

Data Collection and Preprocessing Phase

Date	16 March 2025
Team ID	LTVIP2025TMID25168
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	<p>Project Overview:</p> <p>The Cosmetic Insights project aims to leverage data analytics to understand consumer behavior, market trends, and product preferences within the cosmetics industry. By analyzing raw data from various sources, we seek to extract meaningful insights that will help businesses optimize their strategies and improve customer satisfaction.</p>
Data Collection Plan	<p>Data Collection Plan:</p> <p>To ensure comprehensive and high-quality data, we will collect information from the following sources:</p>

	<ol style="list-style-type: none"> 1. E-commerce Platforms – Product reviews, ratings, and customer feedback from leading online cosmetic retailers (e.g., Sephora, Ulta, Amazon). 2. Social Media Channels – User-generated content, discussions, and trends from platforms like Instagram, Twitter, and TikTok. 3. Surveys & Questionnaires – Direct consumer feedback gathered through structured online and offline surveys. 4. Industry Reports & Market Research – Insights from published cosmetic industry reports, whitepapers, and research publications. 5. Competitor Analysis – Data from competitors' websites, advertisements, and promotional campaigns. 6. Point-of-Sale Data – Sales transaction records from retail stores and e-commerce platforms. 7. Influencer Endorsements – Data from beauty influencers and bloggers who impact purchasing decisions. <p>:</p>
Raw Data Sources Identified	List the raw data sources with relevant details (as a short description).

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions

Dataset 1	Product reviews and ratings from Sephora, Ulta, and Amazon.	[Dataset 1 URL]	CSV	XX GB	Public
Dataset 2	Social media discussions and trends from Instagram, Twitter, TikTok.	[Dataset 2 URL]	JSON	YY GB	Private (with access)
Dataset 3	Consumer survey responses on product preferences.	[Dataset 3 URL]	Excel	ZZ GB	Private (with access)
Dataset 4	Market research reports on the cosmetics industry.	[Dataset 4 URL]	PDF	AA GB	Subscription-based
Dataset 5	Sales transaction records from retail and online stores.	[Dataset 5 URL]	SQL Database	BB GB	Confidential
Dataset 6	Influencer endorsement data and engagement metrics.	[Dataset 6 URL]	CSV	CC GB	Private (with access)