DATA ANALYST PRACTICAL EXERCISE (NW REALITE LTD)

Context

You've joined **NW Realite Ltd**, managing properties across **Nairobi CBD, Westlands, Kilimani**. You'll analyze **rent performance, occupancy, and arrears** using the following tables: locations, properties, units, tenants, leases.
Note: The dataset contains **some dirty records**.

TASK 1: Data Cleaning & Preparation

Goal: Clean and structure data for analysis.

Do:

1. Import the data (Python/Pandas or Excel Power Query).
2. Identify and document issues:
    - Missing/NULLs (e.g., tenant emails, lease end_date).
    - Invalid values/dates (e.g. Negative Rent per Month)
    - Duplicates and outliers.
3. Create the following columns:
    - lease_duration_months (treat NULL end_date as ongoing up to today).
    - lease_status (ongoing, expired, invalid).
    - valid_lease (1 if start_date < end_date OR end_date is NULL and start_date <= today; else 0).
    - annual_rent = rent_per_month * 12.

    Note: Rent cannot be negative, change negative to positive if it exists.
4. Standardize location names to Proper Case.

Deliverables:

- Cleaned CSV/Excel.
- Short note (bullets) listing issues found + cleaning decisions.

TASK 2: Exploratory Analysis & Visualization

Goal: Communicate insights visually (Power BI preferred, Equivalent visualizer acceptable).

Build visuals for:

1. **Total Rent Billed vs Total Arrears by Location.**
    - Define "Rent Billed" as sum of rent_per_month for **valid leases only**.
2. **Occupancy Rate per Property.**
    - Occupancy = (# units with **at least one valid lease**) ÷ (total units) × 100.
3. **Top 3 Properties by Arrears.**
    - Use sum of arrears (include credits).
4. **Average Monthly Rent per Property** and per Location.
5. (Optional) **Valid vs Invalid Lease Count** trend (by start month).

Deliverables:

- 1-page dashboard .pbix.
- 3–5 sentence summary of key insights.

TASK 3: SQL (write queries)

Goal: Test SQL reasoning on our schema.

Assumptions for all questions:

- Treat **only valid_lease = 1** as occupied for occupancy calculations.
- A unit is **occupied** if it has **at least one valid lease**.
- If needed, you may use CTEs and window functions.

Questions (write queries):

1. **Occupancy Underperformers**
   Return property_id, property_name, occupancy_rate for properties with **occupancy < 80%**.
   - Occupancy = occupied units ÷ total units × 100.
2. **Arrears by Location**
   Return location, total_arrears (sum of arrears) grouped by location. Include locations with zero arrears as 0.
3. **Collection Efficiency Leaderboard**
   For each property, compute collection_efficiency = (1 - (SUM(arrears) / NULLIF(SUM(rent_per_month),0))) * 100 using **valid leases only**.
   Return **Top 3** properties by collection_efficiency (handle division by zero).
4. **Data Quality Check – Invalid Leases**
   List leases where end_date < start_date or rent_per_month < 0. Include lease_id, property_name, unit_name, tenant_name, and reason flag ('NEGATIVE_RENT', 'END_BEFORE_START').
5. **Multi-Unit Tenants**
   List tenants who hold **2+ distinct units** (current or historical), showing tenant, count of units, and the properties they span (comma-separated property names).

TASK 4: Insight Communication

Goal: Translate analysis into action.

Write a 150–200 word email/memo to the Head of Property Management covering:

- Overall performance (rent, occupancy).
- Where arrears/inefficiencies concentrate.
- Recommended actions (e.g., lease audits, arrears follow-up cadence, incentives, pricing review, data fixes).

BONUS (Optional) – You may use existing plug and play models

Predictive Modeling (Python)
Build a simple linear regression predicting **next month's collection efficiency** per property using recent aggregates (e.g., rent billed, arrears, occupancy). Provide notebook and a one-paragraph interpretation.