Summary of the Task:

- **Objective**: Create a **Jupyter notebook** that accomplishes the following:
  - Utilize PySpark to read and transform the provided dataset.
  - Generate visualizations that offer meaningful insights; be creative and showcase any interesting findings.
  - Include markdown explanations detailing your transformation methods and the rationale behind your visualizations.
- **How to submit your workdone:** Please upload the your notebook to GitHub and share us your git hub link
- **Dataset Information**: The data can be downloaded from Kaggle:
  - This dataset pertains to credit card transactions and requires data cleaning due to its messy nature.
- **Key Considerations for Data Transformation**:
  - **Handling PII Data**: Clearly explain your chosen methods for managing personally identifiable information (PII).
  - **Data Quality Assurance**: Describe how you identify and process dirty data.
  - **JSON Flattening**: Convert JSON data into a tabular format. The expected columns include:
    - Unnamed: 0
    - trans_date_trans_time (Transaction Time)
    - cc_num (Credit Card Number)
    - merchant (Merchant Name)
    - category (Merchant Category)
    - amt (Transaction Amount)
    - first (Credit Card Owner's First Name)
    - last (Credit Card Owner's Last Name)
    - gender (Credit Card Owner's Gender)
    - street (Credit Card Owner's Street Address)
    - city (Credit Card Owner's City)
    - state (Credit Card Owner's State)
    - zip (Credit Card Owner's Zip Code)
    - lat (Credit Card Owner's Latitude)
    - long (Credit Card Owner's Longitude)
    - city_pop (City Population)
    - job (Credit Card Owner's Job)
    - dob (Credit Card Owner's Date of Birth)
    - trans_num (Transaction Number)
    - merch_lat (Merchant Latitude)
    - merch_long (Merchant Longitude)
    - is_fraud (Fraud Case Indicator)
    - merch_zipcode (Merchant Zipcode)
    - merch_last_update_time (Merchant Last Update Time)
    - merch_eff_time (Merchant Effective Registration Time)
    - cc_bic (Credit Card BIC Code)
  - **Timestamp Conversion**: All time-related columns (`trans_date_trans_time`, `merch_last_update_time`,

`merch_eff_time` columns) must be converted to a human-readable timestamp format in UTC +8 timezone (e.g., YYYY-MM-DD HH:MM .SSSSSS Z).

- **Name Derivation**: Extract first and last name columns from person_name:
  - The expected format is based on "first, last" but you may encounter dirty data. Please process the name based on following example:
    - person_name: "Edward, Sanchez" should result in First: "Edward" and Last: "Sanchez".
- Candidates are expected to identify and clean any dirty data independently.

- **Visualization and Analysis**:
  - Create any relevant charts or transform the data for analytical purposes.
  - You are encouraged to incorporate additional datasets to enhance your analysis.