

Maren Abele

20/06/2022

TED Talk Explorer - Report

TED Talk Explorer - Report



In diesem Report werden die Ergebnisse der Datenanalyse mit Grafiken detaillierter dargestellt und erläutert.

1. Wissenswertes über den Datensatz

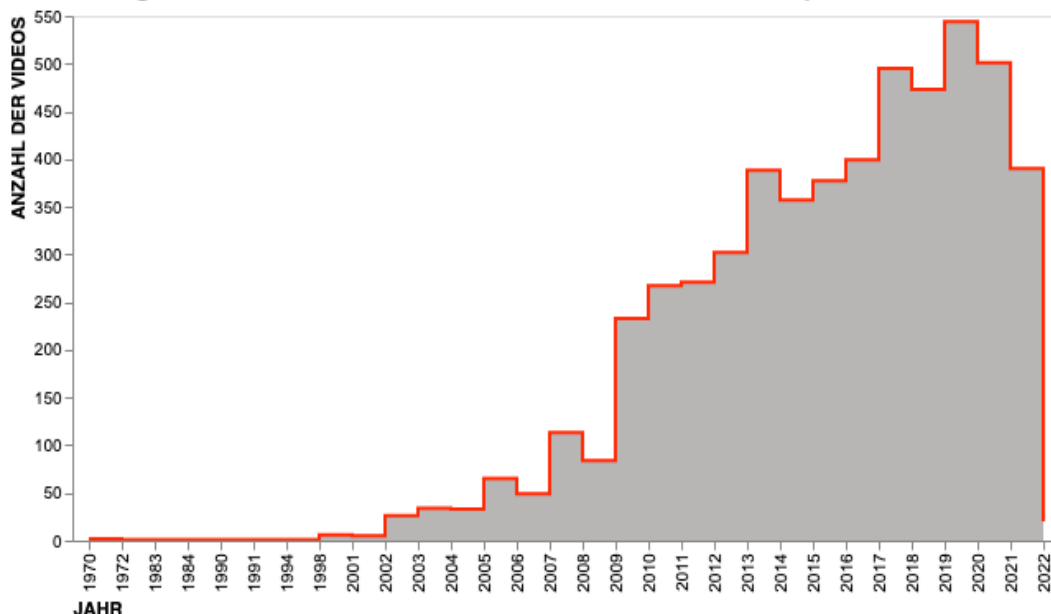
TED steht für Technology, Entertainment und Design. Der Name geht auf die Anfangszeiten zurück, als TED noch eine Konferenz für die im Namen vorkommenden Themen war. Die erste TED Konferenz fand im Jahr 1984 statt. Später wurden die Vorträge beziehungsweise kurzen Talks (Vorträge) auf der Webseite kostenlos zur Verfügung gestellt. Heute findet man eine Vielzahl an Vorträgen zu vielen Themengebieten. TED Talks finden nicht nur zu den ursprünglichen Themen statt. Außerdem gibt es die Vorträge inzwischen in 110 Sprachen.

Für diesen Report wurden Daten aus einem Kaggle Datensatz verwendet. Der Ersteller des Datensatzes heißt Ashish Jangra und ist ein Entrepreneur und KI Trainer in Indien. In der Beschreibung des Datensatzes steht zudem, wie Herr Jangra die Daten zusammengestellt hat. Er hat dafür die Webseite von TED gecrawlt und sich die Daten gezogen, die er in den Datensatz packen wollte. Der Datensatz besteht aus den Variablen: Titel, Autor, Datum, Views, Likes und Links. Auf Grundlage dieser Variablen und Daten wurden Untersuchungen vorgenommen.

2. Verteilung der Videos

Der Datensatz besteht aus insgesamt 5440 Videos.

Verteilung aller veröffentlichten TED Talks über den kompletten Zeitraum



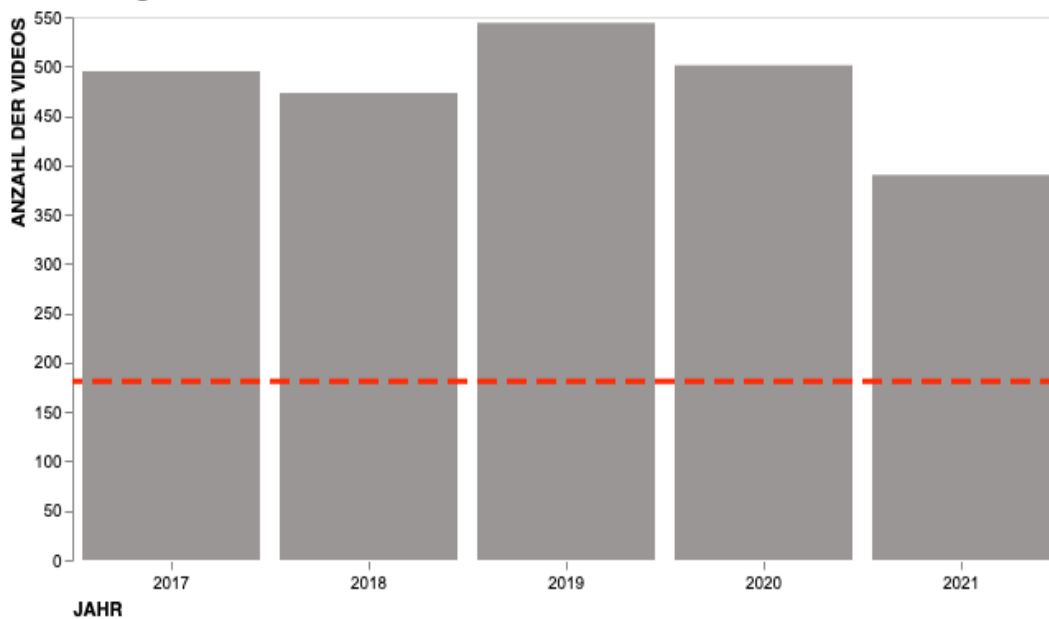
Diese Videos wurde über die Jahre 1970 bis 2022 veröffentlicht. Wie die Grafik zeigt, sind es in den Jahren ab 1970 bis ca. 2022 nur vereinzelte Videos. Ab 2002 sieht man einen deutlichen Anstieg an Videos pro Jahr. Was sich natürlich durch das Internet und die steigende Bekanntheit über die USA heraus erklären lässt. Die erste TED Konferenz fand 1984 statt, deswegen ist zu hinterfragen, was es mit den Videos vor diesem Jahr auf sich hat. Sind diese Talks im nachhinein auf die Webseite gestellt worden? Evtl. gibt es noch ein Video aus dieser Zeit, das zu der Idee der TED Konferenz geführt hat. Diese Videos könnten auch Fehler im System aufweisen und falsch datiert worden sein.

Das erste Video des Datensatzes ist auf Januar 1970 und das letzte Video des Datensatzes auf Februar 2022 datiert.

Im Jahr 2019 wurden 544 Videos veröffentlicht. Hier sieht man auch in der Grafik den höchsten Punkt. Gefolgt von den Jahren 2020 mit 501 und 2017 mit 495 Videos.

Dies zeigt, dass die meisten Videos in den letzten Jahren veröffentlicht wurden. Deswegen werden nun die Veröffentlichungen der letzten 5 Jahre aufgeschlüsselt.

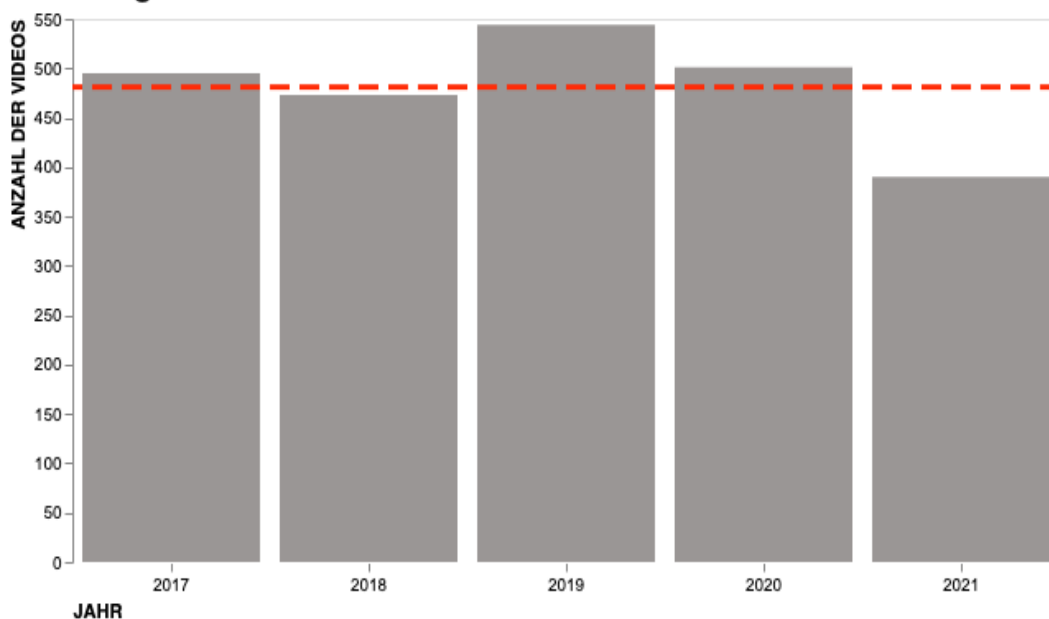
Verteilung aller veröffentlichten TED Talks der letzten 5 Jahre



Die rote Linie zeigt den Durchschnitt der Videos über den kompletten Datensatz. Zieht man alle Jahre heran, wurden durchschnittlich 181 Videos veröffentlicht. Dieser Durchschnitt wurde in den letzten 5 Jahren immer deutlich übertroffen. Dieser Fakt ist nicht weiter verwunderlich, wenn drei dieser 5 Jahre zu den Jahren mit den meisten veröffentlichten Videos gehört.

Deswegen schauen wir uns diese 5 Jahre noch mal etwas genauer an. Zieht man nämlich den durchschnitt dieser 5 Jahre heran, steigt dieser auf 481 Videos pro Jahr. So fällt auf, dass die drei Top Jahre über dem Durchschnitt liegen. 2018 ist etwas schwächer, kratzt aber noch an dem Durchschnitt. Es fällt jedoch auf, dass 2021 deutlich schwächer ist, als die Jahre davor. Es wurden weniger Talks veröffentlicht. Hierfür könnte ein möglicher Grund die Corona-Pandemie sein. Es mussten Maßnahmen eingehalten werden und es durften keine Großveranstaltungen stattfinden. Meistens werden TED Talks vor einem Live-Publikum aufgenommen. Deswegen könnte es hier zu einem Abfall der Veröffentlichungen gekommen sein.

Verteilung aller veröffentlichten TED Talks der letzten 5 Jahre



3. Likes und Views

Zwei weitere Variablen zum detaillierten Analysieren der Talks sind die Likes und Views der Videos. Zieht man alle 5440 des Datensatzes heran ergibt es eine durchschnittliche Like-Anzahl von 62.608 Likes pro Video.

Die Top 3 Videos mit den häufigsten Likes sind "Do schools kill creativity?" von Sir Ken Robinson, "Your body language may shape who you are" von Amy Cuddy und "Inside the mind of a master procrastinator" von Tim Urban. Diese Videos liegen mit 2.100.000, 1.900.000 und 1.800.000 weit über den durchschnittlichen Werten der Likes.

Die schlechtesten Videos gemessen an den Likes sind "Year In Ideas 2015" von unbekannt, "Post-Pandemic Paradise in Rapa Nui" von Far Flung und "Virtual Worlds" von Far Flung. Mit 15, 37 und 39 Likes sind diese Videos weit vom Durchschnitt entfernt.

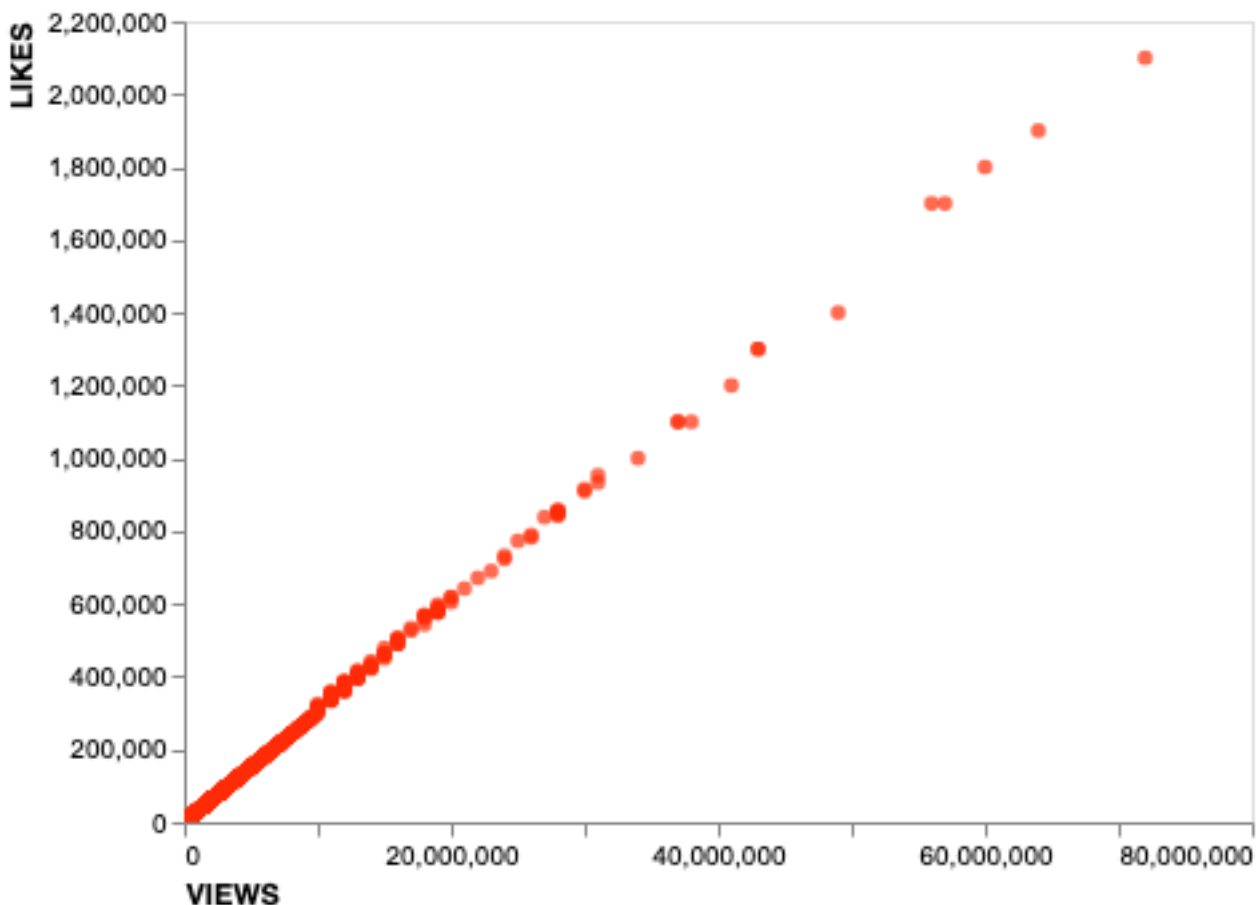
Die durchschnittliche View-Anzahl liegt bei 2.061.576 Aufrufen pro Video. Auch hier werden wieder die Top und die Flop Videos angeschaut. Mit 72.000.000, 64.000.000 und 60.000.000 Aufrufen sind es genau die gleichen Top 3 Videos gemessen an den Likes. Die schlechtesten 3 Videos gemessen an den Likes haben ebenfalls die schlechtesten Aufruf-Zahlen. Mit 532, 1200 und 1300 Aufrufen liegen diese Videos auch weit unter diesem Durchschnitt.

Aufgrund dieser Erkenntnisse kann man annehmen, dass es eine Korrelation zwischen Likes und Views gibt.

Korrelation

Errechnet man den Korrelationskoeffizienten zwischen Views und Likes liegt dieser bei 0.99. Der Korrelationskoeffizient ist eine Einheit für den Grad eines linearen Zusammenhangs. Bei +1 besteht ein perfekter linearer Zusammenhang zwischen den zwei Variablen. Mit 0.99 ist der Zusammenhang zwischen den Views und Likes nahezu perfekt. Dies lässt sich auch gut in der Grafik erkennen.

Korrelation zwischen Views und Likes der TED Talks



Eine gute Ergänzung zum Datensatz wäre eine Like-Quote, die das Verhältnis zwischen Views und Likes direkt angibt. So muss man nicht erst selbst ausrechnen, wie vielen Zuschauern das Video gefallen hat. Das beste Verhältnis sind 36.4%, das schlechteste sind 30.4%. Dies zeigt, dass rund einem Drittel der Zuschauer das Video auch gefällt.

4. Keywords

Da es diese Vielzahl an Videos gibt, die auch aus verschiedensten Themengebieten stammen, wurden die Titel genauer analysiert. Die Titel bestehen aus insgesamt 41.645 Wörtern. Wenn man die gleichen Wörter zählt, kommt man auf 7.160. In diesen Wörtern sind die Top Wörter, solche wie "the", "of", "how", "to". Diese Wörter sagen allerdings nichts über den Inhalt des Videos aus. Diese Wörter werden aussortiert.



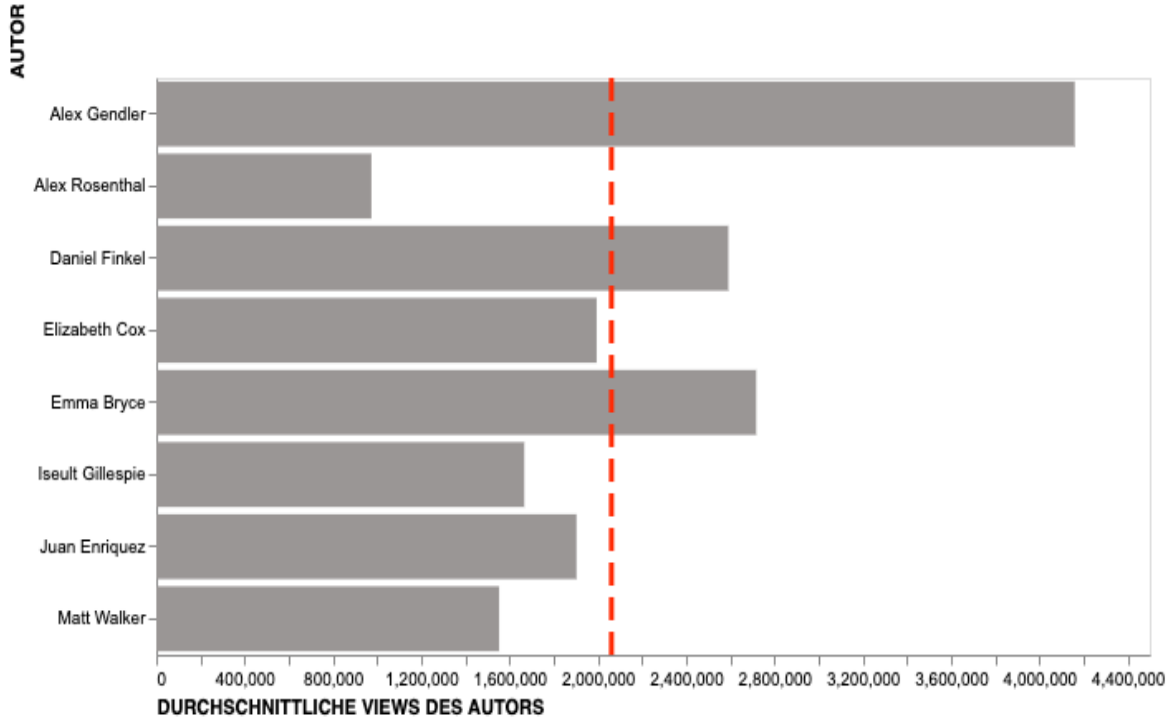
Die Top 10 Wörter in den Titeln sind dann life, world, new, future, change, work, help, climate, art und need. Diese Wörter sind auch gut in der Word Cloud sichtbar, da diese die Häufigkeit der Wörter mit einbezieht und dadurch die Größe der Wörter bestimmt.

	Wörter	Anzahl der Wörter
0	life	180
1	world	149
2	nt	134
3	new	133
4	future	132
5	change	110
6	work	96
7	help	96
8	climate	91
9	art	86
10	need	83
11	power	81
12	brain	80
13	people	78
14	history	76
15	solve	76
16	like	75
17	better	74
18	human	68
19	time	65
20	science	65

5. Autoren

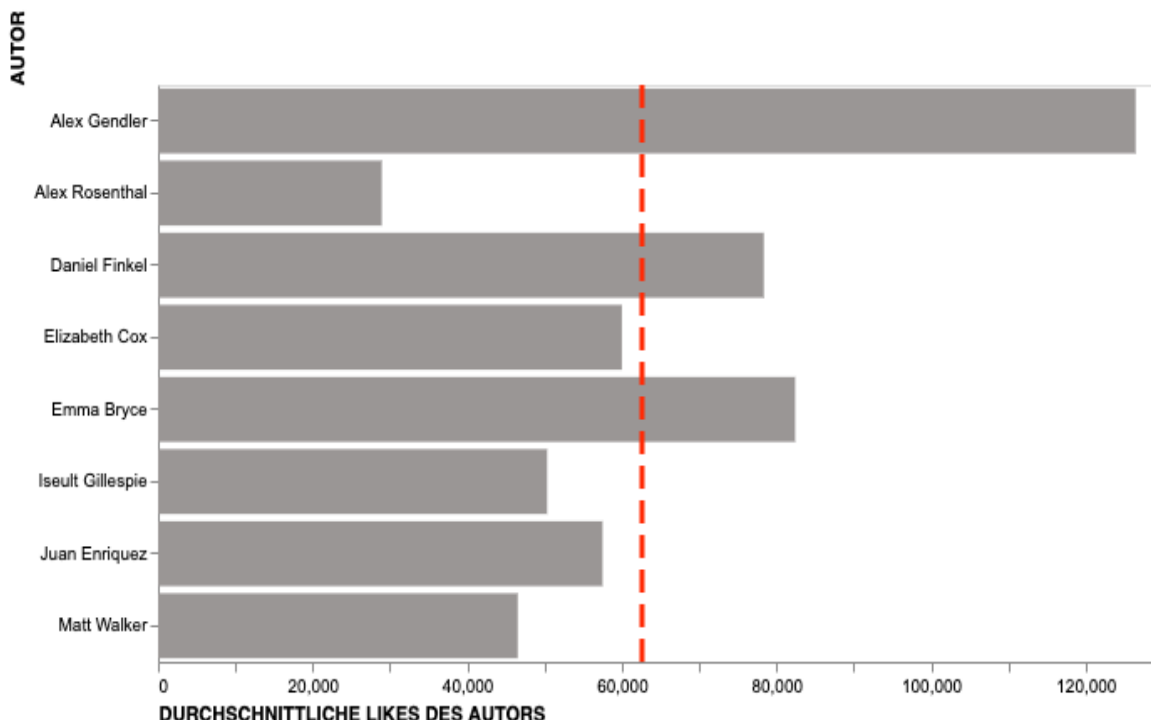
Die TED Talks wurden von 4443 Autoren bzw. Speakern geschrieben und gehalten. Die Autoren mit den meisten Videos sind:
 Alex Gendler mit 45 Videos, Iseult Gillespie mit 33, Matt Walker mit 18, Alex Rosenthal mit 5, Elizabeth Cox mit 13, Emma Bryce mit 12, Daniel Finkel mit 11 und Juan Enriquez mit 11. Für diese 8 Autoren habe ich mich entschieden noch eine genauere Analyse zu erstellen. Da danach 6 Autoren mit einer gleichen Anzahl von Videos (9) folgen, habe ich den Cut bei 8 Autoren gesetzt.

Durchschnittliche View-Anzahl der TOP 8 Autoren



In dieser Grafik werden die durchschnittlichen Views des Autors aufgezeigt. Wie man erkennen kann übertreffen nur 3 dieser 8 Autoren die durchschnittliche View-Anzahl aller Videos von 2.061.576.

Durchschnittliche Like-Anzahl der TOP 8 Autoren



Auch bei den Likes haben es nur Alex Gendler, Daniel Finkel und Emmy Bryce über die durchschnittliche Like-Anzahl von 62.608 geschafft. Das liegt natürlich auch am Zusammenhang zwischen Views und Likes.

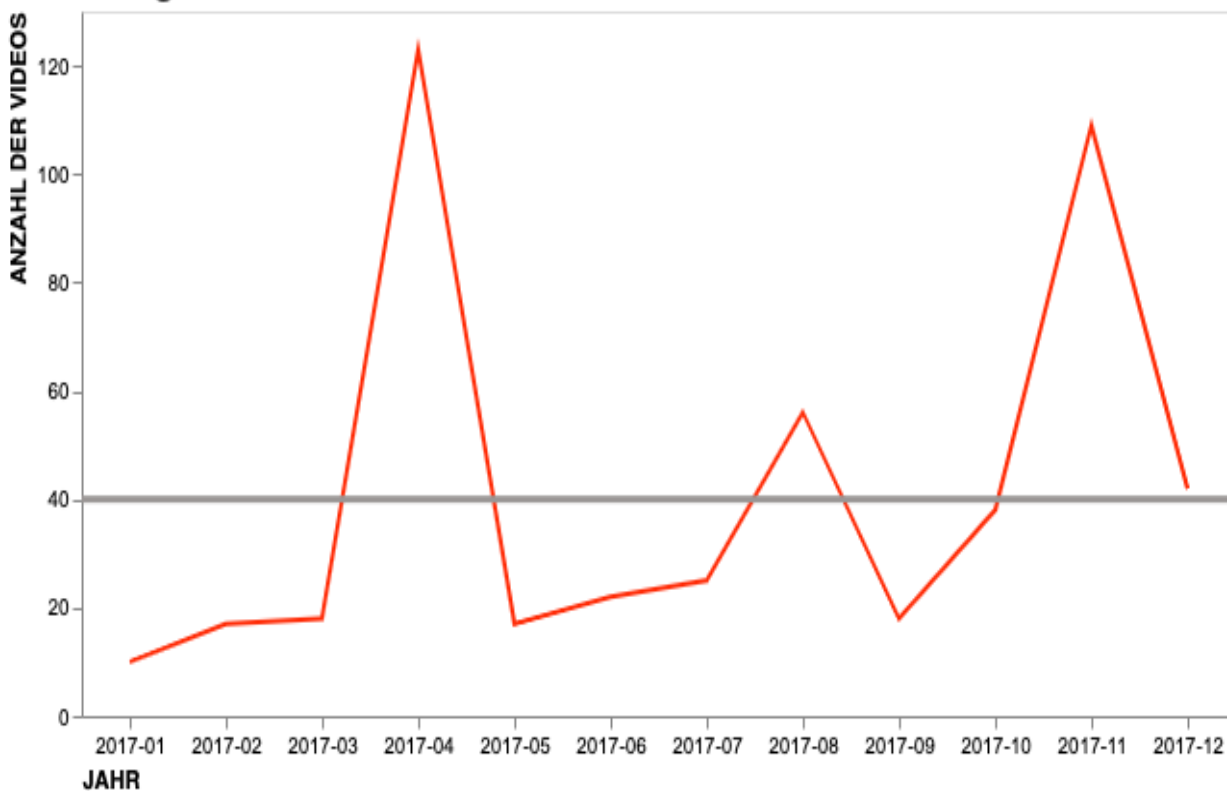
Dennoch gilt, nur weil die Autoren viele Videos herausgebracht haben heißt das nicht, dass die Videos auch gleichzeitig eine gute Performance haben. Deswegen sollte man in Zukunft schauen, welche Autoren eine gute Performance auf ihre Videos haben.

6. Monatsvergleich

Am Anfang wurden schon die Verteilungen der Videos nach Jahren dargestellt. Nun folgt noch eine Aufschlüsselung der Videos für die letzten 5 Jahre nach Monaten.

Die grauen Linien stellen die durchschnittlichen Videos pro Monat da. Es wurden in den letzten 5 Jahren durchschnittlich 40 Videos pro Monat veröffentlicht.

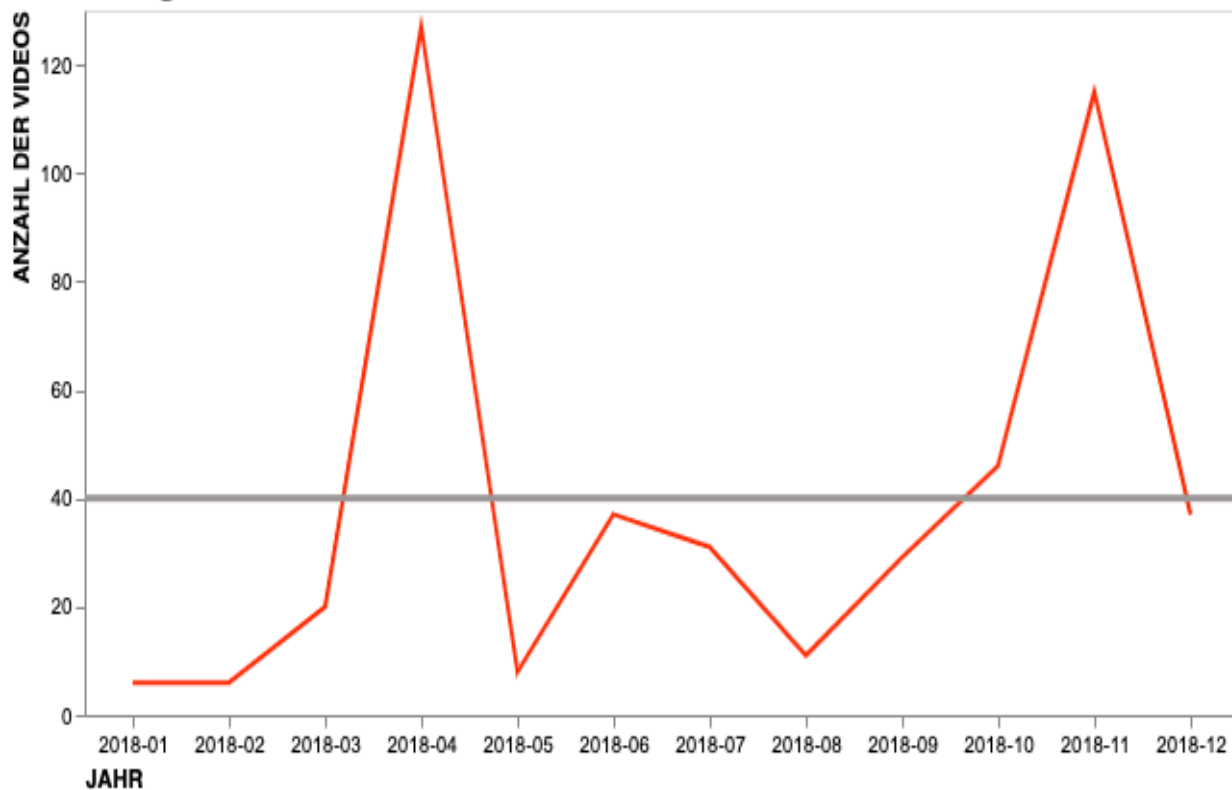
Verteilung der veröffentlichten TED Talks in 2017



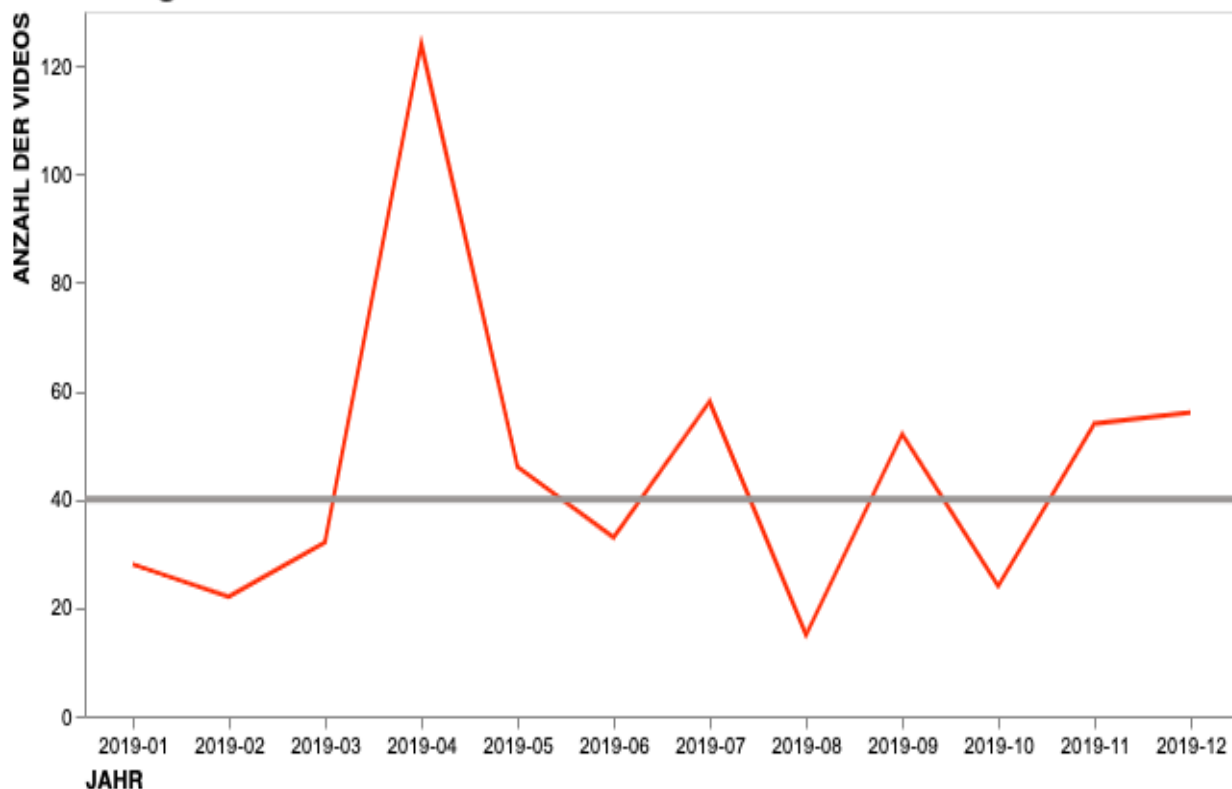
Im Jahr 2017 gibt es 4 Monate die den Durchschnitt übersteigen. Der Monat mit den meisten Videos ist der April gefolgt von November, August und Dezember.

Im Jahr 2018 gibt es ebenfalls 4 Monate die den Durchschnitt übersteigen. Ebenfalls der April, November und Dezember. Aber auch im Oktober.

Verteilung der veröffentlichten TED Talks in 2018



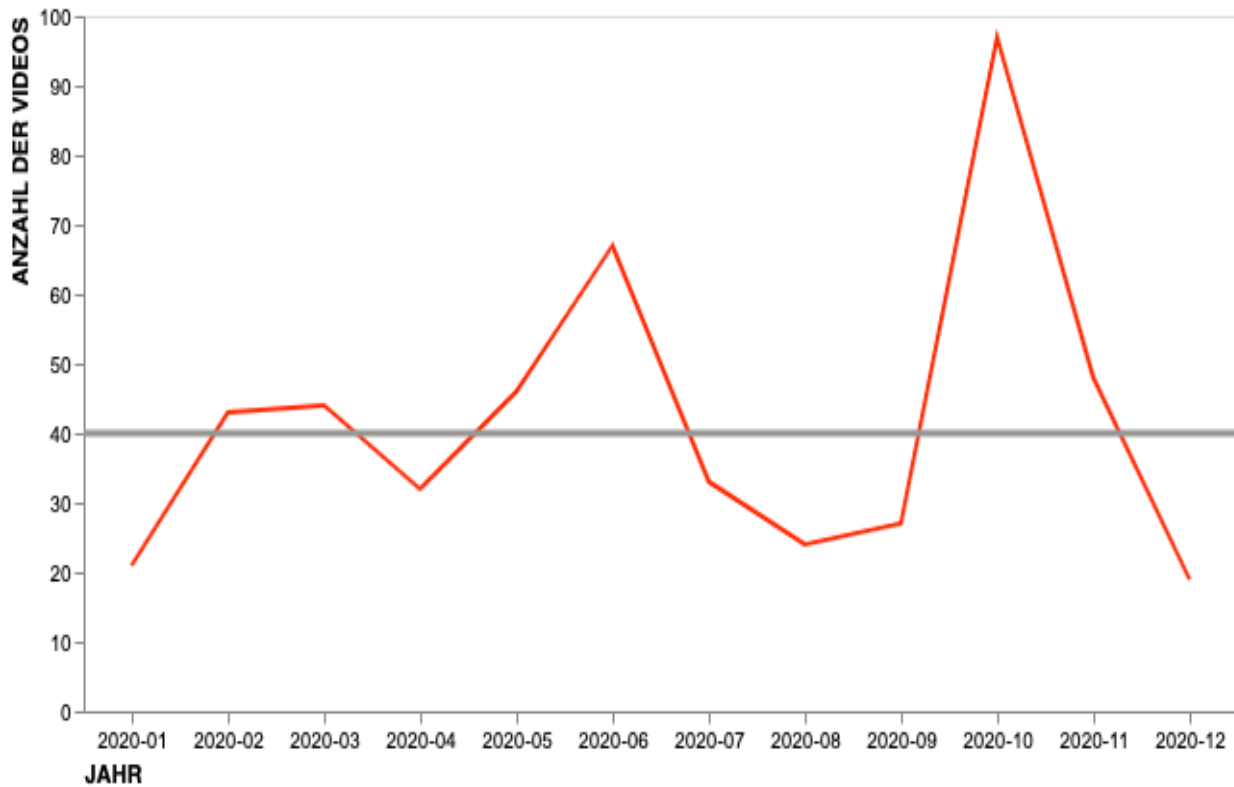
Verteilung der veröffentlichten TED Talks in 2019



In 2019 übersteigen die Monate April, Juli, September und November den durchschnittlichen Wert.

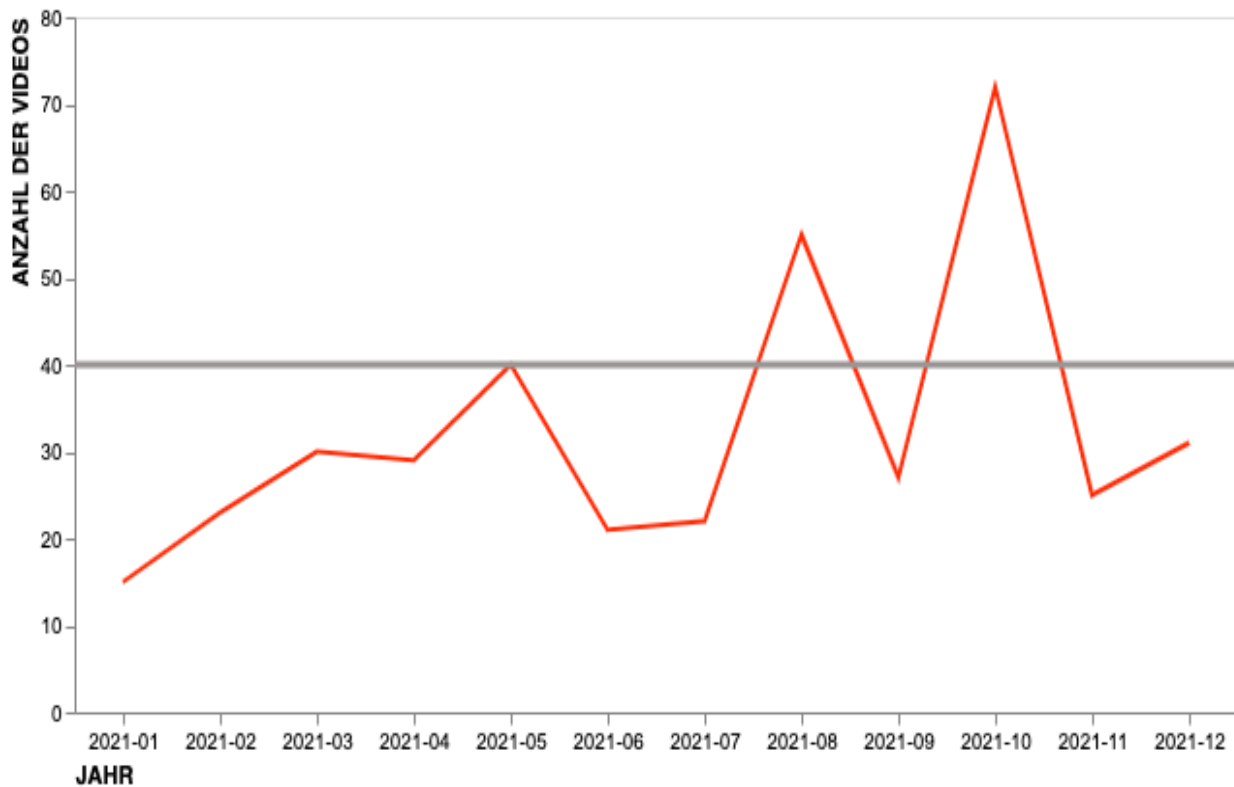
Der Oktober ist 2020 der Monat mit den meisten Videos. Aber auch im Februar, März, Juni und November wurde der Durchschnitt überschritten.

Verteilung der veröffentlichten TED Talks in 2020



2021 wurde 3 mal der Durchschnitt erreicht: im Mai, August und Oktober.

Verteilung der veröffentlichten TED Talks in 2021



Es fällt auf, dass im April und eher am Ende des Jahres die meisten Videos veröffentlicht werden. Das hängt evtl. mit dem Zeitpunkt der eigentlichen Konferenz zusammen. Außerdem lässt es auf ein "Sommer-Loch" schließen, denn in den Monaten Juni, Juli werden eher weniger Videos veröffentlicht. Hier wäre es gut, wenn pro Monat ungefähr gleich viele Videos veröffentlicht werden würden, so gibt es keine Monate an denen es deutlich mehr neue Videos gibt und die Zuschauer überfordern könnten.