Data Assignment 1

September 18, 2016

0.1 Labib Fawaz

0.2 lf2494

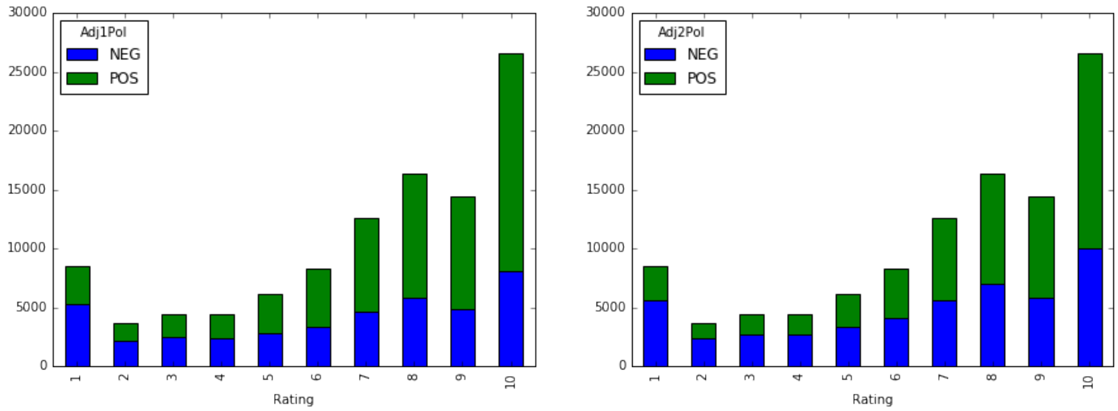
0.2.1 Problem 1

Our hypothesis is that high ratings are more likely to include postive adjectives while low ratings would produce negative ones. We will consider rating 7 and above as positive and ratings below 7 as negative. The graphs below show that Negative adjectives are higher until rating 5 in Adjective 1 and until 6 in Adjective 2.

```
In [43]: import pandas as pd
         import matplotlib.pyplot as plt
         %matplotlib inline

         imdb = pd.read_csv('imdb-adjcoord.csv')
         fig, axs = plt.subplots(1,2)
         imdb.groupby(['Rating','Adj1Pol']).size().unstack().plot(kind='bar',stacked=True, ax=axs[0],fi
         imdb.groupby(['Rating','Adj2Pol']).size().unstack().plot(kind='bar',stacked=True,ax=axs[1],fig

Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x11f72a550>
```



0.2.2 Problem 2

- Feature one: (punctuation) Using an exclamation point tends to be more used in postive texts while quotations are more used in negative ones.
- Feature two: Lenght of the text is shorter in in the postive text than in the negative texts

- Feature three: Using an Emoji in the text seems to indicate that the text is more likely to be positive than negative

```
In [ ]:
```