

Data Collection and Preprocessing Phase

Date	25 march 2025
Team ID	LTVIP2025TMID28445
Project Title	Cosmetic insights
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 cosmetic products by evaluating their ingredients, sustainability, and user feedback. The objective is to provide consumers with insightful and transparent information to make informed purchase decisions.
Data Collection Plan	Data will be collected from multiple sources, including online cosmetic databases, brand websites, e-commerce platforms, and user-generated reviews. Web scraping, APIs, and publicly available datasets will be used.
Raw Data Sources Identified	<p>Kaggle Cosmetic Datasets: Contains product details, ingredients, and pricing information.</p> <ul style="list-style-type: none"> - Brand Websites: Official product information, ingredient transparency reports. - E-commerce Platforms (Amazon, Sephora, etc.): User reviews, ratings,

	<p>and pricing trends.</p> <ul style="list-style-type: none"> - Regulatory Websites (FDA, EWG, etc.): Ingredient safety and compliance data. - Social Media & Beauty Forums: Consumer sentiment and emerging trends.
--	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Description of the data in this source.	Link of Dataset 1	CSV	XX GB	Public
Dataset 2	Description of the data in this source.	Link of Dataset 2	Excel	YY GB	Private (with access)
...