Phase 3

(Artificial intelligence)

**Sentiment Analysis for Marketing**

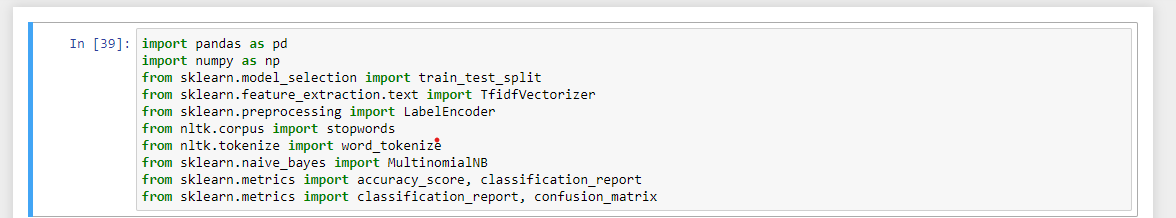
**Loading and preprocessing** a dataset is a Very first step in data analysis, machine learning and artificial intelligence. This process involves several stages, including data cleaning, data integration, data transformation, and data reduction. Let's discuss each of these stages in more detail:

**1. Data Loading:**

- Data loading is the initial step where you acquire the dataset you intend to work with. This can involve importing data from various sources such as CSV files, Excel spreadsheets, databases, web APIs, or other data storage formats.

- You may use libraries and tools like Pandas in Python or read functions in R to load and read the data into a data structure that can be manipulated and analyzed.

**Program:**

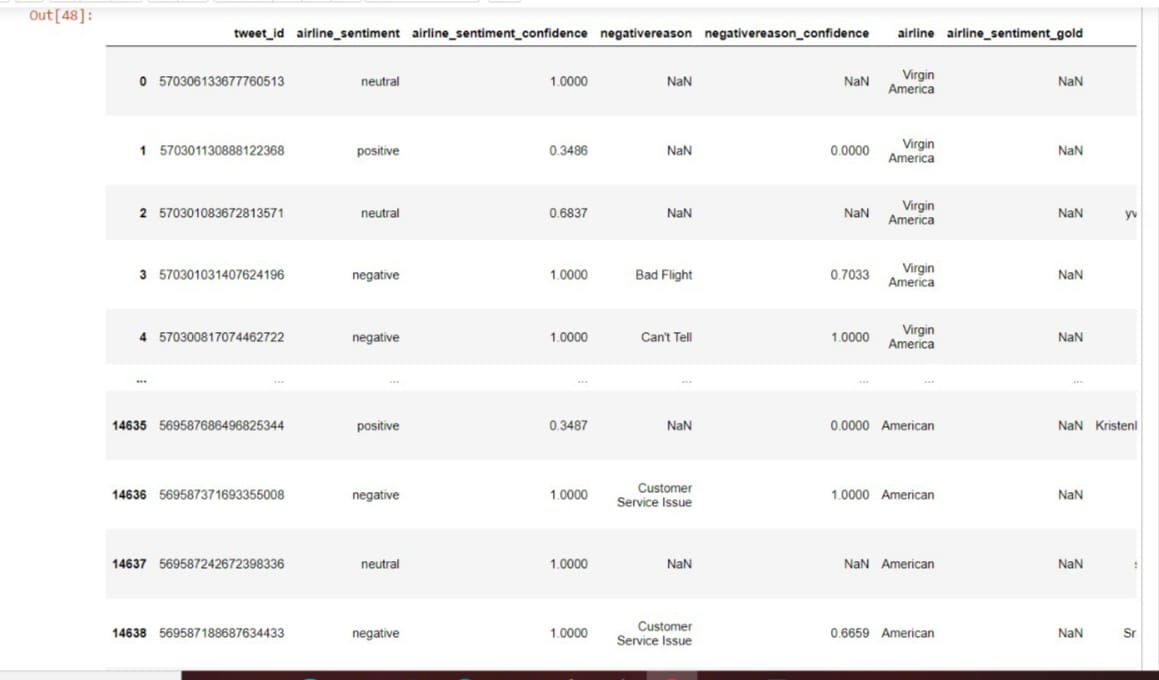


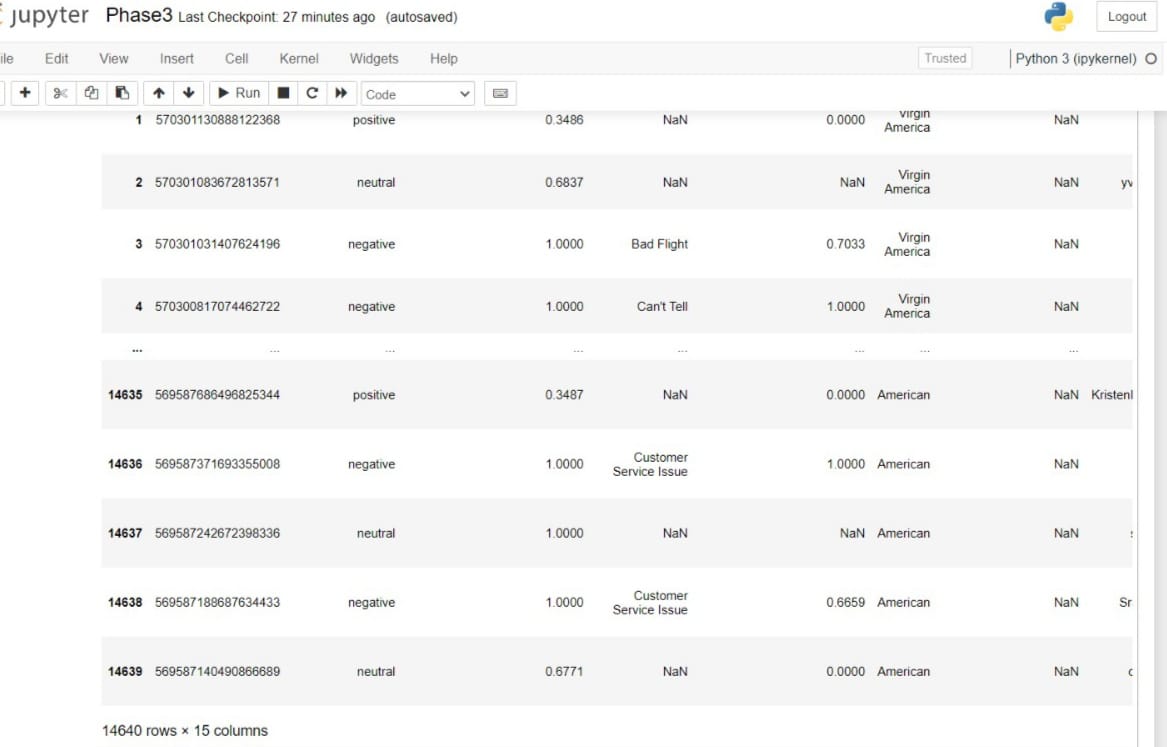
**Importing Required Packages**



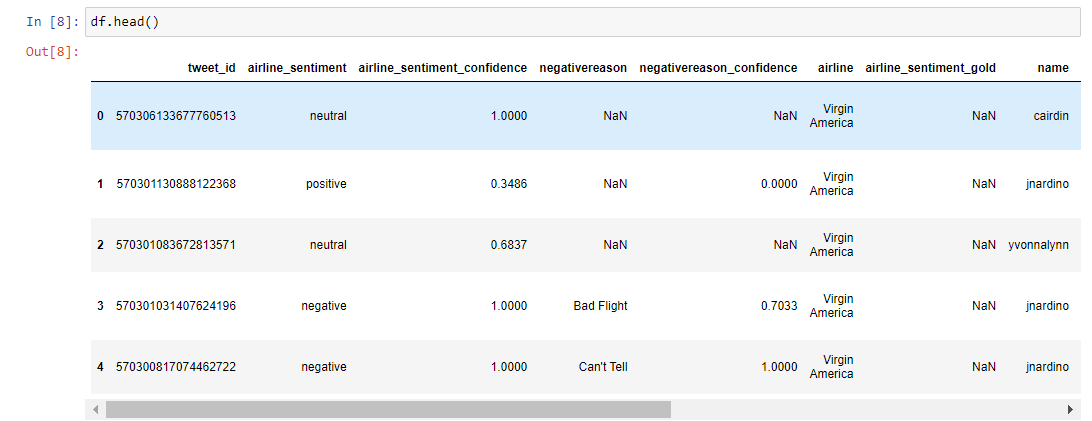
**Reading the CSV Datasets and printing it**

**output:**

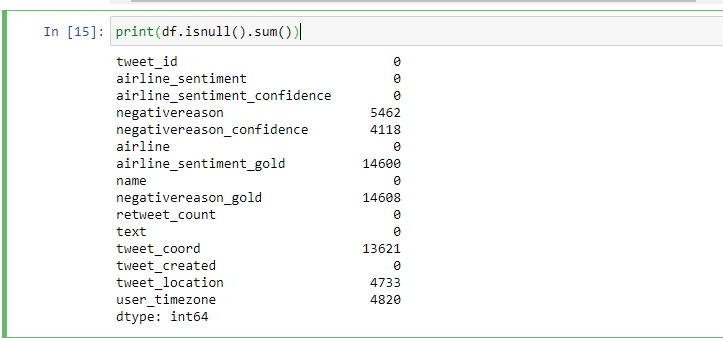




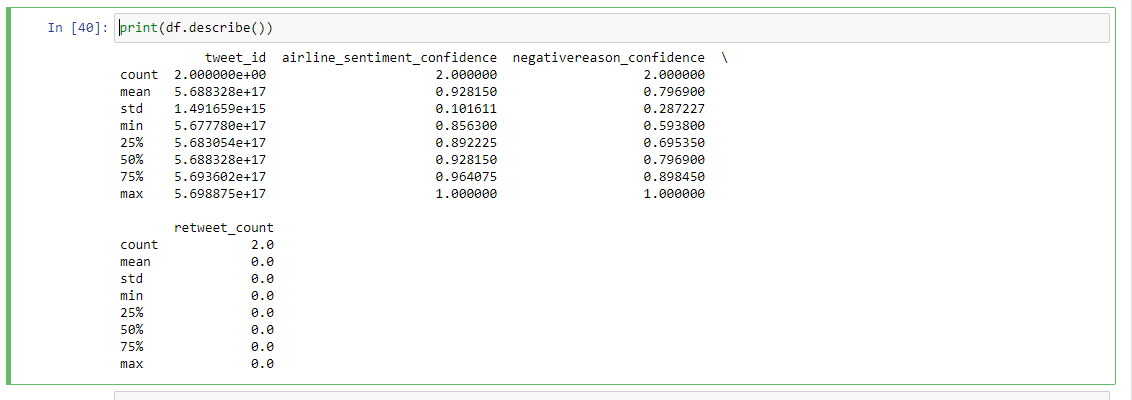
**Printing the output of the Dataset**



**Just printing head of the dataset**

****

**Checking Null values in dataset**

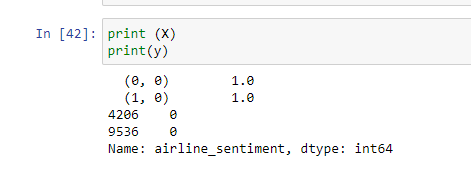
****

**Describing dataset**

**Data Preprocessing**

****

**Removing missing data, converting text to lowercase, Tokenization using TF-IDF, Label encoding for sentiment labels**

****

**Printing the X and y**

**1. Data Cleaning:**

- Data cleaning involves identifying and handling issues with the dataset, such as missing values, duplicates, outliers, and inconsistencies.

- Common data cleaning tasks include:

- Handling missing data by imputation (replacing missing values with estimates) or removal.

- Identifying and dealing with duplicate records.

- Outlier detection and handling (e.g., removing or transforming outliers).

- Standardizing or normalizing data to ensure consistency (e.g., converting categorical data to numerical format).

- Correcting inconsistent or erroneous data entries.

**2. Data Integration:**

- Data integration is the process of combining data from multiple sources into a unified dataset for analysis.

- You might need to merge, join, or concatenate datasets, especially when working with data from diverse sources.

- Data integration may also involve resolving schema conflicts and data format discrepancies.

**3. Data Transformation:**

- Data transformation is about altering the format or structure of the data to make it more suitable for analysis or modeling.

- Common data transformation tasks include:

- Feature engineering: Creating new features from existing ones to capture important information.

- Encoding categorical variables: Converting categorical data (e.g., text labels) into numerical representations using techniques like one-hot encoding or label encoding.

- Scaling or standardizing features: Bringing different features to a common scale to prevent certain features from dominating the analysis.

- Reducing dimensionality: Reducing the number of features through techniques like Principal Component Analysis (PCA) or feature selection.

**4.Data Reduction:**

- Data reduction involves reducing the volume of data while retaining as much relevant information as possible.

- Techniques for data reduction include:

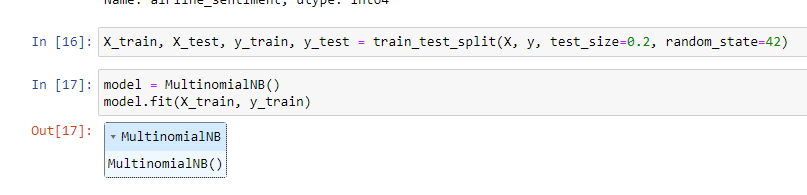
- Aggregation: Summarizing data by grouping and aggregating values.

- Sampling: Using a subset of the data for analysis, especially when dealing with large datasets.

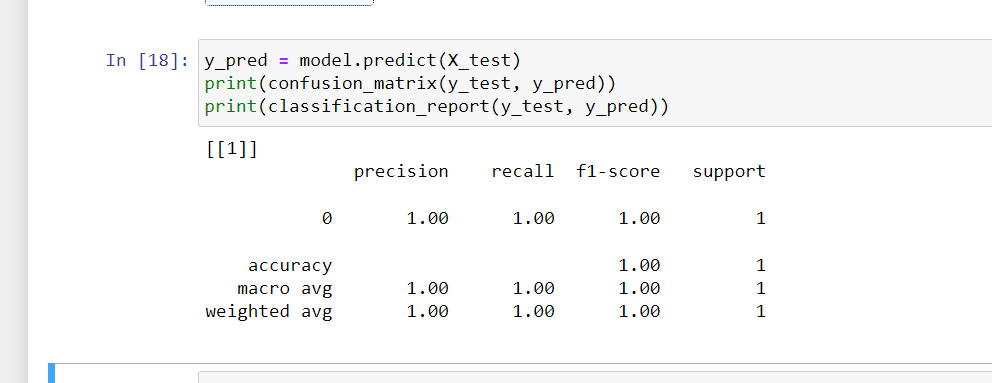
- Dimensionality reduction: Reducing the number of features while preserving as much variance as possible (e.g., using PCA).

- Feature selection: Selecting a subset of the most relevant features for analysis or modeling.

The order in which you perform these steps may vary depending on the specific dataset and the goals of your analysis or machine learning project. The ultimate aim is to prepare the data in a clean, consistent, and informative form that is suitable for further analysis or modeling. Data preprocessing significantly impacts the quality and effectiveness of your data-driven work.

****

**Training and testing**



**Model Prediction**