

In [11]:

```
import xampy as xp
```

# VS

In [12]:

```
import matplotlib.pyplot as plt
```

## Reading data and quick look

- Most text data

In [13]:

```
data = xp.makeData('topcsv.csv')
xp.showInfo(data)
```

Descriptive stats:

	AuthorID
count	5.867660e+05
mean	5.519042e+17
std	2.197354e+17
min	2.388172e+16
25%	3.673389e+17
50%	5.660515e+17
75%	7.800801e+17
max	8.176622e+17

Dataframe information:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 586766 entries, 0 to 586765
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   AuthorID        586766 non-null  int64
1   Author          586766 non-null  object
2   Date            586766 non-null  object
3   Content         572015 non-null  object
4   Attachments     19464 non-null   object
5   Reactions       84827 non-null   object
dtypes: int64(1), object(5)
memory usage: 26.9+ MB
None
```

Dataframe Head:

	AuthorID	Author	Date	Content	Attachments	Reactions
0	741087608263671809	bp#7889	01-Oct-20 12:22 AM	Wow, 999 moments per player in WCF. Aren't the...	NaN	NaN
1	212928885904506880	Nole77#7519	01-Oct-20 12:27 AM	yea very interesting	NaN	NaN
2	212928885904506880	Nole77#7519	01-Oct-20 12:27 AM	wonder if that will increase the value of the ...	NaN	NaN
3	404374804792344589	alxo#6661	01-Oct-20 12:30 AM	if they sell out, then i'd say yes	NaN	NaN
4	212928885904506880	Nole77#7519	01-Oct-20 12:34 AM	Oh it will sell out, but I don't think it'll h...	NaN	NaN

# Looking into the emoji data

- convenient jupyter shows the emoji
- emoji follows by a number in ()
- some posts have a slew of emojis on them
  - do we want to explode these or treat them uniquely
- not all the emojis come through
  - should we encode them or leave them as is
  - are the ones not pulled through exclusive to the TOPSHOT discord
  - are these emojis belonging to more "Elite" members
- 23719 unique combinations of reactions

In [14]:

```
data = xp.dropnaDim(data, 'row')
data
```

Out[14]:

	AuthorID	Author	Date	Content	Attachments	Reactions
63	734842980510335179	kyloren#1260	01-Oct-20 02:28 AM	> @kyloren as soon as maintenance is finished...	https://cdn.discordapp.com/attachments/6061118...	♥ (3)
297	476033925093916672	Usman#6305	01-Oct-20 01:09 PM	I see you Spencer!	https://cdn.discordapp.com/attachments/6061118...	fire (7),nets (3)
840	404374804792344589	alxo#6661	01-Oct-20 03:58 PM	this is awesome. congratulations team and comm...	https://cdn.discordapp.com/attachments/6061118...	♥ (2)
1102	114939077647794178	rudem00se#9384	01-Oct-20 05:13 PM	Those are my lowest id, which I'm pretty excit...	https://cdn.discordapp.com/attachments/6061118...	😊 (2)
1230	212928885904506880	Nole77#7519	01-Oct-20 06:35 PM	I live for this	https://cdn.discordapp.com/attachments/6061118...	□ (7)
...	...	...	...	...	...	...
578481	715303429265817680	GPJ_Junky#2758	05-Mar-21 03:56 PM	HODL those packs people! It is possible ☺	https://cdn.discordapp.com/attachments/6061118...	□ (26),□ (33),□ (7)
581965	422479740801122304	RawDog#0536	05-Mar-21 08:10 PM	Mike Conley 15k LE to the moonnn	https://cdn.discordapp.com/attachments/6061118...	□ (2)
583065	769611775012110386	Ahmazing#5115	05-Mar-21 09:20 PM	Moment is AWESOME	https://cdn.discordapp.com/attachments/6061118...	□ (10),CurryShhhh (1)
585725	345558875665334272	Zemekis#5743	06-Mar-21 01:00 AM	Haliburton confidence, love it, buys top 1 & 3...	https://cdn.discordapp.com/attachments/6061118...	1_ (12),□ (12),□ (1)
586014	396783138057224192	MasterCollector#6853	06-Mar-21 01:55 AM	just got this completed	https://cdn.discordapp.com/attachments/6061118...	□ (4)

4781 rows × 6 columns

## Counting the missing values by column

- attachments are photos and i chose not to download those along
  - going to drop this col
- creating a subset dataframe where reactions are not null? are those the more important messages

In [15]:

```
xp.countMissing(data)
```

```

AuthorID      0
Author        0
Date          0
Content       0
Attachments   0
Reactions     0
dtype: int64

```

## Sentiment Analysis

```
In [16]:
data = xp.socialSentiment(data, 'Content')
```

```
In [17]:
data.head(5)
```

```
Out[17]:
```

	AuthorID	Author	Date	Content	Attachments	Reactions	compound	neg
63	734842980510335179	kyloren#1260	01-Oct-20 02:28 AM	> @kyloren as soon as maintenance is finished...	https://cdn.discordapp.com/attachments/6061118...	♥ (3)	0.6696	0.024
297	476033925093916672	Usman#6305	01-Oct-20 01:09 PM	I see you Spencer!	https://cdn.discordapp.com/attachments/6061118...	fire (7),nets (3)	0.0000	0.000
840	404374804792344589	alxo#6661	01-Oct-20 03:58 PM	this is awesome. congratulations team and comm...	https://cdn.discordapp.com/attachments/6061118...	♥ (2)	0.8516	0.000
1102	114939077647794178	rudem00se#9384	01-Oct-20 05:13 PM	Those are my lowest id, which I'm pretty excit...	https://cdn.discordapp.com/attachments/6061118...	☺ (2)	0.4588	0.171
1230	212928885904506880	Nole77#7519	01-Oct-20 06:35 PM	I live for this	https://cdn.discordapp.com/attachments/6061118...	☐ (7)	0.0000	0.000

```
In [18]:
data.to_csv('sent1.csv')
```

## Senitment column has polarity and then subjectivity

- then we seperated them to their own cols

## Time series sentiment data

```
In [19]:
data = xp.makeData('sent1.csv')
```

## converting date to datetime

```
In [20]:
data['Date'] = data['Date'].astype('datetime64')
```

```
In [21]:
data['just_date'] = data['Date'].dt.date
```

## Plotting time Series

```

groupedTime = data.groupby([data['just_date']]).mean()
In [22]:

groupedTime = groupedTime.drop('AuthorID',axis=1)
In [23]:

groupedTime.head()
In [24]:

Out[24]:

```

	Unnamed: 0	compound	neg	neu	pos
just_date					
2020-10-01	706.400000	0.396000	0.039000	0.735800	0.225200
2020-10-02	1752.800000	-0.064360	0.120400	0.834600	0.045000
2020-10-03	1966.000000	0.785100	0.000000	0.772000	0.228000
2020-10-04	2409.666667	0.309533	0.000000	0.870667	0.129333
2020-10-05	2826.125000	0.224175	0.018375	0.913750	0.067875

```

groupedTime = groupedTime.drop(['Unnamed: 0', 'neg', 'neu', 'pos'],axis=1)
In [25]:

groupedTime.plot(figsize=(40,20))
plt.rc('font', size=30)
plt.suptitle("Sentiment of Topshot Discord General Channel since 10-1-20")
plt.xlabel('Compount Sentiment Score ranging from -1 to 1')
plt.ylabel('Time by months since we collected the data')
plt.savefig('topsent.png')
In [26]:

```



# Use vader for sentiment over text blob, better for social media

In [ ]:

In [ ]: