**Lab – 26**

**Student Name: Navneet P**

**Student ID: AF0411619**

***Topic: Pandas IO and Data Cleaning***

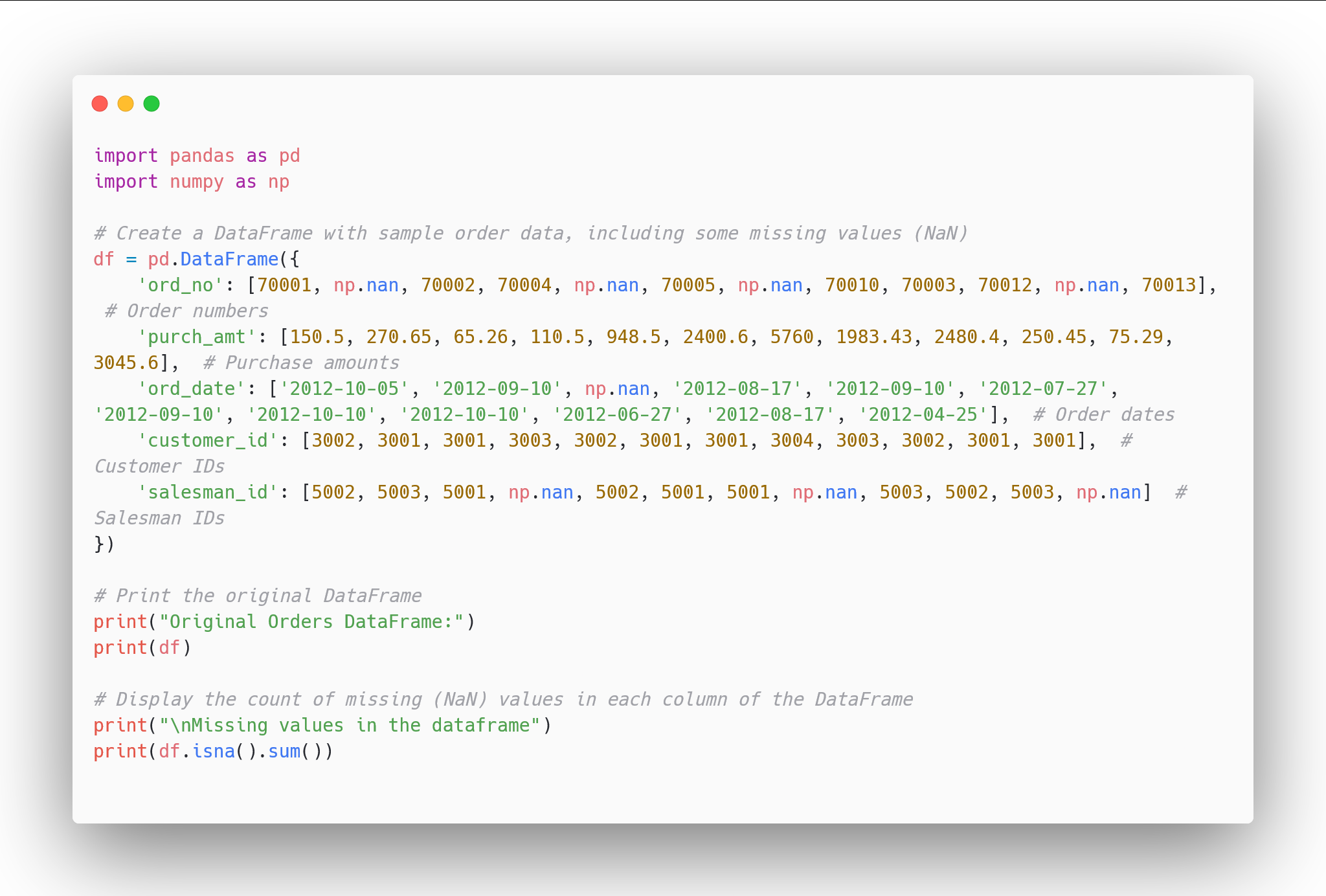
**Data cleaning:** Our data often comes from multiple resources and is not clean. It may contain missing values, duplicates, wrong or undesired formats, etc. Therefore, it is necessary to prepare your data before it is fed to your model. This preparation of the data by identifying and resolving the potential errors, inaccuracies etc.

Functions used in this assignment:

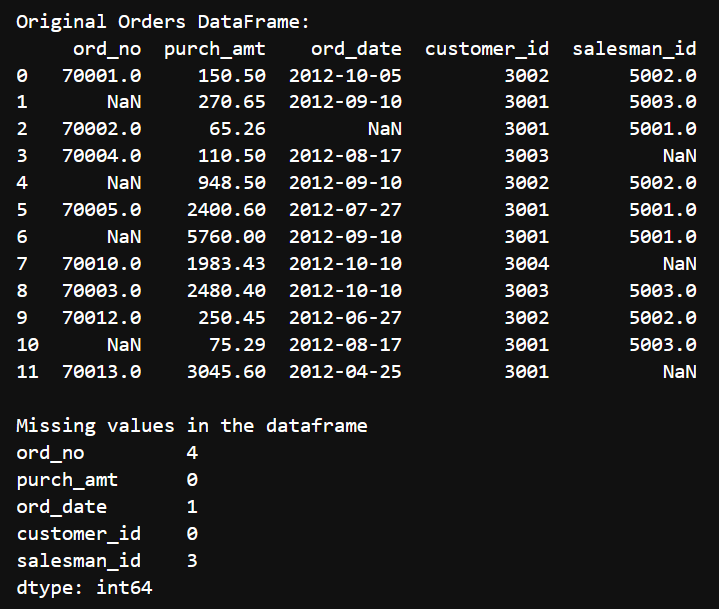
* **numpy.nan**: numpy.nan represents "Not a Number" (NaN) in NumPy. It is used to represent missing or undefined numerical data in arrays. Operations involving nan generally result in nan, making it useful for handling missing data.
* **isna()**: isna() is a Pandas function that detects missing values (NaN) in a DataFrame or Series. It returns a DataFrame/Series of the same shape, with True for missing values and False for non-missing values. It helps in identifying where data is missing.
* **dropna()**: dropna() is a Pandas function that removes rows or columns with missing values (NaN) from a DataFrame or Series. By default, it drops any row that contains a NaN, but it can be customized to target specific columns or rows with missing data. It is useful for cleaning datasets before analysis.

Q1. Count the number of missing values in each column and display them.

Solution:



Output:

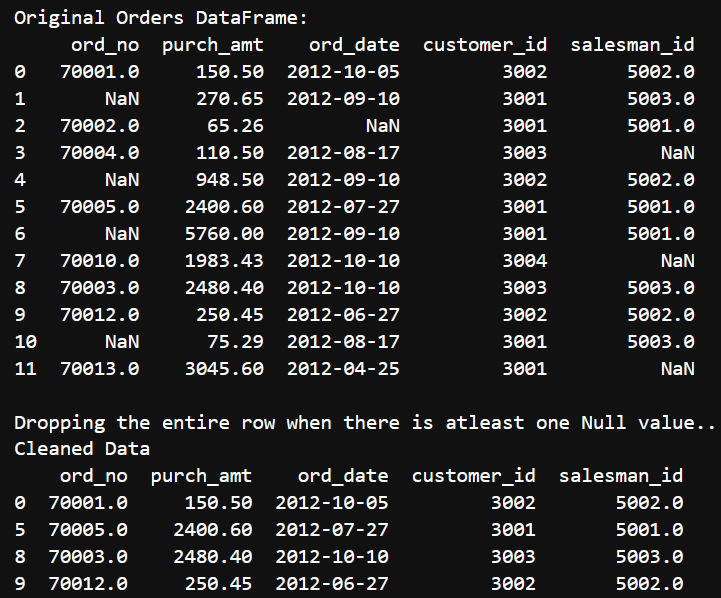


Q2. Drop those rows from the table which has at least one empty value.

Solution:



Output:

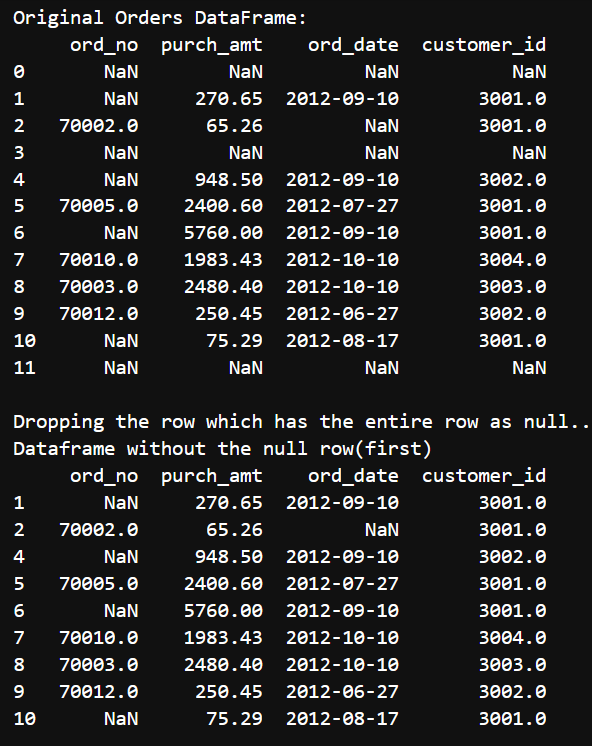


Q3. Drop the rows where all elements of a certain rows are missing.

Solution:



Output:



Q4: Drop those rows which have null values based on specific columns.

Solution:



Output:

