Data Engineer Assignment

The purpose of this assignment is to evaluate your pipeline orchestration, data preprocessing and data mapping skills.

You are provided with three CSV files with tenant details. These tenants are contracted to operate and generate sales.

The CSV files details are as follows:

1. Lease_Details
   This file includes the details about the tenants:

Column Name	Detail
L_ID	Unique ID for Each Tenant
Lease_Type	The type of the tenant
Lease_Status	Current operational status of the tenant
Category	Category of the tenant

2. Lease_Trans
   The file includes the contract details of the tenant

Column Name	**Detail**
L_ID	Unique ID for Each Tenant
Area_PerSqFT	The contractual area that the tenant is currently occupying.
Monthly_Rent	Rent paid by the tenant monthly
Lease_From	Date when the lease started
Lease_To	Date when the lease expires.

3. Lease_Sales
   This file includes the sales generated by the tenant

Column Name	**Detail**
L_ID	Unique ID for Each Tenant
Year	The year when the sale is recorded
Month	The month when the sale was recorded
Source_Main	The sales figure recorded from our main system
Source_Sub	The sales figure from the sub system

The objective is as follows:

1. **Pipeline Orchestration** (60%):

    - Create a process to ingest the CSV files into an SQL Database. This can be done Via Python or ADF.
    - You will be judged based on structure of your approach and your scalability of the pipeline.

    Expected output:
    The method and approach is what we are looking for, however providing the DB Replica is a plus.

2. **Data Preprocessing and Mapping** (40%):

    Create a data map that links all three tables to provide the following:

    - A view that shows Yearly overview of each **current** tenant in terms of:
        - Tenant ID
        - Lease Type
        - Category
        - Area Per Sqft
        - Annual Rent (Monthly rent * 12)
        - Lease From
        - Lease To
        - Annual Sales (If there are no 12 months of sales, should be pro-rated)*
    - A View that shows Category performance that shows of **ALL** Tenants:
        - Rent per Category
        - Area SqFt per Category
        - Sales per Category

    Expected output:

    Code written to generate these views + the table as a view or DataFrame

    If the View can be visualized in a BI tool that is a huge plus

Pushing the code to a Github Repository is also a plus

The README.md file should contain some tips on how to approach this assignment objectives.

For any further information please contact:

Adib.mohammad@nakheel.com

+971 54 566 7155