This dataset comprises over 200 million anonymized UK key worker financial transactions, complete with geodemographic information. It's available in CSV, Parquet, or database table formats. ​[data.cdrc.ac.uk](https://data.cdrc.ac.uk/dataset/salad-money-open-banking-transaction-data?utm_source=chatgpt.com)

**Project Problem Statement: Advanced Financial Transaction Analysis Using Spark**

**Objective:** Analyze extensive financial transaction data to uncover spending patterns, assess financial behaviors, and identify potential areas for financial product development.​

**Problem Statement:** A financial services company aims to gain deeper insights into the spending behaviors of key workers to tailor their product offerings better. They provide a substantial dataset containing detailed transaction records, including:​

* **Transaction ID**
* **Customer ID**
* **Transaction Amount**
* **Transaction Date**
* **Merchant Category**
* **Geographical Information (e.g., Output Area codes)**
* **Demographic Information (e.g., age, sex)**

Your tasks are to:

1. **Load the dataset** into an RDD from the provided format (CSV or Parquet).
2. **Apply transformations** such as:
   * Filtering transactions above a certain amount (e.g., > £500).
   * Mapping data to key-value pairs (e.g., Customer ID → Total Spend).
   * Aggregating data to compute metrics like average spending per customer.
3. **Convert the RDD to a DataFrame** with a well-defined schema.
4. **Perform analyses**, including:
   * Identifying top spending customers.
   * Determining the most common merchant categories.
   * Analyzing spending patterns across different geographical areas.
   * Examining correlations between demographic factors and spending behaviors.
5. **Store the final processed data** in Parquet format for optimized storage and future analysis.

**Expected Deliverables:**

* A Spark application implementing the above workflow.
* Comprehensive insights derived from the data analysis.
* The processed dataset stored in Parquet format.