**PROJECT REPORT**

Submitted by:

Sooraj S

CHN22BT040

Techmaghi Data Science and ML internship

Date: 1-07-2025

1.**Project Overview**

This project involved analyzing a simulated Amazon inventory dataset to identify issues related to stock levels, sales performance, and product demand. Three datasets were provided, but only one was usable for analysis due to data integrity issues in the others. All insights and visualizations presented here are based on that single dataset.

2. **Dataset Summary**

The provided Amazon dataset contained columns was obtained using ‘df.columns’ command, result is as shown:

['Name', 'Author', 'User Rating', 'Reviews', 'Price', 'Year', 'Genre']

3. **Dataset Understanding and Primary Analysis**

Loaded the valid dataset into a Jupyter notebook using pandas and performed basic analysis using: df.info(), df.describe(), df.isnull().sum() and df.head().

Identified key columns relevant to inventory and sales.

4.**Data Cleaning and Preprocessing**

There were no duplicated rows nor any null rows checked and verified using : df.duplicated().sum() and df.isnull().sum().

Standard column names and data types using:

df.columns=df.columns.str.lower().str.replace(' ','\_')

5. **Exploratory Data Analysis(EDA)**

Datasets were grouped by ‘authors’ and and sorted them based on ‘user\_rating’ in descending order. Most rated and reviewed book in this dataset were found using:

df\_rat=df[(df['user\_rating']>4.0)&(df['reviews']>12000)]

df\_rat.sort\_values(by='user\_rating',ascending=False,inplace=True)

6. **visualization**

Imported plotly, matplotlib.pyplot and seaborn into notebook.

Plotted book genre distribution in pie chart:

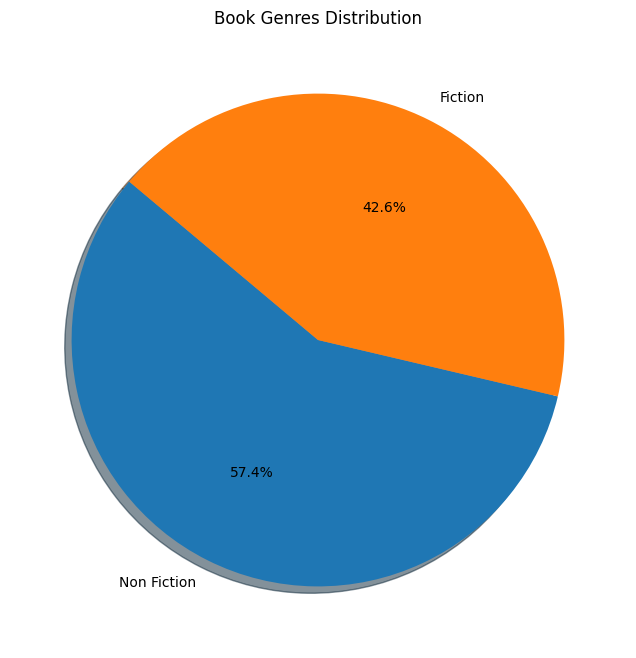


Fig 1

Plotted line graph with ‘price’ and ‘reviews’ in x and y axis respectively:

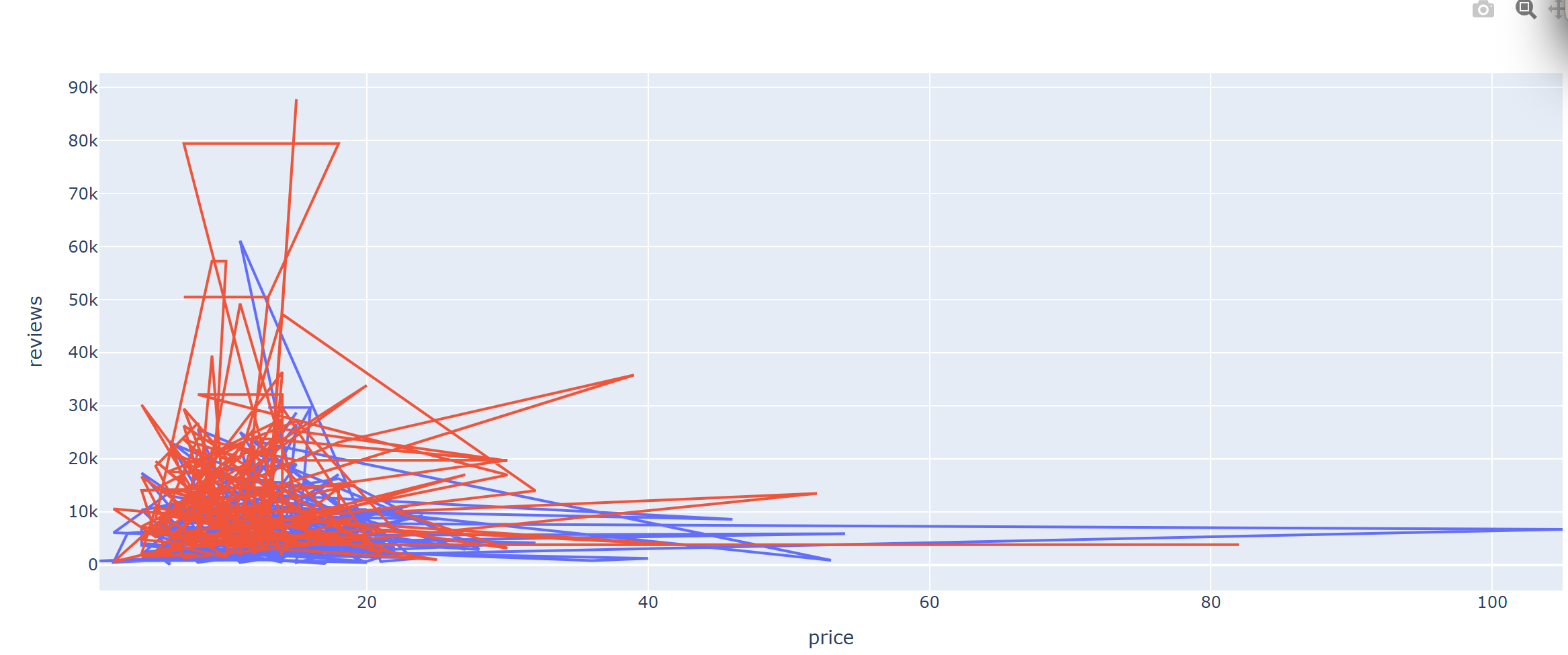


Fig 2

7. **Key Findings**

* From fig 2, we can observe that most responses from customers came to books which had price under 20.
* Books with low price could benefit from a slight price hike to increase revenue as they already have good response and rating.
* Introducing more books to low price range could potentially yield more response from customers within a low time frame.

**Appendix**

Dataset used: https://www.kaggle.com/datasets/sootersaalu/amazon-top-50-bestselling-books-

2009-2019

Code file: https://drive.google.com/drive/folders/10gt-IaCZwtMRCmgflUe4IgPAiIZ6PNTo?usp=sharing