Ex-01\_DS\_Data\_Cleansing AIM

To read the given data and perform data cleaning and save the cleaned data to a file.

# Explanation

Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incompleted , irrelevant , duplicated or improperly formatted. Data cleaning is not simply about erasing data ,but rather finding a way to maximize datasets accuracy without necessarily deleting the information.

# ALGORITHM

## STEP 1

Read the given Data

STEP 2

Get the information about the data

STEP 3

Remove the null values from the data

STEP 4

Save the Clean data to the file

# CODE

Developed by: M.SARAVANAN

Reg.no: 212219223002

## Data\_set.csv

import numpy as np import pandas as pd import io from google.colab import files uploaded = files.upload() dd = pd.read\_csv(io.BytesIO(uploaded['Data\_set.csv'])) print(dd) dd.tail() dd.info() dd.isnull() dd.isnull().sum()

#mode dd['show\_name'] = dd['show\_name'].fillna(dd['aired\_on'].mode()[0]) dd['aired\_on'] = dd['aired\_on'].fillna(dd['aired\_on'].mode()[0])

dd['original\_network'] = dd['original\_network'].fillna(dd['aired\_on'].mode()[0]) dd.head() #mean dd['rating'] = dd['rating'].fillna(dd['rating'].mean()) dd['current\_overall\_rank'] = dd['current\_overall\_rank'].fillna(dd['current\_overall\_rank'].mean()) dd.head() #median dd['watchers'] = dd['watchers'].fillna(dd['watchers'].median()) dd.head() dd.info() dd.isnull().sum()

## Loan\_data.csv

import numpy as np import pandas as pd import io

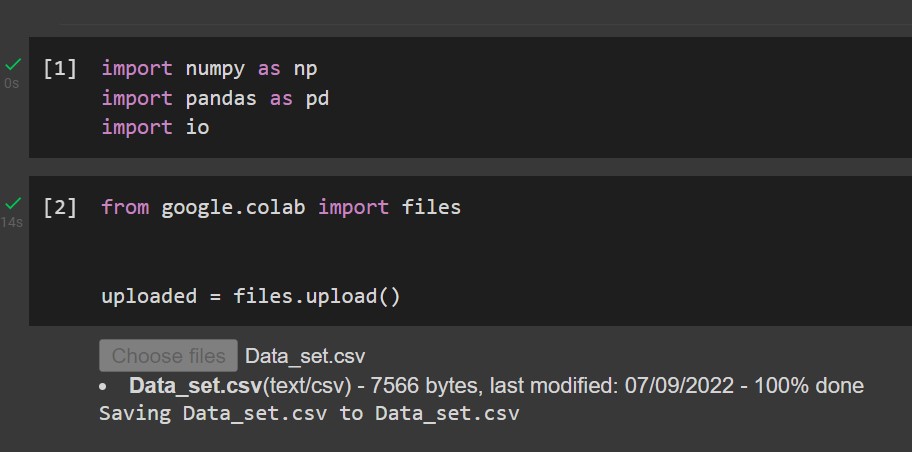
from google.colab import files uploaded = files.upload() de = pd.read\_csv(io.BytesIO(uploaded['Loan\_data.csv'])) print(de) de.tail(10) de.info() de.isnull() de.isnull().sum()

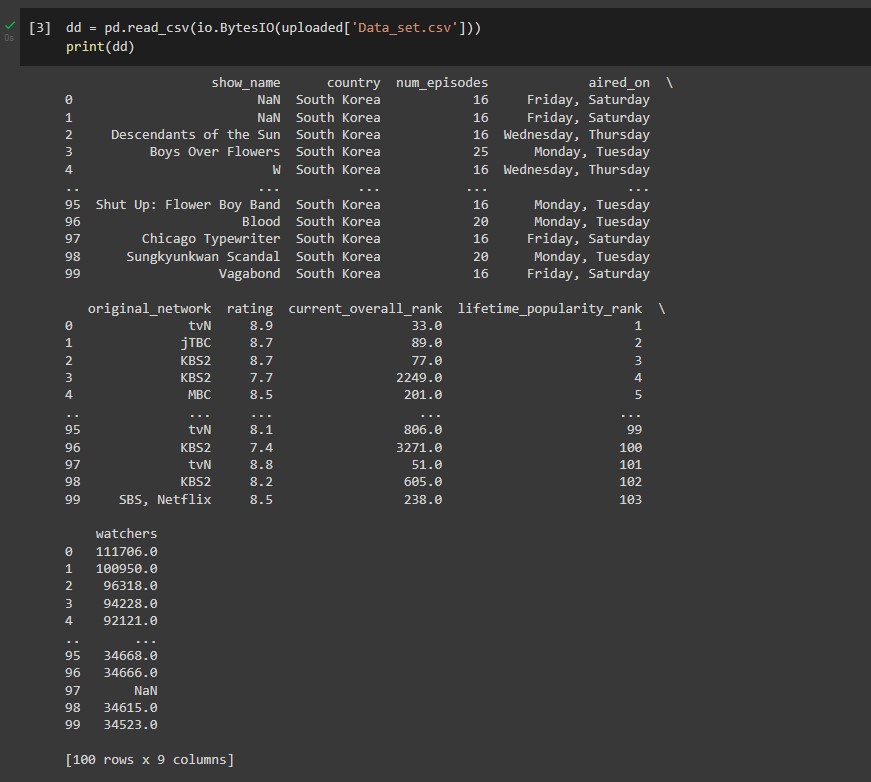
#mode de['Gender'] = de['Gender'].fillna(de['Dependents'].mode()[0]) de['Dependents'] = de['Dependents'].fillna(de['Dependents'].mode()[0]) de['Self\_Employed'] = de['Self\_Employed'].fillna(de['Dependents'].mode()[0]) de.head() #mean

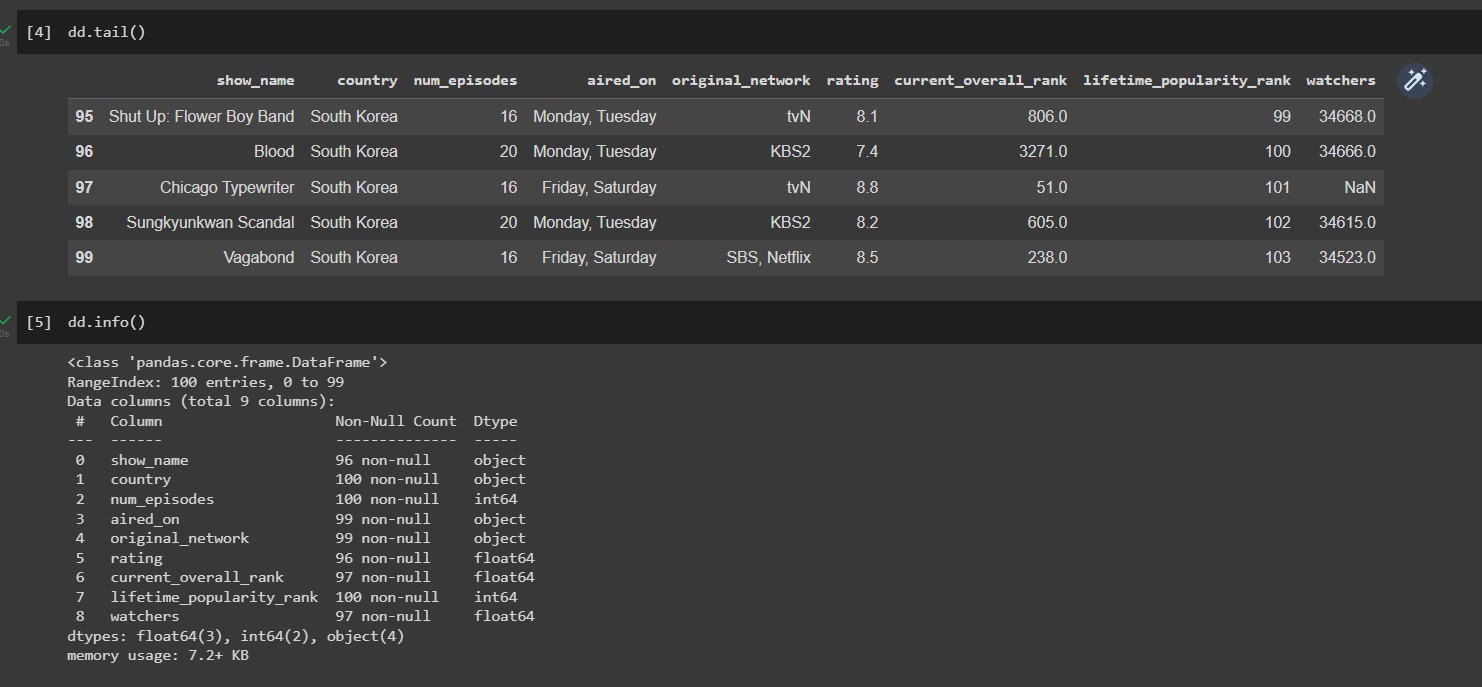
de['LoanAmount'] = de['LoanAmount'].fillna(de['LoanAmount'].mean()) de['Loan\_Amount\_Term'] = de['Loan\_Amount\_Term'].fillna(de['Loan\_Amount\_Term'].mean()) de.tail() #median de['Credit\_History'] = de['Credit\_History'].fillna(de['Credit\_History'].median()) de.head() de.info() de.isnull().sum()

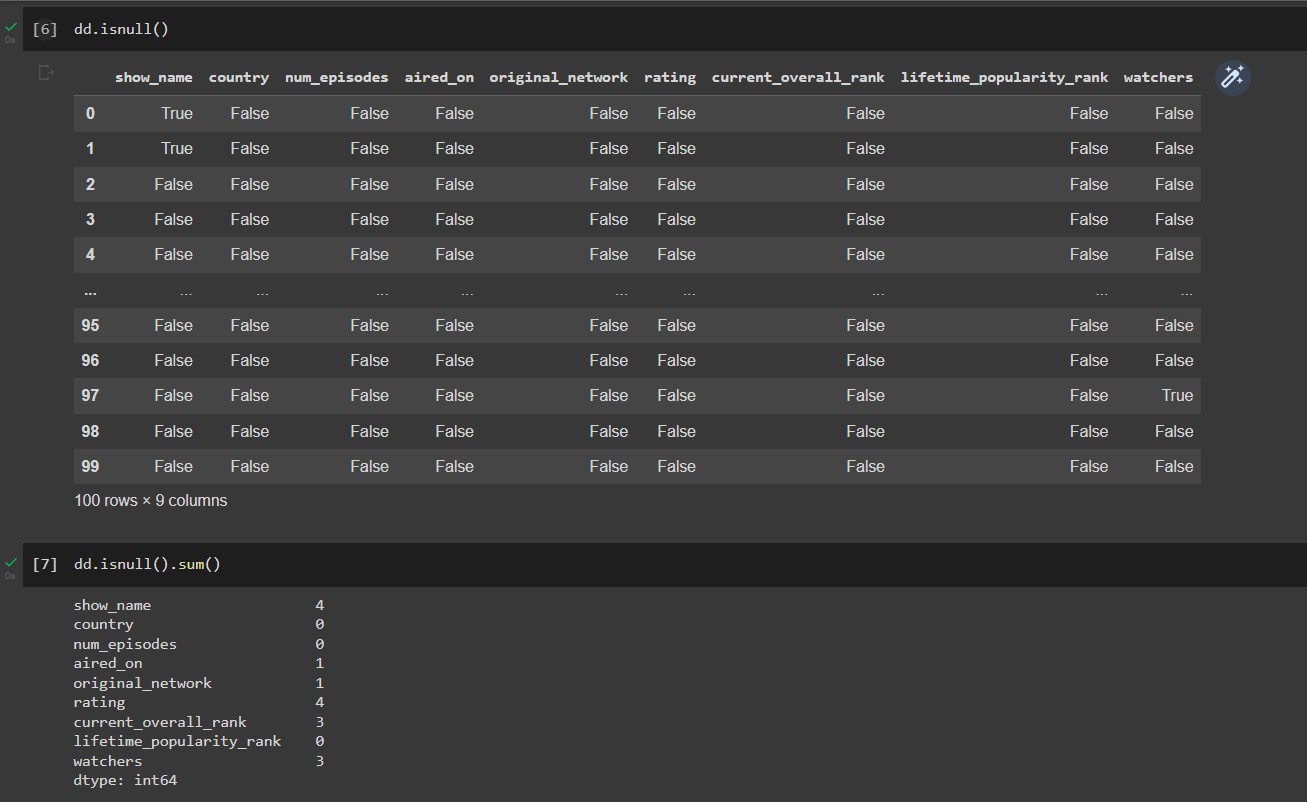
# OUPUT

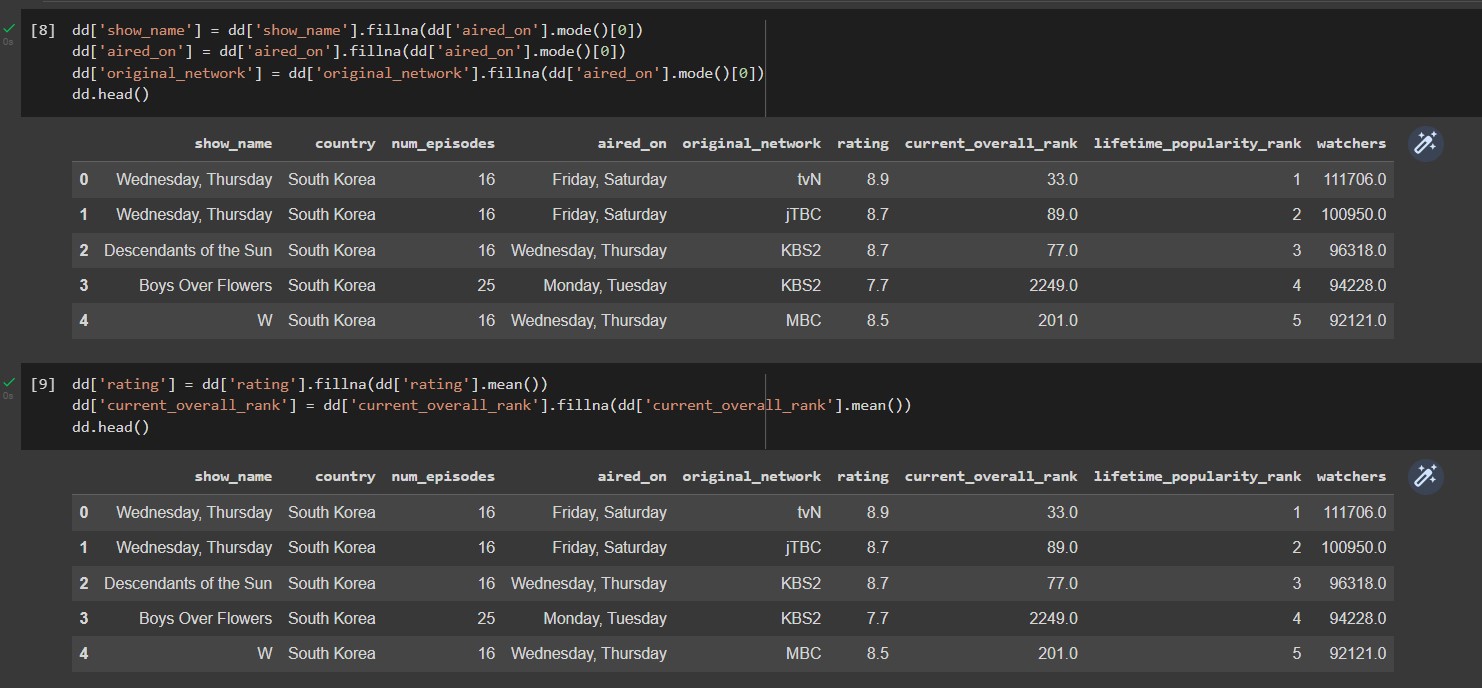
## Data\_set

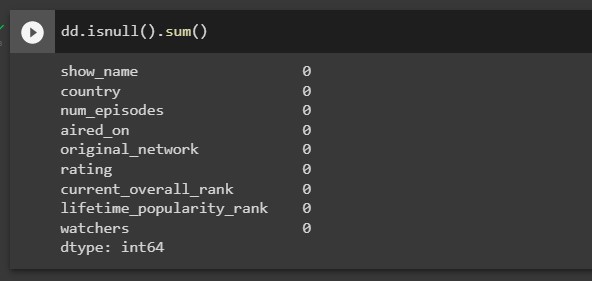
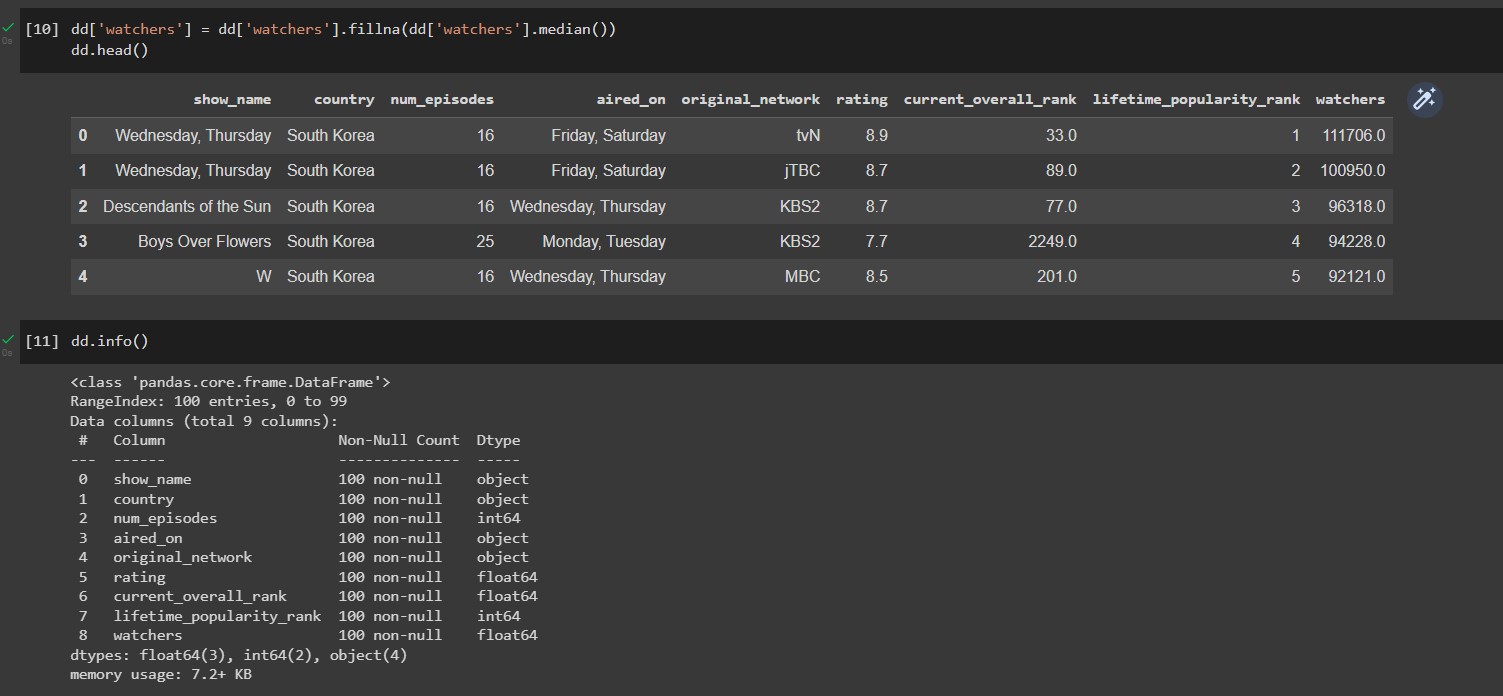




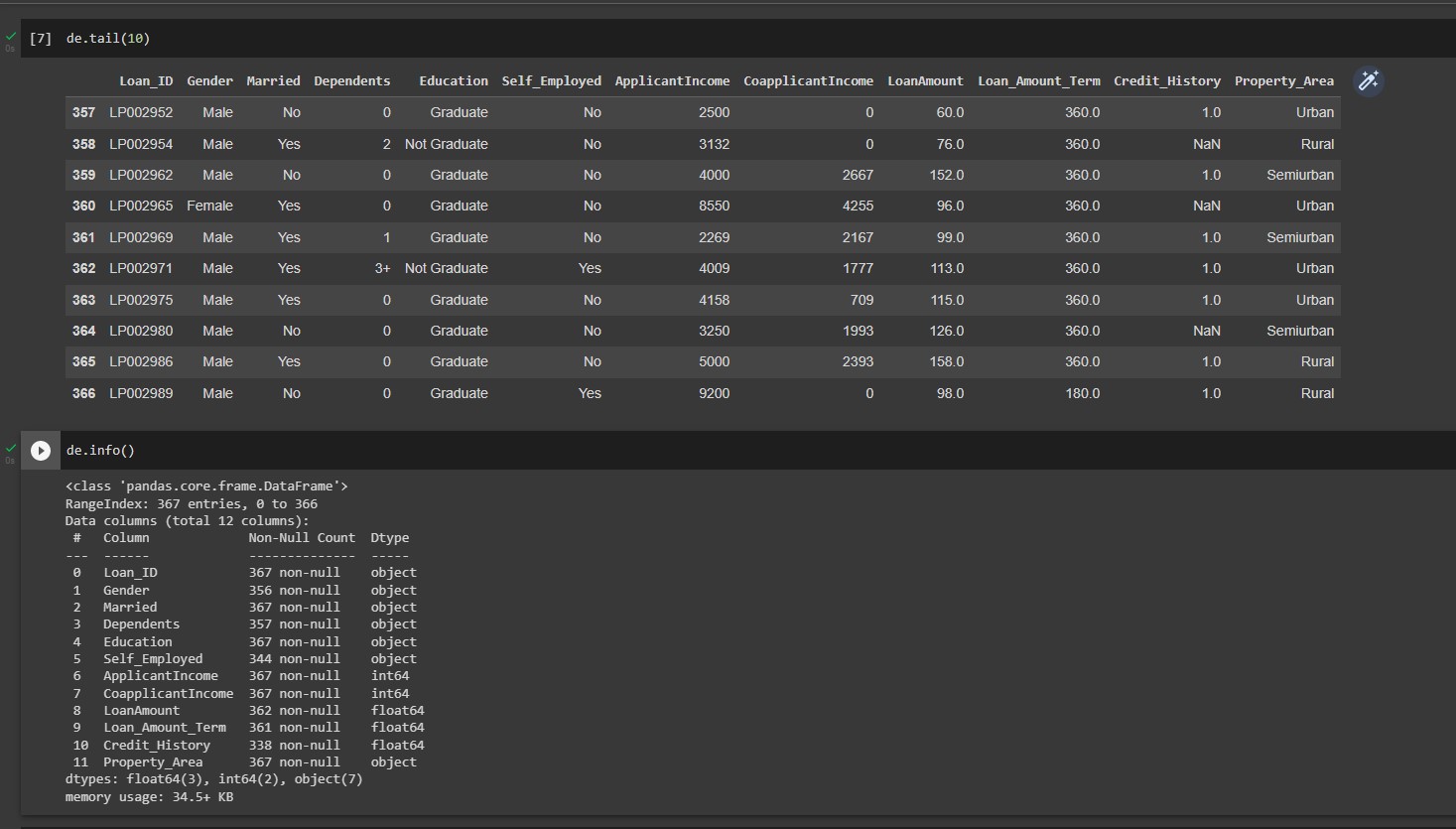
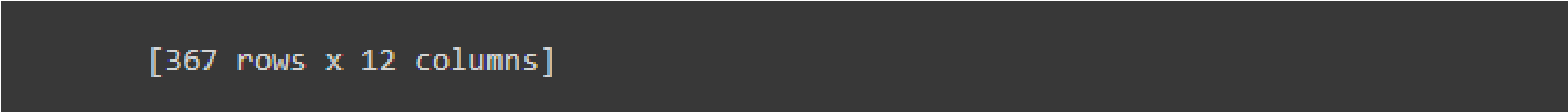
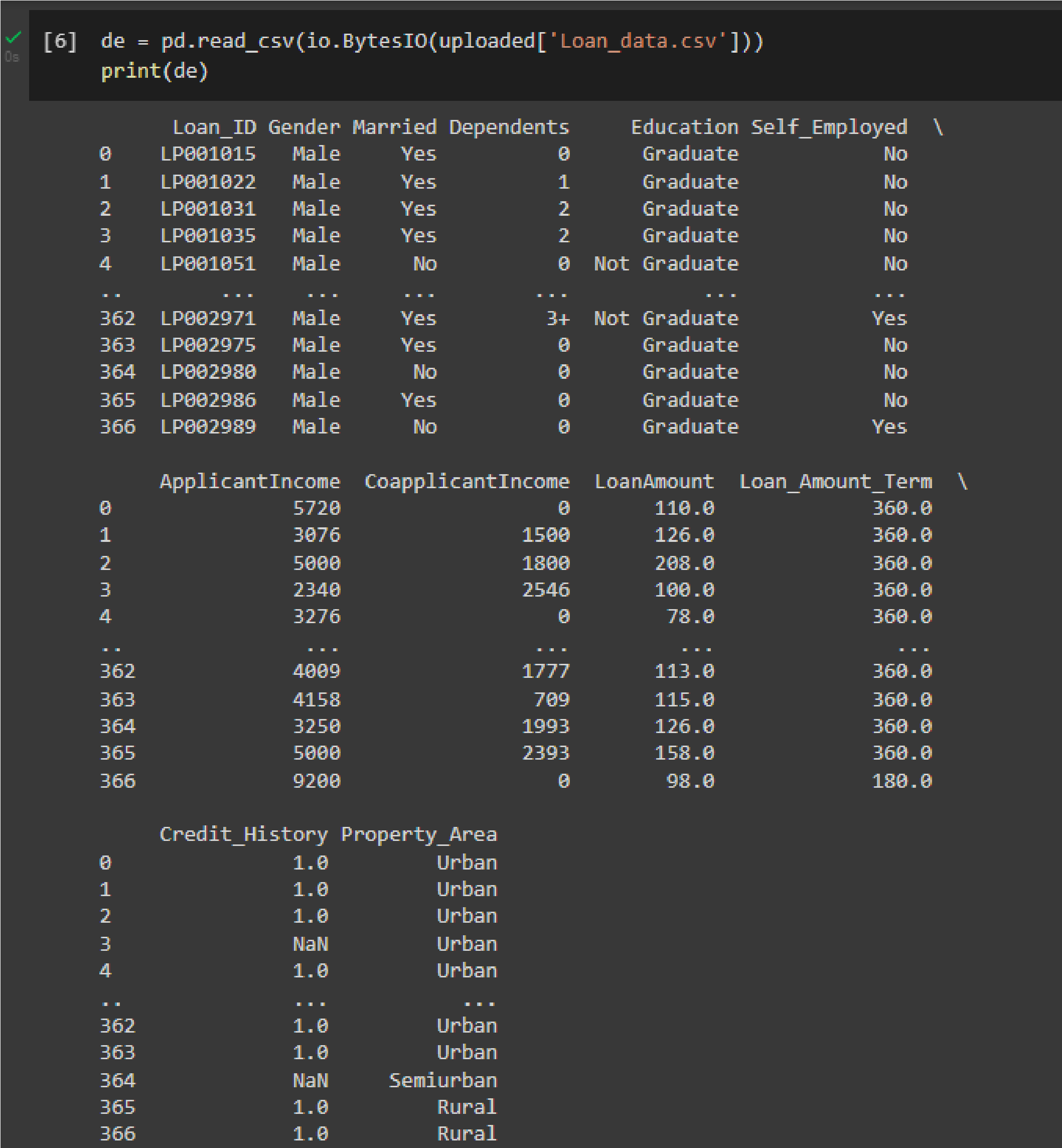
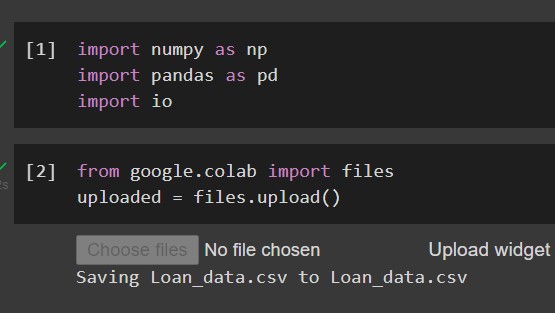


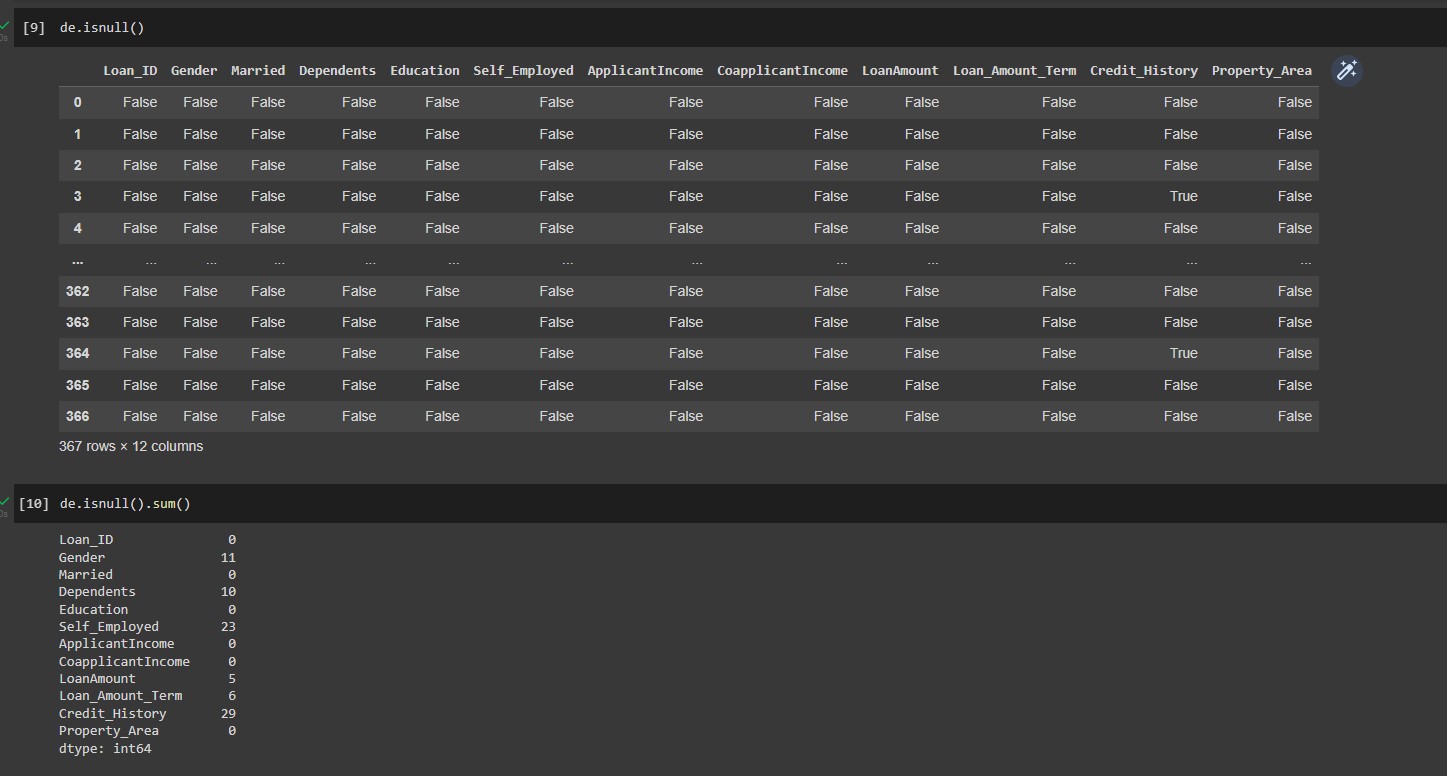


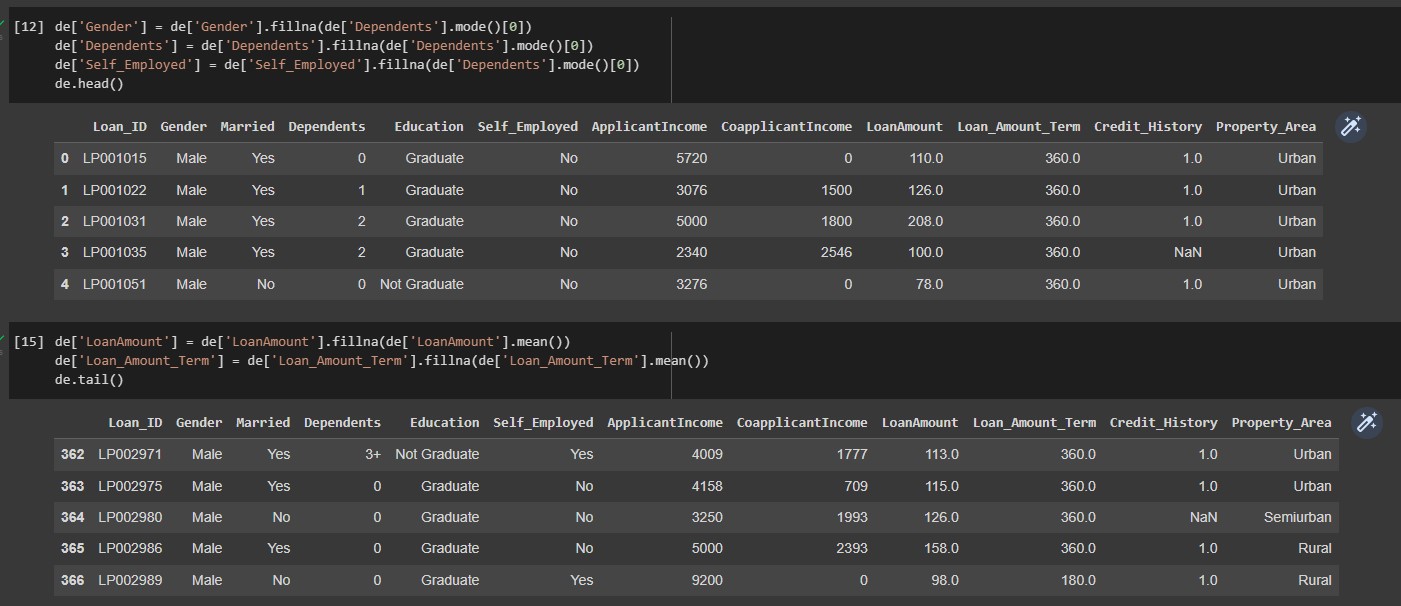


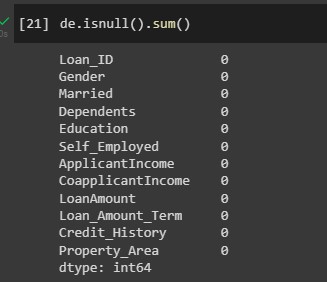
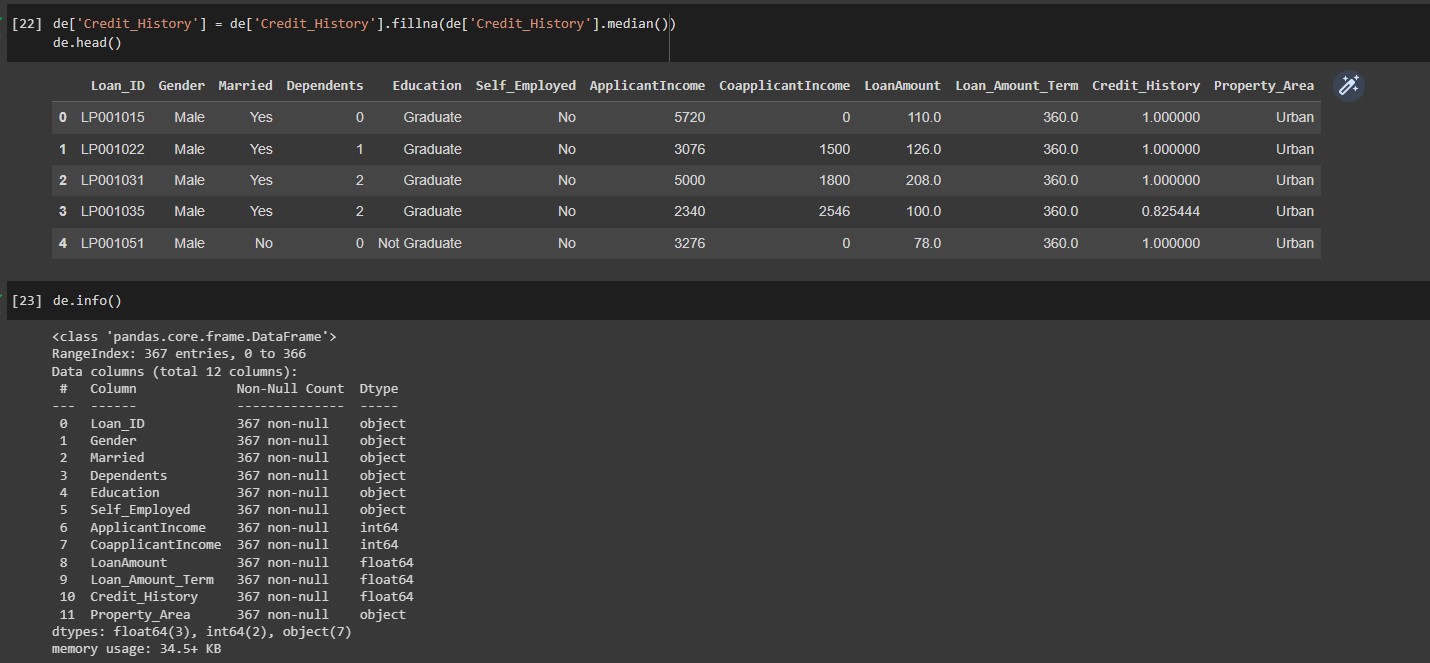


## Loan\_data









R

e

su

l

t

The Dataset is been successfully cleaned and saved into the python file.