P8_2347138

September 15, 2023

```python
import pandas as pd

# Assuming you have a CSV file named 'your_dataset.csv'
data = pd.read_csv("D:\PRATHAM FILES\Web design Project\Python new\Student.csv")

# Use describe() to get summary statistics
description = data.describe()

# Get the shape of the data frame
shape = data.shape

# Get information about the data frame
info = data.info()

# Print the results
print("Summary Statistics:")
print(description)
print("\nShape of the Data Frame:")
print(shape)
print("\nInformation about the Data Frame:")
print(info)

# Check for missing values
missing_values = data.isnull().sum()

# Handle missing values in four different ways
# For example, you can drop rows with missing values:
data_drop = data.dropna()

# Or replace missing values with a specific value:
data_replace = data.fillna("replacement_value")

# Or forward fill missing values:
data_fill_forward = data.ffill()

# Or backward fill missing values:
data_fill_backward = data.bfill()
```

```python
# Print the results
print("Missing Values:")
print(missing_values)
print("\nData Frame after Dropping Rows with Missing Values:")
print(data_drop)
print("\nData Frame after Replacing Missing Values:")
print(data_replace)
print("\nData Frame after Forward Filling Missing Values:")
print(data_fill_forward)
print("\nData Frame after Backward Filling Missing Values:")
print(data_fill_backward)
# For example, group by the 'Nationality' column and calculate the mean of
  'raisedhands' for each group
grouped_data = data.groupby("Nationality")["raisedhands"].mean()

# Print the results
print("Grouped Data:")
print(grouped_data)
# Select 20 random samples from the original data frame
random_samples = data.sample(n=20)

# Create a hierarchical index
hierarchical_indexed_data = random_samples.set_index(["Gender", "Nationality"])

# Print the results
print("Randomly Selected Samples:")
print(random_samples)
print("\nData Frame with Hierarchical Index:")
print(hierarchical_indexed_data)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            98 non-null     object
 1   NationalITy       99 non-null     object
 2   StageID           98 non-null     object
 3   GradeID           98 non-null     object
 4   Semester          99 non-null     object
 5   Relation          98 non-null     object
 6   raisedhands       99 non-null     int64
 7   VisITedResources  99 non-null     int64
 8   AnnouncementsView 99 non-null     int64
 9   Discussion        97 non-null     float64
dtypes: float64(1), int64(3), object(6)
memory usage: 7.9+ KB
```

```
Summary Statistics:
       raisedhands  VisITedResources  AnnouncementsView  Discussion
count   99.000000         99.000000          99.000000   97.000000
mean    29.303030         33.242424          19.252525   45.041237
std     24.004445         30.367508          15.891432   27.065475
min      0.000000          0.000000           0.000000    1.000000
25%     10.000000         10.000000           6.000000   20.000000
50%     20.000000         20.000000          16.000000   40.000000
75%     49.500000         60.000000          31.000000   70.000000
max    100.000000         90.000000          70.000000   99.000000


Shape of the Data Frame:
(99, 10)


Information about the Data Frame:
None
Missing Values:
gender                1
NationalITy           0
StageID               1
GradeID               1
Semester              0
Relation              1
raisedhands           0
VisITedResources      0
AnnouncementsView     0
Discussion            2
dtype: int64


Data Frame after Dropping Rows with Missing Values:
    gender NationalITy        StageID GradeID Semester Relation  raisedhands  \
0        M          KW    lowerlevel    G-04        F   Father           15
1        M          KW    lowerlevel    G-04        F   Father           20
7        M          KW   MiddleSchool   G-07        F   Father           50
9        F          KW   MiddleSchool   G-07        F   Father           70
10       M          KW   MiddleSchool   G-07        F   Father           50
..     ...         ...            ...     ...      ...      ...          ...
94       F          KW    lowerlevel    G-02        F   Father           80
95       F          KW    lowerlevel    G-02        F   Father          100
96       F          KW    lowerlevel    G-02        F   Father           14
97       F          KW    lowerlevel    G-02        F   Father            6
98       F          KW    lowerlevel    G-02        F   Father           10


    VisITedResources  AnnouncementsView  Discussion
0                 16                  2        20.0
1                 20                  3        25.0
7                 10                 15        22.0
9                 80                 25        70.0
```

```
10                  88                  30          80.0
..                 …                   …           …
94                  44                  35          10.0
95                  80                   2          70.0
96                  60                  11          75.0
97                   2                  38          12.0
98                   3                   0          30.0

[93 rows x 10 columns]

Data Frame after Replacing Missing Values:
               gender NationalITy     StageID GradeID Semester  \
0                   M          KW  lowerlevel    G-04        F
1                   M          KW  lowerlevel    G-04        F
2                   M          KW  lowerlevel    G-04        F
3                   M          KW  lowerlevel    G-04        F
4   replacement_value          KW  lowerlevel    G-04        F
..                …           …          …       …        …
94                  F          KW  lowerlevel    G-02        F
95                  F          KW  lowerlevel    G-02        F
96                  F          KW  lowerlevel    G-02        F
97                  F          KW  lowerlevel    G-02        F
98                  F          KW  lowerlevel    G-02        F

             Relation  raisedhands  VisITedResources  AnnouncementsView  \
0              Father           15                16                  2
1              Father           20                20                  3
2              Father           10                 7                  0
3   replacement_value           30                25                  5
4              Father           40                50                 12
..                …            …                 …                  …
94             Father           80                44                 35
95             Father          100                80                  2
96             Father           14                60                 11
97             Father            6                 2                 38
98             Father           10                 3                  0

           Discussion
0                20.0
1                25.0
2   replacement_value
3                35.0
4                50.0
..                …
94               10.0
95               70.0
96               75.0
97               12.0
```

```
98                  30.0

[99 rows x 10 columns]

Data Frame after Forward Filling Missing Values:
    gender NationalITy      StageID GradeID Semester Relation  raisedhands  \
0        M          KW  lowerlevel    G-04        F   Father           15
1        M          KW  lowerlevel    G-04        F   Father           20
2        M          KW  lowerlevel    G-04        F   Father           10
3        M          KW  lowerlevel    G-04        F   Father           30
4        M          KW  lowerlevel    G-04        F   Father           40
..     ...         ...         ...     ...      ...      ...          ...
94       F          KW  lowerlevel    G-02        F   Father           80
95       F          KW  lowerlevel    G-02        F   Father          100
96       F          KW  lowerlevel    G-02        F   Father           14
97       F          KW  lowerlevel    G-02        F   Father            6
98       F          KW  lowerlevel    G-02        F   Father           10

    VisITedResources  AnnouncementsView  Discussion
0                 16                  2        20.0
1                 20                  3        25.0
2                  7                  0        25.0
3                 25                  5        35.0
4                 50                 12        50.0
..               ...                ...         ...
94                44                 35        10.0
95                80                  2        70.0
96                60                 11        75.0
97                 2                 38        12.0
98                 3                  0        30.0

[99 rows x 10 columns]

Data Frame after Backward Filling Missing Values:
    gender NationalITy      StageID GradeID Semester Relation  raisedhands  \
0        M          KW  lowerlevel    G-04        F   Father           15
1        M          KW  lowerlevel    G-04        F   Father           20
2        M          KW  lowerlevel    G-04        F   Father           10
3        M          KW  lowerlevel    G-04        F   Father           30
4        F          KW  lowerlevel    G-04        F   Father           40
..     ...         ...         ...     ...      ...      ...          ...
94       F          KW  lowerlevel    G-02        F   Father           80
95       F          KW  lowerlevel    G-02        F   Father          100
96       F          KW  lowerlevel    G-02        F   Father           14
97       F          KW  lowerlevel    G-02        F   Father            6
98       F          KW  lowerlevel    G-02        F   Father           10

    VisITedResources  AnnouncementsView  Discussion
```

```
0                   16                   2          20.0
1                   20                   3          25.0
2                    7                   0          35.0
3                   25                   5          35.0
4                   50                  12          50.0
..                 ...                 ...           ...
94                  44                  35          10.0
95                  80                   2          70.0
96                  60                  11          75.0
97                   2                  38          12.0
98                   3                   0          30.0

[99 rows x 10 columns]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
d:\PRATHAM FILES\Web design Project\Python new\matplotlib.ipynb Cell 2 line 5
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
  Python%20new/matplotlib.ipynb#X10sZmlsZQ%3D%3D?line=48'>49</a>
  print(data_fill_backward)
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
  Python%20new/matplotlib.ipynb#X10sZmlsZQ%3D%3D?line=49'>50</a> # For example,
  group by the 'Nationality' column and calculate the mean of 'raisedhands' for
  each group
---> <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
  Python%20new/matplotlib.ipynb#X10sZmlsZQ%3D%3D?line=50'>51</a> grouped_data =
  data.groupby("Nationality")["raisedhands"].mean()
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
  Python%20new/matplotlib.ipynb#X10sZmlsZQ%3D%3D?line=52'>53</a> # Print the
  results
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
  Python%20new/matplotlib.ipynb#X10sZmlsZQ%3D%3D?line=53'>54</a> print("Grouped
  Data:")

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
  11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\frame.
  py:8872, in DataFrame.groupby(self, by, axis, level, as_index, sort,
  group_keys, observed, dropna)
   8869 if level is None and by is None:
   8870     raise TypeError("You have to supply one of 'by' and 'level'")
-> 8872 return DataFrameGroupBy(
   8873     obj=self,
   8874     keys=by,
   8875     axis=axis,
   8876     level=level,
   8877     as_index=as_index,
   8878     sort=sort,
   8879     group_keys=group_keys,
   8880     observed=observed,
```

```
8881       dropna=dropna,
8882 )

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
 ↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\groupby\grou
 ↪py:1273, in GroupBy.__init__(self, obj, keys, axis, level, grouper,␣
 ↪exclusions, selection, as_index, sort, group_keys, observed, dropna)
   1270 self.dropna = dropna
   1272 if grouper is None:
-> 1273     grouper, exclusions, obj = get_grouper(
   1274         obj,
   1275         keys,
   1276         axis=axis,
   1277         level=level,
   1278         sort=sort,
   1279         observed=False if observed is lib.no_default else observed,
   1280         dropna=self.dropna,
   1281     )
   1283 if observed is lib.no_default:
   1284     if any(ping._passed_categorical for ping in grouper.groupings):

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
 ↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\groupby\grou
 ↪py:1009, in get_grouper(obj, key, axis, level, sort, observed, validate,␣
 ↪dropna)
   1007         in_axis, level, gpr = False, gpr, None
   1008     else:
-> 1009         raise KeyError(gpr)
   1010 elif isinstance(gpr, Grouper) and gpr.key is not None:
   1011     # Add key to exclusions
   1012     exclusions.add(gpr.key)

KeyError: 'Nationality'
```

```python
import pandas as pd

# Step 1: Read the CSV file and create and understand the data frame using␣
 ↪describe(), shape, and info()
# Replace 'your_dataset.csv' with the actual file path of your dataset.
data = pd.read_csv("D:\PRATHAM FILES\Web design Project\Python new\Student.csv")

# Use describe() to get summary statistics
description = data.describe()

# Get the shape of the data frame
shape = data.shape
```

```python
# Get information about the data frame
info = data.info()

# Print the results
print("Summary Statistics:")
print(description)
print("\nShape of the Data Frame:")
print(shape)
print("\nInformation about the Data Frame:")
print(info)

# Step 2: Find if any missing values (null values) are in the data and print
  ↪the data frame
# Check for missing values
missing_values = data.isnull().sum()

# Print the results
print("Missing Values:")
print(missing_values)

# Step 3: Handle all the rows with missing data in four different ways (delete,
  ↪replace, fill, bfill), and print the data frame
# Handle missing values in four different ways:
# 1. Drop rows with missing values
data_drop = data.dropna()

# 2. Replace missing values with a specific value
data_replace = data.fillna('replacement_value')

# 3. Forward fill missing values
data_fill_forward = data.ffill()

# 4. Backward fill missing values
data_fill_backward = data.bfill()

# Print the results
print("\nData Frame after Dropping Rows with Missing Values:")
print(data_drop)
print("\nData Frame after Replacing Missing Values:")
print(data_replace)
print("\nData Frame after Forward Filling Missing Values:")
print(data_fill_forward)
print("\nData Frame after Backward Filling Missing Values:")
print(data_fill_backward)

# Step 4: Filter based on any column using groupby()
```

```python
# For example, group by the 'Nationality' column and calculate the mean of␣
 ↪'raisedhands' for each group
grouped_data = data.groupby("Nationality")["raisedhands"].mean()

# Print the results
print("Grouped Data:")
print(grouped_data)

# Step 5: Select 20 samples randomly and create a data frame with Hierarchical␣
 ↪Index
# Select 20 random samples from the original data frame
random_samples = data.sample(n=20)

# Create a hierarchical index
hierarchical_indexed_data = random_samples.set_index(['Gender', 'Nationality'])

# Print the results
print("Randomly Selected Samples:")
print(random_samples)
print("\nData Frame with Hierarchical Index:")
print(hierarchical_indexed_data)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            98 non-null     object
 1   NationalITy       99 non-null     object
 2   StageID           98 non-null     object
 3   GradeID           98 non-null     object
 4   Semester          99 non-null     object
 5   Relation          98 non-null     object
 6   raisedhands       99 non-null     int64
 7   VisITedResources  99 non-null     int64
 8   AnnouncementsView 99 non-null     int64
 9   Discussion        97 non-null     float64
dtypes: float64(1), int64(3), object(6)
memory usage: 7.9+ KB
Summary Statistics:
       raisedhands  VisITedResources  AnnouncementsView  Discussion
count    99.000000         99.000000          99.000000   97.000000
mean     29.303030         33.242424          19.252525   45.041237
std      24.004445         30.367508          15.891432   27.065475
min       0.000000          0.000000           0.000000    1.000000
25%      10.000000         10.000000           6.000000   20.000000
50%      20.000000         20.000000          16.000000   40.000000
```

```
75%       49.500000        60.000000       31.000000   70.000000
max      100.000000        90.000000       70.000000   99.000000


Shape of the Data Frame:
(99, 10)


Information about the Data Frame:
None
Missing Values:
gender               1
NationalITy          0
StageID              1
GradeID              1
Semester             0
Relation             1
raisedhands          0
VisITedResources     0
AnnouncementsView    0
Discussion           2
dtype: int64


Data Frame after Dropping Rows with Missing Values:
    gender NationalITy        StageID GradeID Semester Relation  raisedhands  \
0        M          KW     lowerlevel    G-04        F   Father           15
1        M          KW     lowerlevel    G-04        F   Father           20
7        M          KW   MiddleSchool    G-07        F   Father           50
9        F          KW   MiddleSchool    G-07        F   Father           70
10       M          KW   MiddleSchool    G-07        F   Father           50
..      ...         ...            ...     ...      ...      ...          ...
94       F          KW     lowerlevel    G-02        F   Father           80
95       F          KW     lowerlevel    G-02        F   Father          100
96       F          KW     lowerlevel    G-02        F   Father           14
97       F          KW     lowerlevel    G-02        F   Father            6
98       F          KW     lowerlevel    G-02        F   Father           10

    VisITedResources  AnnouncementsView  Discussion
0                 16                  2        20.0
1                 20                  3        25.0
7                 10                 15        22.0
9                 80                 25        70.0
10                88                 30        80.0
..               ...                ...         ...
94                44                 35        10.0
95                80                  2        70.0
96                60                 11        75.0
97                 2                 38        12.0
98                 3                  0        30.0
```

```
[93 rows x 10 columns]

Data Frame after Replacing Missing Values:
              gender NationalITy    StageID GradeID Semester  \
0                  M          KW  lowerlevel   G-04       F
1                  M          KW  lowerlevel   G-04       F
2                  M          KW  lowerlevel   G-04       F
3                  M          KW  lowerlevel   G-04       F
4   replacement_value          KW  lowerlevel   G-04       F
..               ...         ...        ...    ...      ...
94                 F          KW  lowerlevel   G-02       F
95                 F          KW  lowerlevel   G-02       F
96                 F          KW  lowerlevel   G-02       F
97                 F          KW  lowerlevel   G-02       F
98                 F          KW  lowerlevel   G-02       F

              Relation  raisedhands  VisITedResources  AnnouncementsView  \
0               Father           15                16                  2
1               Father           20                20                  3
2               Father           10                 7                  0
3    replacement_value           30                25                  5
4               Father           40                50                 12
..                 ...          ...               ...                ...
94              Father           80                44                 35
95              Father          100                80                  2
96              Father           14                60                 11
97              Father            6                 2                 38
98              Father           10                 3                  0

             Discussion
0                  20.0
1                  25.0
2     replacement_value
3                  35.0
4                  50.0
..                  ...
94                 10.0
95                 70.0
96                 75.0
97                 12.0
98                 30.0

[99 rows x 10 columns]

Data Frame after Forward Filling Missing Values:
   gender NationalITy    StageID GradeID Semester Relation  raisedhands  \
0       M          KW  lowerlevel   G-04       F   Father           15
1       M          KW  lowerlevel   G-04       F   Father           20
```

```
2       M          KW    lowerlevel    G-04       F    Father           10
3       M          KW    lowerlevel    G-04       F    Father           30
4       M          KW    lowerlevel    G-04       F    Father           40
..     …          …         …         …         …        …              …
94      F          KW    lowerlevel    G-02       F    Father           80
95      F          KW    lowerlevel    G-02       F    Father          100
96      F          KW    lowerlevel    G-02       F    Father           14
97      F          KW    lowerlevel    G-02       F    Father            6
98      F          KW    lowerlevel    G-02       F    Father           10


     VisITedResources   AnnouncementsView   Discussion
0                 16                   2         20.0
1                 20                   3         25.0
2                  7                   0         25.0
3                 25                   5         35.0
4                 50                  12         50.0
..                …                   …           …
94                44                  35         10.0
95                80                   2         70.0
96                60                  11         75.0
97                 2                  38         12.0
98                 3                   0         30.0


[99 rows x 10 columns]

Data Frame after Backward Filling Missing Values:
   gender NationalITy     StageID GradeID Semester Relation  raisedhands  \
0       M          KW    lowerlevel    G-04       F    Father           15
1       M          KW    lowerlevel    G-04       F    Father           20
2       M          KW    lowerlevel    G-04       F    Father           10
3       M          KW    lowerlevel    G-04       F    Father           30
4       F          KW    lowerlevel    G-04       F    Father           40
..     …          …         …         …         …        …              …
94      F          KW    lowerlevel    G-02       F    Father           80
95      F          KW    lowerlevel    G-02       F    Father          100
96      F          KW    lowerlevel    G-02       F    Father           14
97      F          KW    lowerlevel    G-02       F    Father            6
98      F          KW    lowerlevel    G-02       F    Father           10


     VisITedResources   AnnouncementsView   Discussion
0                 16                   2         20.0
1                 20                   3         25.0
2                  7                   0         35.0
3                 25                   5         35.0
4                 50                  12         50.0
..                …                   …           …
94                44                  35         10.0
95                80                   2         70.0
```

```
96                    60                    11         75.0
97                     2                    38         12.0
98                     3                     0         30.0

[99 rows x 10 columns]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
d:\PRATHAM FILES\Web design Project\Python new\matplotlib.ipynb Cell 3 line 5
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=53'>54</a>␣
↪print(data_fill_backward)
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=55'>56</a> # Step 4:␣
↪Filter based on any column using groupby()
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=56'>57</a> # For example,␣
↪group by the 'Nationality' column and calculate the mean of 'raisedhands' for␣
↪each group
---> <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=57'>58</a> grouped_data =␣
↪data.groupby("Nationality")["raisedhands"].mean()
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=59'>60</a> # Print the␣
↪results
      <a href='vscode-notebook-cell:/d%3A/PRATHAM%20FILES/Web%20design%20Project
↪Python%20new/matplotlib.ipynb#X11sZmlsZQ%3D%3D?line=60'>61</a> print("Grouped␣
↪Data:")

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\frame.
↪py:8872, in DataFrame.groupby(self, by, axis, level, as_index, sort,␣
↪group_keys, observed, dropna)
   8869 if level is None and by is None:
   8870     raise TypeError("You have to supply one of 'by' and 'level'")
-> 8872 return DataFrameGroupBy(
   8873     obj=self,
   8874     keys=by,
   8875     axis=axis,
   8876     level=level,
   8877     as_index=as_index,
   8878     sort=sort,
   8879     group_keys=group_keys,
   8880     observed=observed,
   8881     dropna=dropna,
   8882 )
```

```
File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
 ↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\groupby\grou
 ↪py:1273, in GroupBy.__init__(self, obj, keys, axis, level, grouper,␣
 ↪exclusions, selection, as_index, sort, group_keys, observed, dropna)
   1270 self.dropna = dropna
   1272 if grouper is None:
-> 1273     grouper, exclusions, obj = get_grouper(
   1274         obj,
   1275         keys,
   1276         axis=axis,
   1277         level=level,
   1278         sort=sort,
   1279         observed=False if observed is lib.no_default else observed,
   1280         dropna=self.dropna,
   1281     )
   1283 if observed is lib.no_default:
   1284     if any(ping._passed_categorical for ping in grouper.groupings):


File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
 ↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\groupby\grou
 ↪py:1009, in get_grouper(obj, key, axis, level, sort, observed, validate,␣
 ↪dropna)
   1007         in_axis, level, gpr = False, gpr, None
   1008     else:
-> 1009         raise KeyError(gpr)
   1010 elif isinstance(gpr, Grouper) and gpr.key is not None:
   1011     # Add key to exclusions
   1012     exclusions.add(gpr.key)


KeyError: 'Nationality'
```

```python
import pandas as pd

# Step 1: Read the CSV file and create and understand the data frame using␣
 ↪describe(), shape, and info()
# Replace 'your_dataset.csv' with the actual file path of your dataset.
data = pd.read_csv("D:\PRATHAM FILES\Web design Project\Python new\Student.csv")

# Use describe() to get summary statistics
description = data.describe()

# Get the shape of the data frame
shape = data.shape

# Get information about the data frame
info = data.info()
```

```python
# Print the results
print("Summary Statistics:")
print(description)
print("\nShape of the Data Frame:")
print(shape)
print("\nInformation about the Data Frame:")
print(info)

# Step 2: Find if any missing values (null values) are in the data and print␣
 ↪the data frame
# Check for missing values
missing_values = data.isnull().sum()

# Print the results
print("Missing Values:")
print(missing_values)

# Step 3: Handle all the rows with missing data in four different ways (delete,␣
 ↪replace, fill, bfill), and print the data frame
# Handle missing values in four different ways:
# 1. Drop rows with missing values
data_drop = data.dropna()

# 2. Replace missing values with a specific value
data_replace = data.fillna('replacement_value')

# 3. Forward fill missing values
data_fill_forward = data.ffill()

# 4. Backward fill missing values
data_fill_backward = data.bfill()

# Print the results
print("\nData Frame after Dropping Rows with Missing Values:")
print(data_drop)
print("\nData Frame after Replacing Missing Values:")
print(data_replace)
print("\nData Frame after Forward Filling Missing Values:")
print(data_fill_forward)
print("\nData Frame after Backward Filling Missing Values:")
print(data_fill_backward)

# Step 4: Filter based on any column using groupby()
# For example, group by the 'NationalITy' column and calculate the mean of␣
 ↪'raisedhands' for each group
grouped_data = data.groupby("NationalITy")["raisedhands"].mean()
print("Grouped Data:")
```

```
print(grouped_data)

# Step 5: Select 20 samples randomly and create a data frame with Hierarchical␣
  ↪Index
# Select 20 random samples from the original data frame
random_samples = data.sample(n=20)

# Create a hierarchical index
hierarchical_indexed_data = random_samples.set_index(['Gender', 'NationalITy'])

# Print the results
print("Randomly Selected Samples:")
print(random_samples)
print("\nData Frame with Hierarchical Index:")
print(hierarchical_indexed_data)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   gender             98 non-null     object
 1   NationalITy        99 non-null     object
 2   StageID            98 non-null     object
 3   GradeID            98 non-null     object
 4   Semester           99 non-null     object
 5   Relation           98 non-null     object
 6   raisedhands        99 non-null     int64
 7   VisITedResources   99 non-null     int64
 8   AnnouncementsView  99 non-null     int64
 9   Discussion         97 non-null     float64
dtypes: float64(1), int64(3), object(6)
memory usage: 7.9+ KB
Summary Statistics:
       raisedhands  VisITedResources  AnnouncementsView  Discussion
count    99.000000         99.000000          99.000000   97.000000
mean     29.303030         33.242424          19.252525   45.041237
std      24.004445         30.367508          15.891432   27.065475
min       0.000000          0.000000           0.000000    1.000000
25%      10.000000         10.000000           6.000000   20.000000
50%      20.000000         20.000000          16.000000   40.000000
75%      49.500000         60.000000          31.000000   70.000000
max     100.000000         90.000000          70.000000   99.000000

Shape of the Data Frame:
(99, 10)

Information about the Data Frame:
```

```
None
Missing Values:
gender              1
NationalITy         0
StageID             1
GradeID             1
Semester            0
Relation            1
raisedhands         0
VisITedResources    0
AnnouncementsView   0
Discussion          2
dtype: int64


Data Frame after Dropping Rows with Missing Values:
   gender NationalITy        StageID GradeID Semester Relation  raisedhands  \
0       M          KW    lowerlevel    G-04        F   Father           15
1       M          KW    lowerlevel    G-04        F   Father           20
7       M          KW  MiddleSchool    G-07        F   Father           50
9       F          KW  MiddleSchool    G-07        F   Father           70
10      M          KW  MiddleSchool    G-07        F   Father           50
..    ...         ...           ...     ...      ...      ...          ...
94      F          KW    lowerlevel    G-02        F   Father           80
95      F          KW    lowerlevel    G-02        F   Father          100
96      F          KW    lowerlevel    G-02        F   Father           14
97      F          KW    lowerlevel    G-02        F   Father            6
98      F          KW    lowerlevel    G-02        F   Father           10


    VisITedResources  AnnouncementsView  Discussion
0                 16                  2        20.0
1                 20                  3        25.0
7                 10                 15        22.0
9                 80                 25        70.0
10                88                 30        80.0
..               ...                ...         ...
94                44                 35        10.0
95                80                  2        70.0
96                60                 11        75.0
97                 2                 38        12.0
98                 3                  0        30.0

[93 rows x 10 columns]

Data Frame after Replacing Missing Values:
          gender NationalITy        StageID GradeID Semester  \
0              M          KW  lowerlevel    G-04        F
1              M          KW  lowerlevel    G-04        F
2              M          KW  lowerlevel    G-04        F
```

```
3                        M        KW  lowerlevel   G-04        F
4   replacement_value            KW  lowerlevel   G-04        F
..                   …    …               …      …        …
94                       F        KW  lowerlevel   G-02        F
95                       F        KW  lowerlevel   G-02        F
96                       F        KW  lowerlevel   G-02        F
97                       F        KW  lowerlevel   G-02        F
98                       F        KW  lowerlevel   G-02        F

              Relation  raisedhands  VisITedResources  AnnouncementsView  \
0               Father           15                16                  2
1               Father           20                20                  3
2               Father           10                 7                  0
3    replacement_value           30                25                  5
4               Father           40                50                 12
..                   …            …                 …                  …
94              Father           80                44                 35
95              Father          100                80                  2
96              Father           14                60                 11
97              Father            6                 2                 38
98              Father           10                 3                  0

              Discussion
0                   20.0
1                   25.0
2      replacement_value
3                   35.0
4                   50.0
..                     …
94                  10.0
95                  70.0
96                  75.0
97                  12.0
98                  30.0

[99 rows x 10 columns]

Data Frame after Forward Filling Missing Values:
   gender NationalITy      StageID GradeID Semester Relation  raisedhands  \
0       M          KW  lowerlevel   G-04        F   Father           15
1       M          KW  lowerlevel   G-04        F   Father           20
2       M          KW  lowerlevel   G-04        F   Father           10
3       M          KW  lowerlevel   G-04        F   Father           30
4       M          KW  lowerlevel   G-04        F   Father           40
..     …           …           …      …        …        …            …
94      F          KW  lowerlevel   G-02        F   Father           80
95      F          KW  lowerlevel   G-02        F   Father          100
96      F          KW  lowerlevel   G-02        F   Father           14
```

```
97      F           KW   lowerlevel    G-02       F   Father              6
98      F           KW   lowerlevel    G-02       F   Father             10


     VisITedResources  AnnouncementsView  Discussion
0                 16                 2        20.0
1                 20                 3        25.0
2                  7                 0        25.0
3                 25                 5        35.0
4                 50                12        50.0
..                ...               ...         ...
94                44                35        10.0
95                80                 2        70.0
96                60                11        75.0
97                 2                38        12.0
98                 3                 0        30.0

[99 rows x 10 columns]

Data Frame after Backward Filling Missing Values:
   gender NationalITy      StageID GradeID Semester Relation  raisedhands  \
0       M           KW   lowerlevel    G-04       F   Father             15
1       M           KW   lowerlevel    G-04       F   Father             20
2       M           KW   lowerlevel    G-04       F   Father             10
3       M           KW   lowerlevel    G-04       F   Father             30
4       F           KW   lowerlevel    G-04       F   Father             40
..     ...         ...          ...     ...     ...      ...            ...
94      F           KW   lowerlevel    G-02       F   Father             80
95      F           KW   lowerlevel    G-02       F   Father            100
96      F           KW   lowerlevel    G-02       F   Father             14
97      F           KW   lowerlevel    G-02       F   Father              6
98      F           KW   lowerlevel    G-02       F   Father             10


     VisITedResources  AnnouncementsView  Discussion
0                 16                 2        20.0
1                 20                 3        25.0
2                  7                 0        35.0
3                 25                 5        35.0
4                 50                12        50.0
..                ...               ...         ...
94                44                35        10.0
95                80                 2        70.0
96                60                11        75.0
97                 2                38        12.0
98                 3                 0        30.0

[99 rows x 10 columns]
Grouped Data:
NationalITy
```

```
Egypt          45.000000
Iran           15.000000
Jordan         33.000000
KW             26.180723
SaudiArabia    60.000000
USA            47.200000
lebanon        50.000000
venzuela       80.000000
Name: raisedhands, dtype: float64
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_3488\2358278301.py in ?()
     63 # Select 20 random samples from the original data frame
     64 random_samples = data.sample(n=20)
     65
     66 # Create a hierarchical index
---> 67 hierarchical_indexed_data = random_samples.set_index(['Gender',␣
  ↪'NationalITy'])
     68
     69 # Print the results
     70 print("Randomly Selected Samples:")


~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.
  ↪11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\pandas\core\frame.
  ↪py in ?(self, keys, drop, append, inplace, verify_integrity)
   5869                     if not found:
   5870                         missing.append(col)
   5871
   5872         if missing:
-> 5873             raise KeyError(f"None of {missing} are in the columns")
   5874
   5875         if inplace:
   5876             frame = self


KeyError: "None of ['Gender'] are in the columns"
```

[ ]:
```

```
```