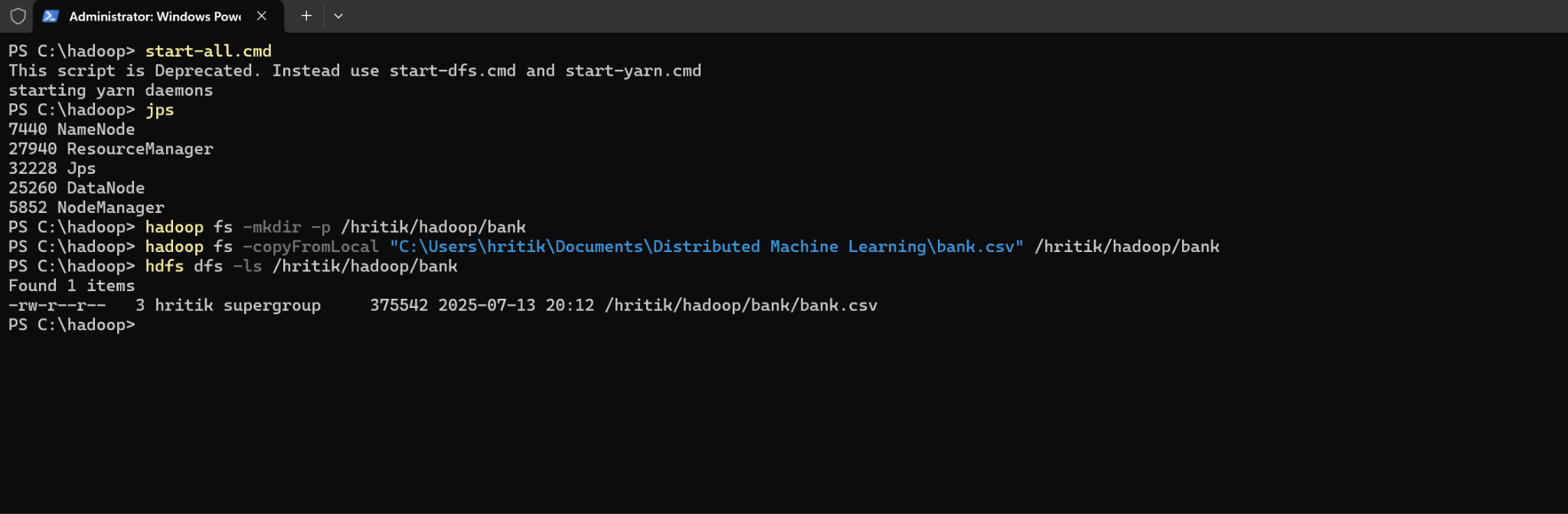
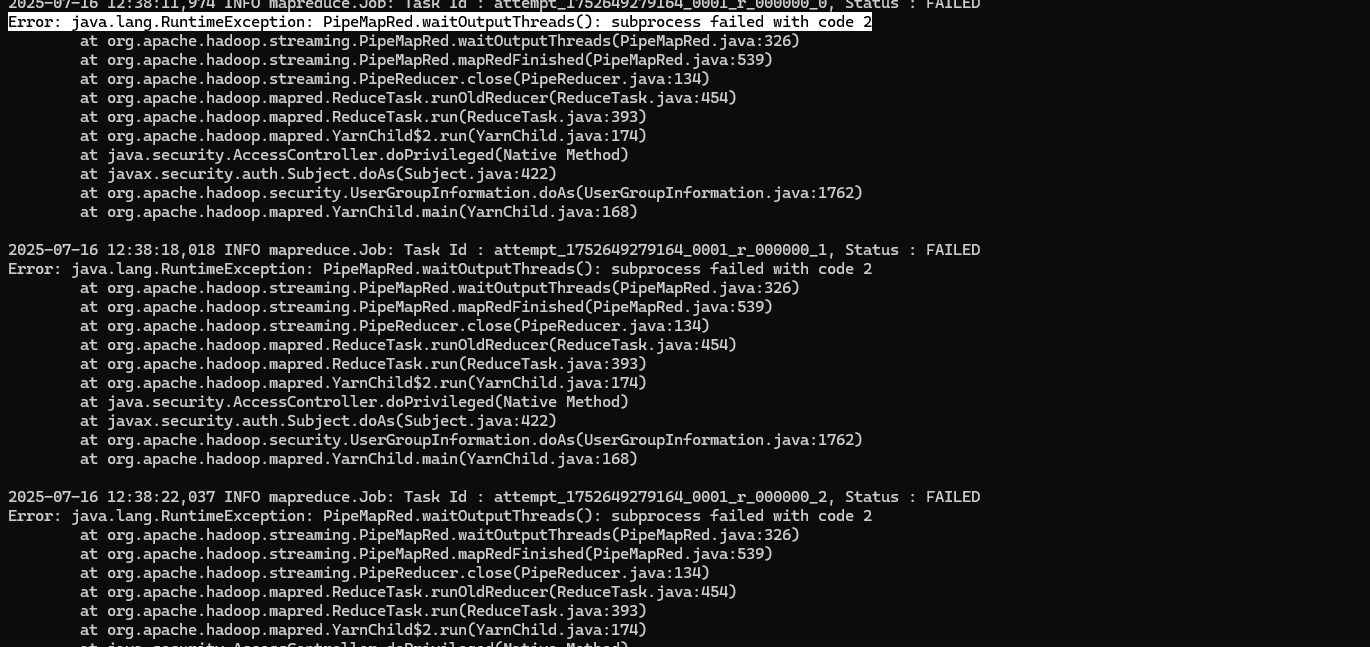
**HADOOP**

1. **Data Ingestion:**

* Create a directory in HDFS and transfer the banking dataset from the local system to the HDFS directory.



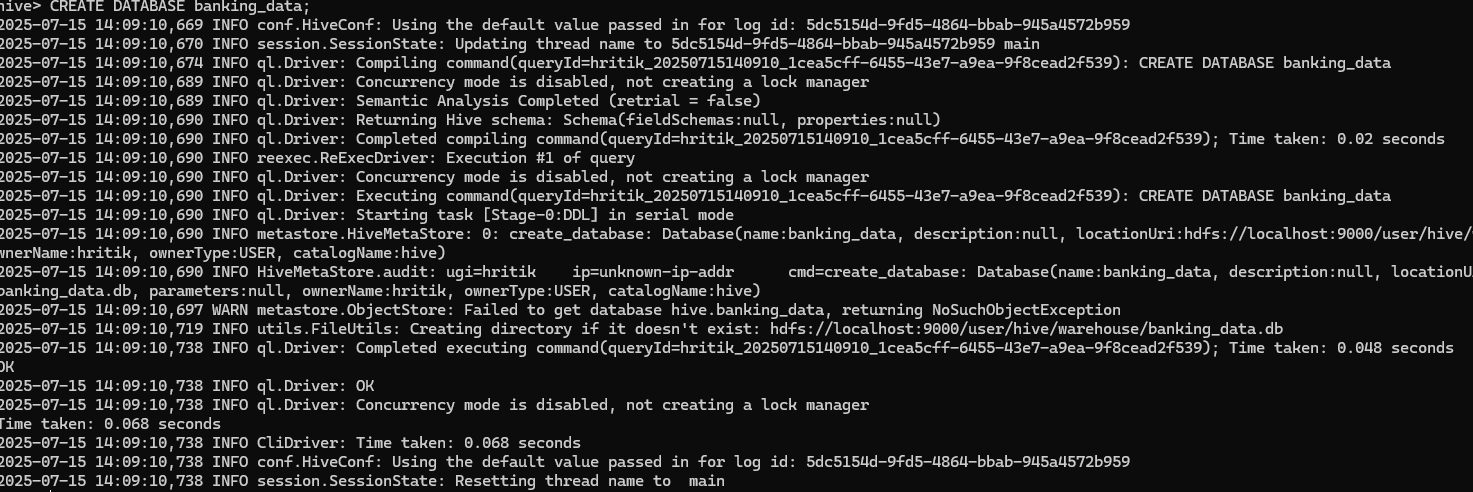
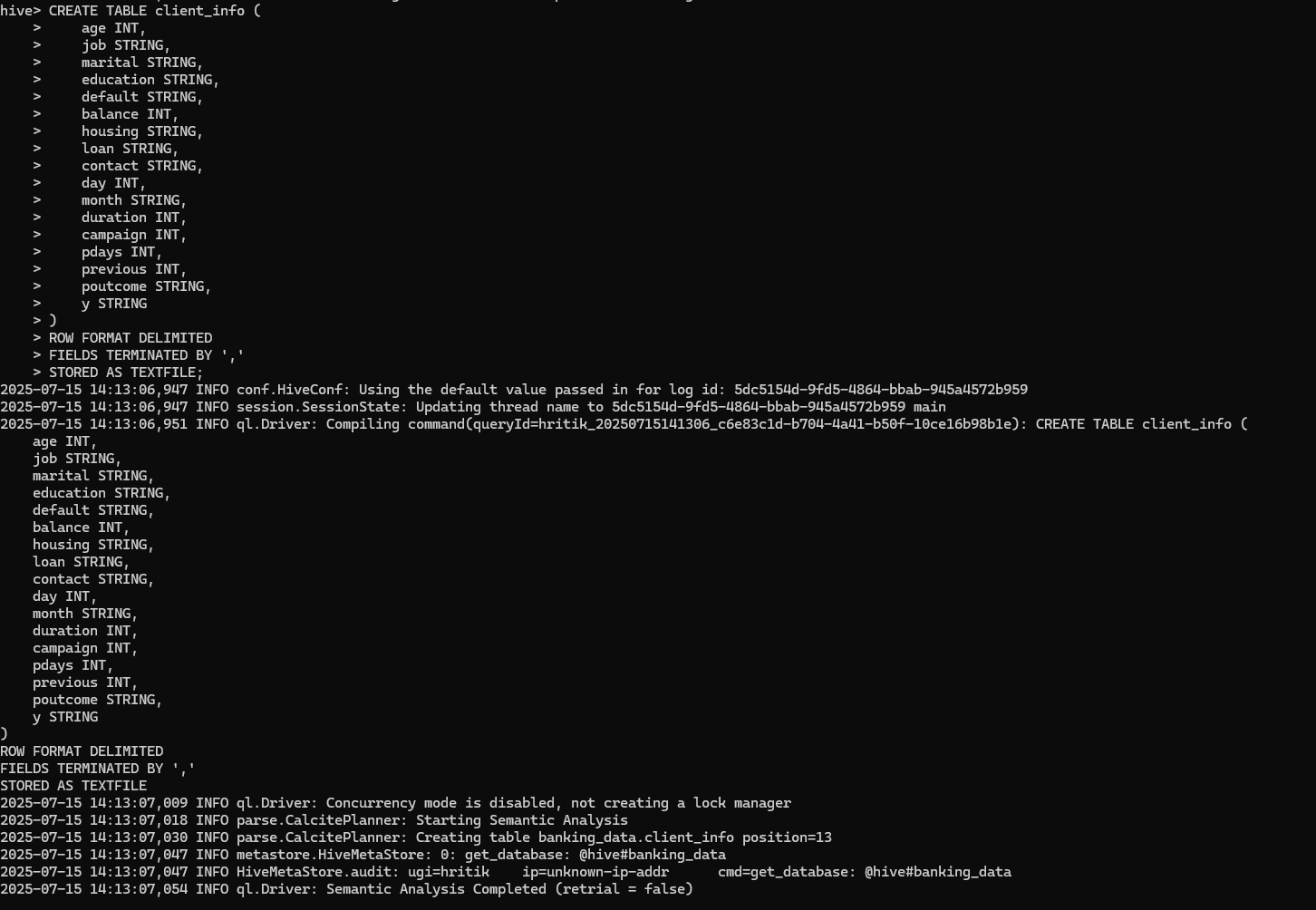
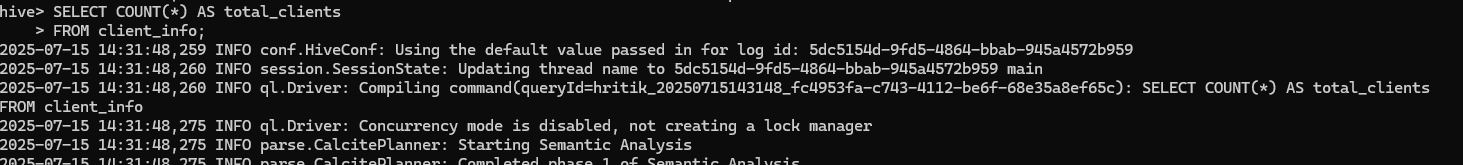
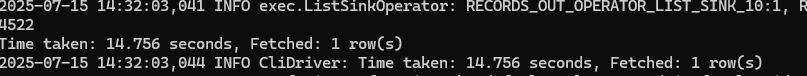
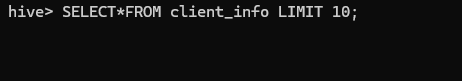
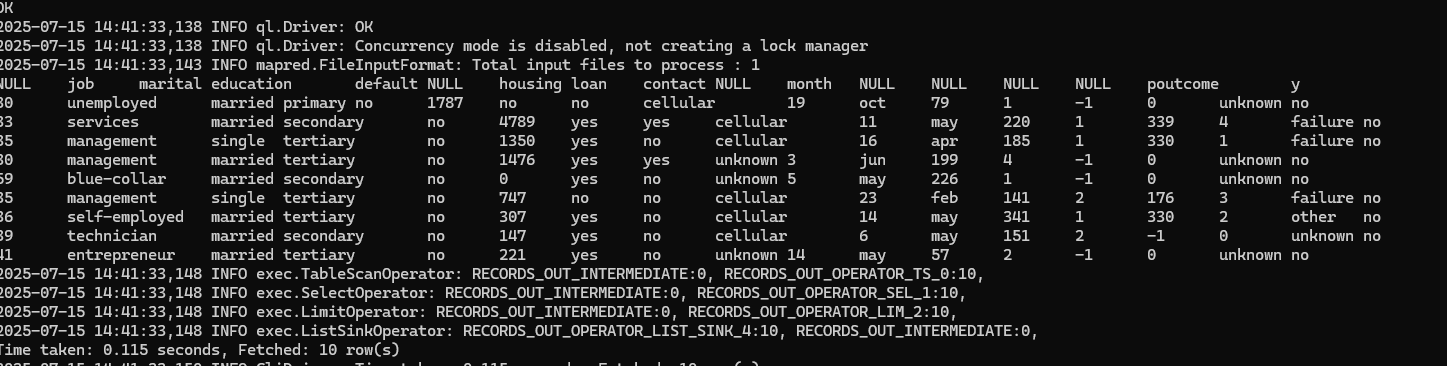
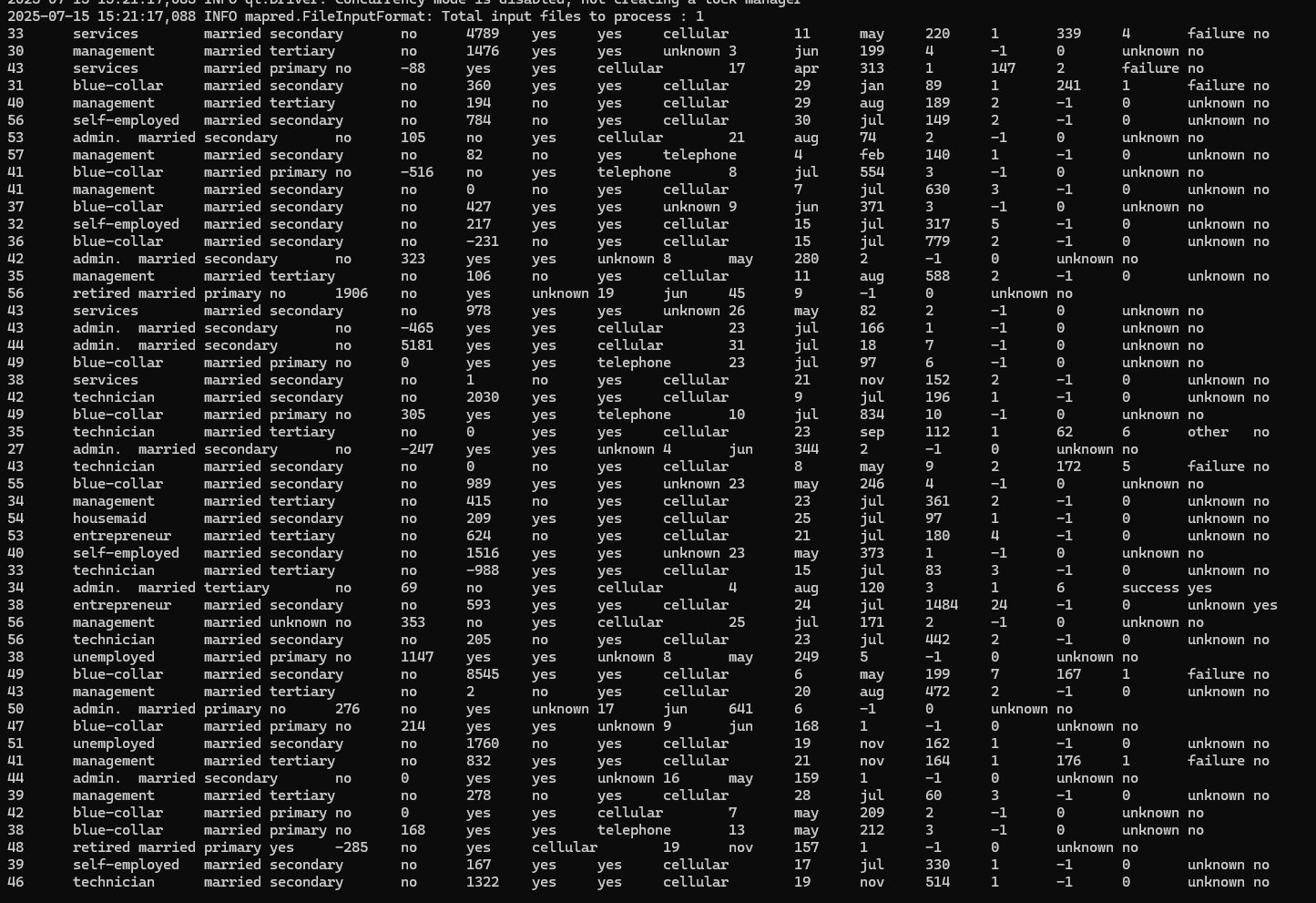
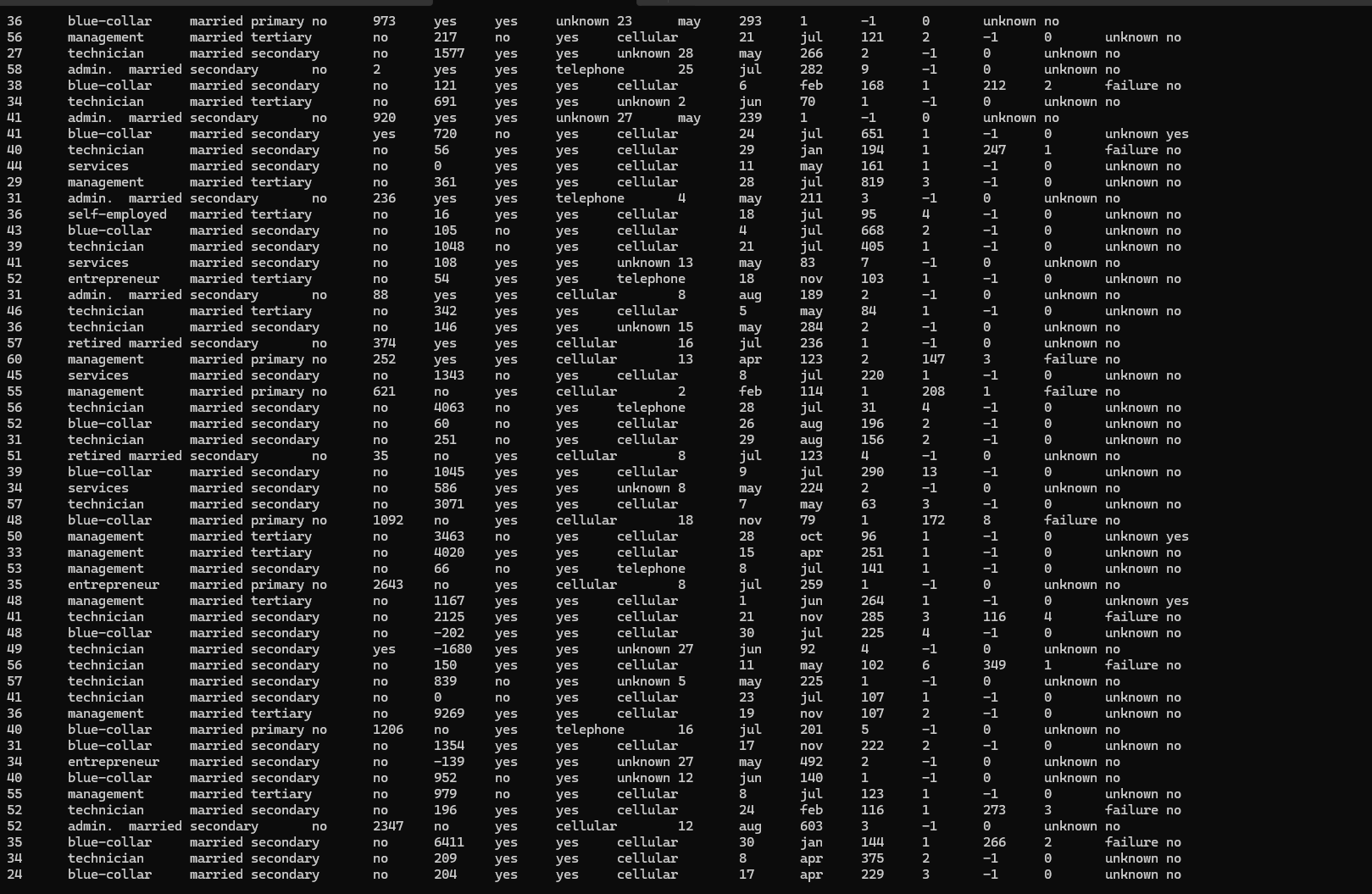
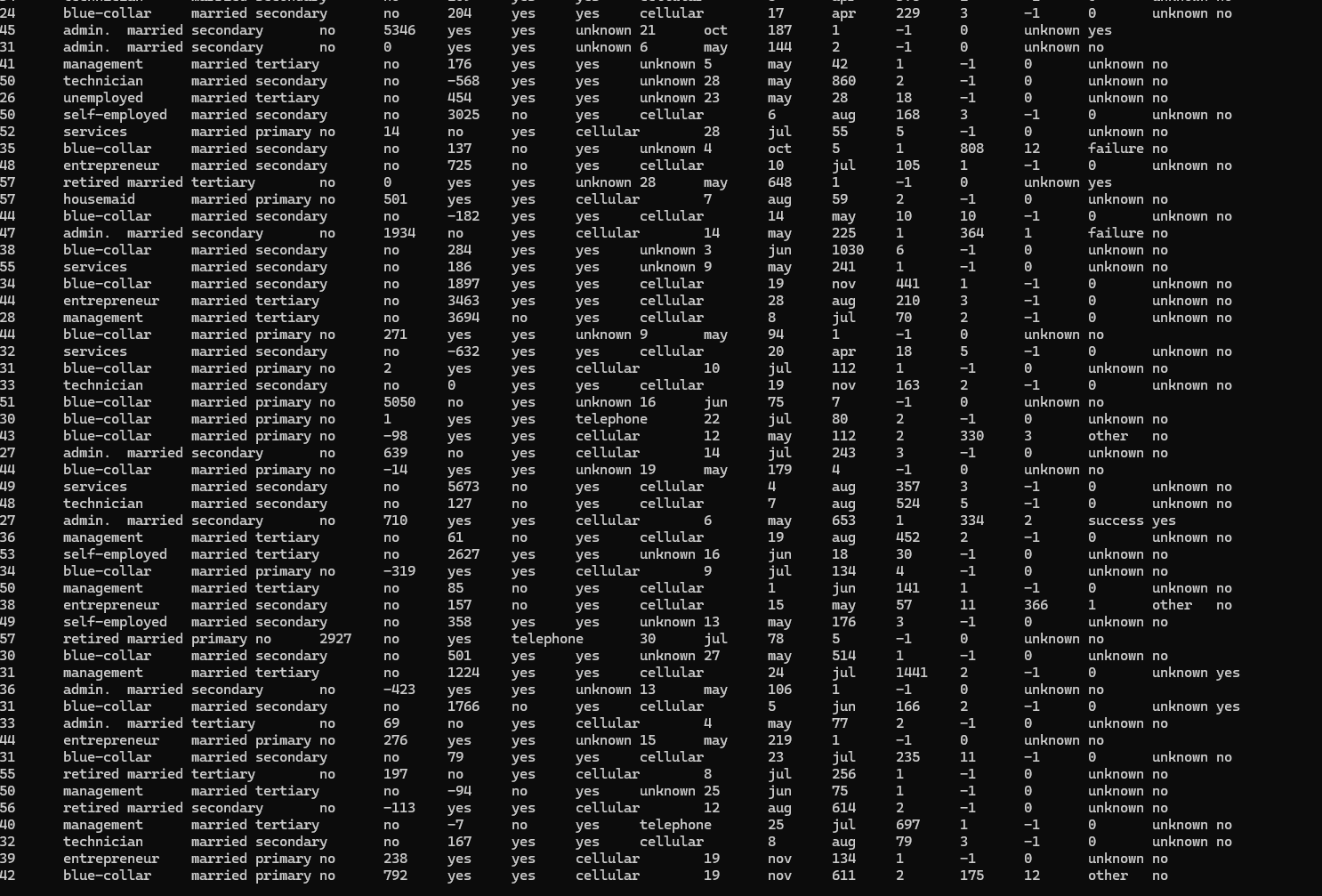
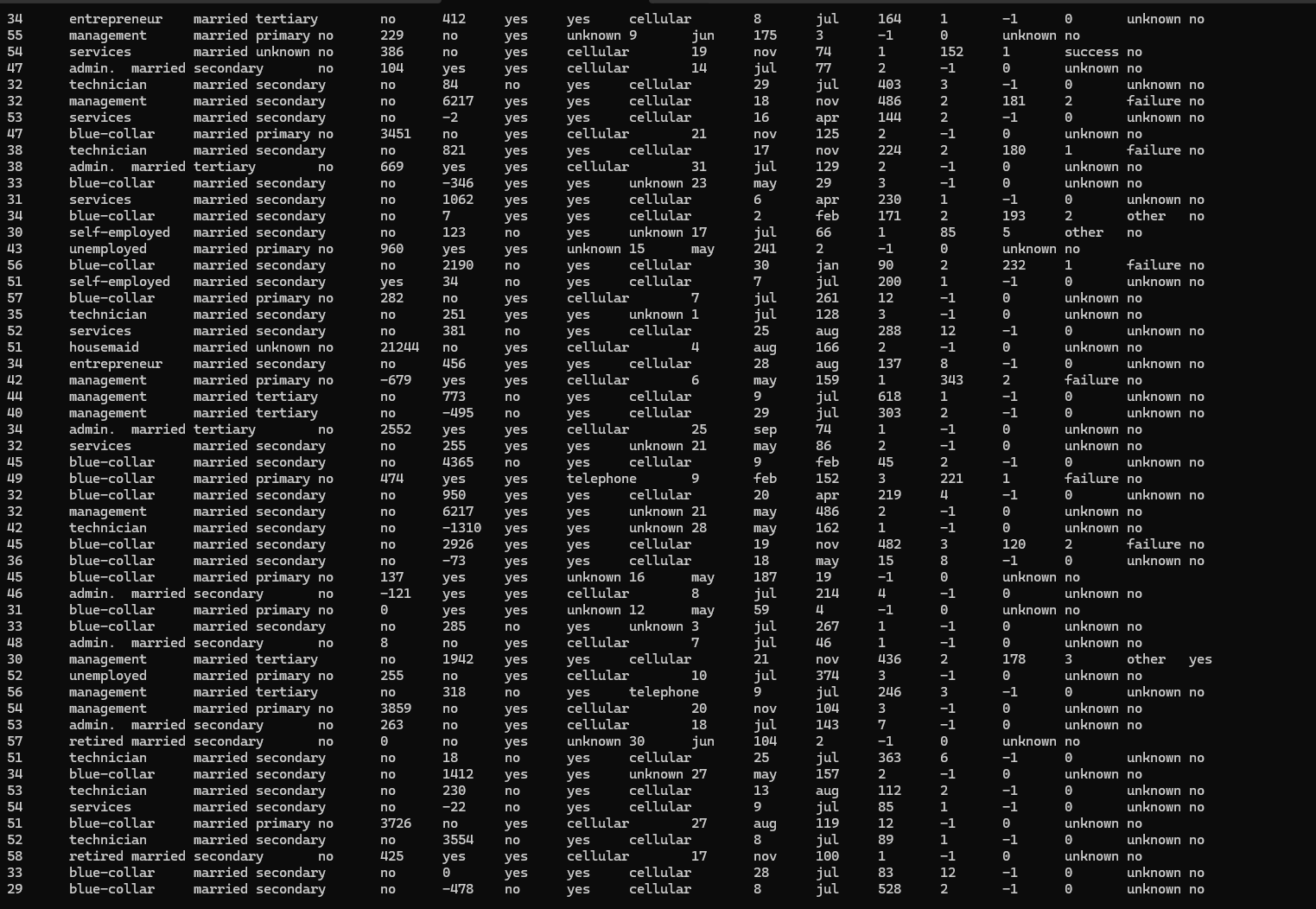
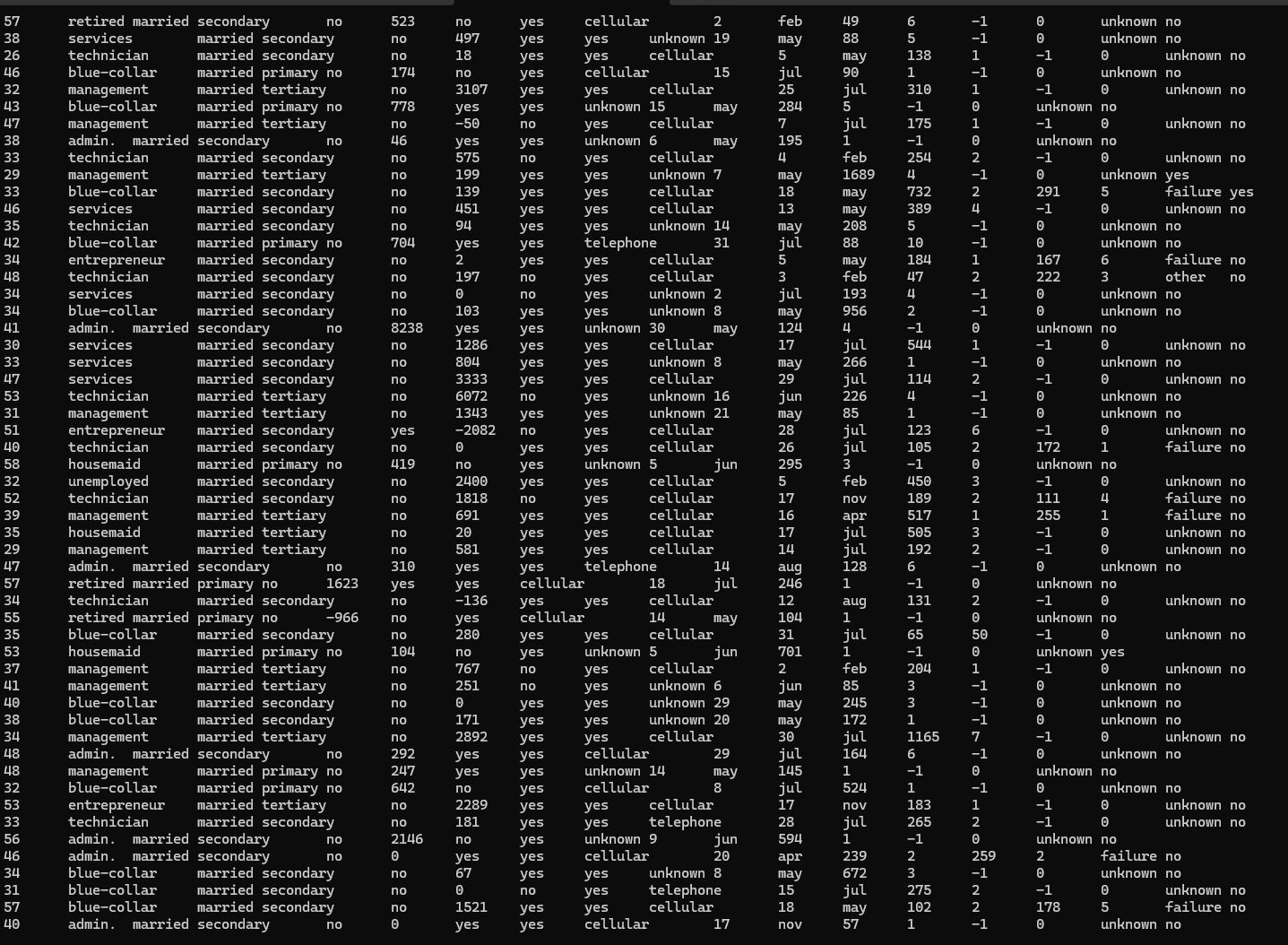
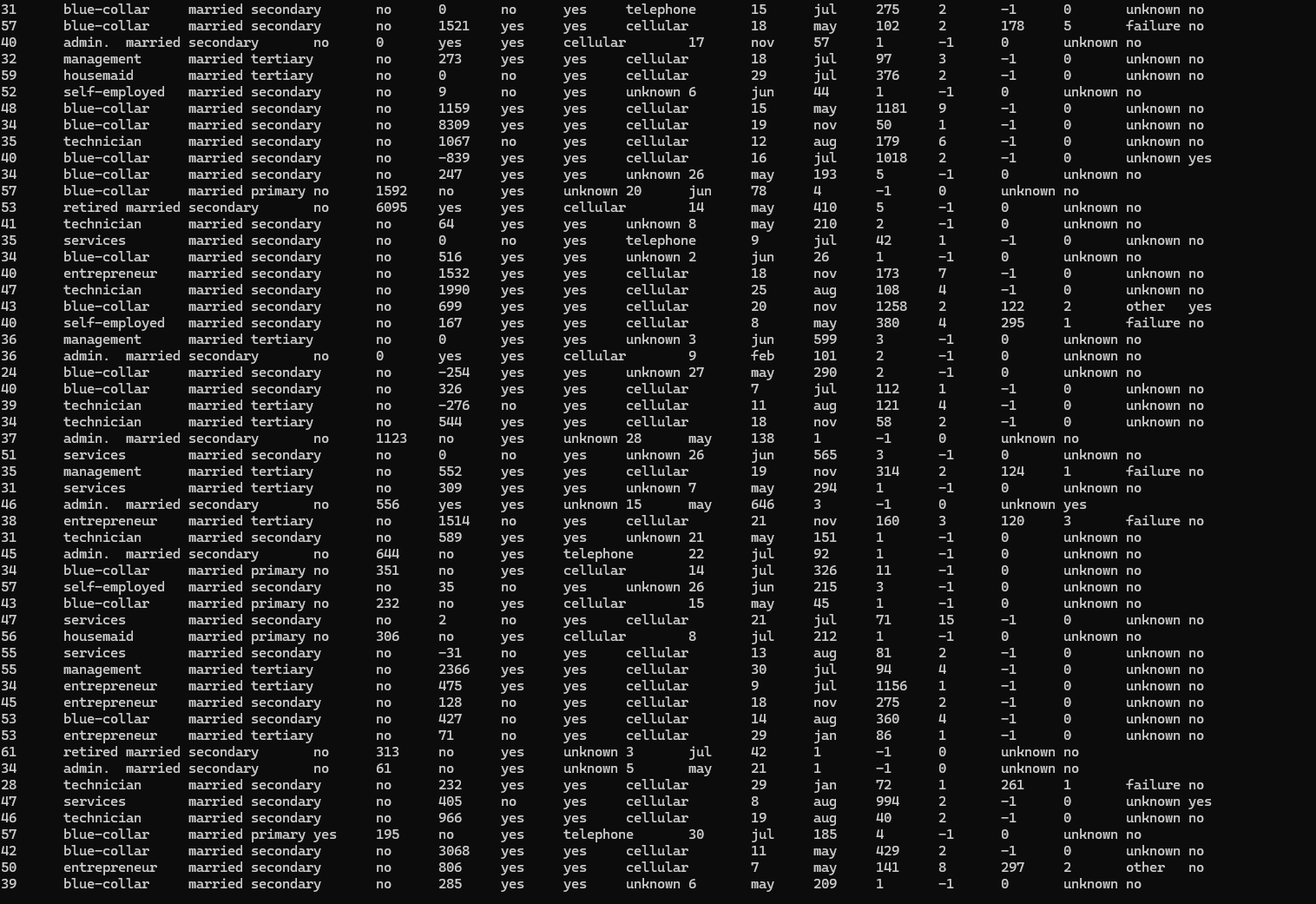
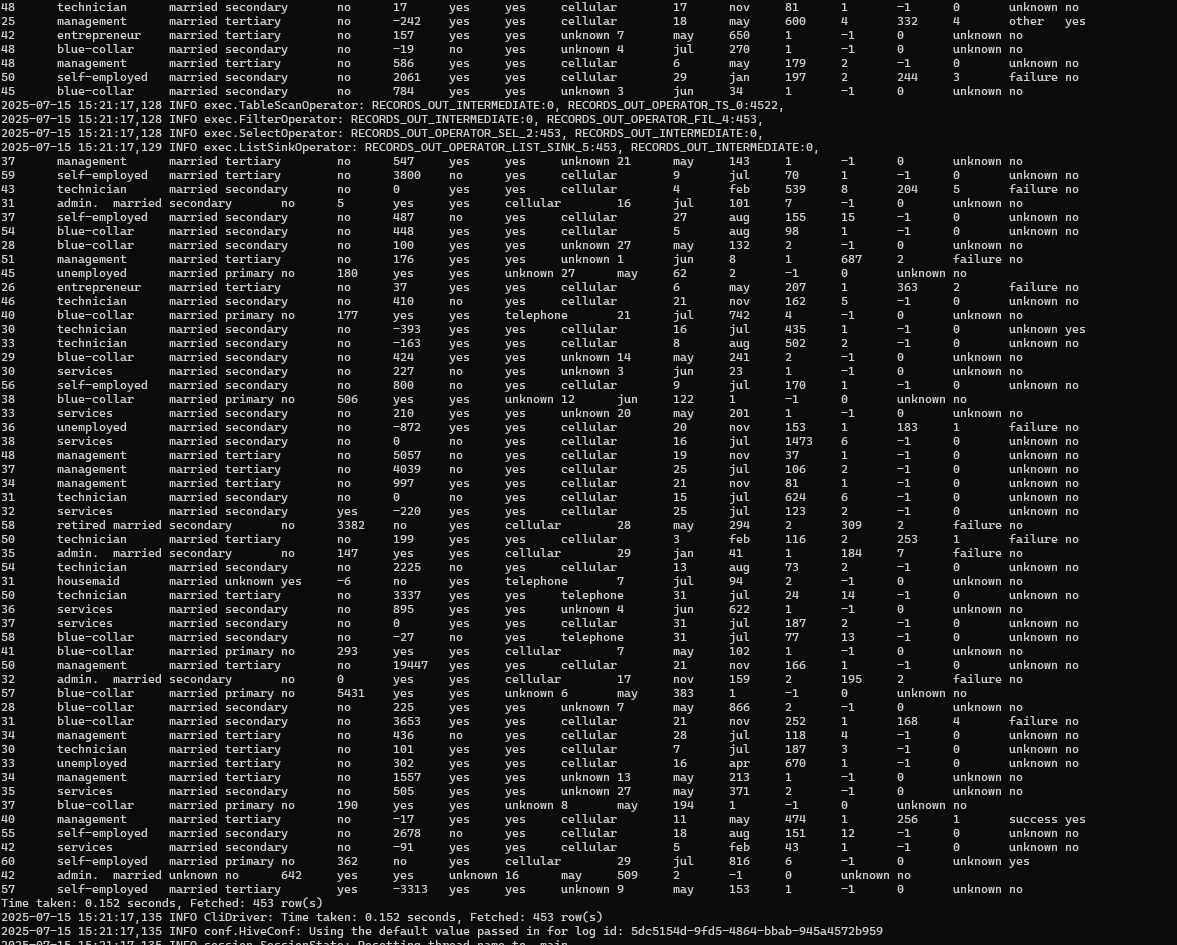
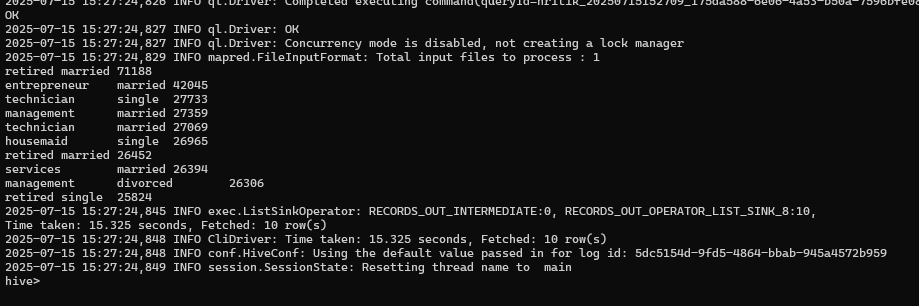
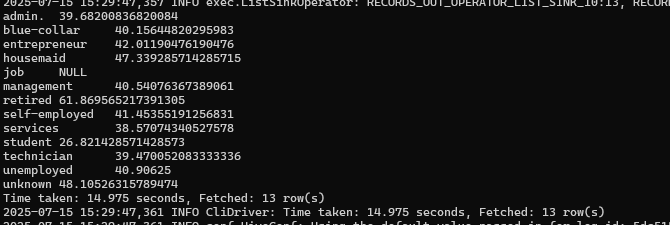
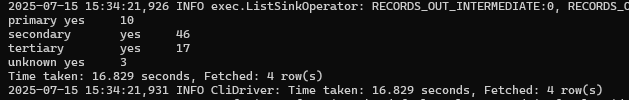
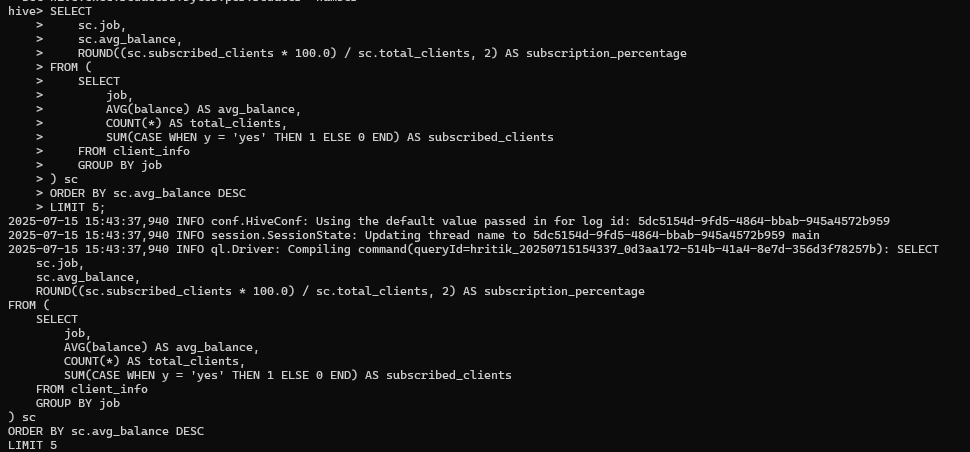
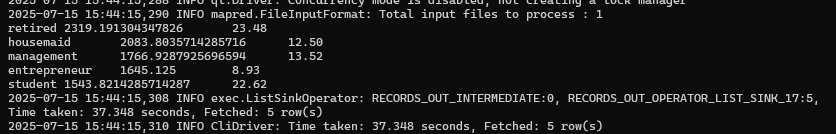
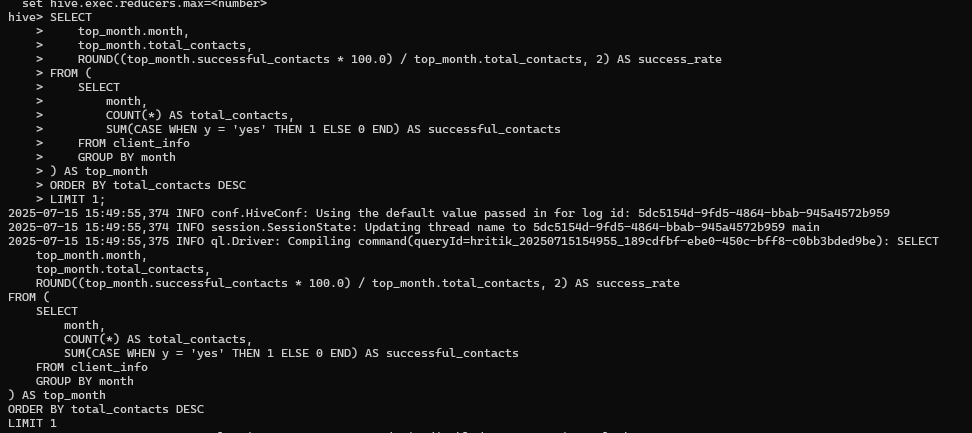
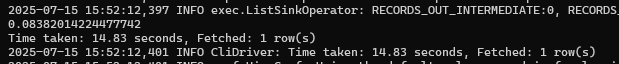
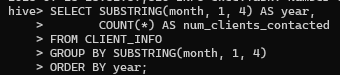
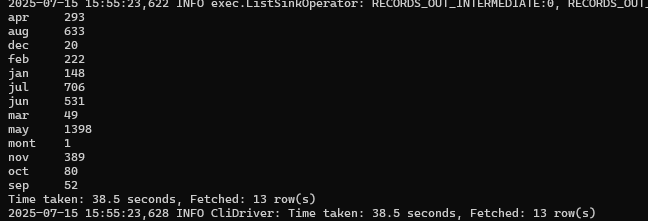
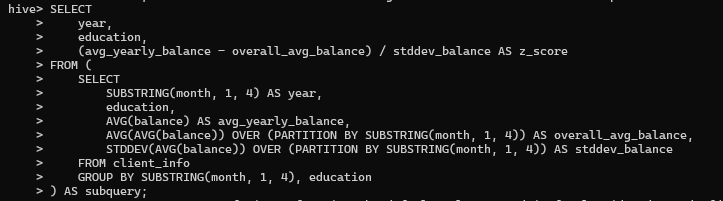
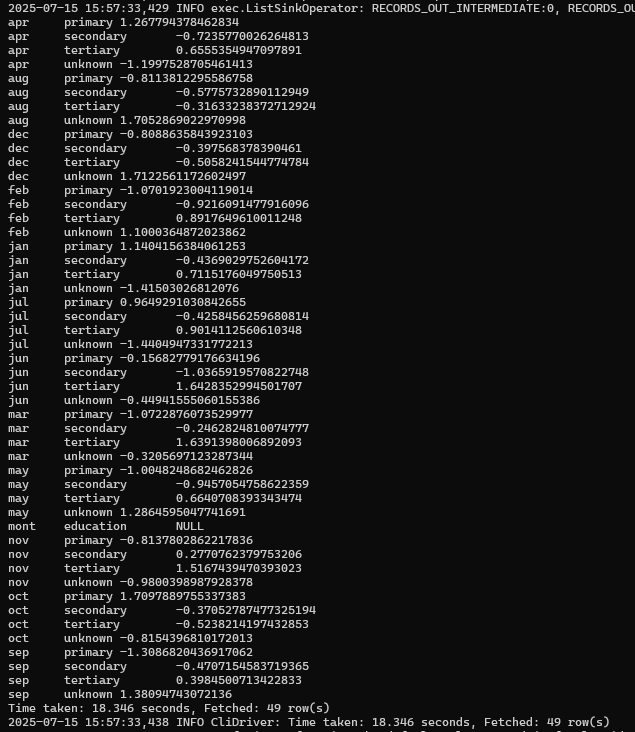
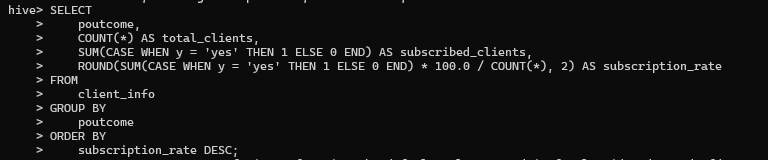
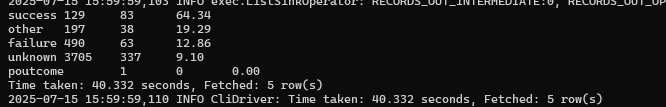
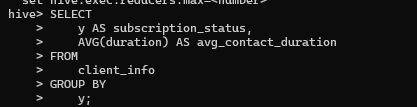
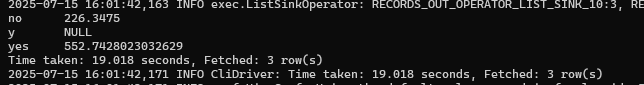
1. **Data Transformation with MapReduce:**

* Write a MapReduce program in Python that calculates the average account balance for each job type.
* 
* (I got this error and i am unable to solve this error in my local windows system.)
* Write another MapReduce program that counts the number of individuals with and without a housing loan in each education category.
* Perform a MapReduce job to determine the number of clients contacted in each month and their subscription status to term deposits ('y' column).

1. **Data Analysis with MapReduce:**

* Analyze the average duration of contact (in seconds) per campaign outcome ('poutcome').
* Examine the relationship between the age of clients and their balance, and present findings in a summarized form.

**HIVE**

1. **Data Ingestion and Table Creation**:
   * Create a Hive database named **banking\_data**.
   * 
   * Define and create a Hive table **client\_info** with appropriate data types for the **bank.csv** dataset.
   * 
   * Load the data from the **bank.csv** file into the **client\_info** table.
   * 
2. **Basic Data Exploration**:
   * Write a HiveQL query to count the total number of clients in the dataset.
   * 
   * 
   * Display the first 10 rows of the dataset.SELECT COUNT(\*) AS total\_clients
   * FROM client\_info;
   * 
   * 
   * ● Total Number of Clients: The query returns a single number representing the total number of clients in the dataset. This number gives you a quick overview of the dataset size, indicating how many client records are available for analysis. So, here we can see that the total number of clients is 4522.
3. **Data Filtering and Sorting**:
   * Retrieve all records of clients who are married and have a personal loan.
   * 
   * 
   * 
   * 
   * 
   * 
   * 
   * 
   * List the top 10 clients with the highest balance, displaying their job, marital status, and balance.
   * 
   * 
4. **Data Aggregation and Grouping**:
   * Calculate the average age of clients for each job category.
   * 
   * 
   * Find the total number of clients for each education level who have defaulted on credit.  
     
   * 
5. **Complex Queries for Insights**:
   * Identify the top 5 job categories with the highest average balance and the percentage of clients in each of these job categories who have subscribed to a term deposit.
   * 
   * 
   * Determine the month with the highest number of contacts and the success rate of the campaign in that month (percentage of clients who subscribed to a term deposit).
   * 
   * 
6. **Correlation Analysis**:
   * Calculate the correlation between age and balance for the clients.
   * 
   * 
7. **Trend Analysis**:
   * Analyze the year-over-year trend in the number of clients contacted.
   * 
   * 
8. **Anomaly Detection**:
   * Identify any unusual patterns in the average yearly balance across different education levels.
   * 
   * 
9. **Advanced Analysis**:
   * Analyze the impact of previous campaign outcomes (**poutcome**) on the current campaign's success. Calculate the subscription rate (to term deposits) for each **poutcome** category.
   * 
   * 
   * Compare the average contact duration for clients who subscribed and who did not subscribe to a term deposit.
   * 
   * 

**Submission Guidelines:**

* Make a copy of this doc file.
* Perform the analysis in your local system using Hadoop and Hive and provide screenshots of both the **code** and the **output** under each question.
* Upload the doc file with other files and submit it in the submission dashboard.