In [2]:
```
import pandas as pd
```

In [4]:
```
df=pd.read_csv('zomato-ipo.csv')
```

In [5]:
```
df
```

Out[5]:

	id	conversation_id	created_at	date	time	timezone	
0	1415522642363224064	1415522329195515904	2021-07-15 04:04:31 UTC	2021-07-15	04:04:31	0	134322552781
1	1415522472628162564	1415522472628162564	2021-07-15 04:03:51 UTC	2021-07-15	04:03:51	0	131226540704
2	1415522463052537860	1415522463052537860	2021-07-15 04:03:48 UTC	2021-07-15	04:03:48	0	132115671699
3	1415522282039136256	1415522282039136256	2021-07-15 04:03:05 UTC	2021-07-15	04:03:05	0	9
4	1415522222912020480	1415522222912020480	2021-07-15 04:02:51 UTC	2021-07-15	04:02:51	0	169
...	...	...	...	...	...	...	
11489	1304028285714575361	1304028285714575361	2020-09-10 12:05:45 UTC	2020-09-10	12:05:45	0	79442440155
11490	1297125683034980353	1297125683034980353	2020-08-22 10:57:16 UTC	2020-08-22	10:57:16	0	100648730787
11491	1263833874192203776	1263833874192203776	2020-05-22 14:07:31 UTC	2020-05-22	14:07:31	0	5
11492	613296543093903360	613296543093903360	2015-06-23 10:44:25 UTC	2015-06-23	10:44:25	0	334

	id	conversation_id	created_at	date	time	timezone	
11493	315823312209080320	315823312209080320	2013-03-24 13:52:00 UTC	2013-03-24	13:52:00	0	117

11494 rows × 36 columns

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
id                    0
conversation_id       0
created_at            0
date                  0
time                  0
timezone              0
user_id               0
username              0
name                  1
place             11485
tweet                 0
language              0
mentions              0
urls                  0
photos                0
replies_count         0
retweets_count        0
likes_count           0
hashtags              0
cashtags              0
link                  0
retweet               0
quote_url         10604
video                 0
thumbnail          8846
near              11494
geo               11494
source            11494
user_rt_id        11494
user_rt           11494
retweet_id        11494
reply_to              0
retweet_date      11494
translate         11494
trans_src         11494
trans_dest        11494
dtype: int64
```

In [7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11494 entries, 0 to 11493
Data columns (total 36 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
```

```
 0   id               11494 non-null  int64
 1   conversation_id  11494 non-null  int64
 2   created_at       11494 non-null  object
 3   date             11494 non-null  object
 4   time             11494 non-null  object
 5   timezone         11494 non-null  int64
 6   user_id          11494 non-null  int64
 7   username         11494 non-null  object
 8   name             11493 non-null  object
 9   place            9 non-null      object
 10  tweet            11494 non-null  object
 11  language         11494 non-null  object
 12  mentions         11494 non-null  object
 13  urls             11494 non-null  object
 14  photos           11494 non-null  object
 15  replies_count    11494 non-null  int64
 16  retweets_count   11494 non-null  int64
 17  likes_count      11494 non-null  int64
 18  hashtags         11494 non-null  object
 19  cashtags         11494 non-null  object
 20  link             11494 non-null  object
 21  retweet          11494 non-null  bool
 22  quote_url        890 non-null    object
 23  video            11494 non-null  int64
 24  thumbnail        2648 non-null   object
 25  near             0 non-null      float64
 26  geo              0 non-null      float64
 27  source           0 non-null      float64
 28  user_rt_id       0 non-null      float64
 29  user_rt          0 non-null      float64
 30  retweet_id       0 non-null      float64
 31  reply_to         11494 non-null  object
 32  retweet_date     0 non-null      float64
 33  translate        0 non-null      float64
 34  trans_src        0 non-null      float64
 35  trans_dest       0 non-null      float64
dtypes: bool(1), float64(10), int64(8), object(17)
memory usage: 3.1+ MB
```

In [6]:

```
df.describe()
```

Out[6]:

	id	conversation_id	timezone	user_id	replies_count	retweets_count	likes_cou
count	1.149400e+04	1.149400e+04	11494.0	1.149400e+04	11494.000000	11494.000000	11494.000
mean	1.410685e+18	1.410286e+18	0.0	5.708234e+17	0.797547	1.030625	10.337
std	1.793884e+16	2.315213e+16	0.0	5.986485e+17	5.733031	6.668306	86.007
min	3.158233e+17	3.158233e+17	0.0	5.194300e+04	0.000000	0.000000	0.000
25%	1.413189e+18	1.413166e+18	0.0	3.352015e+08	0.000000	0.000000	0.000
50%	1.415002e+18	1.414992e+18	0.0	4.567719e+09	0.000000	0.000000	0.000
75%	1.415232e+18	1.415218e+18	0.0	1.229650e+18	0.000000	0.000000	2.000
max	1.415523e+18	1.415522e+18	0.0	1.415316e+18	329.000000	204.000000	4211.000

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

In [7]:
```
l=[]
for col in df.columns:
    l.append(col)
```

Out[7]:
```
Index(['id', 'conversation_id', 'created_at', 'date', 'time', 'timezone',
       'user_id', 'username', 'name', 'place', 'tweet', 'language', 'mentions',
       'urls', 'photos', 'replies_count', 'retweets_count', 'likes_count',
       'hashtags', 'cashtags', 'link', 'retweet', 'quote_url', 'video',
       'thumbnail', 'near', 'geo', 'source', 'user_rt_id', 'user_rt',
       'retweet_id', 'reply_to', 'retweet_date', 'translate', 'trans_src',
       'trans_dest'],
      dtype='object')
```

In [17]:
```
l
```

Out[17]:
```
['id',
 'conversation_id',
 'created_at',
 'date',
 'time',
 'timezone',
 'user_id',
 'username',
 'name',
 'place',
 'tweet',
 'language',
 'mentions',
 'urls',
 'photos',
 'replies_count',
 'retweets_count',
 'likes_count',
 'hashtags',
 'cashtags',
 'link',
 'retweet',
 'quote_url',
 'video',
 'thumbnail',
 'near',
 'geo',
 'source',
 'user_rt_id',
 'user_rt',
 'retweet_id',
 'reply_to',
 'retweet_date',
 'translate',
 'trans_src',
 'trans_dest']
```

In [8]:
```
df.head()
```

Out[8]:

	id	conversation_id	created_at	date	time	timezone	use
0	1415522642363224064	1415522329195515904	2021-07-15 04:04:31 UTC	2021-07-15	04:04:31	0	1343225527813898
1	1415522472628162564	1415522472628162564	2021-07-15 04:03:51 UTC	2021-07-15	04:03:51	0	1312265407047163
2	1415522463052537860	1415522463052537860	2021-07-15 04:03:48 UTC	2021-07-15	04:03:48	0	1321156716990353
3	1415522282039136256	1415522282039136256	2021-07-15 04:03:05 UTC	2021-07-15	04:03:05	0	99065
4	1415522222912020480	1415522222912020480	2021-07-15 04:02:51 UTC	2021-07-15	04:02:51	0	1693990

5 rows × 36 columns

In [41]:
```
df1= df[['created_at',  'name','tweet','replies_count', 'retweets_count','likes_count',
```

In [42]:
```
df1
```

Out[42]:

	created_at	name	tweet	replies_count	retweets_count	likes_count	
0	2021-07-15 04:04:31 UTC	Aman Singh	@ipo_mantra I haven't applied for Zomato IPO -...	1	0	2	
1	2021-07-15 04:03:51 UTC	TRIPURATEER	#PaytmIPO #PaytmIPOnews #zomatoipo #Zomatoshar...	0	0	0	'pay 'zomat
2	2021-07-15 04:03:48 UTC	Ankit Kakkad	#wipro CMP 570😍 Trgt achieved 🚀😊 (Safe players...	0	0	0	['wi 'nifty50'
3	2021-07-15 04:03:05 UTC	Saurabh Chandra	So much love poured by the startup community o...	0	0	0	

	created_at	name	tweet	replies_count	retweets_count	likes_count	
4	2021-07-15 04:02:51 UTC	Dr. Silvia Elaluf-Calderwood	Zomato IPO: India food delivery 'unicorn' open...	0	0	0	['i
...	...	...	...	...	...	...	
11489	2020-09-10 12:05:45 UTC	Nikita Vashisht	Is this why Info Edge's shares zoomed suddenly...	0	0	0	['tradin 'ipo
11490	2020-08-22 10:57:16 UTC	Digital Agents	Business news #84 | Ambience Mall illegal-Pata...	0	0	0	
11491	2020-05-22 14:07:31 UTC	Kushal Bhagia	Wow looks like Mumbai is allowing online order...	1	0	32	
11492	2015-06-23 10:44:25 UTC	zooey m	#TheIndianCapitalist: e-IPO for #Start-ups ht...	0	0	0	['theindi 'start'- '
11493	2013-03-24 13:52:00 UTC	IB Singapore	Indian restaurant guide Zomato Media planning ...	0	0	0	['zo

11494 rows × 8 columns

In [43]:
```
df1.isnull().sum()
```

Out[43]:
```
created_at        0
name              1
tweet             0
replies_count     0
retweets_count    0
likes_count       0
hashtags          0
retweet           0
dtype: int64
```

In [44]:
```
df['tweet'] = df['tweet'].str.replace(',','-')
df['hashtags'] = df['hashtags'].str.replace(',','-')
```

In [45]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11494 entries, 0 to 11493
Data columns (total 8 columns):
```

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   created_at      11494 non-null  object
 1   name            11493 non-null  object
 2   tweet           11494 non-null  object
 3   replies_count   11494 non-null  int64
 4   retweets_count  11494 non-null  int64
 5   likes_count     11494 non-null  int64
 6   hashtags        11494 non-null  object
 7   retweet         11494 non-null  bool
dtypes: bool(1), int64(3), object(4)
memory usage: 639.9+ KB
```

In [46]:
```
df1=df1.dropna()
```

In [47]:
```
df1.isnull().sum()
```

Out[47]:
```
created_at        0
name              0
tweet             0
replies_count     0
retweets_count    0
likes_count       0
hashtags          0
retweet           0
dtype: int64
```

In [48]:
```
df2=df1.iloc[0:9997]
```

In [49]:
```
df2.shape
```

Out[49]:
```
(9997, 8)
```

In [50]:
```
df2.describe()
```

Out[50]:

	replies_count	retweets_count	likes_count
count	9997.000000	9997.000000	9997.000000
mean	0.828048	1.031810	10.596879
std	6.083892	6.838925	90.708587
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	2.000000
max	329.000000	204.000000	4211.000000

In [51]:
```
df2.to_csv("zomatoIpo.xlsx")
```

In [ ]: