

# Pandas\_assignmet\_Kashti\_Fao

March 14, 2022

## 1 Python Chilla 10 hours video

### 1.1 Basics of Python

1.1.1 This file contains the basic concept of python by Ammar Bhai

### 1.2 Pandas Practice

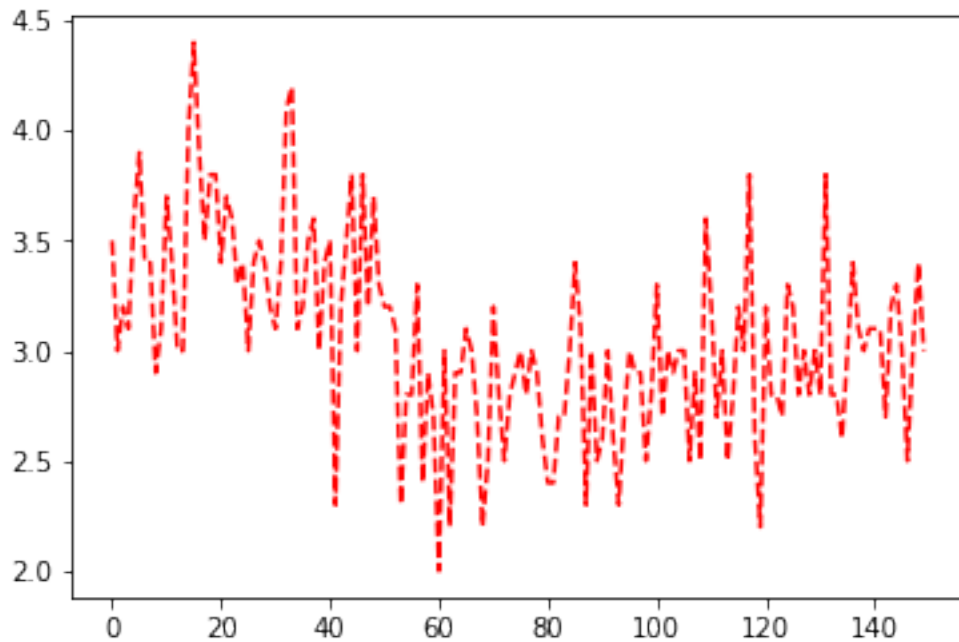
```
[ ]: # Importing Librarires
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
phool = pd.read_csv("iris.csv")
phool
```

```
[ ]:      sepal.length  sepal.width  petal.length  petal.width  variety
0           5.1         3.5         1.4         0.2      Setosa
1           4.9         3.0         1.4         0.2      Setosa
2           4.7         3.2         1.3         0.2      Setosa
3           4.6         3.1         1.5         0.2      Setosa
4           5.0         3.6         1.4         0.2      Setosa
..          ...          ...          ...          ...          ...
145          6.7         3.0         5.2         2.3  Virginica
146          6.3         2.5         5.0         1.9  Virginica
147          6.5         3.0         5.2         2.0  Virginica
148          6.2         3.4         5.4         2.3  Virginica
149          5.9         3.0         5.1         1.8  Virginica
```

[150 rows x 5 columns]

```
[ ]: plt.plot( phool["sepal.width"], "r--")
```

```
[ ]: [<matplotlib.lines.Line2D at 0x1de82debfd0>]
```



```
[ ]: # Object creation
s = pd.Series([1,2,4,5,6,np.NaN])
s
```

```
[ ]: 0    1.0
     1    2.0
     2    4.0
     3    5.0
     4    6.0
     5    NaN
     dtype: float64
```

```
[ ]: dates = pd.date_range("20220314",periods=20)
     dates
```

```
[ ]: DatetimeIndex(['2022-03-14', '2022-03-15', '2022-03-16', '2022-03-17',
                    '2022-03-18', '2022-03-19', '2022-03-20', '2022-03-21',
                    '2022-03-22', '2022-03-23', '2022-03-24', '2022-03-25',
                    '2022-03-26', '2022-03-27', '2022-03-28', '2022-03-29',
                    '2022-03-30', '2022-03-31', '2022-04-01', '2022-04-02'],
                    dtype='datetime64[ns]', freq='D')
```

```
[ ]: df1 = pd.DataFrame(np.random.randn(20,4),index=dates, columns=list("ABCD"))
     df1
```

```
[ ]:
```

	A	B	C	D
2022-03-14	-2.389341	-0.175309	0.212880	1.913759
2022-03-15	1.144009	1.504342	0.705470	1.011790
2022-03-16	-0.342653	0.582746	-0.613556	0.950932
2022-03-17	0.315202	-0.562544	-1.507190	2.031003
2022-03-18	0.077353	0.648974	2.185769	-0.498190
2022-03-19	1.202017	0.134866	-0.956809	1.329111
2022-03-20	-0.388075	-0.075626	-1.060376	0.690637
2022-03-21	1.544423	1.179385	1.979273	0.026136
2022-03-22	1.529805	-0.646925	0.358865	1.380160
2022-03-23	-1.170390	1.458856	0.545326	1.716879
2022-03-24	-1.218228	0.236370	0.974356	0.402824
2022-03-25	0.730549	-0.351083	0.495550	1.459450
2022-03-26	-0.879890	-0.835687	0.488129	-1.261578
2022-03-27	1.351165	-0.414541	1.399763	0.563101
2022-03-28	2.052156	0.709011	-1.519008	0.384681
2022-03-29	0.432916	-1.525424	-1.119159	-0.610072
2022-03-30	-0.231208	0.475551	0.262853	0.642505
2022-03-31	0.793882	-0.024003	-0.025051	1.022415
2022-04-01	-0.504087	-0.757736	-1.629195	0.135687
2022-04-02	-1.233164	1.038869	1.690570	-0.211571

```
[ ]: df1.head(3)
```

```
[ ]:
```

	A	B	C	D
2022-03-14	-2.389341	-0.175309	0.212880	1.913759
2022-03-15	1.144009	1.504342	0.705470	1.011790
2022-03-16	-0.342653	0.582746	-0.613556	0.950932

```
[ ]: df1.tail(3)
```

```
[ ]:
```

	A	B	C	D
2022-03-31	0.793882	-0.024003	-0.025051	1.022415
2022-04-01	-0.504087	-0.757736	-1.629195	0.135687
2022-04-02	-1.233164	1.038869	1.690570	-0.211571

```
[ ]: df1.index
```

```
[ ]: DatetimeIndex(['2022-03-14', '2022-03-15', '2022-03-16', '2022-03-17',
                    '2022-03-18', '2022-03-19', '2022-03-20', '2022-03-21',
                    '2022-03-22', '2022-03-23', '2022-03-24', '2022-03-25',
                    '2022-03-26', '2022-03-27', '2022-03-28', '2022-03-29',
                    '2022-03-30', '2022-03-31', '2022-04-01', '2022-04-02'],
                    dtype='datetime64[ns]', freq='D')
```

```
[ ]: df1.to_numpy()
```

```
[ ]: array([[ -2.38934144, -0.17530894,  0.21287967,  1.91375945],
 [  1.14400855,  1.50434152,  0.70546953,  1.0117896 ],
 [-0.34265302,  0.58274589, -0.61355567,  0.95093156],
 [  0.31520214, -0.56254443, -1.50719039,  2.03100294],
 [  0.07735267,  0.64897402,  2.18576921, -0.4981899 ],
 [  1.20201673,  0.13486608, -0.95680854,  1.32911119],
 [-0.38807516, -0.07562606, -1.06037572,  0.69063689],
 [  1.54442254,  1.17938523,  1.97927312,  0.02613649],
 [  1.5298054 , -0.64692493,  0.35886543,  1.38016041],
 [-1.17038997,  1.4588564 ,  0.54532559,  1.71687898],
 [-1.21822814,  0.23637004,  0.97435556,  0.40282428],
 [  0.73054855, -0.35108256,  0.49555008,  1.45945026],
 [-0.87988968, -0.83568694,  0.48812933, -1.26157771],
 [  1.35116481, -0.4145415 ,  1.39976259,  0.56310111],
 [  2.05215627,  0.70901059, -1.51900798,  0.3846813 ],
 [  0.43291564, -1.52542429, -1.11915916, -0.61007228],
 [-0.23120773,  0.4755512 ,  0.26285272,  0.64250536],
 [  0.7938822 , -0.02400309, -0.02505075,  1.02241469],
 [-0.50408657, -0.75773581, -1.62919475,  0.13568712],
 [-1.23316423,  1.03886855,  1.69056982, -0.21157056]])
```

```
[ ]: df1.describe()
```

```
[ ]:
count    20.000000    20.000000    20.000000    20.000000
mean      0.140822     0.130005     0.143423     0.653983
std       1.170010     0.819297     1.180439     0.881054
min      -2.389341    -1.525424    -1.629195    -1.261578
25%      -0.598037    -0.451542    -0.982700     0.108299
50%       0.196277     0.055431     0.310859     0.666571
75%       1.158511     0.663983     0.772691     1.341873
max       2.052156     1.504342     2.185769     2.031003
```

```
[ ]: #transpose
df1.T
```

```
[ ]:
2022-03-14  2022-03-15  2022-03-16  2022-03-17  2022-03-18  2022-03-19  \
A    -2.389341    1.144009   -0.342653    0.315202    0.077353    1.202017
B    -0.175309    1.504342    0.582746   -0.562544    0.648974    0.134866
C     0.212880    0.705470   -0.613556   -1.507190    2.185769   -0.956809
D     1.913759    1.011790    0.950932    2.031003   -0.498190    1.329111

2022-03-20  2022-03-21  2022-03-22  2022-03-23  2022-03-24  2022-03-25  \
A    -0.388075    1.544423    1.529805   -1.170390   -1.218228    0.730549
B    -0.075626    1.179385   -0.646925    1.458856    0.236370   -0.351083
C    -1.060376    1.979273    0.358865    0.545326    0.974356    0.495550
D     0.690637    0.026136    1.380160    1.716879    0.402824    1.459450
```

	2022-03-26	2022-03-27	2022-03-28	2022-03-29	2022-03-30	2022-03-31	\
A	-0.879890	1.351165	2.052156	0.432916	-0.231208	0.793882	
B	-0.835687	-0.414541	0.709011	-1.525424	0.475551	-0.024003	
C	0.488129	1.399763	-1.519008	-1.119159	0.262853	-0.025051	
D	-1.261578	0.563101	0.384681	-0.610072	0.642505	1.022415	

	2022-04-01	2022-04-02
A	-0.504087	-1.233164
B	-0.757736	1.038869
C	-1.629195	1.690570
D	0.135687	-0.211571

```
[ ]: df1.sort_index(axis=1,ascending=True)
```

```
[ ]:
```

	A	B	C	D
2022-03-14	-2.389341	-0.175309	0.212880	1.913759
2022-03-15	1.144009	1.504342	0.705470	1.011790
2022-03-16	-0.342653	0.582746	-0.613556	0.950932
2022-03-17	0.315202	-0.562544	-1.507190	2.031003
2022-03-18	0.077353	0.648974	2.185769	-0.498190
2022-03-19	1.202017	0.134866	-0.956809	1.329111
2022-03-20	-0.388075	-0.075626	-1.060376	0.690637
2022-03-21	1.544423	1.179385	1.979273	0.026136
2022-03-22	1.529805	-0.646925	0.358865	1.380160
2022-03-23	-1.170390	1.458856	0.545326	1.716879
2022-03-24	-1.218228	0.236370	0.974356	0.402824
2022-03-25	0.730549	-0.351083	0.495550	1.459450
2022-03-26	-0.879890	-0.835687	0.488129	-1.261578
2022-03-27	1.351165	-0.414541	1.399763	0.563101
2022-03-28	2.052156	0.709011	-1.519008	0.384681
2022-03-29	0.432916	-1.525424	-1.119159	-0.610072
2022-03-30	-0.231208	0.475551	0.262853	0.642505
2022-03-31	0.793882	-0.024003	-0.025051	1.022415
2022-04-01	-0.504087	-0.757736	-1.629195	0.135687
2022-04-02	-1.233164	1.038869	1.690570	-0.211571

```
[ ]: df1['A']
```

```
[ ]:
```

2022-03-14	-2.389341
2022-03-15	1.144009
2022-03-16	-0.342653
2022-03-17	0.315202
2022-03-18	0.077353
2022-03-19	1.202017
2022-03-20	-0.388075
2022-03-21	1.544423

```

2022-03-22    1.529805
2022-03-23   -1.170390
2022-03-24   -1.218228
2022-03-25    0.730549
2022-03-26   -0.879890
2022-03-27    1.351165
2022-03-28    2.052156
2022-03-29    0.432916
2022-03-30   -0.231208
2022-03-31    0.793882
2022-04-01   -0.504087
2022-04-02   -1.233164
Freq: D, Name: A, dtype: float64

```

```
[ ]: df1["C"]
```

```

[ ]: 2022-03-14    0.212880
      2022-03-15    0.705470
      2022-03-16   -0.613556
      2022-03-17   -1.507190
      2022-03-18    2.185769
      2022-03-19   -0.956809
      2022-03-20   -1.060376
      2022-03-21    1.979273
      2022-03-22    0.358865
      2022-03-23    0.545326
      2022-03-24    0.974356
      2022-03-25    0.495550
      2022-03-26    0.488129
      2022-03-27    1.399763
      2022-03-28   -1.519008
      2022-03-29   -1.119159
      2022-03-30    0.262853
      2022-03-31   -0.025051
      2022-04-01   -1.629195
      2022-04-02    1.690570
Freq: D, Name: C, dtype: float64

```

```
[ ]: # Row wise selection
      df1[0:3]
```

```

[ ]:           A          B          C          D
      2022-03-14 -2.389341 -0.175309  0.212880  1.913759
      2022-03-15  1.144009  1.504342  0.705470  1.011790
      2022-03-16 -0.342653  0.582746 -0.613556  0.950932

```

```
[ ]: df1.loc[dates[0]]
```

```
[ ]: A    -2.389341
      B    -0.175309
      C     0.212880
      D     1.913759
      Name: 2022-03-14 00:00:00, dtype: float64
```

```
[ ]: df1.loc["20220314":"20220318",["A","B"]]
```

```
[ ]:
      A      B
2022-03-14 -2.389341 -0.175309
2022-03-15  1.144009  1.504342
2022-03-16 -0.342653  0.582746
2022-03-17  0.315202 -0.562544
2022-03-18  0.077353  0.648974
```

```
[ ]: df1.loc["20220314",["A","B"]]
```

```
[ ]: A    -2.389341
      B    -0.175309
      Name: 2022-03-14 00:00:00, dtype: float64
```

```
[ ]: df1.at[dates[19], "A"]
```

```
[ ]: -1.2331642313754636
```

```
[ ]: df1.iloc[: -7]
```

```
[ ]:
      A      B      C      D
2022-03-14 -2.389341 -0.175309  0.212880  1.913759
2022-03-15  1.144009  1.504342  0.705470  1.011790
2022-03-16 -0.342653  0.582746 -0.613556  0.950932
2022-03-17  0.315202 -0.562544 -1.507190  2.031003
2022-03-18  0.077353  0.648974  2.185769 -0.498190
2022-03-19  1.202017  0.134866 -0.956809  1.329111
2022-03-20 -0.388075 -0.075626 -1.060376  0.690637
2022-03-21  1.544423  1.179385  1.979273  0.026136
2022-03-22  1.529805 -0.646925  0.358865  1.380160
2022-03-23 -1.170390  1.458856  0.545326  1.716879
2022-03-24 -1.218228  0.236370  0.974356  0.402824
2022-03-25  0.730549 -0.351083  0.495550  1.459450
2022-03-26 -0.879890 -0.835687  0.488129 -1.261578
```

```
[ ]: df1[df1["A"]>1]
```

```
[ ]:
      A      B      C      D
2022-03-15  1.144009  1.504342  0.705470  1.011790
2022-03-19  1.202017  0.134866 -0.956809  1.329111
2022-03-21  1.544423  1.179385  1.979273  0.026136
```

2022-03-22	1.529805	-0.646925	0.358865	1.380160
2022-03-27	1.351165	-0.414541	1.399763	0.563101
2022-03-28	2.052156	0.709011	-1.519008	0.384681

```
[ ]: df1["B"]>0
```

```
[ ]: 2022-03-14    False
      2022-03-15     True
      2022-03-16     True
      2022-03-17    False
      2022-03-18     True
      2022-03-19     True
      2022-03-20    False
      2022-03-21     True
      2022-03-22    False
      2022-03-23     True
      2022-03-24     True
      2022-03-25    False
      2022-03-26    False
      2022-03-27    False
      2022-03-28     True
      2022-03-29    False
      2022-03-30     True
      2022-03-31    False
      2022-04-01    False
      2022-04-02     True
      Freq: D, Name: B, dtype: bool
```

```
[ ]: df2=df1.copy
```

```
[ ]: df2
```

```
[ ]: <bound method NDFrame.copy of
```

	A	B	C	D
2022-03-14	-2.389341	-0.175309	0.212880	1.913759
2022-03-15	1.144009	1.504342	0.705470	1.011790
2022-03-16	-0.342653	0.582746	-0.613556	0.950932
2022-03-17	0.315202	-0.562544	-1.507190	2.031003
2022-03-18	0.077353	0.648974	2.185769	-0.498190
2022-03-19	1.202017	0.134866	-0.956809	1.329111
2022-03-20	-0.388075	-0.075626	-1.060376	0.690637
2022-03-21	1.544423	1.179385	1.979273	0.026136
2022-03-22	1.529805	-0.646925	0.358865	1.380160
2022-03-23	-1.170390	1.458856	0.545326	1.716879
2022-03-24	-1.218228	0.236370	0.974356	0.402824
2022-03-25	0.730549	-0.351083	0.495550	1.459450
2022-03-26	-0.879890	-0.835687	0.488129	-1.261578
2022-03-27	1.351165	-0.414541	1.399763	0.563101



```

2022-03-28  2.052156  0.709011 -1.519008  0.384681
2022-03-29  0.432916 -1.525424 -1.119159 -0.610072
2022-03-30 -0.231208  0.475551  0.262853  0.642505
2022-03-31  0.793882 -0.024003 -0.025051  1.022415
2022-04-01 -0.504087 -0.757736 -1.629195  0.135687
2022-04-02 -1.233164  1.038869  1.690570 -0.211571>

```

```

[ ]: df2["E"] =_
      ↳["one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one",

```

```

-----
TypeError                                Traceback (most recent call last)
d:\pythonchilla\SartajAhmedPythokChilla\ML_chilla\Pandas_ass.ipynb Cell 26' in_
      ↳<module>
----> <a href='vscode-notebook-cell:/d%3A/pythonchilla/SartajAhmedPythokChilla/
      ↳ML_chilla/Pandas_ass.ipynb#ch0000025?line=0'>1</a> df2["E"] =_
      ↳["one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one", "two", "one",
TypeError: 'method' object does not support item assignment

```

```

[ ]: df2

```

```

[ ]:
      A      B      C      D      E
2022-03-14 -0.388491 -0.472719 -0.729052  1.343839 one
2022-03-15  0.374134 -1.576127  0.977833 -0.339402 two
2022-03-16  0.825673  0.303721  1.009210  1.290154 one
2022-03-17  0.842147 -1.325749 -0.977785 -0.739024 two
2022-03-18 -1.663753  1.518848 -0.786516 -0.952310 one
2022-03-19  1.440243 -0.583804 -1.133892 -0.150511 two
2022-03-20 -1.833655 -1.336432 -0.227743  0.576949 one
2022-03-21  0.167561  0.113478 -0.240099  0.025911 two
2022-03-22  0.296135  2.884198 -0.830483  1.248489 one
2022-03-23  1.639017  2.455900 -1.757085 -0.375882 two
2022-03-24  1.280345  0.910121 -0.028334 -0.513699 one
2022-03-25 -1.326436 -0.754767 -0.434965  0.267053 two
2022-03-26  0.621524 -0.265237  1.436473  0.653887 one
2022-03-27  1.513879 -0.660165 -1.002322 -0.187261 two
2022-03-28  0.681857  0.645532  0.606087  1.152005 one
2022-03-29  0.447837 -1.333515 -0.972332 -2.152906 two
2022-03-30 -1.069359  1.417371 -1.707312 -1.281053 one
2022-03-31  0.399382 -1.788995  0.928380 -1.141412 two
2022-04-01 -0.337783 -0.902655 -1.042334 -0.278996 one
2022-04-02  1.007307  1.452827 -2.155665 -1.526379 two

```

```

[ ]: df2["mean"] =[1.2,1.3,1.5,1.8,1.2,1.3,1.5,1.8,1.2,1.3,1.5,1.8,1.2,1.3,1.5,1.8,1.
      ↳2,1.3,1.5,1.8,]

```

```
df2
```

```
[ ]:
      A      B      C      D      E  mean
2022-03-14 -0.388491 -0.472719 -0.729052  1.343839  one  1.2
2022-03-15  0.374134 -1.576127  0.977833 -0.339402  two  1.3
2022-03-16  0.825673  0.303721  1.009210  1.290154  one  1.5
2022-03-17  0.842147 -1.325749 -0.977785 -0.739024  two  1.8
2022-03-18 -1.663753  1.518848 -0.786516 -0.952310  one  1.2
2022-03-19  1.440243 -0.583804 -1.133892 -0.150511  two  1.3
2022-03-20 -1.833655 -1.336432 -0.227743  0.576949  one  1.5
2022-03-21  0.167561  0.113478 -0.240099  0.025911  two  1.8
2022-03-22  0.296135  2.884198 -0.830483  1.248489  one  1.2
2022-03-23  1.639017  2.455900 -1.757085 -0.375882  two  1.3
2022-03-24  1.280345  0.910121 -0.028334 -0.513699  one  1.5
2022-03-25 -1.326436 -0.754767 -0.434965  0.267053  two  1.8
2022-03-26  0.621524 -0.265237  1.436473  0.653887  one  1.2
2022-03-27  1.513879 -0.660165 -1.002322 -0.187261  two  1.3
2022-03-28  0.681857  0.645532  0.606087  1.152005  one  1.5
2022-03-29  0.447837 -1.333515 -0.972332 -2.152906  two  1.8
2022-03-30 -1.069359  1.417371 -1.707312 -1.281053  one  1.2
2022-03-31  0.399382 -1.788995  0.928380 -1.141412  two  1.3
2022-04-01 -0.337783 -0.902655 -1.042334 -0.278996  one  1.5
2022-04-02  1.007307  1.452827 -2.155665 -1.526379  two  1.8
```

```
[ ]: df2["A"]
```

```
[ ]: 2022-03-14    -0.388491
      2022-03-15     0.374134
      2022-03-16     0.825673
      2022-03-17     0.842147
      2022-03-18    -1.663753
      2022-03-19     1.440243
      2022-03-20    -1.833655
      2022-03-21     0.167561
      2022-03-22     0.296135
      2022-03-23     1.639017
      2022-03-24     1.280345
      2022-03-25    -1.326436
      2022-03-26     0.621524
      2022-03-27     1.513879
      2022-03-28     0.681857
      2022-03-29     0.447837
      2022-03-30    -1.069359
      2022-03-31     0.399382
      2022-04-01    -0.337783
      2022-04-02     1.007307
Freq: D, Name: A, dtype: float64
```

```
[ ]: import seaborn as sns
kashti = sns.load_dataset("titanic")
```

```
[ ]: kashti.head(5)
```

```
[ ]: Unnamed: 0  survived  pclass    sex   age  sibsp  parch    fare embarked \
0            0         0        3   male  22.0    1     0   7.2500         S
1            1         1        1  female 38.0    1     0  71.2833         C
2            2         1        3  female 26.0    0     0   7.9250         S
3            3         1        1  female 35.0    1     0  53.1000         S
4            4         0        3   male  35.0    0     0   8.0500         S
```

```

class  who  adult_male  deck  embark_town  alive  alone
0  Third   man         True  NaN  Southampton    no  False
1  First  woman        False   C    Cherbourg   yes  False
2  Third  woman        False  NaN  Southampton   yes  True
3  First  woman        False   C    Southampton   yes  False
4  Third   man         True  NaN  Southampton    no  True
```

```
[ ]: # Saving Data Frame into csv and excel file
```

```
kashti.to_csv('kashti.csv')
```

```
[ ]: # Basic statistics
```

```
kashti.describe()
```

```
[ ]: Unnamed: 0    survived    pclass    age    sibsp    parch \
count  891.000000  891.000000  891.000000  714.000000  891.000000  891.000000
mean    445.000000    0.383838    2.308642    29.699118    0.523008    0.381594
std     257.353842    0.486592    0.836071    14.526497    1.102743    0.806057
min       0.000000    0.000000    1.000000     0.420000    0.000000    0.000000
25%     222.500000    0.000000    2.000000    20.125000    0.000000    0.000000
50%     445.000000    0.000000    3.000000    28.000000    0.000000    0.000000
75%     667.500000    1.000000    3.000000    38.000000    1.000000    0.000000
max     890.000000    1.000000    3.000000    80.000000    8.000000    6.000000
```

```

fare
count  891.000000
mean    32.204208
std     49.693429
min       0.000000
25%      7.910400
50%     14.454200
75%     31.000000
max     512.329200
```

```
[ ]: kashti.head(5)
```

```
[ ]: Unnamed: 0  survived  pclass    sex   age  sibsp  parch    fare embarked \
0          0          0        3   male  22.0    1    0   7.2500         S
1          1          1        1  female  38.0    1    0  71.2833         C
2          2          1        3  female  26.0    0    0   7.9250         S
3          3          1        1  female  35.0    1    0  53.1000         S
4          4          0        3   male  35.0    0    0   8.0500         S
```

```
      class  who  adult_male deck  embark_town alive  alone
0  Third   man      True   NaN  Southampton   no  False
1  First  woman     False    C    Cherbourg   yes  False
2  Third  woman     False   NaN  Southampton   yes   True
3  First  woman     False    C    Southampton   yes  False
4  Third   man      True   NaN  Southampton   no   True
```

```
[ ]: new =kashti.drop(["deck","embark_town","alone"], axis=1)
```

```
[ ]: new.head(5)
```

```
[ ]: Unnamed: 0  survived  pclass    sex   age  sibsp  parch    fare embarked \
0          0          0        3   male  22.0    1    0   7.2500         S
1          1          1        1  female  38.0    1    0  71.2833         C
2          2          1        3  female  26.0    0    0   7.9250         S
3          3          1        1  female  35.0    1    0  53.1000         S
4          4          0        3   male  35.0    0    0   8.0500         S
```

```
      class  who  adult_male alive
0  Third   man      True    no
1  First  woman     False   yes
2  Third  woman     False   yes
3  First  woman     False   yes
4  Third   man      True    no
```

```
[ ]: new.groupby(["sex","class"]).mean()
```

```
[ ]: Unnamed: 0  survived  pclass    age    sibsp    parch \
sex  class
female First  468.212766  0.968085    1.0  34.611765  0.553191  0.457447
      Second  442.105263  0.921053    2.0  28.722973  0.486842  0.605263
      Third  398.729167  0.500000    3.0  21.750000  0.895833  0.798611
male   First  454.729508  0.368852    1.0  41.281386  0.311475  0.278689
      Second  446.962963  0.157407    2.0  30.740707  0.342593  0.222222
      Third  454.515850  0.135447    3.0  26.507589  0.498559  0.224784

      fare  adult_male
sex  class
female First  106.125798    0.000000
      Second  21.970121    0.000000
```

	Third	16.118810	0.000000
male	First	67.226127	0.975410
	Second	19.741782	0.916667
	Third	12.661633	0.919308

```
[ ]: new.value_counts(["survived"])
```

```
[ ]: survived
0      549
1      342
dtype: int64
```

```
[ ]: new.groupby(["sex", "survived"]).mean()
```

```
[ ]:
      Unnamed: 0    pclass    age    sibsp    parch \
sex  survived
female 0      433.851852  2.851852  25.046875  1.209877  1.037037
      1      428.699571  1.918455  28.847716  0.515021  0.515021
male   0      448.121795  2.476496  31.618056  0.440171  0.207265
      1      474.724771  2.018349  27.276022  0.385321  0.357798

      fare  adult_male
sex  survived
female 0      23.024385    0.000000
      1      51.938573    0.000000
male   0      21.960993    0.959402
      1      40.821484    0.807339
```

```
[ ]: new.groupby(["sex"]).mean()
```

```
[ ]:
      Unnamed: 0  survived    pclass    age    sibsp    parch \
sex
female 430.028662  0.742038  2.159236  27.915709  0.694268  0.649682
male   453.147314  0.188908  2.389948  30.726645  0.429809  0.235702

      fare  adult_male
sex
female 44.479818    0.000000
male   25.523893    0.930676
```

```
[ ]: # Childrens and Ladies first
new[new["age"]<18].groupby(["sex", "class"]).mean()
```

```
[ ]:
      Unnamed: 0  survived  pclass    age    sibsp    parch \
sex  class
female First  524.375000  0.875000    1.0  14.125000  0.500000  0.875000
      Second  368.250000  1.000000    2.0   8.333333  0.583333  1.083333
      Third  373.942857  0.542857    3.0   8.428571  1.571429  1.057143
```

male	First	525.500000	1.000000	1.0	8.230000	0.500000	2.000000
	Second	526.818182	0.818182	2.0	4.757273	0.727273	1.000000
	Third	436.953488	0.232558	3.0	9.963256	2.069767	1.000000

		fare	adult_male
sex	class		
female	First	104.083337	0.000000
	Second	26.241667	0.000000
	Third	18.727977	0.000000
male	First	116.072900	0.250000
	Second	25.659473	0.181818
	Third	22.752523	0.348837

## 2 For FAO dataset

```
[ ]: fao= pd.read_csv("fao.csv")
```

```
[ ]: fao.head(5)
```

```
[ ]: Unnamed: 0  Domain Code          Domain  Area Code      Area \
0              0      RFN  Fertilizers by Nutrient      165  Pakistan
1              1      RFN  Fertilizers by Nutrient      165  Pakistan
2              2      RFN  Fertilizers by Nutrient      165  Pakistan
3              3      RFN  Fertilizers by Nutrient      165  Pakistan
4              4      RFN  Fertilizers by Nutrient      165  Pakistan
```

	Element Code	Element	Item Code	Item	\
0	5157	Agricultural Use	3102	Nutrient nitrogen N (total)	
1	5157	Agricultural Use	3103	Nutrient phosphate P205 (total)	
2	5157	Agricultural Use	3103	Nutrient phosphate P205 (total)	
3	5157	Agricultural Use	3102	Nutrient nitrogen N (total)	
4	5157	Agricultural Use	3102	Nutrient nitrogen N (total)	

	Year Code	Year	Unit	Value	Flag	\
0	1961	1961	tonnes	41659	Qm	
1	1961	1961	tonnes	500	Qm	
2	1962	1962	tonnes	210	Qm	
3	1962	1962	tonnes	41160	Qm	
4	1963	1963	tonnes	67620	Qm	

	Flag	Description
0	Official data from questionnaires and/or natio...	
1	Official data from questionnaires and/or natio...	
2	Official data from questionnaires and/or natio...	
3	Official data from questionnaires and/or natio...	
4	Official data from questionnaires and/or natio...	

```
[ ]: fao.drop(["Flag Description"],axis=1)
```

```
[ ]:      Unnamed: 0 Domain Code      Domain Area Code      Area \
0          0      RFN  Fertilizers by Nutrient      165 Pakistan
1          1      RFN  Fertilizers by Nutrient      165 Pakistan
2          2      RFN  Fertilizers by Nutrient      165 Pakistan
3          3      RFN  Fertilizers by Nutrient      165 Pakistan
4          4      RFN  Fertilizers by Nutrient      165 Pakistan
..      ...      ...      ...      ...      ...
167      167      RFN  Fertilizers by Nutrient      165 Pakistan
168      168      RFN  Fertilizers by Nutrient      165 Pakistan
169      169      RFN  Fertilizers by Nutrient      165 Pakistan
170      170      RFN  Fertilizers by Nutrient      165 Pakistan
171      171      RFN  Fertilizers by Nutrient      165 Pakistan
```

```
      Element Code      Element Item Code \
0          5157  Agricultural Use      3102
1          5157  Agricultural Use      3103
2          5157  Agricultural Use      3103
3          5157  Agricultural Use      3102
4          5157  Agricultural Use      3102
..      ...      ...      ...
167      5157  Agricultural Use      3102
168      5157  Agricultural Use      3104
169      5157  Agricultural Use      3104
170      5157  Agricultural Use      3102
171      5157  Agricultural Use      3103
```

```
      Item Year Code Year      Unit      Value Flag
0      Nutrient nitrogen N (total)      1961 1961 tonnes      41659 Qm
1      Nutrient phosphate P205 (total)      1961 1961 tonnes          500 Qm
2      Nutrient phosphate P205 (total)      1962 1962 tonnes          210 Qm
3      Nutrient nitrogen N (total)      1962 1962 tonnes      41160 Qm
4      Nutrient nitrogen N (total)      1963 1963 tonnes      67620 Qm
..      ...      ...      ...      ...      ...
167      Nutrient nitrogen N (total)      2018 2018 tonnes      3446922 Fm
168      Nutrient potash K20 (total)      2018 2018 tonnes          54173 Fm
169      Nutrient potash K20 (total)      2019 2019 tonnes          47260 Qm
170      Nutrient nitrogen N (total)      2019 2019 tonnes      3505356 Qm
171      Nutrient phosphate P205 (total)      2019 2019 tonnes      1099707 Qm
```

[172 rows x 14 columns]

```
[ ]: fao.value_counts(["Value"])
```

```
[ ]: Value
54173      2
```

```

3446922    2
1257773    2
144        1
678840     1
..
38642      1
40078      1
41160      1
41659      1
3505356    1
Length: 169, dtype: int64

```

```
[ ]: fao.groupby(["Year Code", "Value"]).mean()
```

```
[ ]:
      Unnamed: 0  Area Code  Element Code  Item Code  Year
Year Code Value
1961      500           1         165         5157      3103  1961
      41659           0         165         5157      3102  1961
1962      210           2         165         5157      3103  1962
      41160           3         165         5157      3102  1962
1963      630           5         165         5157      3103  1963
...
2018  1257773        166         165         5157      3103  2018
      3446922        167         165         5157      3102  2018
2019   47260        169         165         5157      3104  2019
      1099707        171         165         5157      3103  2019
      3505356        170         165         5157      3102  2019

```

```
[172 rows x 5 columns]
```

```
[ ]: fao[fao["Year Code"]>2001].groupby(["Value", "Year Code"]).mean()
```

```
[ ]:
      Unnamed: 0  Area Code  Element Code  Item Code  Year
Value  Year Code
15499  2009           139         165         5157      3104  2009
16564  2008           138         165         5157      3104  2008
19427  2012           150         165         5157      3104  2012
19849  2013           151         165         5157      3104  2013
20500  2002           120         165         5157      3104  2002
23944  2003           121         165         5157      3104  2003
25291  2015           157         165         5157      3104  2015
26449  2004           126         165         5157      3104  2004
28013  2011           145         165         5157      3104  2011
28954  2016           162         165         5157      3104  2016
29549  2014           156         165         5157      3104  2014
35301  2005           127         165         5157      3104  2005
36408  2006           132         165         5157      3104  2006

```



37545	2010	144	165	5157	3104	2010
43924	2007	133	165	5157	3104	2007
47260	2019	169	165	5157	3104	2019
54173	2017	163	165	5157	3104	2017
	2018	168	165	5157	3104	2018
565675	2008	136	165	5157	3103	2008
647360	2002	118	165	5157	3103	2002
653181	2011	147	165	5157	3103	2011
669307	2003	123	165	5157	3103	2003
678840	2012	148	165	5157	3103	2012
756700	2010	142	165	5157	3103	2010
761031	2004	124	165	5157	3103	2004
778714	2009	141	165	5157	3103	2009
816044	2007	135	165	5157	3103	2007
871416	2005	129	165	5157	3103	2005
875452	2013	153	165	5157	3103	2013
906749	2006	130	165	5157	3103	2006
936148	2014	154	165	5157	3103	2014
993338	2015	159	165	5157	3103	2015
1099707	2019	171	165	5157	3103	2019
1209145	2016	160	165	5157	3103	2016
1257773	2017	165	165	5157	3103	2017
	2018	166	165	5157	3103	2018
2384980	2002	119	165	5157	3102	2002
2456009	2003	122	165	5157	3102	2003
2570230	2004	125	165	5157	3102	2004
2714053	2007	134	165	5157	3102	2007
2818779	2005	128	165	5157	3102	2005
2850771	2012	149	165	5157	3102	2012
2868984	2006	131	165	5157	3102	2006
2928057	2008	137	165	5157	3102	2008
3139263	2014	155	165	5157	3102	2014
3160507	2015	158	165	5157	3102	2015
3187339	2011	146	165	5157	3102	2011
3239557	2013	152	165	5157	3102	2013
3241907	2016	161	165	5157	3102	2016
3270531	2010	143	165	5157	3102	2010
3446922	2017	164	165	5157	3102	2017
	2018	167	165	5157	3102	2018
3451900	2009	140	165	5157	3102	2009
3505356	2019	170	165	5157	3102	2019

[ ]: