**Smart Bricks – Senior Data Scientist Case Study:**

**Task:**

You are required to create a **data strategy and algorithmic approach** for generating **the following metrics**. The goal is to showcase how you would process raw data into actionable insights for investors.

This should be specific to the Dubai Real Estate market. The idea is that the investor should be able to compare these indexes across a unit, building, area and city level (e.g., 2 Bedroom apartment in Business Bay vs Dubai average)

1. **Sales Supply Index**
   * Tracks the availability of properties for sale over time.
   * ***Formula:*** *Sales Supply Index = (Number of Properties Listed for Sale in the Period/Total Number of Properties in the Market)*
   * **This index should be calculated on a unit, building, area and city level**
2. **Rental Supply Index**
   * Tracks the availability of rental properties over time.
   * ***Formula:*** *Rent Supply Index = (Number of Properties Listed for Rent in the Period/Total Number of Properties in the Market)*
   * **This index should be calculated on a** **unit, building, area and city level**
3. **Rental vs. Sales Supply Index**
   * Compares the trends between rental and sales markets.
   * ***Formula:*** *Rent vs Sales Supply Index = (Number of Properties Listed for Rent / Number of Properties Listed for Sale)*
   * **This index should be calculated on a unit, building, area and city level**
4. **Sales Asking Price (AED/SqFt)**
   * Monitors trends in property prices per square foot for sales listings.
   * ***Formula:*** *Sales Asking Price per Sqft = (Property sale asking price in AED /Property Size in Sqft)*
   * **This should be calculated on a unit, building, area and city level**
5. **Rental Asking Price (AED/SqFt)**
   * Monitors trends in rental prices per square foot
   * ***Formula:*** *Rental Asking Price per Sqft = (Rental asking price in AED /Property Size in Sqft)*
   * **This should be calculation on a building, area and city level**
6. **Sales Asking Price (AED)**
   * Monitors trends in property prices for sales listings.
   * **This should be calculated on a unit, building, area and city level**
7. **Rental Asking Price (AED)**
   * Monitors trends in rental prices
   * **This should be calculation on a unit, building, area and city level**
8. **Gross Yield (%) on Asking Price (AED/SqFt)**
   * Calculates rental yield for properties using price per square foot.
   * ***Formula:*** *Gross Yield On Asking Price (AED/Sqft)= (Annual Rental Asking Price (AED/Sqft) / Sales Asking Price (AED/Sqft)*
   * **This should be calculated on a unit, building, area and city level**
9. **Gross Yield (%) on Asking Price (AED)**
   * Calculates rental yield using absolute rental and sales prices.
   * ***Formula:*** *Gross Yield On Asking Price (AED)= (Annual Rental Asking Price (AED) / Sales Asking Price (AED)*
   * **This should be calculated on a unit, building, area and city level**
10. **Access to Off-Market and Distressed Properties**
    * Suggest how you would leverage data to identify off-market or distressed property opportunities.

**Deliverables:**

1. **Strategy Document (As part of the Notebook):**
   * Explain your approach to collecting, cleaning, and processing data.
   * Define the logic and methods for calculating your selected metrics.
   * Highlight any Dubai-specific adjustments (e.g., handling off-plan properties, community-based segmentation).
2. **Algorithm Implementation:**
   * Provide a working Python notebook (or script) showcasing how you would calculate all of the metrics.
   * Demonstrate how you would handle outliers, missing data, and anomalies in your calculation.
3. **Data Visualization** 
   * Use a visualization tool (e.g., Matplotlib, Plotly, Tableau) to represent trends in the metric(s). The output should be a dashboard showcasing the historical trends of these metrics

**For Each Metric**:

**Step 1: Data Sources**  
Outline the data you would need to calculate these metrics. This may include:

* Dubai Land Department data (transaction records, RERA rental index) – (<https://dubailand.gov.ae/en/open-data/real-estate-data/>)

**Step 2: Algorithm Design**  
Describe your algorithm for calculating these metrics:

* **Input:** Raw data sources (e.g., property listings, rental prices).
* **Processing:** Cleaning, filtering, normalization, and segmentation (e.g., by community, property type).
* **Output:** Final metric value (e.g., Sales Supply Index, AED/SqFt) with relevant visualizations.
* IMPORTANT: Please ensure the data visualization dashboard is dynamic and connected to the Dubai Land Department's API. The dashboard should automatically update as new data becomes available, ensuring it reflects the latest information at all times

**Step 3: Implementation**

* Write and submit a Python implementation for **all metrics**.
* Include comments to explain your code and methodology.

**Submission Guidelines:**

* Submit your **strategy document** (As part of the notebook), **Python notebook/script**, and any visualizations as a ZIP file.
* Include a README file explaining how to run your code and interpret the results.
* **Deadline:** 9am on 27th November 2024
* Please respond to the email with your submission

**Bonus Points:**

* Suggest any new indexes or metrics you recommend would be useful
* Suggesting innovative metrics tailored for Dubai’s off-plan, luxury, or distressed property markets.
* Incorporating machine learning models for trend prediction or anomaly detection.