Project Title: Sales Data Analysis and Visualization
Objective: Analyze sales data to extract insights about revenue trends, top-selling products, customer behavior, and regional sales performance.

Project Overview

In this project, you'll work with a dataset containing sales records of an online retail store. You'll perform data cleaning, exploratory data analysis (EDA), and visualization to identify key trends and patterns.

Skills Demonstrated

Data Cleaning (handling missing values, duplicates, and incorrect formats)
Exploratory Data Analysis (EDA)
Data Visualization (Matplotlib, Seaborn)
Aggregation & Pivot Tables (Pandas)
SQL (optional: for advanced queries)
Report Generation

Project Steps

1. Import Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

2. Load Dataset

```
df = pd.read_csv('sales_data.csv')
df.head()
```

(Ensure you have a dataset with columns like Order ID, Product, Category, Quantity, Price, Total Sales, Customer Location, Date.)

3. Data Cleaning

```
# Check for missing values
df.isnull().sum()

# Fill missing values (example: replace NaN in 'Customer Location' with 'Unknown')
df['Customer Location'].fillna('Unknown', inplace=True)

# Convert 'Date' column to datetime
df['Date'] = pd.to_datetime(df['Date'])

# Remove duplicates
df.drop_duplicates(inplace=True)
```

4. Exploratory Data Analysis (EDA)

A. Revenue Over Time

```
df['Month'] = df['Date'].dt.to_period('M')
monthly_sales = df.groupby('Month')['Total Sales'].sum()

plt.figure(figsize=(12,6))
sns.lineplot(x=monthly_sales.index.astype(str), y=monthly_sales.values, marker='o')
plt.xticks(rotation=45)
plt.title('Monthly Sales Trend')
plt.ylabel('Total Revenue')
plt.show()
```

B. Top 10 Best-Selling Products

```
top_products = df.groupby('Product')['Total Sales'].sum().sort_values(ascending=False).head(10)
```

```
plt.figure(figsize=(10,5))
sns.barplot(x=top_products.values, y=top_products.index, palette='Blues_r')
plt.title('Top 10 Best-Selling Products')
plt.xlabel('Total Sales')
plt.show()
```

C. Sales Distribution by Category

```
category_sales = df.groupby('Category')['Total Sales'].sum()

plt.figure(figsize=(8,6))
plt.pie(category_sales, labels=category_sales.index, autopct='%1.1f%%', startangle=140,
colors=sns.color_palette('pastel'))
plt.title('Sales Distribution by Category')
plt.show()
```

D. Regional Sales Analysis

```
location_sales = df.groupby('Customer Location')['Total
Sales'].sum().sort_values(ascending=False).head(10)

plt.figure(figsize=(12,5))
sns.barplot(x=location_sales.index, y=location_sales.values, palette='coolwarm')
plt.xticks(rotation=45)
plt.title('Top 10 Sales by Region')
plt.ylabel('Total Sales')
plt.show()
```

5. Conclusion & Insights

After analyzing the data, summarize insights such as:

Peak Sales Months: Identify months with highest revenue.
Best-Selling Products: Understand which products drive the most revenue.
Category Trends: Recognize which categories perform best.
Regional Performance: Find which locations contribute the most sales.

Bonus (Optional Enhancements)
SQL Integration: Store and query the data using MySQL.
Dashboard Creation: Use Tableau or Power BI for an interactive dashboard.
Predictive Modeling: Use Machine Learning to predict future sales trends.