MACHINE LEARNING LAB

LAB 1

NIKKI EVANA BLESSY.N

2341459

1. a)Create a Dataset (.csv file)for Student Enrollment and Performance

Which contains the following fields.

StudentId(Unique identifier),

Course_Name: Categorical (Math, Science, History, Art).

Gender: (Male, Female, Non-binary).

Age: Integer (10-25).

Enrollment_Date: Random dates from the last 5 years.

Final_Grade: Float (0.0–100.0).

Attendance_Percentage: Float (50–100).

b)Include Null values in the data set.

c)Minimum Fifteen entries(record) in the data set.

| StudentI | Course_Nam | | | | Final_Grad | |
|----------|------------|--------|-----|-----------------|------------|-----------------------|
| d | е | Gender | Age | Enrollment_Date | е | Attendance_Percentage |
| S0001 | Math | Female | 20 | 1/18/2022 | 66.1 | 76.7 |
| S0002 | Math | Male | 18 | 5/14/2020 | 77.3 | 62.2 |
| S0003 | History | Female | 14 | 1/8/2023 | 98.5 | 73.1 |
| S0004 | Science | Male | 16 | 7/27/2021 | 85.5 | 63.5 |
| S0005 | Science | Female | 20 | 7/29/2024 | 86.6 | 96.3 |
| S0006 | Science | Female | 13 | 6/14/2023 | 38 | 84.4 |
| S0007 | Math | Male | 12 | 5/24/2023 | 64.1 | 61 |
| S0008 | Math | Male | 22 | 11/21/2024 | 83.4 | 66.2 |
| S0009 | Art | Female | 13 | 10/3/2024 | 16.3 | 88.4 |
| S0010 | Math | Male | 21 | 12/14/2021 | 35.5 | 52.8 |
| S0011 | Math | Female | 21 | 3/1/2023 | 67 | 92.12 |
| S0012 | Art | Male | 18 | 1/1/2021 | 70.2 | 90.3 |
| S0013 | Science | Male | 11 | 11/16/2023 | 68.4 | 70.1 |
| S0014 | Science | Female | 24 | 4/25/2020 | 7.1 | 53.3 |

95.7

2. Import the above(Q.No1) csv file using Pandas and display the following

```
import pandas as pd
     import numpy as np
     import seaborn as sns
[5] from google.colab import files
     uploaded = files.upload()
     import pandas as pd
     df = pd.read_csv("Student_Enrollment_and_Performance.csv")
     print(df.head())
                 Student_En...rmance.csv

    Student Enrollment and Performance.csv(text/csv) - 709 bytes, last modified: 12/3/2024 - 100% done

    Saving Student_Enrollment_and_Performance.csv to Student_Enrollment_and_Performance.csv
       StudentId Course_Name Gender Age Enrollment_Date Final_Grade \
S0001 Math Female 20 1/18/2022 66.1
           50002
                        Math Male 18
                                                                      77.3
                                                 5/14/2020
                     History Female 14
Science Male 16
          50003
                                                                      98.5
                                                  1/8/2023
                                                  7/27/2021
                                                                      85.5
           50004
           50005
                      Science Female 20
                                                  7/29/2024
                                                                      86.6
        Attendance Percentage
```

a)Display the first 10 rows, Head, Tail.

```
# First 10 rows
print("First 10 Rows:")
print(df.head(10))

# Head
print("\nHead:")
print(df.head())

# Tail
print("\nTail:")
print(df.tail())
```

| ₹ | First | 10 Rows | : | | | | | |
|---|-----------------------|----------|-------------|--------|-----------------|-------------|------|--|
| | StudentId Course_Name | | Gender | Age | Enrollment_Date | Final_Grade | \ | |
| | 0 | 50001 | Math | Female | 20 | 1/18/2022 | 66.1 | |
| | 1 | 50002 | Math | Male | 18 | 5/14/2020 | 77.3 | |
| | 2 | 50003 | History | Female | 14 | 1/8/2023 | 98.5 | |
| | 3 | 50004 | Science | Male | 16 | 7/27/2021 | 85.5 | |
| | 4 | 50005 | Science | Female | 20 | 7/29/2024 | 86.6 | |
| | 5 | 50006 | Science | Female | 13 | 6/14/2023 | 38.0 | |
| | 6 | 50007 | Math | Male | 12 | 5/24/2023 | 64.1 | |
| | 7 | 50008 | Math | Male | 22 | 11/21/2024 | 83.4 | |
| | 8 | 50009 | Art | Female | 13 | 10/3/2024 | 16.3 | |
| | 9 | 50010 | Math | Male | 21 | 12/14/2021 | 35.5 | |
| | | | | | | | | |
| | At | tendance | _Percentage | | | | | |
| | 0 76.7 | | | | | | | |
| | 1 | 1 62.2 | | | | | | |
| | 2 | 2 73.1 | | | | | | |
| | 3 63.5 | | | | | | | |
| | 4 96.3 | | | | | | | |
| | 5 | 5 84.4 | | | | | | |
| | 6 61.0 | | | | | | | |
| | 7 66.2 | | | | | | | |
| | 8 88.4 | | | | | | | |
| | 9 | | 52.8 | | | | | |

```
Head:
  StudentId Course_Name Gender Age Enrollment_Date Final_Grade \
     50001
                                       1/18/2022
                 Math Female
                               20
                                                        66.1
                                       5/14/2020
1
     50002
                 Math
                         Male
                               18
                                                        77.3
2
     50003
              History Female
                               14
                                        1/8/2023
                                                        98.5
     50004
                         Male 16
                                       7/27/2021
                                                        85.5
              Science
     50005
              Science Female 20
                                      7/29/2024
                                                        86.6
   Attendance_Percentage
0
                  62.2
2
                  73.1
3
                  63.5
4
                  96.3
Tail:
   StudentId Course_Name Gender Age Enrollment_Date Final_Grade
                                          3/1/2023
10
      50011
                  Math Female
                                                         67.0
                               21
11
      50012
                   Art
                         Male 18
                                         1/1/2021
                                                         70.2
12
      S0013 Science
                          Male 11
                                       11/16/2023
                                                         68.4
13
      50014
               Science Female 24
                                        4/25/2020
                                                         7.1
14
      50015
                  Math Female 13
                                         3/7/2020
                                                         63.5
   Attendance_Percentage
10
                  92.12
                  90.30
11
12
                  70.10
13
                  53.30
14
                  95.70
```

b)Display the shape of the data.

```
[7] print("\nShape of the dataset:")
print(df.shape)

Shape of the dataset:
(15, 7)
```

c)Display the columns, Number unique columns and the data available in any of the unique columns.

```
# Display columns
print("\nColumns in the dataset:")
print(df.columns)

# Number of unique values per column
print("\nNumber of unique values per column:")
print(df.nunique())

# unique values for each column
print("\nSample unique values from each column:")
for column in df.columns:
    print(f"{column}: {df[column].unique()[:5]}")
```

```
Columns in the dataset:
Index(['StudentId', 'Course_Name', 'Gender', 'Age', 'Enrollment Date',
       'Final_Grade', 'Attendance_Percentage'],
      dtype='object')
Number of unique values per column:
StudentId
Course_Name
                        2
Gender
                       10
Enrollment Date
                       15
Final Grade
                        15
Attendance_Percentage 15
dtype: int64
Sample unique values from each column:
StudentId: ['S0001' 'S0002' 'S0003' 'S0004' 'S0005']
Course_Name: ['Math' 'History' 'Science' 'Art']
Gender: ['Female' 'Male']
Age: [20 18 14 16 13]
Enrollment_Date: ['1/18/2022' '5/14/2020' '1/8/2023' '7/27/2021' '7/29/2024']
Final_Grade: [66.1 77.3 98.5 85.5 86.6]
Attendance_Percentage: [76.7 62.2 73.1 63.5 96.3]
```

d)Check the Null values in the dataset.

```
print("\nCheck for null values:")
    print(df.isnull().sum())
    Check for null values:
    StudentId
                            0
    Course_Name
                            0
    Gender
                            0
                            0
    Age
    Enrollment_Date
                            0
    Final_Grade
                            0
    Attendance_Percentage
    dtype: int64
```

e)Store the data in data frame and display

```
print("\nDataframe contents:")
print(df)
```

```
[10] Dataframe contents:
        StudentId Course Name
                                         Age Enrollment_Date Final_Grade \
                                Gender
     0
            50001
                          Math
                                Female
                                          20
                                                    1/18/2022
                                                                       66.1
     1
                                                    5/14/2020
            50002
                          Math
                                  Male
                                          18
                                                                       77.3
     2
            50003
                       History
                                Female
                                          14
                                                    1/8/2023
                                                                       98.5
            50004
                       Science
                                  Male
                                          16
                                                    7/27/2021
                                                                       85.5
     4
            50005
                       Science
                                Female
                                          20
                                                    7/29/2024
                                                                       86.6
     5
                       Science
                                Female
                                                    6/14/2023
                                                                       38.0
            50006
                                          13
     6
            50007
                                                                       64.1
                          Math
                                  Male
                                          12
                                                    5/24/2023
                                                  11/21/2024
            50008
                          Math
                                  Male
                                          22
                                                                       83.4
     8
            50009
                           Art
                                Female
                                          13
                                                    10/3/2024
                                                                       16.3
     9
            50010
                                  Male
                                                  12/14/2021
                                                                       35.5
                          Math
                                          21
     10
            50011
                          Math
                                Female
                                          21
                                                     3/1/2023
                                                                       67.0
     11
            50012
                                  Male
                                                     1/1/2021
                                                                       70.2
                           Art
                                          18
     12
            50013
                       Science
                                   Male
                                                  11/16/2023
                                                                       68.4
                                          11
     13
            50014
                       Science Female
                                          24
                                                   4/25/2020
                                                                       7.1
     14
            50015
                          Math
                                Female
                                          13
                                                    3/7/2020
                                                                       63.5
         Attendance_Percentage
     0
                          76.70
     1
                          62.20
     2
                          73.10
                          63.50
     4
                          96.30
     5
                          84.40
     6
                          61.00
                          66.20
     8
                          88.40
     9
                          52.80
```

e)Fill the null values with 'na' and display

```
df_na = df.fillna('na')
print("\nDataframe with null values filled with 'na':")
print(df_na)
```

| | υa | rallanc wr | cii iiuii vaiuc | 3 111154 | WILL | ı 11 0 . | | |
|---|-----------------------|------------|-----------------|----------|------|-----------------|---------------|--|
| 0 | | | Course Name | Gender | | Enrollment_Date | Final Grade \ | |
| , | 0 | 50001 | _ Math | Female | 20 | 1/18/2022 | 66.1 | |
| | 1 | 50002 | Math | Male | 18 | 5/14/2020 | 77.3 | |
| | 2 | 50003 | History | Female | 14 | 1/8/2023 | 98.5 | |
| | 3 | 50004 | Science | Male | 16 | 7/27/2021 | 85.5 | |
| | 4 | 50005 | Science | Female | 20 | 7/29/2024 | 86.6 | |
| | 5 | 50006 | Science | Female | 13 | 6/14/2023 | 38.0 | |
| | 6 | 50007 | Math | Male | 12 | 5/24/2023 | 64.1 | |
| | 7 | 50008 | Math | Male | 22 | 11/21/2024 | 83.4 | |
| | 8 | 50009 | Art | Female | 13 | 10/3/2024 | 16.3 | |
| | 9 | 50010 | Math | Male | 21 | 12/14/2021 | 35.5 | |
| | 10 | 50011 | Math | Female | 21 | 3/1/2023 | 67.0 | |
| | 11 | 50012 | Art | Male | 18 | 1/1/2021 | 70.2 | |
| | 12 | 50013 | Science | Male | 11 | 11/16/2023 | 68.4 | |
| | 13 | 50014 | Science | Female | 24 | 4/25/2020 | 7.1 | |
| | 14 | 50015 | Math | Female | 13 | 3/7/2020 | 63.5 | |
| | | | | | | | | |
| | Attendance_Percentage | | | | | | | |
| | 0 | | 76.70 |) | | | | |
| | 1 | | 62.20 |) | | | | |
| | 2 | 73.10 | | | | | | |
| | 3 | | 63.50 | | | | | |
| | 4 | | 96.30 |) | | | | |
| | 5 | | 84.40 |) | | | | |
| | 6 | | 61.00 | | | | | |
| | 7 | | 66.20 | | | | | |
| | 8 | | 88.40 | | | | | |
| | 9 | | 52.80 |) | | | | |
| | 10 | | 02 12 |) | | | | |

f)Fill the null values with 'mean' and display

```
df_mean = df.fillna(df.mean(numeric_only=True))
print("\nDataframe with null values filled with mean:")
print(df_mean)
```

```
O
    Dataframe with null values filled with mean:
        StudentId Course_Name
                                Gender
                                         Age Enrollment_Date
                                                               Final Grade
₹
    0
            50001
                          Math
                                Female
                                          20
                                                    1/18/2022
                                                                       66.1
                                  Male
                                          18
                                                    5/14/2020
     1
            50002
                          Math
                                                                       77.3
                      History
                                Female
                                                     1/8/2023
     2
            50003
                                          14
                                                                       98.5
     3
                       Science
                                  Male
                                          16
                                                    7/27/2021
                                                                       85.5
            50004
    4
            50005
                       Science
                                Female
                                          20
                                                    7/29/2024
                                                                       86.6
    5
            50006
                       Science
                                Female
                                          13
                                                    6/14/2023
                                                                       38.0
    6
            50007
                          Math
                                  Male
                                          12
                                                    5/24/2023
                                                                       64.1
                                  Male
                                          22
                                                  11/21/2024
                                                                       83.4
            50008
                          Math
                                Female
                                                                       16.3
    8
            50009
                           Art
                                          13
                                                    10/3/2024
    9
            50010
                          Math
                                  Male
                                          21
                                                  12/14/2021
                                                                       35.5
    10
            50011
                          Math
                                Female
                                          21
                                                     3/1/2023
                                                                       67.0
    11
                                  Male
                                                     1/1/2021
            50012
                           Art
                                          18
                                                                       70.2
    12
            50013
                       Science
                                  Male
                                          11
                                                  11/16/2023
                                                                       68.4
    13
            50014
                       Science
                                Female
                                          24
                                                    4/25/2020
                                                                        7.1
    14
            50015
                          Math
                                Female
                                          13
                                                     3/7/2020
                                                                       63.5
         Attendance_Percentage
    0
                          76.70
     1
                          62.20
     2
                          73.10
     3
                          63.50
    4
                          96.30
    5
                          84.40
                          61.00
    6
                          66.20
    8
                          88.40
    q
                          52.80
```

g)Fill the null values with 'median' and display

```
df_median = df.fillna(df.median(numeric_only=True))
print("\nDataframe with null values filled with median:")
print(df_median)
```

| Dataframe with null values filled with median: | |
|---|------------|
| | ↑ ↓ |
| Os StudentId Course_Name Gender Age Enrollment_Date | Final_' |
| 9 50001 Math Female 20 1/18/2022 | 66.1 |
| 1 50002 Math Male 18 5/14/2020 | 77.3 |
| 2 S0003 History Female 14 1/8/2023 | 98.5 |
| 3 S0004 Science Male 16 7/27/2021 | 85.5 |
| 4 S0005 Science Female 20 7/29/2024 | 86.6 |
| 5 S0006 Science Female 13 6/14/2023 | 38.0 |
| 6 S0007 Math Male 12 5/24/2023 | 64.1 |
| 7 S0008 Math Male 22 11/21/2024 | 83.4 |
| 8 S0009 Art Female 13 10/3/2024 | 16.3 |
| 9 S0010 Math Male 21 12/14/2021 | 35.5 |
| 10 S0011 Math Female 21 3/1/2023 | 67.0 |
| 11 S0012 Art Male 18 1/1/2021 | 70.2 |
| 12 S0013 Science Male 11 11/16/2023 | 68.4 |
| 13 S0014 Science Female 24 4/25/2020 | 7.1 |
| 14 S0015 Math Female 13 3/7/2020 | 63.5 |
| | |
| Attendance_Percentage | |
| 0 76.70 | |
| 1 62.20 | |
| 2 73.10 | |
| 3 63.50 | |
| 4 96.30 | |
| 5 84.40 | |
| 6 61.00 | |
| 7 66.20 | |
| 8 88.40 | |
| 9 52.80 | |