*wir haben gesagt, dass ich anhand von den available Genres die Videos zum scrapen auswählen.  
und das es wichtig ist das vorallem die Microgenre gut representiert werden.  
  
Problem, viele Microgenres haben nicht viele Videos.  
  
k=1: 685 columns have at least 1 non-zero entries  
k=2: 681 columns have at least 2 non-zero entries  
k=3: 610 columns have at least 3 non-zero entries  
k=4: 565 columns have at least 4 non-zero entries  
k=5: 534 columns have at least 5 non-zero entries  
k=6: 516 columns have at least 6 non-zero entries  
k=7: 503 columns have at least 7 non-zero entries  
k=8: 486 columns have at least 8 non-zero entries  
k=9: 466 columns have at least 9 non-zero entries  
k=10: 450 columns have at least 10 non-zero entries  
k=11: 442 columns have at least 11 non-zero entries  
k=12: 427 columns have at least 12 non-zero entries  
k=13: 412 columns have at least 13 non-zero entries  
k=14: 408 columns have at least 14 non-zero entries  
k=15: 398 columns have at least 15 non-zero entries  
  
Frage 1: Wie soll ich k festlegen, damit ich schlüsse auf das gesamte Genre schließen kann? Was tu ich mit mit den restlichen Genres, (rausfiltern?)*

**Antwort:  
k =??? bis ca 200 genre**

**wieivie k sind möglich**

**artist/genre berechnen**

*Frage 2: Wieviele Kommentare soll ich pro Video scrapen("neuesten")?*

**50-100 Kommentare -timespampts**  
*Frage 3: Zeitraum von 2023-2025, dadurch sind oft nicht genügend Kommentare in dem Zeitraum verfügbar:  
Wie soll ich das am besten angehen?  
Was tun wenn keine Kommentare verfügbar sind?*

**Algorithm gut dokumentieren und schicken**

**Statistiken berechnungen !!!**

**Nicht augmentieren!!!**

**Einfach in statistic mitaufnehmen!!!**

**Relevant oder neusten Kommentare: mittendrinnen checken was sinnvoll**

**Timestamp mitscrapen!!!**

**Server: documentation hier hochladen**

[**https://cloud.cp.jku.at/index.php/s/ArNjsN7W7psGdbY**](https://cloud.cp.jku.at/index.php/s/ArNjsN7W7psGdbY)