Practical Data Science, A2

Report File

C2. By reading the CSV file (which has the format of two columns, one for the title and one for the comments of each video), we get the following:

```
Title \
0
   Ο Τσίπρας ξεχνά τα αγγλικά του μπροστά στον Κλ...
                           Όλα τα αγγλικά του Τσίπρα
1
2
         ΡΑΔΙΟ ΑΡΒΥΛΑ - Obama - Τσίπρας (22/11/2016)
3
                  Τα τέλεια αγγλικά του Mr Tsipras 😜
         Μεϊμαράκης για Τσίπρα "Άντε γαμήσου μαλάκα"
5
  10 Λόγοι που Αγοράζουμε Μάρκες | Greekonomics #38
6 Πρωθυπουργία, ΛΟΑΤΚΙ Γάμος, Εκκλησία και κράτο...
   Η ερώτηση του δημοσιογράφου στον Τσίπρα για τα...
8 Η υποδοχή του Προέδρου των ΗΠΑ στο Μέγαρο Μαξί...
   Χατζηδάκης: Πώς γίνεται να είσαι ελεύθ. επαγγε...
9
10 Απίστευτο δυστύχημα στη Μύκονο: Αγελάδα κάρφωσ...
11 Ο Αλέξης Τσίπρας στο ετήσιο Συνέδριο του Clint...
12 Σήμερα | Ξεπέρασε κάθε όριο η κόντρα Τσίπρα-Μη...
13
     Αγγλικά Τσίπρα εντός κειμένου νε εκτός κειμένου
14
                Στον Ενικό - Αλ. Τσίπρας - 17.9.2015
15
                        Μητσοτάκης vs Τσίπρα στο CNN
16 Εφιάλτης στην Ευρώπη ο κίνδυνος μίας ακροδεξιά...
17 Τα αγγλικά του Τσίπρα "κούφαναν" τους Ευρωπαίο...
18 Βουλή: Πώς ο Τσίπρας ανάγκασε τον Μητσοτάκη να...
19 Ο Αλέξης Τσίπρας ανακοινώνει την παραίτηση του...
                                            Comments
0
   Δεν γίνεται να ξεχάσεις κάτι που δεν ξέρεις\ηΔ...
1
   Σε δύο μέρες δίνω proficiency και νιώθω έτσι α...
  Αχαχαχαχαχα! Ίδιος! Όσο θυμάμαι τον Τσίπρα που...
  Γελάω τρελά με την αντίδραση του Κλίντον!\nTsi...
  μεγάλες στιγμές σύγχρονου ελληνικού κοινοβουλε...
5
  Εξαιρετικά υψηλής ποιότητας περιεχόμενο, στολι...
  Κλαίω που σε όλες αυτού του τύπου τις συνεντεύ...
7
   Σ' έχουν πάρει στο ψητό όλος ο πλανήτης Αλέξη....
9
                                                 NaN
10
                                                 NaN
11
                                                 NaN
12
13 Είπε τις πυρηνικές πηγές αντί για ανανεώσιμες ...
14 Πολύ σωστή η ερώτηση του Γιώργου...\ηΚαθόλου ο...
15 Η διαφορά μεταξύ ενός Αμόρφωτου και Γελοίου Αν...
17
   Ο Γαρδελης ,ο Ψάλτης και ο Πέτροχείλος στις τ...
18
                                                 NaN
                                                 NaN
[0, 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 17]
```

where the last array corresponds to the non-na values. Note that the Na values occurred because either there were no comments in the video or the comment section of the video had been removed.

Decision Tree classifier made the following predictions about the language comments:

languages_0: ['Greek', 'Greek', 'Greek'

languages_1: ['Greek', 'Greek', 'Greek'

languages_2: ['Greek', 'Greek', 'Greek'

languages_3: ['Greek', 'Greek', 'Greek'

languages_4: ['English', 'English', 'English

languages_5: ['Greek', 'Greek', 'Greek'

languages_6: ['Greek', 'Greek', 'Greek'

languages_7: ['English', 'English', 'English

languages_8: []

languages_9: []

languages_10: []

languages_11: []

languages_12: []

languages_13: ['Greek', 'Greek', 'Greek

languages_14: ['English', 'English', 'English', 'English', 'English', 'English', 'English']

languages_15: ['Greek', 'Greek', 'Greek

languages_16: []

languages_17: ['Greek', 'Greek', 'Greek

languages_18: []

languages_19: []

A good idea would be to make a bar chart to see the total number of comments for each video.

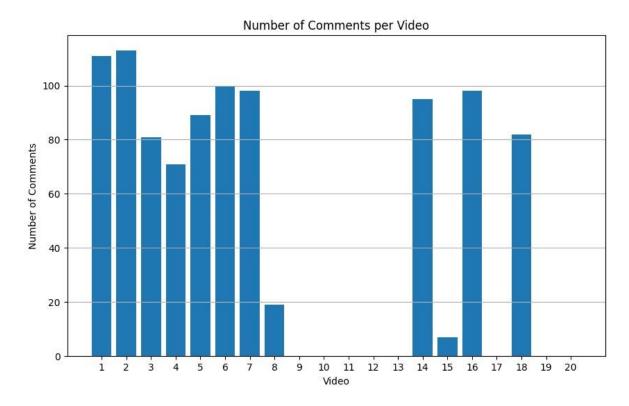


Figure 1: Exploring the number of comments per video.

So we see that most videos have more than 80 comments. Also, we verify here that there are some videos that have no comments (or no comment section, as stated earlier).

The language distribution across all the comments can be seen by this pie chart:

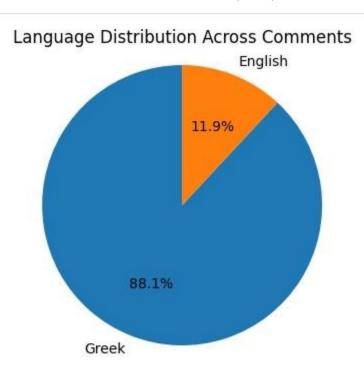


Figure 2: Exploring the variety of languages across all comments.

Suprisingly enough, there were found English comments in our videos. In fact, this is mainly a mistake by the classifier, as it should be classified as Greeklish. As expected though, the vast majority of comments is in Greek.

This pie chart could be made also one for each video, although in our case, because all comments had the same language, there would be no point to do so (all pies would have only one language).

a. The most toxic language is found in these comments:

Καλα μεταξυ μας και εγω με lower ετσι οπως μιλάει ο Κλίντον με την γερικη πλέον φωνή του χρειάστηκε να τ ο ακούσω πολλές φορες και με αρκετή συγκέντρωση για να καταλάβω τι λεει.. Ενα μουρμουρητο ακουγα μον ο στην αρχή.

Αυτό το Playmobil μας κυβερνούσε επί 4,5 χρόνια

Ελα ρε τσιπρααα δεν ξερεις αγγλικα? Ειναι διεθνης γλωσσα συννενοησης. Δεν μπορεις να μην ξερεις αγγλικα και να λες οτι εχεις εστω και μια βασικη ενημερωση για τον συγχρονο κοσμο. Αλλα και ετσι παρε τουλαχιστο μεταφραστη. Ο μπιλ καλα το παιζει. Σου λεει αν ερχεται εδω η θα με πιανει οταν του μιλαω γρηγορα η θα δειξ ω οτι δεν ξερει. Δεν γινεται να μην ξερεις αγγλικα σαν προθυπουργος. Και ο πουτιν που οπου παει μιλαει μον ο στα ρωσικα ΞΕΡΕΙ μια χαρα αγγλικα. Απλα δεν μιλα για προφανεις λογους.

Δεν μπορώ να καταλάβω το εξής: όλοι οι υπόλοιποι πολιτικοί των άλλων χωρών μιλούν πάντα στη δική τους γλώσσα και αυτό το βλέπουμε από τους Ρωσσους, τους Γαλλους μεχρι και τους Ρουμανους. Γιατί λοιπόν ο κύ ριος Τσιπρας ,γνωρίζοντας οτι δεν το εχει με τα αγγλικα, δεν μιλάει την δική μας γλωσσα; Δεν είναι οτι εχουμ ε καμια ασήμαντη γλωσσα ισα ακριβως το αντιθετο! Χαρη στη ελληνική γλώσσα ο δυτικός πολιτισμός οφ ειλει μεγαλο μερος της ύπαρξής του.

Αυτό τον τύπο τον κάνανε κάποιοι ...Πρωθυπουργό..!

Αν καταλαβα καλα, ο Τσιπρας δεν καταλαβε που ηταν το αστειο στην δηλωση Κλιντον, οχι οτι δεν καταλαβε τι ειπε.

Πόσο διήρκεσε αυτό το ρεζιλίκι; θέλω να το απολαύσω από την αρχή! Υπάρχει κάπου ολόκληρο;

ΚΑΙ όμως Τον ψηφίσατε δεύτερη φορά Αυτό φτάνει πού βλέπεις Έναν βλάκα χωρίς λέξεις δε ψηφίζεις Έναν ηλίθιο

ο Μειμαράκης θα μιλούσε πιο καλα αγγλικά ...όπως μιλά στα ελληνικά του!

Έφτασε η βενζίνη 2 ευρώ το λίτρο δεν πειράζει όμως ο Μητσοτάκης μιλάει καλα αγγλικά 41%

Καλα τον κοροϊδεύετε, αλλα μην πατε και τον ψηφίζετε μετά. Και μη μου λετε ολοι όχι δε ψηφίζουμε, που πήγε το 24% χάθηκε σε μια μερα?

Αυτό το άτομο κυβερνούσε κράτος μαντάμ Μερκελ

ΑΓΑΠΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ ΜΕΓΑΛΟΣ ΗΓΕΤΗΣ !!!!!!!!!!!!!!!!!!!!!!!!!!!! ΚΑΛΑ ΜΕ ΚΑΝΕΙΣ ΚΑΙ ΓΕΛΑΩ!!!!!!!

Κι ούτε έκανε προσπάθεια να διευκολύνει το συνομιλητή του, γι 'αυτό ο Τσίπρας έπρεπε να του πει "Μόνικα Λεβίνσκι σπερμ".

Αυτό και αν είναι

Να είσαι άνθρωπος αυτό έχει σημασία και να λες την αλήθεια αυτό μετράει ας την λες και με χάλια αγγλικά.Τ ι με νοιάζει το συντακτικό αν μου κοροϊδεύουν;

Αυτό το πράγμα έγινε πρωθυπουργός χώρας????????? Tragic.

Ο άνθρωπος δεν καταλαβαίνει την διαφορά μεταξύ Show και so ..αυτό τα λέει όλα..

κανακη ιδιος ομπαμα . σε εβαψαν πολυ καλα!!!

Γιάννη αυτό θα πει πρωθυπουργός.

Τα αγγλικά δεν κάνουν τον ηγέτη. Χαρακτηριστικά παραδειγματα: η Μέρκελ, οι Γάλλοι πρόεδροι, ο Ερντογάν και πολλοί άλλοι που βαριέμαι να ψάξω. Γι'αυτό έχουν τους διερμηνείς τους έτσι κι αλλιώς. Αυτό που εντυπω σιαζομαστε όταν ο πρωθυπουργός μας ξέρει άπταιστα αγγλικά, πότε δεν το κατάλαβα

Αυτό το καθίκι από 25/01/2015 μέχρι 28/08/2015 κι από 20/09/2015 μέχρι 07/07/2019 ήταν πρωθυπουρ γός της Ελλάδας!!!

ΤΟ ΕΝΑ Π...... ΚΟΙΤΑΕΙ ΤΟ ΑΛΛΟ!!!!!! ΑΛΛΑ ΑΣ ΜΗΝ ΤΟΝ ΑΔΙΚΟΥΜΕ ΤΟ GO HOME ΤΟ ΛΕΕΙ ΚΑΛΑ!!!!!! ΤΟΣΟ ΚΑΛΑ ΠΟΥ ΤΟ ΕΛΕΓΕ ΓΙΑ ΤΟΝ ΕΑΥΤΟΝ ΤΟΥ!!!!!!!! ΠΗΓΕ ΣΠΙΤΙ ΤΟΥ ΓΙΑ ΤΑ ΚΑΛΑ!!!!!!!! ΜΕ ΦΟΡΑ ΠΟΥ ΕΛΕΓΕ ΚΑΙ Η ΠΟΠΑΡΑ!!!!

Εμένα δε με ενδιαφέρει να μιλάει τέλεια τα.αγγλικα αλλά να μιλάει σωστά την ελληνική γλώσσα και σ; αυτό ή ταν τέλειος!!

Και αυτό το σουργελο κάποιοι τον ψηφίζουν... Αλλά αυτοί είμαστε...

Αυτό είναι το ήθος και ο ορθός Ευγενικός λόγος τού τότε προεδρεύοντος της Βουλής

ΚΑΙ ΠΟΛΎ ΚΑΛΑ ΤΟΥ ΕΙΠΕ!

Και πως θα το κάνουμε αυτό; :-)

καλα του ειπε

Νασαι καλα ..

Ειλικρινά Ό,ΤΙ καλύτερο έχω δει σε βίντεό σου!! Η πληροφορία και η γνώση που μας έδωσες σε αυτό το βίντε ο ήταν πραγματικά απίστευτη, ευχαριστούμε πολύ!!

Μπράβο!!! Με έκανες να νιώσω τύψεις που έχω πιαστεί αρκετές φορές από λίγο έως πολύ κορόιδο των πολυ εθνικών και των διαφημιστών. Αυτό όμως σημαίνει μάθηση. Μέσω της γνώσης την επόμενη φορά θα είμαι πι ο υποψιασμένος και θα είναι δυσκολότερο να με εκμεταλλευτεί ένας έμπορος.

Όλο αυτό μου θύμησε τον μπάρπα που συνάντησε στο φανάρι μια Ferrari και ρώτησε τον τυπά που την οδηγούσε:

Ευχαριστούμε πολύ για αυτό το βίντεο!! Είναι κάτι που τα τελευταία χρόνια το βλέπουμε παντού! Και εγώ τ ο κάνω και σε ευχαριστώ πολύ γιατί με βοήθησε να αλλάξω γνώμη

Καλησπέρα Κοσμά και σε ευχαριστώ για αυτό το βίντεο

Για άλλη μια φορά είδαμε ένα κορυφαίο βίντεο που συνδέει όμορφα τους λιγότερο ή περισσότερο γνωστούς λόγους που συμβαίνει αυτό. Η αλήθεια βέβαια είναι οτι κάποιες φορές αισθάνεσαι ότι κάποιες εταιρίες βγάζο υν προϊόντα που αν τα σχεδίαζες εσύ, θα τα έβγαζες έτσι ή περίπου έτσι. Ταιριάζουν δηλαδή απόλυτα στα κριτήρια σου και τα γούστα σου και αυτός κατά τη γνώμη μου είναι ο πιο "ορθολογικός" παράγοντας που οδηγεί στο να πληρώσουμε κάτι παραπάνω ή ακόμη και στο φανατισμό μερικούς που δεν έχουν κάτι καλύτερο να κάνουν.

Γεια σου Κοσμά, ωραίο βίντεο. Ήθελα να πω τη γνώμη μου για τις μάρκες που βάζουν το σήμα τους "ευδιάκρ ιτα" επάνω στα προϊόντα τους, αλλά θα εστιαστώ στο "μέλος μιας μεγάλης ομάδας" και να μιλήσω για την Εθ νική Γαλλίας που έριξε 14 στο Γιβραλτάρ. Για μένα αυτό είναι σύνδρομο ανωτερότητας, δηλαδή βρήκες ρε φί λε στο Γιβραλτάρ να ρίξεις 14; Θα μπορούσαν να πάρουν σαν παράδειγμα τον ΠΑΟΚ που μετά τα ντόρτια πο υ έριξε στο Γαύρο έκανε fair-play. Πόσα να ρίξεις στο Γαύρο δηλαδή; Από ένα σημείο και μετά δεν έχει πλέον γούστο. Φιλάκια απ' τα ΠΑΟΚια του Σικάγου

Αυτό που μου άρεσε περισσότερο σε αυτό το Κρίμα είναι το γεγονός ότι έγινε συζήτηση απλή στο να γνωρίσ ουμε ένα τόσο δημόσιο πρόσωπο με τις απόψεις του και όχι οι φανφάρες από εδώ και από κει. Με πέθανε το αστειάκι με το "Αλλά αυτοί είστε" κανονικά

Ζωή συγχαρητήρια! Ανεξάρτητα με την πολιτική θέση του καθένα είναι αδιαμφισβήτητο πως η συγκεκριμέν η συζήτηση έγινε πολύ σωστά συγκριτικά με εκείνη στο κανάλι της Μεγκ, η οποία τράβλιζε και κόμπλαρε αφ άνταστα με αποτέλεσμα η συνέντευξη να καταλήξει άκρως πληκτική. Οι ερωτήσεις είναι στοχευμένες έτσι ώ στε να γνωρίσει κανείς την προσωπικότητα ενός ανθρώπου με εξουσία, το οποίο είναι σαφώς σημαντικό εν όψη των εκλογών. Η σημαντικότερη διαφορά με την συνέντευξη του Μητσοτάκη είναι το αίσθημα πως οι συ νομιλητές είναι ίσοι. Αυτό και μόνο χαίρει αναγνώρισης, ανεξάρτητα απο τα πολιτικά πιστεύω. Άμεση, ανθρώ πινη συνέντευξη χωρίς μπουγάδες

Εγώ αυτό καταλαβαίνω.

Συγχαρητήρια Ζωή για τις ερωτήσεις και συγχαρητήρια που φιλοξένησες αυτό τον άνθρωπο

Εξαιρετική συνέντευξη - συζήτηση. Εμπεριστατωμένος λόγος. Όποιος δεν βλέπει διαφορές με τον Μητσοτάκ η, ας βγάλει τα γαλάζια γυαλιά και ας ξαναπροσπαθήσει. Μπράβο σας που αφήσατε ανοιχτά τα σχόλια. Αυτό είναι θάρρος και πολιτικός πολιτισμός.

Χαίρομαι πραγματικά που YouTubers φέρνουν στο κανάλι τους υποψήφιους για τις εκλογές. Το λέω αυτό γι α έναν και μόνο λόγο. Για να μάθουν οι νέοι και οι νέες ποιο κόμμα να ψηφίσουν. Το μέλλον της Ελλάδας είμ αστε εμείς, οι νέοι. Πρέπει όλοι να ψηφίσουμε.

Δεν μιλάω προσωπικά για τον κύριο Α.Τσίπρα , το λέω γενικά , εννοείτε ο καθένας ψηφίζει αυτό που πιστεύε $\iota!$

Και μετά από ένα λεπτό τι έγινε; ;;;;;; Αυτό το επεισόδιο θέλω να δω!!!!!!!

Εγώ όταν βλέπω αυτό βίντεο δε γελάω. Αντιλαμβανόμαστε ότι υπήρξε πρωθυπουργός που διαπραγμευτηκε τη ζωή μας για εκατό χρόνια και δεν ήξερε λέξη αγγλικά.

Μια ώρα να το ακούω αυτό υπογράφω τα πάντα

Το μεγαλύτερο βασανιστήριο της ζωής μου που κατάφερα να δω όλο αυτό που είδα μόλις τώρα σ'αυτό το βί ντεο.

καλα, σιγουρα του τα εχουν γραψει με ελληνικους χαρακτηρες τα αγγλικα. τι να λεμε. αφθονο γελιο Αυτά παθαίνει κανείς όταν βλέπει το σχολείο μόνο απ' έξω. Οπότε προτείνω να δείχνουμε αυτό το βίντεο στα παιδιά μας για να μην κάνουν κοπάνα από το σχολείο!!!

Αυτό το παραμύθι κυβέρνησε την Ελλάδα

Και αυτός αντιπροσωπεύει ένα Κράτος ;;;; !!!! Η μήπως αντιπροσωπεύει - Οχλο ;;;;

Αφήνω αυτό εδώ. Rip IQ

Ο ΖΗΚΟΣ ΠΙΟ ΚΑΛΑ ΤΑ ΜΙΛΑΓΕ !!!

Παρα πολλοί άνθρωποι δυσκολεύονται να συλλάβουν μια ξένη γλώσσα. Δεν φταίνε οι ίδιοι για αυτό αλλά η λ ειτουργία του εγκεφάλου τους.

Φημες λενε οτι βλεπει πολλες ξενες ταινιες με υπότιτλους σε slow motion οποτε καλα τα παει

1:11:00 θα σας έδινα τεράστιο δίκιο κύριε Τσίπρα, αλλά όλα αυτά, πριν γίνετε Πρωθυπουργός, πριν ζητήσετε την ψήφο και την εμπιστοσύνη του ελληνικού λαού, δεν τα γνωρίζατε; Δείχνατε με το δάχτυλο ΠΑΣΟΚ και Ν Δ για τα μνημόνια, ότι είπαν σε όλα ναι αβίαστα και δεν διαπραγματεύτηκαν. Δεν γνωρίζατε τι συνέπειες θα ε ίχε το όχι; Αυτό μεταξύ άλλων σας καταλόγισε ο λαός και δεν σας ξαναψήφισε.

Και όμως αυτό το βλήμα, το ψήφισαν κάποιοι και τον πιστεύουν κιόλας

Καλά τώρα εσείς που είστε η ομάδα αλήθειας θέλετε να λέγεστε και έτσι τι θέλετε να μας πείτε δηλαδή μαύτ ο το βίντεο που ο Τσίπρας λες και μιλάει αγγλικά σα να έχει βγάλει μόνο ένα χρόνο στο φροντιστήριο και ο ά λλος ο Κούλης φυσικά και όπως είναι φανερό να μιλάει σε επίπεδο lower proficiency θα ήταν το πιο σωστό ν α πω . Αλλά πείτε μας αυτό τι σημαίνει αυτή η αντιπαραβολή στα παράθυρα που ο ένας δε ξέρει τη τυφλά το υ από αγγλικά και ο άλλος τα μιλάει σχεδόν άριστα . Ε και Ομάδα Αλήθειας; Τι θέλετε να πείτε ποιητές ;; Ο Καμμένος τι έχει να πειβλέποντας Αυτό?

Τραγικά τα αγγλικά του αλλά ανεξάρτητα από αυτό δεν κρίνεις ένα πρωθυπουργό από τα αγγλικά,όχι ότι ήτ αν καλός σε άλλους τομείς αλλά λέμε τωρα...

Σε αυτό το επίπεδο, ο Κούλης είναι τελευταίος, τον ξεπερνάει ακόμη και το ΚΚΕ.

ενα τετοίο βιντεακι καθε βραδυ και ουτε ελληνική ταινία ουτε τίποτα ...α ρε τσίπρα να είσαι καλα με τις μπο υρδες σου !

Σαν Πενταετία της ντροπής πρέπει να γραφτεί στα βιβλία της ιστορίας αυτό που ζήσαμε.

Αυτό ακριβώς... τίποτα άλλο

Καλα, δε ντρέπεται?

Note that all these comments have been classified as number 4 in the toxicity scale. The LLM did not classify any comment as number 5 in our training exaples, so our classifier did not mark any comment as number 5 in the toxicity scale.

Now we find in which videos these comments belong to:

b. Then we calculate the toxicity rates for each video, to find the video with the most toxic language.

Video 0: Toxicity Rate = 0.08

Video 1: Toxicity Rate = 0.08

Video 2: Toxicity Rate = 0.02

Video 3: Toxicity Rate = 0.07

Video 4: Toxicity Rate = 0.06

Video 5: Toxicity Rate = 0.07

Video 6: Toxicity Rate = 0.07

Video 7: Toxicity Rate = 0.05

Video 8: Toxicity Rate = 0.00

Video 9: Toxicity Rate = 0.00

Video 10: Toxicity Rate = 0.00

Video 11: Toxicity Rate = 0.00

Video 12: Toxicity Rate = 0.00

```
Video 13: Toxicity Rate = 0.12
Video 14: Toxicity Rate = 0.14
Video 15: Toxicity Rate = 0.05
Video 16: Toxicity Rate = 0.00
Video 17: Toxicity Rate = 0.05
Video 18: Toxicity Rate = 0.00
Video 19: Toxicity Rate = 0.00
```

The most toxic video was the video 14 with title Στον Ενικό - Αλ. Τσίπρας - 17.9.2015 and toxicity rate of 14. 29 %.

c. We will check now in which video (if exists) the toxicity is uniform over time. By the term uniform, we mean that all comments within the video should have received the same toxicity rating (the videos with no comments have been excluded).

Indices of videos where toxicity is uniform: [] The video titles are:

We get that no video has been characterized as uniform, and it does not come as a surprise, b ecause most videos have over 80 comments. In other checks we made, for videos with smaller amount of comments, we managed to found videos that had uniform toxicity.

d. Finally, we will find the videos where toxicity increases over time. We can implement this by che cking every comment to have the same or higher toxicity rating than the previous comment. Thi s is also a difficult case to catch when there is a big number of comments.

Indices of videos where toxicity increases over time: [] The video titles are:

Simirarly, we find that no videos satisfy this criteria for our case.