What is Pandas

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

https://pandas.pydata.org/about/index.html

Pandas Series

A Pandas Series is like a column in a table. It is a 1-D array holding data of any type.

Importing Pandas

```
import numpy as np
import pandas as pd
```

Series from lists

```
# string
country = ['India','Pakistan','USA','Nepal','Srilanka']

pd.Series(country)
```

```
0       India
1    Pakistan
2         USA
3       Nepal
4    Srilanka
dtype: object
```

```
# integers
runs = [13,24,56,78,100]

runs_ser = pd.Series(runs)
```

```
# custom index
marks = [67,57,89,100]
subjects = ['maths','english','science','hindi']

pd.Series(marks,index=subjects)
```

```
maths       67
english     57
science     89
hindi      100
dtype: int64
```

```
# setting a name
marks = pd.Series(marks,index=subjects,name='Nitish ke marks')
marks
```

```
maths       67
english     57
science     89
hindi      100
Name: Nitish ke marks, dtype: int64
```

Series from dict

```
marks = {
    'maths':67,
    'english':57,
    'science':89,
    'hindi':100
}

marks_series = pd.Series(marks,name='nitish ke marks')
marks_series
```

```
maths       67
english     57
```

```
science     89
hindi      100
Name: nitish ke marks, dtype: int64
```

Series Attributes

```
# size
marks_series.size
```

```
4
```

```
# dtype
marks_series.dtype
```

```
dtype('int64')
```

```
# name
marks_series.name
```

```
'nitish ke marks'
```

```
# is_unique
marks_series.is_unique

pd.Series([1,1,2,3,4,5]).is_unique
```

```
False
```

```
# index
marks_series.index
```

```
Index(['maths', 'english', 'science', 'hindi'], dtype='object')
```

```
runs_ser.index
```

```
RangeIndex(start=0, stop=5, step=1)
```

```
# values
marks_series.values
```

```
array([ 67,  57,  89, 100])
```

Series using read_csv

```
# with one col
subs = pd.read_csv('/content/subs.csv',squeeze=True)
subs
```

```
0       48
1       57
2       40
3       43
4       44
      ...
360    231
361    226
362    155
363    144
364    172
Name: Subscribers gained, Length: 365, dtype: int64
```

```
# with 2 cols
vk = pd.read_csv('/content/kohli_ipl.csv',index_col='match_no',squeeze=True)
vk
```

```
match_no
1       1
2      23
3      13
4      12
5       1
      ..
211     0
212    20
213    73
214    25
```

```
215     7
Name: runs, Length: 215, dtype: int64
```

```
movies = pd.read_csv('/content/bollywood.csv',index_col='movie',squeeze=True)
movies
```

```
movie
Uri: The Surgical Strike                   Vicky Kaushal
Battalion 609                                Vicky Ahuja
The Accidental Prime Minister (film)         Anupam Kher
Why Cheat India                             Emraan Hashmi
Evening Shadows                           Mona Ambegaonkar
                                                  ...
Hum Tumhare Hain Sanam                     Shah Rukh Khan
Aankhen (2002 film)                       Amitabh Bachchan
Saathiya (film)                               Vivek Oberoi
Company (film)                                  Ajay Devgn
Awara Paagal Deewana                          Akshay Kumar
Name: lead, Length: 1500, dtype: object
```

Series methods

```
# head and tail
subs.head()
```

```
0    48
1    57
2    40
3    43
4    44
Name: Subscribers gained, dtype: int64
```

```
vk.head(3)
```

```
match_no
1     1
2    23
3    13
Name: runs, dtype: int64
```

```
vk.tail(10)
```

```
match_no
206     0
207     0
208     9
209    58
210    30
211     0
212    20
213    73
214    25
215     7
Name: runs, dtype: int64
```

```
# sample
movies.sample(5)
```

```
movie
Arjun: The Warrior Prince    Yudhveer Bakoliya
Viceroy's House (film)         Hugh Bonneville
Joggers' Park (film)           Victor Banerjee
Tere Mere Phere                   Vinay Pathak
Mission Mangal                    Akshay Kumar
Name: lead, dtype: object
```

```
# value_counts -> movies
movies.value_counts()
```

```
Akshay Kumar        48
Amitabh Bachchan    45
Ajay Devgn          38
Salman Khan         31
Sanjay Dutt         26
                    ..
Diganth              1
Parveen Kaur         1
Seema Azmi           1
Akanksha Puri        1
Edwin Fernandes      1
Name: lead, Length: 566, dtype: int64
```

```
# sort_values -> inplace
vk.sort_values(ascending=False).head(1).values[0]
```

```
113
```

```
vk.sort_values(ascending=False)
```

```
match_no
128    113
126    109
123    108
164    100
120    100
       ...
93       0
211      0
130      0
8        0
135      0
Name: runs, Length: 215, dtype: int64
```

```
# sort_index -> inplace -> movies
movies.sort_index(ascending=False,inplace=True)
```

```
movies
```

```
movie
Zor Lagaa Ke...Haiya!            Meghan Jadhav
Zokkomon                       Darsheel Safary
Zindagi Tere Naam           Mithun Chakraborty
Zindagi Na Milegi Dobara        Hrithik Roshan
Zindagi 50-50                      Veena Malik
                                       ...
2 States (2014 film)              Arjun Kapoor
1971 (2007 film)                 Manoj Bajpayee
1920: The Evil Returns              Vicky Ahuja
1920: London                      Sharman Joshi
1920 (film)                     Rajniesh Duggall
Name: lead, Length: 1500, dtype: object
```

```
vk.sort_values(inplace=True)
```

```
vk
```

```
match_no
87       0
211      0
207      0
206      0
91       0
       ...
164    100
120    100
123    108
126    109
128    113
Name: runs, Length: 215, dtype: int64
```

Series Maths Methods

```
# count
vk.count()
```

```
215
```

```
# sum -> product
subs.sum()
```

```
49510
```

```
# mean -> median -> mode -> std -> var
subs.mean()
print(vk.median())
print(movies.mode())
print(subs.std())
print(vk.var())
```

```
24.0
0    Akshay Kumar
```

```
dtype: object
62.6750230372527
688.0024777222343
```

```
# min/max
subs.max()
```

```
396
```

```
# describe
subs.describe()
```

```
count    365.000000
mean     135.643836
std       62.675023
min       33.000000
25%       88.000000
50%      123.000000
75%      177.000000
max      396.000000
Name: Subscribers gained, dtype: float64
```

Series Indexing

```
# integer indexing
x = pd.Series([12,13,14,35,46,57,58,79,9])
x
```

```
0    12
1    13
2    14
3    35
4    46
5    57
6    58
7    79
8     9
dtype: int64
```

```
# negative indexing
x[-1]
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
/usr/local/lib/python3.8/dist-packages/pandas/core/indexes/range.py in
get_loc(self, key, method, tolerance)
    384                 try:
--> 385                     return self._range.index(new_key)
    386                 except ValueError as err:

ValueError: -1 is not in range

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

KeyError                                  Traceback (most recent call last)
                                3 frames
/usr/local/lib/python3.8/dist-packages/pandas/core/indexes/range.py in
get_loc(self, key, method, tolerance)
    385                     return self._range.index(new_key)
    386                 except ValueError as err:
--> 387                     raise KeyError(key) from err
    388             raise KeyError(key)
    389         return super().get_loc(key, method=method,
tolerance=tolerance)

KeyError: -1
```

```
movies
```

```
movie
Zor Lagaa Ke...Haiya!           Meghan Jadhav
Zokkomon                      Darsheel Safary
Zindagi Tere Naam           Mithun Chakraborty
Zindagi Na Milegi Dobara        Hrithik Roshan
Zindagi 50-50                      Veena Malik
                                   ...
2 States (2014 film)             Arjun Kapoor
1971 (2007 film)                Manoj Bajpayee
1920: The Evil Returns             Vicky Ahuja
1920: London                     Sharman Joshi
1920 (film)                    Rajniesh Duggall
Name: lead, Length: 1500, dtype: object
```

```
vk[-1]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.8/dist-packages/pandas/core/indexes/base.py in
get_loc(self, key, method, tolerance)
   3360             try:
-> 3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:

 ---------------------------------- 5 frames ----------------------------------
pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.Int64HashTable.get_item()

pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.Int64HashTable.get_item()

KeyError: -1

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

KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.8/dist-packages/pandas/core/indexes/base.py in
get_loc(self, key, method, tolerance)
   3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:
-> 3363                 raise KeyError(key) from err
   3364
   3365         if is_scalar(key) and isna(key) and not self.hasnans:

KeyError: -1
```

```
marks_series[-1]
```

```
100
```

```
# slicing
vk[5:16]
```

```
match_no
6      9
7     34
8      0
9     21
10     3
11    10
12    38
13     3
14    11
15    50
16     2
Name: runs, dtype: int64
```

```
# negative slicing
vk[-5:]
```

```
match_no
211     0
212    20
213    73
214    25
215     7
Name: runs, dtype: int64
```

```
movies[::2]
```

```
movie
Zor Lagaa Ke...Haiya!         Meghan Jadhav
Zindagi Tere Naam        Mithun Chakraborty
Zindagi 50-50                   Veena Malik
Zinda (film)                    Sanjay Dutt
Zid (2014 film)              Mannara Chopra
                                   ...
3 Storeys                       Aisha Ahmed
3 Deewarein                Naseeruddin Shah
22 Yards                        Barun Sobti
1971 (2007 film)             Manoj Bajpayee
1920: London                  Sharman Joshi
Name: lead, Length: 750, dtype: object
```

```
# fancy indexing
vk[[1,3,4,5]]
```

```
match_no
1     1
3    13
4    12
5     1
Name: runs, dtype: int64
```

```
# indexing with labels -> fancy indexing
movies['2 States (2014 film)']
```

```
'Arjun Kapoor'
```

Editing Series

```
# using indexing
marks_series[1] = 100
marks_series
```

```
maths       67
english    100
science     89
hindi      100
Name: nitish ke marks, dtype: int64
```

```
# what if an index does not exist
marks_series['evs'] = 100
```

```
marks_series
```

```
maths       67
english    100
science     89
hindi      100
sst         90
evs        100
Name: nitish ke marks, dtype: int64
```

```
# slicing
runs_ser[2:4] = [100,100]
runs_ser
```

```
0     13
1     24
2    100
3    100
4    100
dtype: int64
```

```
# fancy indexing
runs_ser[[0,3,4]] = [0,0,0]
runs_ser
```

```
0      0
1     24
2    100
3      0
4      0
dtype: int64
```

```
# using index label
movies['2 States (2014 film)'] = 'Alia Bhatt'
movies
```

```
movie
Zor Lagaa Ke...Haiya!           Meghan Jadhav
Zokkomon                      Darsheel Safary
Zindagi Tere Naam           Mithun Chakraborty
Zindagi Na Milegi Dobara        Hrithik Roshan
Zindagi 50-50                      Veena Malik
                                   ...
2 States (2014 film)               Alia Bhatt
1971 (2007 film)                Manoj Bajpayee
1920: The Evil Returns             Vicky Ahuja
1920: London                     Sharman Joshi
1920 (film)                   Rajniesh Duggall
Name: lead, Length: 1500, dtype: object
```

Copy and Views

Series with Python Functionalities

```
# len/type/dir/sorted/max/min
print(len(subs))
print(type(subs))
print(dir(subs))
print(sorted(subs))
print(min(subs))
print(max(subs))
```

```
365
<class 'pandas.core.series.Series'>
['T', '_AXIS_LEN', '_AXIS_ORDERS', '_AXIS_REVERSED', '_AXIS_TO_AXIS_NUMBER', '_HANDLED_TYPES', '__abs__', '__add__', '__
[33, 33, 35, 37, 39, 40, 40, 40, 40, 42, 42, 43, 44, 44, 44, 45, 46, 46, 48, 49, 49, 49, 49, 50, 50, 50, 51, 54, 56, 56,
33
396
```

```
# type conversion
list(marks_series)
```

```
[67, 100, 89, 100, 90, 100]
```

```
dict(marks_series)
```

```
{'maths': 67,
 'english': 100,
 'science': 89,
 'hindi': 100,
 'sst': 90,
 'evs': 100}
```

```
# membership operator

'2 States (2014 film)' in movies
```

```
True
```

```
'Alia Bhatt' in movies.values
```

```
True
```

```
movies
```

```
movie
Zor Lagaa Ke...Haiya!            Meghan Jadhav
Zokkomon                       Darsheel Safary
Zindagi Tere Naam            Mithun Chakraborty
Zindagi Na Milegi Dobara         Hrithik Roshan
Zindagi 50-50                       Veena Malik
                                    ...
2 States (2014 film)                 Alia Bhatt
1971 (2007 film)                 Manoj Bajpayee
1920: The Evil Returns              Vicky Ahuja
1920: London                      Sharman Joshi
1920 (film)                     Rajniesh Duggall
Name: lead, Length: 1500, dtype: object
```

```
# looping
for i in movies.index:
  print(i)
```

```
Zor Lagaa Ke...Haiya!
Zokkomon
Zindagi Tere Naam
Zindagi Na Milegi Dobara
Zindagi 50-50
Zindaggi Rocks
Zinda (film)
Zila Ghaziabad
Zid (2014 film)
Zero (2018 film)
Zeher
Zed Plus
Zameer: The Fire Within
```

```
Zameen (2003 film)
Zamaanat
Yuvvraaj
Yuva
Yun Hota Toh Kya Hota
Youngistaan
Yeh Saali Aashiqui
Yeh Mera India
Yeh Lamhe Judaai Ke
Yeh Khula Aasmaan
Yeh Jawaani Hai Deewani
Yeh Hai India
Yeh Hai Bakrapur
Yeh Dooriyan
Yeh Dil
Yatra (2007 film)
Yamla Pagla Deewana: Phir Se
Yamla Pagla Deewana
Yakeen (2005 film)
Yadvi – The Dignified Princess
Yaaram (2019 film)
Ya Rab
Xcuse Me
Woodstock Villa
Woh Lamhe...
Why Cheat India
What's Your Raashee?
What the Fish
Well Done Abba
Welcome to Sajjanpur
Welcome Back (film)
Welcome 2 Karachi
Welcome (2007 film)
Wedding Pullav
Wedding Anniversary
Waris Shah: Ishq Daa Waaris
War Chhod Na Yaar
Waqt: The Race Against Time
Wanted (2009 film)
Wake Up Sid
Wake Up India
Wajah Tum Ho
Waiting (2015 film)
Waisa Bhi Hota Hai Part II
Wah Taj
```

```
# Arithmetic Operators(Broadcasting)
100 + marks_series
```

```
maths      167
english    200
science    189
hindi      200
sst        190
evs        200
Name: nitish ke marks, dtype: int64
```

```
# Relational Operators

vk >= 50
```

```
match_no
1      False
2      False
3      False
4      False
5      False
       ...
211    False
212    False
213     True
214    False
215    False
Name: runs, Length: 215, dtype: bool
```

Boolean Indexing on Series

```
# Find no of 50's and 100's scored by kohli
vk[vk >= 50].size
```

```
50
```

```
# find number of ducks
vk[vk == 0].size
```

```
9
```

```
# Count number of day when I had more than 200 subs a day
subs[subs > 200].size
```

```
59
```

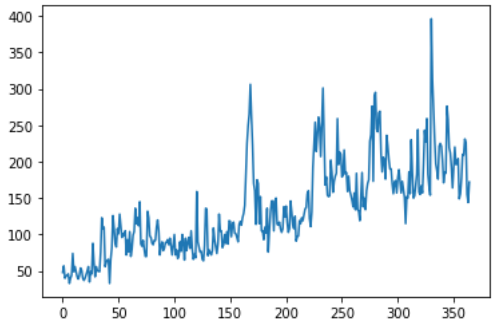
```
# find actors who have done more than 20 movies
num_movies = movies.value_counts()
num_movies[num_movies > 20]
```

```
Akshay Kumar        48
Amitabh Bachchan    45
Ajay Devgn          38
Salman Khan         31
Sanjay Dutt         26
Shah Rukh Khan      22
Emraan Hashmi       21
Name: lead, dtype: int64
```

Plotting Graphs on Series

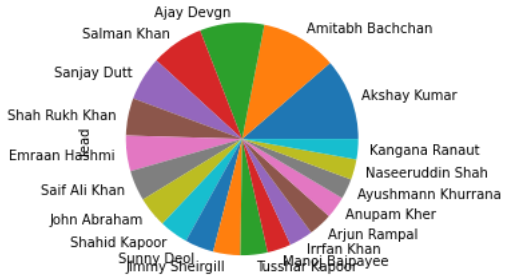
```
subs.plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f54e0531a60>
```




```
movies.value_counts().head(20).plot(kind='pie')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f54e04f6850>
```



Some Important Series Methods

```
# astype
# between
# clip
# drop_duplicates
# isnull
# dropna
# fillna
# isin
# apply
# copy
```

```
import numpy as np
import pandas as pd
```

```
subs = pd.read_csv('/content/subs.csv',squeeze=True)
subs
```

```
0       48
1       57
2       40
3       43
4       44
      ...
360    231
361    226
362    155
363    144
364    172
Name: Subscribers gained, Length: 365, dtype: int64
```

```
vk = pd.read_csv('/content/kohli_ipl.csv',index_col='match_no',squeeze=True)
vk
```

```
match_no
1       1
2      23
3      13
4      12
5       1
       ..
211     0
212    20
213    73
214    25
215     7
Name: runs, Length: 215, dtype: int64
```

```
movies = pd.read_csv('/content/bollywood.csv',index_col='movie',squeeze=True)
movies
```

```
movie
Uri: The Surgical Strike                   Vicky Kaushal
Battalion 609                                Vicky Ahuja
The Accidental Prime Minister (film)         Anupam Kher
Why Cheat India                             Emraan Hashmi
Evening Shadows                           Mona Ambegaonkar
                                                 ...
Hum Tumhare Hain Sanam                      Shah Rukh Khan
Aankhen (2002 film)                       Amitabh Bachchan
Saathiya (film)                              Vivek Oberoi
Company (film)                                 Ajay Devgn
Awara Paagal Deewana                         Akshay Kumar
Name: lead, Length: 1500, dtype: object
```

```
# astype
import sys
sys.getsizeof(vk)
```

```
3456
```

```
sys.getsizeof(vk.astype('int16'))
```

```
2166
```

```
# between
vk[vk.between(51,99)].size
```

```
43
```

```
# clip
subs
```

```
0       48
1       57
2       40
3       43
4       44
      ...
360    231
361    226
362    155
363    144
364    172
Name: Subscribers gained, Length: 365, dtype: int64
```

```
subs.clip(100,200)
```

```
0      100
1      100
2      100
3      100
4      100
      ...
360    200
361    200
362    155
363    144
364    172
Name: Subscribers gained, Length: 365, dtype: int64
```

```
# drop_duplicates
temp = pd.Series([1,1,2,2,3,3,4,4])
temp
```

```
0    1
1    1
2    2
3    2
4    3
5    3
6    4
7    4
dtype: int64
```

```
temp.drop_duplicates(keep='last')
```

```
1    1
3    2
5    3
7    4
dtype: int64
```

```
temp.duplicated().sum()
```

```
4
```

```
vk.duplicated().sum()
```

```
137
```

```
movies.drop_duplicates()
```

```
movie
Uri: The Surgical Strike                   Vicky Kaushal
Battalion 609                                Vicky Ahuja
The Accidental Prime Minister (film)         Anupam Kher
Why Cheat India                            Emraan Hashmi
Evening Shadows                         Mona Ambegaonkar
                                                  ...
Sssshhh...                              Tanishaa Mukerji
Rules: Pyaar Ka Superhit Formula                 Tanuja
Right Here Right Now (film)                       Ankit
Talaash: The Hunt Begins...                Rakhee Gulzar
The Pink Mirror                           Edwin Fernandes
Name: lead, Length: 566, dtype: object
```

```
temp = pd.Series([1,2,3,np.nan,5,6,np.nan,8,np.nan,10])
temp
```

```
0     1.0
1     2.0
2     3.0
3     NaN
4     5.0
5     6.0
6     NaN
7     8.0
8     NaN
9    10.0
dtype: float64
```

```
temp.size
```

```
10
```

```
temp.count()
```

```
7
```

```
# isnull
temp.isnull().sum()
```

```
3
```

```
# dropna
temp.dropna()
```

```
0     1.0
1     2.0
2     3.0
4     5.0
5     6.0
7     8.0
9    10.0
dtype: float64
```

```
# fillna
temp.fillna(temp.mean())
```

```
0     1.0
1     2.0
2     3.0
3     5.0
4     5.0
5     6.0
6     5.0
7     8.0
8     5.0
9    10.0
dtype: float64
```

```
# isin
vk[(vk == 49) | (vk == 99)]
```

```
match_no
82    99
86    49
Name: runs, dtype: int64
```

```
vk[vk.isin([49,99])]
```

```
match_no
82    99
86    49
Name: runs, dtype: int64
```

```
# apply
movies
```

```
movie
Uri: The Surgical Strike                   Vicky Kaushal
Battalion 609                                Vicky Ahuja
The Accidental Prime Minister (film)         Anupam Kher
Why Cheat India                             Emraan Hashmi
Evening Shadows                           Mona Ambegaonkar
                                                 ...
Hum Tumhare Hain Sanam                     Shah Rukh Khan
Aankhen (2002 film)                       Amitabh Bachchan
Saathiya (film)                               Vivek Oberoi
Company (film)                                 Ajay Devgn
Awara Paagal Deewana                         Akshay Kumar
Name: lead, Length: 1500, dtype: object
```

```
movies.apply(lambda x:x.split()[0].upper())
```

```
movie
Uri: The Surgical Strike                   VICKY
Battalion 609                              VICKY
The Accidental Prime Minister (film)     ANUPAM
Why Cheat India                           EMRAAN
```

```
Evening Shadows                          MONA
                                        ...
Hum Tumhare Hain Sanam                   SHAH
Aankhen (2002 film)                    AMITABH
Saathiya (film)                          VIVEK
Company (film)                           AJAY
Awara Paagal Deewana                    AKSHAY
Name: lead, Length: 1500, dtype: object
```

```
subs
```

```
0       48
1       57
2       40
3       43
4       44
      ...
360    231
361    226
362    155
363    144
364    172
Name: Subscribers gained, Length: 365, dtype: int64
```

```
subs.apply(lambda x:'good day' if x > subs.mean() else 'bad day')
```

```
0       bad day
1       bad day
2       bad day
3       bad day
4       bad day
         ...
360    good day
361    good day
362    good day
363    good day
364    good day
Name: Subscribers gained, Length: 365, dtype: object
```

```
subs.mean()
```

```
135.64383561643837
```

```
# copy
```

```
vk
```

```
match_no
1       1
2      23
3      13
4      12
5       1
       ..
211     0
212    20
213    73
214    25
215     7
Name: runs, Length: 215, dtype: int64
```

```
new = vk.head()
```

```
new
```

```
match_no
1     1
2    23
3    13
4    12
5     1
Name: runs, dtype: int64
```

```
new[1] = 1
```

```
new = vk.head().copy()
```

```
new[1] = 100
```

```
new
```

```
match_no
1    100
2     23
3     13
4     12
5      1
Name: runs, dtype: int64
```

```
vk
```

```
match_no
1       1
2      23
3      13
4      12
5       1
       ..
211     0
212    20
213    73
214    25
215     7
Name: runs, Length: 215, dtype: int64
```