Group members:

1. Islam MD Shariful 1720601
2. Abdella Mame Abdo 1714883
3. Mohamed Moubarak Mohamed Misbahou Mkouboi 1820705

In [31]:
```
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import json
import re
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
from tweepy import OAuthHandler
from tweepy import Stream
from tweepy.streaming import StreamListener
import os
#os.chdir('F:/SEM_5/NATURAL LANGUAGE PROCESSING/Assignment 3')

consumer_key = "UQ8XFSgXLOSnpm8FLt2ggi0c4"
consumer_secret = "9suCc5INoh69KRflUER83K2kAINYXTkT2gm2jO8LttPKENVJ4M"
access_token = "1111498145177849856-P64xYnQ7LMata8ULOHeMRSmngqsaKA"
access_secret= "ICWLB2vC4CIMrNaGgx6BqbJHqJX2RSCFdAwcCYsxYR4sr"



auth= OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_secret)
api=tweepy.API(auth)


file= open('flood_raw.dat','a')


class Mylistener(StreamListener):
    def __init__(self, api=None):
        super(StreamListener,self).__init__()
        self.num_tweet = 0

    def on_data(self,data):
        try:
            with open('flood_filtered.dat','a') as f:
                tweet=json.loads(data)

                if tweet ['lang']=="en":
                    file.write(data)
                    file.write('\n')

                if tweet['lang']=='en' :
                    if self.num_tweet<2000:
                        print(json.dumps(tweet["text"],indent=4))
                        f.write(tweet["text"])
                        f.write("\n")
                        self.num_tweets += 1

                return True
        except BaseException as e:
            print("Error on_data: %s" % str(e))
        return True

def on_error(self,status):
```

```
        print(status)
        return True
def on_status(self,status):
    if status.retweeted_status=='true':
        return
    print(status)

mytwitter_stream = Stream(auth,Mylistener())
mytwitter_stream.filter(track=['flood', 'floodmalaysia', 'malaysia', 'flood2021',
file.close()
print("done")
```

```
"RT @malaysia_covid: Weekly Imported Cases in Malaysia\n\n13/12 - 19/12 : 196
\n20/12 - 26/12 : 687\n27/12 - 2/1 : 1,982\n3/1 - 9/1 : 2,189\n10/1 -\u2026"
Error on_data: 'Mylistener' object has no attribute 'num_tweets'
"RT @MSuppasitTrends: [TRENDS UPDATE]\n\nGolden song with MEW #MewXTheGoldenS
ong \n1\ufe0f\u20e3 Thailand \ud83c\uddf9\ud83c\udded\n5\ufe0f\u20e3 Worldwid
e \ud83c\udf0e \n6\ufe0f\u20e3 Vietnam \ud83c\uddfb\ud83c\uddf3 \n7\ufe0f\u20
e3 Malaysi\u2026"
Error on_data: 'Mylistener' object has no attribute 'num_tweets'
"RT @FishGuyKai: This is an arresting species and one the world\u2019s great
s. A must see insect before you die sort of deal. Trogonoptera brooki\u2026"
Error on_data: 'Mylistener' object has no attribute 'num_tweets'
"RT @terpusing2: Goh Jin wei - World jr Champ, gold youth olympic\nLee Zii Ji
a - All england champ, world No 7\n\nSatu kerugian buat Malaysia."
Error on_data: 'Mylistener' object has no attribute 'num_tweets'
"RT @GhaDaJW: Jackson Wang advertisement for Pepsin in Malaysia \ud83c\uddf2
\ud83c\uddfe \u2764\ufe0f\ud83d\udc99!! \ud83e\udd29\ud83d\udd25\ud83d\udd25
\n\nIs he going for Asia? Or global \ud83d\ude35\u2665\ufe0f\u2665\ufe0f\n\n#
JacksonWang #\uc7ad\uc2a8 #\u738b\u5609\u5c14 #\u2026"
Error on_data: 'Mylistener' object has no attribute 'num_tweets'
```

In [35]:
```
posts = api.user_timeline(screen_name="Malaysia", count = 100, lang ="en", tweet_

#  Print last 5 tweets
print("Show the 5 recent tweets:\n")
i=1
for tweet in posts[:5]:
    print(str(i) +') '+ tweet.full_text + '\n')
    i= i+1

# Create a dataframe with a column called Tweets
df = pd.DataFrame([tweet.full_text for tweet in posts], columns=['Tweets'])
df.head()
```

```
Show the 5 recent tweets:

1) @realDonaldTrump You’re amazing! 💪

2) You’ll always be a champion in the eyes of Malaysians. Wish you happy retire
ment, Datuk @LeeChongWei! Best of health to you too!

3) @LazadaMY @amein_nur_iman There seem to be an error with paying with my wall
et after top up. All my orders get cancelled citing "system default reason".

4) Congratulations Hassan! https://t.co/FYLh7K5KUc (https://t.co/FYLh7K5KUc)

5) Follow Automotive Enthusiast + Next Top Instagrammer https://t.co/cVnXrM484p
(https://t.co/cVnXrM484p) as he travels Japan! #iamthespeedhunter
```

Out[35]:

	Tweets
0	@realDonaldTrump You’re amazing! 💪
1	You’ll always be a champion in the eyes of Mal...
2	@LazadaMY @amein_nur_iman There seem to be an ...
3	Congratulations Hassan! https://t.co/FYLh7K5KUc
4	Follow Automotive Enthusiast + Next Top Instag...

In [36]:
```
# Create a function to clean the tweets
def cleanTxt(text):
 text = re.sub('@[A-Za-z0-9]+', '', text)
 text = re.sub('#', '', text)
 text = re.sub('RT[\s]+', '', text)
 text = re.sub('https?:\/\/\S+', '', text)

 return text


# Clean the tweets
df['Tweets'] = df['Tweets'].apply(cleanTxt)

df

# Create a function to get the subjectivity
def getSubjectivity(text):
   return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
   return  TextBlob(text).sentiment.polarity

# Create two new columns 'Subjectivity' & 'Polarity'
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

df

df.info
```

Out[36]:
```
<bound method DataFrame.info of
Tweets  Subjectivity  Polarity
0                                  You’re amazing! 💪      0.900000  0.750000
1   You’ll always be a champion in the eyes of Mal...      0.650000  1.000000
2    _nur_iman There seem to be an error with payi...      0.500000  0.500000
3                           Congratulations Hassan!      0.000000  0.000000
4   Follow Automotive Enthusiast + Next Top Instag...      0.250000  0.312500
..                                                ...           ...       ...
94  A good gov't needs a strong opposition, says M...      0.666667  0.566667
95  UM pharmacy programme can't be axed without fu...      0.550000  0.350000
96  Ex-spy chief to be released tomorrow, says she...      0.550000  0.000000
97  Rafizi says his bid for PKR No 2 is to safegua...      0.000000  0.000000
98  PM limits ministers to 10 aides, Dzulkefly con...      0.000000  0.000000

[99 rows x 3 columns]>
```

In [38]:
```
# word cloud visualization
allWords = ' '.join([twts for twts in df['Tweets']])
wordCloud = WordCloud(width=600, height=400, random_state=21, max_font_size=110).


plt.imshow(wordCloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```

In [39]:
```
# Create a function to compute negative (-1), neutral (0) and positive (+1) analy
def getAnalysis(score):
    if score < 0:
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    elif score>0:
        if score >=0.4:
            return "Booster"
        else:
            return "positive"

df['Analysis'] = df['Polarity'].apply(getAnalysis)
# Show the dataframe
df
```

Out[39]:

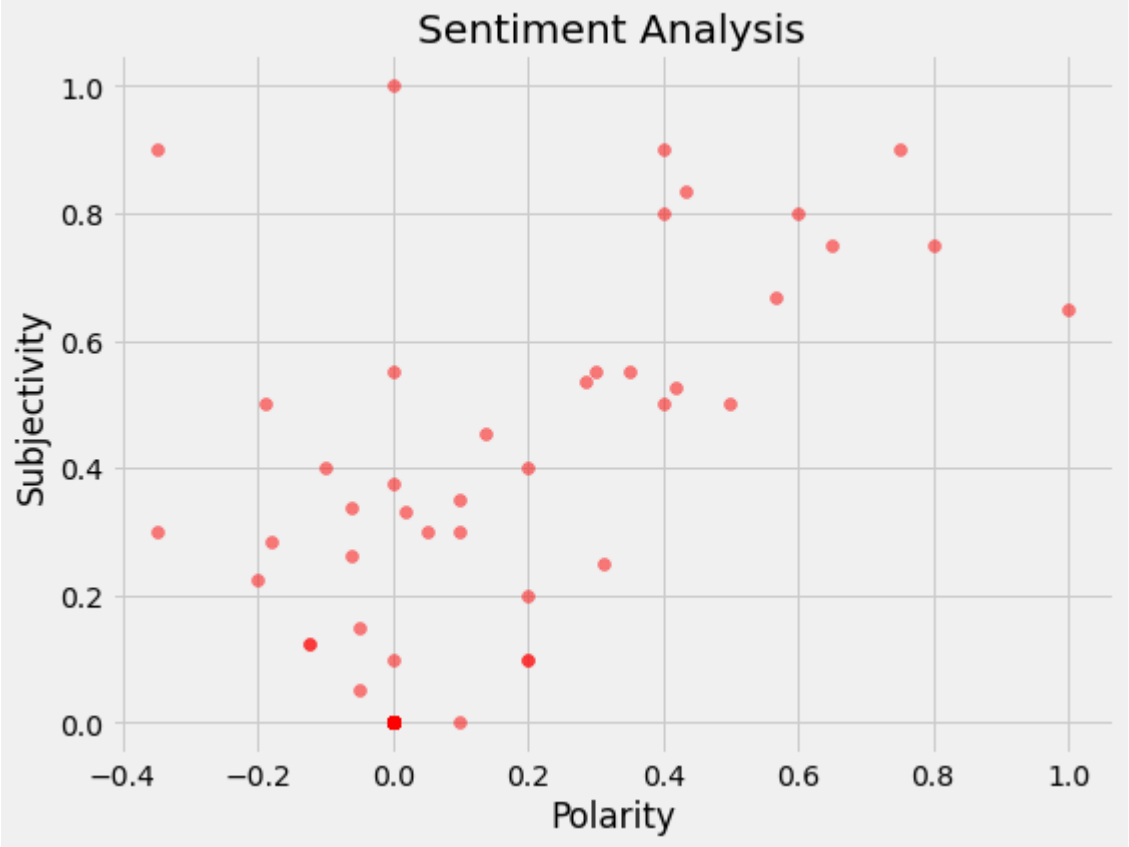
	Tweets	Subjectivity	Polarity	Analysis
0	You’re amazing! 💪	0.900000	0.750000	Booster
1	You’ll always be a champion in the eyes of Mal...	0.650000	1.000000	Booster
2	_nur_iman There seem to be an error with payi...	0.500000	0.500000	Booster
3	Congratulations Hassan!	0.000000	0.000000	Neutral
4	Follow Automotive Enthusiast + Next Top Instag...	0.250000	0.312500	positive
...	...	...	...	...
94	A good gov't needs a strong opposition, says M...	0.666667	0.566667	Booster
95	UM pharmacy programme can't be axed without fu...	0.550000	0.350000	positive
96	Ex-spy chief to be released tomorrow, says she...	0.550000	0.000000	Neutral
97	Rafizi says his bid for PKR No 2 is to safegua...	0.000000	0.000000	Neutral
98	PM limits ministers to 10 aides, Dzulkefly con...	0.000000	0.000000	Neutral

99 rows × 4 columns

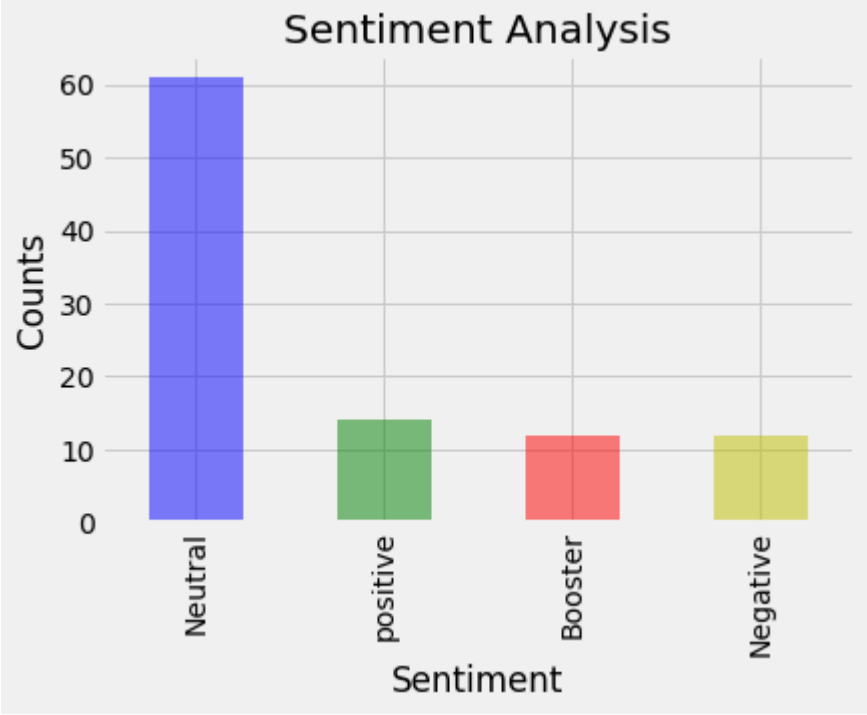
In [40]:
```
# Printing positive tweets
print('Printing positive tweets:\n')
j=1
sortedDF = df.sort_values(by=['Polarity']) #Sort the tweets
for i in range(0, sortedDF.shape[0] ):
  if( sortedDF['Analysis'][i] == 'Positive'):
    print(str(j) + ') '+ sortedDF['Tweets'][i])
    print()
    j= j+1

# Printing negative tweets
print('Printing negative tweets:\n')
j=1
sortedDF = df.sort_values(by=['Polarity'],ascending=False) #Sort the tweets
for i in range(0, sortedDF.shape[0] ):
  if( sortedDF['Analysis'][i] == 'Negative'):
    print(str(j) + ') '+sortedDF['Tweets'][i])
    print()
    j=j+1

# Plotting
plt.figure(figsize=(8,6))
for i in range(0, df.shape[0]):
  plt.scatter(df["Polarity"][i], df["Subjectivity"][i], color='red',alpha=0.5)
# plt.scatter(x,y,color)
plt.title('Sentiment Analysis')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()
```

```
Printing positive tweets:

Printing negative tweets:

1) Low voter turnout due to voters' perception, says EC

2) Azmin: There were a lot of green Birkins, but Halimah failed to see it

3) Sixteen dead, 26 missing in Hokkaido quake

4) International court says it has jurisdiction over alleged crimes against Roh
ingya

5) Maszlee's appointment as IIUM president puzzling and appalling

6) Zuraida: Gov't looking into reducing Forest City foreign ownership

7) Right-wing sites swamp Sweden with 'junk news' in tight election race

8) 'Little Napoleons' abusing your name, Kg Bkt Kuda residents tell PM

9) PM: SST and actual GST collections not much difference to gov't revenue

10) Foreign vehicles with unsettled summonses to be barred from leaving M'sia s
oon
```

```
11) Probes into 1MDB expected to be wrapped up by year end - report

12) Licence for foreign deep sea fishing boat to be abolished
```



In [42]:
```
# Print the percentage of positive tweets
ptweets = df[df.Analysis == 'Positive']
ptweets = ptweets['Tweets']
ptweets

round( (ptweets.shape[0] / df.shape[0]) * 100 , 1)

# Print the percentage of negative tweets
ntweets = df[df.Analysis == 'Negative']
ntweets = ntweets['Tweets']
ntweets

round( (ntweets.shape[0] / df.shape[0]) * 100, 1)

# Show the value counts
df['Analysis'].value_counts()

# Plotting and visualizing the counts
plt.title('Sentiment Analysis')
plt.xlabel('Sentiment')
plt.ylabel('Counts')
df['Analysis'].value_counts().plot(kind = 'bar',alpha=0.5,color=['b','g','r','y']
plt.show()
```



In [ ]: