```
import pandas as pd
!pip install matplotlib
```

```
Defaulting to user installation because normal site-packages is not
writeable
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\
lib\site-packages (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.21 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in c:\programdata\anaconda3\
lib\site-packages (from matplotlib) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\
anaconda3\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\
lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

```
df = pd.read_csv("/Users/admin/Downloads/StudentsPerformance.csv")
df
```

```
     gender race/ethnicity parental level of education
lunch  \
0    female        group B           bachelor's degree      standard

1    female        group C                some college      standard

2    female        group B             master's degree      standard

3      male        group A          associate's degree  free/reduced

4      male        group C                some college      standard

..      ...            ...                         ...           ...

995  female        group E             master's degree      standard

996    male        group C                 high school  free/reduced

997  female        group C                 high school  free/reduced
```

```
998  female        group D                some college      standard

999  female        group D                some college  free/reduced


    test_preparation_course  math_score  reading_score  writing_score
0                      none        72.0           72.0           74.0
1                 completed        69.0           90.0           88.0
2                      none        90.0           95.0           93.0
3                      none        47.0           57.0           44.0
4                      none        76.0           78.0           75.0
..                      ...         ...            ...            ...
995               completed        88.0           99.0           95.0
996                    none        62.0           55.0           55.0
997               completed        59.0           71.0           65.0
998               completed        68.0           78.0           77.0
999                    none        77.0           86.0           86.0


[1000 rows x 8 columns]
```

```
df.isnull().sum()
```

```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test_preparation_course        0
math_score                     9
reading_score                  5
writing_score                  6
dtype: int64
```

```
df.dropna()
```

```
     gender race/ethnicity parental level of education
lunch  \
0    female        group B           bachelor's degree      standard

1    female        group C                some college      standard
```

```
2    female        group B             master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E             master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test_preparation_course  math_score  reading_score  writing_score
0                      none        72.0           72.0           74.0
1                 completed        69.0           90.0           88.0
2                      none        90.0           95.0           93.0
3                      none        47.0           57.0           44.0
4                      none        76.0           78.0           75.0
..                      ...         ...            ...            ...
995               completed        88.0           99.0           95.0
996                    none        62.0           55.0           55.0
997               completed        59.0           71.0           65.0
998               completed        68.0           78.0           77.0
999                    none        77.0           86.0           86.0

[980 rows x 8 columns]
```

```
 math_score_mean = df["math score"].mean()
df["math score"] = df["math score"].fillna(math_score_mean)
```

```
--------------------------------------------------------------------------
-----
```

```
KeyError                                  Traceback (most recent call
last)
File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\indexes\
base.py:3805, in Index.get_loc(self, key)
   3804 try:
-> 3805     return self._engine.get_loc(casted_key)
   3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

File pandas\\_libs\\hashtable_class_helper.pxi:7081, in
pandas._libs.hashtable.PyObjectHashTable.get_item()

File pandas\\_libs\\hashtable_class_helper.pxi:7089, in
pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'math score'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call
last)
Cell In[5], line 1
----> 1 math_score_mean = df["math score"].mean()
      2 df["math score"] = df["math score"].fillna(math_score_mean)

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\
frame.py:4102, in DataFrame.__getitem__(self, key)
   4100 if self.columns.nlevels > 1:
   4101     return self._getitem_multilevel(key)
-> 4102 indexer = self.columns.get_loc(key)
   4103 if is_integer(indexer):
   4104     indexer = [indexer]

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\indexes\
base.py:3812, in Index.get_loc(self, key)
   3807     if isinstance(casted_key, slice) or (
   3808         isinstance(casted_key, abc.Iterable)
   3809         and any(isinstance(x, slice) for x in casted_key)
   3810     ):
   3811         raise InvalidIndexError(key)
-> 3812     raise KeyError(key) from err
   3813 except TypeError:
   3814     # If we have a listlike key, _check_indexing_error will
raise
   3815     #  InvalidIndexError. Otherwise we fall through and re-
raise
   3816     #  the TypeError.
```

```
   3817     self._check_indexing_error(key)

KeyError: 'math score'
```

```
 math_score_mean = df["math_score"].mean()
df["math_score"] = df["math_score"].fillna(math_score_mean)

df
```

```
     gender race/ethnicity parental level of education
lunch  \
0    female        group B           bachelor's degree      standard

1    female        group C                some college      standard

2    female        group B             master's degree      standard

3      male        group A          associate's degree  free/reduced

4      male        group C                some college      standard

..      ...            ...                         ...           ...

995  female        group E             master's degree      standard

996    male        group C                 high school  free/reduced

997  female        group C                 high school  free/reduced

998  female        group D                some college      standard

999  female        group D                some college  free/reduced


    test_preparation_course  math_score  reading_score  writing_score

0                      none        72.0           72.0           74.0

1                 completed        69.0           90.0           88.0

2                      none        90.0           95.0           93.0

3                      none        47.0           57.0           44.0

4                      none        76.0           78.0           75.0

..                      ...         ...            ...            ...

995               completed        88.0           99.0           95.0

996                    none        62.0           55.0           55.0
```

```
997                completed        59.0           71.0           65.0

998                completed        68.0           78.0           77.0

999                     none        77.0           86.0           86.0


[1000 rows x 8 columns]
```

```
df.boxlot()
```

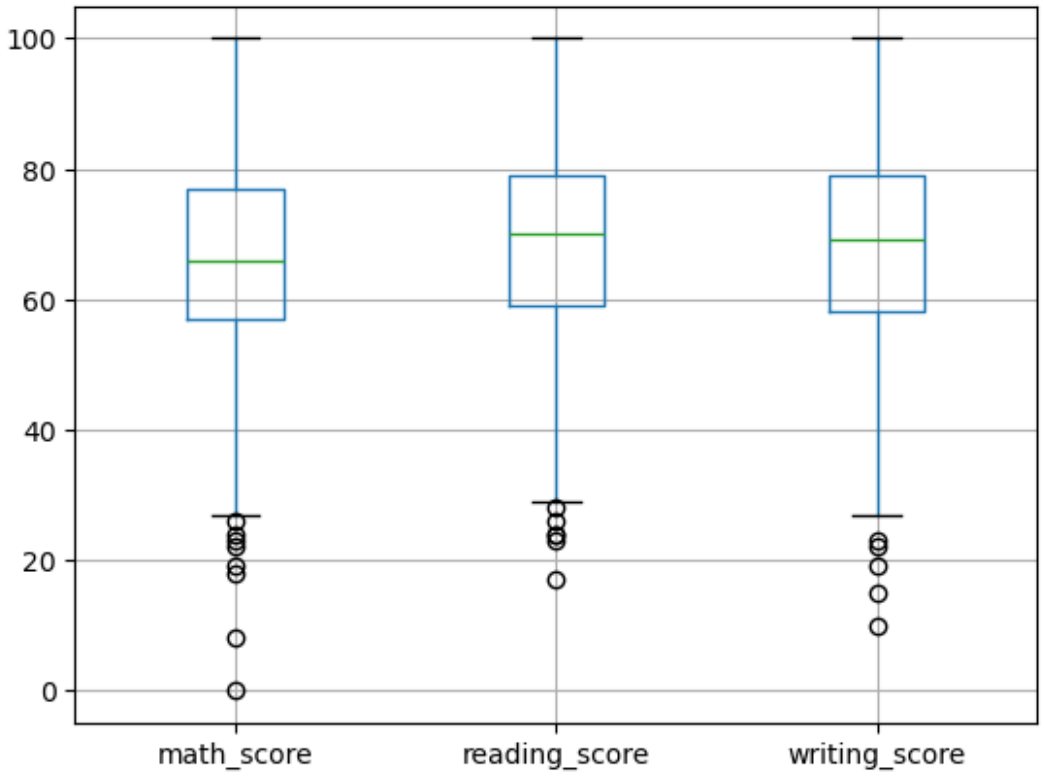
```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_12776\2877535902.py in ?()
----> 1 df.boxlot()

C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\generic.py in ?(self, name)
   6295             and name not in self._accessors
   6296             and self._info_axis._can_hold_identifiers_and_holds_name(name)
   6297         ):
   6298             return self[name]
-> 6299         return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'boxlot'
```

```
df.boxplot()
```

```
<Axes: >
```




```
dfnewdf = df[df["math_score"] > 20]
newdf
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
Cell In[10], line 2
      1 dfnewdf = df[df["math_score"] > 20]
----> 2 newdf

NameError: name 'newdf' is not defined
```

```
newdf = df[df["math_score"] > 30]
newdf
```

```
     gender race/ethnicity parental level of education
lunch  \
0    female        group B           bachelor's degree      standard

1    female        group C                some college      standard

2    female        group B             master's degree      standard

3      male        group A          associate's degree  free/reduced
```

```
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E             master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test_preparation_course  math_score  reading_score  writing_score
0                      none        72.0           72.0           74.0
1                 completed        69.0           90.0           88.0
2                      none        90.0           95.0           93.0
3                      none        47.0           57.0           44.0
4                      none        76.0           78.0           75.0
..                      ...         ...            ...            ...
995               completed        88.0           99.0           95.0
996                    none        62.0           55.0           55.0
997               completed        59.0           71.0           65.0
998               completed        68.0           78.0           77.0
999                    none        77.0           86.0           86.0

[984 rows x 8 columns]
```

```
newdf.plot()
```

```
<Axes: >
```

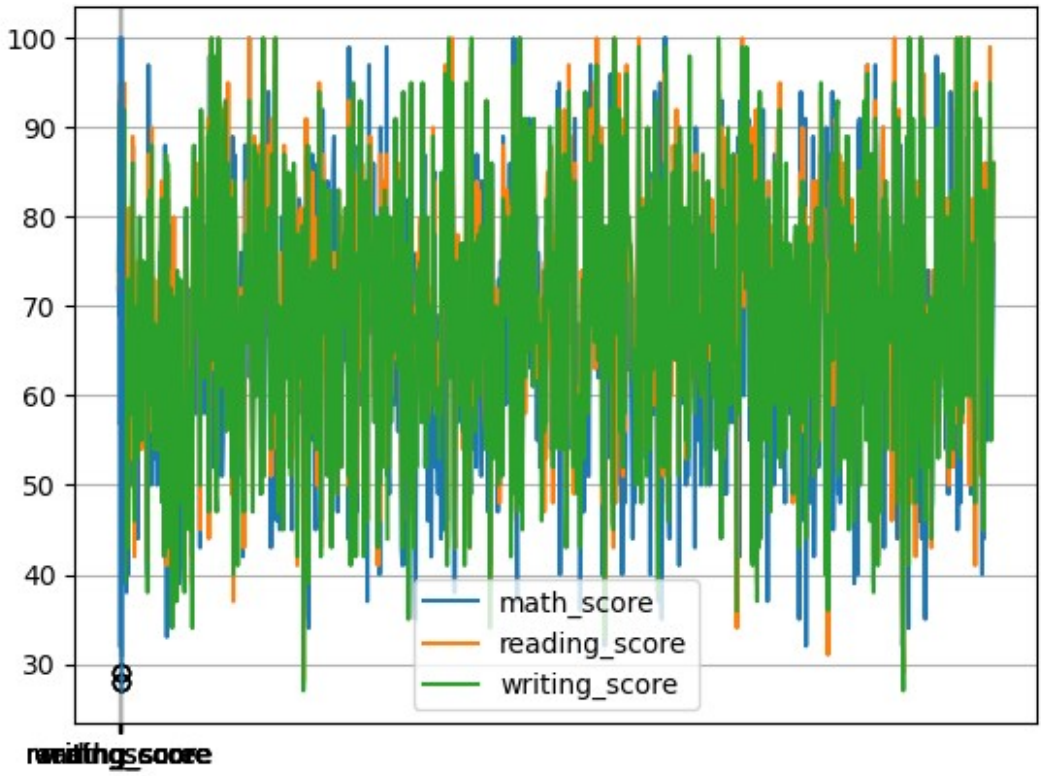
```
newdf.boxplot()
```

```
<Axes: >
```

```
%matplotlib inline
newdf.boxplot()
```

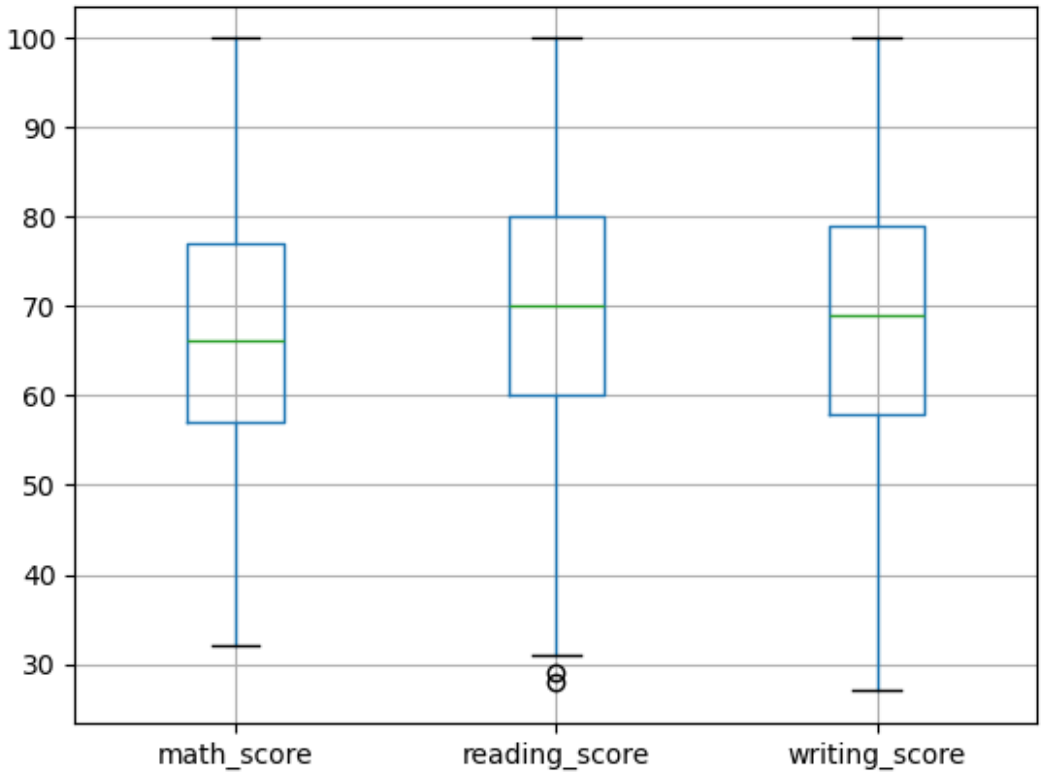
```
<Axes: >
```

```
import matplotlib.pyplot as plt
newdf.boxplot()
plt.show()
```




```
newdf.boxplot()
```

```
<Axes: >
```


100
90
80
70
60
50
40
30
math_score
reading_score
writing_score