**Extract movie reviews for any movie from IMDB and perform sentimental analysis**

Extracting [*Black and Blue*](https://www.imdb.com/title/tt7390646/?ref_=tt_urv)  moviereview from IMDB

We are using Chrome WebDriver from selenium package to access the IMDB review page.

Selenium Exceptions are used while navigating through review page.

import requests

from bs4 import BeautifulSoup as bsoup

from selenium import webdriver

from selenium.common.exceptions import NoSuchElementException

from selenium.common.exceptions import ElementNotVisibleException

from selenium.common.exceptions import ElementNotInteractableException

*Opens Chrome browser and loads review page*

browser = webdriver.Chrome("G:\Assingments\Text Mining\chromedriver.exe")

browser.get(url) # url = “url for review page”

*To navigate through the review page*

while (i>=0):

try:

button = browser.find\_element\_by\_xpath('//\*[@id="load-more-trigger"]')

button.click()

time.sleep(3)

except NoSuchElementException:

break

except ElementNotVisibleException:

print("Reached End")

break

except ElementNotInteractableException:

break

Clicks the load more button after waiting for 3 seconds to get more reviews.

If the button label (xpath element) is not found/wrong,

NoSuchElementException of selenium will catch and comes out of loop

If the button label is not visible/end of review is reached.

ElementNotVisibleException will catch and comes out of loop

*Loads entire source page and it is covered into BeautifulSoup object*

response = browser.page\_source

browser.close() # closes browser

soup\_obj = bsoup(response, "html.parser")

BeautifulSoup object is used to extract reviews, reviewer name, review title etc.

HTML objects are used to extract required information.

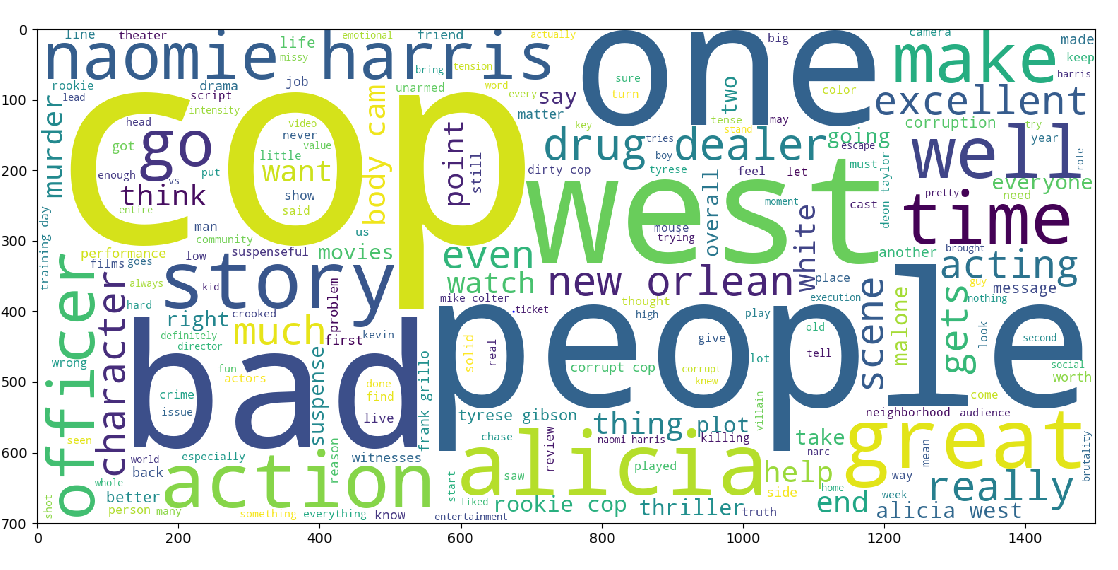
‘*display-name-link’* attribute of *span class tag* for reviewer name

‘*title’* attribute of *anchor tag a* for reviewer title.

‘*text’* attribute of *div class tag a* for reviewer title.

After cleaning up the data and removing stopwords WordClouds are plotted.

Corpus level WordClouds

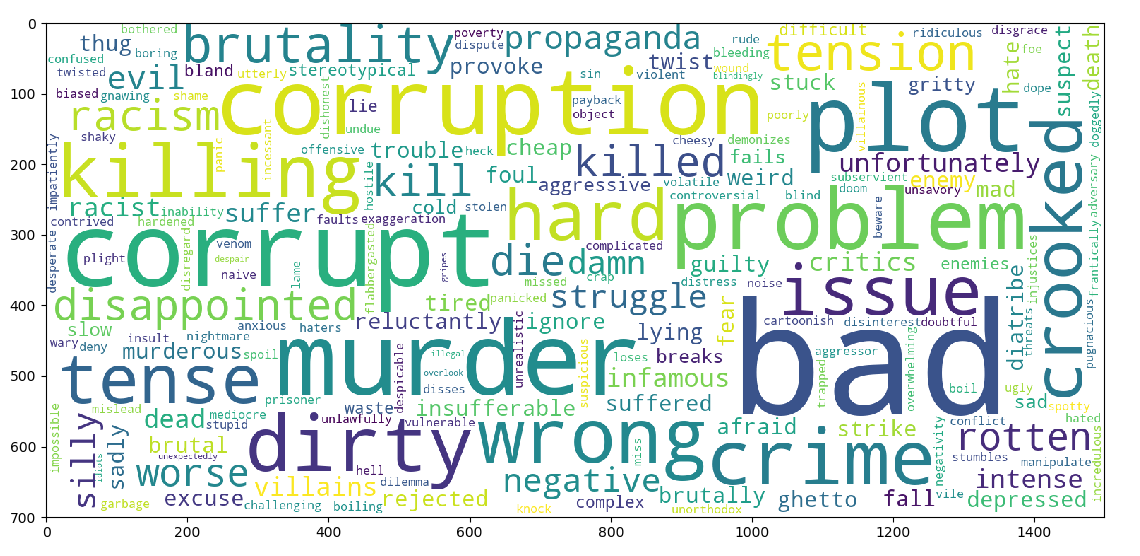


It appears that people are talking about movie roles, movie story and artists

Positive WordCloud



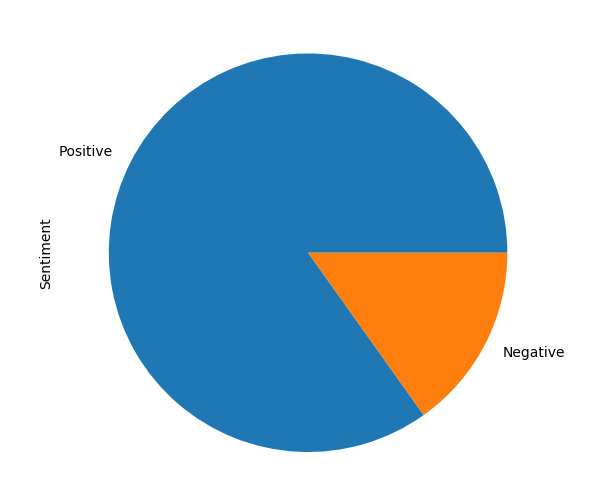
Negative WordCloud



From above WordCloud, we can’t infer much about the movie. As the movie is about corrupt cops and people are talking about the movie, most of the words will be categorized as bad.

We will have to perform Contextual Reading to find out what exactly people are talking

Sentiment Analysis using TextBlob



We can see [*Black and Blue*](https://www.imdb.com/title/tt7390646/?ref_=tt_urv) movie has good feedback