

# Welcome to the Ultimate Pandas Introduction Youtube Tutorial

Learn the basics on the pandas python library in this tutorial!

This notebook was created as part of the youtube tutorial [which can be found here.](#)

- Check out other videos [on my channel here.](#)

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

## Series

```
mydata = ['Boat', 'Car', 'Bike', 'Truck']
```

```
myseries1 = pd.Series(mydata)
```

```
print(myseries1)
```

```
0    Boat
1     Car
2    Bike
3   Truck
dtype: object
```

```
mydata = [1, 55, 99, 43]
myseries2 = pd.Series(mydata)
print(myseries2)
```

```
0     1
1    55
2    99
3    43
dtype: int64
```

## DataFrame

```
mydfdata = [('Boat', 1), ('Car', 55), ('Bike', 99), ('Truck', 43)]
```

```
mydf = pd.DataFrame(mydfdata, columns=['thing', 'count'])
```

```
mydf
```

	thing	count
0	Boat	1
1	Car	55
2	Bike	99
3	Truck	43

```
mydf.dtypes
```

```
thing    object
count    int64
dtype: object
```

## Reading in Data

```
df =
pd.read_csv('../input/mr-beast-youtube-video-statistics/MrBeast_youtub
e_stats.csv')
```

## Inspect The Data

```
# Head / Tail
```

```
df.head()
```

	id	title \
0	TQHEJj68Jew	I Got Hunted By A Real Bounty Hunter
1	00NgUctWoLQ	Extreme \$1,000,000 Hide And Seek
2	NaN	MrBeast
3	ayXxwJJId_c	I Bought The World's Largest Mystery Box! ...
4	cExLQ1o2pDw	First To Rob Bank Wins \$100,000

	description \
0	Sign up for Current w/ my Creator Code "BEAST"...
1	I didn't expect that to happen at the end I wa...
2	Accomplishments - Raised \$20000000 To Plant 20...
3	I cant believe I spent over \$500000 on mystery...
4	I didnt think he would actually rob the bank.....

	publishTime	kind_stats	duration_seconds
viewCount \			
0	2021-04-24 20:00:00+00:00	youtube#video	861
84717282.0			
1	2021-12-18 21:00:00+00:00	youtube#video	729
32090178.0			
2	2012-02-20 00:43:50+00:00	NaN	0
NaN			
3	2021-04-03 20:00:01+00:00	youtube#video	709
101745632.0			

4 2021-09-26 20:00:06+00:00 youtube#video 482  
50008942.0

```
likeCount commentCount
thumbnails.default.url \
0 2876493.0 128922.0
https://i.ytimg.com/vi/TQHEJj68Jew/default.jpg
1 2125183.0 73593.0
https://i.ytimg.com/vi/00NgUctWoLQ/default.jpg
2 NaN NaN
https://yt3.ggpht.com/yt3/AKed0LTctGKJ32CdDLiS...
3 3110824.0 162796.0
https://i.ytimg.com/vi/ayXxwJJId_c/default.jpg
4 2359606.0 120621.0
https://i.ytimg.com/vi/cExLQ1o2pDw/default.jpg
```

```
... thumbnails.high.width thumbnails.high.height
contentDetails.duration \
0 ... 480.0 360.0
PT14M21S
1 ... 480.0 360.0
PT12M9S
2 ... NaN NaN
NaN
3 ... 480.0 360.0
PT11M49S
4 ... 480.0 360.0
PT8M2S
```

```
contentDetails.dimension \
0 2d
1 2d
2 NaN
3 2d
4 2d
```

```
topicDetails.topicCategories
snippet.defaultLanguage \
0 ['https://en.wikipedia.org/wiki/Lifestyle_(soc...
NaN
1 ['https://en.wikipedia.org/wiki/Lifestyle_(soc...
en
2 NaN
NaN
3 ['https://en.wikipedia.org/wiki/Lifestyle_(soc...
NaN
4 ['https://en.wikipedia.org/wiki/Lifestyle_(soc...
NaN
```

```
localizations.en.title \
```

0		NaN
1	Extreme \$1,000,000 Hide And Seek	
2		NaN
3		NaN
4		NaN

	localizations.en.description	snippet.tags	\
0		NaN	NaN
1	I didn't expect that to happen at the end I wa...		NaN
2		NaN	NaN
3		NaN	NaN
4		NaN	NaN

	contentDetails.contentRating.ytRating
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

[5 rows x 26 columns]

df.tail()

	id	title	\
242	yeHjsYQ076A	Remember When COD Was Fun?	
243	X7doE4h6W64	Insane Gun Sync - 7 Hours To Make	
244	AIytwdufwW8	MY MESSAGE TO COD YOUTUBERS (Watch till end plz)	
245	560kH6ocYZU	L0114R - Biblical Creeper Post for Post @L0114R	
246	zi7tZ-2PhLk	How Much Money Do You Make As An Uber Driver??...	

	description	\
242	Yup, another very pointless video > Subscribe?...	
243	Look at this - <a href="http://gyazo.com/8852509d2350db...">http://gyazo.com/8852509d2350db...</a>	
244	The beginning of the video is a little off top...	
245	His channel - <a href="https://www.youtube.com/user/the...">https://www.youtube.com/user/the...</a>	
246	In this video I talk about being an uber drive...	

	publishTime	kind_stats	duration_seconds
viewCount	\		
242	2015-04-26 21:26:36+00:00	youtube#video	21616312.0
243	2015-06-21 12:22:34+00:00	youtube#video	13415740.0
244	2015-05-29 20:07:48+00:00	youtube#video	29218502.0
245	2015-05-15 16:48:54+00:00	youtube#video	10916627.0
246	2015-05-31 18:31:25+00:00	youtube#video	30418773.0

	likeCount	commentCount	
	thumbnails.default.url	\	
242	574.0	145.0	
	https://i.ytimg.com/vi/yeHjsYQ076A/default.jpg		
243	637.0	105.0	
	https://i.ytimg.com/vi/X7doE4h6W64/default.jpg		
244	508.0	117.0	
	https://i.ytimg.com/vi/AIytwdufwW8/default.jpg		
245	430.0	134.0	
	https://i.ytimg.com/vi/560kH6ocYZU/default.jpg		
246	370.0	59.0	
	https://i.ytimg.com/vi/zi7tZ-2PhLk/default.jpg		

	... thumbnails.high.width	thumbnails.high.height	\
242	...	480.0	360.0
243	...	480.0	360.0
244	...	480.0	360.0
245	...	480.0	360.0
246	...	480.0	360.0

	contentDetails.duration	contentDetails.dimension	\
242	PT3M36S	2d	
243	PT2M14S	2d	
244	PT4M52S	2d	
245	PT1M49S	2d	
246	PT5M4S	2d	

	topicDetails.topicCategories	\
242	['https://en.wikipedia.org/wiki/Action-adventu...	
243	['https://en.wikipedia.org/wiki/Action-adventu...	
244	['https://en.wikipedia.org/wiki/Action-adventu...	
245	['https://en.wikipedia.org/wiki/Action-adventu...	
246	['https://en.wikipedia.org/wiki/Action-adventu...	

	snippet.defaultLanguage	localizations.en.title	\
242	NaN	NaN	
243	NaN	NaN	
244	NaN	NaN	
245	NaN	NaN	
246	NaN	NaN	

	localizations.en.description	\
242	NaN	
243	NaN	
244	NaN	
245	NaN	
246	NaN	

	snippet.tags	\
--	--------------	---

```

242 ['How', 'much', 'money', 'does', 'make', 'blac...
243 ['How', 'much', 'money', 'does', 'make', 'blac...
244 ['How', 'much', 'money', 'does', 'make', 'blac...
245 ['How', 'much', 'money', 'does', 'make', 'blac...
246 ['How', 'much', 'money', 'does', 'make', 'blac...

```

```

contentDetails.contentRating.ytRating
242 NaN
243 NaN
244 NaN
245 NaN
246 NaN

```

```
[5 rows x 26 columns]
```

```
df.dtypes
```

```

id object
title object
description object
publishTime object
kind_stats object
duration_seconds int64
viewCount float64
likeCount float64
commentCount float64
thumbnails.default.url object
thumbnails.default.width float64
thumbnails.default.height float64
thumbnails.medium.url object
thumbnails.medium.width float64
thumbnails.medium.height float64
thumbnails.high.url object
thumbnails.high.width float64
thumbnails.high.height float64
contentDetails.duration object
contentDetails.dimension object
topicDetails.topicCategories object
snippet.defaultLanguage object
localizations.en.title object
localizations.en.description object
snippet.tags object
contentDetails.contentRating.ytRating object
dtype: object

```

```
df.describe()
```

	duration_seconds	viewCount	likeCount	commentCount	\
count	247.000000	2.460000e+02	2.440000e+02	245.000000	
mean	1886.234818	5.084638e+07	1.610498e+06	75562.783673	

std	8642.188728	4.062672e+07	1.879529e+06	85356.537127
min	0.000000	1.395400e+04	3.700000e+02	59.000000
25%	425.500000	1.406442e+07	4.348780e+05	21552.000000
50%	776.000000	4.904782e+07	1.358232e+06	60701.000000
75%	923.000000	8.125392e+07	2.294378e+06	99670.000000
max	85686.000000	1.787632e+08	1.904943e+07	726861.000000

	thumbnails.default.width	thumbnails.default.height \
count	246.0	246.0
mean	120.0	90.0
std	0.0	0.0
min	120.0	90.0
25%	120.0	90.0
50%	120.0	90.0
75%	120.0	90.0
max	120.0	90.0

	thumbnails.medium.width	thumbnails.medium.height \
count	246.0	246.0
mean	320.0	180.0
std	0.0	0.0
min	320.0	180.0
25%	320.0	180.0
50%	320.0	180.0
75%	320.0	180.0
max	320.0	180.0

	thumbnails.high.width	thumbnails.high.height
count	246.0	246.0
mean	480.0	360.0
std	0.0	0.0
min	480.0	360.0
25%	480.0	360.0
50%	480.0	360.0
75%	480.0	360.0
max	480.0	360.0

## Columns and Rows

```
df['viewCount']
```

0	84717282.0
1	32090178.0
2	NaN
3	101745632.0
4	50008942.0
	...
242	16312.0

```
243      15740.0
244      18502.0
245      16627.0
246      18773.0
```

```
Name: viewCount, Length: 247, dtype: float64
```

```
df.loc[4]
```

```
id
cExLQ1o2pDw
title                                     First To
Rob Bank Wins $100,000
description                               I didnt think he would
actually rob the bank.....
publishTime                             2021-
09-26 20:00:06+00:00
kind_stats
youtube#video
duration_seconds
482
viewCount
50008942.0
likeCount
2359606.0
commentCount
120621.0
thumbnails.default.url
https://i.ytimg.com/vi/cExLQ1o2pDw/default.jpg
thumbnails.default.width
120.0
thumbnails.default.height
90.0
thumbnails.medium.url
https://i.ytimg.com/vi/cExLQ1o2pDw/mqdefault.jpg
thumbnails.medium.width
320.0
thumbnails.medium.height
180.0
thumbnails.high.url
https://i.ytimg.com/vi/cExLQ1o2pDw/hqdefault.jpg
thumbnails.high.width
480.0
thumbnails.high.height
360.0
contentDetails.duration
PT8M2S
contentDetails.dimension
2d
topicDetails.topicCategories
['https://en.wikipedia.org/wiki/Lifestyle_(soc...
```



```
snippet.defaultLanguage
NaN
localizations.en.title
NaN
localizations.en.description
NaN
snippet.tags
NaN
contentDetails.contentRating.ytRating
NaN
Name: 4, dtype: object

df = df.set_index('id')
```

## Subsetting Data

```
df.shape
(247, 25)

# Subsetting Columns
df = df[['title', 'description', 'publishTime',
         'duration_seconds', 'viewCount', 'likeCount',
         'commentCount']]

# Subsetting using loc
df_subset1 = df.loc[df['viewCount'] > 1_000_000]

df_subset2 = df.query('viewCount > 1000000')

df = df.loc[~df['likeCount'].isna()]
```

## Casting dtypes

```
df['viewCount'] = df['viewCount'].astype('int')
df['likeCount'] = df['likeCount'].astype('int')

df['publishTime'] = pd.to_datetime(df['publishTime'])

df['likeCount'] = pd.to_numeric(df['likeCount'].astype('str'))
```

## Creating new column

```
df['like_to_view_ratio'] = df['likeCount'] / df['viewCount']
```

## Adding new Row

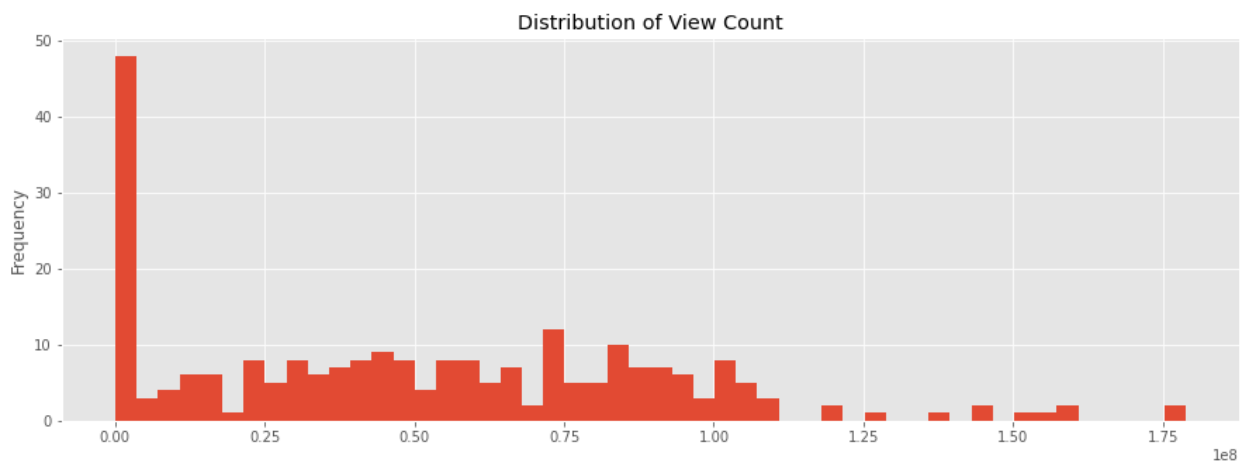
```
df_to_append = df.tail(1)
df_concat = pd.concat([df, df_to_append])
```

## Plot Examples

```
import matplotlib.pyplot as plt
plt.style.use('ggplot')

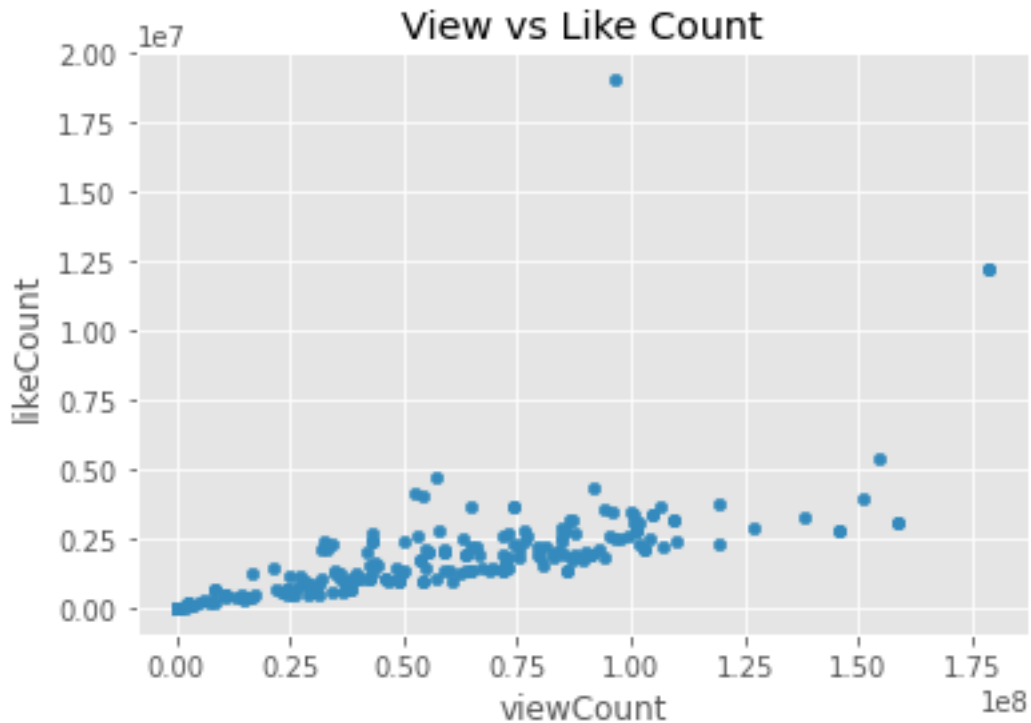
df['viewCount'].plot(kind='hist', bins=50,
                    title='Distribution of View Count',
                    figsize=(15, 5))

<AxesSubplot:title={'center': 'Distribution of View Count'},
ylabel='Frequency'>
```



```
df.plot(kind='scatter', x='viewCount', y='likeCount', title='View vs
Like Count')

<AxesSubplot:title={'center': 'View vs Like Count'},
xlabel='viewCount', ylabel='likeCount'>
```



```
df.query('likeCount > 10000000')
```

title \	
id	
0e3GPea1Tyg	\$456,000 Squid Game In Real Life!
0e3GPea1Tyg	\$456,000 Squid Game In Real Life!
dBx0YE2j55U	Make This Video The Most Liked Video On Youtube

description \	
id	
0e3GPea1Tyg	MAKE SURE YOU WATCH UNTIL GLASS BRIDGE IT'S IN...
0e3GPea1Tyg	MAKE SURE YOU WATCH UNTIL GLASS BRIDGE IT'S IN...
dBx0YE2j55U	WHAT IS UP LOGANG, MAKE SURE YOU LIKE THE VIDE...

likeCount \	publishTime	duration_seconds	viewCount
id			
0e3GPea1Tyg	2021-11-24 21:00:01+00:00	1542	178763228
12250732			
0e3GPea1Tyg	2021-11-24 21:00:01+00:00	1542	178763228
12250732			
dBx0YE2j55U	2019-01-16 22:05:01+00:00	791	96523287
19049431			

commentCount	like_to_view_ratio
id	

0e3GPea1Tyg	571946.0	0.068530
0e3GPea1Tyg	571946.0	0.068530
dBx0YE2j55U	277088.0	0.197356

## Save our output

```
df.to_csv('processed_data.csv')
```