REALTIME AMAZON SALES DASHBOARD

PROJECT REPORT

**Overview**

This project is a real-time data pipeline and visualization dashboard that monitors and analyzes Amazon product sales using the **RapidAPI Amazon Real-Time Data API**, **Python**, **MySQL**, and **Tableau**.

It consists of 4 key components:

1. **Data Source (RapidAPI - Amazon API)**
2. **ETL Pipeline (Python API script with scheduler)**
3. **Storage (MySQL database)**
4. **Visualization (Tableau dashboard)**

A diagram of a sales dashboard

AI-generated content may be incorrect.

**1. Data Source — RapidAPI (Amazon Real-Time Data)**

**API Used:**

**Real-Time Amazon Data API** via RapidAPI website Link - https://rapidapi.com/letscrape-6bRBa3QguO5/api/real-time-amazon-data

**Api Endpoints Used:**

| **Endpoint** | **Purpose** |
| --- | --- |
| /products-by-category | Fetch product listings by category ID |
| /product-reviews | (Optional) To fetch reviews per ASIN |
| /product-details | (Optional) To fetch detailed metadata |

**Authentication:**

* x-rapidapi-key: API key

**Rate Limits:**

* Requests **100 requests/month**  on free tier.
* Set **delay (time.sleep(1))** and **limit to 5 pages per category** to control usage.

2. Data Pipeline — Python API Code

* Python Libraries Used:

requests # To call the API

pandas # For data processing and cleaning

sqlalchemy # For writing data into MySQL

schedule # To automate data fetch hourly

re, time, random # Utility libraries

**Script Functionality:**

* Fetches product data from multiple categories (5 pages/category)
* Cleans and transforms:
  + Price → float
  + Sales volume (e.g., 20K+) → numeric (e.g., 20000)
  + Boolean fields (is\_prime)
* Calculates:
  + quantity based on price
  + total\_sales = price × quantity
* Appends location from a predefined list (used for map analysis)
* Saves cleaned data to **MySQL**
* Scheduled to run every 6 hours using schedule module

Script File:

Fetch\_Clean\_Mysql\_schedule.py

**3. Database — MySQL**

**Table: amazon\_products**

| **Column** | **Type** | **Description** |
| --- | --- | --- |
| product\_id | TEXT | Unique Amazon ASIN |
| title | TEXT | Product title |
| category | TEXT | Category name |
| price | FLOAT | Product price |
| quantity | INT | Simulated purchase quantity |
| total\_sales | FLOAT | price × quantity |
| currency | TEXT | Usually "USD" |
| rating | FLOAT | Star rating (1–5) |
| total\_reviews | INT | Number of reviews |
| sales\_volume | INT | Cleaned volume (e.g., 20000) |
| is\_prime | BOOLEAN | True/False |
| delivery | TEXT | Delivery info |
| location | TEXT | Simulated US cities for geo mapping |
| url | TEXT | Product URL |

**4. Visualization — Tableau Dashboard**

**Dashboard File:** Amazon\_sales\_dashboard\_tableau.twb

**Insights Provided:**

| **Sheet** | **Visualization Type** | **Description** |
| --- | --- | --- |
| Sales by Category | Bar Chart | Total sales by product category |
| Sales by City (Map) | Filled Map | City-level distribution of sales |
| Top 10 Products Sold | Horizontal Bar | Highest total sales products |
| Top 10 Rated Products | Horizontal Bar | Products with highest avg rating |
| KPI Boxes | Text | Total Products, Sales, Avg Price, Avg Rating |
| Category/City Filters | Filter Panel | Interactive filtering |

A screenshot of a computer

AI-generated content may be incorrect.

**5.Future Improvements (Optional)**

* Add sentiment analysis from product reviews
* Add product image thumbnails in dashboard
* Enable live Tableau connection to MySQL for real-time refresh
* Use Airflow for more scalable scheduling
* Add cloud storage/backup (e.g., GCP bucket or AWS S3)