Data Collection and Preprocessing Phase

Date	20 June 2024
Team ID	739968
Project Title	Movie Box Office Gross Prediction
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	"Provide a brief overview of the machine learning project and its objectives. For instance: "This project aims to predict the box office gross of movies using machine learning algorithms. The objectives are to identify key factors influencing a movie's financial success and develop a predictive model that can estimate the gross revenue based on these factors.""
Data Collection Plan	"Mention from which sources the data are going to be collected. For example:

	"The data will be collected from various sources, including movie databases such as IMDb and Box Office Mojo, social media platforms like Twitter for sentiment analysis, and financial records from industry reports"
Raw Data Sources Identified	"List the raw data sources with relevant details (as a short description). For instance: IMDb: Provides detailed information about movies, including release dates, genres, cast, and crew. Box Office Mojo: Offers comprehensive box office revenue data. Twitter: Allows for sentiment analysis based on public opinion and trends. Industry Reports: Contains financial data and market analysis reports"

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Movie Box Office Gross Prediction	https://www.kaggle.com/tmdb/tmdb-movie-metadata	CSV	9 MB	Public