

## PROMPT1:

```
prompt = f"""
    You are a data cleaning assistant.

    Your task is to clean and standardize the following record:
    {record.to_dict()}

    Follow these instructions strictly:

    1. Missing Fields:
        - If any required fields (e.g., city, borough, zip code, date,
        incident_address) are missing, attempt to infer them using available
        fields such as `address`, `agency_name`, `location`, or `borough`.
        - If a value cannot be inferred:
            - For date fields, replace missing values (e.g., None, NaN, Null)
            with the default format "YYYY-MM-DD" or "YYYY-MM-DDTHH:MM:SS".
            - For text or categorical fields, replace missing values with the
            string "Missing Values".

    2. Formatting:
        - Correct all spelling errors.
        - Use title case for proper nouns such as names and boroughs.
        - Use sentence case for description fields.
        - Standardize the `incident_address` field to the format: "Street,
        Borough, NY".
        - Convert all date/time fields to ISO 8601 format: "YYYY-MM-DD" or
        "YYYY-MM-DDTHH:MM:SS".

    3. Categorization:
        - Retain the original complaint type in a new field named
        `original_complaint_type`.
        - Classify the `complaint_type` into a broad category (e.g., "Noise",
        "Sanitation", "Public Safety", "Parking", etc.) based on the description.
        - Add a new field `noise_flag` with value `"Noisy"` or `"Non-Noise"`
        depending on whether the complaint is noise-related.

    Return only the cleaned record as a dictionary, without any
    explanations or commentary.

    """
```

Prompt2:

```
prompt = f"""
    You are a Python testing expert.

    Given the cleaned NYC 311 service request record below, generate Python
    unit tests using assert statements. These tests must validate the data
    quality based on the following rules:

    1. Required fields such as 'created_date', 'complaint_type', 'borough',
    and 'incident_address' must not be null.
    2. 'created_date' must be in ISO 8601 format and a valid datetime.
    3. 'Noise_flag' must be one of: "Noise", "Non-Noise".
    4. 'borough' must be one of: "Manhattan", "Brooklyn", "Queens",
    "Bronx", "Staten Island".
    5. 'incident_address' must follow the format: "Street, Borough, NY".
    6. 'city' must not be null.
    7. If 'latitude' and 'longitude' are present, they must be valid floats
    within NYC bounds.

    Return only the Python code and no explanations
    strat with function method no imports no backticks

    Record:
    {record}
    """
```