Day 4 Explore Pandas

Required Libraries Pandas

In [41…

```
import pandas as pd # Library
import numpy as np # may be contanins some lists,dict and tuples
import os # for Data export or saving and more...
```

Data frame creation

In [8]:

```
data = {'Name': ['A', 'B'], 'Age': [24, 27]}
df = pd.DataFrame(data)
```

In [10…

```
print(df)
```

```
  Name  Age
0    A   24
1    B   27
```

Loading another new data frames

In [13…

```
data1 = {'Name': ['Sandesh' ,'Saroj', 'Santosh', 'Shambhu'], 'Age': [18,
df = pd.DataFrame(data1)
print(df)
```

```
      Name  Age
0  Sandesh   18
1    Saroj   16
2  Santosh   20
3  Shambhu   21
```

In [16…

```
import pandas as pd
import numpy as np

data = {
    'Name': ['Sandesh', 'Saroj', 'Santosh', 'Shambhu', 'Prakash', 'Bimal
             'Ram', 'Shyam', 'Hari', 'Gita', 'Rita', 'Sita', 'Anita', 'K
    'Age': np.random.randint(15, 22, size=14),  # Now 14 ages to match 1
    'Class': np.random.choice(['IX', 'X', 'XI', 'XII'], size=14),
    'Faculty': np.random.choice(['Science', 'Management', 'Arts', 'Educa
    'Address': np.random.choice(['Kathmandu', 'Pokhara', 'Bhaktapur', 'L
                                 'Biratnagar', 'Dharan', 'Butwal'], size=
    'Grade': np.random.choice(['A+', 'A', 'B+', 'B', 'C+', 'C'], size=14
    'Gender': ['Male', 'Male', 'Male', 'Male', 'Male', 'Male',
               'Male', 'Male', 'Male', 'Female', 'Female', 'Female', 'Fe
    'Email': [name.lower().replace(' ', '') + '@school.edu.np' for name
    'Phone': ['98' + str(np.random.randint(10000000, 99999999)) for _ in
}
```

```python
df = pd.DataFrame(data)
print(df)
```

```
       Name  Age Class     Faculty     Address Grade  Gender  \
0   Sandesh   15   XII        Arts      Butwal     A    Male   
1     Saroj   20   XII        Arts    Lalitpur     C    Male   
2   Santosh   18    IX        Arts      Butwal     B    Male   
3   Shambhu   19   XII     Science      Dharan     B    Male   
4   Prakash   20   XII  Management     Pokhara    C+    Male   
5     Bimal   21    IX   Education      Butwal     B    Male   
6       Ram   21     X   Education    Lalitpur    B+    Male   
7     Shyam   20    XI   Education  Biratnagar    B+    Male   
8      Hari   18    XI   Education   Kathmandu     A    Male   
9      Gita   19    XI   Education   Kathmandu    A+  Female   
10     Rita   21    XI  Management      Dharan     C  Female   
11     Sita   15     X  Management     Pokhara    A+  Female   
12    Anita   20   XII   Education     Pokhara     C  Female   
13  Krishna   19     X        Arts      Dharan    B+  Female   

                    Email       Phone  
0   sandesh@school.edu.np  9834575496  
1     saroj@school.edu.np  9848487258  
2   santosh@school.edu.np  9881356261  
3   shambhu@school.edu.np  9888576996  
4   prakash@school.edu.np  9895646646  
5     bimal@school.edu.np  9843075797  
6       ram@school.edu.np  9818826272  
7     shyam@school.edu.np  9882169900  
8      hari@school.edu.np  9891289747  
9      gita@school.edu.np  9890163813  
10     rita@school.edu.np  9868378916  
11     sita@school.edu.np  9865775037  
12    anita@school.edu.np  9814586880  
13  krishna@school.edu.np  9899778822  
```

In [16…

```python
import pandas as pd
import numpy as np

data = {
    'Name': ['Sandesh', 'Saroj', 'Santosh', 'Shambhu', 'Prakash', 'Bimal
             'Ram', 'Shyam', 'Hari', 'Gita', 'Rita', 'Sita', 'Anita', 'K
    'Age': np.random.randint(15, 22, size=14),  # Now 14 ages to match 1
    'Class': np.random.choice(['IX', 'X', 'XI', 'XII'], size=14),
    'Faculty': np.random.choice(['Science', 'Management', 'Arts', 'Educa
    'Address': np.random.choice(['Kathmandu', 'Pokhara', 'Bhaktapur', 'L
                                 'Biratnagar', 'Dharan', 'Butwal'], size=
    'Grade': np.random.choice(['A+', 'A', 'B+', 'B', 'C+', 'C'], size=14
    'Gender': ['Male', 'Male', 'Male', 'Male', 'Male', 'Male',
               'Male', 'Male', 'Male', 'Female', 'Female', 'Female', 'Fe
    'Email': [name.lower().replace(' ', '') + '@school.edu.np' for name
    'Phone': ['98' + str(np.random.randint(10000000, 99999999)) for _ in
}

df = pd.DataFrame(data)
print(df)
```

```
       Name  Age Class     Faculty     Address Grade  Gender  \
0   Sandesh   15   XII        Arts      Butwal     A    Male   
1     Saroj   20   XII        Arts    Lalitpur     C    Male   
2   Santosh   18    IX        Arts      Butwal     B    Male   
3   Shambhu   19   XII     Science      Dharan     B    Male   
4   Prakash   20   XII  Management     Pokhara    C+    Male   
5     Bimal   21    IX   Education      Butwal     B    Male   
6       Ram   21     X   Education    Lalitpur    B+    Male   
7     Shyam   20    XI   Education  Biratnagar    B+    Male   
8      Hari   18    XI   Education   Kathmandu     A    Male   
9      Gita   19    XI   Education   Kathmandu    A+  Female   
10     Rita   21    XI  Management      Dharan     C  Female   
11     Sita   15     X  Management     Pokhara    A+  Female   
12    Anita   20   XII   Education     Pokhara     C  Female   
13  Krishna   19     X        Arts      Dharan    B+  Female   

                    Email       Phone  
0   sandesh@school.edu.np  9834575496  
1     saroj@school.edu.np  9848487258  
2   santosh@school.edu.np  9881356261  
3   shambhu@school.edu.np  9888576996  
4   prakash@school.edu.np  9895646646  
5     bimal@school.edu.np  9843075797  
6       ram@school.edu.np  9818826272  
7     shyam@school.edu.np  9882169900  
8      hari@school.edu.np  9891289747  
9      gita@school.edu.np  9890163813  
10     rita@school.edu.np  9868378916  
11     sita@school.edu.np  9865775037  
12    anita@school.edu.np  9814586880  
13  krishna@school.edu.np  9899778822  
```

export the data frame as csv using this library

In [17…

```
df.to_csv('student_data.csv', index=False)
print("CSV file 'student_data.csv' created successfully!") # check dat w
```

```
CSV file 'student_data.csv' created successfully!
```

In [18…

```
import os
os.getcwd()
```

Out[18…

```
'C:\\Users\\DELL'
```

For User choice location or dyanmaically located path

In [22…

```
df = pd.DataFrame(data)

folder_path = r"C:\Users\DELL\Desktop\A\day 4"
os.makedirs(folder_path, exist_ok=True)
file_path = os.path.join(folder_path, "student_data.csv")
df.to_csv(file_path, index=False)
print(f"CSV saved to: {folder_path}")
```

```
CSV saved to: C:\Users\DELL\Desktop\A\day 4
```

Data Cleaning

df.isnull(), df.dropna(), df.fillna() df.duplicated(), df.drop_duplicates() Type conversions: df.astype() Rename columns: df.rename()Sample Raw Data -> Need to clean

In [24…

```
data_raw = {
    'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve', 'Frank', 'Alice'
    'Age': [25, np.nan, 35, 45, 28, None, 25, 33],
    'Gender': ['Female', 'Male', 'Male', 'Male', 'Female', None, 'Female
    'Grade': ['A', 'B', 'A', 'B', np.nan, 'C', 'A', 'B'],
    'Email': ['alice@mail.com', 'bob@mail.com', 'charlie@mail.com',
              'david@mail.com', 'eve@mail.com', 'frank@mail.com',
              'alice@mail.com', 'unknown@mail.com']
}

df = pd.DataFrame(data_raw)
print(df)
```

```
      Name   Age  Gender Grade             Email
0    Alice  25.0  Female     A    alice@mail.com
1      Bob   NaN    Male     B      bob@mail.com
2  Charlie  35.0    Male     A  charlie@mail.com
3    David  45.0    Male     B    david@mail.com
4      Eve  28.0  Female   NaN      eve@mail.com
5    Frank   NaN    None     C    frank@mail.com
6    Alice  25.0  Female     A    alice@mail.com
7     None  33.0  Female     B  unknown@mail.com
```

In [25…

```
# 1. Check for missing values
print(df.isnull().sum())
```

```
Name      1
Age       2
Gender    1
Grade     1
Email     0
dtype: int64
```

In [26…

```
# 2. Drop rows with missing Name or Email (essential fields)
df = df.dropna(subset=['Name', 'Email'])
```

In [35…

```
print(df.isnull().sum())
print("\n")
print(df)
```

```
Name      0
Age       0
Gender    0
Grade     0
Email     0
dtype: int64


      Name   Age  Gender Grade              Email
0    Alice  25.0  Female     A    alice@mail.com
1      Bob  31.6    Male     B      bob@mail.com
2  Charlie  35.0    Male     A  charlie@mail.com
3    David  45.0    Male     B    david@mail.com
4      Eve  28.0  Female     A      eve@mail.com
5    Frank  31.6  Female     C    frank@mail.com
6    Alice  25.0  Female     A    alice@mail.com
```

In [36…

```
df['Age'] = df['Age'].fillna(df['Age'].mean())
# 3. Fill missing Age with mean
print(df)
```

```
      Name   Age  Gender Grade              Email
0    Alice  25.0  Female     A    alice@mail.com
1      Bob  31.6    Male     B      bob@mail.com
2  Charlie  35.0    Male     A  charlie@mail.com
3    David  45.0    Male     B    david@mail.com
4      Eve  28.0  Female     A      eve@mail.com
5    Frank  31.6  Female     C    frank@mail.com
6    Alice  25.0  Female     A    alice@mail.com
```

In [32…

```
# 4. Fill missing Gender and Grade with mode (most frequent value)
df['Gender'] = df['Gender'].fillna(df['Gender'].mode()[0])
df['Grade'] = df['Grade'].fillna(df['Grade'].mode()[0])
print(df)
```

```
      Name   Age  Gender Grade             Email
0    Alice  25.0  Female     A    alice@mail.com
1      Bob  31.6    Male     B      bob@mail.com
2  Charlie  35.0    Male     A  charlie@mail.com
3    David  45.0    Male     B    david@mail.com
4      Eve  28.0  Female     A      eve@mail.com
5    Frank  31.6  Female     C    frank@mail.com
6    Alice  25.0  Female     A    alice@mail.com
```

In [37…

```
# 5. Drop duplicate rows based on Name + Email
df = df.drop_duplicates(subset=['Name', 'Email'])
print(df)
```

```
      Name   Age  Gender Grade             Email
0    Alice  25.0  Female     A    alice@mail.com
1      Bob  31.6    Male     B      bob@mail.com
2  Charlie  35.0    Male     A  charlie@mail.com
3    David  45.0    Male     B    david@mail.com
4      Eve  28.0  Female     A      eve@mail.com
5    Frank  31.6  Female     C    frank@mail.com
```

In [38…

```
# 6. Convert Age to int (after filling NaNs)
df['Age'] = df['Age'].astype(int)
print(df)
```

```
      Name  Age  Gender Grade             Email
0    Alice   25  Female     A    alice@mail.com
1      Bob   31    Male     B      bob@mail.com
2  Charlie   35    Male     A  charlie@mail.com
3    David   45    Male     B    david@mail.com
4      Eve   28  Female     A      eve@mail.com
5    Frank   31  Female     C    frank@mail.com
```

In [92…

```
# 7. Rename columns (optional)
df = df.rename(columns={'Name': 'Student_Name', 'Grade': 'Final_Grade'})
print(df)
```

```
   Student_Name  Age Class     Faculty     Address  Gender  \
0       SANDESH   21     X     Science  Biratnagar    Male   
1         SAROJ   20    IX     Science   Bhaktapur    Male   
2       SANTOSH   20     X  Management    Lalitpur    Male   
3       SHAMBHU   15   XII   Education     Pokhara    Male   
4       PRAKASH   15     X  Management   Bhaktapur    Male   
5         BIMAL   17    IX  Management   Bhaktapur    Male   
6           RAM   18     X  Management   Bhaktapur    Male   
7         SHYAM   19    IX     Science     Pokhara    Male   
8          HARI   17    XI   Education      Dharan    Male   
9          GITA   18    IX   Education      Butwal  Female   
10         RITA   16    IX  Management  Biratnagar  Female   
11         SITA   21   XII     Science   Bhaktapur  Female   
12        ANITA   16    XI  Management     Pokhara  Female   
13      KRISHNA   21   XII        Arts  Biratnagar  Female   

                    Email       Phone Age_Group Age_Category  \
0   sandesh@school.edu.np  9883693228     Adult  Young Adult   
1     saroj@school.edu.np  9854449472     Adult  Young Adult   
2   santosh@school.edu.np  9840939497     Adult  Young Adult   
3   shambhu@school.edu.np  9824465240      Teen         Teen   
4   prakash@school.edu.np  9891912195      Teen         Teen   
5     bimal@school.edu.np  9875648619      Teen         Teen   
6       ram@school.edu.np  9834722002      Teen         Teen   
7     shyam@school.edu.np  9832606286      Teen         Teen   
8      hari@school.edu.np  9872607344      Teen         Teen   
9      gita@school.edu.np  9883414030      Teen         Teen   
10     rita@school.edu.np  9876036580      Teen         Teen   
11     sita@school.edu.np  9838854765     Adult  Young Adult   
12    anita@school.edu.np  9833800166      Teen         Teen   
13  krishna@school.edu.np  9895928153     Adult  Young Adult   

                           Identity  Gender_Encoded  Final_Grade_A+  \
0   SANDESH <sandesh@school.edu.np>               1            True   
1       SAROJ <saroj@school.edu.np>               1           False   
2   SANTOSH <santosh@school.edu.np>               1           False   
3   SHAMBHU <shambhu@school.edu.np>               1           False   
4   PRAKASH <prakash@school.edu.np>               1           False   
5       BIMAL <bimal@school.edu.np>               1           False   
6           RAM <ram@school.edu.np>               1           False   
7       SHYAM <shyam@school.edu.np>               1            True   
8         HARI <hari@school.edu.np>               1           False   
9         GITA <gita@school.edu.np>               0           False   
10        RITA <rita@school.edu.np>               0           False   
11        SITA <sita@school.edu.np>               0            True   
12      ANITA <anita@school.edu.np>               0           False   
13  KRISHNA <krishna@school.edu.np>               0            True   
```

```
     Final_Grade_B  Final_Grade_C  Final_Grade_C+  Grade_B  Grade_C  Grad
e_D
0            False          False           False    False     True    Fa
lse
1            False          False           False    False    False     T
rue
2            False           True           False    False     True    Fa
lse
3            False          False            True     True    False    Fa
lse
4            False          False           False    False    False     T
rue
5             True          False           False    False     True    Fa
lse
6            False           True           False     True    False    Fa
lse
7            False          False           False    False    False     T
rue
8            False          False            True    False     True    Fa
lse
9            False          False            True    False    False     T
rue
10           False          False            True    False    False     T
rue
11           False          False           False    False    False    Fa
lse
12           False          False            True    False    False     T
rue
13           False          False           False    False    False     T
rue
```

Final check

print(df)

In [42…

```python
folder_path = r"C:\Users\DELL\Desktop\A\day 4\cleaned"
os.makedirs(folder_path, exist_ok=True)
file_path = os.path.join(folder_path, "cleaned_student_data.csv")
df.to_csv(file_path, index=False)

print(f"CSV saved successfully at: {file_path}")
```

```
CSV saved successfully at: C:\Users\DELL\Desktop\A\day 4\cleaned\cleaned
_student_data.csv
```

Step 3: Transformations & Aggregations

In [44…

```python
# 1. Create Age Group column
df['Age_Group'] = df['Age'].apply(lambda x: 'Teen' if x < 20 else 'Adult
# 2. Uppercase column values (for show)
df['Student_Name'] = df['Student_Name'].apply(lambda x: x.upper())
print(df)
```

```
  Student_Name  Age  Gender Final_Grade              Email Age_Group
0        ALICE   25  Female           A     alice@mail.com     Adult
1          BOB   31    Male           B       bob@mail.com     Adult
2      CHARLIE   35    Male           A   charlie@mail.com     Adult
3        DAVID   45    Male           B     david@mail.com     Adult
4          EVE   28  Female           A       eve@mail.com     Adult
5        FRANK   31  Female           C     frank@mail.com     Adult
```

In [47…

```
# 1. Average age by gender
print(df.groupby('Gender')['Age'].mean())

# 2. Count of grades per gender
print(df.groupby('Gender')['Final_Grade'].value_counts())

# 3. How many students in each Age_Group
print(df['Age_Group'].value_counts())
```

```
Gender
Female    28.0
Male      37.0
Name: Age, dtype: float64
Gender  Final_Grade
Female  A              2
        C              1
Male    B              2
        A              1
Name: count, dtype: int64
Age_Group
Adult    6
Name: count, dtype: int64
```

In [48…

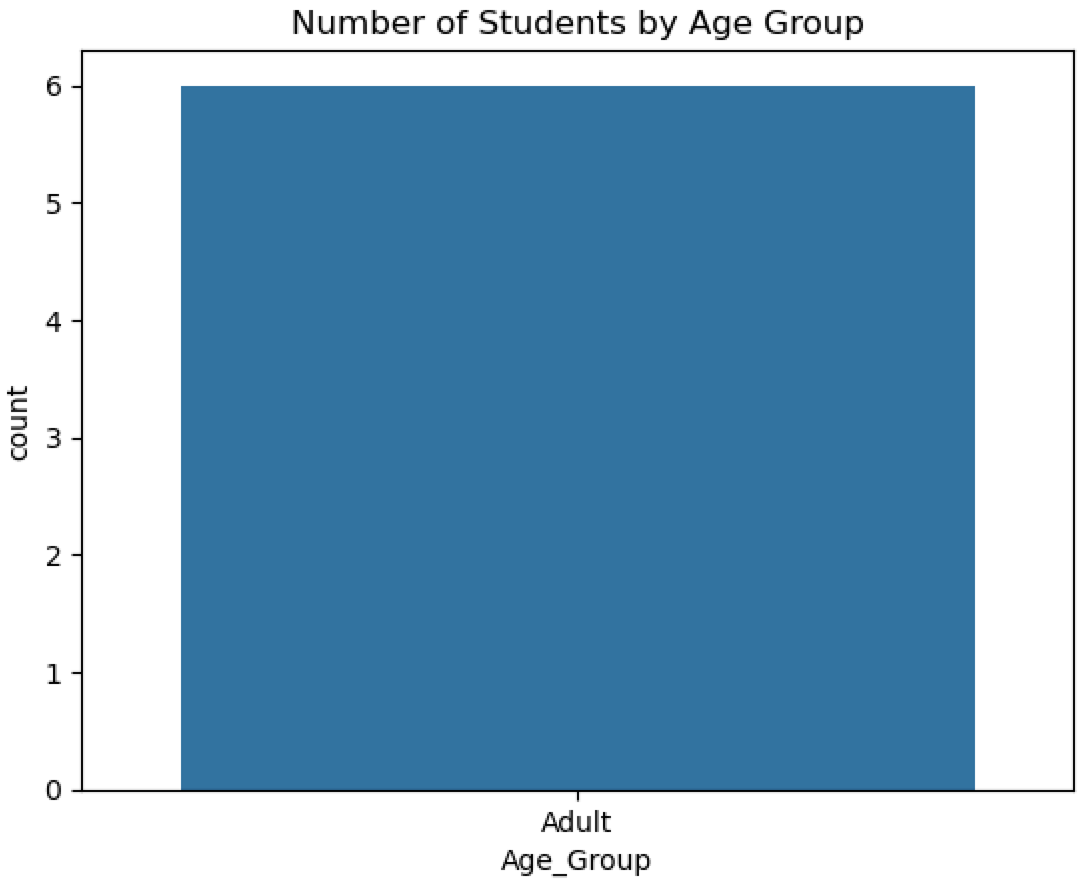
```
# Sort students by age (descending)
df_sorted = df.sort_values(by='Age', ascending=False)
print(df_sorted)

# Most common grades
print(df['Final_Grade'].value_counts())
```

```
  Student_Name  Age  Gender Final_Grade             Email Age_Group
3        DAVID   45    Male           B    david@mail.com     Adult
2      CHARLIE   35    Male           A  charlie@mail.com     Adult
1          BOB   31    Male           B      bob@mail.com     Adult
5        FRANK   31  Female           C    frank@mail.com     Adult
4          EVE   28  Female           A      eve@mail.com     Adult
0        ALICE   25  Female           A    alice@mail.com     Adult
Final_Grade
A    3
B    2
C    1
Name: count, dtype: int64
```

For Some Visulization from Above

In [50…

```python
import seaborn as sns
import matplotlib.pyplot as plt

# Bar plot of student counts by Age_Group
sns.countplot(x='Age_Group', data=df)
plt.title("Number of Students by Age Group")
plt.show()
```



Step 4: Exploratory Data Analysis (EDA)

In [52…

```python
# Basic structure and summary
print(df.info())
print("\n")
print(df.describe(include='all'))
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6 entries, 0 to 5
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   Student_Name  6 non-null      object
 1   Age           6 non-null      int32 
 2   Gender        6 non-null      object
 3   Final_Grade   6 non-null      object
 4   Email         6 non-null      object
 5   Age_Group     6 non-null      object
dtypes: int32(1), object(5)
memory usage: 312.0+ bytes
None


       Student_Name        Age  Gender Final_Grade            Email Age_G
roup
count             6   6.000000       6           6                6     
6
unique            6        NaN       2           3                6     
1
top           ALICE        NaN  Female           A  alice@mail.com     A
dult
freq              1        NaN       3           3                1     
6
mean            NaN  32.500000     NaN         NaN              NaN     
NaN
std             NaN   6.978539     NaN         NaN              NaN     
NaN
min             NaN  25.000000     NaN         NaN              NaN     
NaN
25%             NaN  28.750000     NaN         NaN              NaN     
NaN
50%             NaN  31.000000     NaN         NaN              NaN     
NaN
75%             NaN  34.000000     NaN         NaN              NaN     
NaN
max             NaN  45.000000     NaN         NaN              NaN     
NaN
```

df.describe() shows summary stats (mean, std, min, max, etc.)

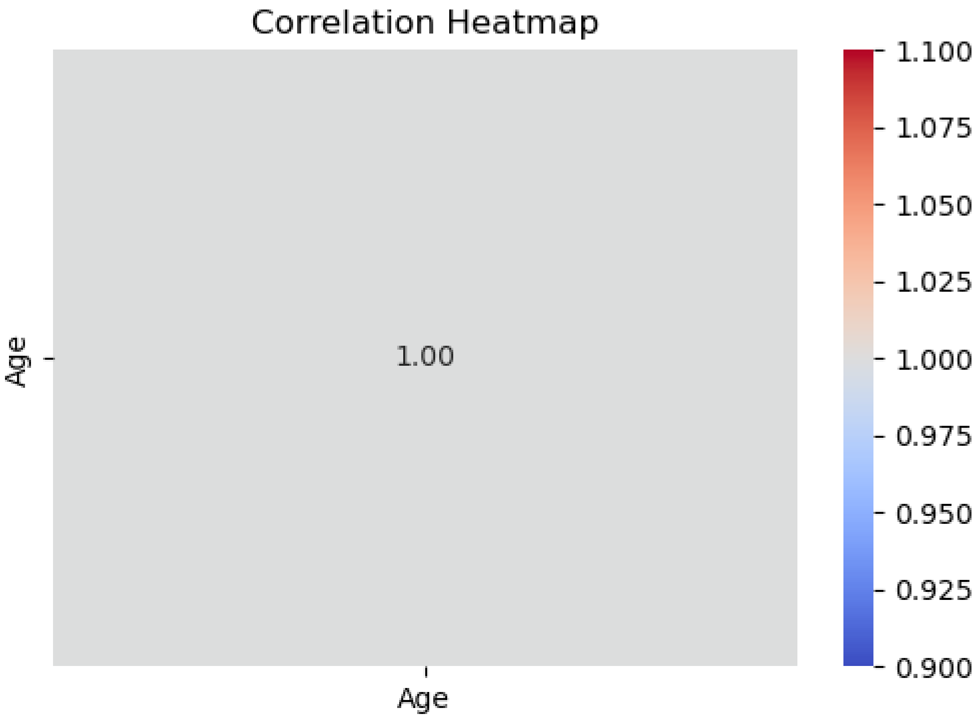
In [53…

```
corr = df.corr(numeric_only=True)
print(corr)
```

```
     Age
Age  1.0
```

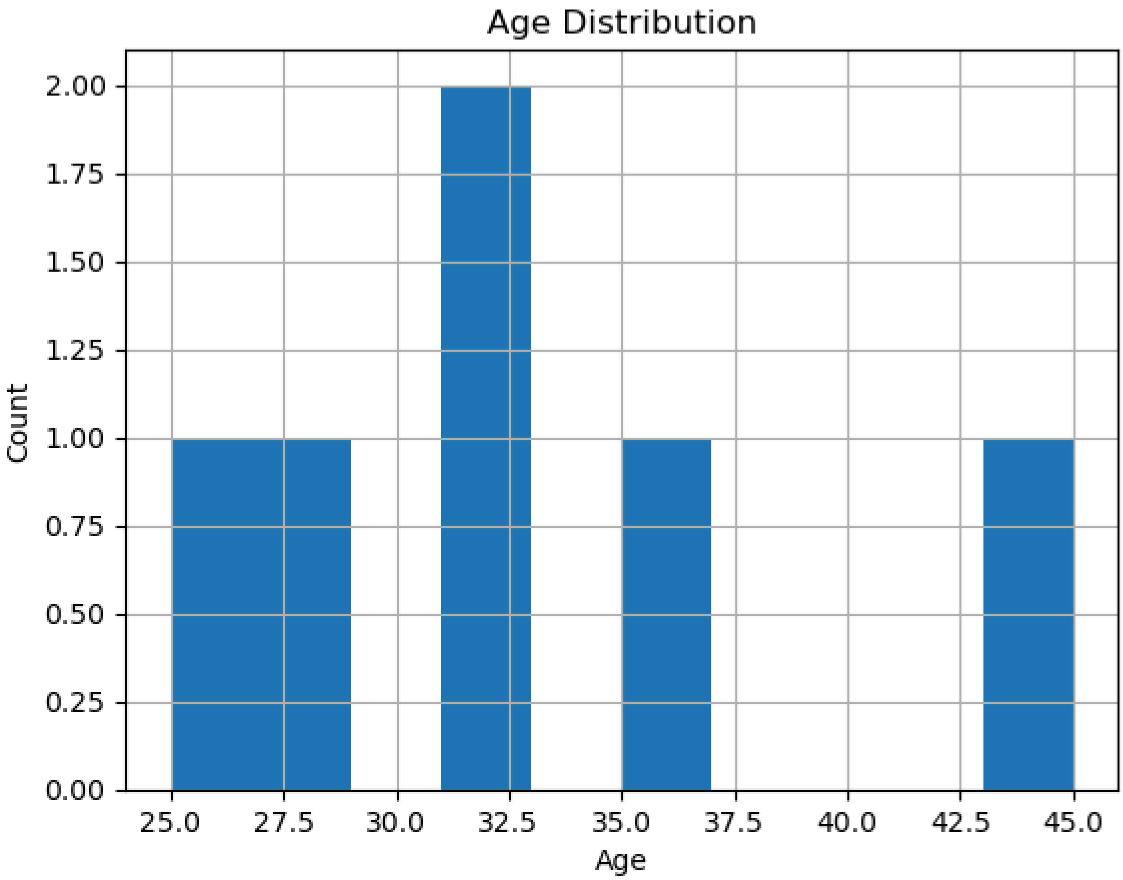
In [54…

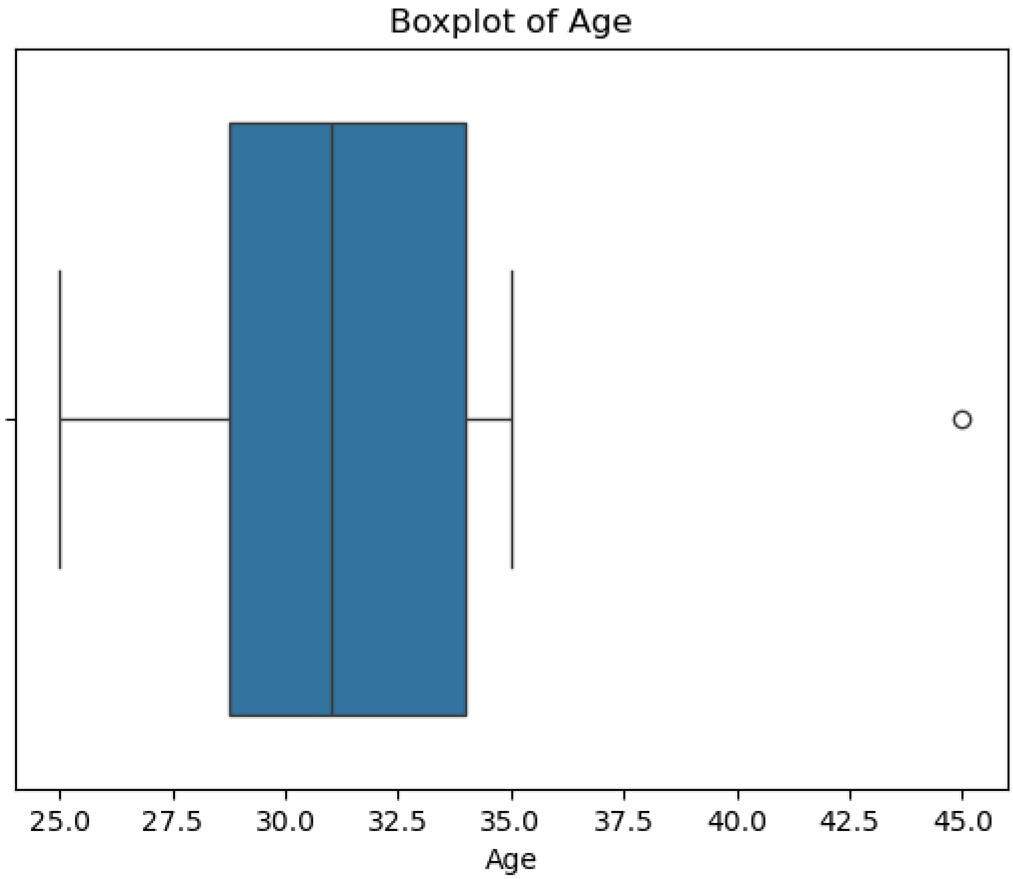
```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



In [55…

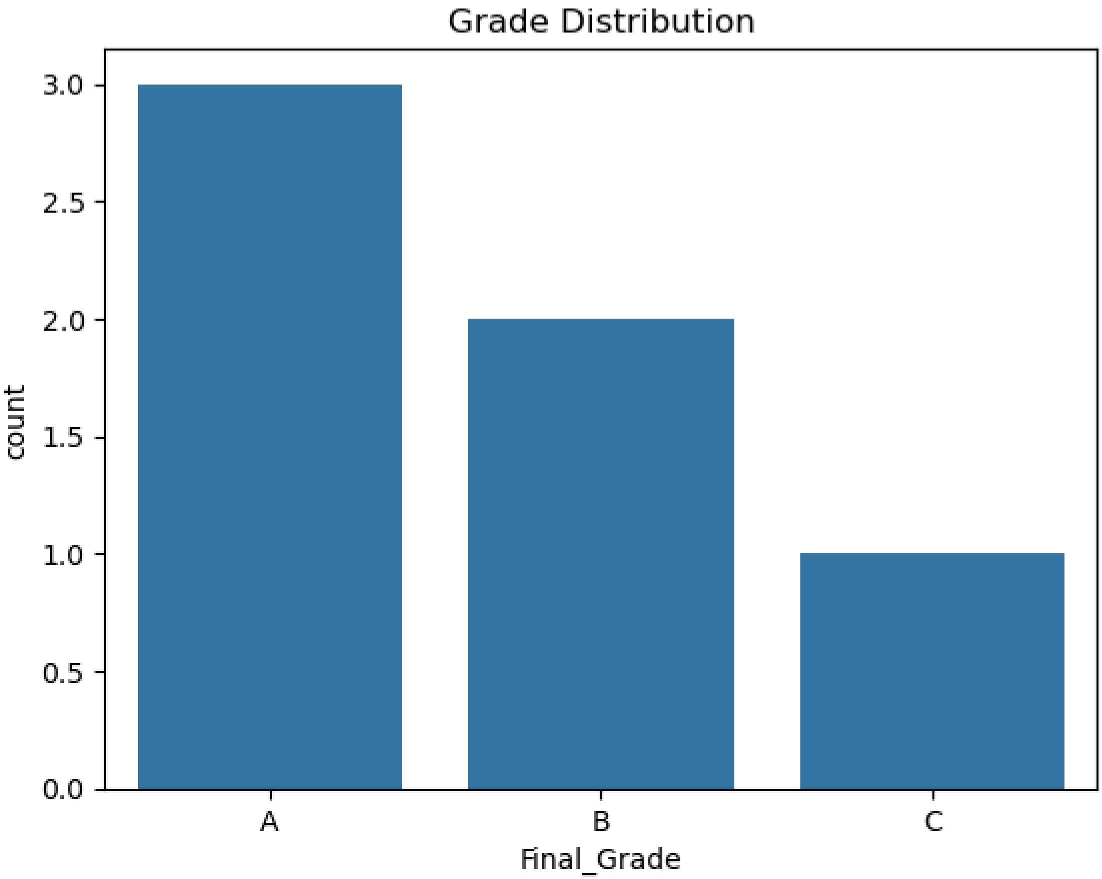
```
# Histogram for Age
df['Age'].hist(bins=10)
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()

# Box plot to spot outliers
sns.boxplot(data=df, x='Age')
plt.title("Boxplot of Age")
plt.show()
```

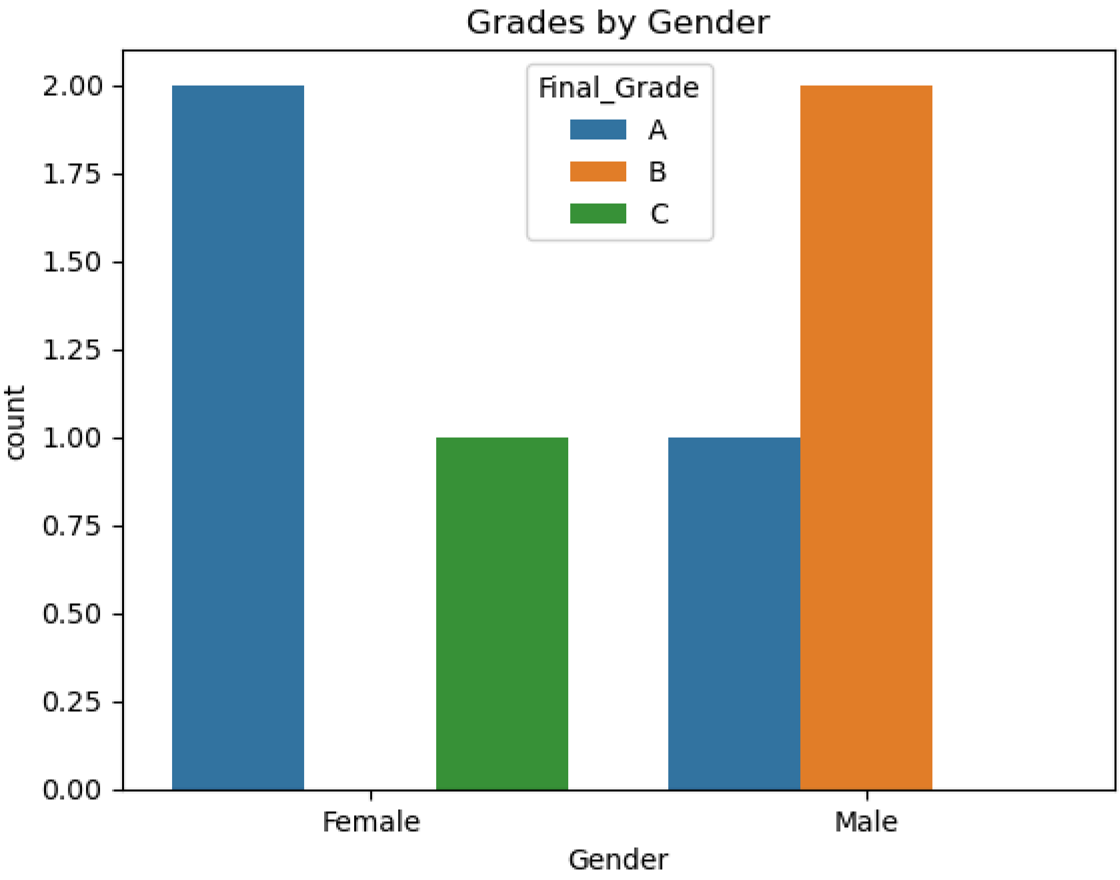

Age Distribution
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Count
25.0
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
Age


Boxplot of Age
25.0
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
Age

In [56…

```
sns.countplot(x='Final_Grade', data=df)
plt.title("Grade Distribution")
plt.show()

sns.countplot(x='Gender', hue='Final_Grade', data=df)
plt.title("Grades by Gender")
plt.show()
```





In [57…

```
data = {
    'Name': ['Sandesh', 'Saroj', 'Santosh', 'Shambhu', 'Prakash', 'Bimal
             'Ram', 'Shyam', 'Hari', 'Gita', 'Rita', 'Sita', 'Anita', 'K
    'Age': np.random.randint(15, 22, size=14),  # Now 14 ages to match 1
    'Class': np.random.choice(['IX', 'X', 'XI', 'XII'], size=14),
    'Faculty': np.random.choice(['Science', 'Management', 'Arts', 'Educa
    'Address': np.random.choice(['Kathmandu', 'Pokhara', 'Bhaktapur', 'L
                                 'Biratnagar', 'Dharan', 'Butwal'], size=
    'Grade': np.random.choice(['A+', 'A', 'B+', 'B', 'C+', 'C'], size=14
    'Gender': ['Male', 'Male', 'Male', 'Male', 'Male', 'Male',
               'Male', 'Male', 'Male', 'Female', 'Female', 'Female', 'Fe
    'Email': [name.lower().replace(' ', '') + '@school.edu.np' for name
    'Phone': ['98' + str(np.random.randint(10000000, 99999999)) for _ in
}

df = pd.DataFrame(data)
print(df)
```

```
       Name  Age Class     Faculty     Address Grade  Gender  \
0   Sandesh   21     X     Science  Biratnagar    A+    Male   
1     Saroj   20    IX     Science   Bhaktapur     A    Male   
2   Santosh   20     X  Management    Lalitpur     C    Male   
3   Shambhu   15   XII   Education     Pokhara    C+    Male   
4   Prakash   15     X  Management   Bhaktapur     A    Male   
5     Bimal   17    IX  Management   Bhaktapur     B    Male   
6       Ram   18     X  Management   Bhaktapur     C    Male   
7     Shyam   19    IX     Science     Pokhara    A+    Male   
8      Hari   17    XI   Education      Dharan    C+    Male   
9      Gita   18    IX   Education      Butwal    C+  Female   
10     Rita   16    IX  Management  Biratnagar    C+  Female   
11     Sita   21   XII     Science   Bhaktapur    A+  Female   
12    Anita   16    XI  Management     Pokhara    C+  Female   
13  Krishna   21   XII        Arts  Biratnagar    A+  Female   

                 Email       Phone  
0   sandesh@school.edu.np  9883693228  
1     saroj@school.edu.np  9854449472  
2   santosh@school.edu.np  9840939497  
3   shambhu@school.edu.np  9824465240  
4   prakash@school.edu.np  9891912195  
5     bimal@school.edu.np  9875648619  
6       ram@school.edu.np  9834722002  
7     shyam@school.edu.np  9832606286  
8      hari@school.edu.np  9872607344  
9      gita@school.edu.np  9883414030  
10     rita@school.edu.np  9876036580  
11     sita@school.edu.np  9838854765  
12    anita@school.edu.np  9833800166  
13  krishna@school.edu.np  9895928153  
```

In [58…

```python
# 1. Check for missing values
print(df.isnull().sum())

# 2. Drop rows with missing Name or Email (essential fields)
df = df.dropna(subset=['Name', 'Email'])

# 3. Fill missing Age with mean
df['Age'] = df['Age'].fillna(df['Age'].mean())

# 4. Fill missing Gender and Grade with mode (most frequent value)
df['Gender'] = df['Gender'].fillna(df['Gender'].mode()[0])
df['Grade'] = df['Grade'].fillna(df['Grade'].mode()[0])

# 5. Drop duplicate rows based on Name + Email
df = df.drop_duplicates(subset=['Name', 'Email'])

# 6. Convert Age to int (after filling NaNs)
```

```
df['Age'] = df['Age'].astype(int)

# 7. Rename columns (optional)
df = df.rename(columns={'Name': 'Student_Name', 'Grade': 'Final_Grade'})

# Final check
print(df)
```

```
Name       0
Age        0
Class      0
Faculty    0
Address    0
Grade      0
Gender     0
Email      0
Phone      0
dtype: int64
   Student_Name  Age Class     Faculty     Address Final_Grade  Gender  
\
0       Sandesh   21     X     Science  Biratnagar          A+    Male   
1         Saroj   20    IX     Science   Bhaktapur           A    Male   
2       Santosh   20     X  Management    Lalitpur           C    Male   
3       Shambhu   15   XII   Education     Pokhara          C+    Male   
4       Prakash   15     X  Management   Bhaktapur           A    Male   
5         Bimal   17    IX  Management   Bhaktapur           B    Male   
6           Ram   18     X  Management   Bhaktapur           C    Male   
7         Shyam   19    IX     Science     Pokhara          A+    Male   
8          Hari   17    XI   Education      Dharan          C+    Male   
9          Gita   18    IX   Education      Butwal          C+  Female   
10         Rita   16    IX  Management  Biratnagar          C+  Female   
11         Sita   21   XII     Science   Bhaktapur          A+  Female   
12        Anita   16    XI  Management     Pokhara          C+  Female   
13      Krishna   21   XII        Arts  Biratnagar          A+  Female   

                    Email       Phone  
0   sandesh@school.edu.np  9883693228  
1     saroj@school.edu.np  9854449472  
2   santosh@school.edu.np  9840939497  
3   shambhu@school.edu.np  9824465240  
4   prakash@school.edu.np  9891912195  
5     bimal@school.edu.np  9875648619  
6       ram@school.edu.np  9834722002  
7     shyam@school.edu.np  9832606286  
8      hari@school.edu.np  9872607344  
9      gita@school.edu.np  9883414030  
10     rita@school.edu.np  9876036580  
11     sita@school.edu.np  9838854765  
12    anita@school.edu.np  9833800166  
13  krishna@school.edu.np  9895928153  
```

In [60…

```
# 1. Create Age Group column
df['Age_Group'] = df['Age'].apply(lambda x: 'Teen' if x < 20 else 'Adult

# 2. Uppercase column values (for show)
```

```
df['Student_Name'] = df['Student_Name'].apply(lambda x: x.upper())
print(df)
```

```
   Student_Name  Age Class     Faculty     Address Final_Grade  Gender
\
0       SANDESH   21     X     Science  Biratnagar          A+    Male
1         SAROJ   20    IX     Science   Bhaktapur           A    Male
2       SANTOSH   20     X  Management    Lalitpur           C    Male
3       SHAMBHU   15   XII   Education     Pokhara          C+    Male
4       PRAKASH   15     X  Management   Bhaktapur           A    Male
5         BIMAL   17    IX  Management   Bhaktapur           B    Male
6           RAM   18     X  Management   Bhaktapur           C    Male
7         SHYAM   19    IX     Science     Pokhara          A+    Male
8          HARI   17    XI   Education      Dharan          C+    Male
9          GITA   18    IX   Education      Butwal          C+  Female
10         RITA   16    IX  Management  Biratnagar          C+  Female
11         SITA   21   XII     Science   Bhaktapur          A+  Female
12        ANITA   16    XI  Management     Pokhara          C+  Female
13      KRISHNA   21   XII        Arts  Biratnagar          A+  Female

                    Email       Phone Age_Group
0   sandesh@school.edu.np  9883693228     Adult
1     saroj@school.edu.np  9854449472     Adult
2   santosh@school.edu.np  9840939497     Adult
3   shambhu@school.edu.np  9824465240      Teen
4   prakash@school.edu.np  9891912195      Teen
5     bimal@school.edu.np  9875648619      Teen
6       ram@school.edu.np  9834722002      Teen
7     shyam@school.edu.np  9832606286      Teen
8      hari@school.edu.np  9872607344      Teen
9      gita@school.edu.np  9883414030      Teen
10     rita@school.edu.np  9876036580      Teen
11     sita@school.edu.np  9838854765     Adult
12    anita@school.edu.np  9833800166      Teen
13  krishna@school.edu.np  9895928153     Adult
```

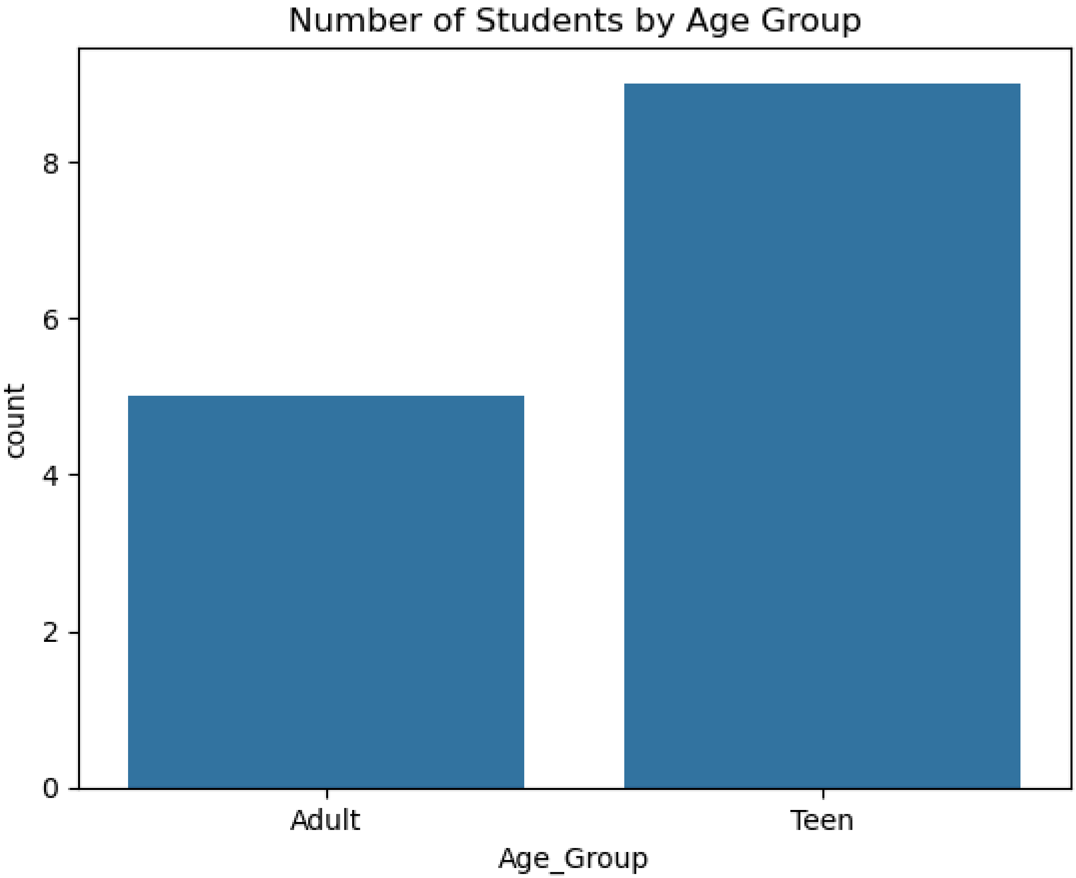
In [62…

```
# 1. Average age by gender
print(df.groupby('Gender')['Age'].mean())

# 2. Count of grades per gender
print(df.groupby('Gender')['Final_Grade'].value_counts())

print("\n")
# 3. How many students in each Age_Group
print(df['Age_Group'].value_counts())
```

```
Gender
Female    18.4
Male      18.0
Name: Age, dtype: float64
Gender  Final_Grade
Female  C+             3
        A+             2
Male    A              2
        A+             2
        C              2
        C+             2
        B              1
Name: count, dtype: int64


Age_Group
Teen     9
Adult    5
Name: count, dtype: int64
```

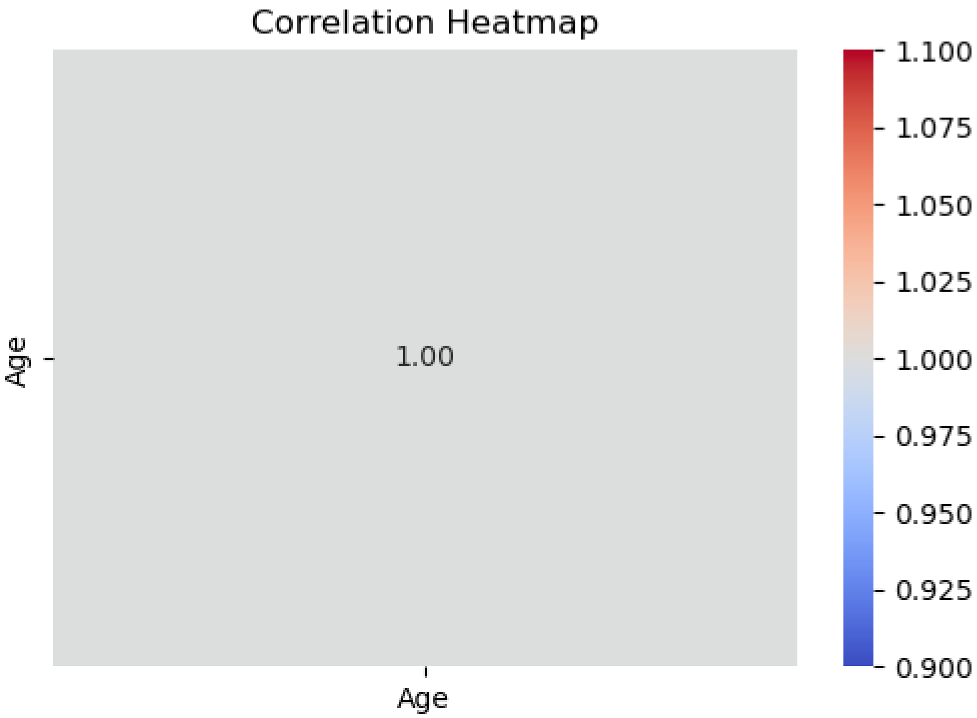
In [63…

```python
import seaborn as sns
import matplotlib.pyplot as plt

# Bar plot of student counts by Age_Group
sns.countplot(x='Age_Group', data=df)
plt.title("Number of Students by Age Group")
plt.show()
```



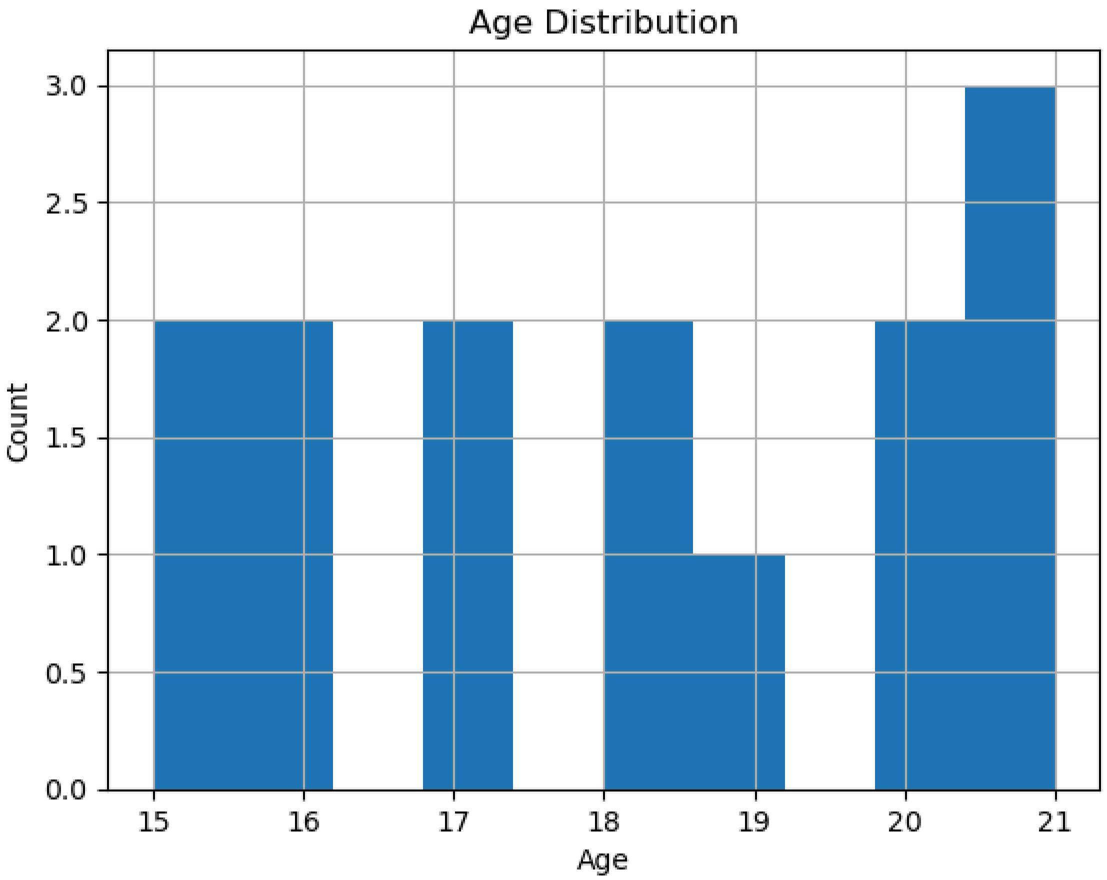
In [64…

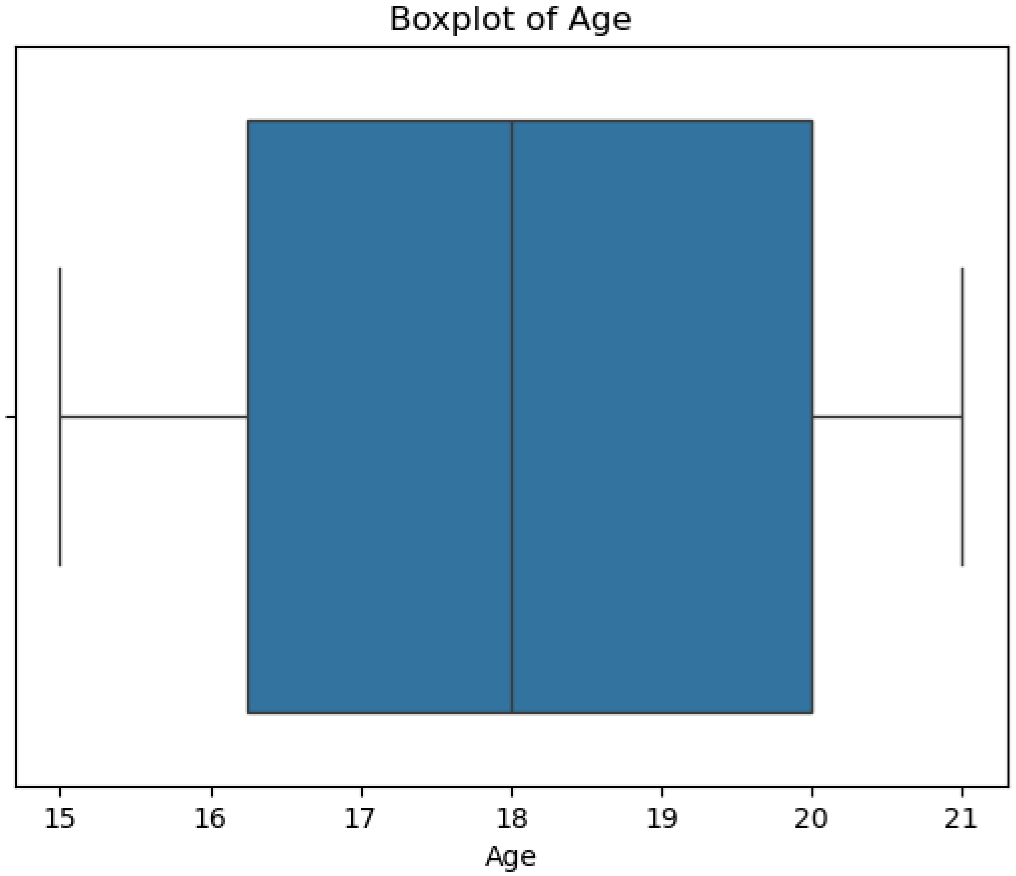
```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



In [65…

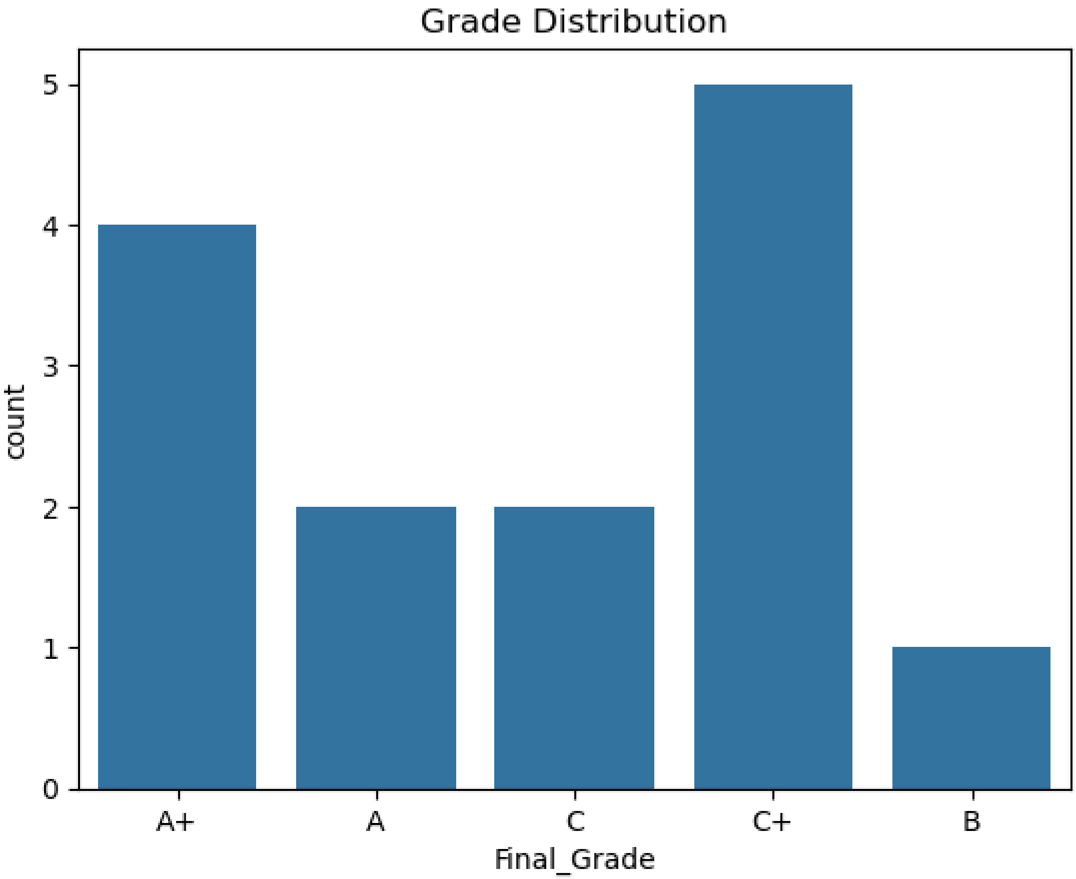
```
# Histogram for Age
df['Age'].hist(bins=10)
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()

# Box plot to spot outliers
sns.boxplot(data=df, x='Age')
plt.title("Boxplot of Age")
plt.show()
```

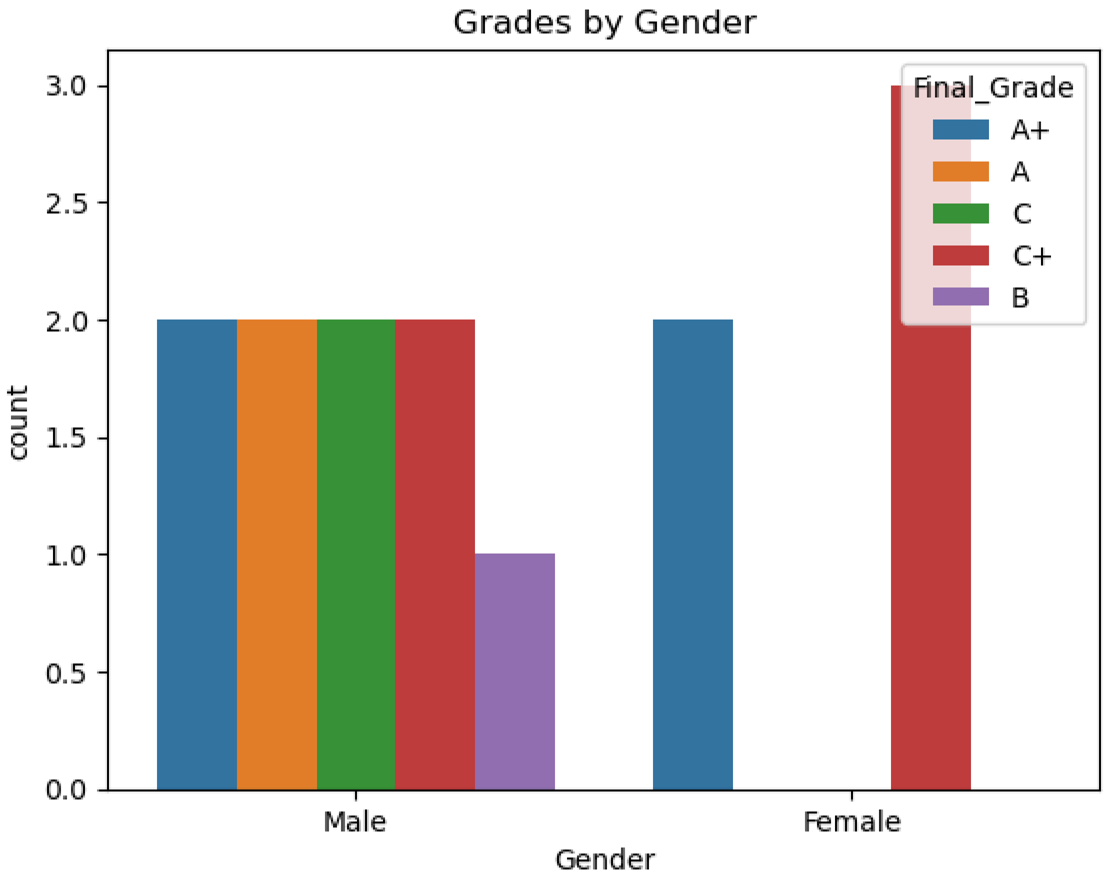

Age Distribution
Count
3.0
2.5
2.0
1.5
1.0
0.5
0.0
15
16
17
18
19
20
21
Age


Boxplot of Age
15
16
17
18
19
20
21
Age

In [66…

```
sns.countplot(x='Final_Grade', data=df)
plt.title("Grade Distribution")
plt.show()

sns.countplot(x='Gender', hue='Final_Grade', data=df)
plt.title("Grades by Gender")
plt.show()
```





In [69…

```
# Average age per grade
print(df.groupby('Final_Grade')['Age'].mean())


print("\n")

# Count of students by gender and grade
print(df.groupby(['Gender', 'Final_Grade']).size())
```

```
Final_Grade
A     17.5
A+    20.5
B     17.0
C     19.0
C+    16.4
Name: Age, dtype: float64


Gender  Final_Grade
Female  A+             2
        C+             3
Male    A              2
        A+             2
        B              1
        C              2
        C+             2
dtype: int64
```

In [70…

```
# Average and max age per gender
print(df.groupby('Gender')['Age'].agg(['mean', 'max', 'min']))
```

```
        mean  max  min
Gender
Female  18.4   21   16
Male    18.0   21   15
```

In [72…

```
def age_group(age):
    if age < 20:
        return "Teen"
    elif age < 30:
        return "Young Adult"
    else:
        return "Adult"

df['Age_Category'] = df['Age'].apply(age_group)
print(df)
```

```
   Student_Name  Age Class     Faculty     Address Final_Grade  Gender  
\
0       SANDESH   21     X     Science  Biratnagar          A+    Male
1         SAROJ   20    IX     Science   Bhaktapur           A    Male
2       SANTOSH   20     X  Management    Lalitpur           C    Male
3       SHAMBHU   15   XII   Education     Pokhara          C+    Male
4       PRAKASH   15     X  Management   Bhaktapur           A    Male
5         BIMAL   17    IX  Management   Bhaktapur           B    Male
6           RAM   18     X  Management   Bhaktapur           C    Male
7         SHYAM   19    IX     Science     Pokhara          A+    Male
8          HARI   17    XI   Education      Dharan          C+    Male
9          GITA   18    IX   Education      Butwal          C+  Female
10         RITA   16    IX  Management  Biratnagar          C+  Female
11         SITA   21   XII     Science   Bhaktapur          A+  Female
12        ANITA   16    XI  Management     Pokhara          C+  Female
13      KRISHNA   21   XII        Arts  Biratnagar          A+  Female

                    Email       Phone Age_Group Age_Category
0   sandesh@school.edu.np  9883693228     Adult  Young Adult
1     saroj@school.edu.np  9854449472     Adult  Young Adult
2   santosh@school.edu.np  9840939497     Adult  Young Adult
3   shambhu@school.edu.np  9824465240      Teen         Teen
4   prakash@school.edu.np  9891912195      Teen         Teen
5     bimal@school.edu.np  9875648619      Teen         Teen
6       ram@school.edu.np  9834722002      Teen         Teen
7     shyam@school.edu.np  9832606286      Teen         Teen
8      hari@school.edu.np  9872607344      Teen         Teen
9      gita@school.edu.np  9883414030      Teen         Teen
10     rita@school.edu.np  9876036580      Teen         Teen
11     sita@school.edu.np  9838854765     Adult  Young Adult
12    anita@school.edu.np  9833800166      Teen         Teen
13  krishna@school.edu.np  9895928153     Adult  Young Adult
```

In [74…

```python
# Combine name and email into a new column
df['Identity'] = df.apply(lambda row: f"{row['Student_Name']} <{row['Ema
print(df)
```

```
   Student_Name  Age Class     Faculty     Address Final_Grade  Gender  
\
0       SANDESH   21     X     Science  Biratnagar          A+    Male
1         SAROJ   20    IX     Science   Bhaktapur           A    Male
2       SANTOSH   20     X  Management    Lalitpur           C    Male
3       SHAMBHU   15   XII   Education     Pokhara          C+    Male
4       PRAKASH   15     X  Management   Bhaktapur           A    Male
5         BIMAL   17    IX  Management   Bhaktapur           B    Male
6           RAM   18     X  Management   Bhaktapur           C    Male
7         SHYAM   19    IX     Science     Pokhara          A+    Male
8          HARI   17    XI   Education      Dharan          C+    Male
9          GITA   18    IX   Education      Butwal          C+  Female
10         RITA   16    IX  Management  Biratnagar          C+  Female
11         SITA   21   XII     Science   Bhaktapur          A+  Female
12        ANITA   16    XI  Management     Pokhara          C+  Female
13      KRISHNA   21   XII        Arts  Biratnagar          A+  Female

                    Email       Phone Age_Group Age_Category  \
0   sandesh@school.edu.np  9883693228     Adult  Young Adult   
1     saroj@school.edu.np  9854449472     Adult  Young Adult   
2   santosh@school.edu.np  9840939497     Adult  Young Adult   
3   shambhu@school.edu.np  9824465240      Teen         Teen   
4   prakash@school.edu.np  9891912195      Teen         Teen   
5     bimal@school.edu.np  9875648619      Teen         Teen   
6       ram@school.edu.np  9834722002      Teen         Teen   
7     shyam@school.edu.np  9832606286      Teen         Teen   
8      hari@school.edu.np  9872607344      Teen         Teen   
9      gita@school.edu.np  9883414030      Teen         Teen   
10     rita@school.edu.np  9876036580      Teen         Teen   
11     sita@school.edu.np  9838854765     Adult  Young Adult   
12    anita@school.edu.np  9833800166      Teen         Teen   
13  krishna@school.edu.np  9895928153     Adult  Young Adult   

                           Identity  
0   SANDESH <sandesh@school.edu.np>  
1       SAROJ <saroj@school.edu.np>  
2   SANTOSH <santosh@school.edu.np>  
3   SHAMBHU <shambhu@school.edu.np>  
4   PRAKASH <prakash@school.edu.np>  
5       BIMAL <bimal@school.edu.np>  
6           RAM <ram@school.edu.np>  
7       SHYAM <shyam@school.edu.np>  
8         HARI <hari@school.edu.np>  
9         GITA <gita@school.edu.np>  
10        RITA <rita@school.edu.np>  
11        SITA <sita@school.edu.np>  
12      ANITA <anita@school.edu.np>  
13  KRISHNA <krishna@school.edu.np>  
```

In [75…

```
# Count number of missing values per column
missing_per_col = df.apply(lambda col: col.isnull().sum(), axis=0)
print(missing_per_col)
```

```
Student_Name    0
Age             0
Class           0
Faculty         0
Address         0
Final_Grade     0
Gender          0
Email           0
Phone           0
Age_Group       0
Age_Category    0
Identity        0
dtype: int64
```

In [76…

```
max_lengths = df.select_dtypes(include='object').apply(lambda col: col.s
print(max_lengths)
```

```
Student_Name     7
Class            3
Faculty         10
Address         10
Final_Grade      2
Gender           6
Email           21
Phone           10
Age_Group        5
Age_Category    11
Identity        31
dtype: int64
```

Pandas for Machine Learning (How does the Machine Learning models can use Data)

In [78…

```
# Confirm no missing data remains
print(df.isnull().sum())
```

```
Student_Name    0
Age             0
Class           0
Faculty         0
Address         0
Final_Grade     0
Gender          0
Email           0
Phone           0
Age_Group       0
Age_Category    0
Identity        0
dtype: int64
```

In [83…

```
print(df.columns)
```

```
Index(['Student_Name', 'Age', 'Class', 'Faculty', 'Address', 'Gender',
'Email',
       'Phone', 'Age_Group', 'Age_Category', 'Identity', 'Gender_Encode
d',
       'Final_Grade_A+', 'Final_Grade_B', 'Final_Grade_C', 'Final_Grade_
C+'],
      dtype='object')
```

In [87…

```
print(df.columns.tolist())
```

```
['Student_Name', 'Age', 'Class', 'Faculty', 'Address', 'Gender', 'Emai
l', 'Phone', 'Age_Group', 'Age_Category', 'Identity', 'Gender_Encoded',
'Final_Grade_A+', 'Final_Grade_B', 'Final_Grade_C', 'Final_Grade_C+']
```

In [88…

```
df['Grade'] = np.random.choice(['A', 'B', 'C', 'D'], size=len(df))
```

In [89…

```
df = pd.get_dummies(df, columns=['Grade'], drop_first=True)
```

In [90…

```
print(df)
```

```
   Student_Name  Age Class     Faculty     Address  Gender  \
0       SANDESH   21     X     Science  Biratnagar    Male   
1         SAROJ   20    IX     Science   Bhaktapur    Male   
2       SANTOSH   20     X  Management    Lalitpur    Male   
3       SHAMBHU   15   XII   Education     Pokhara    Male   
4       PRAKASH   15     X  Management   Bhaktapur    Male   
5         BIMAL   17    IX  Management   Bhaktapur    Male   
6           RAM   18     X  Management   Bhaktapur    Male   
7         SHYAM   19    IX     Science     Pokhara    Male   
8          HARI   17    XI   Education      Dharan    Male   
9          GITA   18    IX   Education      Butwal  Female   
10         RITA   16    IX  Management  Biratnagar  Female   
11         SITA   21   XII     Science   Bhaktapur  Female   
12        ANITA   16    XI  Management     Pokhara  Female   
13      KRISHNA   21   XII        Arts  Biratnagar  Female   

                    Email       Phone Age_Group Age_Category  \
0   sandesh@school.edu.np  9883693228     Adult  Young Adult   
1     saroj@school.edu.np  9854449472     Adult  Young Adult   
2   santosh@school.edu.np  9840939497     Adult  Young Adult   
3   shambhu@school.edu.np  9824465240      Teen         Teen   
4   prakash@school.edu.np  9891912195      Teen         Teen   
5     bimal@school.edu.np  9875648619      Teen         Teen   
6       ram@school.edu.np  9834722002      Teen         Teen   
7     shyam@school.edu.np  9832606286      Teen         Teen   
8      hari@school.edu.np  9872607344      Teen         Teen   
9      gita@school.edu.np  9883414030      Teen         Teen   
10     rita@school.edu.np  9876036580      Teen         Teen   
11     sita@school.edu.np  9838854765     Adult  Young Adult   
12    anita@school.edu.np  9833800166      Teen         Teen   
13  krishna@school.edu.np  9895928153     Adult  Young Adult   

                           Identity  Gender_Encoded  Final_Grade_A+  \
0   SANDESH <sandesh@school.edu.np>               1            True   
1       SAROJ <saroj@school.edu.np>               1           False   
2   SANTOSH <santosh@school.edu.np>               1           False   
3   SHAMBHU <shambhu@school.edu.np>               1           False   
4   PRAKASH <prakash@school.edu.np>               1           False   
5       BIMAL <bimal@school.edu.np>               1           False   
6           RAM <ram@school.edu.np>               1           False   
7       SHYAM <shyam@school.edu.np>               1            True   
8         HARI <hari@school.edu.np>               1           False   
9         GITA <gita@school.edu.np>               0           False   
10        RITA <rita@school.edu.np>               0           False   
11        SITA <sita@school.edu.np>               0            True   
12      ANITA <anita@school.edu.np>               0           False   
13  KRISHNA <krishna@school.edu.np>               0            True   
```

```
    Final_Grade_B  Final_Grade_C  Final_Grade_C+  Grade_B  Grade_C  Grad
e_D
0           False          False           False    False     True     Fa
lse
1           False          False           False    False    False      T
rue
2           False           True           False    False     True     Fa
lse
3           False          False            True     True    False     Fa
lse
4           False          False           False    False    False      T
rue
5            True          False           False    False     True     Fa
lse
6           False           True           False     True    False     Fa
lse
7           False          False           False    False    False      T
rue
8           False          False            True    False     True     Fa
lse
9           False          False            True    False    False      T
rue
10          False          False            True    False    False      T
rue
11          False          False           False    False    False     Fa
lse
12          False          False            True    False    False      T
rue
13          False          False           False    False    False      T
rue
```

Step 7 Exporig to diffrent file Extension.like csv ,pdf,excel,sql

In [91…

```python
import os

folder_path = r"C:\Users\DELL\Desktop\A\day 6"  # Change to your desirea
os.makedirs(folder_path, exist_ok=True)          # Create folder if not

file_name = "cleaned_data.csv"
file_path = os.path.join(folder_path, file_name)

df.to_csv(file_path, index=False)                # Export CSV without r
print(f"CSV saved successfully at: {file_path}")
```

```
CSV saved successfully at: C:\Users\DELL\Desktop\A\day 6\cleaned_data.cs
v
```

Thanks you for Watching this ,This is my learning journey of AI and ML with strong foundation. All rights reserved Sandesh Bhatta 2025.
github:https://github.com/sandeshbhatta495/AI.git website:www.sandeshbhatta495.com.np