**Project Description:**

This project focuses on analyzing the **Amazon Best Seller Software** dataset (available in CSV format on Kaggle) using Python and MySQL. The goal is to clean and transform the data, store it in a relational database, and extract valuable insights using analytical techniques.

**Key Steps:**

1. **Dataset Acquisition:**
   * Download the **Amazon Best Seller Software CSV dataset** from Kaggle.
   * Save it locally to work with using Python.
2. **Python File Handling:**
   * Use Python’s file handling methods to read the CSV file.
   * Handle file reading errors and edge cases to ensure reliability.
3. **Data Cleaning & Sanitization:**
   * Load the dataset into a **Pandas DataFrame**.
   * Perform cleaning operations such as:
     + Removing or handling missing and null values
     + Stripping unwanted characters (e.g., currency symbols, special characters)
     + Converting data types (e.g., prices to float, dates to datetime format)
     + Removing duplicates and irrelevant rows
     + Standardizing column names
4. **MySQL Integration:**
   * Design a suitable schema in **MySQL** to store the cleaned data efficiently.
   * Connect to the MySQL database using **Python (MySQL Connector or SQLAlchemy)**.
   * Insert the cleaned data into the appropriate MySQL tables.
5. **Analytical Computation:**
   * Use Python libraries to perform key analytical tasks such as:
     + Identifying top-rated software products
     + Analyzing pricing trends
     + Finding the most reviewed software categories
     + Calculating average ratings and review counts per category or brand
     + Comparing free vs paid software performance
6. **Console Output:**
   * Use pandas, numpy, and other relevant libraries to display insights in the console.
   * Present data summaries such as:
     + Top 10 best-selling software by rating
     + Average price of software per category
     + Rating vs. review count correlation
     + Outliers and anomalies in pricing or ratings