future use.
2. **Data Cleaning**:
    - Removed rows with missing or invalid data (e.g., rows with null values or corrupted entries like `Order Date` starting with "Or").
    - Ensured proper data types for numerical and date columns:
        - `Quantity Ordered`: Converted to integer.
        - `Price Each`: Converted to float.
        - `Order Date`: Converted to datetime.
3. **Feature Engineering**:
    - Created new columns:
        - **Month**: Extracted from `Order Date`.
        - **Sales**: Calculated as `Quantity Ordered * Price Each`.
        - **City**: Extracted and combined city and state from `Purchase Address`.
4. **Analysis and Visualization**:
    - Answered key business questions using data aggregation and visualizations.

Key Findings

1. Best Month for Sales

- **Insight**: December was the best month for sales, generating the highest revenue.
- **Reason**: Likely due to the holiday season and year-end shopping.
- **Visualization**: Bar chart showing monthly sales trends.

2. City with the Highest Sales

- **Insight**: San Francisco (CA) had the highest sales.
- **Reason**: Likely due to a combination of population density, purchasing power, and market penetration.
- **Visualization**: Bar chart comparing sales across cities.

3. Optimal Advertising Times

- **Insight**: The best times to advertise are:
    - **11:00 AM**: Mid-morning break period.
    - **7:00 PM**: Evening relaxation time.
- **Recommendation**: Schedule advertisements around these times to maximize customer engagement.
- **Visualization**: Line chart showing sales by hour.

4. Products Most Often Sold Together

- **Insight**: Certain products were frequently purchased together, indicating complementary relationships (e.g., accessories with main products).
- **Example**: Identified the top 10 most common product combinations.
- **Visualization**: Text-based analysis of product combinations.

5. Top-Selling Products

- **Insight**: AAA Batteries (4-pack) sold the most units, likely due to low price and high utility.
- **Additional Insight**: Lower-priced products had higher sales volume.
- **Visualization**:
    - Bar chart showing quantity sold by product.
    - Dual-axis plot comparing product quantity and average price.

Business Impact

1. **Strategic Planning**:
    - Identify peak months and allocate resources accordingly to maximize revenue.
2. **Advertising Strategy**:
    - Schedule campaigns during high-traffic hours for better ROI.
3. **Product Bundling**:
    - Leverage insights on frequently purchased product combinations to create promotional bundles.
4. **Stock Management**:
    - Ensure sufficient inventory of high-demand products like AAA Batteries (4-pack).

Tools and Technologies Used

- **Programming Language**: Python
- **Libraries**:
    - **Pandas**: Data manipulation and cleaning.
    - **Matplotlib** and **Seaborn**: Data visualization.
    - **Counter and Combinations**: Analyzing product combinations.
- **Data Format**: CSV files for input and output.

Conclusion

This project provided actionable insights into sales performance, customer behavior, and product relationships, empowering data-driven decision-making for inventory management, marketing, and sales strategy.