Project Overview

The objective of this project is to analyze customer sales data using advanced Pandas techniques to identify sales trends, top performing customers, and regional performance. The project focuses on transforming raw transactional data into meaningful business insights through aggregation, filtering, pivot tables, and visualizations.

This analysis helps understand purchasing behavior and supports data-driven business decision-making.

Datasets Used

Sales Dataset (sales_data.csv)

- Rows: 100
- Columns: 5
- Description: Contains transactional sales information including date, product, quantity, price, customer ID, region, and total sales.

Customer Churn Dataset (customer_churn.csv)

- Rows: 500
- Columns: 4+
- Description: Contains customer subscription and churn-related attributes.
- Note: This dataset was used only for reference and understanding customer behavior. Core sales analysis was performed using sales_data.csv.

Setup Instructions

Environment Requirements

Python 3.x

Jupyter Notebook

Required Libraries

- pandas
- numpy
- matplotlib
- seaborn

Setup Steps

- Download the datasets from the platform
- Place CSV files in the project directory
- Open the Jupyter Notebook
- Run all cells sequentially

Code Structure

The project follows a clean and modular structure:

- Data loading and exploration
- Data cleaning and preprocessing
- Feature engineering
- Aggregation and analysis
- Visualization
- Insight generation

This structure ensures readability, maintainability, and professional coding standards.

Data Cleaning & Preparation

The following preprocessing steps were applied:

- Standardized column names
- Converted date column to datetime format
- Extracted year, month, and day from date
- Cleaned product names using string operations
- Verified missing values and data types

These steps ensured consistency and prevented runtime errors during analysis.

Technical Details

Pandas Operations Used

- groupby() for aggregations
- pivot_table() for multi-dimensional summaries
- Boolean indexing for multi-condition filtering
- dt accessor for datetime feature extraction
- String methods (str.strip(), str.title())

Aggregations Performed

- Total revenue calculation
- Customer-wise revenue
- Region-wise revenue
- Monthly sales trends

Analysis & Key Insights

Total Sales Performance

- The dataset shows strong overall revenue generation.
- Average order value indicates consistent customer spending behavior.

Customer Analysis

- A small percentage of customers contribute a large portion of total revenue.
- Identifying these customers is critical for retention strategies.

Regional Sales Analysis

- Certain regions consistently outperform others.
- Region-based sales variation highlights opportunities for targeted marketing.

Seasonal & Monthly Trends

- Sales demonstrate clear monthly patterns.
- Peak sales occur during specific periods, indicating seasonality.

High-Value Transactions

- High-value orders are concentrated in specific regions and months.
- These transactions represent premium customer segments.

Visual Documentation

The following visualizations were created:

- Monthly sales trend line chart
- Region-wise sales bar chart
- Top customers revenue comparison chart

These visuals effectively communicate patterns and trends in the data.

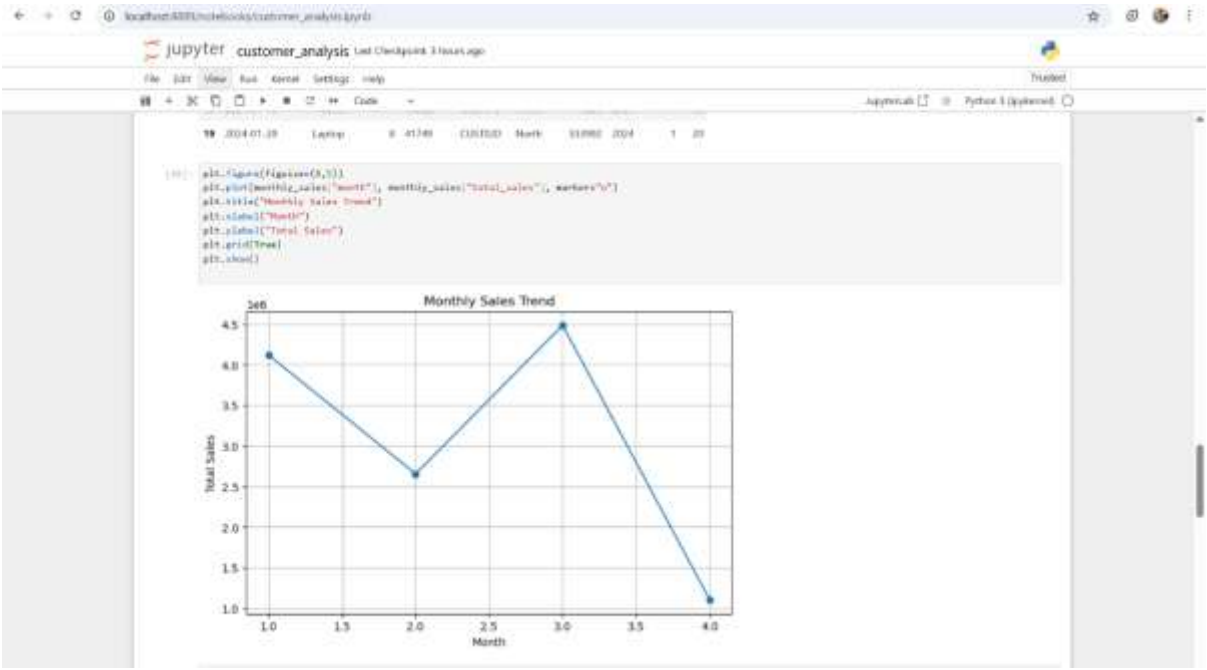

FIG 1: Shows monthly sales trend

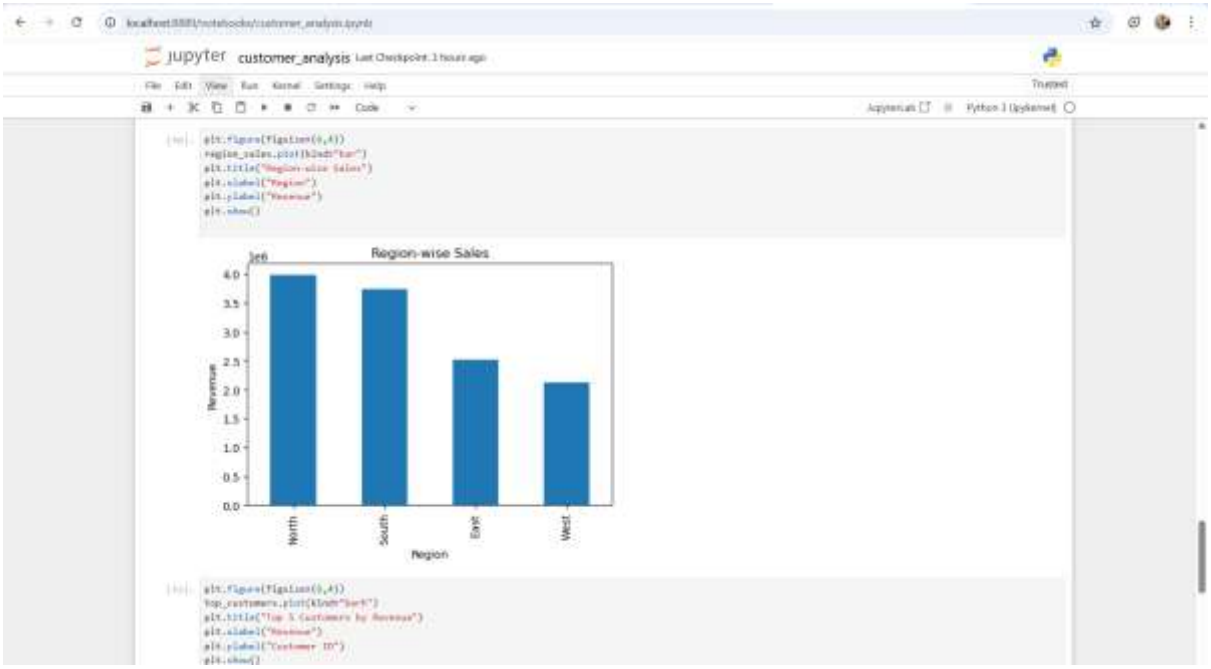

FIG 2: Shows Region-wise Sales

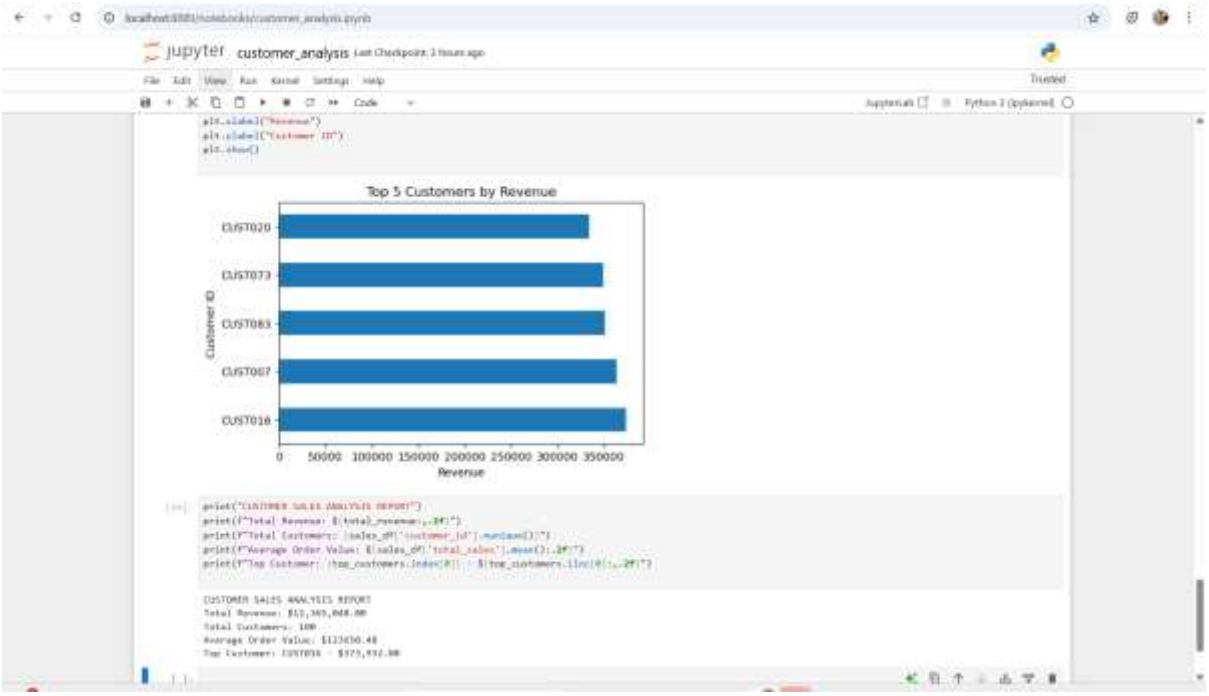

FIG 3: Shows Top Customers Revenue

Testing Evidence

To validate correctness:

- Column names and data types were verified using .info()
- Missing values were checked using. isnull().sum()
- Aggregation results were manually cross-verified
- Multiple test filters were applied to ensure accurate outputs

All analysis steps produced consistent and logical results.