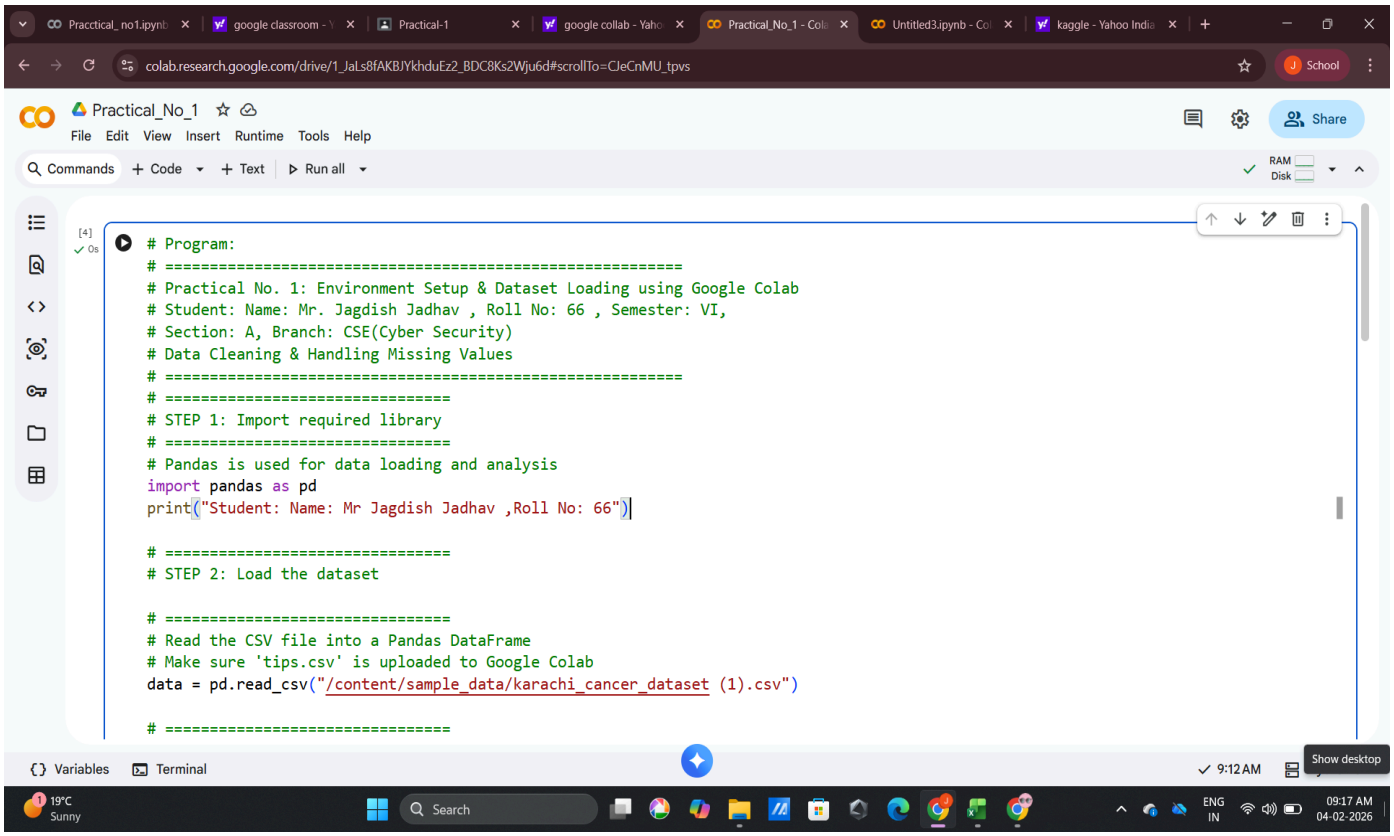
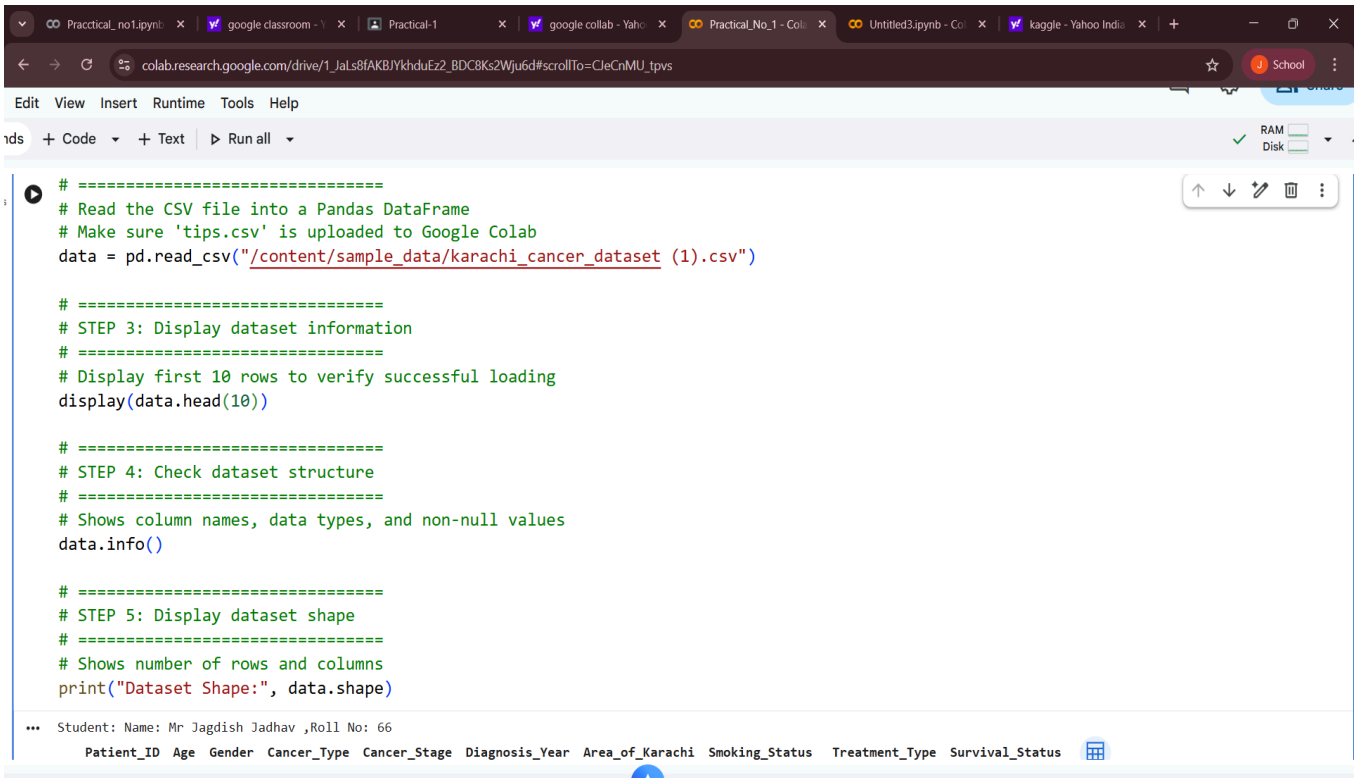


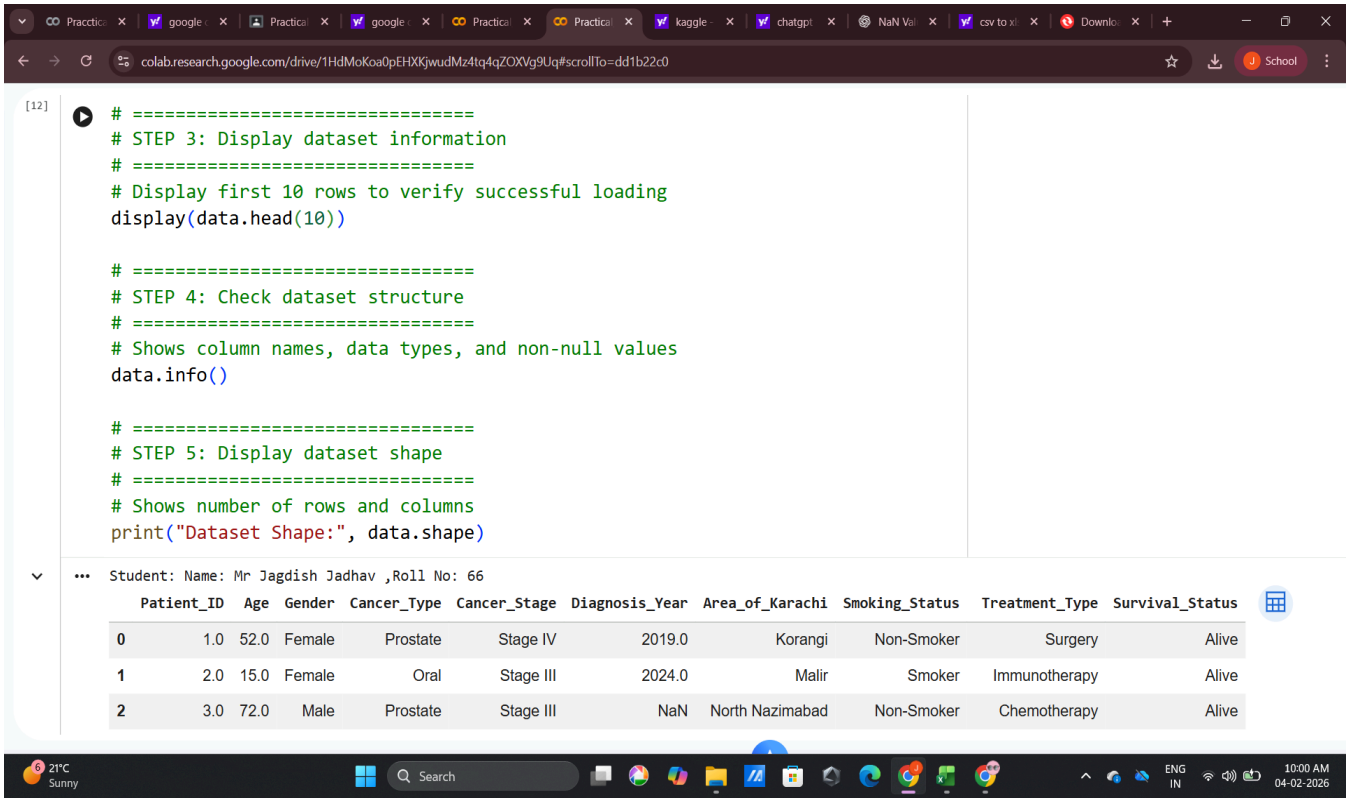
```
# Program:
# ============================================================
# Practical No. 1: Environment Setup & Dataset Loading using Google Colab
# Student: Name: Mr. Jagdish Jadhav , Roll No: 66 , Semester: VI,
# Section: A, Branch: CSE(Cyber Security)
# Data Cleaning & Handling Missing Values
# ============================================================
# ================================
# STEP 1: Import required library
# ================================
# Pandas is used for data loading and analysis
import pandas as pd
print("Student: Name: Mr Jagdish Jadhav ,Roll No: 66")

# ================================
# STEP 2: Load the dataset

# ================================
# Read the CSV file into a Pandas DataFrame
# Make sure 'tips.csv' is uploaded to Google Colab
data = pd.read_csv("/content/sample_data/karachi_cancer_dataset (1).csv")

# ================================
```



```
# ================================
# Read the CSV file into a Pandas DataFrame
# Make sure 'tips.csv' is uploaded to Google Colab
data = pd.read_csv("/content/sample_data/karachi_cancer_dataset (1).csv")

# ================================
# STEP 3: Display dataset information
# ================================
# Display first 10 rows to verify successful loading
display(data.head(10))

# ================================
# STEP 4: Check dataset structure
# ================================
# Shows column names, data types, and non-null values
data.info()

# ================================
# STEP 5: Display dataset shape
# ================================
# Shows number of rows and columns
print("Dataset Shape:", data.shape)
```

```
Student: Name: Mr Jagdish Jadhav ,Roll No: 66
```

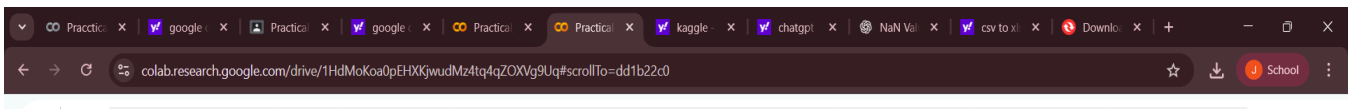
Patient_ID	Age	Gender	Cancer_Type	Cancer_Stage	Diagnosis_Year	Area_of_Karachi	Smoking_Status	Treatment_Type	Survival_Status



```
[12]
# ================================
# STEP 3: Display dataset information
# ================================
# Display first 10 rows to verify successful loading
display(data.head(10))

# ================================
# STEP 4: Check dataset structure
# ================================
# Shows column names, data types, and non-null values
data.info()

# ================================
# STEP 5: Display dataset shape
# ================================
# Shows number of rows and columns
print("Dataset Shape:", data.shape)
```

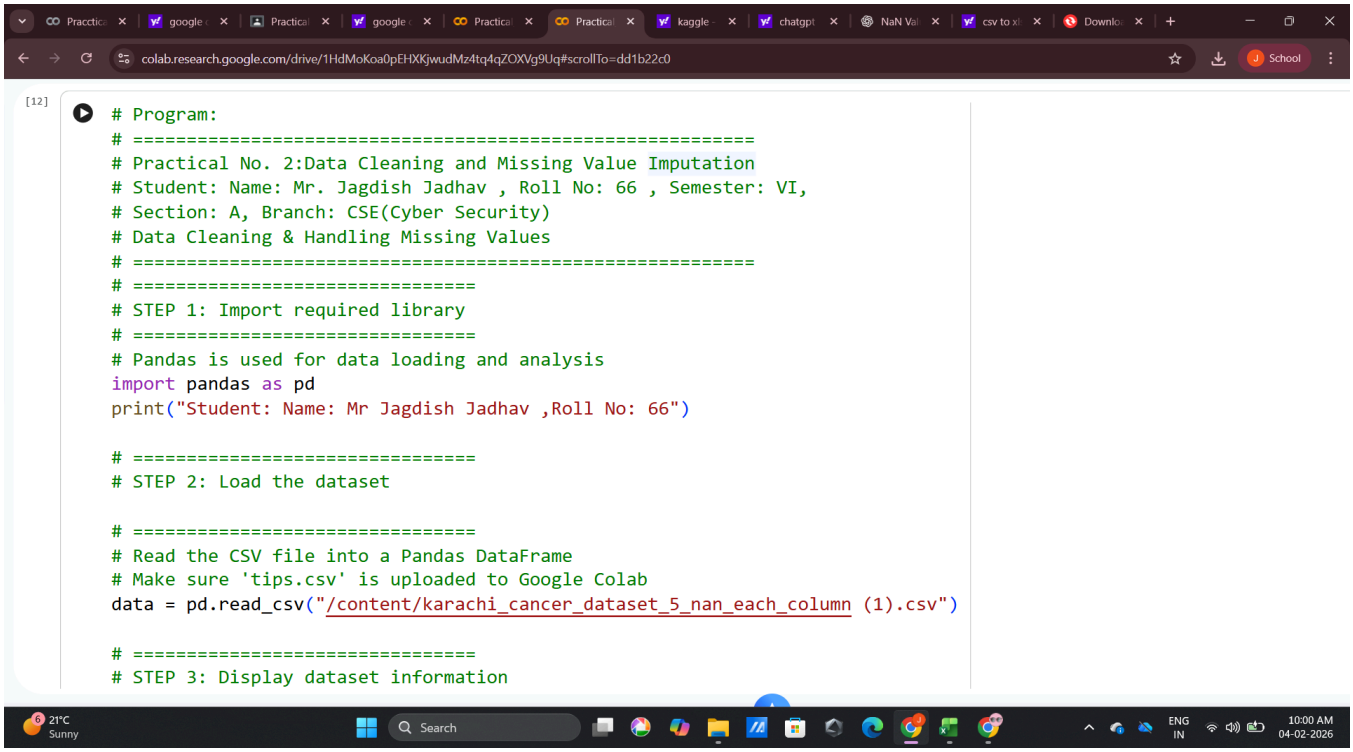
```
Student: Name: Mr Jagdish Jadhav ,Roll No: 66
```

	Patient_ID	Age	Gender	Cancer_Type	Cancer_Stage	Diagnosis_Year	Area_of_Karachi	Smoking_Status	Treatment_Type	Survival_Status
0	1.0	52.0	Female	Prostate	Stage IV	2019.0	Korangi	Non-Smoker	Surgery	Alive
1	2.0	15.0	Female	Oral	Stage III	2024.0	Malir	Smoker	Immunotherapy	Alive
2	3.0	72.0	Male	Prostate	Stage III	NaN	North Nazimabad	Non-Smoker	Chemotherapy	Alive

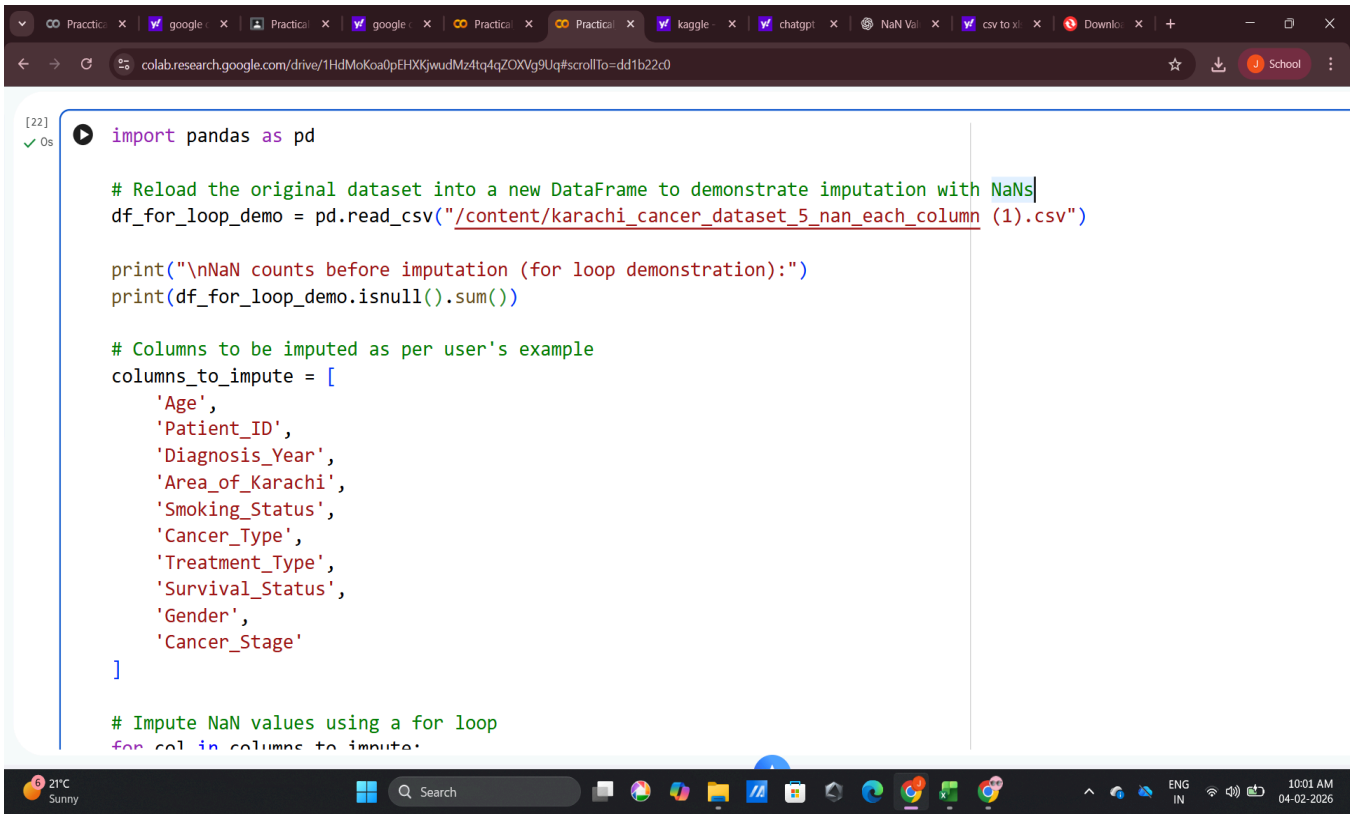


3	4.0	61.0	Female	Prostate	NaN	2020.0	Lyari	Non-Smoker	Radiation Therapy	Deceased
4	5.0	NaN	Male	Leukemia	Stage I	2019.0	Clifton	Non-Smoker	Immunotherapy	Deceased
5	6.0	83.0	Male	Oral	Stage IV	2019.0	NaN	Non-Smoker	NaN	Alive
6	7.0	87.0	Male	Prostate	Stage II	2015.0	Lyari	NaN	Immunotherapy	Alive
7	8.0	75.0	NaN	Liver	Stage I	2016.0	Saddar	Smoker	Chemotherapy	Alive
8	9.0	75.0	Female	Colorectal	Stage III	2022.0	Saddar	Non-Smoker	Surgery	Deceased
9	10.0	88.0	Male	Breast	NaN	2017.0	Malir	NaN	Radiation Therapy	Deceased

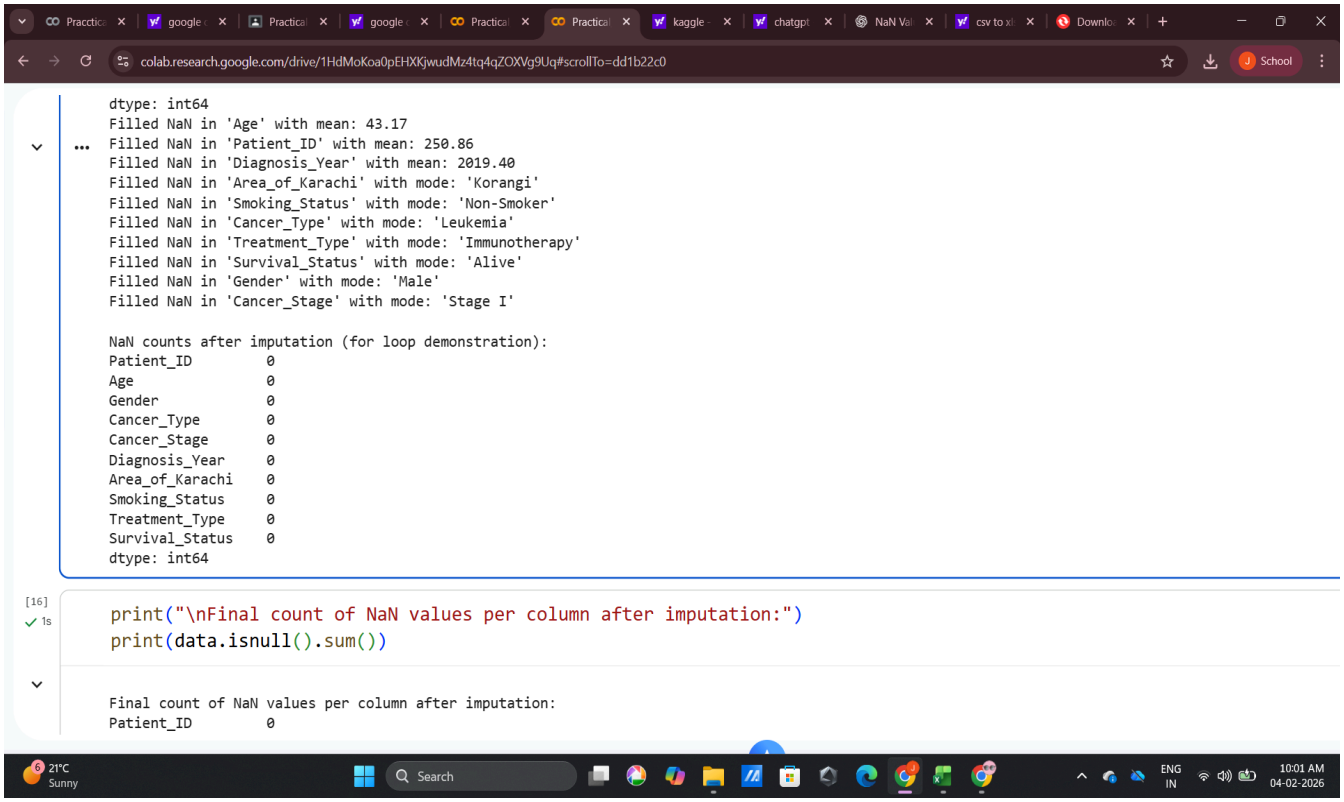
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Patient_ID       495 non-null    float64
 1   Age              492 non-null    float64
 2   Gender           492 non-null    object
 3   Cancer_Type      491 non-null    object
 4   Cancer_Stage     491 non-null    object
 5   Diagnosis_Year   494 non-null    float64
 6   Area_of_Karachi  490 non-null    object
 7   Smoking_Status   488 non-null    object
 8   Treatment_Type   492 non-null    object
 9   Survival_Status  490 non-null    object
dtypes: float64(3), object(7)
memory usage: 39.2+ KB
Dataset Shape: (500, 10)
```

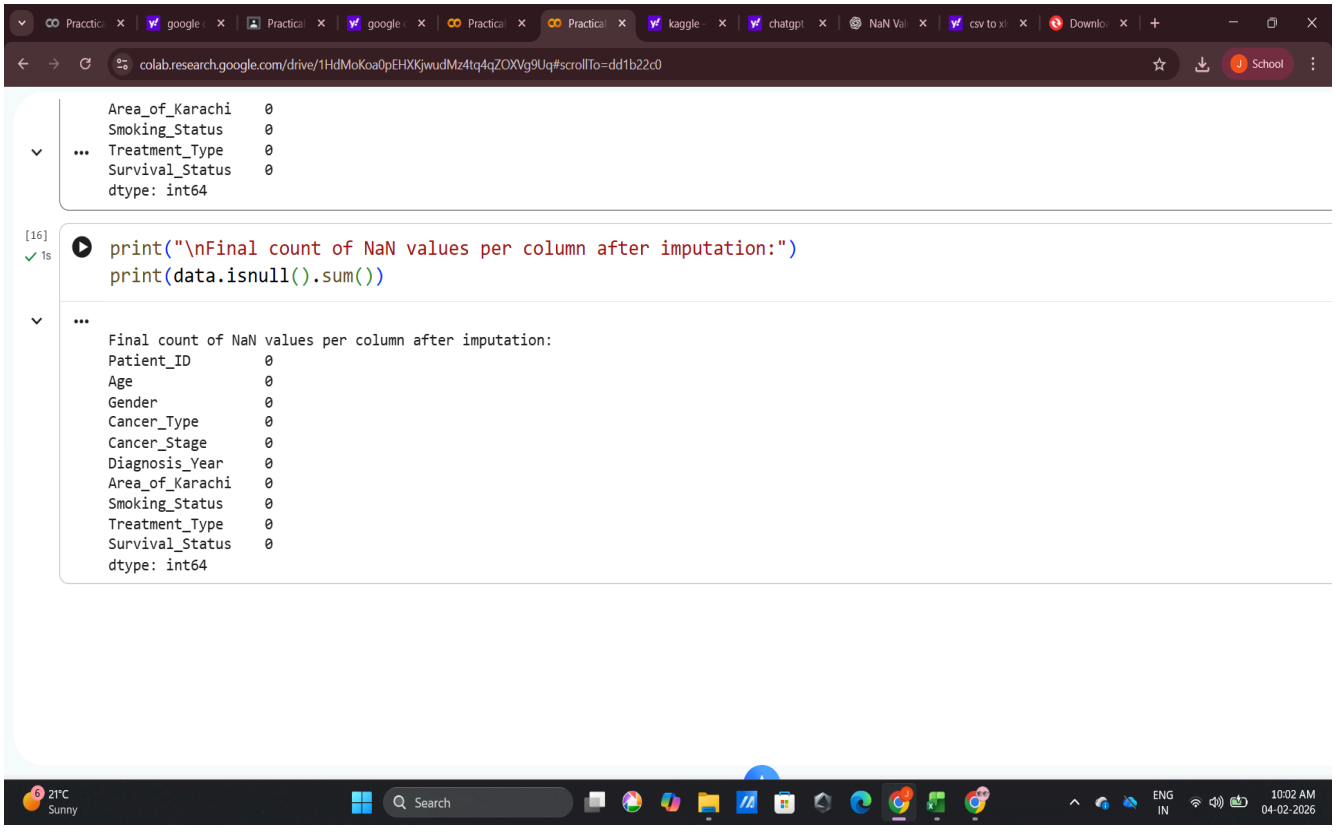


```
[12]
# Program:
# ============================================================
# Practical No. 2:Data Cleaning and Missing Value Imputation
# Student: Name: Mr. Jagdish Jadhav , Roll No: 66 , Semester: VI,
# Section: A, Branch: CSE(Cyber Security)
# Data Cleaning & Handling Missing Values
# ============================================================
# ================================
# STEP 1: Import required library
# ================================
# Pandas is used for data loading and analysis
import pandas as pd
print("Student: Name: Mr Jagdish Jadhav ,Roll No: 66")

# ================================
# STEP 2: Load the dataset

# ================================
# Read the CSV file into a Pandas DataFrame
# Make sure 'tips.csv' is uploaded to Google Colab
data = pd.read_csv("/content/karachi_cancer_dataset_5_nan_each_column (1).csv")

# ================================
# STEP 3: Display dataset information
```



```
[22]
✓ 0s
import pandas as pd

# Reload the original dataset into a new DataFrame to demonstrate imputation with NaNs
df_for_loop_demo = pd.read_csv("/content/karachi_cancer_dataset_5_nan_each_column (1).csv")

print("\nNaN counts before imputation (for loop demonstration):")
print(df_for_loop_demo.isnull().sum())

# Columns to be imputed as per user's example
columns_to_impute = [
    'Age',
    'Patient_ID',
    'Diagnosis_Year',
    'Area_of_Karachi',
    'Smoking_Status',
    'Cancer_Type',
    'Treatment_Type',
    'Survival_Status',
    'Gender',
    'Cancer_Stage'
]

# Impute NaN values using a for loop
for col in columns_to_impute:
```


colab.research.google.com/drive/1HdMoKoa0pEHXKjwudMz4tq4qZOXVg9Uq#scrollTo=dd1b22c0
dtype: int64
Filled NaN in 'Age' with mean: 43.17
Filled NaN in 'Patient_ID' with mean: 250.86
Filled NaN in 'Diagnosis_Year' with mean: 2019.40
Filled NaN in 'Area_of_Karachi' with mode: 'Korangi'
Filled NaN in 'Smoking_Status' with mode: 'Non-Smoker'
Filled NaN in 'Cancer_Type' with mode: 'Leukemia'
Filled NaN in 'Treatment_Type' with mode: 'Immunotherapy'
Filled NaN in 'Survival_Status' with mode: 'Alive'
Filled NaN in 'Gender' with mode: 'Male'
Filled NaN in 'Cancer_Stage' with mode: 'Stage I'
NaN counts after imputation (for loop demonstration):
Patient_ID 0
Age 0
Gender 0
Cancer_Type 0
Cancer_Stage 0
Diagnosis_Year 0
Area_of_Karachi 0
Smoking_Status 0
Treatment_Type 0
Survival_Status 0
dtype: int64
[16]
1s
print("\nFinal count of NaN values per column after imputation:")
print(data.isnull().sum())
Final count of NaN values per column after imputation:
Patient_ID 0
21°C
Sunny
Search
ENG
IN
10:01 AM
04-02-2026


colab.research.google.com/drive/1HdMoKoa0pEHXKjwudMz4tq4qZOXVg9Uq#scrollTo=dd1b22c0
Area_of_Karachi 0
Smoking_Status 0
Treatment_Type 0
Survival_Status 0
dtype: int64
[16]
1s
print("\nFinal count of NaN values per column after imputation:")
print(data.isnull().sum())
Final count of NaN values per column after imputation:
Patient_ID 0
Age 0
Gender 0
Cancer_Type 0
Cancer_Stage 0
Diagnosis_Year 0
Area_of_Karachi 0
Smoking_Status 0
Treatment_Type 0
Survival_Status 0
dtype: int64
21°C
Sunny
Search
ENG
IN
10:02 AM
04-02-2026