Reshape pandas dataframe using melt

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491h

Enter your authorization code:
··········
Mounted at /content/drive
```

```
import pandas as pd
df = pd.read_csv("/content/drive/My Drive/pandas/pandas/11_melt/weather.csv")
df
```

	day	chicago	chennai	berlin
0	Monday	32	75	41
1	Tuesday	30	77	43
2	Wednesday	28	75	45
3	Thursday	22	82	38
4	Friday	30	83	30
5	Saturday	20	81	45
6	Sunday	25	77	47

```
df1= pd.melt(df, id_vars=["day"])
df1
```

	day	variable	value
0	Monday	chicago	32
1	Tuesday	chicago	30
2	Wednesday	chicago	28
3	Thursday	chicago	22
4	Friday	chicago	30
5	Saturday	chicago	20
6	Sunday	chicago	25
7	Monday	chennai	75
8	Tuesday	chennai	77
9	Wednesday	chennai	75
10	Thursday	chennai	82
11	Friday	chennai	83
12	Saturday	chennai	81
13	Sunday	chennai	77
14	Monday	berlin	41
15	Tuesday	berlin	43
16	Wednesday	berlin	45
17	Thursday	berlin	38
18	Friday	berlin	30
19	Saturday	berlin	45
20	Sunday	berlin	47

```
df1[df1['variable']=='chennai']
```

	day	variable	value
7	Monday	chennai	75
8	Tuesday	chennai	77
9	Wednesday	chennai	75
10	Thursday	chennai	82
11	Friday	chennai	83
12	Saturday	chennai	81
13	Sunday	chennai	77

variable and value can be changed

```
melted = pd.melt(df, id_vars=["day"], var_name='city', value_name='temperature')
melted
```

	day	city	temperature
0	Monday	chicago	32
1	Tuesday	chicago	30
2	Wednesday	chicago	28
3	Thursday	chicago	22
4	Friday	chicago	30
5	Saturday	chicago	20
6	Sunday	chicago	25
7	Monday	chennai	75
8	Tuesday	chennai	77
9	Wednesday	chennai	75

```
df_unmelted = melted.pivot(index='day', columns='city')

df_unmelted
```

	temperature		
city	berlin	chennai	chicago

```
df_unmelted.columns = df_unmelted.columns.droplevel()
```

```
df_unmelted
```

city	berlin	chennai	chicago
day			
Friday	30	83	30
Monday	41	75	32
Saturday	45	81	20
Sunday	47	77	25
Thursday	38	82	22
Tuesday	43	77	30
Wednesday	45	75	28

```
df_unmelted.index.name=None
```

```
df_unmelted
```

city	berlin	chennai	chicago
Friday	30	83	30
Monday	41	75	32

```
df_unmelted.columns.name=None
```

```
df_unmelted
```

	berlin	chennai	chicago
Friday	30	83	30
Monday	41	75	32
Saturday	45	81	20
Sunday	47	77	25
Thursday	38	82	22
Tuesday	43	77	30
Wednesday	45	75	28

```
import pandas as pd
df=pd.read_csv('http://files.zillowstatic.com/research/public_v2/invt_fs/Metro_invt_fs_
```

```
df.head()
```

	RegionID	SizeRank	RegionName	RegionType	StateName	2017-10-31	2017-11-30	2017-12-31	2018-01-31	2018-02-28	2018-03-3
0	102001	0	United States	Country	NaN	1660657.0	1600667.0	1526619.0	1491269.0	1429228.0	1431831.
1	394913	1	New York, NY	Msa	NY	82960.0	79438.0	75253.0	73101.0	71028.0	72208.
2	753899	2	Los Angeles-Long Beach-	Msa	CA	25402.0	23886.0	22022.0	21475.0	20597.0	21047.

```
df.shape
```

```
(121, 38)
```

Dallas-Fort

```
df_new=pd.melt(frame=df,id_vars=['RegionID',  'SizeRank', 'RegionName', 'RegionType', '
```

```
df_new.head()
```

	RegionID	SizeRank	RegionName	RegionType	StateName	year_val	count
0	102001	0	United States	Country	NaN	2017-10-31	1660657.0
1	394913	1	New York, NY	Msa	NY	2017-10-31	82960.0
2	753899	2	Los Angeles-Long Beach-Anaheim, CA	Msa	CA	2017-10-31	25402.0

```
df_new.shape
```

```
(3993, 7)
```

```
df_new.dtypes
```

```
RegionID        int64
SizeRank        int64
RegionName     object
RegionType     object
StateName      object
year_val       object
count         float64
dtype: object
```

```
df_new['year_val'].str.split('-').head()
```

```
0    [2017, 10, 31]
1    [2017, 10, 31]
2    [2017, 10, 31]
3    [2017, 10, 31]
4    [2017, 10, 31]
Name: year_val, dtype: object
```

```
df_new[['year','month','date']]=df_new['year_val'].str.split('-',expand=True)
```

```
df_new.head()
```

	RegionID	SizeRank	RegionName	RegionType	StateName	year_val	count	year	month	date
0	102001	0	United States	Country	NaN	2017-10-31	1660657.0	2017	10	31
1	394913	1	New York, NY	Msa	NY	2017-10-31	82960.0	2017	10	31
2	753899	2	Los Angeles-Long	Msa	CA	2017-10-	25402.0	2017	10	31

```
df_new.drop(columns=['year_val'],inplace=True,axis=1)
```

```
df_new.head()
```

	RegionID	SizeRank	RegionName	RegionType	StateName	count	year	month	date
0	102001	0	United States	Country	NaN	1660657.0	2017	10	31
1	394913	1	New York, NY	Msa	NY	82960.0	2017	10	31
2	753899	2	Los Angeles-Long Beach-Anaheim, CA	Msa	CA	25402.0	2017	10	31
3	394463	3	Chicago, IL	Msa	IL	47461.0	2017	10	31
4	394514	4	Dallas-Fort Worth, TX	Msa	TX	29470.0	2017	10	31