Role: Junior Data Analyst
Tools: Python (Pandas, Seaborn, Matplotlib), Excel, Git

Project Overview

The goal of this project is to analyze retail sales data across products, categories, regions, and time periods. The objective was to extract insights that can help the business make informed decisions on inventory, promotions, and regional strategies.

Datasets Used

sales_data.csv: Contains 1200 transactions with Date, Region, Product_ID, Units_Sold, Revenue
product_lookup.csv: Maps Product_ID to Product_Name and Category

Steps Performed

Data Preparation

1. Merged sales and product data
2. Parsed dates and extracted monthly periods
3. Created new fields like Month

Analysis

1. Total revenue and units sold by product
2. Revenue breakdown by region and category
3. Monthly trend analysis
4. Pivot tables for product-region combinations

Visualization

1. Monthly revenue trend (line chart)
2. Revenue heatmap: product vs region
3. Revenue by product category (bar plot)

Key Insights

1. Smartphones, Laptops, and Running Shoes were top performers by revenue
2. The South and West regions consistently outperformed others
3. Electronics was the best-selling category overall
4. Revenue steadily increased month-over-month, peaking in June

Folder Structure

```
Retail-Sales-Analysis/
├── data/
│   ├── sales_data.csv
│   └── product_lookup.csv
├── images/
│   ├── monthly_trend.png
│   ├── heatmap.png
│   └── top_products.png
├── retail_analysis.ipynb
└── README.md
```

Future Scope

1. Build a Streamlit dashboard
2. Replicate analysis using SQL
3. Create a forecast model for monthly sales
4. Add an Excel dashboard version

How to Run

Clone the repo
Open retail_analysis.ipynb
Run all cells after installing required packages (pandas, matplotlib, seaborn)