**🧠 Project Title: LocalBiz Insights**

there are 4 sources for the projects:

Dataset 1: product\_reviews.csv

Dataset 2: transactions (MySQL table)

Dataset 3: customers (MySQL table)

Dataset 4: products.csv

**🎯 Business Objective:**

Help small local business owners gain **data-driven insights** by analyzing their customer reviews, sales data, and product pricing from **CSV files and a MySQL database**.

**💼 Real-World Problem It Solves:**

Many small businesses lack easy access to tools for:

* Understanding **customer sentiment**
* Analyzing **sales trends**
* Monitoring **competitive pricing**
* Making decisions based on **real-time data**

**🛠️ Key Technical Features**

| **Component** | **Description** | **Tools/Skills** |
| --- | --- | --- |
| **Data Sources** | Load product reviews from CSV and sales/transactions from MySQL | pandas, SQLAlchemy, PyMySQL |
| **Data Processing** | Clean, filter, and enrich datasets (e.g., review sentiment, revenue) | pandas, TextBlob |
| **Data Join** | Merge sales and reviews on product/customer ID | pandas.merge() |
| **Visualization** | Sales over time, sentiment breakdown, top-performing products | matplotlib, seaborn, plotly |
| **Testing** | Unit tests for core logic | pytest |
| **CI/CD** | GitHub Actions to run tests and lint on every push | .github/workflows/ci.yml |
| **Documentation** | Clean README with screenshots, examples, and setup guide | Markdown, optional mkdocs |

**🗂 Project Structure**

graphql

CopyEdit

localbiz-insights/

│

├── data/ # Sample CSVs or downloaded data

├── src/ # Core Python scripts

│ ├── \_\_init\_\_.py

│ ├── fetch\_data.py # Load from MySQL and CSV

│ ├── analyze.py # Filter, clean, enrich, join

│ └── visualize.py # Generate plots/reports

├── tests/ # Unit tests with pytest

│ └── test\_analyze.py

├── .github/workflows/ci.yml # GitHub Action for CI/CD

├── requirements.txt # Dependencies

├── README.md

└── main.py # Entrypoint CLI

**🔄 Workflow**

**1. fetch\_data.py**

* Read transaction and customer data from **MySQL**
* Load product reviews and prices from **CSV**

**2. analyze.py**

* Clean missing values
* Enrich data (e.g., compute sentiment using TextBlob)
* Join datasets by product/customer ID

**3. visualize.py**

* Time series plot of sales
* Bar chart of average sentiment by product
* Pie chart of customer distribution

**4. main.py**

* CLI entry to run steps (e.g., python main.py --analyze)

**5. tests/**

* Unit tests for analysis, joins, sentiment scoring

**6. ci.yml**

* Automatically run pytest and flake8 on push

**✅ Example Use Cases**

* “Which product has the highest sales and the best sentiment?”
* “Which time of year sees the most revenue?”
* “Do negative reviews impact sales?”

**💡 Bonus Ideas**

* Export reports as Excel or PDF
* Add a lightweight Streamlit dashboard
* Dockerize the app for easy deployment

Would you like me to generate the **initial code templates** for these components so you can start building right away?