

# Multi-Source Sales Data Pipeline & Business Insights Dashboard

Built by **Vajja Sri Nikitha** for Flipkart Task-1

Last Updated: **17 July 2025**

## Overview

This project implements a **custom sales data pipeline** that ingests, cleans, analyzes, and visualizes sales-related data from multiple sources: **CSV, JSON, and Excel**. It is designed to be:

- Modular
- Insight-rich
- Visualization-capable
- Extensible for large-scale or real-time pipelines

## File Structure

Sales\_Data\_Pipeline\_Main.py # Main Python script with the complete pipeline  
custom\_merged\_data.csv # Auto-generated merged dataset (output)  
auto\_insights\_dashboard.png # Auto-generated visualizations (output)  
custom\_insights\_report.txt # Auto-generated insights (output) **Features**

- Multi-source data loading (CSV, JSON, Excel)
- Auto-cleaning and type conversion
- Smart merging using auto-detected keys
- Insight extraction: stats, revenue, category analysis, time series
- Visualizations: histograms, bar charts, time plots, correlation matrix
- Exports: Merged CSV, PNG dashboard, Text report
- Custom handling for quirky real-world datasets

## Requirements

Install dependencies via pip: `pip install pandas numpy matplotlib seaborn tabulate openpyxl`

### Usage 1. Configure Your File Paths

Edit the following in the `__main__` block:

```
your_csv_file = r"path\to\your\sales_data.csv" your_json_file =  
r"path\to\your\product_metadata.json" your_excel_file =  
r"path\to\your\region_info.xlsx"
```

### 2. Run the Pipeline

python

Sales\_Data\_Pipeline\_Main.py

You will be prompted whether to generate visualizations.

## Output Files

After successful execution, you'll get:

- custom\_merged\_data.csv: Merged, cleaned dataset
- custom\_insights\_report.txt: Readable report with stats and findings
- auto\_insights\_dashboard.png: Visual dashboard

## Example Insights Extracted

### DATA OVERVIEW

Total Records: 2000

Total Columns: 14

Numeric Columns: 9

Text Columns: 5

Date Columns: 0

### CATEGORICAL ANALYSIS GENDER:

Unique values: 2 Top

categories:

Female: 1016 Male:

984

- Total records, column types
- Financial summary: total revenue, average cost, etc.
- Top categories by count
- Daily performance over time
- Strong correlations between metrics

## Sample Visuals



- Sales over time plot
- Category distribution
- Revenue distribution histogram
- Correlation heatmap

## **Error Handling**

The pipeline uses specific exceptions:

- Handles missing or corrupt files gracefully
- Warns about unsupported formats or merge conflicts
- Logs failed conversions per column

## **Version Control**

This project is under Git version control:

- All major updates and commits are tracked
- Code is production-ready and AI-clean

## **Author**

**Vajja Sri Nikitha**

*Flipkart Data Pipeline Task-1 Project*