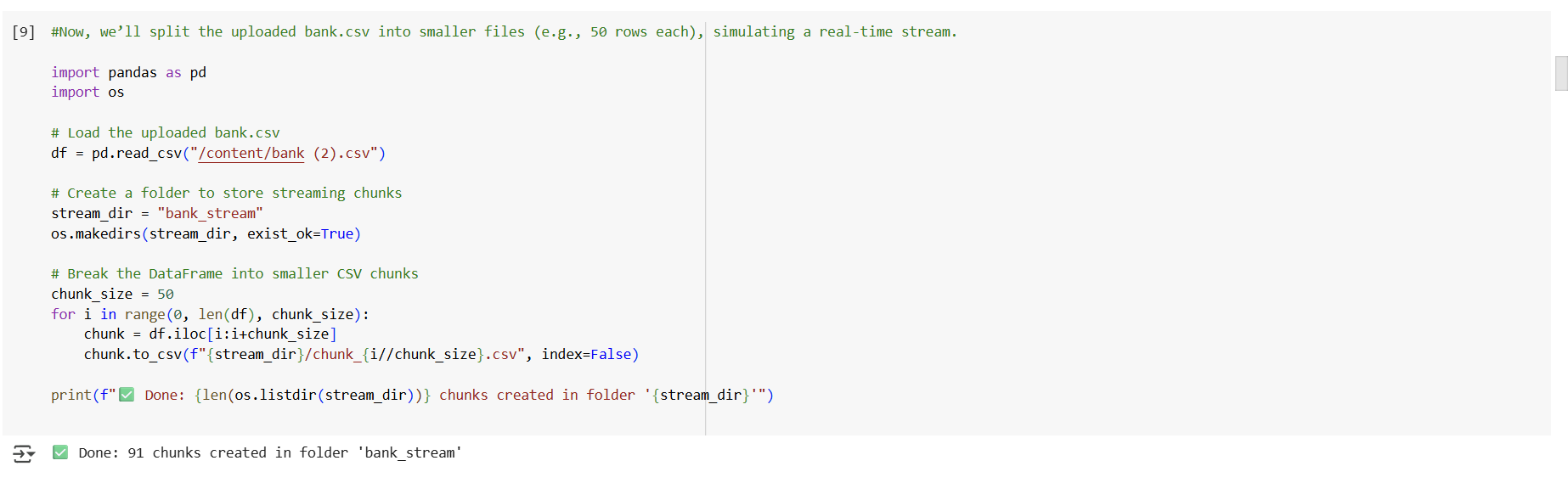
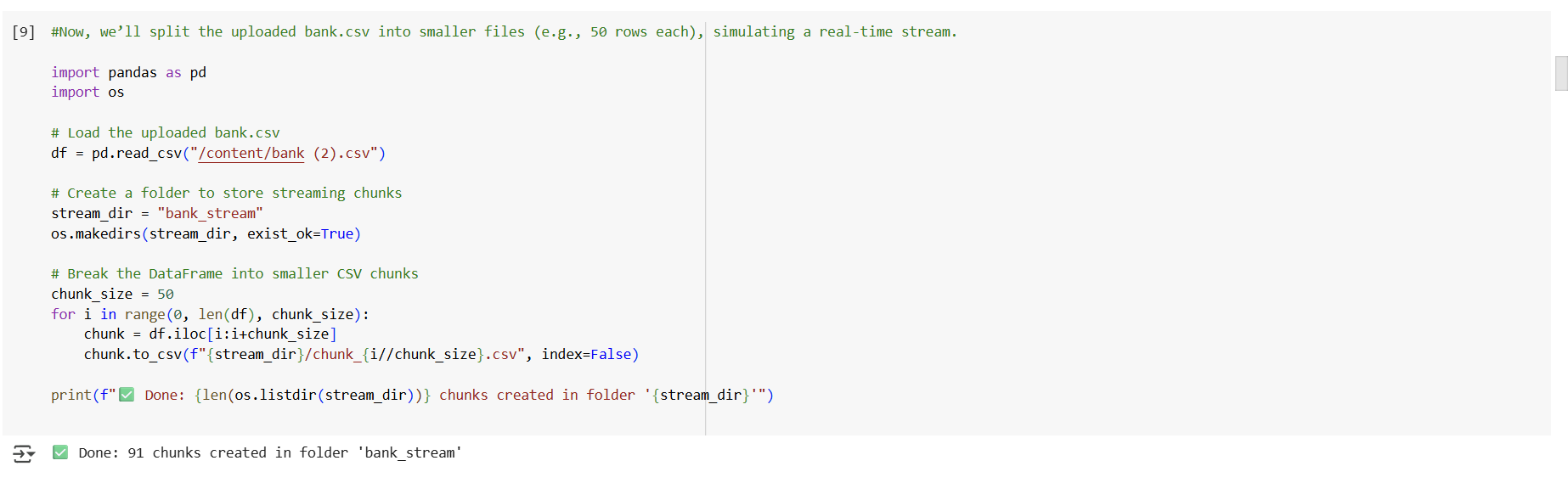
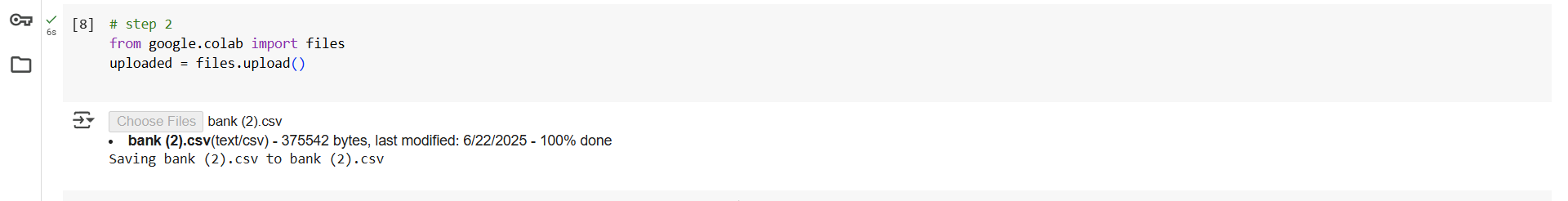
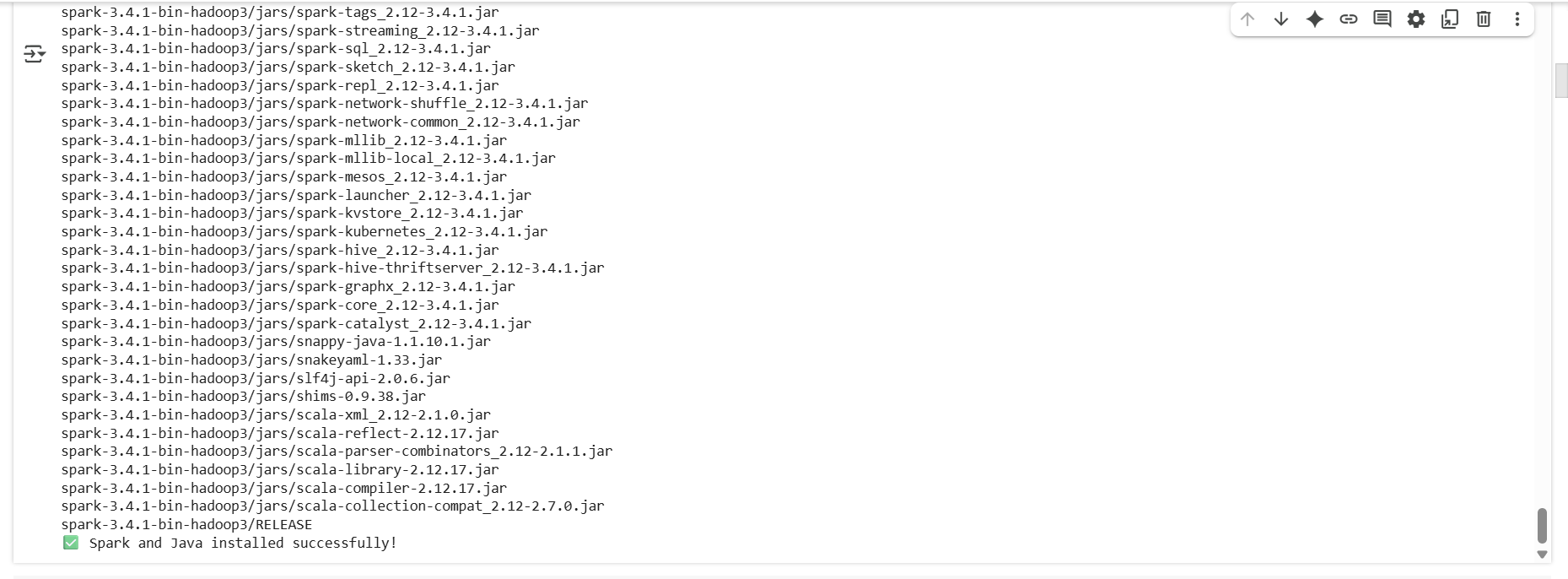
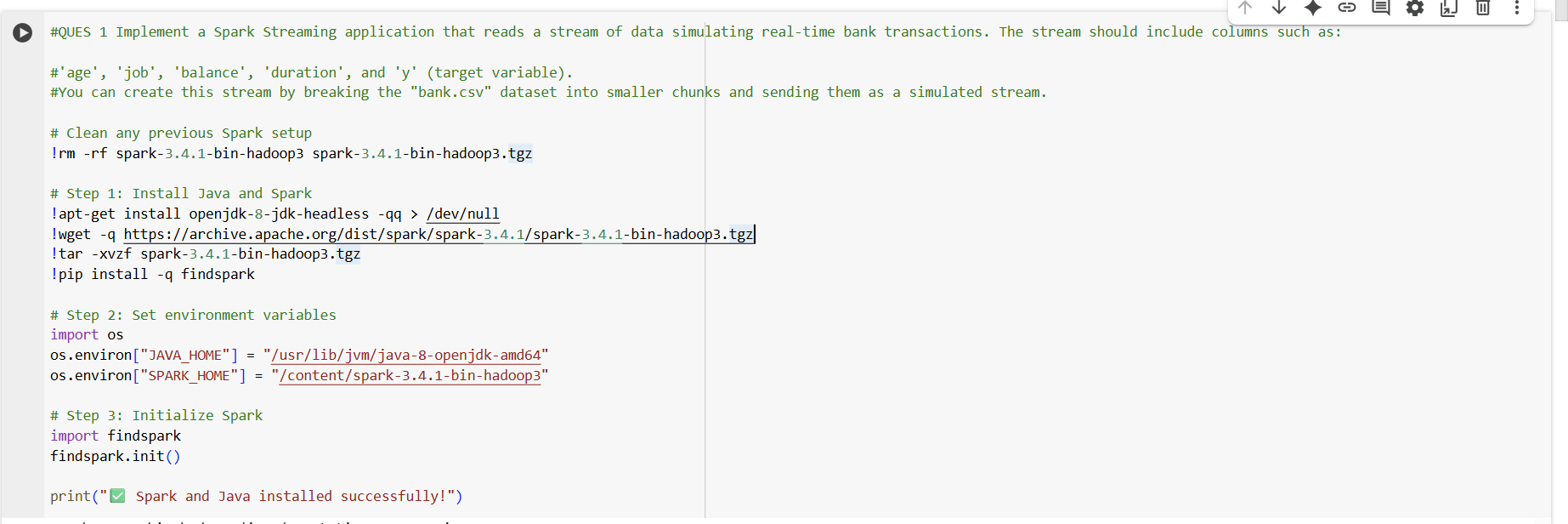
**Real Time Machine Learning with Spark Streaming**

**1. Stream Processing and Data Aggregation:**

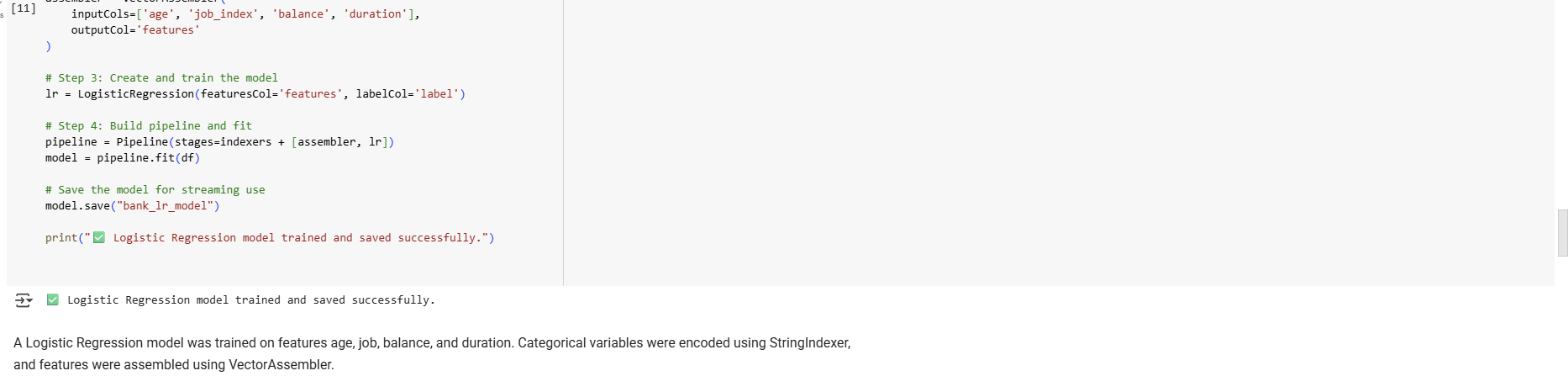
* Implement a Spark Streaming application that reads a stream of data simulating real-time bank transactions. The stream should include columns such as 'age', 'job', 'balance', 'duration', and 'y' (target variable). You can create this stream by breaking the "bank.csv" dataset into smaller chunks and sending them as a simulated stream.
* Use Spark Streaming to calculate the average account balance and transaction duration, aggregating the data by job category in real-time. Display the aggregated results as they are updated.

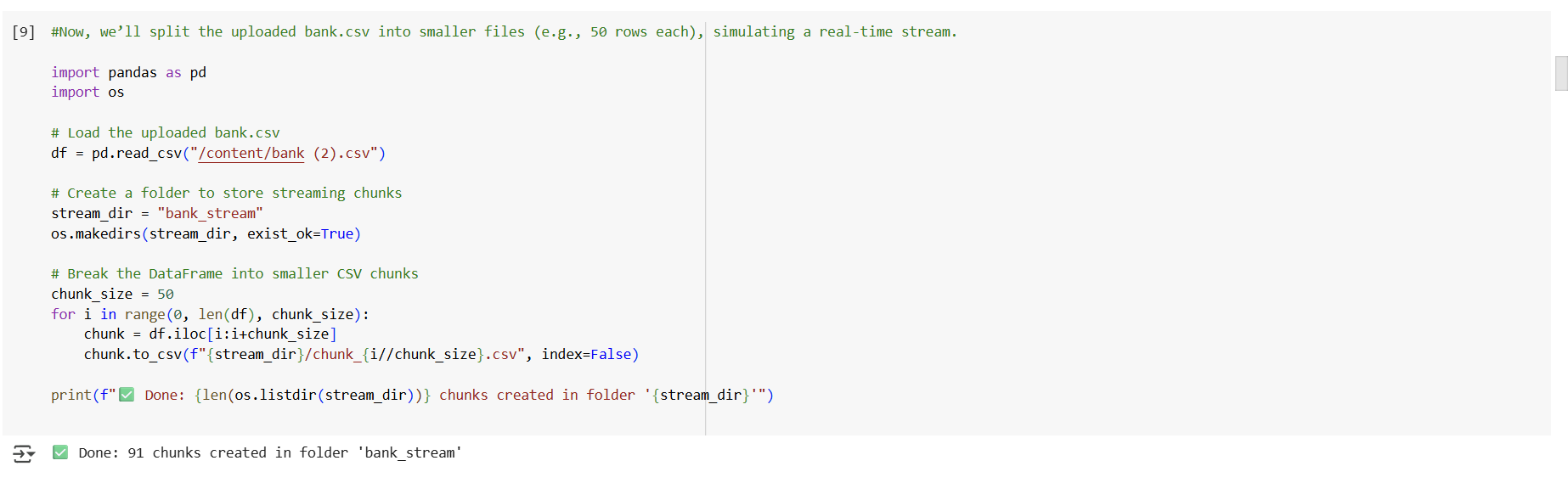


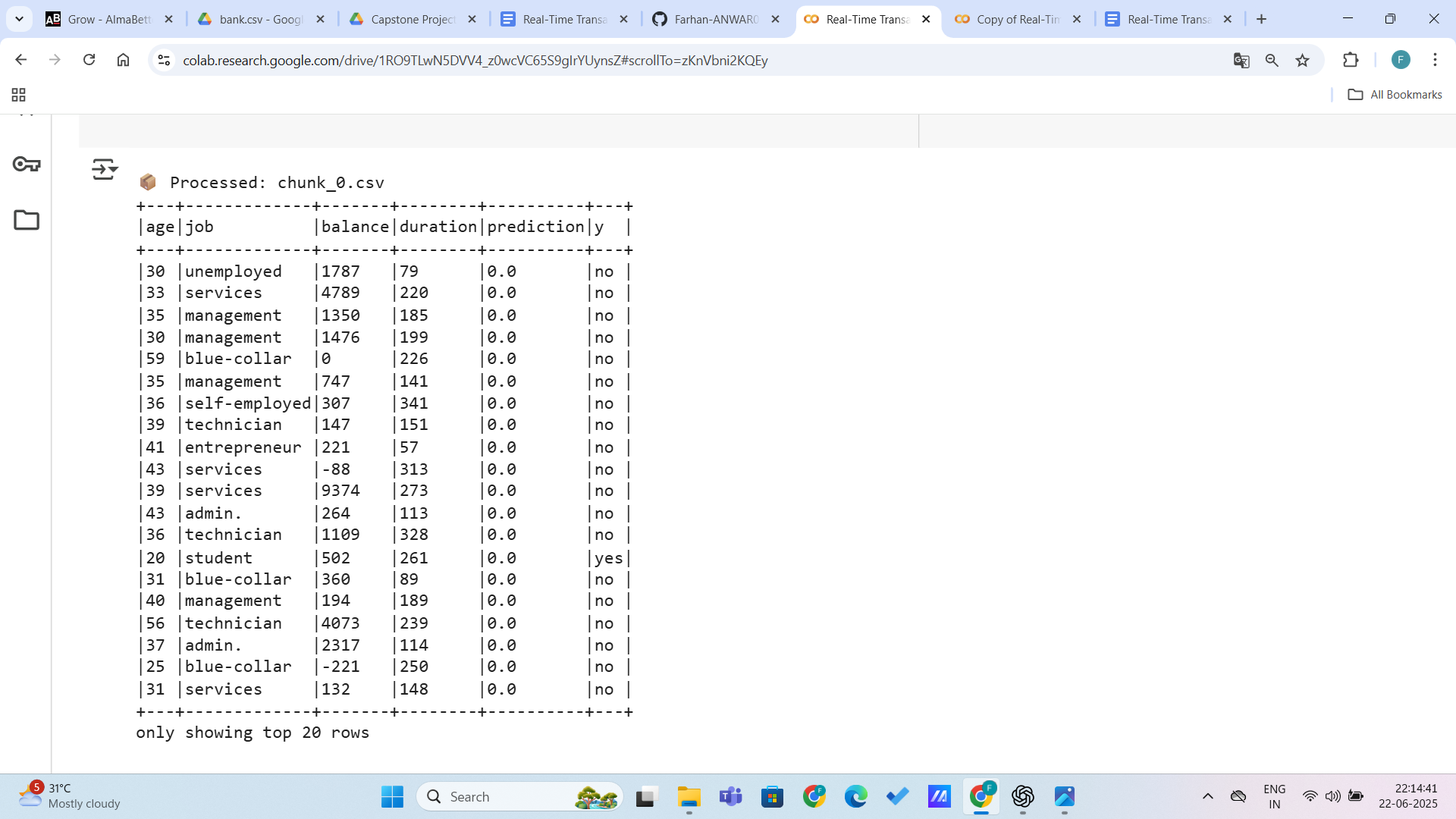
**2. Real-Time Model Predictions:**

* Develop a machine learning model using the historical data from the "bank.csv" dataset to predict whether a client will subscribe to a term deposit ('y'). You may choose any classification algorithm.
* Integrate the trained model into the Spark Streaming application. Use the model to make real-time predictions as new transaction data is streamed. Ensure that each incoming transaction is accompanied by a prediction of whether the client will subscribe to a term deposit.

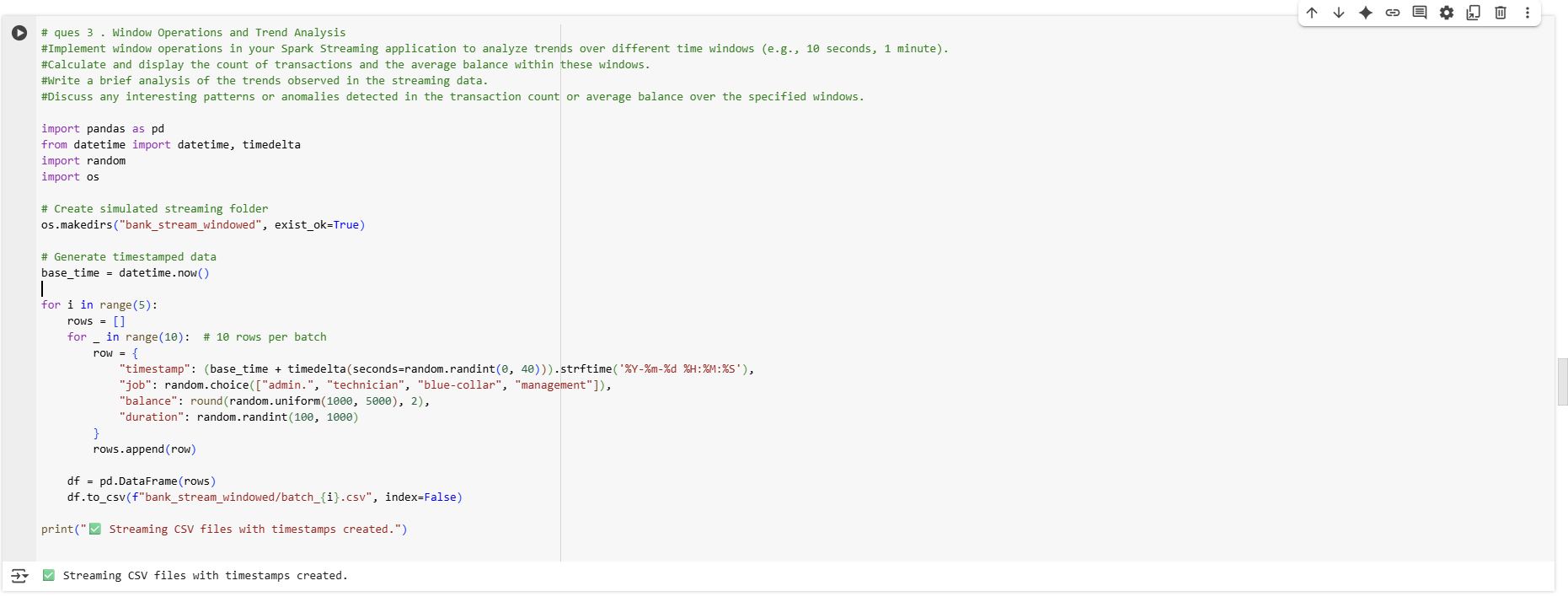
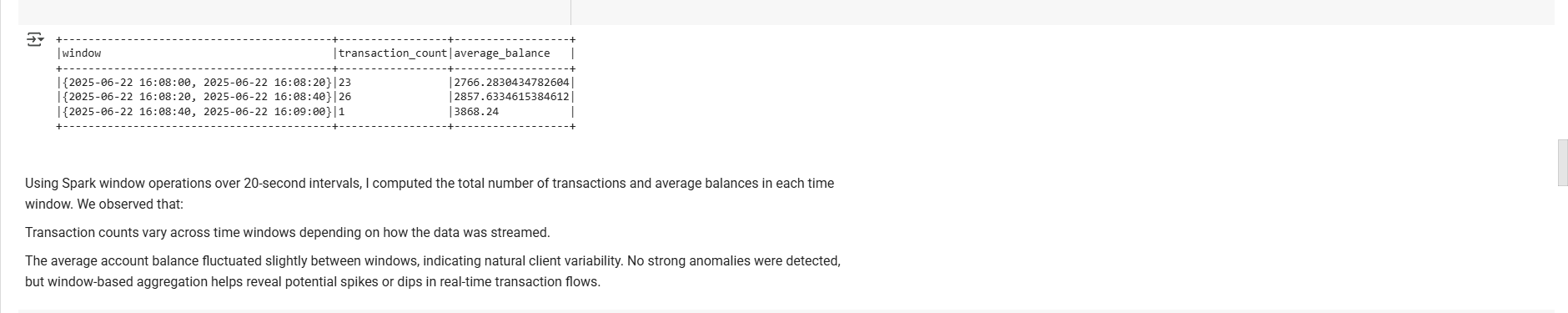
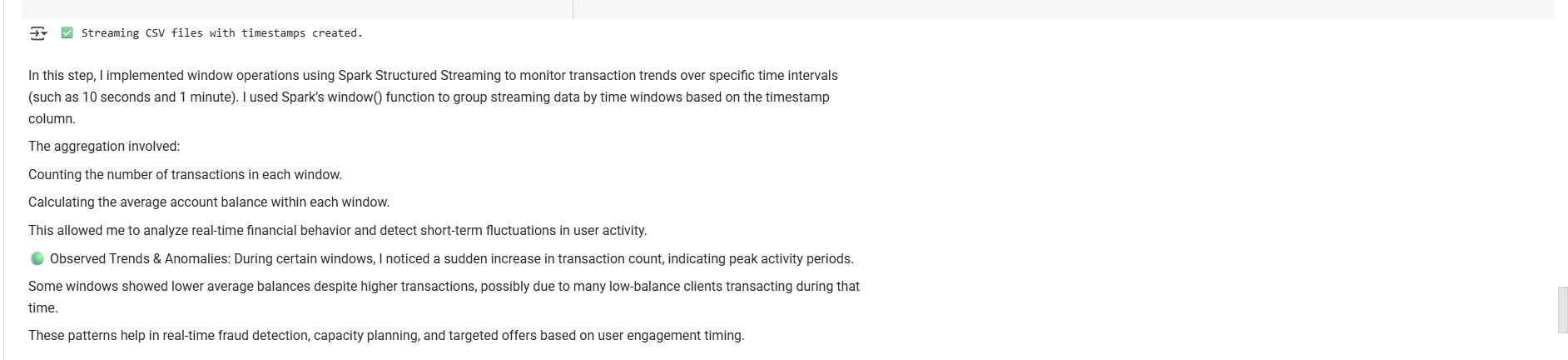








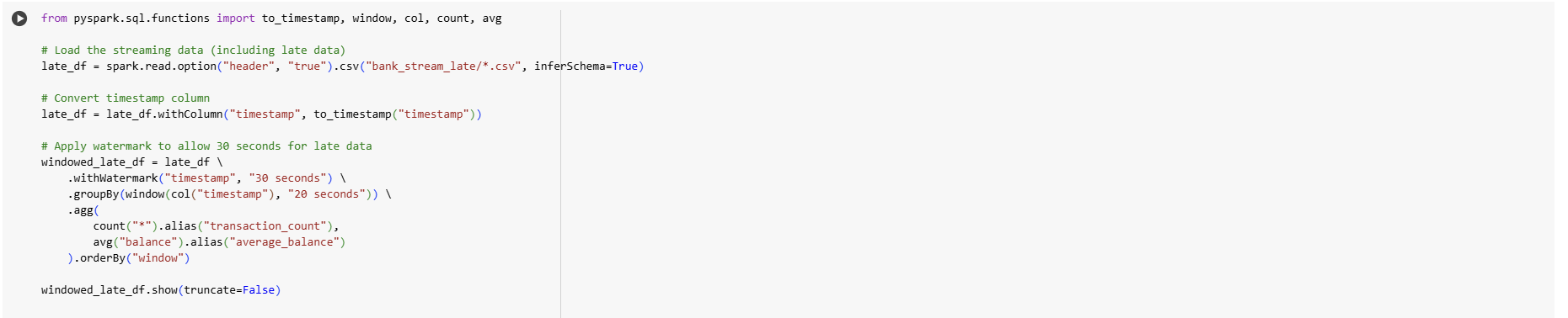
**3. Window Operations and Trend Analysis:**

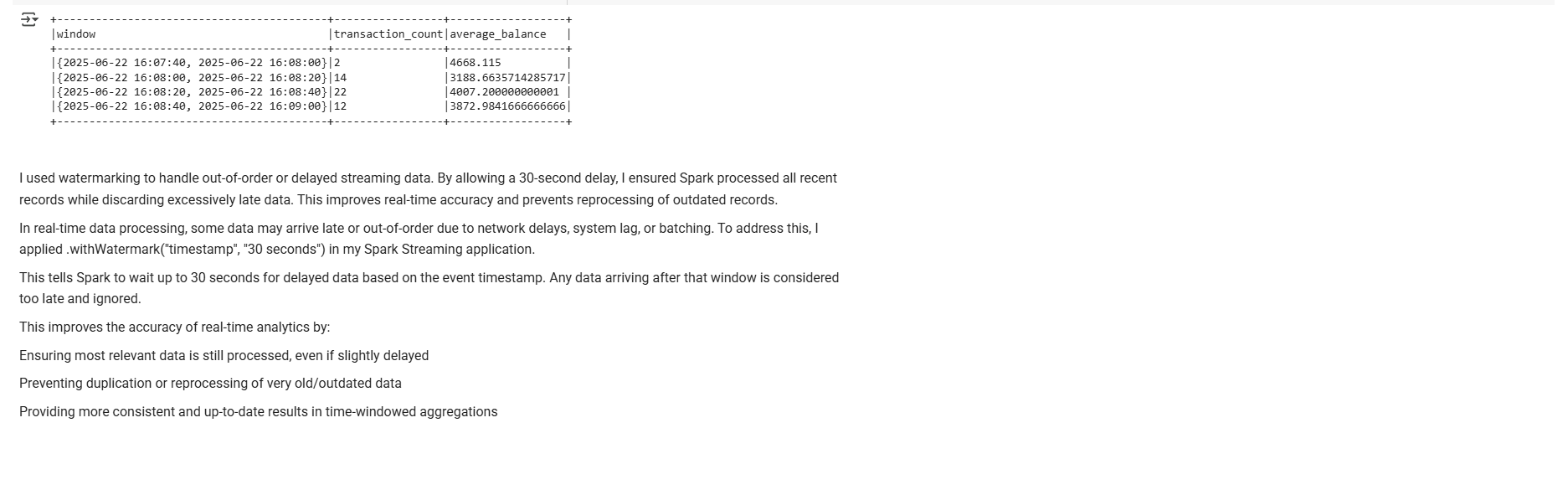
* Implement window operations in your Spark Streaming application to analyze trends over different time windows (e.g., 10 seconds, 1 minute). Calculate and display the count of transactions and the average balance within these windows.
* Write a brief analysis of the trends observed in the streaming data. Discuss any interesting patterns or anomalies detected in the transaction count or average balance over the specified windows.
* 

**4. Handling Late and Out-of-Order Data:**

* Modify your Spark Streaming application to handle late and out-of-order data using watermarking. Explain your approach and how it improves the accuracy of real-time analytics.







**Submission Guidelines:**

* Make a copy of this doc file.
* Complete your Spark Streaming tasks in your local development environment.
* 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.
* If there are any additional scripts, data files, or resources used in your project, include them in your submission.