In [ ]:

```
                  Analysis of comparing user posts with more comments
                   begining with AsK HN or Show HN from Hacker News


Hacker News is one of the popular sites where technology related storie
s or posts are voted and commented.
In this project we will compare and explore two types of posts, Ask HN
and Show HN from the Hackers News data.
Comparing these two types of posts helps to determine which post receiv
es more comments on average and to
see if posts created at a certain time receive more comments on average
.
```

In [53]:

```
# First we read the data

from csv import reader
opened_file = open('HackerNews.csv', encoding= 'utf8')
read_file = reader(opened_file)
hn = list(read_file)
print(hn[0])
print('\n')
print(hn[:6])
```

```
['id', 'title', 'url', 'num_points', 'num_comments', 'author', 'created
_at']


[['id', 'title', 'url', 'num_points', 'num_comments', 'author', 'create
d_at'], ['12579008', 'You have two days to comment if you want stem cel
ls to be classified as your own', 'http://www.regulations.gov/document?
D=FDA-2015-D-3719-0018', '1', '0', 'altstar', '9/26/2016 3:26'], ['1257
9005', 'SQLAR  the SQLite Archiver', 'https://www.sqlite.org/sqlar/doc/
trunk/README.md', '1', '0', 'blacksqr', '9/26/2016 3:24'], ['12578997',
'What if we just printed a flatscreen television on the side of our box
es?', 'https://medium.com/vanmoof/our-secrets-out-f21c1f03fdc8#.ietxmez
43', '1', '0', 'pavel_lishin', '9/26/2016 3:19'], ['12578989', 'algorit
```

```
hmic music', 'http://cacm.acm.org/magazines/2011/7/109891-algorithmic-c
omposition/fulltext', '1', '0', 'poindontcare', '9/26/2016 3:16'], ['12
578979', 'How the Data Vault Enables the Next-Gen Data Warehouse and Da
ta Lake', 'https://www.talend.com/blog/2016/05/12/talend-and-Â\x93the-d
ata-vaultÂ\x94', '1', '0', 'markgainor1', '9/26/2016 3:14']]
```

In [74]:
```
#deleting the header column from the list of lists and check by printin
g the first 5 rows

from csv import reader
opened_file = open('HackerNews.csv', encoding= 'utf8')
read_file = reader(opened_file)
hn = list(read_file)
hn = hn[1:]
print(hn[:5])
```

```
[['12579008', 'You have two days to comment if you want stem cells to b
e classified as your own', 'http://www.regulations.gov/document?D=FDA-2
015-D-3719-0018', '1', '0', 'altstar', '9/26/2016 3:26'], ['12579005',
'SQLAR  the SQLite Archiver', 'https://www.sqlite.org/sqlar/doc/trunk/R
EADME.md', '1', '0', 'blacksqr', '9/26/2016 3:24'], ['12578997', 'What
if we just printed a flatscreen television on the side of our boxes?',
'https://medium.com/vanmoof/our-secrets-out-f21c1f03fdc8#.ietxmez43',
'1', '0', 'pavel_lishin', '9/26/2016 3:19'], ['12578989', 'algorithmic
music', 'http://cacm.acm.org/magazines/2011/7/109891-algorithmic-compos
ition/fulltext', '1', '0', 'poindontcare', '9/26/2016 3:16'], ['1257897
9', 'How the Data Vault Enables the Next-Gen Data Warehouse and Data La
ke', 'https://www.talend.com/blog/2016/05/12/talend-and-Â\x93the-data-v
aultÂ\x94', '1', '0', 'markgainor1', '9/26/2016 3:14']]
```

In [58]:
```
#Here, we extract the two posts by first identifying the posts begining
 either with Ask HN or Show HN.

ask_posts=[]
show_posts=[]
other_posts=[]

for row in hn:
    x = row[1]
```

```
   # x = x.lower()

    if x.lower().startswith('ask hn') is True:
        ask_posts.append(row)

    elif x.lower().startswith('show hn') is True:
         show_posts.append(row)

    else:
        other_posts.append(row)

print(len(ask_posts))
print(len(show_posts))

print(ask_posts[:6])
print('\n')
print(show_posts[:8])
print(len(other_posts))
```

```
9139
10158
[['12578908', 'Ask HN: What TLD do you use for local development?', '',
'4', '7', 'Sevrene', '9/26/2016 2:53'], ['12578522', 'Ask HN: How do yo
u pass on your work when you die?', '', '6', '3', 'PascLeRasc', '9/26/2
016 1:17'], ['12577908', 'Ask HN: How a DNS problem can be limited to a
geographic region?', '', '1', '0', 'kuon', '9/25/2016 22:57'], ['125778
70', 'Ask HN: Why join a fund when you can be an angel?', '', '1', '3',
'anthony_james', '9/25/2016 22:48'], ['12577647', 'Ask HN: Someone uses
stock trading as passive income?', '', '5', '2', '00taffe', '9/25/2016
21:50'], ['12576946', 'Ask HN: How hard would it be to make a cheap, ha
ckable phone?', '', '2', '1', 'hkt', '9/25/2016 19:30']]


[['12578335', 'Show HN: Finding puns computationally', 'http://puns.sam
ueltaylor.org/', '2', '0', 'saamm', '9/26/2016 0:36'], ['12578182', 'Sh
ow HN: A simple library for complicated animations', 'https://christine

cha.github.io/choreographer.js/', '1', '0', 'christinecha', '9/26/2016
```

```
cha.github.io/choreographer-js/', '1', '0', 'christinecha', '9/26/2016
0:01'], ['12578098', 'Show HN: WebGL visualization of DNA sequences',
'http://grondilu.github.io/dna.html', '1', '0', 'grondilu', '9/25/2016
23:44'], ['12577991', 'Show HN: Pomodoro-centric, heirarchical project
management with ES6 modules', 'https://github.com/jakebian/zeal', '2',
'0', 'dbranes', '9/25/2016 23:17'], ['12577142', 'Show HN: Jumble  Essa
ys on the go #PaulInYourPocket', 'https://itunes.apple.com/us/app/jumbl
e-find-startup-essay/id1150939197?ls=1&mt=8', '1', '1', 'ryderj', '9/2
5/2016 20:06'], ['12576813', 'Show HN: Learn Japanese Vocab via multipl
e choice questions', 'http://japanese.vul.io/', '1', '1', 'soulchild3
7', '9/25/2016 19:06'], ['12576627', 'Show HN: Turning a Trello list in
to a shared helpdesk', 'https://boardthreads.com/', '1', '0', 'fiatja
f', '9/25/2016 18:32'], ['12576090', 'Show HN: Markov chain Twitter bo
t. Trained on comments left on Pornhub', 'https://twitter.com/botsonast
y', '3', '1', 'keepingscore', '9/25/2016 16:50']]
273823
```

In [83]:
```
# Comparing which post receives more comments on average.

# Comparing the average for ask posts
total_ask_comments = 0

for posts in ask_posts:
    num_comments = int(posts[4])
    total_ask_comments += num_comments

avg_ask_comments = total_ask_comments / len(ask_posts)

print('Average Number of Comments of Ask posts: ', avg_ask_comments)
```

```
Average Number of Comments on Ask posts:  10.393478498741656
```

In [84]:
```
# Comparing the average for show posts

total_show_comments = 0

for posts in show_posts:
    num_comments = int(posts[4])
    total_show_comments += num_comments
```

```
avg_show_comments = total_show_comments / len(show_posts)

print('Average Number of Comments on Show posts: ', avg_show_comments)
```

```
Average Number of Comments on Show posts:  4.886099625910612
```

In [89]:

```
#Here we create two dictionaries to calculate the amount of ask posts c
reated in each
#hour of the day and the avg number of coments for the ask posts along
 with the no. of coments received


import datetime as dt

result_list = []

for post in ask_posts:
    created_at = post[6]
    num_comments = int(post[4])
    result_list.append([created_at, num_comments])

counts_by_hour = {}
comments_by_hour = {}

for each_row in result_list:
    date = each_row[0]
    comment = each_row[1]
    date = dt.datetime.strptime(date, "%m/%d/%Y %H:%M")
    time = date.strftime("%H")
    if time not in counts_by_hour:
        counts_by_hour[time] = 1
        comments_by_hour[time] = comment
    else:
        counts_by_hour[time] += 1
        comments_by_hour[time] += comment

comments_by_hour
```

Out[89]:

```
{'02': 2996,
```

```
 '01': 2089,
 '22': 3372,
 '21': 4500,
 '19': 3954,
 '17': 5547,
 '15': 18525,
 '14': 4972,
 '13': 7245,
 '11': 2797,
 '10': 3013,
 '09': 1477,
 '07': 1585,
 '03': 2154,
 '23': 2297,
 '20': 4462,
 '16': 4466,
 '08': 2362,
 '00': 2277,
 '18': 4877,
 '12': 4234,
 '04': 2360,
 '06': 1587,
 '05': 1838}
```

In [90]:
```
# below we calculate the average number of comments per post created du
ring each hour of the day.

avg_by_hour = []

for hr in comments_by_hour:
    avg_by_hour.append([hr, round(comments_by_hour[hr] / counts_by_hour
[hr],3)])

avg_by_hour
```

Out[90]:
```
[['02', 11.138],
 ['01', 7.408],
 ['22', 8.804],
 ['21', 8.687],
```

```
['19', 7.163],
['17', 9.45],
['15', 28.676],
['14', 9.692],
['13', 16.318],
['11', 8.965],
['10', 10.684],
['09', 6.653],
['07', 7.013],
['03', 7.948],
['23', 6.697],
['20', 8.749],
['16', 7.713],
['08', 9.191],
['00', 7.565],
['18', 7.943],
['12', 12.38],
['04', 9.712],
['06', 6.782],
['05', 8.794]]
```

Since the above makes it hard to identify the hours with the higest values, we finish by sorting the list of lists and print the five highest values, which would be easier to read. swap_avg_by_hour = [] for hr in avg_by_hour: swap_avg_by_hour.append([hr[1],hr[0]]) swap_avg_by_hour [[5.578, '09'], [14.741, '13'], [13.441, '10'], [13.234, '14'], [16.796, '16'], [7.985, '23'], [9.411, '12'], [11.46, '17'], [38.595, '15'], [16.009, '21'], [21.525, '20'], [23.81, '02'], [13.202, '18'], [7.796, '03'], [10.087, '05'], [10.8, '19'], [11.383, '01'], [6.746, '22'], [10.25, '08'], [7.17, '04'], [8.127, '00'], [9.023, '06'], [7.853, '07'], [11.052, '11']

In [92]:
```
Here we print the top 5 hours for ask posts comments

sorted_swap = sorted(swap_avg_by_hour, reverse=True)

print("Top 5 Hours for Ask Posts Comments")

for avg, hr in sorted_swap[:5]:
    print("{}: {:.2f} average comments per post".format(
        dt.datetime.strptime(hr, "%H").strftime("%H:%M"),avg)
         )
```

```
Top 5 Hours for Ask Posts Comments
15:00: 28.68 average comments per post
13:00: 16.32 average comments per post
12:00: 12.38 average comments per post
02:00: 11.14 average comments per post
10:00: 10.68 average comments per post
```

In [ ]: