Sales Analysis Project

Project Overview

This project is a comprehensive sales data analysis that aims to uncover meaningful insights from transactional records.
The analysis not only summarizes the overall sales performance but also digs deeper into monthly trends, regional contributions, product popularity,
and customer behavior.

The motivation behind this project was to demonstrate how raw business data can be cleaned, transformed, and analyzed to answer real-world business
questions that are useful for strategy and decision-making.

Dataset Information

- File Used: sales-dataset.xlsx
- Type of Data: Historical sales transactions
- Key Features:
  - Date → Purchase date
  - Product → Product name/category
  - Quantity → Number of units sold
  - Sales → Revenue generated
  - Region → Geographic location of sale
  - Customer → Customer/segment information

The dataset was checked for missing values, inconsistencies, and formatting errors. Data types were standardized (e.g., date converted into datetime format for time-based analysis).

Business Questions Answered

1. What is the total revenue generated across all sales?
2. Which month recorded the highest sales?
3. What are the top-selling products?
4. Which region contributes the most to sales?
5. How do sales trends vary across months?
6. Which customer segments drive the highest revenue?

Tools, Libraries & Concepts Used

- Python Programming
- Libraries:
  - Pandas → Data cleaning & manipulation
  - Matplotlib & Seaborn → Data visualization

- NumPy → Numerical computations
- SQL Concepts Applied:
  - Filtering data (like SQL WHERE clauses)
  - Aggregations (similar to SQL GROUP BY and SUM)
  - Sorting & ranking results (like SQL ORDER BY)

Key Analysis & Functions Performed

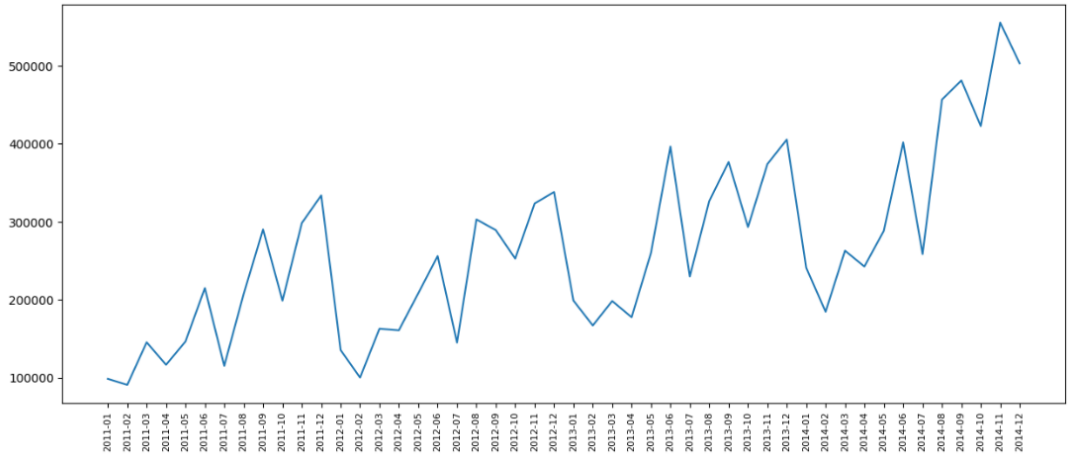
1. **Data Cleaning & Preparation**
- Removed duplicates and missing values
- Renamed columns for consistency
- Converted sales and dates into correct formats

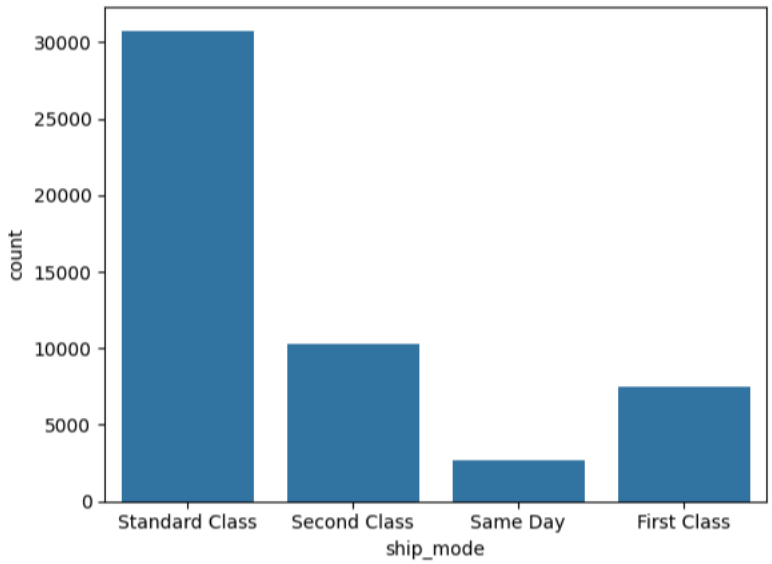
2. Exploratory Data Analysis (EDA)
- Shape, column info, and descriptive statistics
- Distribution of sales by product, region, and month

3. Custom Functions Created
- Function to calculate monthly sales
- Function to identify top N products
- Function to measure regional performance
- Function for customer segmentation

4. Business Insights Extracted
- Seasonal patterns (e.g., certain months peak higher)
- Best-selling products and under performers
- High-revenue regions vs low-performing ones
- Customer groups that bring maximum sales

Visualizations





Conclusions

From this project, we can conclude that:
- Certain months significantly outperform others, showing seasonality.
- A small group of products contributes to the majority of sales.
- Regional analysis highlights opportunities to focus on high-performing markets.
- Customer segmentation reveals which groups are most valuable to the business.

This demonstrates how data analytics can turn raw numbers into actionable insights, making it highly relevant for business strategy.

Future Improvements

- Connect the analysis to a SQL database instead of Excel
- Automate reporting with scheduled scripts
- Build an interactive dashboard using Streamlit/Power BI/Tableau
- Add forecasting techniques for future sales prediction