Data Collection and Preprocessing Phase

Date	7 December 2024
Team ID	xxxxxx
Project Title	Uncovering the Gaming Industry's Hidden Gems: A Comprehensive Analysis of Video Game Sales
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

Section	Description
Data Overview	The dataset consists of over 16,500 video games with sales data above 100,000 units. It includes columns such as Rank, Name, Platform, Year, Genre, Publisher, and Sales by region (e.g., North America, Europe).
Data Cleaning	Handle missing values by checking for NaN entries and filling them with appropriate values or dropping rows/columns. Remove duplicates and correct inconsistencies in data (e.g., typos in game names).
Data Transformation	Apply data filtering to focus on specific platforms, genres, or time periods. Sort data by sales rank or year and create pivot tables to summarize sales by region or genre. Create new calculated columns (e.g., total global sales).
Data Type Conversion	Ensure columns are in the correct data type (e.g., convert sales columns to numerical types, year columns to integers). Check and update data types as needed to facilitate accurate analysis.
Column Splitting and Merging	Split columns such as 'Publisher' into separate columns for better categorization or merge related columns (e.g., merging region sales columns into a 'Total Sales' column).
Data Modeling	Define relationships between columns and tables if necessary (e.g., platform to game, publisher to genre). Create calculated measures for key metrics like average sales per game or total

	sales by platform.
Save Processed Data	Save the cleaned and transformed dataset to a file format (e.g., CSV, Excel) for future use in Tableau or other analysis tools. Ensure it's stored in a well-organized folder structure for easy retrieval.