Exploratory Data Analysis (EDA): Understanding the Sales Landscape

Geographic Sales Performance Analysis

Which city achieved the highest total sales, and what was the total sales amount for that city?

1. Overview

This analysis explores geographic sales performance across multiple cities using exploratory data analysis (EDA). By aggregating total sales figures city-wise, we aim to uncover which city contributed the highest to the overall revenue. Visualizing this information not only provides insights into regional performance but also supports strategic decisions for resource allocation, marketing investment, and operational optimization.

2. Goal

- Identify the city with the highest total sales.
- Quantify the total sales amount associated with each city.
- Visualize sales distribution across cities using an intuitive bar chart.
- Support business strategy through location-based performance insights.

3. Business Challenge

- Sales performance is uneven across regions, and decision-makers lack clarity on high-performing cities.
- Without clear geographic insights, marketing and inventory investments may be misaligned.
- Limited visibility into revenue contribution by location may affect expansion or scaling plans.

4. Analysis Approach

- Load and preprocess the dataset to ensure accurate and complete data.
- Group sales data by city and compute total sales for each location.
- Visualize the data using a bar chart to highlight the top-performing city.
- Format the chart for readability and presentation (currency formatting, labels, and grid).

- Derive insights from the chart and apply them to inform business decisions.

Importing libraries

```
In [9]:  import pandas as pd
         import matplotlib.pyplot as plt
```

Loading the clean dataframe (post-ETL process)

```
In [11]:  df = pd.read_csv("C:\\Monthly_Sales\\cleaned_data.csv")
```

```
In [12]:  df.head()
```

Out[12]:

	Order ID	Product Name	Units Purchased	Unit Price	Order Date	Delivery Address	Month	Month Name	Year	Da Wee
0	160155	Alienware Monitor	1	400.99	2024-01-01 05:04:00	765 Ridge St, Portland, OR 97035	1	January	2024	M
1	151041	AAA Batteries (4-pack)	1	4.99	2024-01-01 05:04:00	964 Lakeview St, Atlanta, GA 30301	1	January	2024	M
2	146765	AAA Batteries (4-pack)	1	4.99	2024-01-01 05:20:00	546 10th St, San Francisco, CA 94016	1	January	2024	M
3	145617	Amana Washing Machine	1	600.00	2024-01-01 05:24:00	961 Meadow St, Portland, OR 97035	1	January	2024	M
4	156535	Lightning Charging Cable	2	14.95	2024-01-01 05:45:00	451 Elm St, Los Angeles, CA 90001	1	January	2024	M

Replacing 'Boston (\rA)' with 'Boston (MA)'

```
In [14]:  import re
```

```python
df['City'] = df['City'].str.replace(r'Boston\s+\(\rA\)', 'Boston  (MA)', regex=True
```

Total Sales for each City

```python
In [16]: City_sales = df.groupby('City', observed=False)['Total Sales'].sum()
         City_sales
```

```
Out[16]: City
         Atlanta  (GA)          3146660.06
         Austin  (TX)           2182210.62
         Boston  (MA)           4169731.86
         Dallas  (TX)           3195316.61
         Los Angeles  (CA)      6347340.83
         New York City  (NY)    5206533.98
         Portland  (ME)          543656.46
         Portland  (OR)         2053496.39
         San Francisco  (CA)    9395524.82
         Seattle  (WA)          3115289.81
         Name: Total Sales, dtype: float64
```

Plotting Sales by City

```python
In [18]: import matplotlib.ticker as ticker

         ax = City_sales.plot(kind='bar', title="City Sales")

         ax.set_xlabel('City')
         ax.set_ylabel('Total Sales in USD ($)')

         ax.get_yaxis().set_major_formatter(plt.FuncFormatter(lambda x, _: f'{int(x):,}'))

         plt.grid(linewidth=0.1)

         plt.savefig(r"C:/Users/DELL/OneDrive - COVENANT UNIVERSITY/Desktop/1. Retail Sales

         plt.show()
```
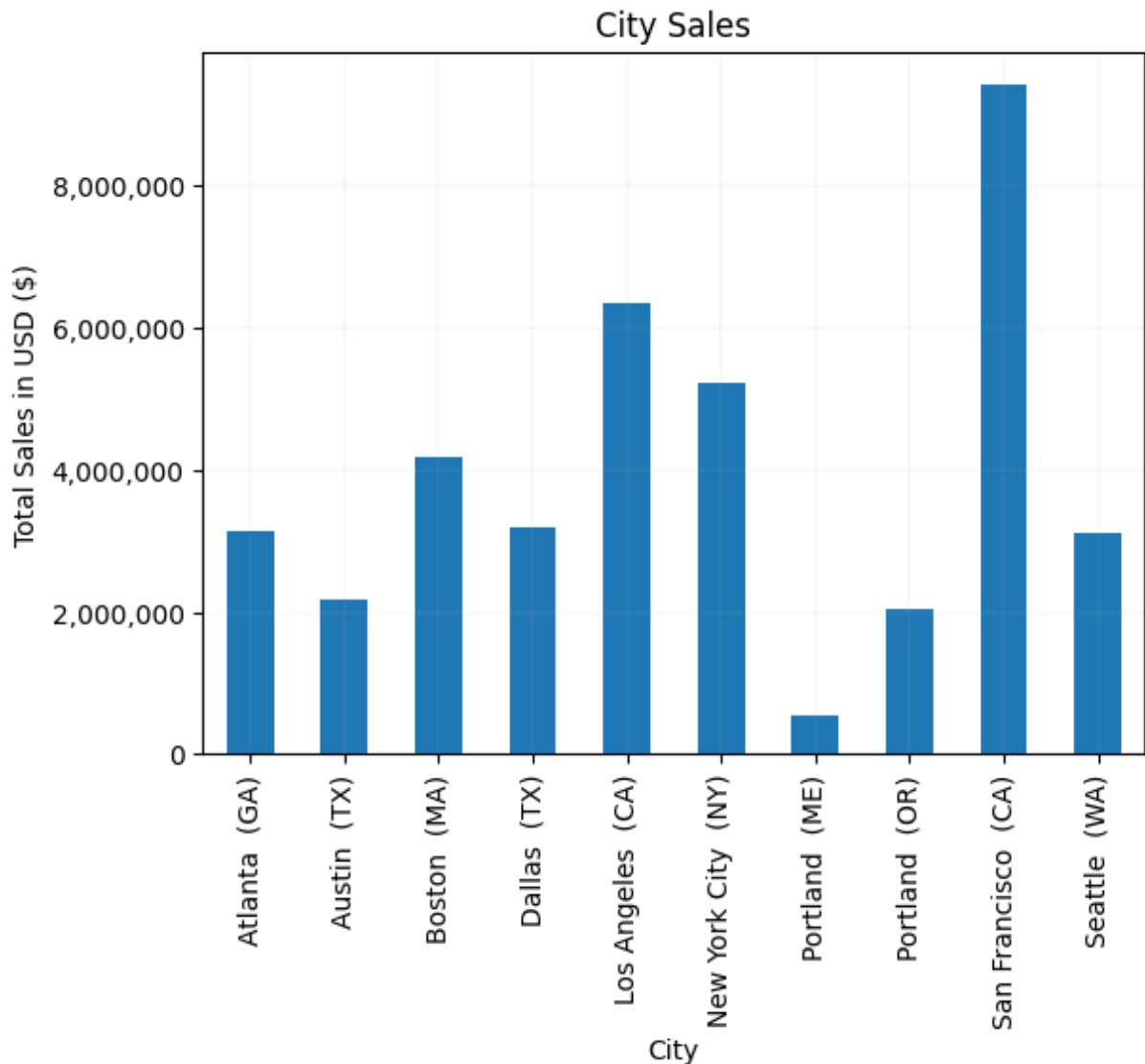
Key Insights

1. San Francisco (CA) recorded the highest total sales, pulling in approximately $9.40M, significantly outpacing all other cities.

2. Los Angeles (CA) and New York City (NY) followed with $6.35M$ and $5.21M$ respectively, making California and New York powerhouses of revenue generation.

3. Portland (ME) posted the lowest total sales at $543K, over 17 times lower than San Francisco.

There is a notable performance gap between top-tier and mid/lower-tier cities. Cities like Austin (TX), Portland (OR), and Seattle (WA) are in the mid-range, suggesting room for growth.

Strategic Recommendations

1. Double Down on High-Performing Cities: Focus marketing spend and promotions in San Francisco, LA, and NYC, these cities are already strong, and pushing further can yield compounding gains.

2. Analyze Seasonal Trends: Investigate month-by-month growth patterns to pinpoint peak sales months, then align product launches and campaigns with those windows.

3. Activate Underperforming Markets: Design city-specific campaigns to boost engagement in Portland (ME) and Austin (TX) — consider pricing tweaks, localized offers, or market research.

4. Segment Customers by Region: Perform deeper customer segmentation to uncover regional preferences — what works in SF might flop in Portland.

5. Forecast Sales Using Monthly Trends: Build a monthly forecasting model to proactively manage supply chain, inventory, and budgeting — reducing waste and maximizing responsiveness.

In [ ]: