

Ejercicio 5

5.1 – El numero de clusters viene dado por el mismo dataframe (no tenemos que encontrar un numero de clusters previamente), ya que los tweets vienen divididos en dos clusters definidos, sentimiento positivo o negativo.

5.2 – No se observan elementos vacíos. (Lo que si he observado es que hay palabras que se componen solo de números, que según el contexto y un texto solo formado por elementos numéricos que podrían ser considerados para remover.

Ej 7.

Sentiment 0: One, tm, got,day, im, today, work, still, sad, night, need...

Sentiment 4: amp, work, lol, one, still, today, time,love, need, wish...

Vemos que hay ciertas palabras que se repiten, y que tiene cierta lógica que lo hagan, ya que no indican un sentimiento (como today, day etc). Vemos que donde varían son en palabras como "love" y "sad"