IE_PANDAS

Name: ie_pandas
Author: Group4
Version: 0.0

ie_pandas is a library that contains a simplified data frame object.

```
from ie_pandas import DataFrame
```

Documentation

(see the following section for the package requirements for ie_pandas)

Summary:

The ie_pandas library contains the DataFrame object which takes an input of a dictionary of lists or numpy.arrays and produces a data frame with the following capabilities:

- Setting a column item to a new value
- Getting a specific column item
- Returning a specific row
- Returning the column names
- Returning the number of rows/columns
- Returning the sum/median/min/max of a numeric/integer column
- Representation of the dataframe (and components) in table form

Example input:

```
dictionary = {
    "pet": np.array(["cat", "dog", "mouse"]),
    "age": np.array([1, 2, 3]),
    "weight": np.array([1.0, 2.0, 3.0]),
    "sick": np.array([True, True, False]),
}

df = DataFrame(dictionary)
```

Setting a column item to a new value:

```
df["pet"] = np.array(["frog", "lizard", "turtle"])
```

Getting a specific column item:

```
df[["pet"]]
```

pet
frog
lizard
turtle

```
In [743]: df[["weight","age"]]
Out[743]: [array([1., 2., 3.]), array([1, 2, 3])]
```

weight	age
1.0	1
2.0	2
3.0	3

Note that the desired columns must be represented as a list of column names, and the output is an array as well as a visual table representation.

Item values can also be changed using the following syntax:

```
df[["pet"]][0][0] = "frog2"
```

Returning a specific row:

```
In [735]: df.get_row([1])
Out[735]: ['lizard', 2, 2.0, True]
```

pet	age	weight	sick
lizard	2	2.0	True

```
In [723]: df.get_row([0,2])
Out[723]: [['frog', 1, 1.0, True], ['turtle', 3, 3.0, False]]
```

pet	age	weight	sick
frog	1	1.0	True
turtle	3	3.0	False

Note that the desired rows must be represented as a list of indices, and the output is an array as well as a visual table representation.

Returning column names:

```
In [724]: df.col_names()
Out[724]: ['pet', 'age', 'weight', 'sick']
```

Returning number of rows/columns:

```
In [13]: df.ncols()
Out[13]: 4
```

```
In [14]: df.nrows()
Out[14]: 3
```

Returning the sum/median/min/max of a numeric/integer column:

```
In [15]: df.sum()
Out[15]: [6, 6.0]
```

age	weight
6	6.0

```
In [727]: df.median()
Out[727]: [2, 2.0]
```

age	weight
2	2.0

```
In [728]: df.max()
Out[728]: [3, 3.0]
```

age	weight
3	3.0

```
In [729]: df.min()
Out[729]: [1, 1.0]
```

age	weight
1	1.0

Note that the output is a list as well as a visual table representation

Representation of the dataframe in table form:

```
In [16]: df
Out[16]: {'pet': array(['frog', 'lizard', 'turtle'], dtype='<U6'), 'age': array([1, 2, 3]), 'weight': array([1., 2., 3.]), 'sick': ar
         ray([ True,  True, False])}
```

pet	age	weight	sick
frog	1	1.0	True
lizard	2	2.0	True
turtle	3	3.0	False

Note that the output is a dictionary as well as a visual table representation

---

Requirements

ie_pandas requires the following Python packages:

- NumPy, for array objects and basic math functions
- Statistics, for more advanced math functions
- Matplotlib, for visualization components

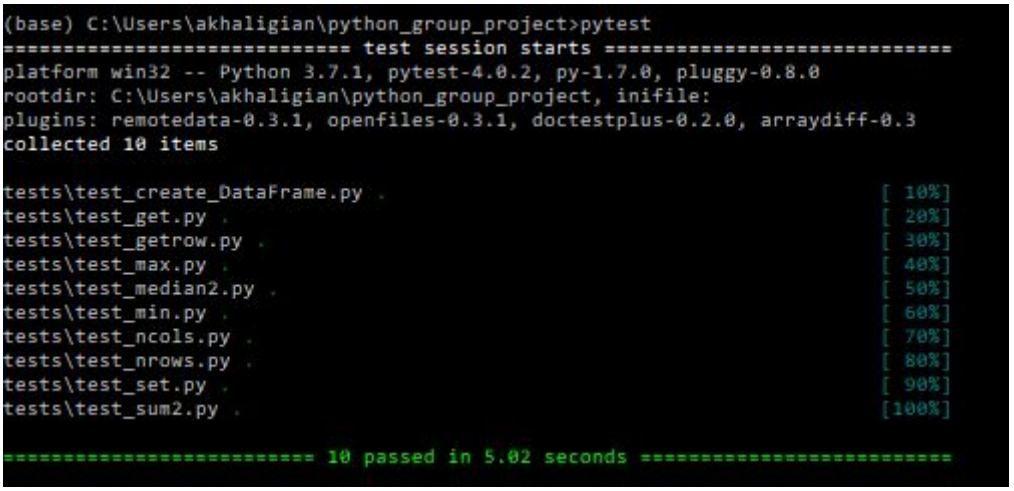
Installation

The library can be installed using the following command:

```
$ pip install --editable .
```

Testing

The tests can be run using pytest:


(base) C:\Users\akhaligian\python_group_project>pytest
test session starts
platform win32 -- Python 3.7.1, pytest-4.0.2, py-1.7.0, pluggy-0.8.0
rootdir: C:\Users\akhaligian\python_group_project, inifile:
plugins: remotedata-0.3.1, openfiles-0.3.1, doctestplus-0.2.0, arraydiff-0.3
collected 10 items
tests\test_create_DataFrame.py . [10%]
tests\test_get.py . [20%]
tests\test_getrow.py . [30%]
tests\test_max.py . [40%]
tests\test_median2.py . [50%]
tests\test_min.py . [60%]
tests\test_ncols.py . [70%]
tests\test_nrows.py . [80%]
tests\test_set.py . [90%]
tests\test_sum2.py . [100%]
10 passed in 5.02 seconds