```
In [1]:  import pandas as pd
         import matplotlib
         from pandas import DataFrame
         import json
         import excel
         import tweepy as tw


         API_KEY = "xloKoOYLYwftMGE1JcBaJ0csx"
         API_SECRET_KEY = "xG4MdEYo5W6SGhTRb0SYGYvGE5gzHbckbuwenIO6kN6xPORZ6
         ACCESS_TOKEN = "1340243550898774016-7am491fPZXjo0M0gzl7OVOhFlmOyw5"
         ACCESS_TOKEN_SECRET = "dC74oCfNJQwvcgn5YyfyVLHAQsA1mudbQue17hezhYz3

         auth = tw.OAuthHandler(API_KEY, API_SECRET_KEY)
         auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
         api = tw.API(auth, wait_on_rate_limit=True)


         count = 10000

         search_query = "#INDvsAUSTest -filter:retweets"


         result = tw.Cursor(api.search,
                            q=search_query,
                            since="2000-01-01").items(count)

         print("starting")
         tweets_copy = []
         tweets_df = pd.DataFrame()
         for tweet in result:
             hashtags = []

             try:
                 for hashtag in tweet.entities["hashtags"]:
                     hashtags.append(hashtag["text"])
                 text = api.get_status(id=tweet.id, tweet_mode='extended').f
             except:
                 pass
             tweets_df = tweets_df.append(pd.DataFrame({'user_name': tweet.u
                                                        'user_location': twe
                                                        'user_description':
                                                        'user_verified': twe
                                                        'date': tweet.create
                                                        'text': text,
                                                        'retweet_count': twe
                                                        'followers_count': t
                                                        'len': len(text),
                                                        'lang': tweet.lang,
                                                        'hashtags': [hashtag
                                                        'source': tweet.sour
             tweets_df = tweets_df.reset_index(drop=True)

         # show the dataframe
         print("done")

         with open('tweets.json', 'w') as json_file:
             json.dump(tweets_df.to_json(orient="table"), json_file)
```
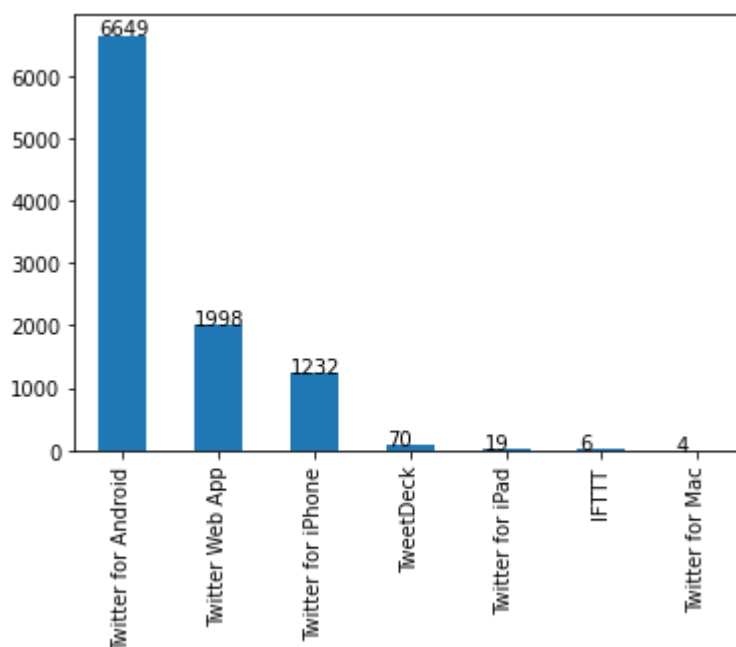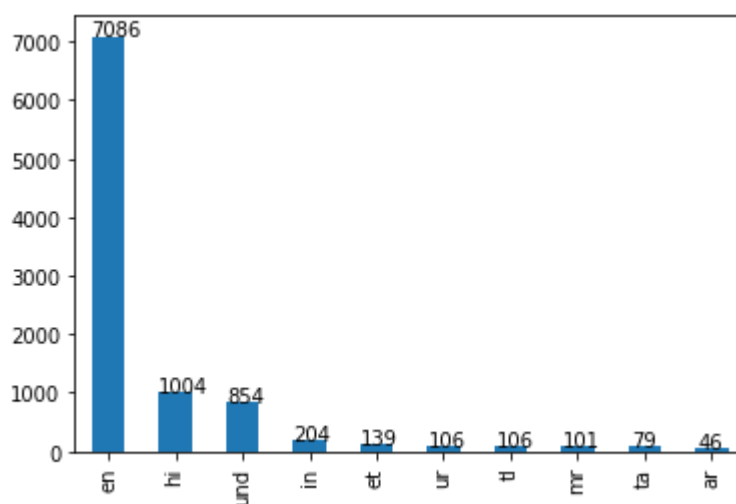
done

In [2]: tweets.df

Out[2]:

	user_name	user_location	user_description	user_verified	date	
0	Ottko	Madhya Pradesh, India	Phir Bhi Dil Hai Hindustani\nBusiness Entrepre...	False	2020-12-22 13:12:31	What hap we don' proper play
1	crictweets		Analysis of current cricket.\nLatest Cricket N...	False	2020-12-22 13:05:38	What sho the openir for 🇮🇳? #B
2	Oneindia Bengali	India	A Bengali news portal brings news in Bengali f...	False	2020-12-22 12:56:12	সঙ্কটে ভারত! কা আগে অসে
3	Shaleen Rao	Earth..Opinions are Personal..	Passionate about Music, Movies & Cricket.Stint...	False	2020-12-22 12:55:24	Former Aus Cric predicti
4	Oneindia Bengali	India	A Bengali news portal brings news in Bengali f...	False	2020-12-22 12:52:02	শামির হাড্ডাহাড্ডি সিরাঞ
...	...	...	...	...	...	
9995	Malik Hafeez Awan	Islamabad, Pakistan	ElectronicEngineer, CSS Aspirant, CricketingFa...	False	2020-12-19 06:20:10	Indians are rap #Adelaic
9996	Aditya	India	Ancora Imparo	False	2020-12-19 06:20:08	Let's forg nightma blame it o
9997	karthick P	chennai	journalist @pttvonlinenews	Tweets &Views are...	False	2020-12-19 06:20:08	மிகக் கு ரன்னில் அ இந்தியா!\n#I
9998	Marcus Aurelias		Positive and friendly	False	2020-12-19 06:20:04	Well just you thoug the year c
9999	Abhishek	Bhārat, that is India. भारत ।	Please don't follow.	False	2020-12-19 06:20:04	#Viratl team's test victory c

10000 rows × 12 columns

In [131]:
```
ax = pd.value_counts(tweets_df['source']).head(7).plot.bar()
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_heig
```
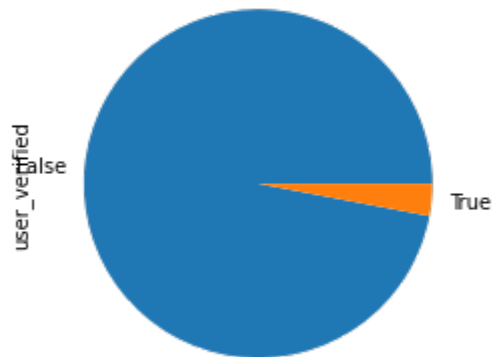

In [132]:
```
ax = pd.value_counts(tweets_df['lang']).head(10).plot.bar()
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_heig
```

In [134]:
```python
pd.value_counts(tweets_df['user_verified']).plot.pie()
```

Out[134]: `<AxesSubplot:ylabel='user_verified'>`

In [135]:
```python
tweets_df['followers_count'] = tweets_df['followers_count'].astype(
tweets_df['len'] = tweets_df['len'].astype(int)
```

In [136]:
```python
tweets_df.sort_values(by=['followers_count'], inplace=True, ascendi
```
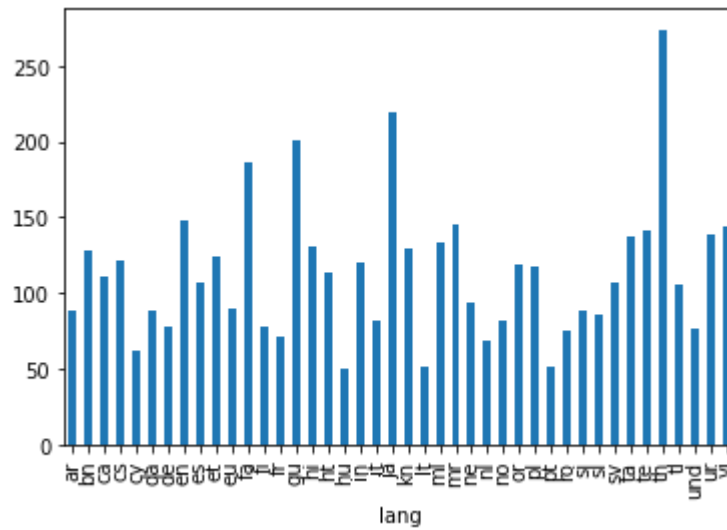
In [137]: `tweets_df`

Out[137]:

	user_name	user_location	user_description	user_verified	date	
7507	AajTak	India	AajTak covers breaking news, latest news in po...	True	2020-12-19 07:31:12	
2161	AajTak	India	AajTak covers breaking news, latest news in po...	True	2020-12-19 15:26:47	ए
999	AajTak	India	AajTak covers breaking news, latest news in po...	True	2020-12-20 05:04:40	ज
628	AajTak	India	AajTak covers breaking news, latest news in po...	True	2020-12-20 10:41:49	इस
8947	ABP News	India	Follow for latest news alerts from India	True	2020-12-19 06:37:37	#I
...	...	...	...	...	...	
5070	ameenalsharef			False	2020-12-19 08:52:23	
2705	Ekaansh			False	2020-12-19 13:36:47	
3317	Bclontia_7	Rohtak, India	💻Engineer\nPassionate 🏏 ⚽ Fan\n#ALL_LIVES_MATTE...	False	2020-12-19 11:32:58	O
3873	Baskar Nagarajan	Indian	Music buff.Great fan of Raja sir(Ilaiyaraja)	False	2020-12-19 10:30:16	1
2853	Abu		Football player	False	2020-12-19 12:58:58	Wa

10000 rows × 12 columns

In [138]: `tweets_df.sort_values(by=['len'], inplace=True, ascending=False)`

In [139]:
```python
mean = tweets_df.groupby('lang')['len'].mean()
tweets_df.groupby('lang')['len'].mean().plot(kind='bar')
```

Out[139]: <AxesSubplot:xlabel='lang'>

In [140]: print(mean)

```
lang
ar      88.043478
bn     127.277778
ca     111.250000
cs     121.500000
cy      62.333333
da      88.428571
de      77.166667
en     148.402907
es     107.320000
```

In [141]: `tweets_df.sort_values(by=['followers_count'], inplace=True, ascendi`

In [142]: `tweets_df.user_name.head(30).unique()`

Out[142]:
```
array(['AajTak', 'ABP News', 'Wasim Akram', 'Zee News English',
       'Shoaib Akhtar', 'MANOJ TIWARY', 'Sorabh Pant', 'Geo News U
rdu',
       'Republic', 'DNA'], dtype=object)
```

In [143]: `tweets_df.sort_values(by=['retweet_count'], inplace=True, ascending`

In [144]: `tweets_df`

Out[144]:
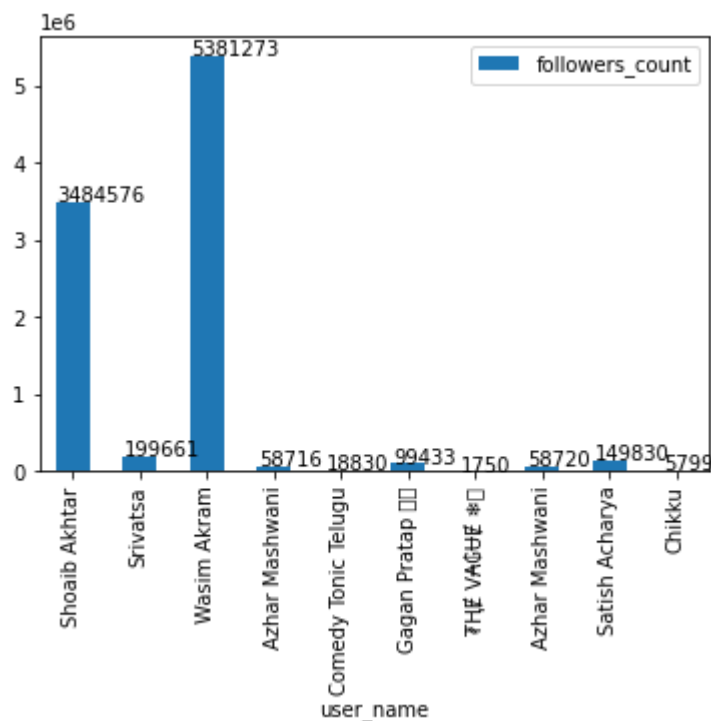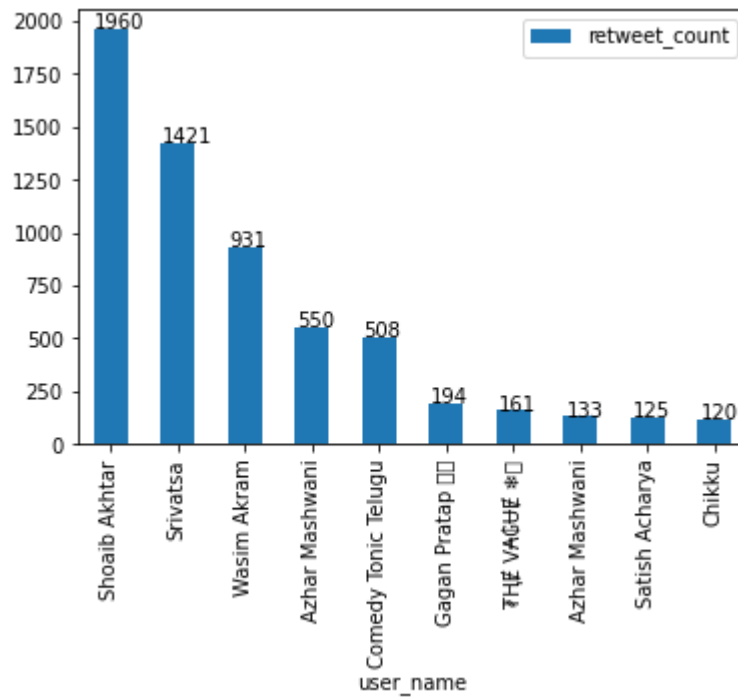
	user_name	user_location	user_description	user_verified	date	
6923	Shoaib Akhtar		YouTube: https://t.co /cj87mk4skE Instagram: ht...	True	2020-12-19 07:44:08	I woke up &a saw the s 369. I could
7581	Srivatsa	Bengaluru	Congressman | Politics | Karnataka | 🇮🇳🇮🇳🇮🇳🇮🇳 | P...	True	2020-12-19 07:29:22	@KanganaT is Jay Sh 'Worst Prod
7443	Wasim Akram		Pakistan's former Cricket Captain, Internation...	True	2020-12-19 07:32:39	Wait happened? went to p h
2306	Azhar Mashwani	Pakistan	Focal Person (Digital Media) to CM Punjab || F...	True	2020-12-19 15:01:17	😂 #INDvsAUS #AUS https:. /KbfG
3846	Comedy Tonic Telugu	Hyderabad, India	Do Follow us on Facebook- \nhttps://t.co /Zl5Fz...	False	2020-12-19 10:32:55	#INDvsAUS https: /j57hY75
...	...	...	...	...	...	
8758	BabaBrainwash	Mumbai	WannabeSinger, Nationalist, Die Hard AndazApna...	False	2020-12-19 06:43:43	#INDvsAUS Abey Shas nikaalo
680	KHEL UPDATES	Doha, Qatar	Follow us for comprehensive coverage on #2022W...	False	2020-12-20 09:23:03	Call on #Sy Test wo rushe #COV

	user_name	user_location	user_description	user_verified	date	
4887	Stephenv		Writer into cricket, film and tv	False	2020-12-19 09:03:36	Dear In cricket fans have been w
9880	Sagar Maverick		Machine Learning and Deep Learning enthusiast...	False	2020-12-19 06:21:34	लगता है र कोहली के ब मुंडन पे जान
6455	Gentleman		Luv POM KLEMENTIEFF	False	2020-12-19 07:54:13	I assure you with this India wil

In [130]:
```python
ax = tweets_df.head(10).plot(x='user_name', y='retweet_count', kind
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_heig

ax = tweets_df.head(10).plot(x='user_name', y='followers_count', ki
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_heig
```


In [ ]: