Data Collection and Preprocessing Phase

Date	8 July 2024
Team ID	739978
Project Title	Rhythmic Revenue: Unveiling The Future Of Music Sales With Machine Learning
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	This project aims to leverage machine learning techniques to predict future music sales. By analyzing historical data on music formats, sales metrics, and the number of records, we intend to build a robust predictive model that can offer insights into future trends in the music industry.
Data Collection Plan	Data will be collected from various sources, including historical music sales databases, industry reports, and publicly available datasets on music sales and trends. We will also use online APIs to gather real-time data.
Raw Data Sources Identified	we have identified three key data sources: a comprehensive Music Sales Database containing historical sales data by format, year, and number of records; detailed Industry Reports offering insights into annual sales trends; and an Online Music API providing real-time data on current music sales. These diverse and rich data sources will be integral in building an accurate and robust predictive model.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises details as(Format,Metric, No of Records,Year,Value(Actual))	https://www.kaggle.com/datasets/thedevastator/music-sales-by-format-and-year/data	CSV	122 kb	Public