

# 臺灣科技大學 社群媒體分析實務 期末作業

電機工程系 M10507514 蘇振驊

# 第一題

- 在三個粉絲頁中共計有多少fans
- 139624776093433粉絲頁,
- 100489613340306粉絲頁
- 1448357445384692粉絲頁

```

import pyes
import json

def pp(o):
    print json.dumps(o, indent=1)

#host = '114.32.24.166:9200'
host = '127.0.0.1:9200'

fid1 = '139624776093433'
fid2 = '1448357445384692'
fid3 = '100489613340306'

def fansNumber(fid):
    conn = pyes.es.ES(host)
    tq = pyes.query.TermQuery(field="fid", value=fid)
    tagg = pyes.aggs.TermsAgg('likes_count', field= 'likes_count')
    qsearch = pyes.query.Search(tq)
    qsearch.agg.add(tagg)
    result = conn.search(query=qsearch , indices='facebook_nested' , doc_types='fanpage')

    for fanNum in result.aggs["likes_count"]["buckets"]:
        print fid + '粉絲頁' + ', ' + 'fans = ' + str(fanNum["key"])
    #print json.dumps(result.aggs,indent=2)

fansNumber(fid1)
fansNumber(fid2)
fansNumber(fid3)

```

```

139624776093433粉絲頁, fans = 580765
1448357445384692粉絲頁, fans = 60673
100489613340306粉絲頁, fans = 1025150

```

## 第二題

- 在三個粉絲頁中共計有多少likes
- 139624776093433粉絲頁,
- 100489613340306粉絲頁
- 1448357445384692粉絲頁

```
import pyes

fid1 = '139624776093433'
fid2 = '1448357445384692'
fid3 = '100489613340306'
conn = pyes.es.ES(server=[('http', 'localhost', 9200)])

def NumberOfLike(fid):

    bq = pyes.query.BoolQuery()
    # BoolQuery本身是一個Query的組合，可以使用add_must(), add_must_not(), add_should()來使用
    bq.add_must(pyes.query.TermQuery(field="fid", value = fid)) #(field, term)
    #bq.add_must_not(pyes.query.TermQuery("name", "john")) #(field, term)
    result = conn.search(query=bq , indices='facebook_nested' , doc_types='like')
    print '粉絲頁'+ fid +'的全部的like :'
    print len(result)
```

```
NumberOfLike(fid1)
NumberOfLike(fid2)|
NumberOfLike(fid3)
```

```
粉絲頁139624776093433的全部的like :
1400259
粉絲頁1448357445384692的全部的like :
388321
粉絲頁100489613340306的全部的like :
7719913
```

## 第三題

- 在三個粉絲頁中共計有多少post
- 139624776093433粉絲頁,
- 100489613340306粉絲頁
- 1448357445384692粉絲頁

## 第三題

- 在三個粉絲頁中共計有多少post
- 139624776093433粉絲頁,
- 100489613340306粉絲頁
- 1448357445384692粉絲頁

原始数据			
_index	_type	_id	_score
{ .kibana	visualization	post-per-page	1
"index": "facebook_nested",			
"type": "post",	visualization	crawled-fanpages	1
"id": "168423776607421_517571445025984",			1
"version": 1,	visualization	Comments-count	1
"score": 1,			1
"source": {	visualization	Likes-count	1
"fid": "168423776607421",	New-Visualization		1
"created_time": "2014-02-18T01:33:19+0000",			1
"updated_time": "2014-02-18T01:33:19+0000",			1
"fname": "台灣IBM",	4.4.1		1
"status_type": "added_photos",	-fanpage		1
"link":			1
"https://www.facebook.com/IBMTaiwan/photos/a.168447479938384.3420			1
type=3",	search	only-post	1
"icon": "https://www.facebook.com/images/icons/photo.gif",			1
"is_hidden": false,	pattern	facebook_an	1
"privacy": {	sualization	timeline-per-page	1
"allow": "",			1
"deny": "",			1
"description": "",		115361891807579_1053743904636035	1
"value": "",			1
"friends": "",		203507002993446_858155294195277	1
},			1
"created_at_linux_timestamp": 1392687199000,		496932930422946_730924503690453	1
"type": "photo",			1
"object_id": "517571221692673",		61891807579_1064548370222255	1
"picture": "https://scontent.xx.fbcdn.net/hphotos-xfal/v/t1.0-			1
0/s130x130/1798800_517571221692673_1239743767_n.jpg?		6831	1
oh=1fd7032cc3ed19850184968c4dbdada3&oe=577E9202",		662317	1
"is_expired": false,			1
"story_tags": {		444264112262451_515392995149562	1
"0": [			1
"post		222708744426746_1008605502503729	1

```
import pyes

conn = pyes.es.ES(server=[('http', 'localhost', 9200)])
bq = pyes.query.BoolQuery()
# BoolQuery本身是一個Query的組合，可以使用add_must(), add_must_not(), add_should()來使用。

bq.add_must(pyes.query.TermQuery(field="fid", value='139624776093433')) #(field, term)
#bq.add_must_not(pyes.query.TermQuery("name","john")) #(field, term)
result = conn.search(query=bq , indices='facebook_nested' , doc_types='post')

print "粉絲頁1396247760934332的全部的post :"
print len(result)
```

粉絲頁1396247760934332的全部的post :  
1120



```
import pyes

conn = pyes.es.ES(server=[('http', 'localhost', 9200)])
bq = pyes.query.BoolQuery()
# BoolQuery本身是一個Query的組合，可以使用add_must(), add_must_not(), add_should()來使用。

bq.add_must(pyes.query.TermQuery(field="fid", value='1448357445384692')) #(field, term)
#bq.add_must_not(pyes.query.TermQuery("name", "john")) #(field, term)
result = conn.search(query=bq, indices='facebook_nested', doc_types='post')

print "粉絲頁1448357445384692的全部的post :"
print len(result)
```

粉絲頁1448357445384692的全部的post :  
969

```
import pyes

conn = pyes.es.ES(server=[('http', 'localhost', 9200)])
bq = pyes.query.BoolQuery()
# BoolQuery本身是一個Query的組合，可以使用add_must(), add_must_not(), add_should()來使用。

bq.add_must(pyes.query.TermQuery(field="fid", value='100489613340306')) #(field, term)
#bq.add_must_not(pyes.query.TermQuery("name", "john")) #(field, term)
result = conn.search(query=bq, indices='facebook_nested', doc_types='post')

print "粉絲頁100489613340306的全部的post :"
print len(result)
```

```
粉絲頁100489613340306的全部的post :
4238
```

## 第四題

- like from unique users

```
原始数据
{ 查询 5 个分片中用的 5 个, 121640926 命中, 耗时 12.240 秒
  "index": "facebook_nested",
  "_type": "like",
  "facebook_nested_like": {
    "fid_pid_uid": "193072760827_10154428611190828_941815425911054",
    "fid": "193072760827",
    "uid": "941815425911054",
    "fname": "統一星巴克咖啡同好會",
    "fid_pid": "193072760827_10154428611190828",
    "created_at_linux_timestamp": "1458916848007",
    "username": "941815425911054",
    "timestamp": "2016-03-25T22:40:48+0800",
    "@version": "1",
    "@timestamp": "2016-03-31T00:41:28.310Z"
  }
}
facebook_nested like 101615286547831_968881086487909_1532160613744488
facebook_nested like 101615286547831_968881086487909_819696671449247
facebook_nested like 101615286547831_968881086487909_576815772495146
facebook_nested like 101615286547831_968881086487909_712393022231668
facebook_nested like 193072760827_10154366038515828_1649497885288614
facebook_nested like 101615286547831_968881086487909_432220370318894
facebook_nested like 101615286547831_968881086487909_566348080185391
facebook_nested like 193072760827_10154366038515828_938173609608333
facebook_nested like 101615286547831_968881086487909_856809791031467
facebook_nested like 193072760827_10154366038515828_10153114872236568
facebook_nested like 101615286547831_968881086487909_447852238745284
facebook_nested like 101615286547831_968881086487909_1254506561232927
facebook_nested like 193072760827_10154366038515828_1113877471980300
facebook_nested like 193072760827_10154428611190828_941815425911054
```

```
import pyes

user1 = '10153114872236568'
user2 = '1532160613744488'
user3 = '819696671449247'

conn = pyes.es.ES(server=[('http', 'localhost', 9200)])

def NumberOfLike(uid):

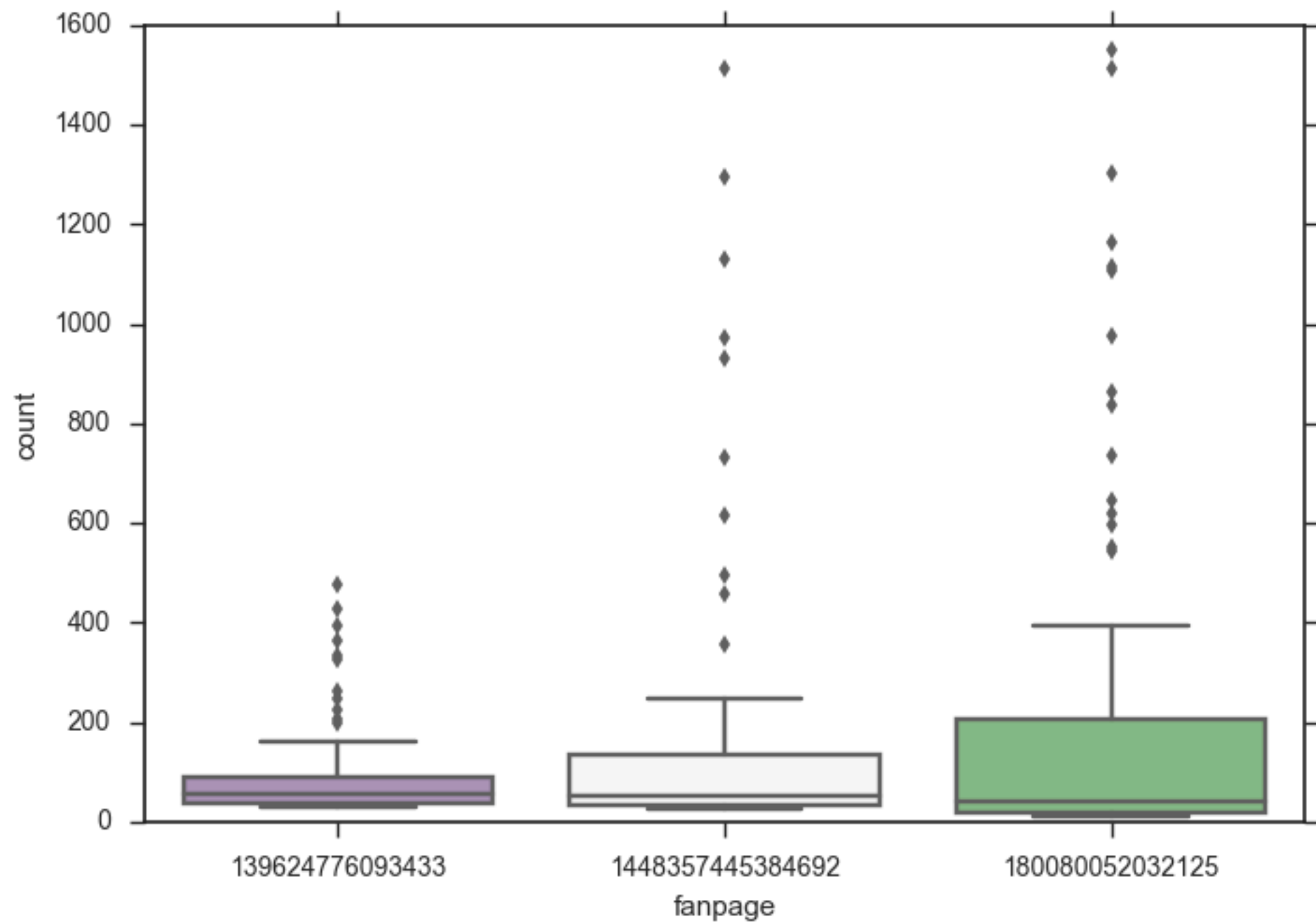
    bq = pyes.query.BoolQuery()
    # BoolQuery本身是一個Query的組合，可以使用add_must(), add_must_not(), add_should()來使用。
    bq.add_must(pyes.query.TermQuery(field="uid", value = uid)) #(field, term)
    #bq.add_must_not(pyes.query.TermQuery("name", "john")) #(field, term)
    result = conn.search(query=bq, indices='facebook_nested', doc_types='like')
    print 'User ' + uid + ' 按了幾個粉絲頁的like :'
    print len(result)

NumberOfLike(user1)
NumberOfLike(user2)
NumberOfLike(user3)
```

```
User 10153114872236568 按了幾個粉絲頁的like :
84
User 1532160613744488 按了幾個粉絲頁的like :
67
User 819696671449247 按了幾個粉絲頁的like :
8
```

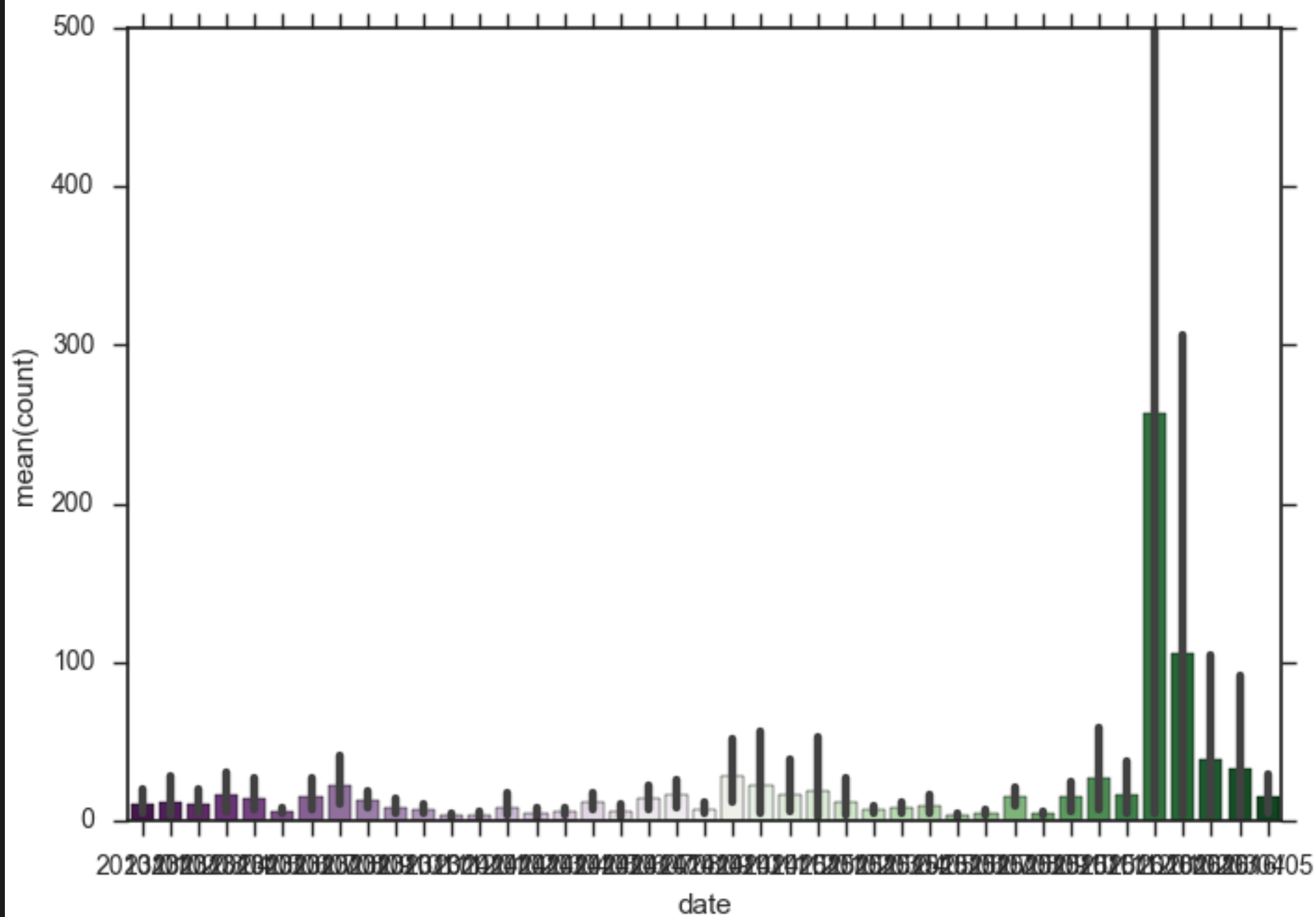
## 第五題

- Comment analytic
- 1. 三個不同粉絲頁的不同的post裡面的comment數量

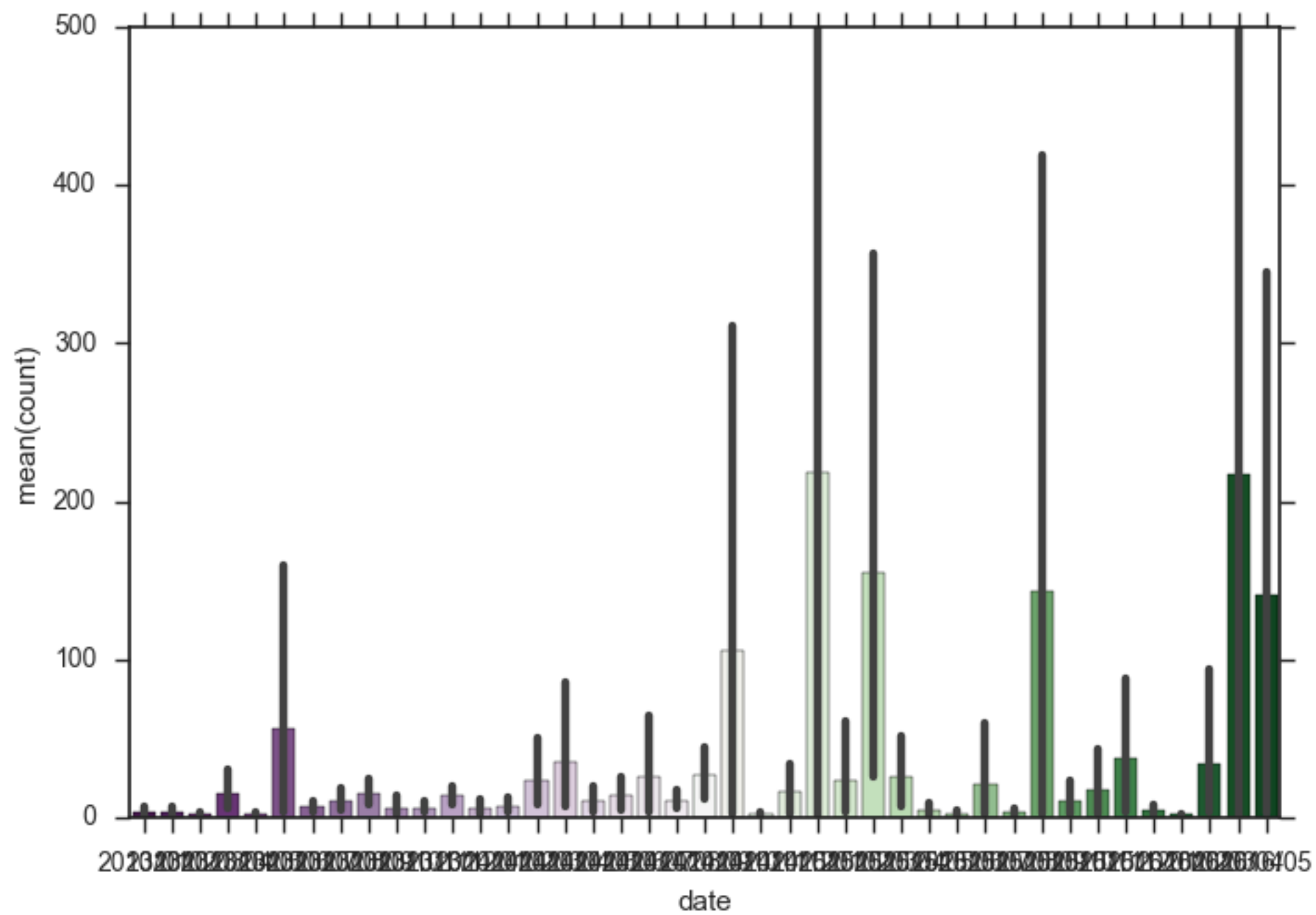


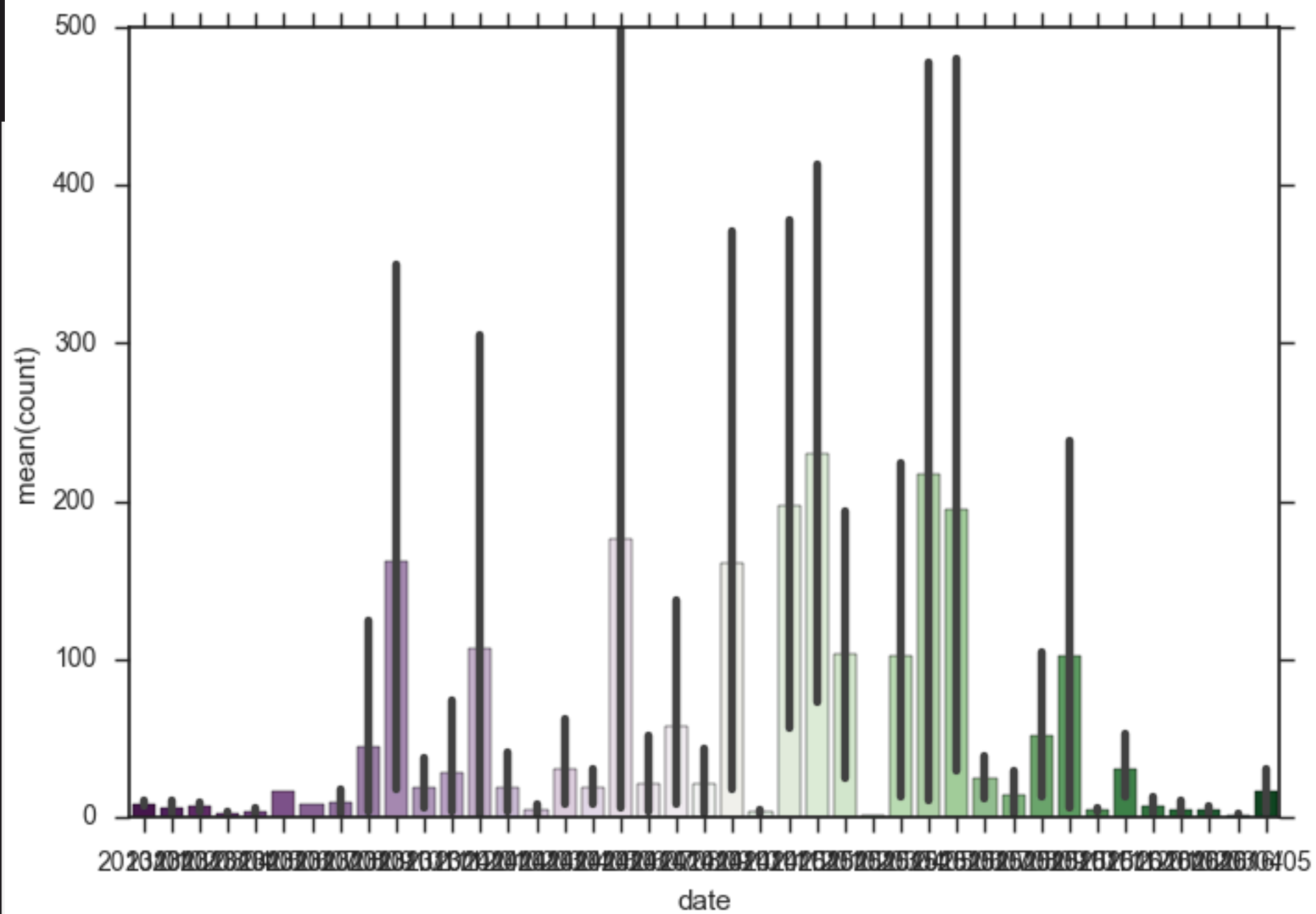
# 第五題

- Comment analytic
- 2. 粉絲頁每個月的變化









- 後續我會繼續研究怎麼樣把users, share的次數找出來，因為目前看資料是在下一層，沒辦法直接query出來。

```

"tid": "154929428406",
"created_time": "2015-02-26T08:50:19+0000",
"updated_time": "2015-02-26T08:50:19+0000",
"indexable": true,
"link": "https://www.facebook.com/154929428406",
"status_type": "added_photos",
"link": "https://www.facebook.com/154929428406",
"facebook_nested_post": {
  "id": "10152667316588407",
  "type": "photo",
  "object_id": "10152846268348407",
  "picture": "https://static.xx.fbcdn.net/rsrc.php/v1/ymqF1qZ1e9.jpg?_nc_cat=103&_nc_ohc=29c26a776dce6025&_nc_ht=static.xx.fbcdn.net&_nc_w=108&_nc_h=108&_nc_rsrc=1",
  "shares": {
    "count": 10
  },
  "is_expired": false,
  "story_tags": {
    "0": {
      "offset": 0,
      "name": "1922防疫達人",
      "length": 8,
      "id": "154929428406",
      "type": "page"
    }
  }
}

```

# 結語

- 透過這學期的期中跟期末作業，我學會了一些資料庫的基本知識，也學會了基本的Query概念，以及怎麼樣使用Query。
- 還有學到了python，雖然還不是太熟悉，但是從課堂上知道了很多相關的資源像是：*SciPy2016*、*numpy*的函式庫。Google GCP CPB100、SVD演算法的介紹。