```
#Adil Muhammed Ashraf PK
#AM.EN.U4CSE19003
#S5 CSE-A


def update_media(x):
    if(x>=12):
        return 12
    else:
        return x

import pandas as pd
pd.options.mode.chained_assignment = None
data=pd.read_csv('/content/sample_data/Adil.csv')
data['is_verified']=data['is_verified'].replace({False : 0})
data['is_verified']=data['is_verified'].replace({True : 1})
data['is_private']=data['is_private'].replace({False : 0})
data['is_private']=data['is_private'].replace({True : 1})
data.drop('fbid',inplace=True, axis=1)
data.drop('seo_category_infos',inplace=True, axis=1)
data.drop('business_phone_number',inplace=True,·axis=1)
data.drop('business_email',inplace=True, axis=1)
data['is_professional_account']=data['is_professional_account'].replace({False : 0})
data['is_professional_account']=data['is_professional_account'].replace({True : 1})
data['has_anonymous_profile_picture']=data['has_anonymous_profile_picture'].replace({False : 0})
data['has_anonymous_profile_picture']=data['has_anonymous_profile_picture'].replace({True : 1})
data['has_clips']=data['has_clips'].replace({False : 0})
data['has_clips']=data['has_clips'].replace({True : 1})
data.fillna(0, inplace=True)
data.drop('username.1',inplace=True, axis=1)
for i in range(0,len(data['username'])):
    if(data['external_url'][i]!=0):
        data['external_url'][i]=1
data['name_length'] = data['full_name'].str.len()
data['media_count_of_data_collected']=data['media_count'].apply(lambda x: update_media(x))
data['average_likes']=data['total_likes']/data['media_count_of_data_collected']
```

```
data.fillna(0, inplace=True)
data.to_csv('cleaned.csv',index=False)
data.head()
```

	username	full_name	is_verified	has_anonymous_profile_picture	biography_len	external_url	followers	fo
0	shkb_shaz	شكيب شاهز MuhammedShakkeeb	0	0	68	0	1177	
1	sayoojbkumar	$🅱k	0	0	51	1	853	
2	devadath__pramod_	Devadath 브	0	0	0	0	307	
3	muhd____ramshad	Ramshad	0	0	0	0	371	
4	s_h_i_n_z_	Shins	0	0	0	0	598	

```
df = data
```

```
print(df.isnull().sum())
```

```
username                        0
full_name                       0
is_verified                     0
has_anonymous_profile_picture   0
biography_len                   0
external_url                    0
followers                       0
following                       0
has_clips                       0
highlight_count                 0
is_professional_account         0
is_private                      0
media_count                     0
total_tags                      0
total_likes                     0
total_comments                  0
name_length                     0
```

```
media_count_of_data_collected    0
average_likes                    0
dtype: int64
```

```
normalized=pd.read_csv('cleaned.csv')
normalized.drop('username',inplace=True,axis=1)
normalized.drop('full_name',inplace=True,axis=1)
normalized = normalized.iloc[:,1:]
x = normalized.iloc[:,1:]
normalized.iloc[:,1:] = (x-x.min())/ (x.max() - x.min())
data=normalized
data
```

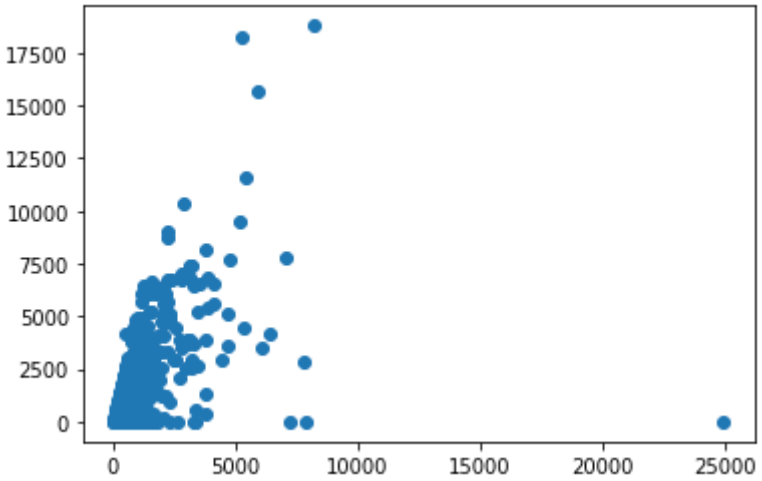
	has_anonymous_profile_picture	biography_len	external_url	followers	following	has_clips	highlight_count	is_professi
0	0	0.453333	0.0	0.046878	0.155214	1.0	0.192308	
1	0	0.340000	1.0	0.033907	0.121070	0.0	0.269231	
2	0	0.000000	0.0	0.012050	0.041542	0.0	0.019231	
3	0	0.000000	0.0	0.014612	0.088633	0.0	0.038462	
4	0	0.000000	0.0	0.023699	0.023616	0.0	0.000000	
...	...	...	...	...	...	...	...	
489	0	0.000000	0.0	0.040272	0.105136	0.0	0.076923	
490	0	0.700000	0.0	0.044115	0.097738	0.0	0.173077	
491	0	0.000000	0.0	0.281305	0.157917	1.0	0.269231	
492	0	0.253333	0.0	0.064532	0.119078	1.0	0.134615	
493	0	1.000000	0.0	0.013811	0.034429	1.0	0.076923	

494 rows × 16 columns

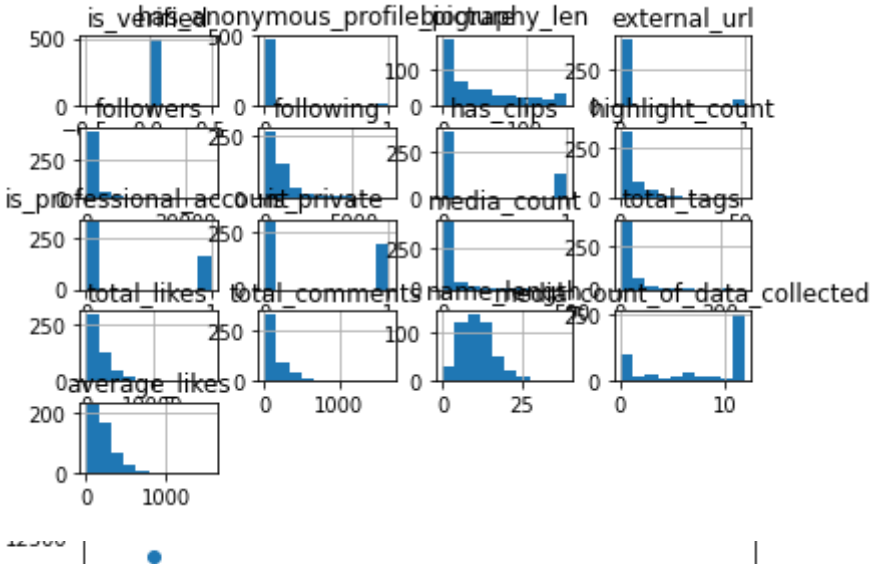
```
plt.scatter(df['followers'], df['total_likes'])
plt.show()

plt.scatter(df['followers'], df['media_count'])
plt.show()

plt.scatter(df['media_count'], df['total_likes'])
plt.show()
```

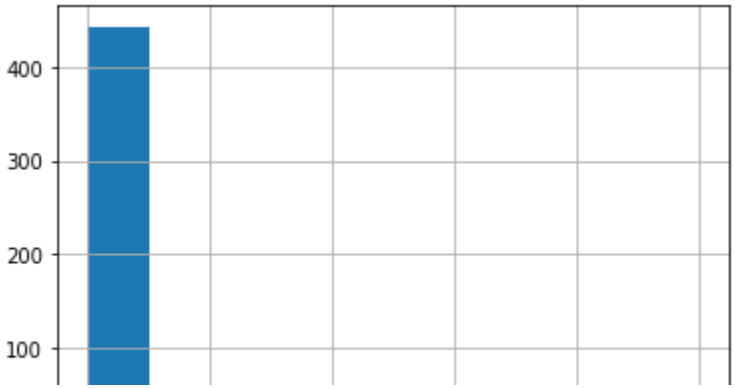



```
df.hist()
plt.show()
```



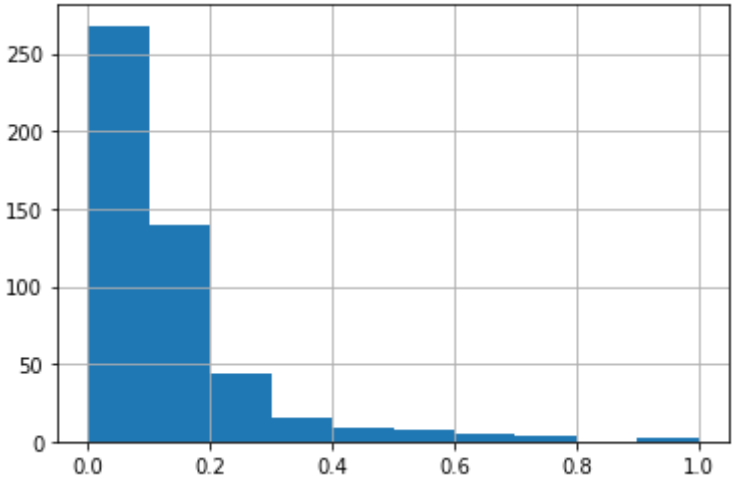

```
data.followers.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df0a2c0d0>
```



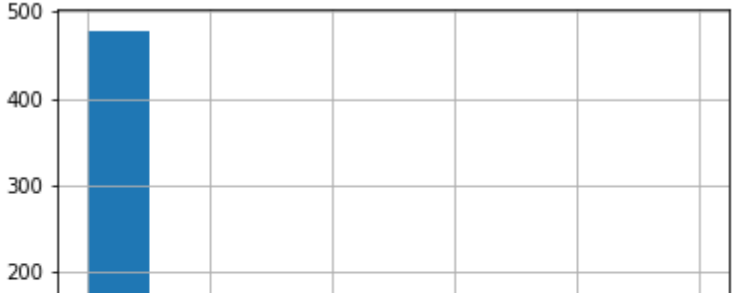

```
data.following.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df834cf90>
```



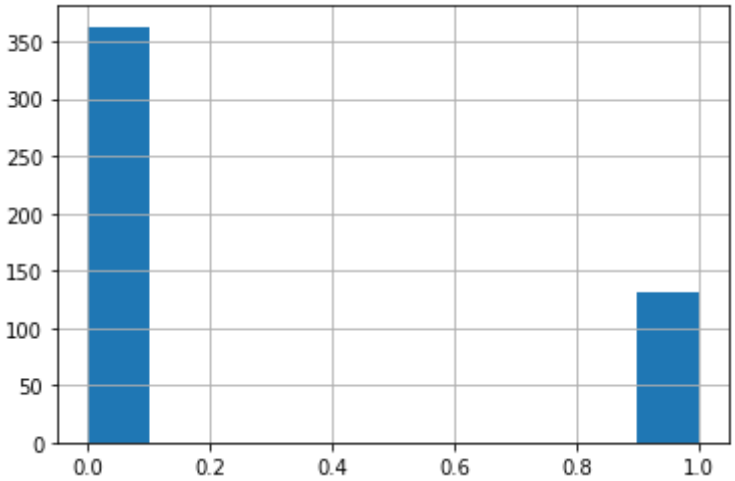

```
data.has_anonymous_profile_picture.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df6594890>
```



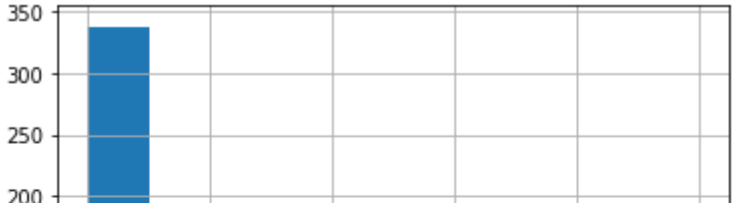

```
data.has_clips.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df1200f10>
```



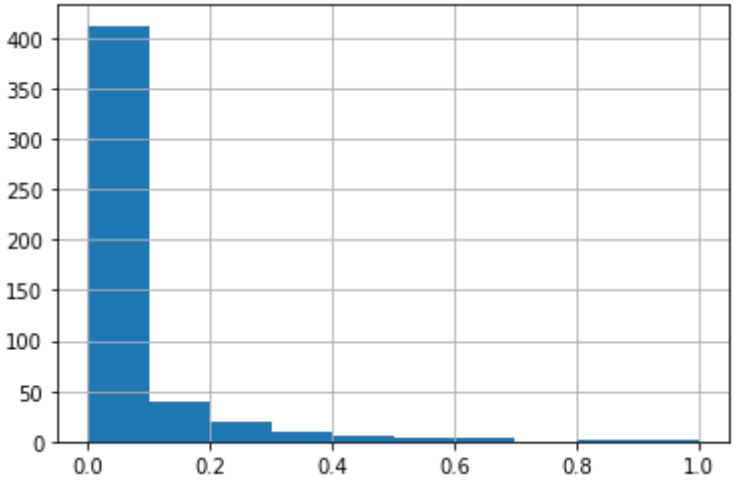

```
data.highlight_count.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df7b78bd0>
```



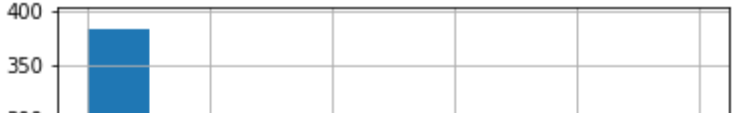

```
data.media_count.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df7901d10>
```



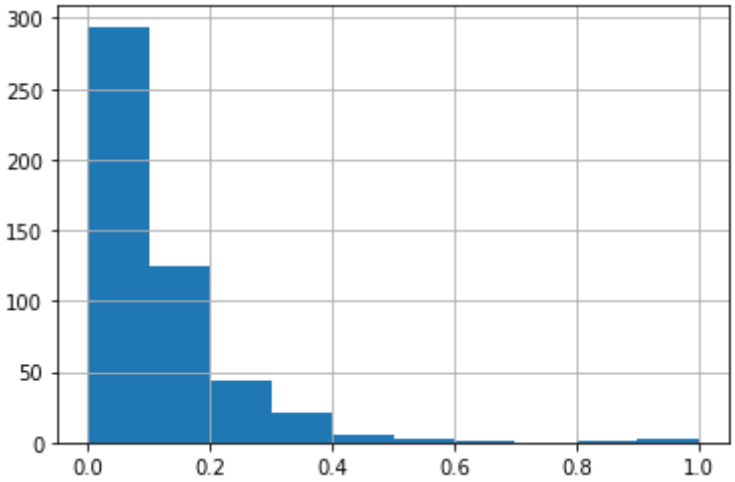

```
data.total_tags.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df5d66390>
```




```
data.total_likes.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df79daf90>
```




```
data.total_comments.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0df62b4810>
```

```
data.describe()
```

	has_anonymous_profile_picture	biography_len	external_url	followers	following	has_clips	
count	494.000000	494.000000	494.000000	494.000000	494.000000	494.000000	
mean	0.032389	0.296127	0.080972	0.046907	0.133858	0.265182	
std	0.177210	0.305922	0.273068	0.065401	0.138687	0.441878	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.013333	0.000000	0.015663	0.051359	0.000000	
50%	0.000000	0.193333	0.000000	0.030404	0.094252	0.000000	
75%	0.000000	0.505000	0.000000	0.050400	0.157277	1.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	

✓ 0s completed at 16:37