In [203]:

```
import pandas as pd
import numpy as np
import re
import os
```

In [204]:

```
os.listdir("C:/Users/jctep/OneDrive/Documents/GitHub/text-mining-computer-vision/
files = os.listdir("C:/Users/jctep/OneDrive/Documents/GitHub/text-mining-compute
```

In [205]:

```
data = pd.read_csv(str("Datos/"+files[0]), sep=" ", header=None, index_col=None)
data.head()
```

Out[205]:

	0
0	23/02/2017
1	21/11/2016
2	12/02/2017
3	06/06/2016
4	04/05/2018

In [34]:

```
str("Datos/"+files[0])
```

Out[34]:

```
'Datos/D1.txt'
```

In [54]:

```
li = []
for filename in files:
    df = pd.read_csv(str("Datos/"+filename), index_col = None, header = None)
    li.append(df)

data = pd.concat(li, axis = 0, ignore_index = True)
data.columns = ['raw_date']
```

In [55]:

```
len(data)
```

Out[55]:

```
21000
```

In [117]:

```
data.head()
```

Out[117]:

	raw_date
0	23/02/2017
1	21/11/2016
2	12/02/2017
3	06/06/2016
4	04/05/2018

In [197]:
```
data.raw_date = data.raw_date.replace(regex=r'[Jj]an', value='01')
data.raw_date = data.raw_date.replace(regex=r'[Ff]eb', value='02')
data.raw_date = data.raw_date.replace(regex=r'[Mm]ar', value='03')
data.raw_date = data.raw_date.replace(regex=r'[Aa]pr', value='04')
data.raw_date = data.raw_date.replace(regex=r'[Mm]ay', value='05')
data.raw_date = data.raw_date.replace(regex=r'[Jj]un', value='06')
data.raw_date = data.raw_date.replace(regex=r'[Jj]ul', value='07')
data.raw_date = data.raw_date.replace(regex=r'[Aa]ug', value='08')
data.raw_date = data.raw_date.replace(regex=r'[Ss]ep', value='09')
data.raw_date = data.raw_date.replace(regex=r'[Oo]ct', value='10')
data.raw_date = data.raw_date.replace(regex=r'[Nn]ov', value='11')
data.raw_date = data.raw_date.replace(regex=r'[Dd]ec', value='12')
data.raw_date = data.raw_date.str.replace('.','/')
data.raw_date = data.raw_date.str.replace('-','/')
```

In [198]:
```
df= data.raw_date.str.split('/',expand= True)
```

In [199]:
```
df.columns = ['day','month','year']
```

In [200]:
```
df.to_csv('fulldata.csv',index=None)
```

In [206]:
```
df['day'].astype(int).mean()
df['month'].astype(int).mean()
df['year'].astype(int).mean()
```

Out[206]:
```
2016.6869047619048
```

In [210]:
```
df['day'] = df['day'].astype(int)
df['month'] = df['month'].astype(int)
df['year'] = df['year'].astype(int)
```

In [211]:
```
df.describe()
```

Out[211]:

	day	month	year
count	21000.000000	21000.000000	21000.000000
mean	15.624762	6.466476	2016.686905
std	8.782500	3.468418	1.441601
min	1.000000	1.000000	2014.000000
25%	8.000000	3.000000	2015.000000
50%	16.000000	6.000000	2017.000000
75%	23.000000	10.000000	2018.000000
max	31.000000	12.000000	2019.000000

In [212]:

```
df.mean()
```

Out[212]:

```
day        15.624762
month       6.466476
year     2016.686905
dtype: float64
```

In [ ]:

```
```