# **📘 Story Dataset Analysis Report**

## **📝 Project Overview**

This report provides an analysis of the **My Stories** **dataset**. The dataset includes information about various stories, such as their genre, country, format, status, age group, likes, views, price, pages, awards, publisher, language, and author. The aim of this project was to gain insights into the patterns of stories, readership, and other metrics using Python (Pandas, NumPy, Matplotlib).

## **📊 Dataset Summary**

* **Shape:** 1000000 rows × 12 columns
* **Main columns:** Title, Author, Genre, Country, Format, Status, AgeGroup, Language, Pages, Price, Likes, Views, Awards
* **Unique values:**
  + Genres: 6
  + Countries: 5
  + Formats: 4
  + Age Groups: 3
* **Missing values:** None

## **📊 Key Findings**

### **1️⃣ Stories per Genre**

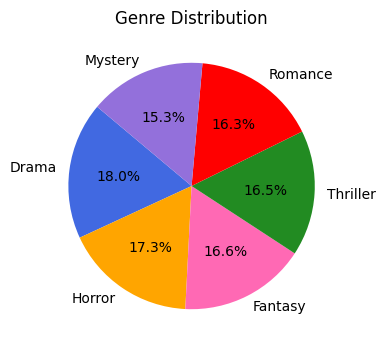
### 

**✅ Observation:-** Horror is the most common genre in the dataset.

| **2️⃣ Top 10 Highest Priced Stories**  **3️⃣ Country + Genre Combination (Highest Total Price)**   **✅ Observation:-** UK Romance stories contributed the highest revenue. |
| --- |

## **🔍 Exploratory Data Analysis**

### **1️⃣ Genre Distribution**



**📊 Insight**:- This chart provides a clear visual breakdown of the story distribution across genres. It illustrates how categories like Drama, Horror, Mystery, Romance, Thriller, and Fantasy contribute to the dataset, helping identify which genres dominate and where diversity exists.

✅ **Observation:-** The dataset is balanced across genres, with Drama and Horror slightly leading.

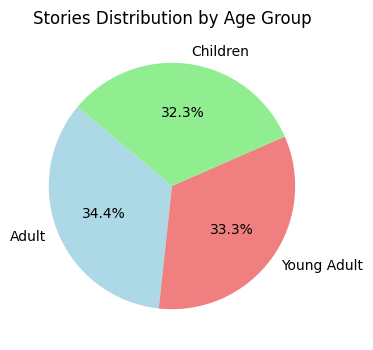
### **2️⃣ Total Likes and Views by Genre**

### 

**📊 Insight**:- This chart compares total likes and views across genres to highlight engagement patterns. It showcases which genres resonate most with readers and helps reveal where audience interest is strongest.

✅ **Observation:-** Drama gathered the most likes and views, showing its popularity among readers.

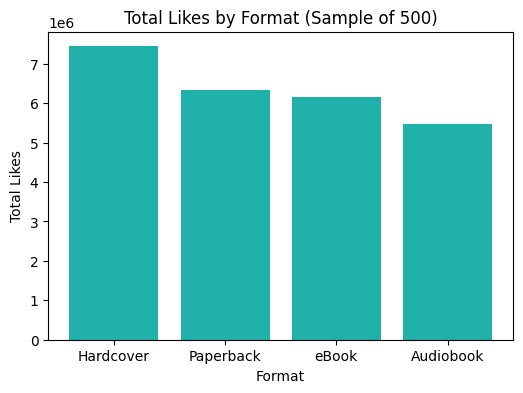
### **3️⃣ Stories Distribution by Age Group**



**📊 Insight:-** These charts provide a detailed view of story distribution by age group. They show how content is tailored to Adults, Children, and Young Adults, offering insights into the dataset’s target audience composition.

✅ **Observation:-** Adult stories are slightly more common, followed by Young Adult and Children.

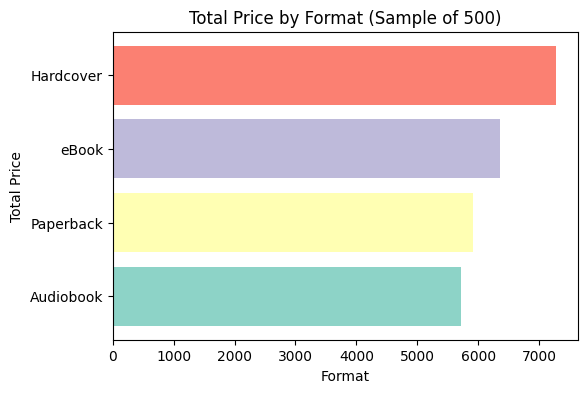
### **4️⃣ Likes by Format**



**📊 Insight:-** This chart presents total likes by format and reveals how reader engagement varies across Hardcover, Paperback, eBook, and Audiobook formats, highlighting preferred reading mediums.

✅ **Observation:-**Hardcover and Paperback formats received the highest total likes in the sample.

### **5️⃣ Price by Format**



**📊 Insight:-** This chart illustrates total price contribution by format, showcasing the revenue share of Hardcover, Paperback, eBook, and Audiobook, and helping identify the most commercially impactful formats.

**✅ Observation:-** Hardcover format contributed the highest total price in the sample.

## **✅ Conclusion**

* The dataset is clean with no missing data.
* Horror and Drama are the most common genres.
* Hardcover and eBook formats are popular for both views and likes.
* UK Romance stories generate the highest revenue.

## **💡 Suggestions**

* Explore trends over time if publication years are available.
* Predict popularity (likes/views) based on story features.
* Add sentiment or text analysis on descriptions or reviews.