**Data Analytics Project Journal**

| **Date:** 03/07/2025 | **Topic: Process Data from Dirty to Clean** | | |
| --- | --- | --- | --- |
| 1. Data Profiling & Initial Checks**:** | Scan for missing values, duplicates, and inconsistent formats  Use filters and conditional formatting to highlight anomalies  Identify risks to data integrity (e.g., outliers, incorrect data types) | | |
| 2. Standardizing Formats**:** | Convert date formats to a consistent structure (e.g., MM/DD/YYYY)  Normalize text casing (e.g., UPPER, LOWER, PROPER)  Clean up inconsistent category labels (e.g., “bike”, “Bike”, “BIKE” → “Bike”) | | |
| 3. Handling Missing & Incorrect Data | Use IFERROR, ISBLANK, and IF formulas to flag and fix missing entries  Apply logic to fill gaps (e.g., default values, inferred entries)  Remove or correct invalid entries using filters and manual review | | |
| 4. Removing Duplicates | Use Remove Duplicates tool or formulas like COUNTIF to identify repeated rows  Decide whether to keep, merge, or delete based on context | | |
| 5. Creating Helper Columns | Add calculated fields (e.g., Duration, Category Flag, Region)  Use formulas like LEFT, RIGHT, MID, LEN, and TEXT to extract or reshape data | | |
| 6. Verifying Cleaning Results**:** | Compare row counts before and after cleaning  Document changes made and rationale  Prepare a summary report for stakeholders | | |