

PAC 4 - Informe de resultats

En aquest arxiu hi haurà un petit resum dels resultats obtinguts a l'executar el programari desenvolupat per a resolució de la PAC 4 de l'assignatura de Programació de la ciència de dades. Únicament es mostraran els resultats que es demanen explícitament a l'enunciat de la PAC

Exercici 1.2

Resultat: Els 5 primers registres del dataset.

```
{'sentiment': '0', 'id': '1467810369', 'date': 'Mon Apr 06 22:19:45 PDT 2009',  
'query': 'NO_QUERY', 'user': '_TheSpecialOne_', 'text': '@switchfoot  
http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of  
Third Day to do it. ;D"}  
  
{'sentiment': '0', 'id': '1467810672', 'date': 'Mon Apr 06 22:19:49 PDT 2009',  
'query': 'NO_QUERY', 'user': 'scotthamilton', 'text': "is upset that he can't  
update his Facebook by texting it... and might cry as a result School today also.  
Blah!"}  
  
{'sentiment': '0', 'id': '1467810917', 'date': 'Mon Apr 06 22:19:53 PDT 2009',  
'query': 'NO_QUERY', 'user': 'mattycus', 'text': '@Kenichan I dived many times for  
the ball. Managed to save 50% The rest go out of bounds'}  
  
{'sentiment': '0', 'id': '1467811184', 'date': 'Mon Apr 06 22:19:57 PDT 2009',  
'query': 'NO_QUERY', 'user': 'ElleCTF', 'text': 'my whole body feels itchy and  
like its on fire '}  
  
{'sentiment': '0', 'id': '1467811193', 'date': 'Mon Apr 06 22:19:57 PDT 2009',  
'query': 'NO_QUERY', 'user': 'Karoli', 'text': "@nationwideclass no, it's not  
behaving at all. i'm mad. why am i here? because I can't see you all over there.  
"}  
}
```

Exercici 2.2

Resultat: 5 últimes files del data set una vegada s'ha fet una neteja del camp "text".

```
{'sentiment': '4', 'id': '1685554684', 'date': 'Sun May 03 00:55:31 PDT 2009',  
'query': 'NO_QUERY', 'user': 'Heidi_vb', 'text': 'morn waitin 4 sun 2 reappear  
cool havin 2 chase around chillin som coffee jst takin nice n slow haha'}  
  
{'sentiment': '4', 'id': '1685554730', 'date': 'Sun May 03 00:55:31 PDT 2009',  
'query': 'NO_QUERY', 'user': 'KrissyJill', 'text': 'still 3am good grief good  
night'}  
  
{'sentiment': '4', 'id': '1685554798', 'date': 'Sun May 03 00:55:32 PDT 2009',  
'query': 'NO_QUERY', 'user': 'dwmooook', 'text': 'morning world taking parents meal  
'}
```

```
today ahem forgot ruby anniversary last week blame lisa'}

{'sentiment': '4', 'id': '1685554800', 'date': 'Sun May 03 00:55:32 PDT 2009',
 'query': 'NO_QUERY', 'user': 'chastinevegas', 'text': 'drive ins funjust saying'}

{'sentiment': '4', 'id': '1685554911', 'date': 'Sun May 03 00:55:34 PDT 2009',
 'query': 'NO_QUERY', 'user': 'bearyhwang', 'text': 'like twitter page background'}
```

Exercici 3

Resultat: 10 primeres paraules del vocabulari dels tuits.

```
0
00
000
0000
0000000000
0007
001
0010
0015
0019
```

Exercici 4.1

Resultado: Element número 20 del data set una vegada s'ha afegit el diccionari de freqüències.

```
{'sentiment': '0', 'id': '1467813985', 'date': 'Mon Apr 06 22:20:37 PDT 2009',
 'query': 'NO_QUERY', 'user': 'quanvu', 'text': 'alydesigns day didnt get much
done', 'words': {'alydesigns': 1, 'day': 1, 'didnt': 1, 'get': 1, 'much': 1,
 'done': 1}}
```

Anàlisi de dades

Exercici 5.1

Quants clústers tenim en el nostre dataset?

Resultat: Hi ha un total de 2 clústers. Un té "sentiment" = 0 i l'altre "sentiment" = 4.

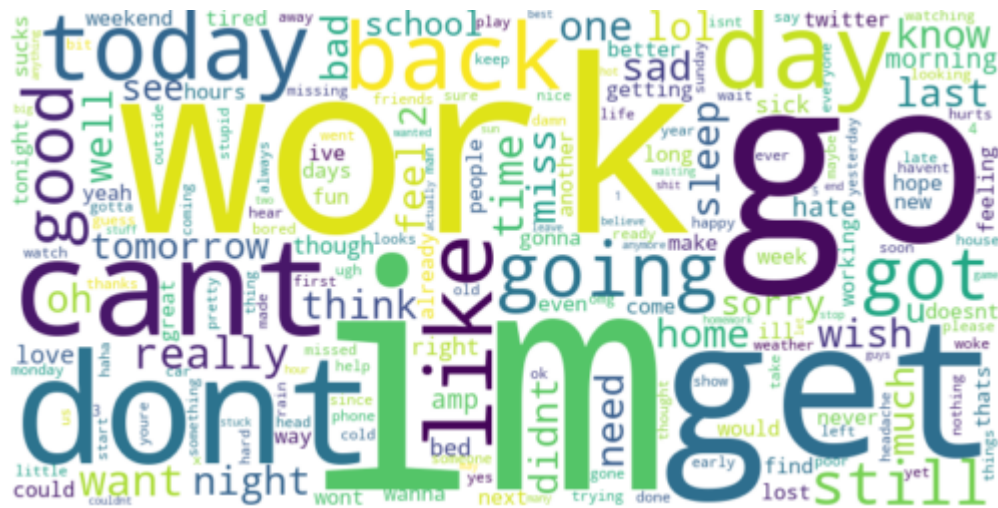
Exercici 5.2

Tenim elements buits en les columnes text? Si és així, quin és el percentatge?

Exercici 5.3

Resultat:

Per "sentiment" 0



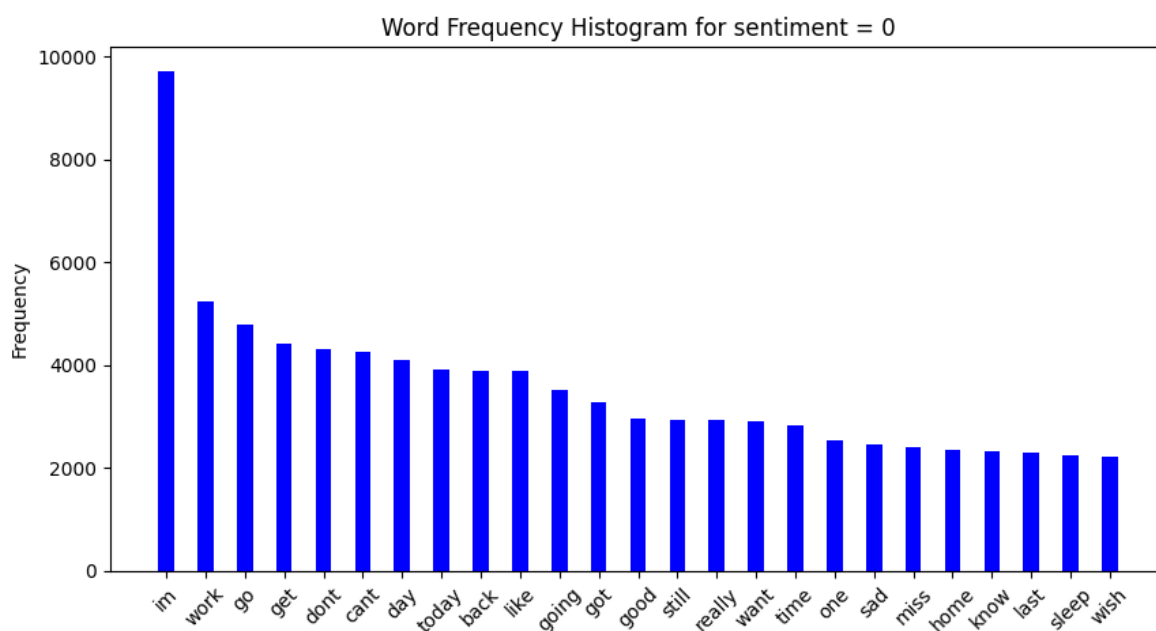
Per "sentiment" 4



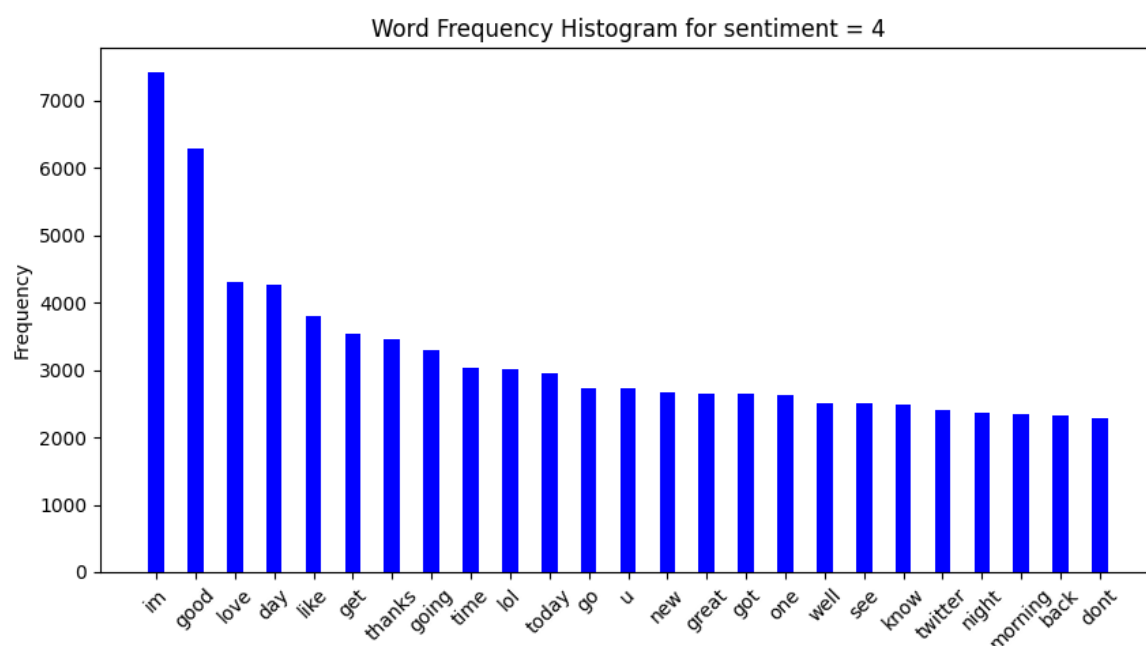
Exercici 6

Resultat:

Per "sentiment" 0



Per "sentiment" 4



Exercici 7

a. Quines són les paraules més utilitzades en les crítiques positives?

Resultat: En les crítiques positives (cluster 4), les paraules més freqüents en les crítiques positives són 'im', 'good', 'love', 'day', 'like', entre d'altres. Aquestes paraules s'acostumen a utilitzar per mostrar emocions positives.

b. Quines són les paraules més utilitzades en les crítiques negatives?

Resultat: Les paraules 'im', 'work', 'go', 'get', 'dont', que sovint tenen connotacions negatives o indiquen obligacions, predominen en les crítiques negatives.

c. Hi ha paraules que apareguin tant en les crítiques positives com en les negatives?

Resultat: Sí, hi ha paraules que apareixen en ambdues crítiques com 'im', 'go', 'get', 'day', 'like', entre d'altres. Aquestes paraules són d'ús freqüent a la llengua i el seu significat pot variar depenent del context en què s'utilitzin.

d. A partir de la word cloud, què es pot deduir sobre el sentiment general de cada grup?

Resultat: El sentiment general a cada grup es reflexa per les paraules més freqüents. Per al grup de crítiques positives, s'utilitzen paraules amb connotacions positives mentre que en el cas de crítiques negatives passa el contrari.