```
In [1]: import pandas as pd
```

```
In [3]: df = pd.read_csv('data/data.csv')
```

Items to fix

- remove unnamed column
- In `month`, make 'Last 30 Days' into May 2020
- In `month`, remove new lines and tabs
- Make `month` a date time

```
In [6]: df.head()
```

Out[6]:

	Unnamed: 0	game_id	game_title	month	peak_users	picture_url
0	0	920120	FLIP FLAPPERS: Pure Storage	Last 30 Days	0	https://steamcdn-a.akamaihd.net/steam/apps/920...
1	1	919890	Navyblue and the Spectrum Killers	Last 30 Days	0	https://steamcdn-a.akamaihd.net/steam/apps/919...
2	2	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tMarch 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
3	3	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tFebruary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
4	4	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tJanuary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...

Remove unnamed column

```
In [7]:  df.columns
```

```
Out[7]:  Index(['Unnamed: 0', 'game_id', 'game_title', 'month', 'peak_users',
                'picture_url'],
               dtype='object')
```

```
In [10]:  df.drop(['Unnamed: 0'], axis=1, inplace=True)
```

```
In [11]:  df.head()
```

Out[11]:

	game_id	game_title	month	peak_users	picture_url
0	920120	FLIP FLAPPERS: Pure Storage	Last 30 Days	0	https://steamcdn-a.akamaihd.net/steam/apps/920...
1	919890	Navyblue and the Spectrum Killers	Last 30 Days	0	https://steamcdn-a.akamaihd.net/steam/apps/919...
2	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tMarch 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
3	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tFebruary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
4	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tJanuary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...

In month, make 'Last 30 Days' into May 2020

```
In [12]:  df.replace('Last 30 Days', 'May 2020', inplace=True)
```

```
In [13]:  df.head()
```

Out[13]:

	game_id	game_title	month	peak_users	picture_url
0	920120	FLIP FLAPPERS: Pure Storage	May 2020	0	https://steamcdn-a.akamaihd.net/steam/apps/920...
1	919890	Navyblue and the Spectrum Killers	May 2020	0	https://steamcdn-a.akamaihd.net/steam/apps/919...
2	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tMarch 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
3	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tFebruary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
4	919890	Navyblue and the Spectrum Killers	\n\t\t\t\t\tJanuary 2019\n\t\t\t\t	1	https://steamcdn-a.akamaihd.net/steam/apps/919...

In `month`, remove new lines and tabs

```
In [23]: df['month'].replace('\n', '', inplace=True, regex=True)
```

```
In [24]: df['month'].replace('\t', '', inplace=True, regex=True)
```

```
In [25]: df.head()
```

Out[25]:

	game_id	game_title	month	peak_users	picture_url
0	920120	FLIP FLAPPERS: Pure Storage	May 2020	0	https://steamcdn-a.akamaihd.net/steam/apps/920...
1	919890	Navyblue and the Spectrum Killers	May 2020	0	https://steamcdn-a.akamaihd.net/steam/apps/919...
2	919890	Navyblue and the Spectrum Killers	March 2019	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
3	919890	Navyblue and the Spectrum Killers	February 2019	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
4	919890	Navyblue and the Spectrum Killers	January 2019	1	https://steamcdn-a.akamaihd.net/steam/apps/919...

Make `month` a date time

Format is May 2020 This would be '%B %Y'

https://docs.python.org/3/library/datetime.html#strftime-and-strptime-behavior (https://docs.python.org/3/library/datetime.html#strftime-and-strptime-behavior)

```
In [26]: df['month'] = pd.to_datetime(df['month'], format='%B %Y')
```

In [27]: `df.head()`

Out[27]:

	game_id	game_title	month	peak_users	picture_url
0	920120	FLIP FLAPPERS: Pure Storage	2020-05-01	0	https://steamcdn-a.akamaihd.net/steam/apps/920...
1	919890	Navyblue and the Spectrum Killers	2020-05-01	0	https://steamcdn-a.akamaihd.net/steam/apps/919...
2	919890	Navyblue and the Spectrum Killers	2019-03-01	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
3	919890	Navyblue and the Spectrum Killers	2019-02-01	1	https://steamcdn-a.akamaihd.net/steam/apps/919...
4	919890	Navyblue and the Spectrum Killers	2019-01-01	1	https://steamcdn-a.akamaihd.net/steam/apps/919...

Saving file

In [28]: `df.to_csv('data/data_clean.csv')`