

YoutubeIN-Input

April 11, 2020

```
[1]: import pandas as pd
      from datetime import date
      import calendar
      import json
      import matplotlib.pyplot as plt
```

```
[2]: raw_file = '/Users/amishra/DEV/DataEngineering.Labs.AirflowProject/
      ↪DataEngg-Airflow/raw/INvideos.csv'
```

```
[3]: print(raw_file)
      df = pd.read_csv(raw_file)
      df['month'] = df.apply(lambda row: calendar.month_name[int(row.trending_date.
      ↪split(".")[2])], axis=1)
      df
```

/Users/amishra/DEV/DataEngineering.Labs.AirflowProject/DataEngg-Airflow/raw/INvideos.csv

```
[3]:      video_id trending_date \
0      kzwfHumJyYc      17.14.11
1      zUZ1z7FwLc8      17.14.11
2      10L1hZ9qa58      17.14.11
3      N1vE8iiEg64      17.14.11
4      kJzGHOPVQHQ      17.14.11
...      ...      ...
37347  iNHecA3PJCo      18.14.06
37348  dpPmPbhcs1M      18.14.06
37349  mV6aztP58f8      18.14.06
37350  qxqDNP1bDEw      18.14.06
37351  wERgpPK44w0      18.14.06
```

	title	channel_title \
0	Sharry Mann: Cute Munda (Song Teaser) Parmi...	Lokdhun Punjabi
1	, , ... HJ NEWS	
2	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	TFPC
3	Eruma Saani Tamil vs English	Eruma Saani
4	why Samantha became EMOTIONAL @ Samantha naga ...	Filmylooks
...

37347	-	...	RDC Rajasthani	
37348		Seetha Flowers Ep# 364		Flowers TV
37349	Bhramanam I Episode 87 - 12 June 2018 I Mazhav...		Mazhavil Manorama	
37350	Nua Bohu Full Ep 285 13th June 2018 Odia...			Tarang TV
37351	Ee Nagarani Emaindi Trailer Tharun Bhascke...		Suresh Productions	

	category_id	publish_time	\
0	1	2017-11-12T12:20:39.000Z	
1	25	2017-11-13T05:43:56.000Z	
2	24	2017-11-12T15:48:08.000Z	
3	23	2017-11-12T07:08:48.000Z	
4	24	2017-11-13T01:14:16.000Z	
...	
37347	23	2018-06-13T08:01:11.000Z	
37348	24	2018-06-13T11:30:04.000Z	
37349	24	2018-06-13T05:00:02.000Z	
37350	24	2018-06-13T15:07:49.000Z	
37351	24	2018-06-10T04:29:54.000Z	

	tags	views	likes	\
0	sharry mann "sharry mann new song" "sharry man...	1096327	33966	
1	" ... 590101 735			
2	Stylish Star Allu Arjun @ ChaySam Wedding Rece...	473988	2011	
3	Eruma Saani "Tamil Comedy Videos" "Films" "Mov...	1242680	70353	
4	Filmylooks "latest news" "telugu movies" "telu...	464015	492	
...	
37347	twinkle vaishnav comedy "twinkle vaishnav" "tw...	214378	3291	
37348	flowers serials "actress" "malayalam serials" "...	406828	1726	
37349	mazhavil manorama "bhramanam full episode" "gt...	386319	1216	
37350	tarang "tarang tv" "tarang tv online" "tarang ...	130263	698	
37351	Ee Nagarani Emaindi "Ee Nagarani Emaindi T...	1278249	22466	

	dislikes	comment_count	\
0	798	882	
1	904	0	
2	243	149	
3	1624	2684	
4	293	66	
...	
37347	404	196	
37348	478	1428	
37349	453	697	
37350	115	65	
37351	1609	1205	

	thumbnail_link	comments_disabled	\
0	https://i.ytimg.com/vi/kzwfHumJyYc/default.jpg	False	

1	https://i.ytimg.com/vi/zUZ1z7FwLc8/default.jpg	True
2	https://i.ytimg.com/vi/10L1hZ9qa58/default.jpg	False
3	https://i.ytimg.com/vi/N1vE8iiEg64/default.jpg	False
4	https://i.ytimg.com/vi/kJzGH0PVQHQ/default.jpg	False
...
37347	https://i.ytimg.com/vi/iNHecA3PJCo/default.jpg	False
37348	https://i.ytimg.com/vi/dpPmPbhcs1M/default.jpg	False
37349	https://i.ytimg.com/vi/mV6aztP58f8/default.jpg	False
37350	https://i.ytimg.com/vi/qxqDNP1bDEw/default.jpg	False
37351	https://i.ytimg.com/vi/wERgpPK44w0/default.jpg	False

	ratings_disabled	video_error_or_removed	\
0	False	False	
1	False	False	
2	False	False	
3	False	False	
4	False	False	
...	
37347	False	False	
37348	False	False	
37349	False	False	
37350	False	False	
37351	False	False	

	description	month
0	Presenting Sharry Mann latest Punjabi Song Cu...	November
1	, , ... November	
2	Watch Stylish Star Allu Arjun @ ChaySam Weddin...	November
3	This video showcases the difference between pe...	November
4	why Samantha became EMOTIONAL @ Samantha naga ...	November
...
37347	PRG Music & RDC Rajasthani presents ...	June
37348	Flowers - A R Rahman Show,Book your Tickets He...	June
37349	Subscribe to Mazhavil Manorama now for your da...	June
37350	Nuabohu : Story of a rustic village girl who w...	June
37351	Check out Ee Nagaranki Emaindi Trailer #EeNag...	June

[37352 rows x 17 columns]

```
[4]: #df1 = pd.read_json('/Users/amishra/DEV/DataEngineering.Labs.AirflowProject/
↳DataEngg-Airflow/youtube-new/US_category_id.json')
#df1.items

data = None
with open('/Users/amishra/DEV/DataEngineering.Labs.AirflowProject/
↳DataEngg-Airflow/raw/US_category_id.json') as f:
    data = json.load(f)
```

```

category_list = list()
for c in data['items']:
    category_list.append([c['id'],c['snippet']['title']])

df11 = pd.DataFrame(category_list, columns = ['category_id', 'category_title'])
df11

```

```

[4]:
category_id      category_title
0           1      Film & Animation
1           2      Autos & Vehicles
2          10           Music
3          15      Pets & Animals
4          17           Sports
5          18      Short Movies
6          19      Travel & Events
7          20           Gaming
8          21      Videoblogging
9          22      People & Blogs
10         23           Comedy
11         24      Entertainment
12         25      News & Politics
13         26      Howto & Style
14         27      Education
15         28  Science & Technology
16         29  Nonprofits & Activism
17         30           Movies
18         31      Anime/Animation
19         32      Action/Adventure
20         33           Classics
21         34           Comedy
22         35      Documentary
23         36           Drama
24         37           Family
25         38           Foreign
26         39           Horror
27         40      Sci-Fi/Fantasy
28         41           Thriller
29         42           Shorts
30         43           Shows
31         44      Trailers

```

```

[5]: df.groupby('month').count()

```

```

[5]:
month      video_id  trending_date  title  channel_title  category_id \
April      4479      4479      4479      4479      4479

```

December	5803	5803	5803	5803	5803
February	4818	4818	4818	4818	4818
January	5464	5464	5464	5464	5464
June	2507	2507	2507	2507	2507
March	5369	5369	5369	5369	5369
May	5656	5656	5656	5656	5656
November	3256	3256	3256	3256	3256

	publish_time	tags	views	likes	dislikes	comment_count	\
month							
April	4479	4479	4479	4479	4479	4479	
December	5803	5803	5803	5803	5803	5803	
February	4818	4818	4818	4818	4818	4818	
January	5464	5464	5464	5464	5464	5464	
June	2507	2507	2507	2507	2507	2507	
March	5369	5369	5369	5369	5369	5369	
May	5656	5656	5656	5656	5656	5656	
November	3256	3256	3256	3256	3256	3256	

	thumbnail_link	comments_disabled	ratings_disabled	\
month				
April	4479		4479	4479
December	5803		5803	5803
February	4818		4818	4818
January	5464		5464	5464
June	2507		2507	2507
March	5369		5369	5369
May	5656		5656	5656
November	3256		3256	3256

	video_error_or_removed	description
month		
April	4479	4430
December	5803	5702
February	4818	4757
January	5464	5357
June	2507	2475
March	5369	5290
May	5656	5597
November	3256	3183

```
[6]: df1 = pd.DataFrame(columns = ['month', 'views'])
df1
```

```
[6]: Empty DataFrame
Columns: [month, views]
Index: []
```

```
[7]: # Aggregate the data according to trending time in calender year and category
```

```
[8]: df.groupby('category_id').count()
```

```
[8]:
```

	video_id	trending_date	title	channel_title	publish_time	\
category_id						
1	1658	1658	1658	1658	1658	
2	72	72	72	72	72	
10	3858	3858	3858	3858	3858	
15	3	3	3	3	3	
17	731	731	731	731	731	
19	8	8	8	8	8	
20	66	66	66	66	66	
22	2624	2624	2624	2624	2624	
23	3429	3429	3429	3429	3429	
24	16712	16712	16712	16712	16712	
25	5241	5241	5241	5241	5241	
26	845	845	845	845	845	
27	1227	1227	1227	1227	1227	
28	552	552	552	552	552	
29	105	105	105	105	105	
30	16	16	16	16	16	
43	205	205	205	205	205	

	tags	views	likes	dislikes	comment_count	thumbnail_link	\
category_id							
1	1658	1658	1658	1658	1658	1658	
2	72	72	72	72	72	72	
10	3858	3858	3858	3858	3858	3858	
15	3	3	3	3	3	3	
17	731	731	731	731	731	731	
19	8	8	8	8	8	8	
20	66	66	66	66	66	66	
22	2624	2624	2624	2624	2624	2624	
23	3429	3429	3429	3429	3429	3429	
24	16712	16712	16712	16712	16712	16712	
25	5241	5241	5241	5241	5241	5241	
26	845	845	845	845	845	845	
27	1227	1227	1227	1227	1227	1227	
28	552	552	552	552	552	552	
29	105	105	105	105	105	105	
30	16	16	16	16	16	16	
43	205	205	205	205	205	205	

	comments_disabled	ratings_disabled	video_error_or_removed	\
category_id				
1	1658	1658	1658	

2	72	72	72
10	3858	3858	3858
15	3	3	3
17	731	731	731
19	8	8	8
20	66	66	66
22	2624	2624	2624
23	3429	3429	3429
24	16712	16712	16712
25	5241	5241	5241
26	845	845	845
27	1227	1227	1227
28	552	552	552
29	105	105	105
30	16	16	16
43	205	205	205

	description	month
category_id		
1	1649	1658
2	69	72
10	3857	3858
15	3	3
17	712	731
19	8	8
20	66	66
22	2494	2624
23	3407	3429
24	16462	16712
25	5173	5241
26	837	845
27	1180	1227
28	548	552
29	105	105
30	16	16
43	205	205

```
[9]: #join_df = df.merge(df11, left_on='category_id', right_on='category_id')
#join_df = df.join(df11, lsuffix='category_id', rsuffix='category_id')
#df.head()
df12 = df11.astype({'category_id': int})
df12.info()
merged_inner = pd.merge(df, df12, on='category_id')
merged_inner
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
```

Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype
0	category_id	32 non-null	int64
1	category_title	32 non-null	object

dtypes: int64(1), object(1)
memory usage: 640.0+ bytes

[9]:

	video_id	trending_date	\
0	kzwfHumJyYc	17.14.11	
1	eP05M5DE01I	17.14.11	
2	MF9ovM9y0PE	17.14.11	
3	xjFKfAudCOo	17.14.11	
4	tzlF8oZd2BU	17.14.11	
...	
37347	17E0kTvARsA	18.04.06	
37348	17E0kTvARsA	18.05.06	
37349	17E0kTvARsA	18.05.06	
37350	17E0kTvARsA	18.06.06	
37351	17E0kTvARsA	18.07.06	

title \

0	Sharry Mann: Cute Munda (Song Teaser) Parmi...
1	Tiger Zinda Hai Official Trailer Salman Kh...
2	Punyalan Private Limited Official Teaser 2 J...
3	Aruvi - Official Teaser Arun Prabu Bindhu ...
4	engagement , ...
...	...
37347	Golak Bugni Bank Te Batua Full Movie (HD) Ha...
37348	Golak Bugni Bank Te Batua Full Movie (HD) Ha...
37349	Golak Bugni Bank Te Batua Full Movie (HD) Ha...
37350	Golak Bugni Bank Te Batua Full Movie (HD) Ha...
37351	Golak Bugni Bank Te Batua Full Movie (HD) Ha...

	channel_title	category_id	publish_time	\
0	Lokdhun Punjabi	1	2017-11-12T12:20:39.000Z	
1	YRF	1	2017-11-07T06:01:50.000Z	
2	DREAMS N BEYOND	1	2017-11-12T04:30:00.000Z	
3	Dream Warrior Pictures	1	2017-11-09T11:31:14.000Z	
4	Murari Ki Kocktail	1	2017-11-11T11:35:45.000Z	
...	
37347	Rhythm Boyz	30	2018-05-31T13:30:04.000Z	
37348	Rhythm Boyz	30	2018-05-31T13:30:04.000Z	
37349	Rhythm Boyz	30	2018-05-31T13:30:04.000Z	
37350	Rhythm Boyz	30	2018-05-31T13:30:04.000Z	
37351	Rhythm Boyz	30	2018-05-31T13:30:04.000Z	

		tags	views	likes \
0	sharry mann "sharry mann new song" "sharry man...		1096327	33966
1	Tiger Zinda Hai Trailer "Tiger Zinda Hai Offic...		35885754	829362
2	dreams n beyond "dreamsnbeyond" "ranjith shank...		249423	7861
3	Aruvi "Aruvi Tamil Movie" "Aruvi Movie" "Aruvi...		696515	12397
4	rajasthani comedy marwari comedy hariyani come...		269593	3931
...				
37347	full movie "full punjabi movie" "golak" "bugni...		4638549	60148
37348	full movie "full punjabi movie" "golak" "bugni...		5661965	67232
37349	full movie "full punjabi movie" "golak" "bugni...		5661965	67232
37350	full movie "full punjabi movie" "golak" "bugni...		6909059	75326
37351	full movie "full punjabi movie" "golak" "bugni...		7398655	77789

	dislikes	comment_count \
0	798	882
1	61195	101117
2	68	364
3	449	385
4	390	640
...		
37347	3160	3441
37348	3784	3851
37349	3784	3851
37350	4534	4279
37351	4785	4489

	thumbnail_link	comments_disabled \
0	https://i.ytimg.com/vi/kzwfHumJyYc/default.jpg	False
1	https://i.ytimg.com/vi/eP05M5DE01I/default.jpg	False
2	https://i.ytimg.com/vi/MF9ovM9yOPE/default.jpg	False
3	https://i.ytimg.com/vi/xjFKfAudCOo/default.jpg	False
4	https://i.ytimg.com/vi/tzlF8oZd2BU/default.jpg	False
...		
37347	https://i.ytimg.com/vi/l7E0kTvARsA/default.jpg	False
37348	https://i.ytimg.com/vi/l7E0kTvARsA/default.jpg	False
37349	https://i.ytimg.com/vi/l7E0kTvARsA/default.jpg	False
37350	https://i.ytimg.com/vi/l7E0kTvARsA/default.jpg	False
37351	https://i.ytimg.com/vi/l7E0kTvARsA/default.jpg	False

	ratings_disabled	video_error_or_removed \
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...		
37347	False	False

37348	False	False
37349	False	False
37350	False	False
37351	False	False

	description	month \
0	Presenting Sharry Mann latest Punjabi Song Cu...	November
1	Stay updated! Subscribe Now: https://goo.gl/...	November
2	Here is the #PunyalanPrivateLimited Official...	November
3	Proudly Presenting the Official Teaser of ARUV...	November
4	Kaka ki ek aur karastaani\nYe natural health a...	November
...
37347	Rhythm Boyz Entertainment & Hayre Omjee Studio...	June
37348	Rhythm Boyz Entertainment & Hayre Omjee Studio...	June
37349	Rhythm Boyz Entertainment & Hayre Omjee Studio...	June
37350	Rhythm Boyz Entertainment & Hayre Omjee Studio...	June
37351	Rhythm Boyz Entertainment & Hayre Omjee Studio...	June

	category_title
0	Film & Animation
1	Film & Animation
2	Film & Animation
3	Film & Animation
4	Film & Animation
...	...
37347	Movies
37348	Movies
37349	Movies
37350	Movies
37351	Movies

[37352 rows x 18 columns]

```
[10]: df13 = merged_inner.groupby(['month', 'category_title']).count()
df13
```

```
[10]:
```

	category_title	video_id	trending_date	title	channel_title \
month	category_title				
April	Autos & Vehicles	11	11	11	11
	Comedy	502	502	502	502
	Education	87	87	87	87
	Entertainment	2145	2145	2145	2145
	Film & Animation	117	117	117	117
...	
November	Pets & Animals	2	2	2	2
	Science & Technology	60	60	60	60
	Shows	23	23	23	23

	Sports	55	55	55	55
	Travel & Events	3	3	3	3

	category_id	publish_time	tags	views	likes	\
month	category_title					
April	Autos & Vehicles	11	11	11	11	
	Comedy	502	502	502	502	
	Education	87	87	87	87	
	Entertainment	2145	2145	2145	2145	
	Film & Animation	117	117	117	117	
...	
November	Pets & Animals	2	2	2	2	
	Science & Technology	60	60	60	60	
	Shows	23	23	23	23	
	Sports	55	55	55	55	
	Travel & Events	3	3	3	3	

	dislikes	comment_count	thumbnail_link	\
month				
April				
	11	11	11	
	502	502	502	
	87	87	87	
	2145	2145	2145	
	117	117	117	
...	
November				
	2	2	2	
	60	60	60	
	23	23	23	
	55	55	55	
	3	3	3	

	comments_disabled	ratings_disabled	\
month			
April			
	11	11	
	502	502	
	87	87	
	2145	2145	
	117	117	
...	
November			
	2	2	
	60	60	
	23	23	
	55	55	
	3	3	

	video_error_or_removed	description
month		
category_title		

April	Autos & Vehicles	11	11
	Comedy	502	495
	Education	87	84
	Entertainment	2145	2116
	Film & Animation	117	117
...		...	
November	Pets & Animals	2	2
	Science & Technology	60	60
	Shows	23	23
	Sports	55	55
	Travel & Events	3	3

[119 rows x 16 columns]

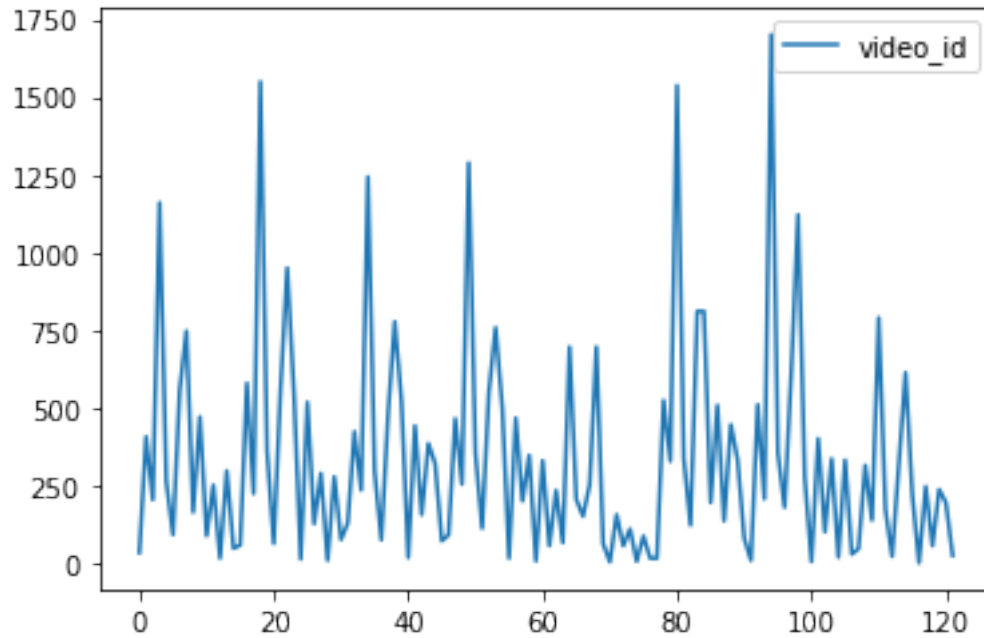
```
[11]: df13 = pd.read_csv('/Users/amishra/DEV/DataEngineering.Labs.AirflowProject/
↳DataEngg-Airflow/agg/US_aggregated.csv')
#df13
df14 = df13.iloc[:, 0:3]
df14.columns.str.strip()
df14
```

```
[11]:      month      category_title  video_id
0      April      Autos & Vehicles        34
1      April              Comedy       407
2      April              Education       204
3      April      Entertainment     1160
4      April      Film & Animation       264
..      ...                      ...      ...
117   November      People & Blogs       247
118   November      Pets & Animals        56
119   November  Science & Technology       237
120   November              Sports       195
121   November      Travel & Events        25
```

[122 rows x 3 columns]

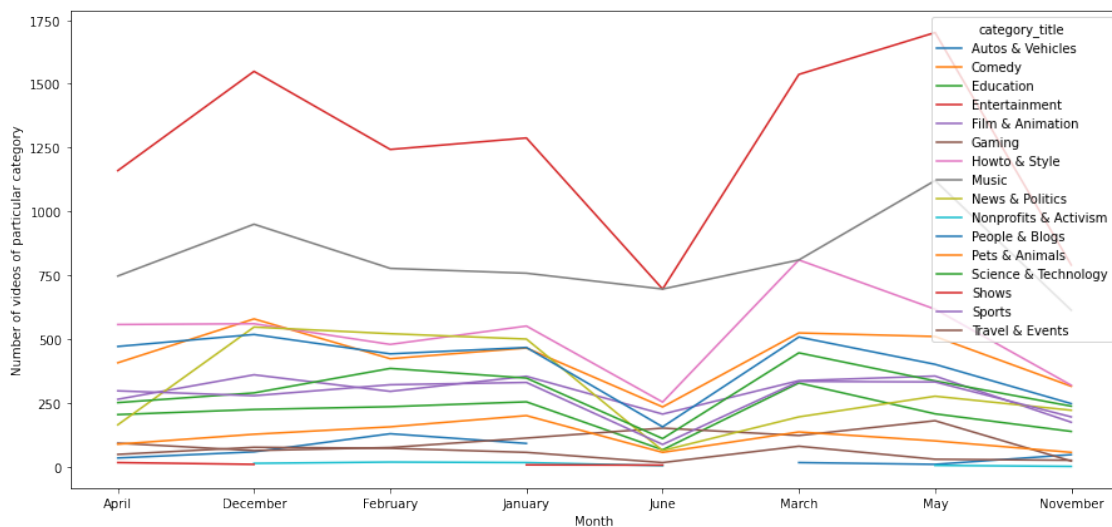
```
[12]: df14.plot.line()
```

```
[12]: <matplotlib.axes._subplots.AxesSubplot at 0x109ca2f50>
```



```
[13]: fig, ax = plt.subplots(figsize=(15,7))
df14.groupby(['month', 'category_title']).sum()['video_id'].unstack().plot(ax=ax)
ax.set_xlabel('Month')
ax.set_ylabel('Number of videos of particular category')
```

```
[13]: Text(0, 0.5, 'Number of videos of particular category')
```



```
[14]: report_name = "/Users/amishra/DEV/AirflowProject/youtube_analysis_" + str(date.  
      ↪today().strftime("%Y%m%d"))  
      report_for_2018 = fig.savefig(report_name, dpi=300, bbox_inches='tight')
```

```
[ ]:
```

```
[ ]:
```