Data Collection and Preprocessing Phase

Date	16 JUNE 2025
Team ID	SWTID1749825524
Project Title	Deepfruitveg: Automated Fruit and Veg Identification
Maximum Marks	2 Marks

Data Quality Report:

This report outlines the quality issues identified in the dataset used for the project 'DeepFruitVeg: Automated Fruit and Vegetable Identification', along with their severity and planned resolutions.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Google Drive / Kaggle (Image Dataset)	Some images are mislabeled or placed in incorrect folders (class mismatch).	High	Manually verify folder names and restructure dataset to follow strict class labeling.
Google Drive / Kaggle (Image Dataset)	Uneven class distribution (some fruits/vegetables have significantly more images).	Moderate	Apply data augmentation (flipping, rotation) to balance classes.
Google Drive / Kaggle (Image Dataset)	Image resolutions are inconsistent.	Low	Resize all images to a uniform shape (e.g., 64x64) during preprocessing.
Google Drive / Kaggle (Image Dataset)	Presence of corrupted or unreadable image files.	Low	Use exception handling during loading to skip and log corrupted files.
Google Drive / Kaggle (Image Dataset)	Dataset lacks metadata like location, freshness, or ripeness.	Low	Not applicable for this stage; may enhance with extended datasets in future versions.