Data Collection and Preprocessing Phase

Date	02 Oct 2025
Team ID	LTVIP2025TMIDS63456
Project Title	Analysis of Medium App Reviews from Google Play Store
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	Machine-learning project to analyse sentiment in Medium-app user reviews from Google Play Store, surface feature-improvement opportunities, track sentiment shifts after updates, and compare against competitor apps.
Data Collection Plan	*Primary*: Download the **"Medium App Reviews" Kaggle dataset (CSV)**, which aggregates Google Play review text, rating scores, and timestamps.*Supplementary*: Pull Medium-app **metadata & changelog** directly from Google Play pages (for version context).Optionally collect **competitor-app review datasets** (e.g., Wattpad, Pocket) from Kaggle or directly via the Play Store for benchmarking.
Raw Data Sources Identified	1. Medium App Reviews Dataset (CSV) 2. Medium-app metadata & changelog (HTML/JSON) 3. Competitor-app review datasets (CSV/JSON)

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Medium App Reviews	Consolidated Google Play reviews for the Medium app, including review ID, user name, rating, review text, date, and helpful-vote counts.	`https://github .com/Ashwin-S-Narayanan-2005/analysis-of-medium-app-reviews-from-google-play-store/blob/main/5.%20Project%20Executable%20Files/dataset.csv`	CSV	~50 MB	Public
Medium App Metadata	Metadata such as app version, installs, permissions, and last updated date, scraped from Google Play.	`https://github .com/Ashwin-S-Narayanan-2005/analysis-of-medium-app-reviews-from-google-play-store/blob/main/5.%20Project%20Executable%20Files/dataset.csv`	JSON / CSV	~5 MB	Public
Medium App Background Image	A background JPG related to the Medium app, used for presentation or optional image-based analysis.	`https://github .com/Ashwin-S-Narayanan-2005/analysis-of-medium-app-reviews-from-google-play-store/blob/main/5.%20Project%20Executable%20Files/static/images/background.jpg`	JPG	~500 KB	Public