

Day 5 Lesson Plan: Spy Mission – Data Cleaning

Objective

Students will identify and correct data issues such as missing values, duplicate entries, and inconsistent formatting in order to prepare a dataset for analysis.

Word Bytes

- **Data Cleaning** – The process of correcting or removing inaccurate, inconsistent, or incomplete data.
- **Duplicate** – An identical entry that appears more than once in a dataset.
- **Inconsistent Format** – A mismatch in how information is presented, such as “January 2” vs. “1/2.”

Materials

- “Spy Mission” Intro Slide or video (can be a teacher-led story)
- Corrupted Dataset Handout (printed or digital)
- Colored pens/highlighters or editing tools
- “Mission Report” reflection sheet (optional)

Activities

Hook: Spy Mission Briefing

- Announce the day’s challenge like a spy-themed operation.
 - “Agents, our organization’s data has been compromised. Your task is to clean the files before time runs out. Accuracy is critical.”

Mini-Lesson: What Makes Data Messy?

- Review common issues in real-world datasets:
 - Missing values (empty fields in a survey)
 - Duplicate entries (same person recorded twice)
 - Inconsistent formatting (e.g., different date formats, varied capitalization)

- Show examples and ask:
 - Why might messy data cause problems?
 - Who needs clean data to do their job well?

Mission Activity: Clean the Corrupted Dataset

- Students work individually or in teams to spot and fix errors.
- Encourage highlighting errors before correcting them.
- Optional: Add a timer or small “data detective” rewards.

Wrap-Up Discussion: Why Clean Data Matters

- Use these prompts to guide conversation:
 - What was the most frustrating error to fix?
 - How could a mistake in a dataset affect real-life decisions (e.g., medical records, financial systems)?
 - What strategies helped you stay organized while cleaning?

Optional Extension: Challenge Round

Provide a second mini-dataset with sneakier errors (e.g., swapped columns, false duplicates) for early finishers.

Word Bytes Dictionary Update

Students add today’s terms with examples:

- Data Cleaning – Fixing typos and removing repeated rows in a spreadsheet.
- Duplicate – Seeing the same student name and grade listed twice.
- Inconsistent Format – Some birthdays are “12-05,” others say “December 5.”