

## Data Collection and Preprocessing Phase

Date	June 2025
Team ID	Team-739774
Project Title	Amazon kindle store review analysis
Maximum Marks	2 Marks

### Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

### Data Collection Plan Template

Section	Description
Project Overview	The project aims to analyze and predict product ratings and trends on the Amazon Kindle Store using machine learning techniques. Objectives include identifying top-rated books, understanding key factors influencing customer reviews, and developing models for rating predictions.
Data Collection Plan	Data will be collected from multiple sources such as customer reviews, product metadata, Kindle bestseller lists, author profiles, sentiment analysis from social media, expert literary reviews, and sales rankings.
Raw Data Sources Identified	<p><b>Amazon Customer Reviews:</b> User-generated reviews and ratings for Kindle books.</p> <p><b>Product Metadata:</b> Book titles, authors, genres, pricing, page count, and publication dates from Amazon.</p> <p><b>Bestseller Lists:</b> Data on top-performing Kindle books updated regularly by Amazon.</p> <p><b>Author Profiles:</b> Information on authors' popularity, number of published books, and engagement</p> <p><b>Social Media Mentions:</b> Public sentiment and discussions related to Kindle books on platforms like Twitter and Reddit</p>

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Access Permissions
Dataset 1	In this project we have used .csv data. This data is downloaded from kaggle.com.	<a href="https://www.kaggle.com/datasets/bharadwaj6/kindle-reviews">https://www.kaggle.com/datasets/bharadwaj6/kindle-reviews</a>	CSV	Public
Dataset 2	Access Google Drive with a Google account (for personal use) or Google Workspace account	<a href="https://www.kaggle.com/datasets/bharadwaj6/kindle-reviews">https://www.kaggle.com/datasets/bharadwaj6/kindle-reviews</a>	Excel	Private (with access)