

Pandas : Basic commands

```
import pandas as pd
```

1. To read CSV :

```
dataframe = pd.read_csv("file path")
```

To read XLS :

```
dataframe = pd.read_xls("file path")
```

To read TXT :

```
dataframe = pd.read_table("file path")
```

1. To view first n columns :

```
dataframe.head(n)
```

To view last n columns :

```
dataframe.tail(n)
```

1. To get shape of dataframe :

```
dataframe.shape
```

1. To get index of dataframe :

```
dataframe.index
```

1. To get all the column names of the dataframe :

```
dataframe.columns
```

1. Get all the datatypes present in the dataframe :

```
dataframe.dtypes
```

1. To get features of numerical data in dataframe :

```
dataframe.describe()
```

1. To get a concise summary of dataframe :

```
dataframe.info ()
```

1. Lookup data using labels

```
dataframe.at[ <row no.> , '<column name>']
```

Look up data using integer indices

```
dataframe.iat[<row no.> , <column-name> ]
```

1. Get number of null values in dataframe

```
dataframe.isnull().sum()
```

Let us apply all these commands to a dataset for example :

Dataset : <https://www.kaggle.com/datasets/srolka/ecommerce-customers>

```
dataframe1 = pd.read_csv("Ecommerce_Customers.csv")
```

```
dataframe1.head()
```

	Email \
0	mstephenson@fernandez.com
1	hduke@hotmail.com
2	pallen@yahoo.com
3	riverarebecca@gmail.com
4	mstephens@davidson-herman.com

	Address	Avatar
0	835 Frank Tunnel\nWrightmouth, MI 82180-9605	Violet
1	4547 Archer Common\nDiazchester, CA 06566-8576	DarkGreen
2	24645 Valerie Unions Suite 582\nCobbborough, D...	Bisque
3	1414 David Throughway\nPort Jason, OH 22070-1220	SaddleBrown
4	14023 Rodriguez Passage\nPort Jacobville, PR 3...	MediumAquaMarine

	Avg. Session Length	Time on App	Time on Website	Length of Membership \
0	34.497268	12.655651	39.577668	4.082621
1	31.926272	11.109461	37.268959	2.664034
2	33.000915	11.330278	37.110597	4.104543
3	34.305557	13.717514	36.721283	3.120179
4	33.330673	12.795189	37.536653	4.446308

	Yearly Amount Spent
0	587.951054
1	392.204933
2	487.547505
3	581.852344
4	599.406092

```
dataframe1.tail()
```

	Email \
495	lewisjessica@craig-evans.com
496	katrina56@gmail.com
497	dale88@hotmail.com
498	cwilson@hotmail.com
499	hannahwilson@davidson.com

	Address	Avatar
495	4483 Jones Motorway Suite 872\nLake Jamiefurt,...	Tan
496	172 Owen Divide Suite 497\nWest Richard, CA 19320	PaleVioletRed
497	0787 Andrews Ranch Apt. 633\nSouth Chadburgh, ...	Cornsilk
498	680 Jennifer Lodge Apt. 808\nBrendacheater, TX...	Teal
499	49791 Rachel Heights Apt. 898\nEast Drewboroug...	DarkMagenta

	Avg. Session Length	Time on App	Time on Website	Length of Membership \
495	33.237660	13.566160	36.417985	3.746573
496	34.702529	11.695736	37.190268	3.576526
497	32.646777	11.499409	38.332576	4.958264
498	33.322501	12.391423	36.840086	2.336485
499	33.715981	12.418808	35.771016	2.735160

	Yearly Amount Spent
495	573.847438
496	529.049004
497	551.620145
498	456.469510
499	497.778642

```
dataframe1.shape
```

```
(500, 8)
```

```
dataframe1.columns
```

```
Index(['Email', 'Address', 'Avatar', 'Avg. Session Length', 'Time on App',  
      'Time on Website', 'Length of Membership', 'Yearly Amount  
Spent'],  
      dtype='object')
```

```
dataframe1.dtypes
```

```
Email          object  
Address         object  
Avatar          object  
Avg. Session Length  float64  
Time on App       float64  
Time on Website    float64  
Length of Membership float64  
Yearly Amount Spent float64  
dtype: object
```

```
dataframe1.describe()
```

	Avg. Session Length	Time on App	Time on Website \
count	500.000000	500.000000	500.000000
mean	33.053194	12.052488	37.060445
std	0.992563	0.994216	1.010489
min	29.532429	8.508152	33.913847
25%	32.341822	11.388153	36.349257
50%	33.082008	11.983231	37.069367
75%	33.711985	12.753850	37.716432
max	36.139662	15.126994	40.005182

	Length of Membership	Yearly Amount Spent
count	500.000000	500.000000
mean	3.533462	499.314038
std	0.999278	79.314782
min	0.269901	256.670582
25%	2.930450	445.038277
50%	3.533975	498.887875
75%	4.126502	549.313828
max	6.922689	765.518462

```
dataframe1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 500 entries, 0 to 499
```

```
Data columns (total 8 columns):
```

#	Column	Non-Null Count	Dtype
---	-----	-----	-----

```
0    Email      500 non-null    object
1    Address    500 non-null    object
2    Avatar     500 non-null    object
3    Avg. Session Length  500 non-null    float64
4    Time on App  500 non-null    float64
5    Time on Website  500 non-null    float64
6    Length of Membership  500 non-null    float64
7    Yearly Amount Spent  500 non-null    float64
```

```
dtypes: float64(5), object(3)
```

```
memory usage: 31.4+ KB
```

```
dataframe1.iat[(1,2)]
```

```
'DarkGreen'
```

```
dataframe1.at[4,"Yearly Amount Spent"]
```

```
599.4060920457634
```

```
dataframe1["Length of Membership"]
```

```
0      4.082621
1      2.664034
2      4.104543
3      3.120179
4      4.446308
```

```
...
495     3.746573
496     3.576526
497     4.958264
498     2.336485
499     2.735160
```

```
Name: Length of Membership, Length: 500, dtype: float64
```

```
dataframe1["Email"]
```

```
0      mstephenson@fernandez.com
1      hduke@hotmail.com
2      pallen@yahoo.com
3      riverarebecca@gmail.com
4      mstephens@davidson-herman.com
```

```
...
495     lewisjessica@craig-evans.com
496     katrina56@gmail.com
497     dale88@hotmail.com
498     cwilson@hotmail.com
499     hannahwilson@davidson.com
```

```
Name: Email, Length: 500, dtype: object
```

```
dataframe1.isnull().sum()
```

```
Email          0
Address        0
Avatar         0
Avg. Session Length  0
Time on App    0
Time on Website 0
Length of Membership 0
Yearly Amount Spent 0
dtype: int64
```

Notebook created by :Shivang Kainthola (<https://www.linkedin.com/in/shivang-kainthola-2835151b9/>)