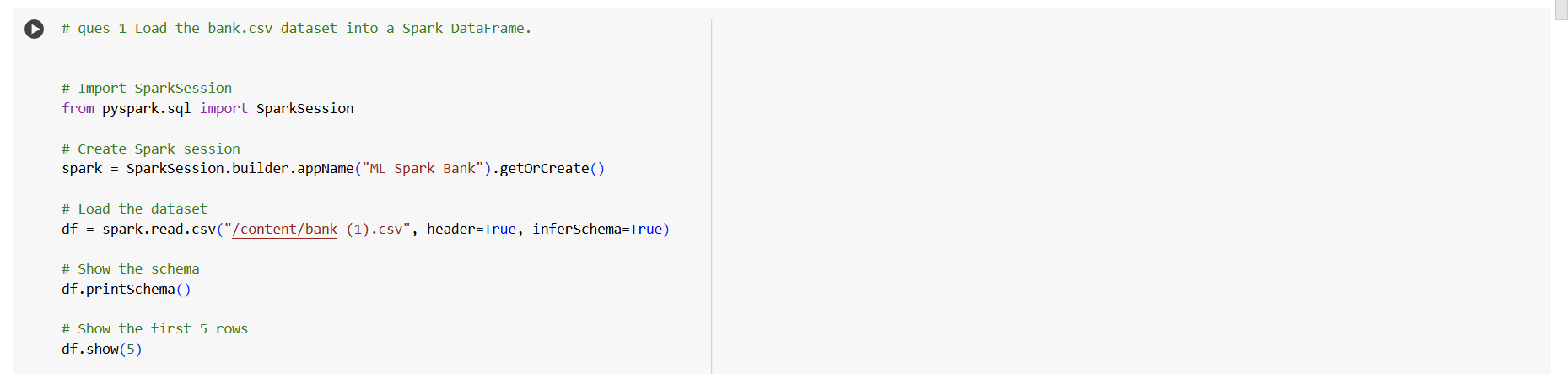
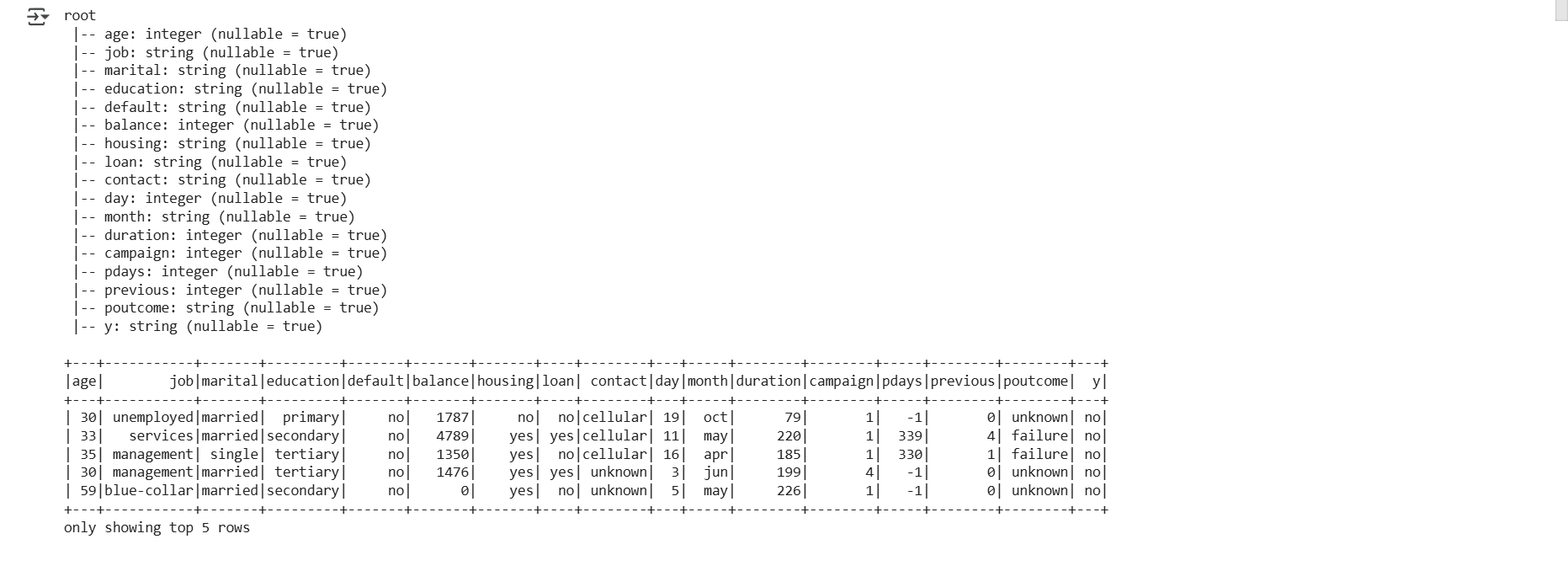
**Machine Learning with Spark ML**

**Objective**

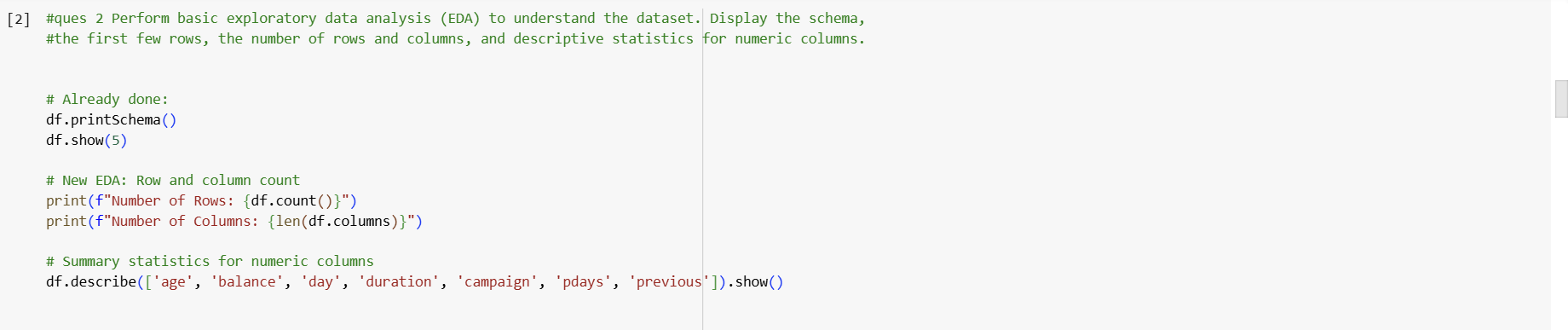
Apply the concepts of Spark ML to build and evaluate a machine learning model using the **bank.csv** dataset. The goal is to predict whether a client will subscribe to a term deposit (binary classification).

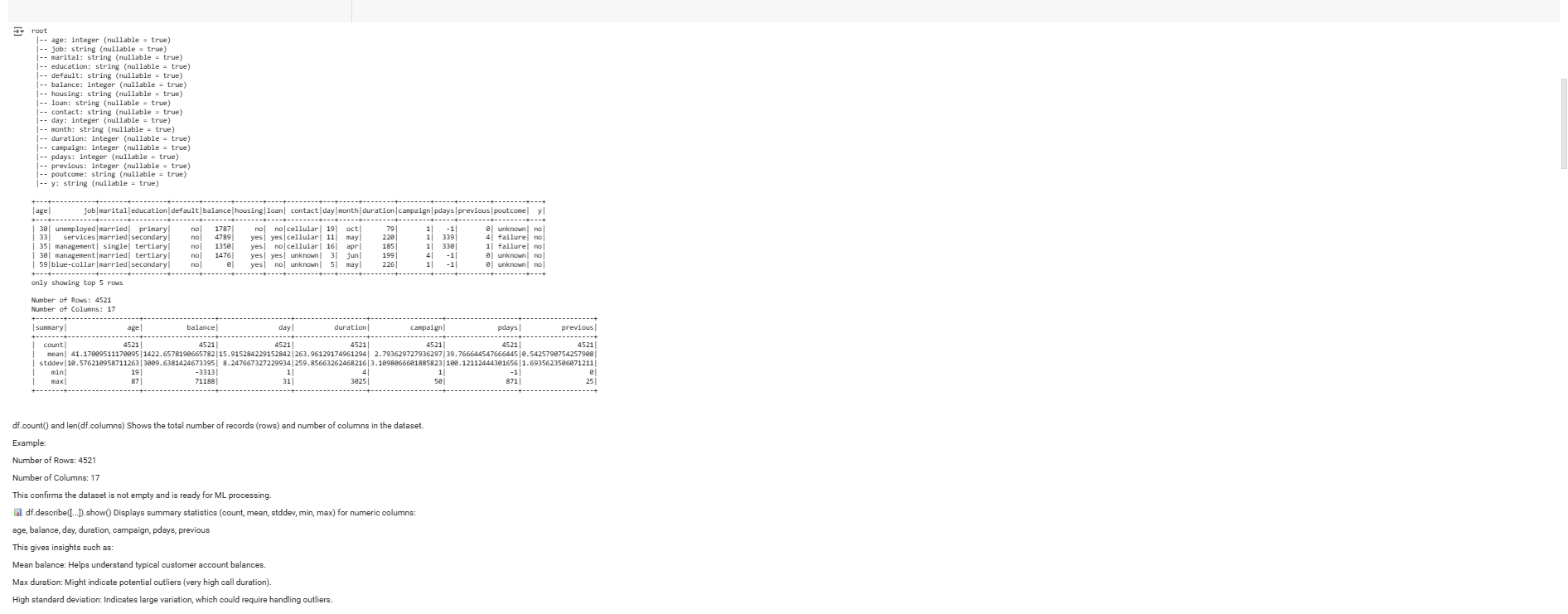
1. **Data Loading and Initial Exploration**:
   * Load the **bank.csv** dataset into a Spark DataFrame.



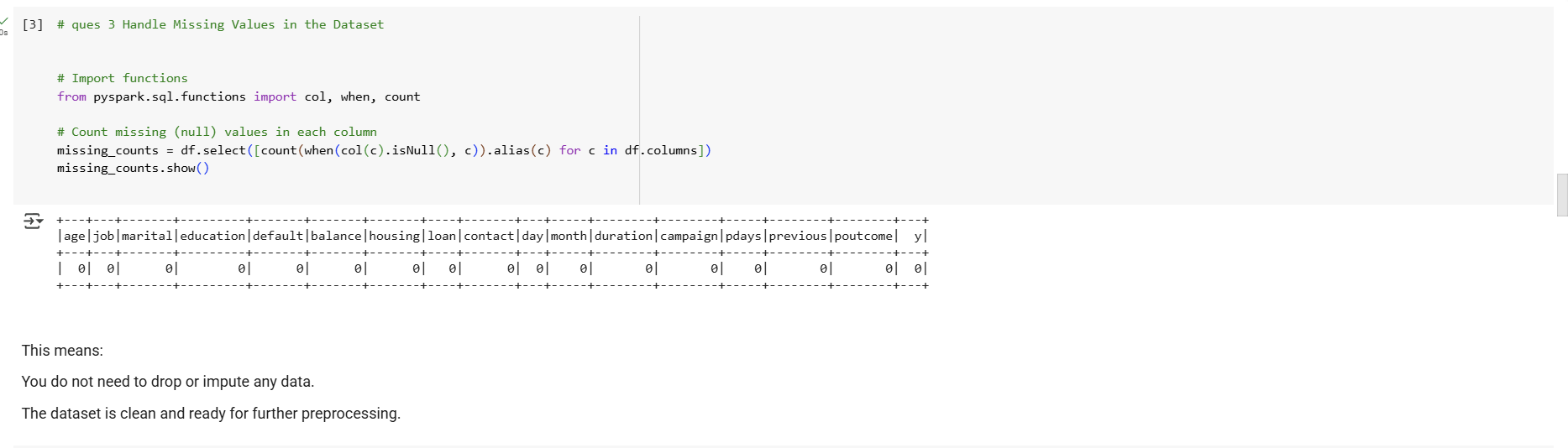


* + Perform basic exploratory data analysis (EDA) to understand the dataset. Display the schema and the first few rows.





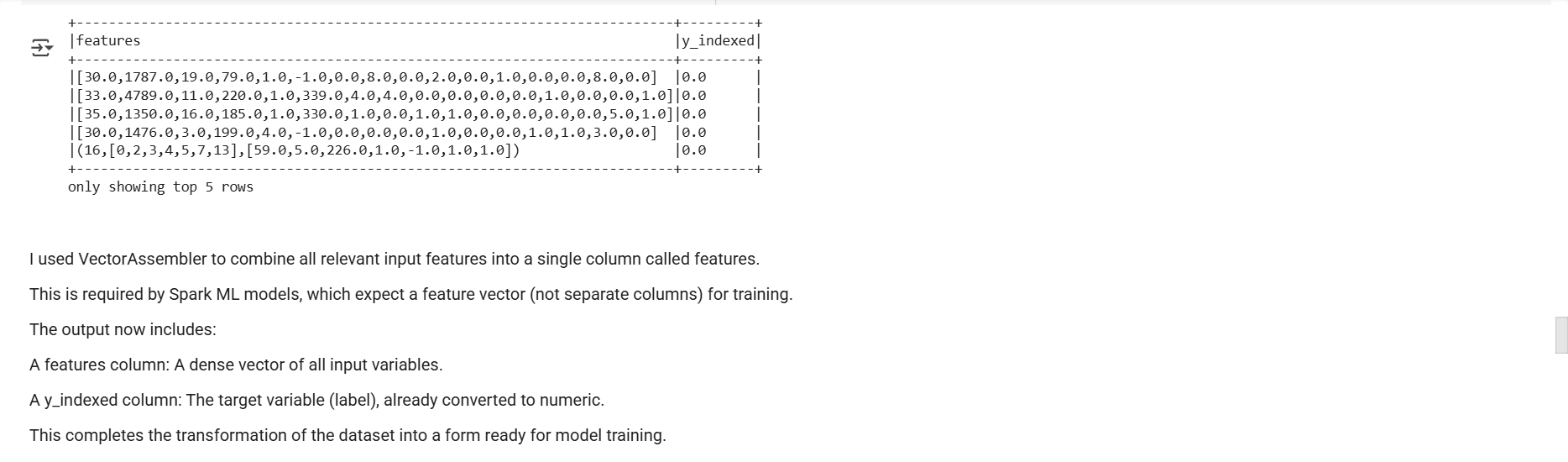
1. **Data Preprocessing**:
   * Handle missing values in the dataset.



* + Handle Outliers in the dataset

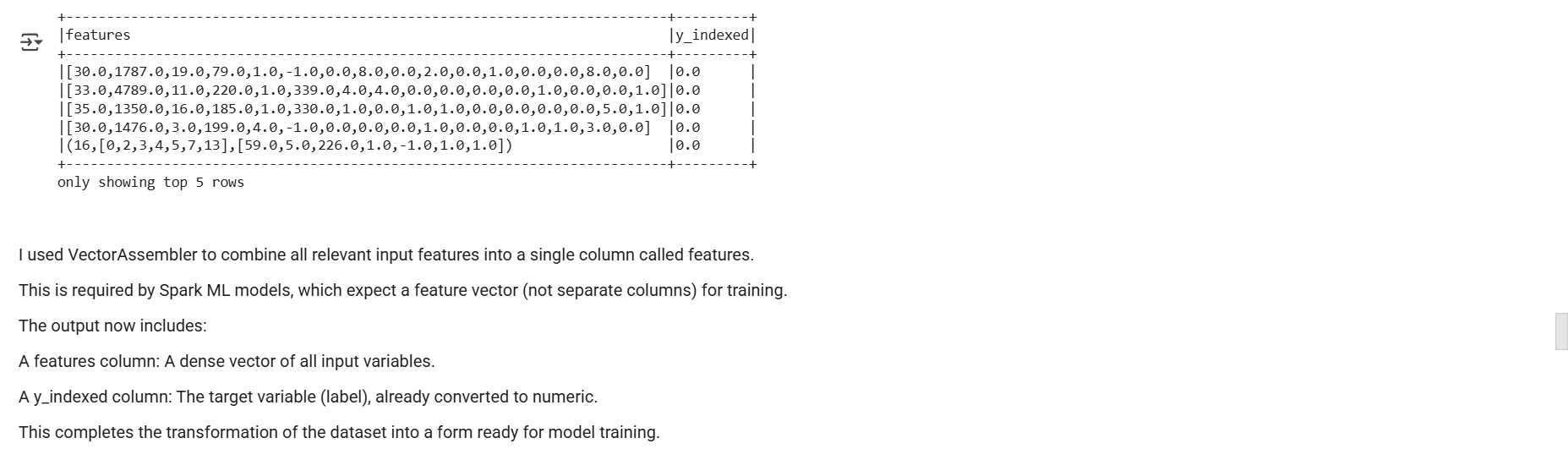
1. 
   * Convert categorical variables to numerical format using StringIndexer or OneHotEncoder.



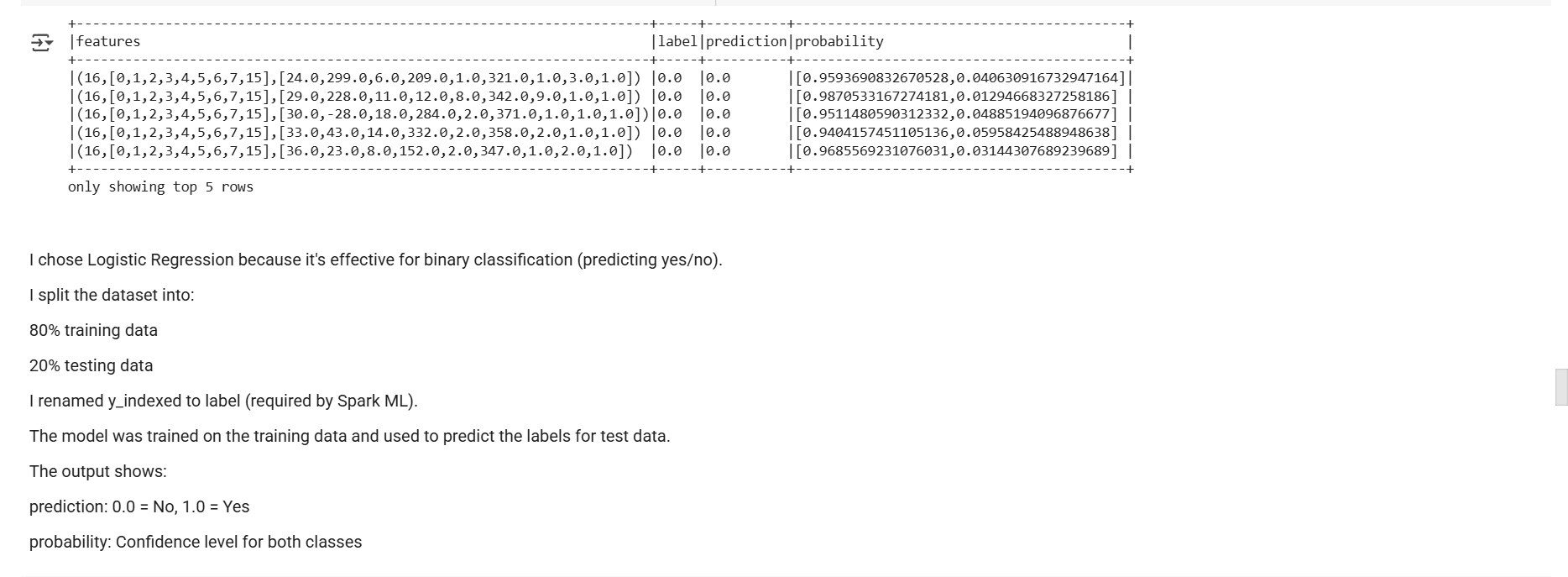
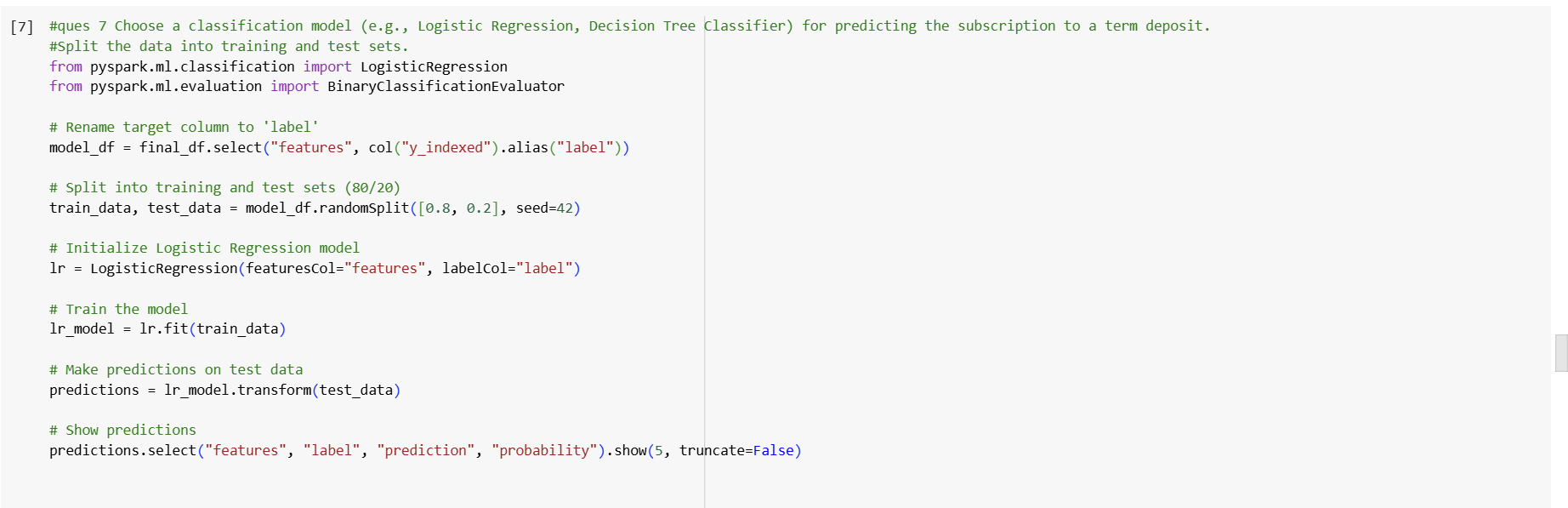


1. **Feature Engineering and Data Transformation**:
   * Create a feature vector using VectorAssembler by combining all feature columns.



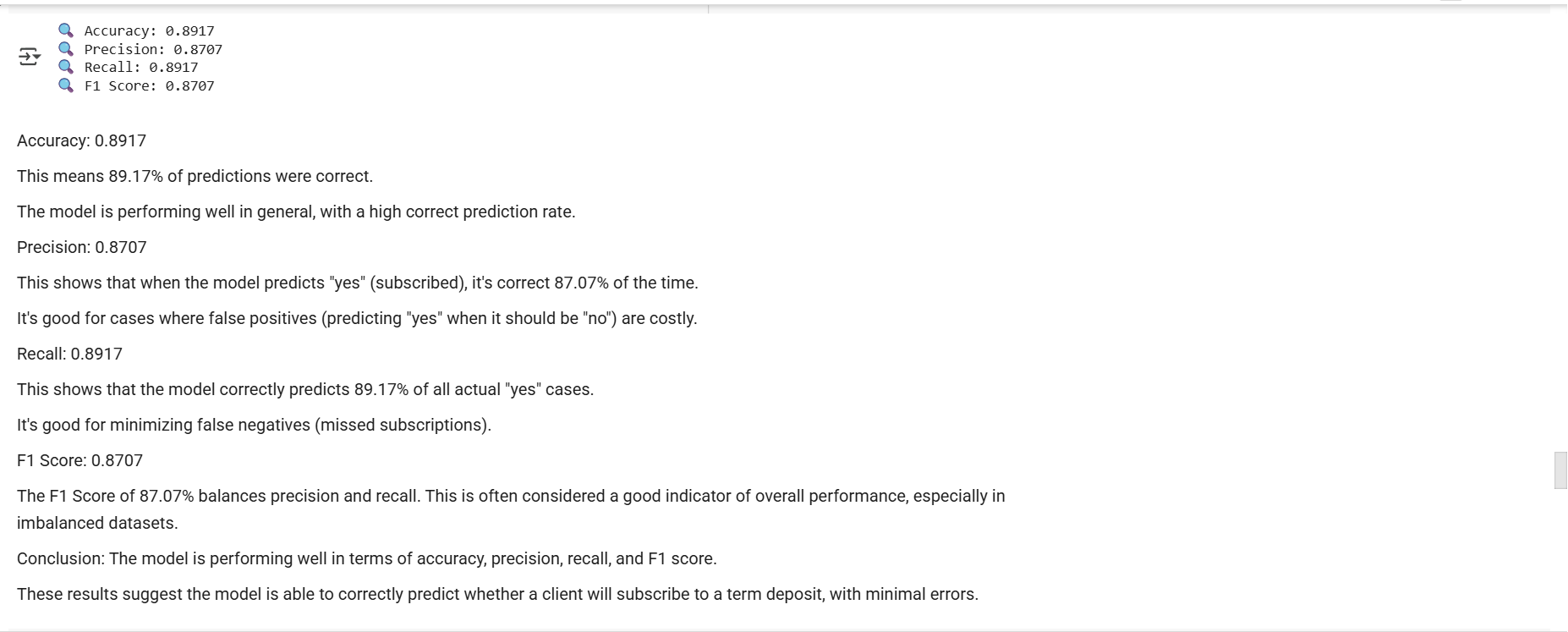


1. **Model Training and Selection**:
   * Choose a classification model (e.g., Logistic Regression, Decision Tree Classifier) for predicting the subscription to a term deposit.
   * Split the data into training and test sets.
   * Train the model on the training dataset.



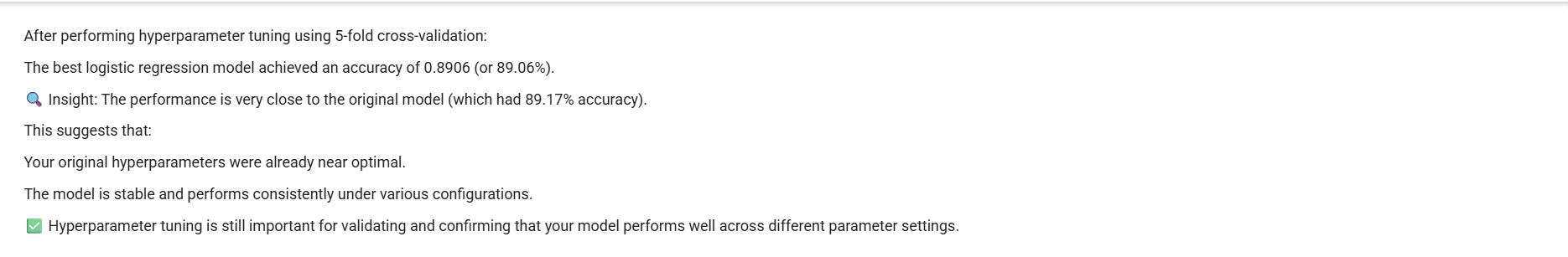
1. **Model Evaluation**:
   * Evaluate the model on the test dataset using appropriate metrics.

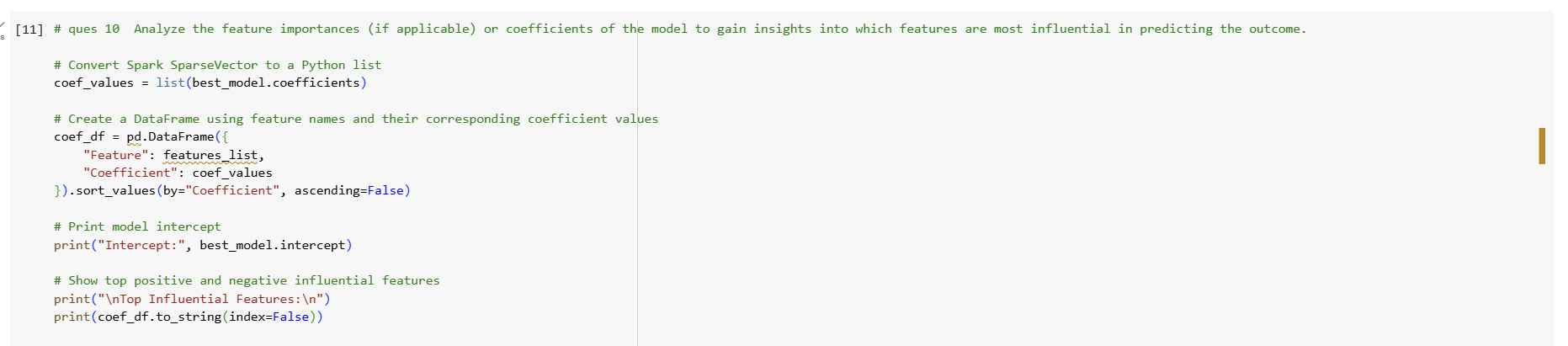
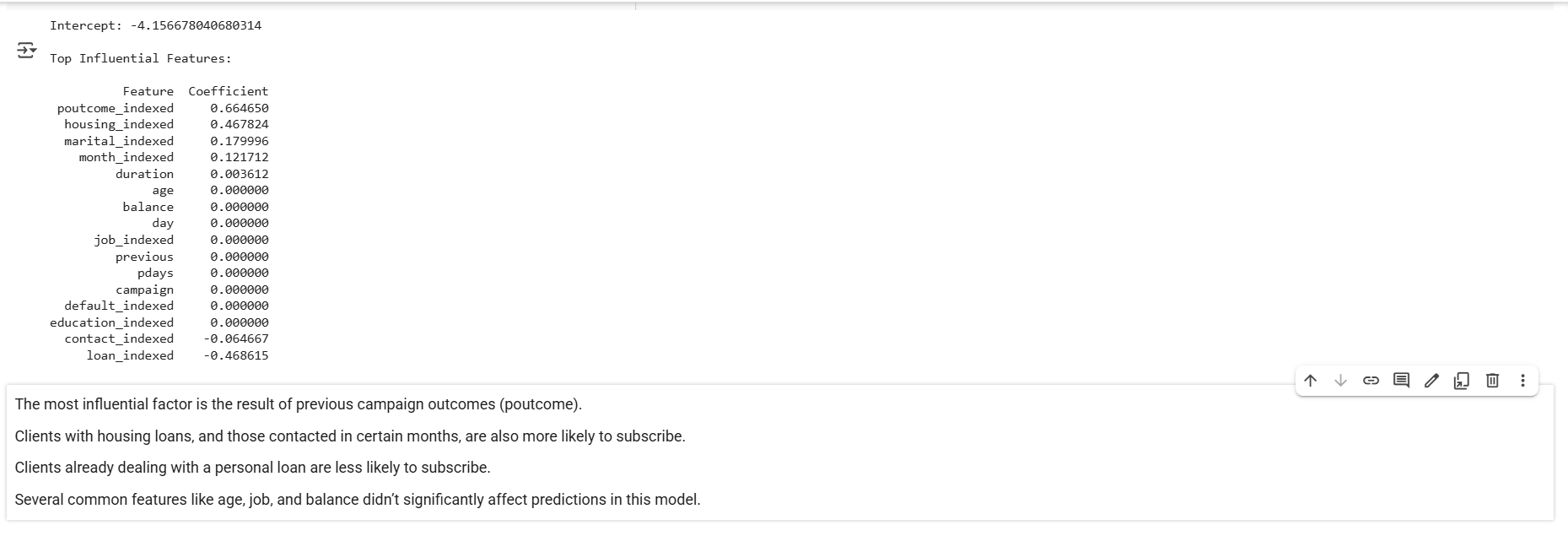




1. **Hyperparameter Tuning**:
   * Perform hyperparameter tuning (using **ParamGridBuilder** and **CrossValidator**.)





1. **Advanced Analysis:**
   * Analyze the feature importances (if applicable) or coefficients of the model to gain insights into which features are most influential in predicting the outcome.
   * 
   * 

**Submission Guidelines:**

* You have the flexibility to perform the Spark data analysis either in Google Colab or in your local environment. Choose the option that is most convenient for you.
* If you choose to work in a local environment, please ensure you take screenshots of both the code and the output for each question. Make a copy of this doc file and include the screenshots in your copied document.
* If you opt for Google Colab, ensure to include the Colab file directly as part of your submission.
* Once your project is complete and your documentation is ready, please upload the document along with any other required files to the submission dashboard.

Colab link - [Copy of Predictive Modeling for Banking Trends.ipynb](https://colab.research.google.com/drive/1SWS1HwdBHLKUIDFcKiKBFevHmPWCwfgZ?usp=sharing)