Data Collection and Preprocessing Phase

Date	22 November 2024
Team ID	SWTID1727420425
Project Title	Analysis amazon cell phone review with nlp technique
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 utilizes two datasets for analyzing Amazon cell phone reviews: one containing product metadata (e.g., name, brand, price) and another with customer reviews (e.g., ratings, review text, review date). The review dataset provides insights into customer feedback, while the item dataset complements it with product details. Together, these datasets enable the application of NLP techniques to understand customer sentiments, trends, and preferences, forming the foundation for comprehensive analysis
Data Collection Plan	The data collection plan for this project involves gathering two key datasets: product metadata and customer reviews from Amazon. Product metadata includes details such as product names, brands, and specifications, while the review dataset contains review text, ratings, and dates. The data is sourced through web scraping or public datasets, ensuring it is comprehensive and representative of user feedback. Proper filtering is applied to include only relevant

	cell phone reviews, forming the basis for meaningful analysis using NLP techniques.
Raw Data Sources Identified	The raw data sources for this project include publicly available datasets and web-scraped information from Amazon. These sources provide detailed customer reviews, ratings, review dates, and metadata such as product names, brands, and specifications. The identified datasets are rich in content, enabling comprehensive analysis of customer sentiments and trends using NLP techniques while ensuring relevance to cell phone products.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	I have two dataset reviews and items.	https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews	CSV	Review=24800KB Items=226KB	Public