* **Rows**: 10
* **Columns**: 8 (customer\_id, name, gender, country, signup\_date, age, income, purchase\_count)
* Data looks already cleaned:
  + gender standardized (male, female)
  + country normalized (India, United States)
  + signup\_date parsed to proper format (YYYY-MM-DD)
  + age, income, purchase\_count are numeric
* A few **nulls remain** (1 missing signup\_date, 1 missing country, 2 missing gender).

**Next Steps (for GitHub-ready project)**

I’ll create this folder structure for you:

task1\_data\_cleaning/

│── README.md

│── requirements.txt

│

├── data/

│ ├── raw/ # raw dataset

│ └── processed/ # cleaned dataset

│ └── cleaned\_task1.csv

│

├── notebooks/

│ └── task1\_cleaning.ipynb

│

└── src/

└── clean\_data.py

* **README.md** → Explains the task, steps, and results.
* **requirements.txt** → Libraries needed (pandas, numpy, jupyter).
* **task1\_cleaning.ipynb** → Notebook with before/after cleaning + explanations.
* **clean\_data.py** → Script to re-run the cleaning pipeline.
* **data/processed/cleaned\_task1.csv** → Your cleaned file.