

## 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	3 Marks

### Data Quality Report Template

Data Source	Data Quality Issue	Severity	Resolution Plan
Video Game Sales Data	Missing values in columns (e.g., Publisher, Genre)	High	Identify missing data using <code>pandas.isnull()</code> and fill them with placeholder values or drop rows/columns as needed.
Video Game Sales Data	Duplicated records	Moderate	Use <code>pandas.drop_duplicates()</code> to remove duplicated entries and ensure data uniqueness.
Video Game Sales Data	Inconsistent data formats (e.g., different date formats)	High	Standardize date columns using <code>pandas.to_datetime()</code> and ensure

			uniform formatting for all records.
Video Game Sales Data	Outliers in sales data	High	Use statistical methods (e.g., IQR or Z-score) to detect and handle outliers by capping or removing them.
Video Game Sales Data	Inaccurate or misspelled game names and genres	Moderate	Apply text normalization (e.g., <code>str.lower()</code> ) and use string matching algorithms or dictionaries for correction.
Video Game Sales Data	Unstructured or mixed data types in columns	Low	Use <code>pandas.astype()</code> to convert columns to the appropriate data type, ensuring consistent format for analysis.
Video Game Sales Data	Missing or incorrect region-specific sales data	High	Impute missing values with the mean or median of the specific region or use data from reliable external sources.