In [1]:

```
import findspark
findspark.init()
import pyspark as ps
import warnings
from pyspark.sql import SQLContext
```

In [2]:

```
sc = ps.SparkContext('local[8]')
spark = SQLContext(sc)
```

In [3]:

```
import seaborn as sn
sn.set_style("darkgrid")

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

In [4]:

```
df = spark.read.json("hdfs://localhost:9000/input_dir/tweets.json")
df.createOrReplaceTempView("tweets")
```

In [12]:

```
fav = spark.sql("SELECT user.name, text, user.favourites_count FROM twe
ets")
fav = fav.filter(fav.text.contains('beer'))
fav = fav.dropDuplicates(['name'])
fav.count()
```

Out[12]:

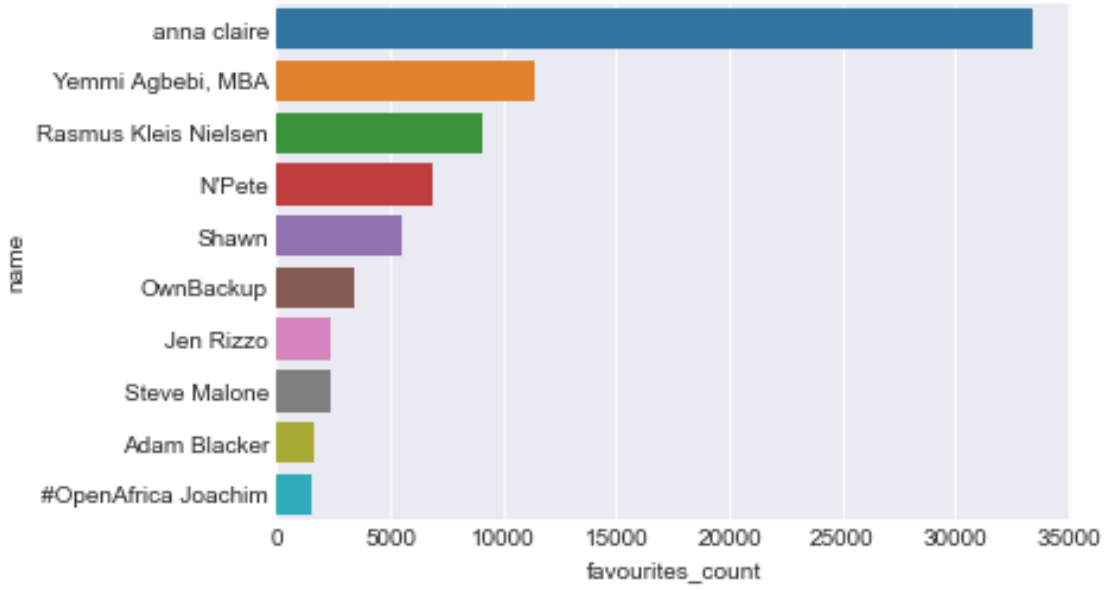
20

In [13]:

```
fav = fav.toPandas()
Top_fav = fav.query('favourites_count > 1000').reset_index().sort_value
s('favourites_count', ascending=False)
sn.barplot(x='favourites_count', y='name', data=Top_fav)
```

Out[13]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1ad0f24a240>
```



In [17]:

```
city = spark.sql("SELECT user.location as city FROM tweets WHERE user.l
ocation != 'null' and user.location != 'United States' and  user.locati
on != 'USA'")
city.createOrReplaceTempView("citys")
city_count = city.groupBy('city').count()
city_count.count()
```

Out[17]:

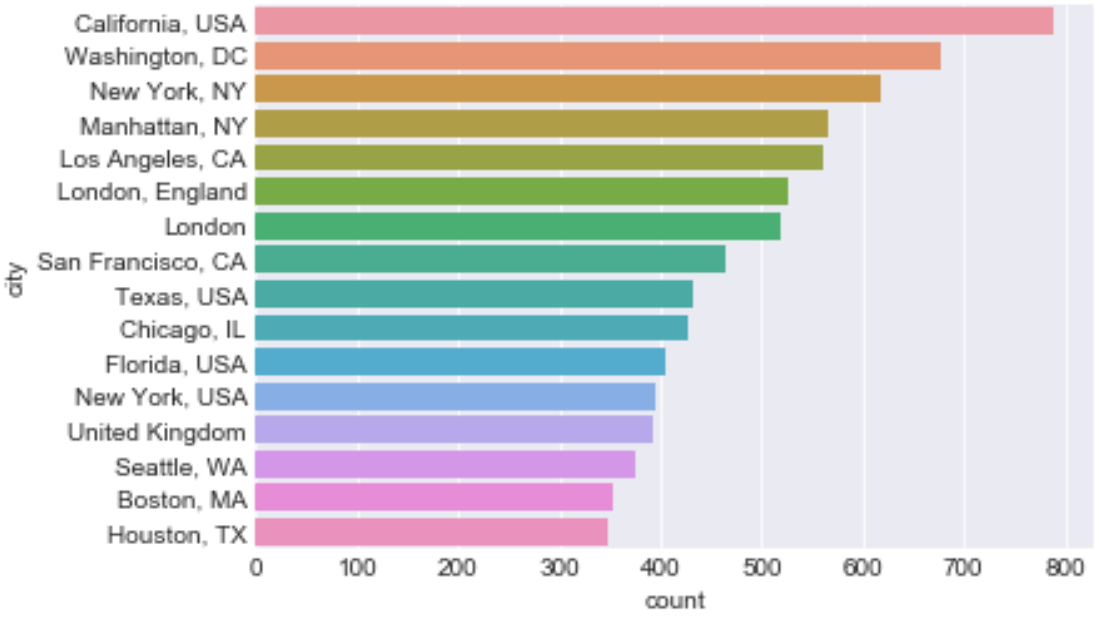
```
24927
```

In [18]:

```
city_count = city_count.toPandas()
Top_city = city_count.query('count > 340').reset_index().sort_values('c
ount', ascending=False)
sn.barplot(x='count', y='city', data=Top_city)
```

Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1ad0fc52b38>
```



In [21]:

```
lang = spark.sql("SELECT user.lang as langs FROM tweets WHERE user.lang
 != 'en' and user.lang != 'en-gb' and user.lang != 'en-GB'")
lang_count = lang.groupBy('langs').count()
lang_count.count()
```

Out[21]:

```
44
```

In [22]:

```
lang_count = lang_count.toPandas()
Top_lang = lang_count.query('count > 100').reset_index().sort_values('c
ount', ascending=False)
sn.barplot(x='count', y='langs', data=Top_lang)
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1ad0fa72a58>
```