

trends

May 3, 2024

```
[2]: import pandas as pd
import numpy as np
```

```
[3]: tr = pd.read_csv("purchased_entries_clean.csv")
print(tr.shape)
print(tr.info())
```

```
(4614853, 6)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4614853 entries, 0 to 4614852
Data columns (total 6 columns):
#   Column          Dtype
---  -
0   event_time      object
1   event_type      object
2   category_id     float64
3   category_code   object
4   brand           object
5   price          float64
dtypes: float64(2), object(4)
memory usage: 211.3+ MB
None
```

```
[4]: tr
```

```
[4]:
```

		event_time	event_type	category_id	\
0	2019-10-01	00:02:14+00:00	purchase	2.053014e+18	
1	2019-10-01	00:04:37+00:00	purchase	2.053014e+18	
2	2019-10-01	00:07:07+00:00	purchase	2.053014e+18	
3	2019-10-01	00:09:26+00:00	purchase	2.053014e+18	
4	2019-10-01	00:09:54+00:00	purchase	2.053014e+18	
...		
4614848	2020-03-19	08:54:20+00:00	purchase	2.232732e+18	
4614849	2020-03-19	08:54:21+00:00	purchase	2.232732e+18	
4614850	2020-03-19	08:54:21+00:00	purchase	2.232732e+18	
4614851	2020-03-19	08:54:21+00:00	purchase	2.232732e+18	
4614852	2020-03-		NaN	NaN	

	category_code	brand	price
0	electronics.smartphone	samsung	130.76
1	electronics.smartphone	apple	642.69
2	furniture.bathroom.toilet	santeri	54.42
3	electronics.audio.headphone	apple	189.91
4	electronics.audio.headphone	apple	161.98
...
4614848	construction.tools.light	apple	962.02
4614849	construction.tools.light	samsung	169.46
4614850	appliances.personal.massager	lg	448.40
4614851	construction.tools.light	samsung	144.70
4614852	NaN	NaN	NaN

[4614853 rows x 6 columns]

```
[5]: tr[tr.duplicated(subset = 'event_time')]
```

```
[5]:
```

	event_time	event_type	category_id \
22	2019-10-01 02:23:05+00:00	purchase	2.053014e+18
34	2019-10-01 02:26:05+00:00	purchase	2.053014e+18
59	2019-10-01 02:31:46+00:00	purchase	2.053014e+18
67	2019-10-01 02:33:01+00:00	purchase	2.053014e+18
76	2019-10-01 02:34:23+00:00	purchase	2.053014e+18
...
4614840	2020-03-19 08:53:59+00:00	purchase	2.053014e+18
4614843	2020-03-19 08:54:04+00:00	purchase	2.232732e+18
4614846	2020-03-19 08:54:17+00:00	purchase	2.232732e+18
4614850	2020-03-19 08:54:21+00:00	purchase	2.232732e+18
4614851	2020-03-19 08:54:21+00:00	purchase	2.232732e+18

	category_code	brand	price
22	electronics.audio.headphone	jbl	20.57
34	electronics.smartphone	samsung	197.43
59	electronics.smartphone	samsung	286.86
67	electronics.tablet	wacom	78.51
76	appliances.kitchen.washer	lg	308.65
...
4614840	electronics.audio.headphone	asus	463.31
4614843	sport.bicycle	xiaomi	24.97
4614846	furniture.bedroom.bed	sv	282.89
4614850	appliances.personal.massager	lg	448.40
4614851	construction.tools.light	samsung	144.70

[1083759 rows x 6 columns]

```
[6]: tr[tr['event_time'] == '2019-10-01 02:23:05+00:00']
```

```
[6]:          event_time event_type  category_id \
21  2019-10-01 02:23:05+00:00  purchase  2.053014e+18
22  2019-10-01 02:23:05+00:00  purchase  2.053014e+18
```

```
          category_code  brand  price
21      electronics.smartphone  huawei  111.82
22  electronics.audio.headphone    jbl   20.57
```

```
[7]: new_tr = tr[tr.duplicated('event_time', keep=False)]
new_tr
```

```
[7]:          event_time event_type  category_id \
21      2019-10-01 02:23:05+00:00  purchase  2.053014e+18
22      2019-10-01 02:23:05+00:00  purchase  2.053014e+18
33      2019-10-01 02:26:05+00:00  purchase  2.053014e+18
34      2019-10-01 02:26:05+00:00  purchase  2.053014e+18
58      2019-10-01 02:31:46+00:00  purchase  2.053014e+18
...
4614845  2020-03-19 08:54:17+00:00  purchase  2.232732e+18
4614846  2020-03-19 08:54:17+00:00  purchase  2.232732e+18
4614849  2020-03-19 08:54:21+00:00  purchase  2.232732e+18
4614850  2020-03-19 08:54:21+00:00  purchase  2.232732e+18
4614851  2020-03-19 08:54:21+00:00  purchase  2.232732e+18
```

```
          category_code  brand  price
21      electronics.smartphone  huawei  111.82
22      electronics.audio.headphone    jbl   20.57
33      electronics.audio.headphone  ritmix    0.88
34      electronics.smartphone  samsung  197.43
58      electronics.smartphone    apple 1634.51
...
4614845  construction.tools.light    apple 1376.05
4614846  furniture.bedroom.bed        sv   282.89
4614849  construction.tools.light  samsung  169.46
4614850  appliances.personal.massager    lg   448.40
4614851  construction.tools.light  samsung  144.70
```

[1888851 rows x 6 columns]

```
[8]: trr = new_tr[['event_time', 'category_code']]
```

```
[9]: grouped = trr.groupby('event_time', as_index=False).agg(list)
grouped
```

```
[9]:          event_time \
0      2019-10-01 02:23:05+00:00
1      2019-10-01 02:26:05+00:00
```

```

2      2019-10-01 02:31:46+00:00
3      2019-10-01 02:33:01+00:00
4      2019-10-01 02:34:23+00:00
...
805087 2020-03-19 08:53:44+00:00
805088 2020-03-19 08:53:59+00:00
805089 2020-03-19 08:54:04+00:00
805090 2020-03-19 08:54:17+00:00
805091 2020-03-19 08:54:21+00:00

```

```

                                category_code
0      [electronics.smartphone, electronics.audio.he...
1      [electronics.audio.headphone, electronics.smar...
2      [electronics.smartphone, electronics.smartphone]
3      [electronics.video.tv, electronics.tablet]
4      [electronics.smartphone, appliances.kitchen.wa...
...
805087 [construction.tools.light, construction.tools...
805088 [construction.tools.light, electronics.audio.h...
805089      [construction.tools.light, sport.bicycle]
805090 [construction.tools.light, furniture.bedroom.bed]
805091 [construction.tools.light, appliances.personal...

```

[805092 rows x 2 columns]

```
[10]: l = grouped['category_code'].tolist()
```

```
[11]: from mlxtend.preprocessing import TransactionEncoder
```

```

te = TransactionEncoder()
te_ary = te.fit(l).transform(l)
df = pd.DataFrame(te_ary, columns=te.columns_)
df

```

```

[11]:      accessories.bag  accessories.umbrella  accessories.wallet  \
0                False                False                False
1                False                False                False
2                False                False                False
3                False                False                False
4                False                False                False
...
805087            False                False                False
805088            False                False                False
805089            False                False                False
805090            False                False                False
805091            False                False                False

```

	apparel.belt	apparel.costume	apparel.dress	apparel.glove	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
805087	False	False	False	False	
805088	False	False	False	False	
805089	False	False	False	False	
805090	False	False	False	False	
805091	False	False	False	False	

	apparel.jacket	apparel.jeans	apparel.jumper	...	kids.swing	\
0	False	False	False	...	False	
1	False	False	False	...	False	
2	False	False	False	...	False	
3	False	False	False	...	False	
4	False	False	False	...	False	
...	
805087	False	False	False	...	False	
805088	False	False	False	...	False	
805089	False	False	False	...	False	
805090	False	False	False	...	False	
805091	False	False	False	...	False	

	kids.toys	medicine.tools.tonometer	sport.bicycle	sport.diving	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
805087	False	False	False	False	
805088	False	False	False	False	
805089	False	False	True	False	
805090	False	False	False	False	
805091	False	False	False	False	

	sport.ski	sport.snowboard	sport.tennis	sport.trainer	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
...	
805087	False	False	False	False	

805088	False	False	False	False
805089	False	False	False	False
805090	False	False	False	False
805091	False	False	False	False

	stationery.cartridge
0	False
1	False
2	False
3	False
4	False
...	...
805087	False
805088	False
805089	False
805090	False
805091	False

[805092 rows x 136 columns]

```
[13]: from mlxtend.frequent_patterns import fpgrowth
      fpgrowth(df, min_support=0.1, use_colnames=True)
```

```
[13]:      support      itemsets
0  0.186892  (electronics.smartphone)
1  0.614512  (construction.tools.light)
```

No itemset is bought together frequently according to fpgrowth, except for electronics.smartphone as one group, and construction.tools.light as a separate group.

```
[ ]:
```