

# Explorative Datenanalyse: Niedersächsische Hochschulen auf YouTube

## Fallstudie im Wintersemester 2024/2025

### Inhaltsverzeichnis

Einführung . . . . .	1
Forschungsfragen . . . . .	1
Aufgabenstellung . . . . .	2
Informationen zum Datensatz . . . . .	2
Aufgabe 1: Deskriptive Analysen . . . . .	3
Aufgabe 2: Explorative Analysen . . . . .	4
Hinweise zur Fallstudie . . . . .	5
Bearbeitung . . . . .	5
Abgabe . . . . .	6
Kurzvorstellung der Ergebnisse . . . . .	6

### Einführung

#### Forschungsfragen

YouTube ist eine wichtige Plattform für die Ansprache von Studieninteressierten, Studierenden sowie weiteren Zielgruppen wie Alumni, Unternehmen und der breiten Öffentlichkeit. Die Art und Weise, wie Hochschulen ihre Inhalte präsentieren, kann starken Einfluss auf deren Wahrnehmung und Attraktivität haben. In dieser Fallstudie sollt ihr untersuchen, wie verschiedene Hochschulen in Deutschland - insbesondere die in Niedersachsen ansässigen und große Fernuniversitäten - YouTube für ihre Zwecke nutzen. Hierzu stehen euch Metadaten von über 4.100 YouTube-Videos zur Verfügung.

Dabei stehen die folgenden Forschungsfragen im Fokus eurer Analysen:

1. Wie aktiv sind deutsche Hochschulen und speziell die niedersächsischen Hochschulen sowie die größten Fernunis in ihrer Nutzung von YouTube als Kommunikations- und Marketingplattform?
2. Wie erfolgreich sind die Hochschulen auf der Plattform im Vergleich zueinander?

## Aufgabenstellung

Beantwortet die Fragen in den folgenden beiden Aufgabenteilen. Erstellt für jede Frage eine oder mehrere Visualisierungen mit `ggplot2` und führt die notwendigen Transformationen mit `dplyr` und anderen Tidyverse-Paketen durch. Speichert die Ergebnisse in einem R-Skript und verwendet Kommentare, um eure Lösungen zu den jeweiligen Fragen zuzuordnen.

Anschließend verfasst ihr eine zusammenhängende schriftliche Abhandlung in Form eines wissenschaftlichen Artikels auf Basis eurer Erkenntnisse. Nutzt Quarto zur Erstellung des Artikels. Überlegt euch eine sinnvolle Struktur und setzt diese mit Markdown um. Wählt die Analysen aus den Aufgabenteilen aus, die eurer Meinung nach am besten zur Forschungsfrage passen. Verfeinert die Visualisierungen, die in eurem Artikel verwendet werden, und passt Achsenbeschriftungen, Farben etc. an, so dass sie veröffentlichungswürdig sind. Verfasst zu jeder Analyse eine Interpretation und stellt die einzelnen Erkenntnisse in einen logischen Zusammenhang. Euer Artikel sollte nicht nur eine Aneinanderreihung von Visualisierungen sein, sondern ein zusammenhängender Text mit einer kurzen Einleitung, aufeinander aufbauenden Erkenntnissen aus den Analysen sowie einem Schlussteil mit einer kurzen Diskussion. Eine Aufarbeitung des Standes des Wissens aus der Literatur ist für diese Fallstudie nicht notwendig.

## Informationen zum Datensatz

Der Datensatz beinhaltet Metadaten zu YouTube-Videos der größten niedersächsischen Hochschulen sowie der vier größten privaten Hochschulen deutschlandweit. Die Daten wurden am 29.11.2024 abgerufen. Laut [Wikipedia](#) sind die in der Tabelle aufgeführten Hochschulen die größten in Niedersachsen, gemessen an der Studierendenzahl. Die Hochschule Braunschweig wurde trotz mehr als 10.000 Studierenden nicht berücksichtigt.

Hochschule	Anzahl Studierende	YouTube Channel ID
<a href="#">Universität Hannover</a>	28.925	<a href="#">UCC0Eg0d1gGkpChelGo4-b4g</a>
<a href="#">Universität Göttingen</a>	28.614	<a href="#">UCzg-z2TL0Ks4Efz5o0z7AxQ</a>
<a href="#">TU Braunschweig</a>	17.709	<a href="#">UC8X4NAyIUr9Q12hVUOoqyhQ</a>
<a href="#">Universität Oldenburg</a>	15.635	<a href="#">UCbVtIjTlrNYqozDxIJPjHHw</a>
<a href="#">Universität Osnabrück</a>	13.640	<a href="#">UCN_aFLAXQEKZIa96PXxbi0Q</a>
<a href="#">Hochschule Osnabrück</a>	13.620	<a href="#">UCblmQHzIyzVHnKHwu0nISrw</a>

Die erfolgreichsten privaten Hochschulen in ganz Deutschland mit mehr als 10.000 Studierenden sind laut [Wikipedia](#):

Hochschule	Anzahl Studierende	YouTube Channel ID
<a href="#">International University</a>	85.483	<a href="#">UCR4ZvrahpeSe1TBrmObSkSw</a>
<a href="#">FOM</a>	58.836	<a href="#">UCdxxZI5F0u0l_wY8Bimsnow</a>
<a href="#">Hochschule Fresenius</a>	17.045	<a href="#">UCq9E5LlsAWjzK5RkvhWUZKQ</a>
<a href="#">Hamburger Fern-Hochschule</a>	13.629	<a href="#">UCSoSuBd4s-xJWHxJfqQCuvQ</a>

Folgende Variablen sind im Datensatz enthalten:

Variablenname	Beschreibung
<code>yt_id</code>	Eindeutige ID eines Videos
<code>channel_id</code>	Eindeutige ID des YouTube-Channels
<code>channel_title</code>	Name des YouTube-Channels
<code>published_at</code>	Zeitpunkt der Videoveröffentlichung
<code>url</code>	Link zum Video auf YouTube
<code>title</code>	Titel des Videos
<code>description</code>	Kurze Beschreibung des Videos
<code>duration_seconds</code>	Dauer des Videos in Sekunden
<code>views</code>	Anzahl der Ansichten zum Zeitpunkt des Datenabrufs
<code>language</code>	Die Sprache des Videos
<code>thumbnail</code>	URL zum Vorschaubild des Videos
<code>likes</code>	Anzahl der Likes des Videos zum Zeitpunkt des Datenabrufs
<code>comment_count</code>	Anzahl der Kommentare zum Zeitpunkt des Datenabrufs

## Aufgabe 1: Deskriptive Analysen

Die folgenden Aufgaben sind rein deskriptiv und können mit einer geeigneten Visualisierung beantwortet werden:

1. Wie viele Videos hat jede Hochschule veröffentlicht? Erstellt eine Visualisierung, die die Anzahl der Videos nach Hochschulen zeigt.
2. Wie viele Videos wurden von jeder Hochschule in den letzten Jahren ab 2015 inklusive veröffentlicht? Erstellt eine Visualisierung, die einen schnellen Vergleich ermöglicht.
3. Wie verhält sich die durchschnittliche Anzahl der Views über die verschiedenen Hochschulen hinweg? Gibt es hier klare Spitzenreiter?

4. Welche Uhrzeiten werden von den Hochschulen für die Veröffentlichung ihrer Videos bevorzugt?
5. Welche Wochentage sind für die Veröffentlichung von Videos besonders beliebt?
6. Wie lang sind die Videos durchschnittlich? Gibt es hier Unterschiede zwischen den Hochschulen?
7. Wie ist die Verteilung der Likes der Videos? Gibt es hier Unterschiede zwischen den Hochschulen?
8. Wie entwickelt sich die durchschnittliche Anzahl Likes für jede Hochschule über die letzten Jahre ab 2015 inklusive? Erstellt eine Visualisierung, die einen schnellen Vergleich zulässt.
9. Visualisiert den Anteil an deutschsprachigen und englischsprachigen Videos für jede Hochschule.
10. Gibt es einen Zusammenhang zwischen der Anzahl Views und den Likes, die ein Video erhält?
11. Ermittelt für jede Hochschule die neue Kennzahl *Views per Subscriber* und stellt sie vergleichend dar.
12. In welchem Zeitabstand veröffentlichen die Hochschulen neue Videos auf ihren Kanälen?

## Aufgabe 2: Explorative Analysen

Für die folgenden Aufgaben sollt ihr neben deskriptiven Darstellungen auch Hypothesen aufstellen, die eure Beobachtungen erklären könnten. Versucht, die Hypothesen wenn möglich mit weiteren Analysen zu untersuchen:

1. Betrachtet Videos mit besonders hohen Zahlen bei Likes, Views und Kommentaren. Stellt Hypothesen auf, warum diese Videos so erfolgreich sind. Welche Eigenschaften könnten dafür verantwortlich sein?
2. Welche Rolle spielen Videotitel und Videobeschreibungen für die Anzahl der Views, Likes und Kommentare? Entwickelt eine Hypothese und überprüft sie anhand der Daten (z.B. könnten bestimmte Schlagwörter die Reichweite erhöhen oder mehr Interaktionen fördern, möglicherweise spielt auch die Länge eine Rolle).
3. Gibt es einen Zusammenhang zwischen der Videodauer und der Anzahl der Views, Likes oder Kommentare? Stellt Hypothesen auf, um die beobachteten Zusammenhänge oder das Fehlen solcher Zusammenhänge zu erklären.
4. Gibt es Unterschiede zwischen den niedersächsischen Hochschulen und den privaten Fernunis in Bezug auf ihre Reichweite, die Art der Inhalte und die Interaktionsrate (Likes und Kommentare)?

5. Untersucht, wie sich die Nutzung und der Stellenwert der YouTube-Plattform im Laufe der Zeit für die Hochschule Osnabrück und die International University entwickelt haben. Stellt dazu Hypothesen auf zieht dazu auch Informationen hinzu, die ihr zu den beiden Hochschulen recherchiert und bringt sie in einem Zusammenhang mit den vorliegenden Daten.

## Hinweise zur Fallstudie

Beachtet bitte unbedingt die folgenden Hinweise zur Fallstudie.

## Bearbeitung

- Beantwortet *jede* Frage unter Zuhilfenahme von mindestens einer geeigneten Visualisierung. Eure Aufgabe ist es, eine zielführende Visualisierungsform für die jeweilige Frage und Antwort zu finden.
- Führt die explorative Datenanalyse ausschließlich mit R und dem RStudio durch. Für sämtliche Datentransformationen und Visualisierungen verwendet ihr Funktionen aus dem Tidyverse.
- Für Funktionen, die nicht im Tidyverse enthalten sind, dürft ihr auch andere Pakete verwenden. Achtet darauf, dass ihr die Pakete *zu Beginn* eures Skripts mit `library()` einbindet.
- Eine explorative Datenanalyse lebt von eurer Kreativität und Ausdauer. Gebt euch nicht mit der ersten Analyse zufrieden, die ihr erstellt. Das gilt insbesondere, aber nicht nur, für offene Fragestellungen. Probiert mehrere Ansätze aus. Wenn ihr ein interessantes Muster entdeckt, geht dem tiefer auf den Grund. Obwohl alle Teams die gleichen Aufgaben haben, wird es viele verschiedene Lösungswege geben.
- Wenn ihr bei einer Aufgabe nicht sicher seid, wie sie gemeint ist, oder Informationen fehlen, trifft sinnvolle Annahmen und dokumentiert sie in eurem Bericht.
- Lesbarer Code ist besser als schwer verständlicher Code. Verwendet den Pipe-Operator, verteilt lange Ausdrücke auf mehrere Zeilen, rückt den Code ein und fügt Kommentare hinzu, wo es angebracht ist.
- Reproduzierbarkeit ist in der Wissenschaft von großer Bedeutung. Stellt sicher, dass eure R-Skripte ohne zusätzlichen Aufwand auf einem anderen Computer ausgeführt werden können. Vermeidet also unbedingt absolute Pfadangaben.
- Es ist in Ordnung, ab und an Code zu kopieren und sich von anderen inspirieren zu lassen, solange der kopierte Code verstanden und erklärt werden kann und die Quelle als Kommentar angegeben wird. Dies schließt auch Sprachmodelle wie ChatGPT ein.

- *Make it work, make nice!* Erst die Funktion, dann die Ästhetik! Verliert euch nicht in Details wie Achsenformatierung oder Farbpaletten, bevor ihr die korrekte Datenbasis erstellt und die geeignete Visualisierungsform gewählt habt. Eine unpassende Visualisierung, selbst wenn sie optisch ansprechend ist, wird euch nicht helfen. Denkt umgekehrt! In eurem Bericht sollte aber alles glänzen.

## Abgabe

- Führt den Code eurer Datenanalyse für die Aufgabenteile 1 und 2 für die Abgabe in *ein* R-Skript zusammen. Verwendet Kommentare, um die Fragen, zu dem der jeweilige Code gehört, in eurem Skript zu kennzeichnen. Das Skript muss von oben nach unten fehlerfrei ausführbar sein.
- Verwendet für die Bearbeitung der Fallstudie das bereitgestellte R-Projekt. Für die Abgabe gebt ihr die beiden Dateien `case_study.R` und `article.qmd` über ILIAS ab. Achtet darauf, dass die Daten relativ zu diesen Dateien referenziert werden und im Ordner `data/` abgelegt sind.

## Kurzvorstellung der Ergebnisse

- Im Prüfungsgespräch sollt ihr euren mit Quarto verfassten Artikel und die darin enthaltenen Ergebnisse präsentieren.
- Bereitet für die Vorstellung eine kurze Präsentation vor. Die Präsentation darf nicht mehr als 5 Folien beinhalten.
- Der Zeitrahmen für die Präsentation darf 15 Minuten nicht überschreiten, anschließend sind 15 Minuten für Fragen vorgesehen.
- Jedes Gruppenmitglied sollte ungefähr den gleichen Sprechanteil haben.
- Gebt eure Präsentation im PDF-Format zusammen mit den anderen Dateien ab.

Viel Spaß und Erfolg bei der Bearbeitung der Fallstudie!