

Amazon x TUM Project Study

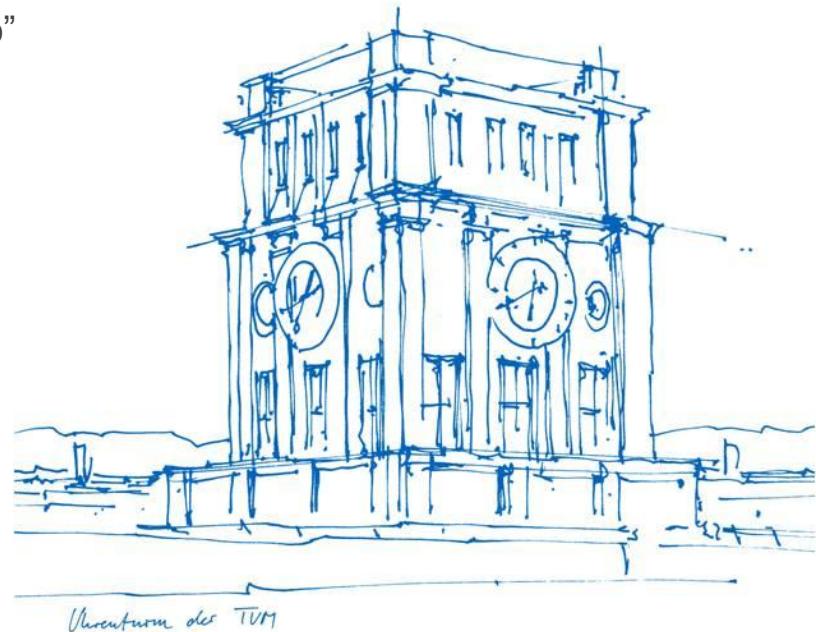
Title: “The Analysis of Consumer Behavior and Video Content Preferences of Young User Groups of Amazon Prime Video”

Supervisor: Prof. Dr. Christoph Ungemach
TUM School of Management
Chair of Marketing

Students:

Chernevich Mariia, Harutyunyan Narine,
Homola Daniel, Petrova Gabriela,
Simonazzi Filippo

Apr, 2021, Munich, Germany



Agenda



- 1 Project Introduction
- 2 Survey: Design & Results
- 3 IMDb Data
- 4 YouTube Data
- 5 Summary & Discussion



1

Project Introduction

2

Survey: Design & Results

3

IMDb Data

4

YouTube Data

5

Summary & Discussion

Project Details



Students

Chernevich Mariia, M.Sc Management and Technology

Harutyunyan Narine, M.Sc Consumer Science

Homola Daniel, M.Sc Data Engineering and Analytics

Simonazzi Filippo, M.Sc Mathematics in Data Science

Petrova Gabriela, M.Sc Consumer Science

Supervisors

Prof. Dr. Christoph Ungemach, TU Munich

Pavel Pankratov, Amazon

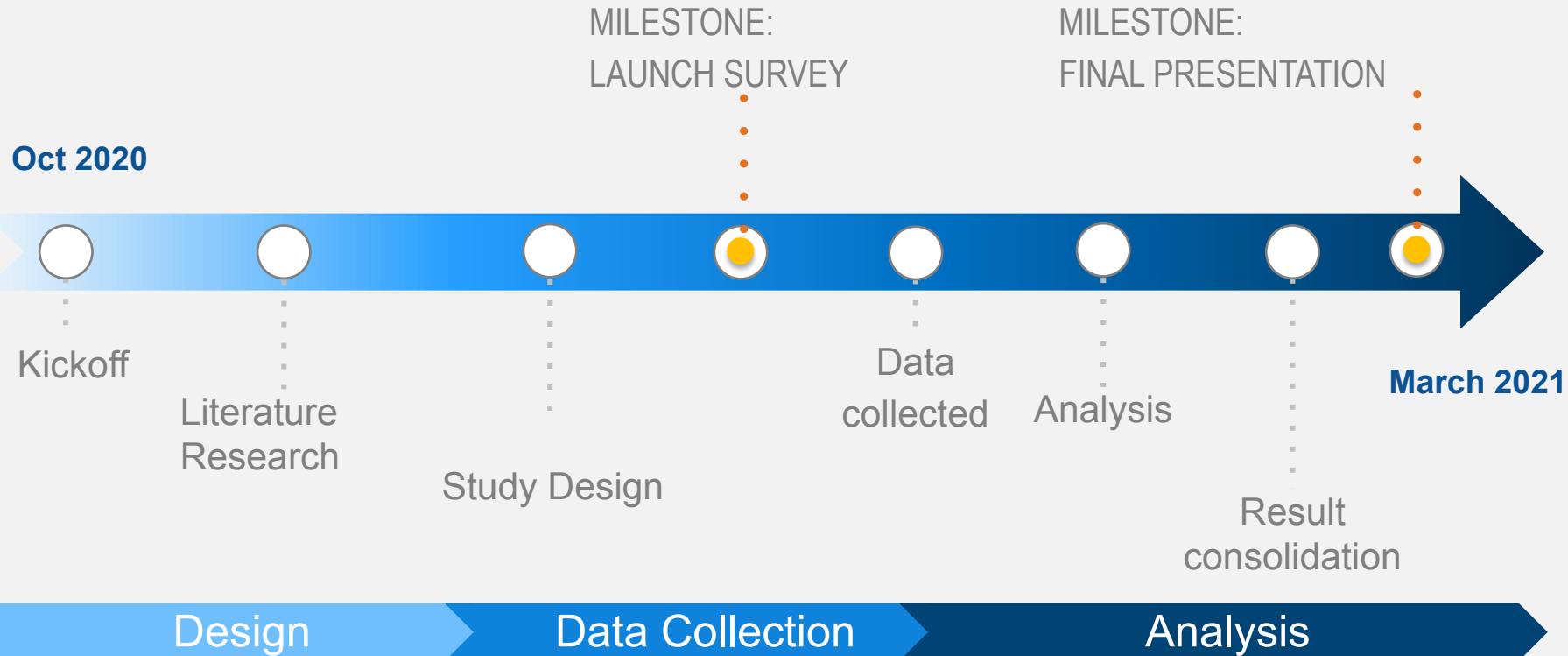
Tools

Communication Channels: Slack + Amazon Chime + Zoom

Workspace: Confluence + Miro



Project Timeline: Survey



Problem: Young Audience as underrepresented user group of APV



- The audience of Amazon Prime Video (APV) skews more towards adult audience
APV is included in Prime Membership which adult users initially subscribe for shopping, and receive APV as a benefit

- The young audience is underrepresented on APV & is more active on competitors' platforms
The acquisition of more young audience is important for several reasons:
 - Early adopters of new content
 - Influencing "upwards" older user groups



1

18-24 years old

- Have their own accounts, available data to analyze their content preferences and video streaming behavior

2

14-17 years old

- **No sufficient data for this user group about their content preferences**
→ they do not have their own separate accounts

- Their preferences and differences to the 18-24 y.o. Young Adult user group is unclear
→ essential for providing the content they need

Project goals

- Understand whether and in what ways **Video Content Preferences** and **Video Consumption Behavior** of **14-17 y.o.** user group **differs** from **18-24 y.o.** user group.
- Conduct **Youtube content analysis** as one of the main competitor platforms, identify well-performing content categories on Youtube for young audience, underrepresented on APV.
- Analyze 14-17 y.o. and 18-24 y.o. user groups' **preferences for Influencers**, as an important factor influencing young audience's content choices.
- Explore potential ways of **expanding young-audience-appealing content** on Amazon Prime Video, especially aimed at **new user acquisition**.

Research Directions



Research Direction. Awareness

This section aims to understand to what extent German young audience is aware of Amazon Prime Video as part of Prime Membership.

Research Direction. Attitude

This direction aims to explore young audience's opinion towards overall APV & Youtube and their specific features.

Research Direction. Consumption

This direction aims to identify the frequency and amount of consumption of Amazon Prime Video and Youtube among 2 young user groups.

Research Directions



Research Direction. Content Preferences

This section tries to explore the content preferences for young audience, namely, favorite content differences for 14-17 y.o. and 18-24 y.o. user groups, top content categories on Youtube, and potential ways of expanding underrepresented content on APV.

Research Direction. Content Characteristics

This direction aims to understand the preferred content characteristics for youth, such as video length.

Research Direction. Influencer Preference

This direction aims to understand 14-17 y.o. and 18-24 y.o. user groups favorite influencers, attitude towards them, as well as top industries for influencers.



For the project, the following methods are used:

Survey: Designing a survey and accessing a representative sample of German young audience:
14-17 years old and 18-24 years old.

Youtube & IMDb Data: Gathering and analyzing detailed YouTube catalogue and demand data
and IMDb demand data tied to young audience.



- 1 Project Introduction
- 2 Survey: Design & Results**
- 3 IMDb Data
- 4 YouTube Data
- 5 Summary & Discussion

Generation Z

- Born between 1995 – 2010
- Lives are tightly scheduled
- Virtual world = natural environment
- Digital-centric
- Fast-paced multimedia
- Lower ability to focus

Changes

- Audience fragmentation
- Content offerings fragmentation
- Binge-watching
- Production practice
- Cliff-hangers & recaps undermined
- Digital distribution paths
- New forms of marketing

Consumption

- “On-demand culture”
- Consume content at any time
- TV screens and laptops
- Poor data transfer
- Confusion of the seemingly indefinite platform offer
- Challenge - find the “right” platform

Amazon Prime Video

- Preferred content platform potential
- One of the most popular platform among 18-30 y.o.
- Amazon’s recommender system is one of the most famous examples of collaborative filtering

Past Studies: JIM Studies 2020



Title: Research on media usage by children and adolescents

Sample: N= 1,200 respondents across Germany

Age: 12 - 19

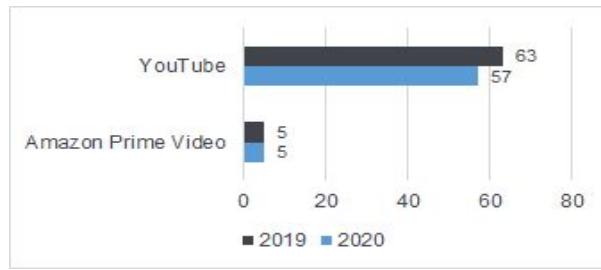
Interviews: by telephone or online

Dates: June 8 and July 20, 2020

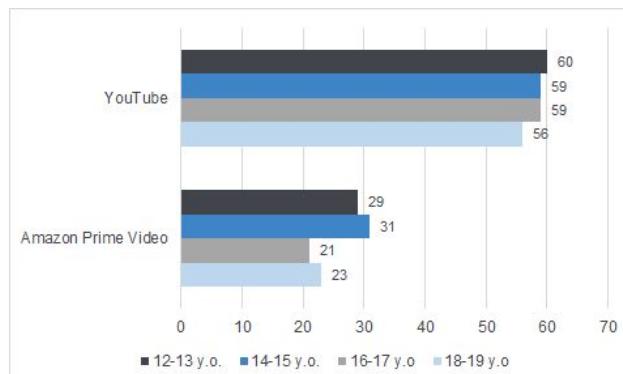
Conducted by "Medienpädagogische Forschungsverbund Südwest".



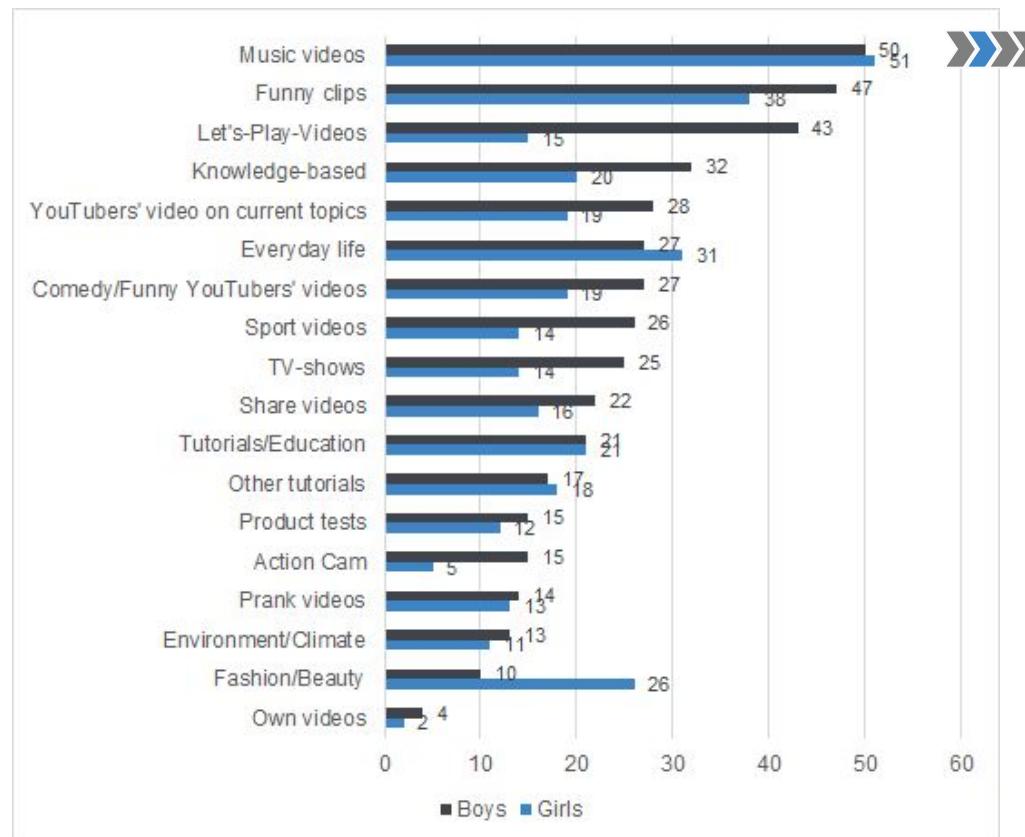
Literature Review, JIM-Studie 2020



Favourite Internet Offerings 2020 vs 2019
- up to 3 mentions -



Regular usage of TV-Shows, Series, Films on the Internet



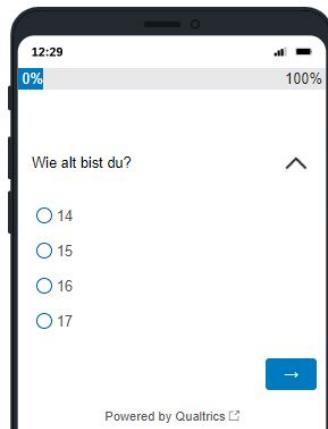
Regular Youtube Usage

Challenge: 14-17 parental consent
Agency: TUM -> Cint
Pre-Launch: 10 - 13.02.2021
Launch: 16 - 18.02.2021
2 age groups: 14-17 & 18-24
Valid responses: 464 and 442

Exclusion Criteria:

- Quota Full
- Early Screen-Out
- Quality Terminate

- PC and MP experience
- Average length - 10 mins



Survey: Organisation + Structure

- Quota Full**
 - Population in Germany (~ 5.5 million Young Adults and 3 million 15-17) -> 385 minimum sample
 - Population -> age groups -> + 3%
- Early Screen-Out criteria**
 - Children in the age of 14-17
 - Parental consent
 - Active YouTube consumption
- Quality Terminate**
 - Quality check question

Then Branch If:

If Wie oft schaust du YouTube / Amazon Prime Video normalerweise pro Woche? YouTube - Bitte auswählen: - Ich schaue generell nicht
Is Selected Edit Condition

Move Duplicate Options Collapse Delete

End of Survey Move Duplicate Customize Delete

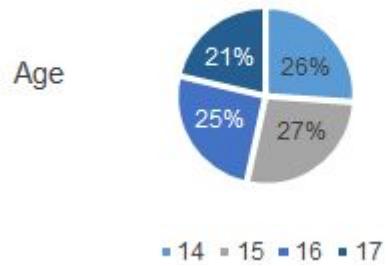
- Single-/multiple-choice, matrix tables (Likert scale), side-by-side, text entry questions
- Number of options: 2 - 15
- Force response
- Randomisation
- Display logic

Q8 Wie oft schaust du Videos, Filme oder Serien in diesen Kategorien normalerweise auf YouTube bzw. Amazon Prime?

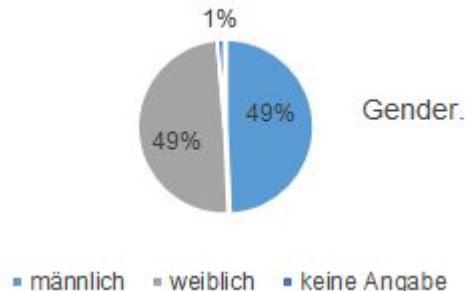
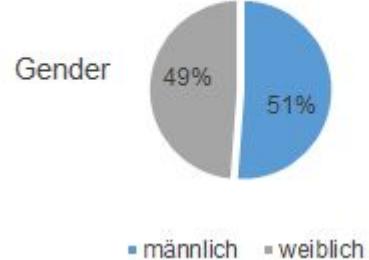
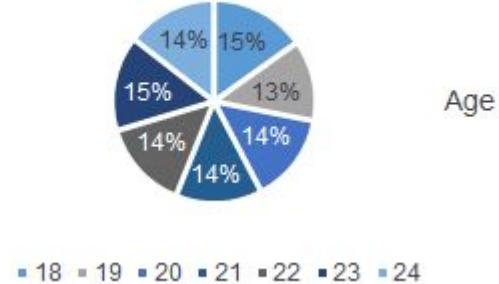
	YouTube					Amazon Prime Video				
	nie	selten	gelegentlich	oft	ständig	nie	selten	gelegentlich	oft	ständig
↳ Autos & Fahrzeuge	<input type="radio"/>									
↳ Bildung	<input type="radio"/>									
↳ Gaming	<input type="radio"/>									
↳ NGOs (Organisationen wie UNO, WWF, Greenpeace) & Aktivismus	<input type="radio"/>									
↳ Haustiere & Tiere	<input type="radio"/>									

Survey: Demographics

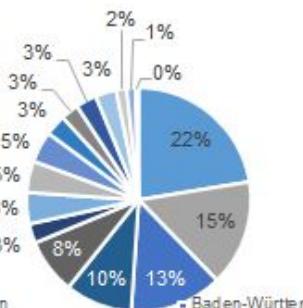
14 -17 age group



18 - 24 age group



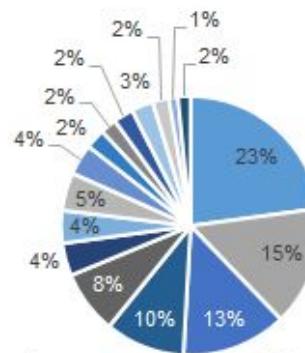
14 -17 age group



Bundesländer

- Nordrhein-Westfalen
- Bayern
- Baden-Württemberg
- Hessen
- Sachsen
- Sachsen-Anhalt
- Rheinland-Pfalz
- Brandenburg
- Mecklenburg-Vorpommern
- Saarland
- Bremen
- Hamburg
- Thüringen

18 - 24 age group

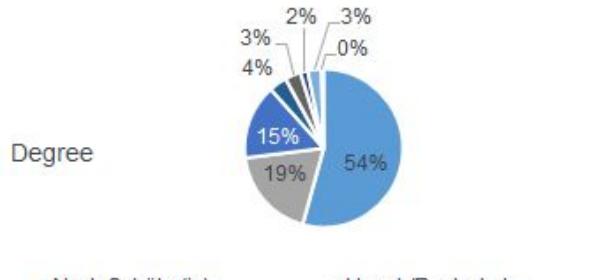
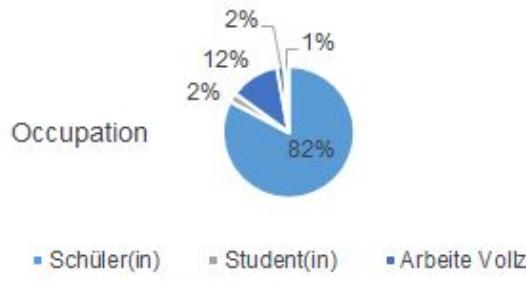


Bundesländer

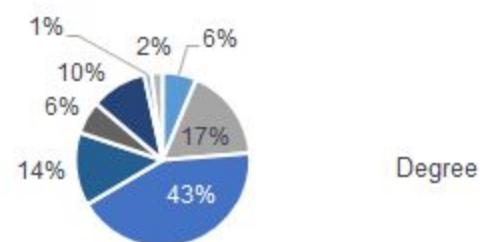
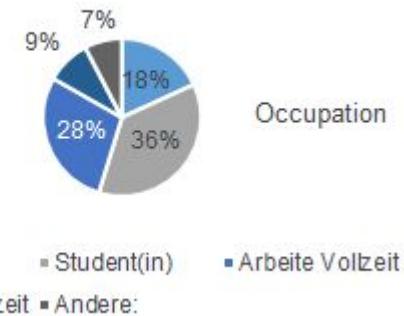
- Nordrhein-Westfalen
- Bayern
- Baden-Württemberg
- Niedersachsen
- Hessen
- Sachsen
- Sachsen-Anhalt
- Rheinland-Pfalz
- Brandenburg
- Thüringen
- Bremen
- Hamburg
- Saarland
- Berlin

Survey: Demographics

14 -17 age group



18 - 24 age group



Initial Questions

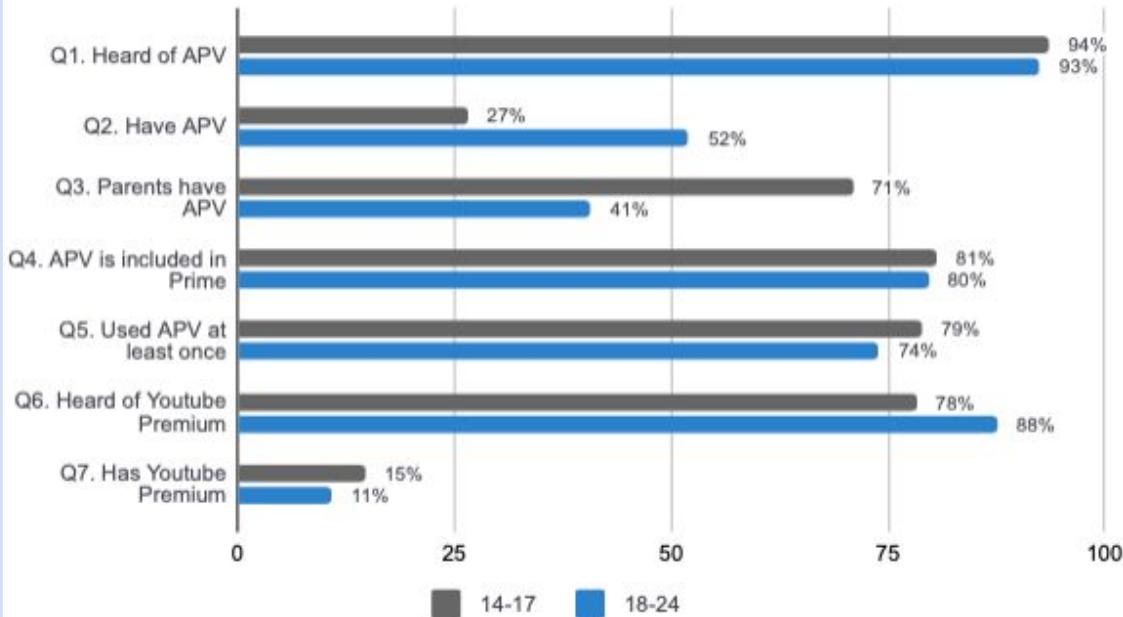
- Does APV need to increase Awareness among Teens and YAs?
- Do users know that APV is included in Prime Membership?
- Is there a potential for new user acquisition through increasing Awareness?

Survey Results: Awareness

APV Awareness

- **94% & 93%** (Q1) of users heard of APV
 - APV has generally high awareness
 - Awareness is not a major constraint for new user acquisition
- **19% & 20%** does not know that APV is included in Prime (Q4); **21% & 26%** have not used APV at least once (Q5).
 - Some potential for new user acquisition through increasing awareness
 - Increase the emphasis of "How to register for APV" in Marketing Campaigns or increasing interest towards APV

Survey Results: Awareness (% of answered "Yes")



Survey Results: Awareness

Teen Accounts

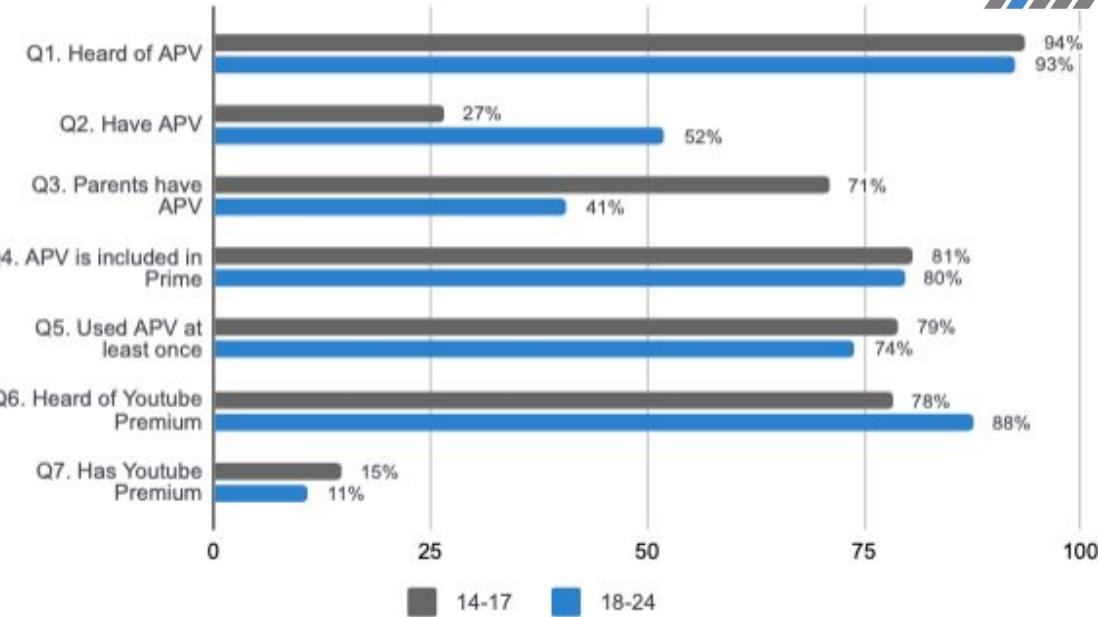
- **27% of Teens & 52% of YA report to have their own Prime accounts (Q2).**
- **Parents of 71 % of Teens & 41 % of YA have Prime account (Q3).**

→ Parent/Own Account Trade-off

→ **"Teen Login" for DE:** Teen has a separate login as part of the main account of their parent.

- As a strategy for new user acquisition/new registrations
- Better Content Preference Analysis for Teens

Survey Results: Awareness (% of answered "Yes")



Survey Results: Awareness

Youtube

85% Teen & 89 % YA of active YT users report that they do not have YPremium (Q7).

-> watch Youtube with ads

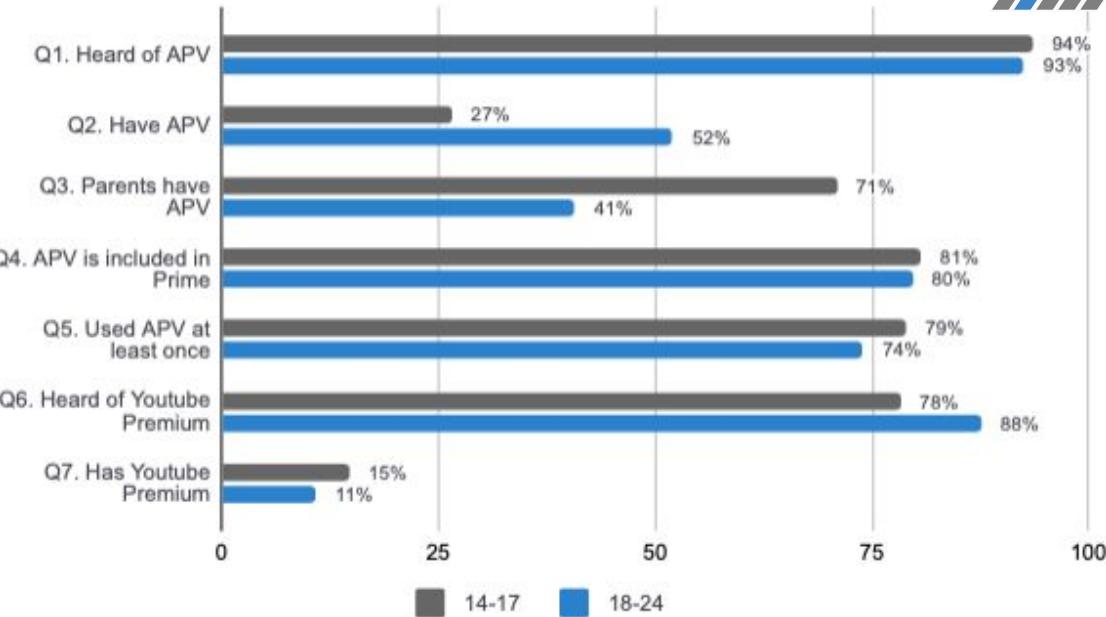
→ "Top content of Youtube, ad-free" positioning - competitive advantage

Example: Positioning for PVD

"Prime Video Direct" - platform for studios, distributors, independent filmmakers

"Prime Video Direct for Influencer Content/Top Youtube Content": Platform for Youtube-specific, user-generated content that is underrepresented on APV

Survey Results: Awareness (% of answered "Yes")



Initial Questions

- What do users think of APV & Youtube?
- What suggestions for improvement do users have for APV?

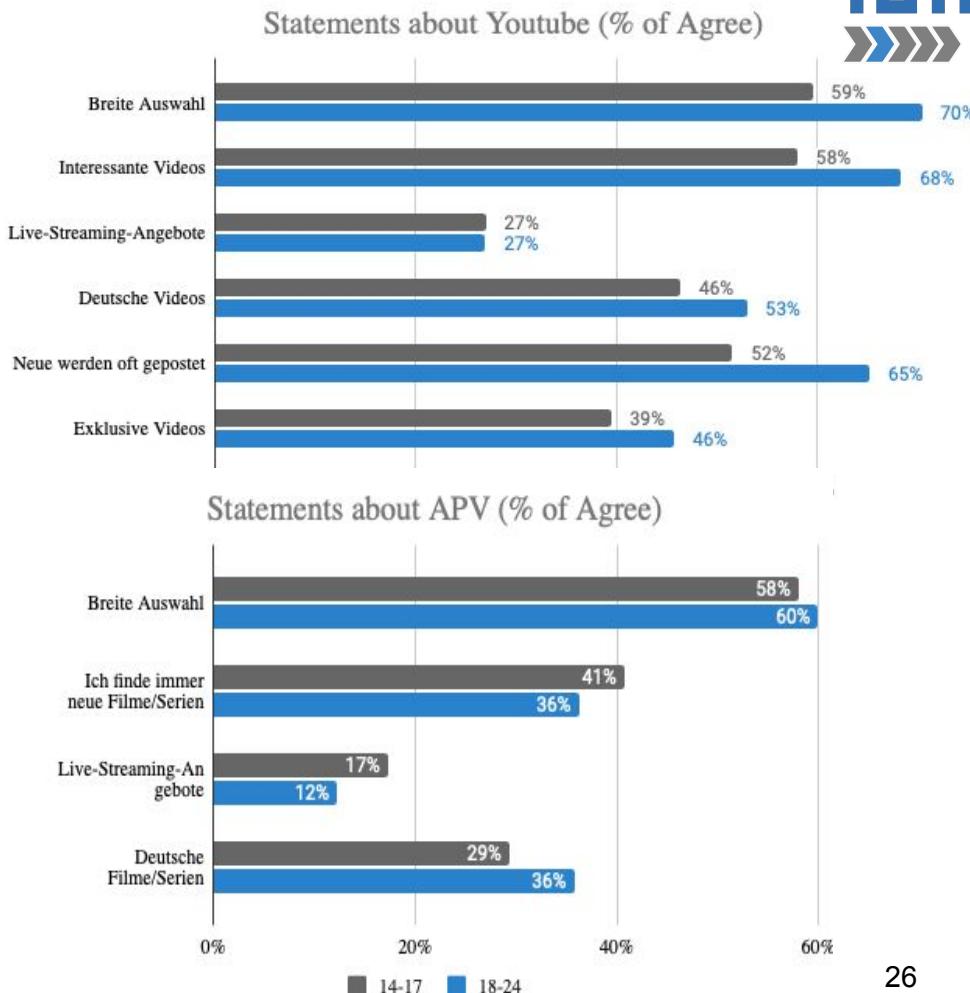
Survey Results: Attitude

Statements based on Nielsen (2020) report of what features are important for VoD Platforms for users

High variety of content: Young Adults (70%) agree with the statement that YT has a high variety of content significantly more than Teens (59 %).

Young Adults (70%) agree with the statement that YT has more variety of content than APV (60 %).

Both user groups: Youtube has **more Live-streams, more new content and more german content** than APV.



Open-ended question: What Recommendations do you have for Amazon Prime Video?



➤ Response Rates:

30 % of 14-17 user group has some recommendation, 21 % - no recommendations, the rest - no response/invalid

37 % of 18-24 user group has some recommendation, 9 % - no recommendations, the rest - no response/invalid

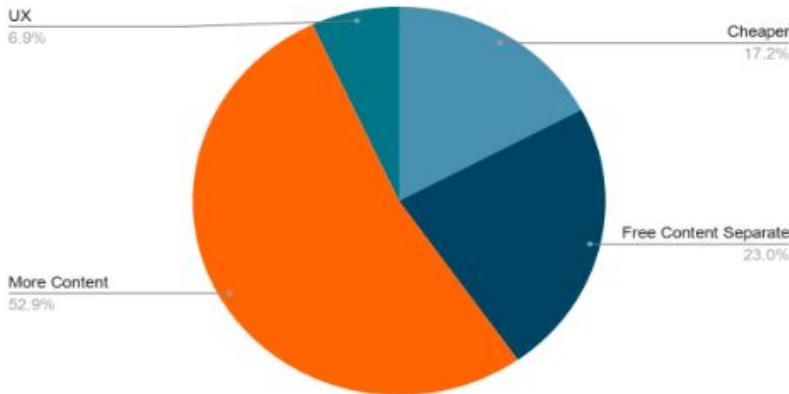
Based on Qualitative Coding, the recommendations were summarized in 4 main categories:

- **Free content:** More free content, Separate Free and Paid Content more clearly
- **Cheaper** prices for available content
- **More Content** in addition to already available content
- **User Experience** and User Design Improvements

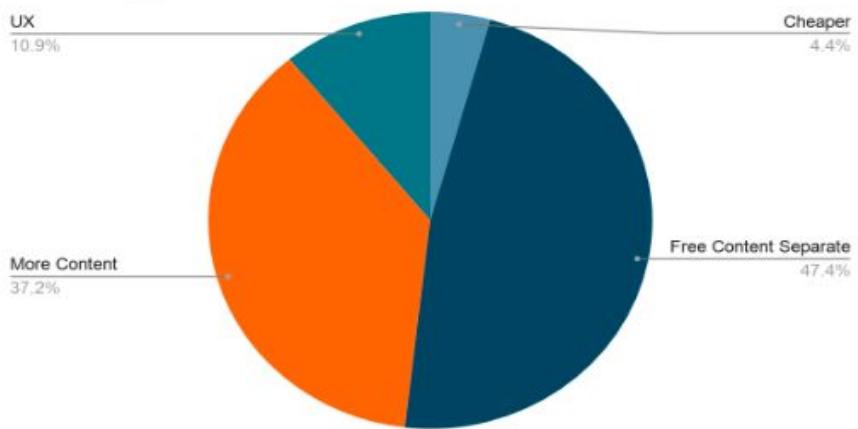
Open-ended question: What Recommendations do you have for Amazon Prime Video?



14-17 Suggestions



18-24 Suggestions



YA, Top Reply (48 %)
More free content or
separating free & paid
content better

--> Already being implemented
on APV

For Teens = 53 % Content

- "content is old, make more modern"
- "more interesting"
- "more german content"

-> to be further validated with quantitative research

For YA = 37 % Content

- general suggestion: "more content"
- not emphasized "old & modern" aspect

-> to be further validated with quantitative research

Initial Questions

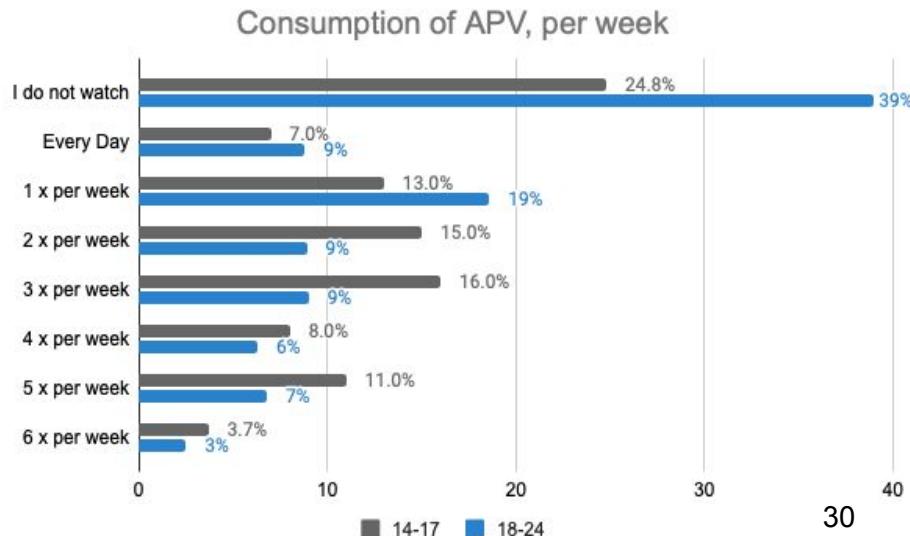
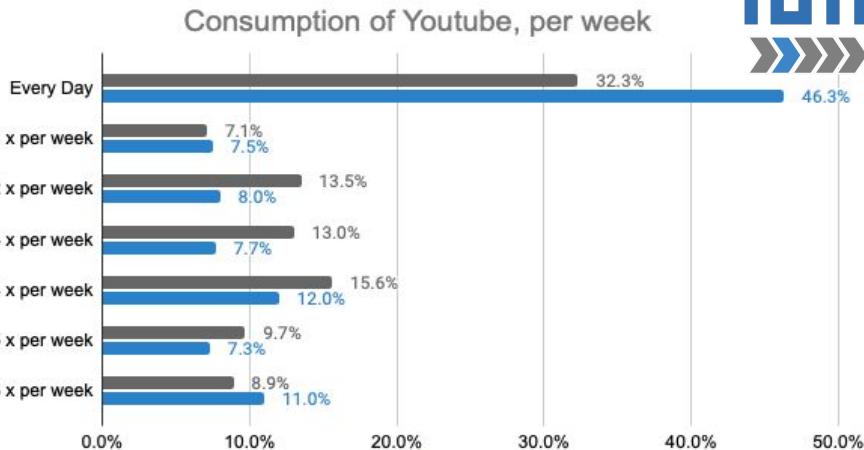
- How much do both user groups watch **Youtube & APV** per week?
- Are Teens active consumers of APV?

Survey Results: Consumption

- Youtube: YA watch Youtube more than Teens
(M=10.1 vs M=7.4 hours per week)
 - Focusing on Youtube content categories underrepresented in APV

- APV: Teens watch APV more than YA
(M=4.4 vs M=5.2 hours per week)

- 27% reported to have their own accounts (Q2)
 - Teens as an active user group
 - Focusing on content acquisition specifically for Teens' content preferences

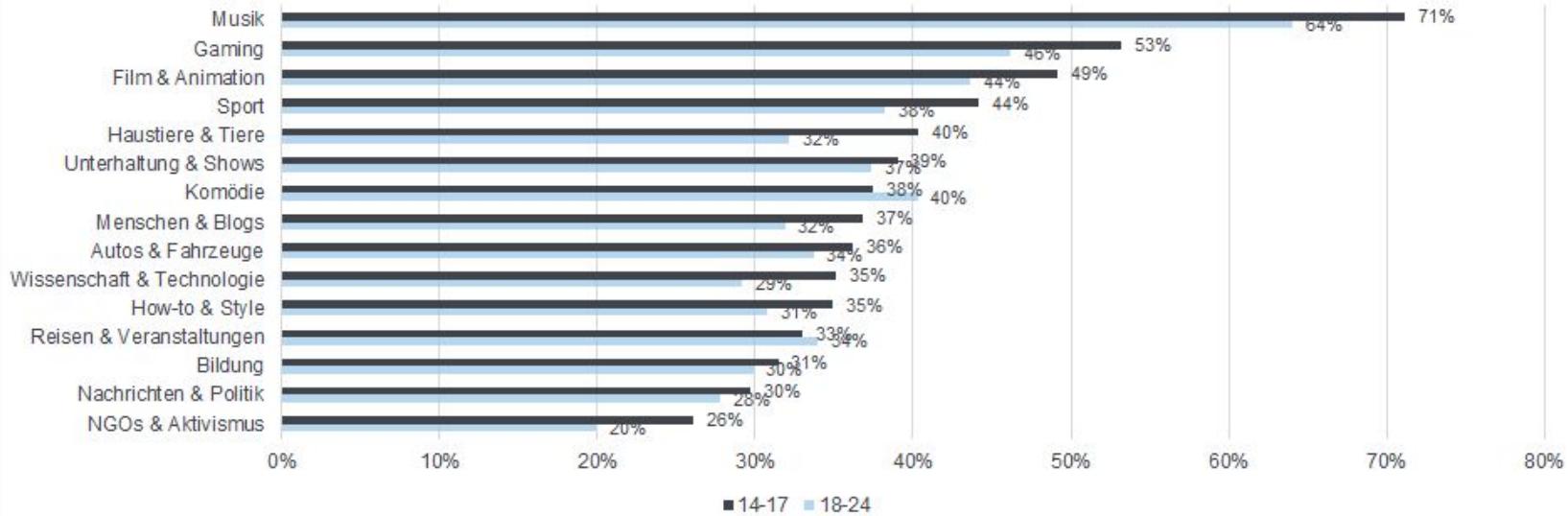


Initial Questions

- What content do Teens find interesting on YouTube and Amazon Prime Video?
- How do Teens and Young Adults differ in terms of content preferences?

Survey Results: Content Preferences

Topics that respondents find somewhat and very interesting



Teens (14-17 y.o.)

1. Music 71 %
2. Gaming 53 %
3. Film & Animation 49 %
4. Sport 44%
5. Animals & Pets 40 %

Young Adults (18-24 y.o.)

1. Music 64 %
2. Gaming 46 %
3. Film & Animation 44%
4. Comedy 40 %
5. Sports 38%

Survey Results: Difference in Content between Teens & YA



Following categories show the biggest significant differences

Teens are significantly more interested in

- Film & Animation
- Gaming (top category of interest for both groups)
- People and Blogs
- Animals & pets
- How To
- Sports
- NGOs (overall lower interest)

Young Adults show significantly higher interest in the category

- Comedy

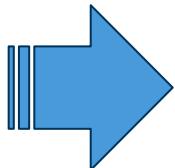
Survey Results: Consumption on YouTube & Amazon



- **Gaming**
- **Cooking (How To subcategory)**
- **People & Blogs**
- **Animals & Pets**
- **Sports**



- Action
- Romance
- Thriller
- Comedy
- SciFi
- Animation
- Anime
- Growing Up
- Criminal



Those 5 categories are such where Amazon can look into expanding the content, as there might be a lack of variety thus causing people to stay on YouTube for consuming those content categories.

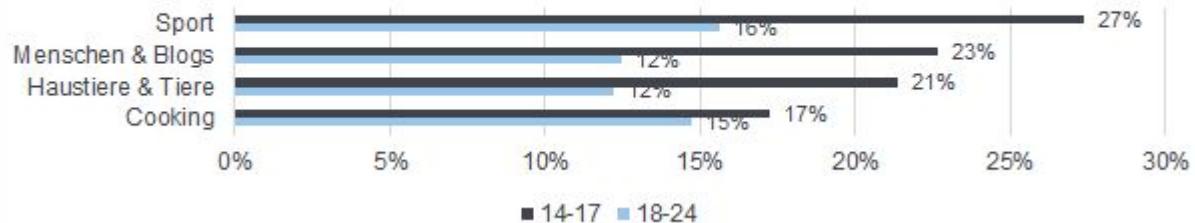
Survey Results: Amazon Content

- **Gaming ***
- Cooking
- People & Blogs
- Animals & pets
- Sports

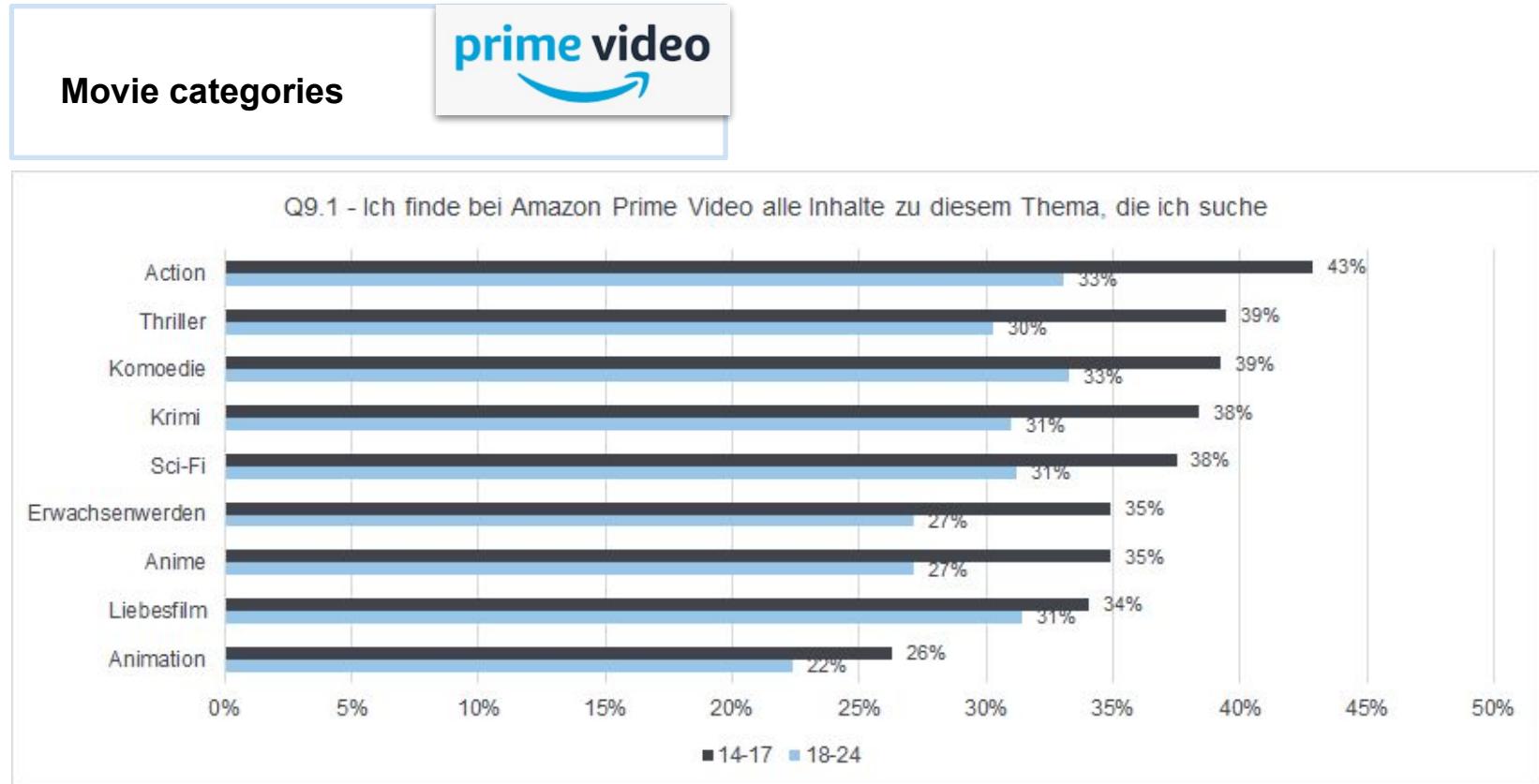


- For all those categories, a low percent of respondents said they find all they want on Amazon Prime Video, leading to the idea that expanding the variety of content would be useful.
- However, for Sports the demand here seems to be in the Young Adult group, as nearly a third of Teens say they find all they're looking for.

Q9.1 - Ich finde bei Amazon Prime Video alle Inhalte zu diesem Thema, die ich suche



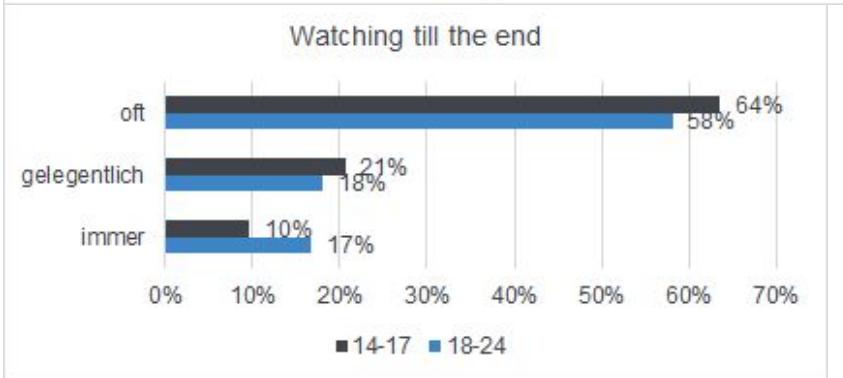
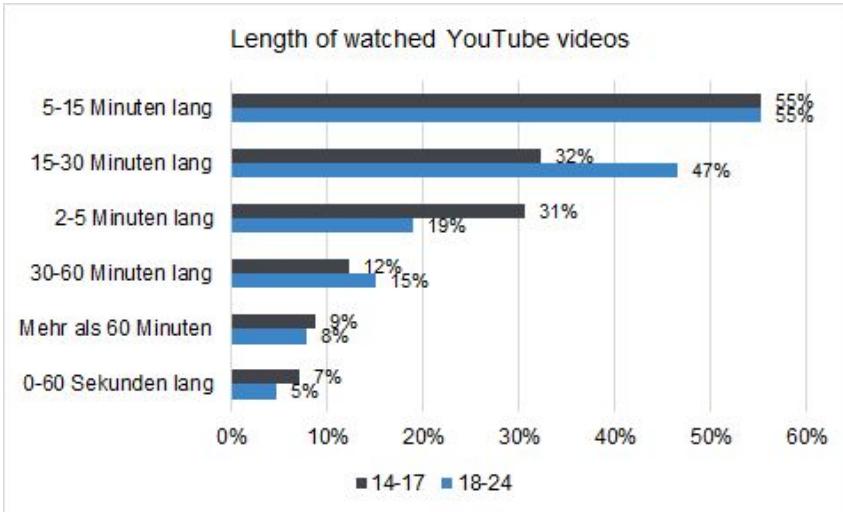
Survey Results: Amazon Content



Initial Questions

- Do Teens tend to watch shorter videos?
- Do Teens usually watch videos till the end?
- Do Teens watch live streams often?

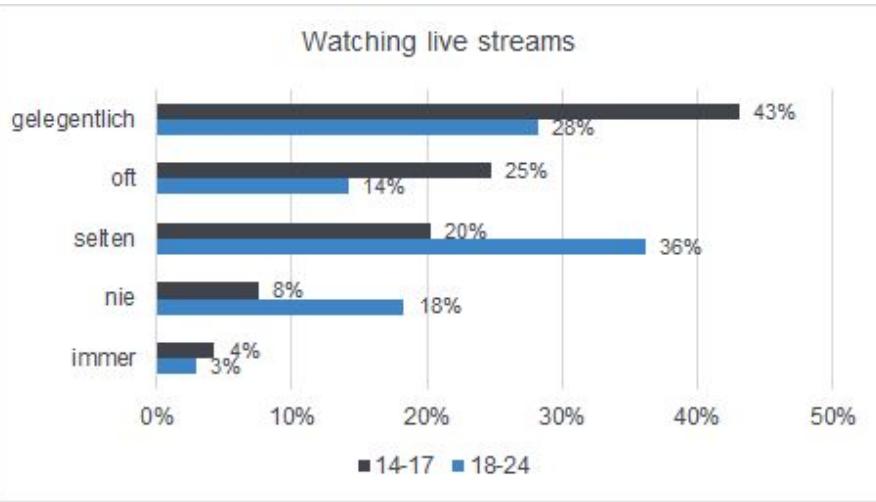
Survey Results: Content Characteristics



- 55% in both age groups report watching videos that are between 5 and 15 minutes
- Only 7 % of Teens watch videos under a minute regularly
 - Amazon can still compete with shorter content
 - No apparent threat of Teens switching to only clips < 1 Min

- Mostly, people watch the videos until the end, so there seems to be no obvious attention span shortage.

Survey Results: Content Characteristics



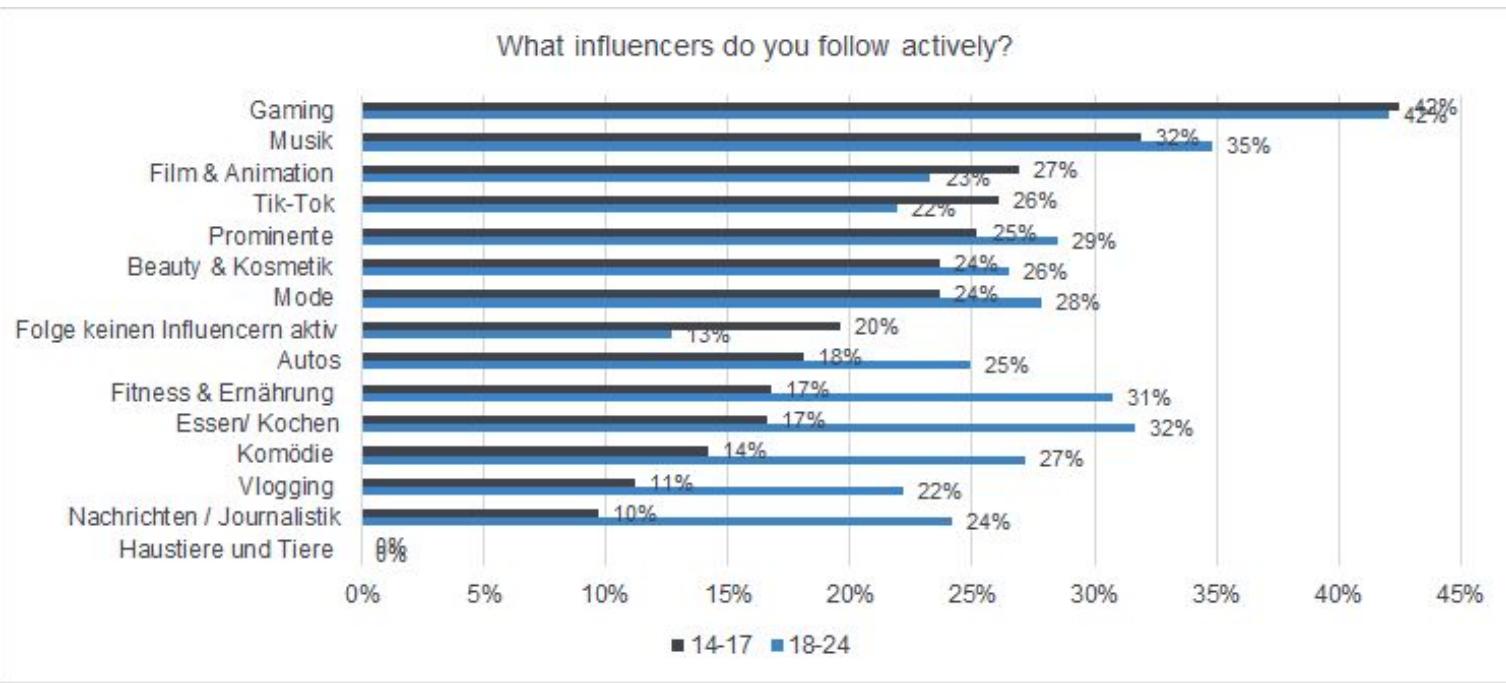
- There is no clear trend that Teens ($M=2.98$, $sd=0.96$) watch live streams constantly, however they do watch them more often than Young Adults ($M=2.47$, $sd=1.04$)

- In conclusion, Teens do have a tendency to lean towards typical YouTube content by watching significantly more live streams, liking short videos, and having a higher reported interest in Blogs & Vlogs content, which is typically user-generated and a quintessence of social media platforms.
- Amazon might profit from introducing certain elements of this type of content on their platform, depending on feasibility & most importantly deeper research into transferability.

Initial Questions

- Influencers in which categories are most popular among Teens & YA?
- Who are the top influencers for Teens & YA?
- How do Teens & YA engage with their favorite influencers?

Survey Results: Influencers



- Teens follow most actively influencers in the areas of **Gaming, Music, Film & Animation, Tik tok & celebrities**.
- However the biggest differences to Young Adults are in the categories **News, Fitness, Cooking, Comedy & Vlogging**.

Open-ended question: Who are your favorite influencers?

TOP Influencers, 14-17 y.o. user group

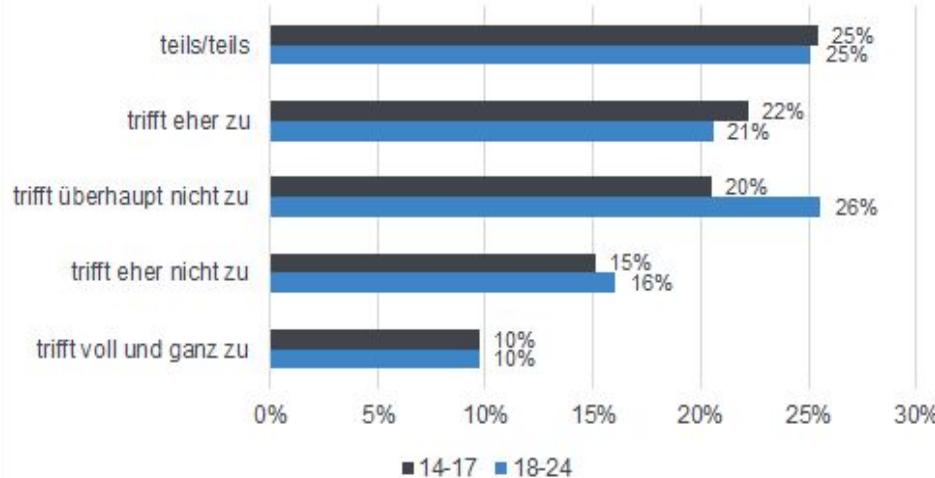
1. BibisBeautyPalace (Bianca Heinicke) - 5.6 %
Beauty & Cosmetics
2. Gronkh - 4.7 %
Gaming
3. Dagi Bee - 2.1 %
Beauty & Cosmetics
4. Paluten - 1.7 %
Gaming
5. Montana Black - 1.2%
Gaming

TOP Influencers, 18-24 y.o. user group

1. BibisBeautyPalace (Bianca Heinicke) - 3.8 %
Beauty & Cosmetics
2. Montanablack - 3.1 %
Gaming
3. Pamela Reif - 2.2 %
Fitness & Nutrition
- = 4. JP Performance - 2 %
Cars
- = 4. Paluten - 2 %
Gaming
- = 5. Sascha Huber - 1.8 %
Fitness & Nutrition
- = 5. Jindaois - 1.8 %
Lifestyle
- = 5. Dagi Bee - 1.8 % Beauty & Cosmetics

Survey Results: Engaging with influencers

Interest in seeing a documentary about favorite influencer



- Only 10% of Teens are very interested in seeing a documentary about their favorite influencer, another 22% being relatively interested.
- Teens report higher emotional connection with their favorite influencers compared to Young Adults, which would suggest they have a higher engagement and thus might be drawn by collaboration as initially presumed.

- Given the spread of liking for various influencers, it is unlikely that a collaboration with a single influencer will result in massive audience acquisition despite a substantial investment.
- Amazon might achieve higher ROI from other types of collaboration, e.g. "Prime Video Direct for Top Youtube Content" , or with Twitch since Gaming is the most popular category of influencers and is a type of content vastly underrepresented on Amazon Prime Video.



- 1 Project Introduction
- 2 Survey: Design & Results
- 3 IMDb Data**
- 4 YouTube Data
- 5 Summary & Discussion

- 3.1 Introduction to IMDb Dataset and KPIs
- 3.2 Which genres are the most important for teenagers?
- 3.3 Survey validation using IMDb data
- 3.4 Actors Analysis
- 3.5 Comparison between IMDb and APV

IMDb Overview:



- **PRO**
 - Age Granularity
 - Assumption: voters from 0 - 17 are considered to be at least 13/14 years old
- **CON**
 - Very skewed towards males
 - Few votes from teenagers, we decided to consider only a subset of the titles: only the ones with at least 30 votes from teenagers (~2500)

IMDb

Age splits

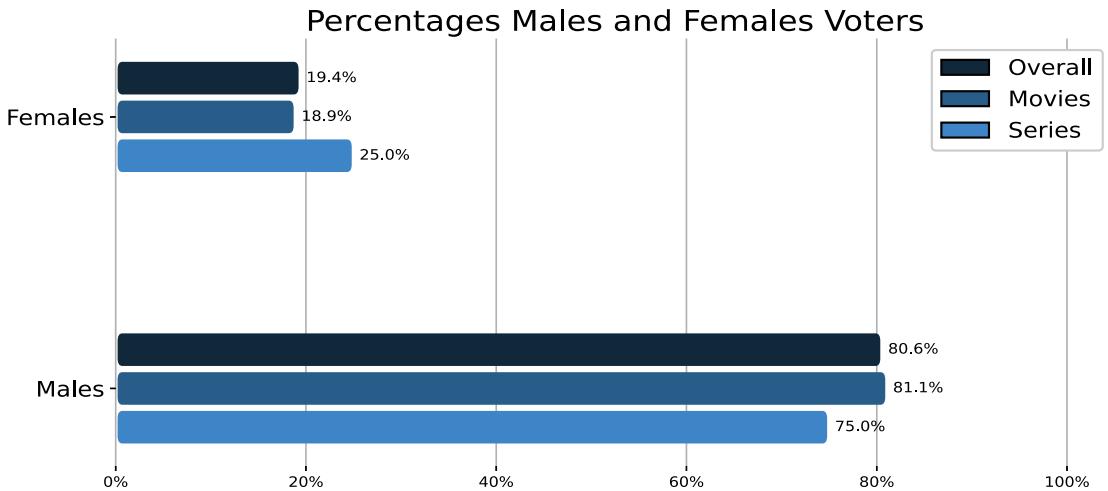
- 0 - 17
- 18 - 29
- 30 - 44
- 45+

- PRO

- Age Granularity
- Assumption: voters from 0 - 18 are considered to be at least 13/14 years old

- CON

- Skewed towards males
- Few votes from teenagers, we decided to consider only a subset of the titles: only the ones with at least 30 votes from teenagers (~2500)

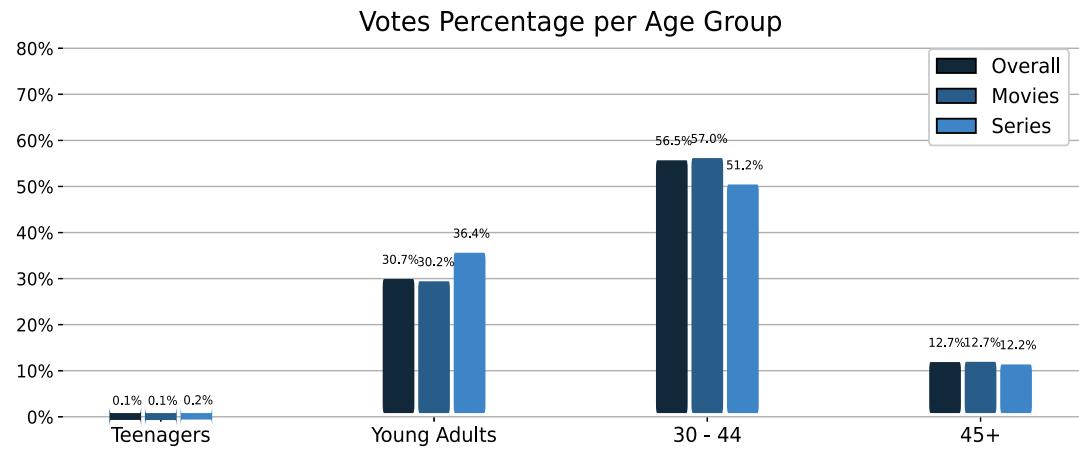


- PRO

- Age Granularity
- Assumption: voters from 0 - 18 are considered to be at least 13/14 years old

- CON

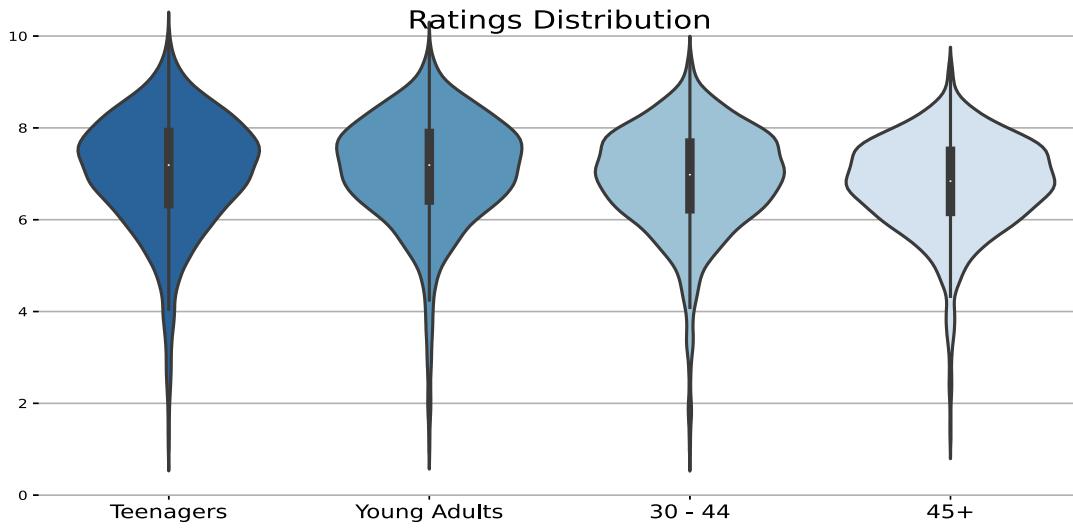
- Very skewed towards males
- Few votes from teenagers, we decided to consider only a subset of the titles: only the ones with at least 30 votes from teenagers
- Final dataset has 2544 titles



IMDb KPIs Overview:

- **Ratings**

- It tends to be similar among different age groups, especially teenagers and young adults



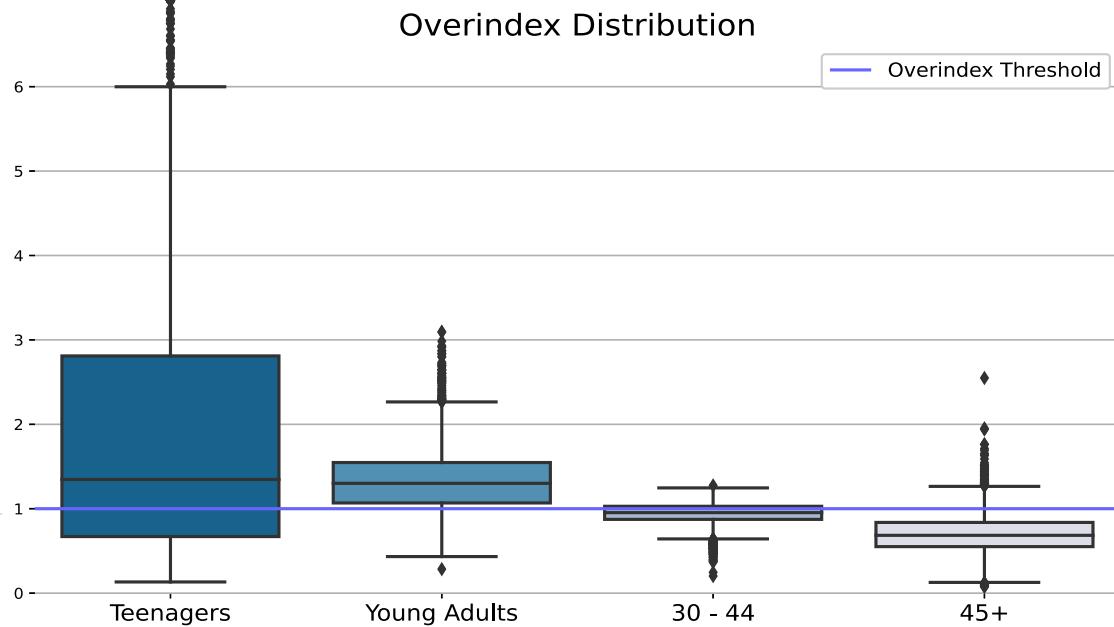
IMDb KPIs Overview:

- **Ratings**

- It tends to be similar among different age groups, especially teenagers and young adults

- **Overindex**

- It measures how engaging a title is for a specific age group
- Higher than 1 means that the title received proportionally more votes
- For teenagers, it has many outliers due to numerical reasons



IMDb KPIs Overview:

- **Ratings**

- It tends to be similar among different age groups, especially teenagers and young adults

- **Overindex**

- It measures how much a title was engaging for a specific age group
- Higher than 1 means that the title received proportionally more votes
- For teenagers, it has many outliers due to numerical reasons

- **Rank Overindex**

- It assigns a rank to each title based on his overindex
- Used to control outliers

	title	oi_0_17	oi_18_29	rank_oi_0_17_main	rank_oi_18_29_main
30096	Dexter	0.195370	1.016131	2535	2023
428136	Mr. Robot	0.952736	1.390640	1598	1015
58656	Stranger Things	2.563076	1.318010	719	1216
1224733	Wheels	127.001096	0.973891	1	2107

Analysis

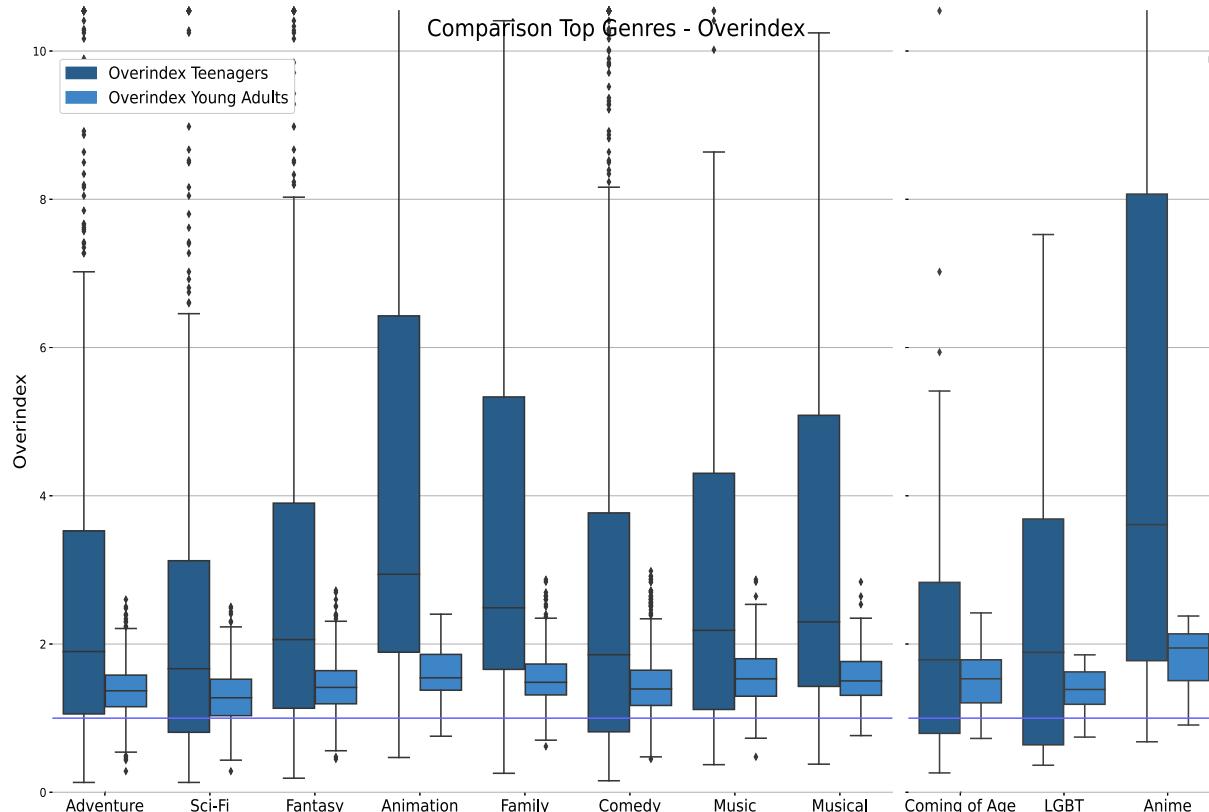
- Main indicator: Overindex
- Rank Overindex is used to make sure that the evaluation is not biased by outliers
- Ratings are, for most of the genres, not so different between age groups, in particular teenagers and young adults
- Correlation is used to understand whether there is a significant difference between teenagers and young adults.

Top Genres for Teenagers

- Adventure
- Animation
- Anime
- Comedy
- Coming of Age
- Family
- Fantasy
- *LGBT*
- Music / Musical
- Sci - Fi

Comparison Teenagers - Young Adults

Overindex



- Overindex

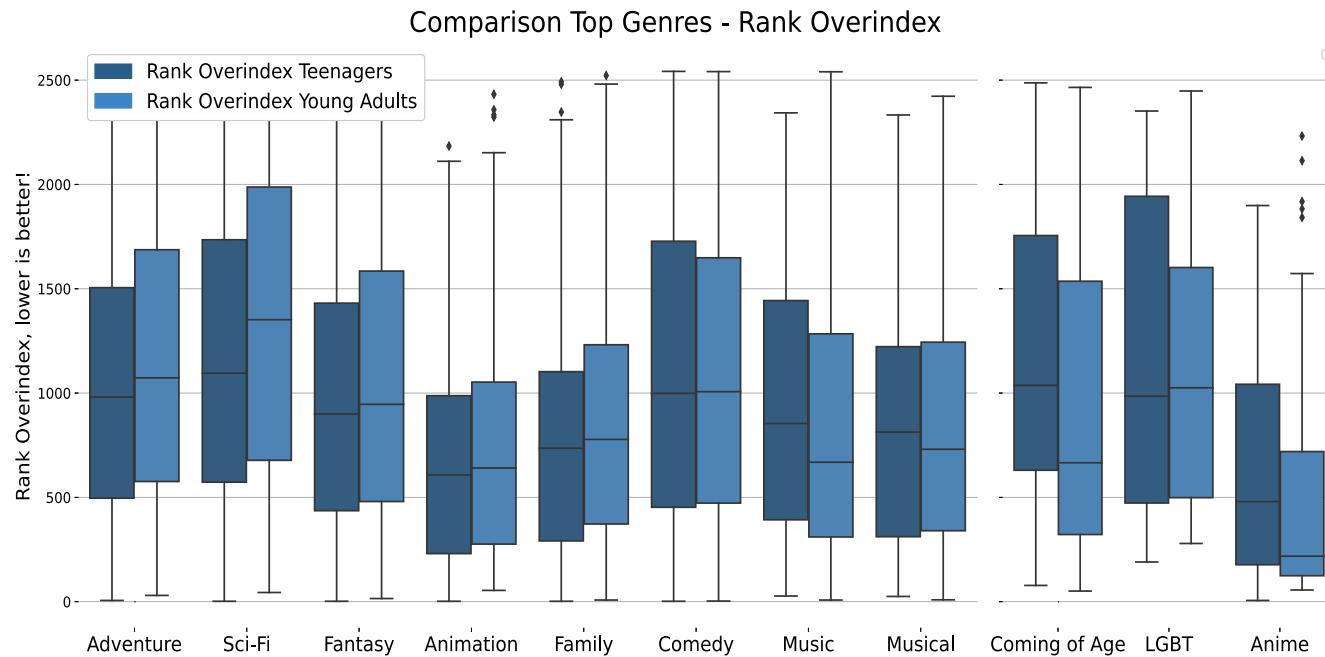
The first quantile is already higher than 1 for almost all these genres.

The boxplot highlights how this metric is unstable for teenagers.

Animation and Anime are the genres with the highest overindex, but also with the highest variance.

Comparison Teenagers - Young Adults

Rank Overindex



- Rank Overindex

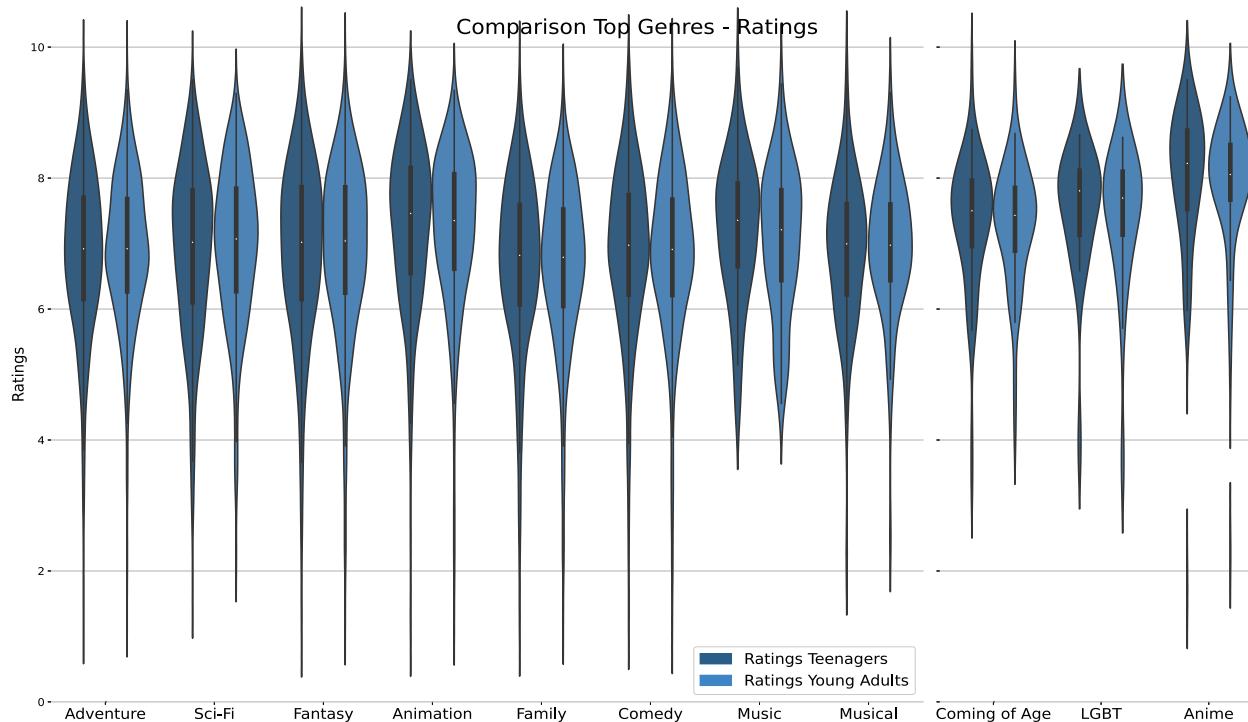
The lower the better!

The boxplot allows to understand the difference between teenagers and young adults

For example, Anime has a very low rank for both, but it's more consistent for young adults.

Comparison Teenagers - Young Adults

Ratings



- Ratings**

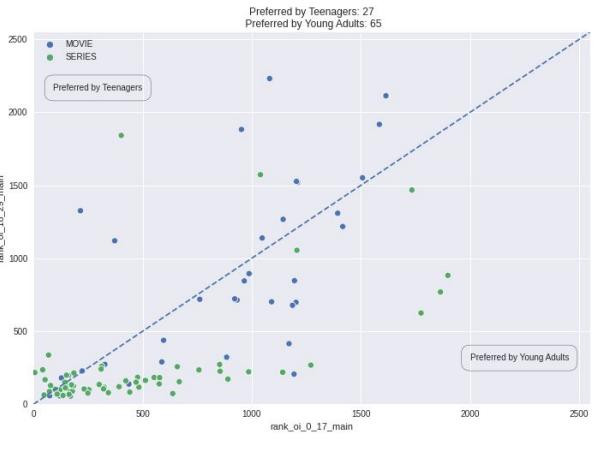
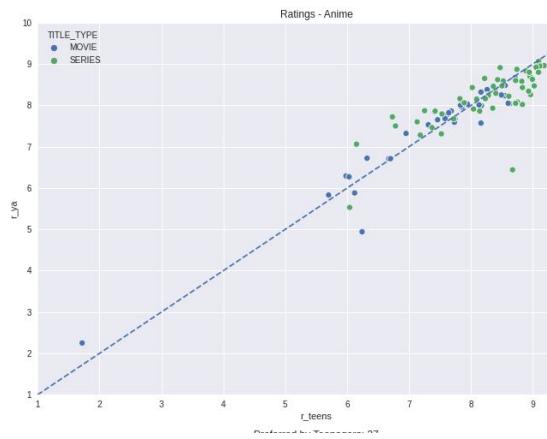
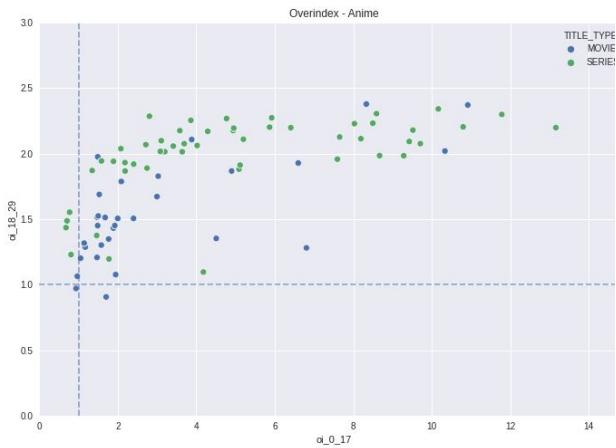
Not great difference between teenagers and young adults.

Teenagers tend to rate movies with votes higher than 9 more than other age groups do.

Comparison Teenagers - Young Adults - Example

Why is one genre good for teenagers? Is it different for young adults?

Anime



Overindex

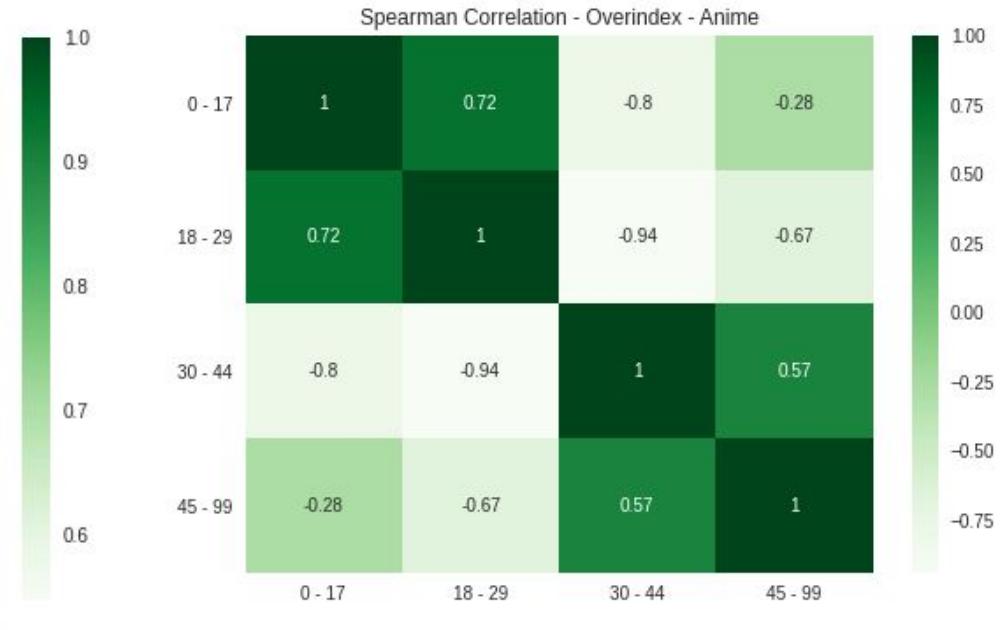
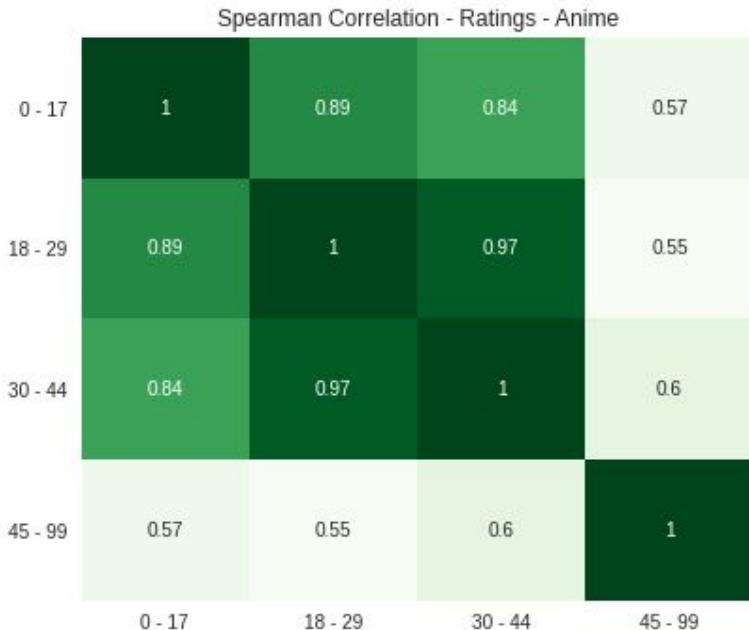
Ratings

Rank Overindex

Comparison Teenagers - Young Adults - Example

Why is one genre good for teenagers? Is it different for young adults?

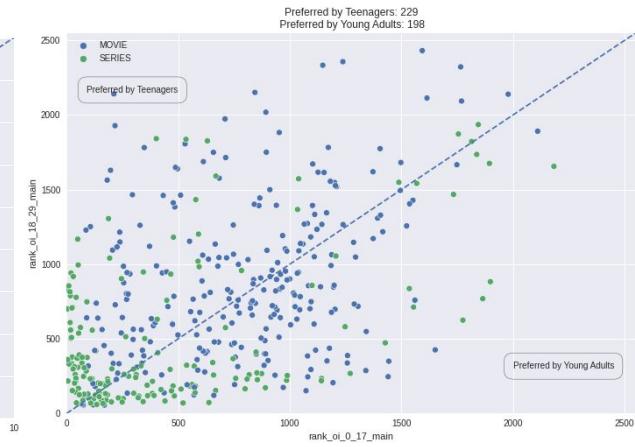
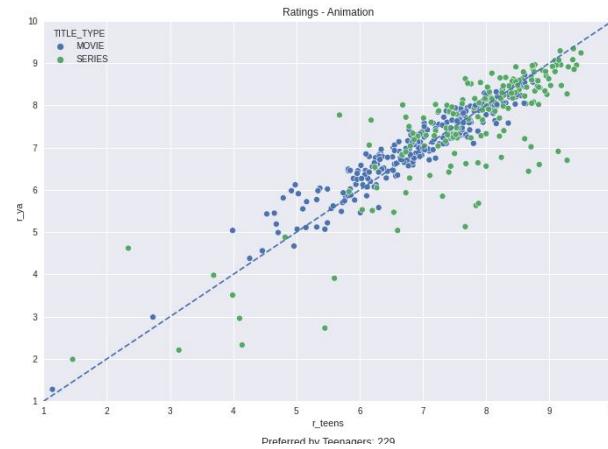
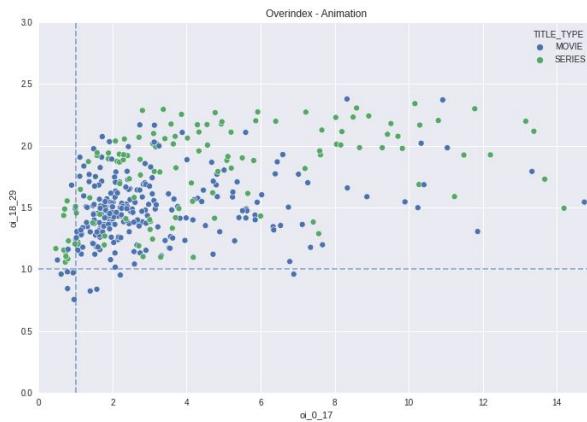
Anime



Comparison Teenagers - Young Adults - Example

Why is one genre good for teenagers? Is it different for young adults?

Animation



Overindex

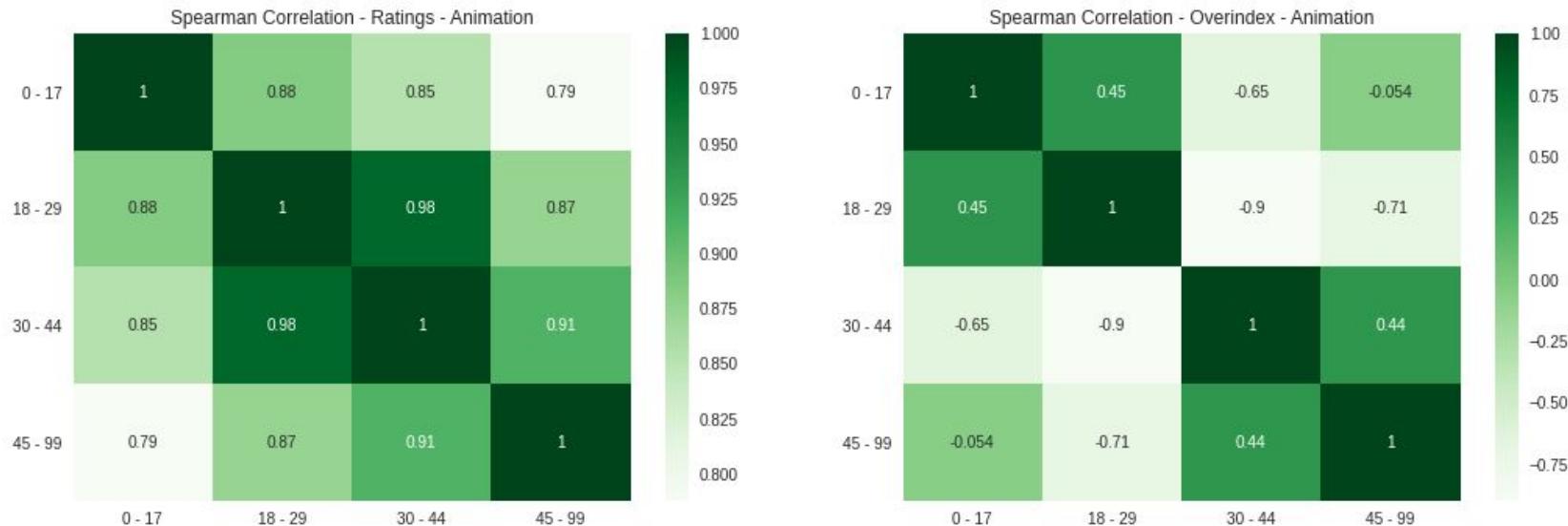
Ratings

Rank Overindex

Comparison Teenagers - Young Adults - Example

Why is one genre good for teenagers? Is it different for young adults?

Animation



Survey - IMDb Validation

Mapping YouTube Categories to IMDb

- Define a set of keywords based on IMDb

```
gaming_keywords = ['video-game', 'arcade-game', 'videogame', 'gaming', 'playstation',
                   'based-on-a-videogame', 'gaming-chair', 'online-gaming', 'video-game-console',
                   'video-game-character', 'video-game-controller', 'video-game-culture',
                   'video-game-in-a-movie', 'video-game-world', 'video-gamer']
```

- Filter IMDb dataset and count number of keywords per title.

- Considering plot and synopsis of each title, compute frequency of keywords.
- Filter whether a title belongs to a category using both the total number of keywords and the number of occurrences in plots

title	IMDb Keywords									num_keywords	gaming	playstation	videogame	tot_plot	tconst	tot
	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain								
Wreck-It Ralph	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain	8.0	2.0	0.0	5.0	7.0	tt1772341	15.0	
Pixels	arcade-game	video-gamer	playing-a-video-game	reference-to-playstation-4	video-game-world	centipede-video-game	video-game-culture	0	7.0	0.0	0.0	5.0	5.0	tt2120120	12.0	
Jumanji: Welcome to the Jungle	video-game-world	arcade-game	video-game-console	0	0	0	0	0	3.0	0.0	0.0	4.0	4.0	tt2283362	7.0	
Casino	0	0	0	0	0	0	0	0	0.0	6.0	0.0	0.0	6.0	tt0112641	6.0	

Survey - IMDb Validation

Mapping YouTube Categories to IMDb

- Define a set of keywords based on IMDb

```
gaming_keywords = ['video-game', 'arcade-game', 'videogame', 'gaming', 'playstation',
                   'based-on-a-videogame', 'gaming-chair', 'online-gaming', 'video-game-console',
                   'video-game-character', 'video-game-controller', 'video-game-culture',
                   'video-game-in-a-movie', 'video-game-world', 'video-gamer']
```

- Filter IMDb dataset and count number of keywords per title.

- Considering plot and synopsis of each title, compute frequency of keywords.
- Filter whether a title belongs to a category using both the total number of keywords and the number of occurrences in plots

title	IMDb Keywords									num_keywords	gaming	playstation	videogame	tot_plot	tconst	tot
	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain								
Wreck-It Ralph	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain	8.0	2.0	0.0	5.0	7.0	tt1772341	15.0	
Pixels	arcade-game	video-gamer	playing-a-video-game	reference-to-playstation-4	video-game-world	centipede-video-game	video-game-culture	0	7.0	0.0	0.0	5.0	5.0	tt2120120	12.0	
Jumanji: Welcome to the Jungle	video-game-world	arcade-game	video-game-console	0	0	0	0	0	3.0	0.0	0.0	4.0	4.0	tt2283362	7.0	
Casino	0	0	0	0	0	0	0	0	0.0	6.0	0.0	0.0	6.0	tt0112641	6.0	

Survey - IMDb Validation

Mapping YouTube Categories to IMDb

- Define a set of keywords based on IMDb

```
gaming_keywords = ['video-game', 'arcade-game', 'videogame', 'gaming', 'playstation',
                   'based-on-a-videogame', 'gaming-chair', 'online-gaming', 'video-game-console',
                   'video-game-character', 'video-game-controller', 'video-game-culture',
                   'video-game-in-a-movie', 'video-game-world', 'video-gamer']
```

- Filter IMDb dataset and count number of keywords per title.

- Considering plot and synopsis of each title, compute frequency of keywords.
- Filter whether a title belongs to a category using both the total number of keywords and the number of occurrences in plots

title	IMDb Keywords									num_keywords	gaming	playstation	videogame	tot_plot	tconst	tot
	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain								
Wreck-It Ralph	arcade-game	video-gamer	video-game-character	video-game-glitch	video-game-world	video-game-characters	video-game-in-a-movie	video-game-villain	8.0	2.0	0.0	5.0	7.0	tt1772341	15.0	
Pixels	arcade-game	video-gamer	playing-a-video-game	reference-to-playstation-4	video-game-world	centipede-video-game	video-game-culture	0	7.0	0.0	0.0	5.0	5.0	tt2120120	12.0	
Jumanji: Welcome to the Jungle	video-game-world	arcade-game	video-game-console	0	0	0	0	0	3.0	0.0	0.0	4.0	4.0	tt2283362	7.0	
Casino	0	0	0	0	0	0	0	0	0.0	6.0	0.0	0.0	6.0	tt0112641	6.0	

Survey - IMDb Validation

YouTube Categories on IMDb

Gaming

- Highest overindex for both age groups
- Ratings are slightly worse for young adults - mean below 7
- Only 8 titles were mapped as gaming

Music

- Music / Musical are some of the genres with the highest overindex and ratings on IMDb too

NGOs

- Lack of interest for both age groups
- Also the ratings are low - mean ~6.5
- Few titles were mapped as NGOs

Sport

- IMDb data shows that young adults are more interested
- Survey results showed that new sports content on APV is requested mainly by young adults

Comedy

- Survey results show that young adults are more interested than teenagers
- Instead, IMDb overindex is slightly higher for teenagers
- The average rating is below 7 for both categories

Survey - IMDb Validation

Film & Animation Categories on IMDb

Survey Results

- Action
- Romance
- Thriller
- Comedy
- SciFi
- **Animation**
- **Anime**
- **Growing Up**
- Criminal



IMDb Comparison

- All the categories proposed in the survey tend to engage with both teenagers and young adults
- Bold categories are the ones that have best overindex and ratings among them
- **LGBT** was not included in the survey, but it tends to have a high overindex and good ratings.

Actors Analysis

Does the age of the cast influence the success of a title?

Assumption

The age of the cast plays a role in the engagement with Teenagers

Analysis

- Computed median cast age for each title
- Both linear and polynomial regression models, considering overindex and ratings for both Teenagers and Young Adults
- We run a different model for each genre to narrow down the possible confounding factors

Results

- Most of the results were not statistically significant, which means that age was not enough to explain the dependent variable
- For the few genres for which the models are significant, the explained variation is extremely low
- Based on IMDb data, the assumption **does not hold**.

Actors Analysis

Is there a relation between having stars in the cast and success?

Star Data

- STARMETER

It's a list of the top 100 actors **at the moment** on IMDb.

We don't have older lists, therefore the classification as star is not perfect by any means.

- TOP-10000 Actors

It's a list of the ~10000 best actors voted by the german audience. Again, 10000 actors are way too many to clearly define what a star is.

Analysis

Two different approaches:

- *Consider titles that include stars, how do the KPIs change?*
- *Consider the best titles (i.e. highest overindex), who are the stars that played in most of them?*

None of them lead to significant results.

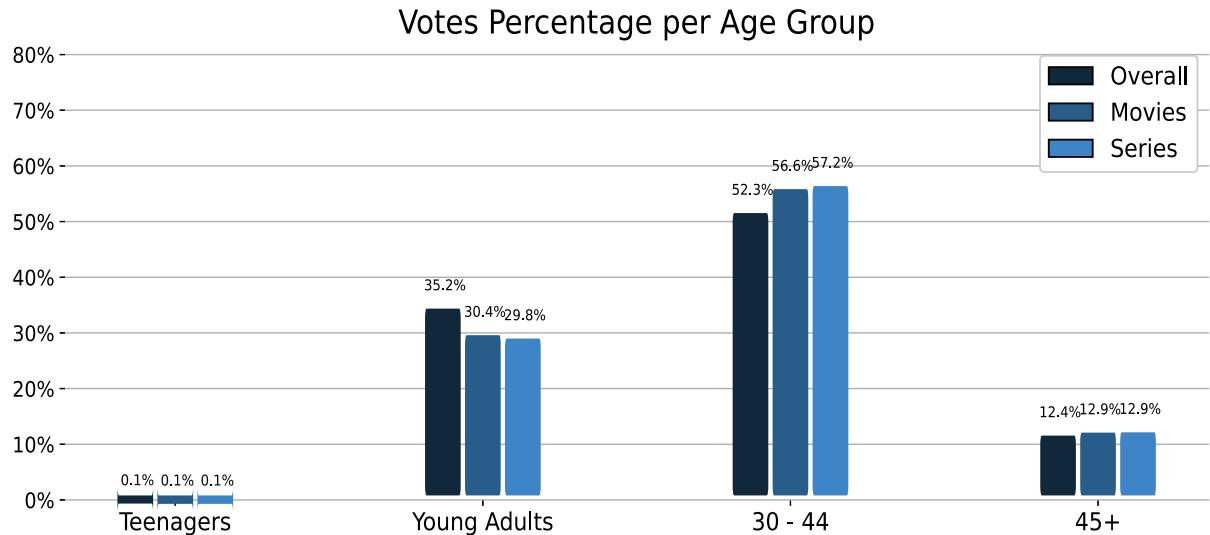
APV Comparison:

- APV Dataset

It's a subset of the IMDb dataset, thus we have the same problem with the data (males and n° of votes teenagers)

It contains 364 titles that can be found on APV

Overindex on APV is lower for teenagers, which validates the assumption of this research project!



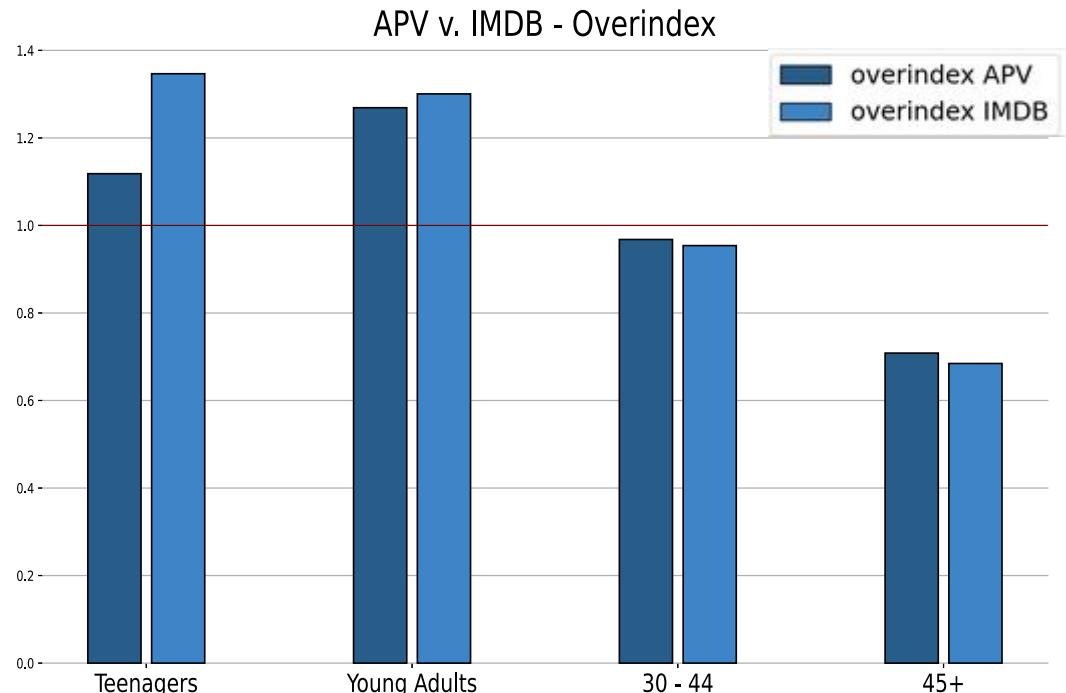
APV Comparison:

- **APV Dataset**

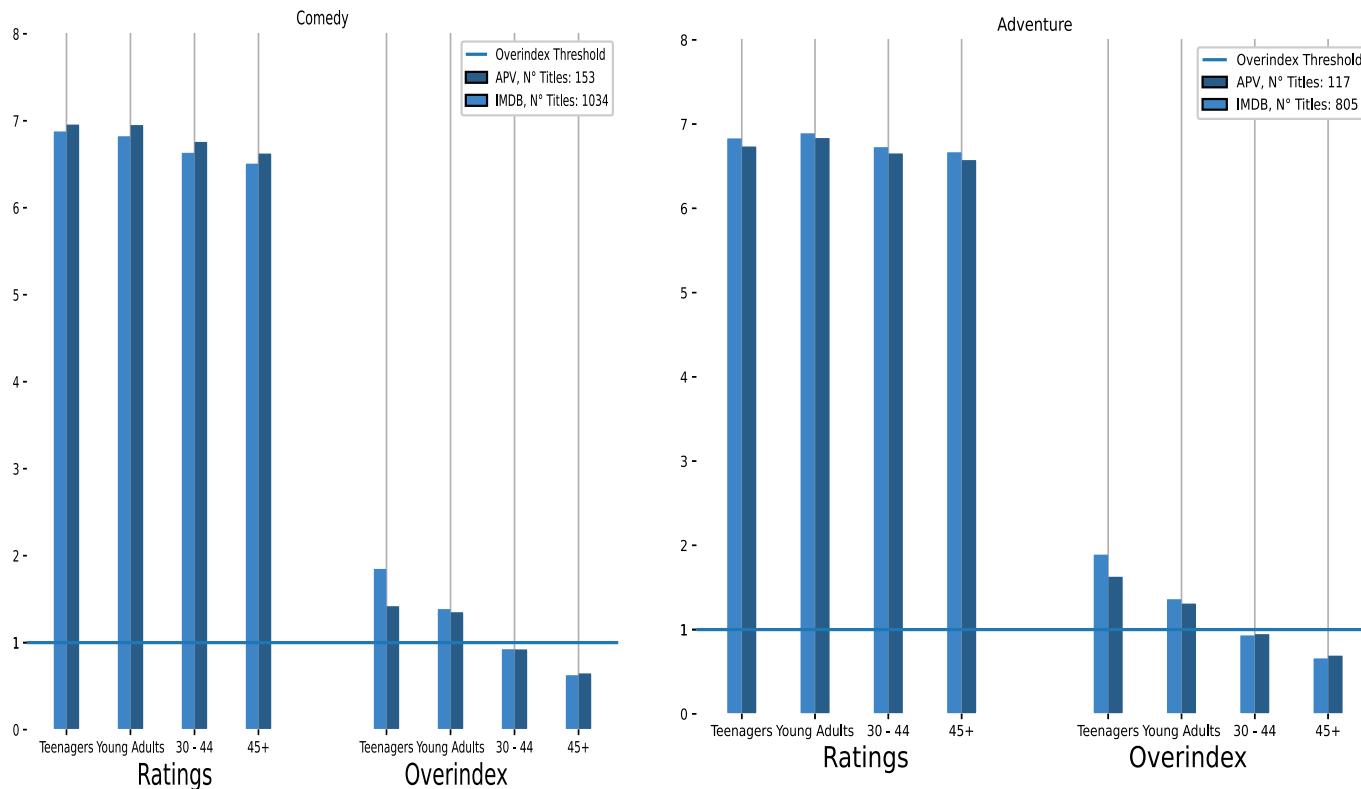
It's a subset of the IMDB dataset, thus we have the same problem with the data (males and n° of votes teenagers)

It contains 364 titles that can be found on APV

Overindex on APV is lower for teenagers, which validates the assumption of this research project!



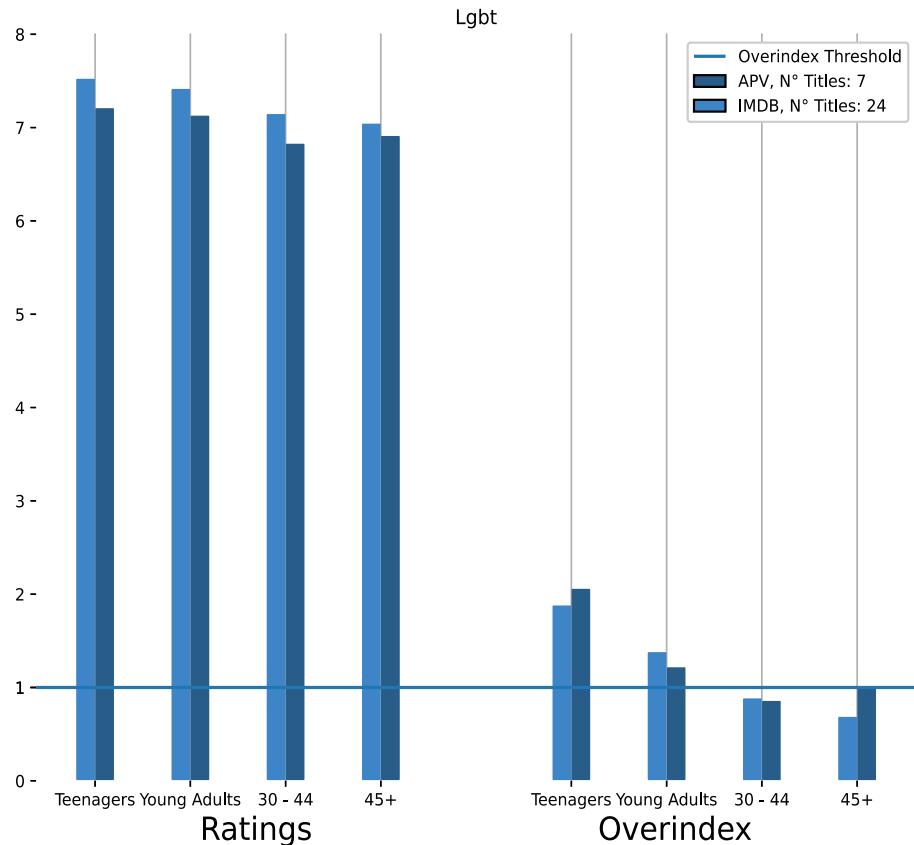
APV Comparison - Focus on Top Genres



Comedy & Adventure

- They are well represented on APV
- However, overindex on IMDb is higher for both teenagers and young adults

APV Comparison - Focus on Top Genres



Results

The following genres performed slightly worse in APV compared to IMDB

- Adventure
- Comedy
- Sci - Fi

LGBT

- It is represented by a few titles, but it shows good results in terms of overindex and ratings

Agenda



- 1 Project Introduction
- 2 Survey: Design & Results
- 3 IMDb Data
- 4 YouTube Data
- 5 Summary & Discussion

- 4.1 YouTube dataset creation/data acquisition
- 4.2 YouTube Premium+Originals analysis
- 4.3 YouTube Channel KPIs, German YouTube analysis
- 4.4 YouTube Case Study: Comedy (Germany)

YouTube Dataset Creation/Data Acquisition

Data from the period January 2021–March 2021

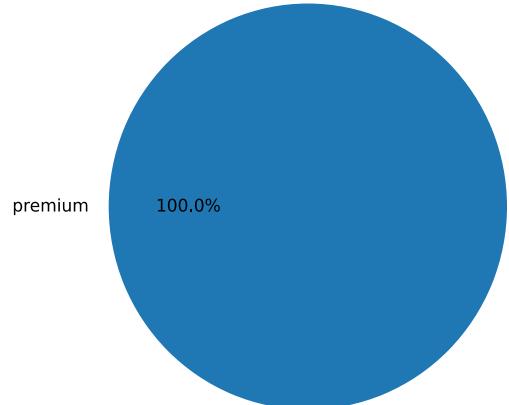
1. YouTube German channel dataset (Noxinfluencer.com)
 - Scrapped German channels in each category from noxinfluencer.com
2. YouTube German channel dataset: Top-level (YouTube API)
 - Acquired channel data from YouTube API using the channel IDs from noxinfluencer.com dataset
 - Collected multiple times: time series data in the period January 2021 - March 2021
3. YouTube German video dataset: Low-level (YouTube API)
 - Acquired video data from YouTube API using the channel IDs of top 100 channels by category from noxinfluencer.com dataset
4. YouTube Premium + Originals dataset (IMDb, manual annotations, and YouTube API)
 - Scrapped IMDb titles tagged as YouTube Premium
 - Annotated IMDb titles to YouTube links manually
 - Acquired video data from YouTube API using the annotations (video IDs) and merged with IMDb data

YouTube Data Analysis: Premium+Originals

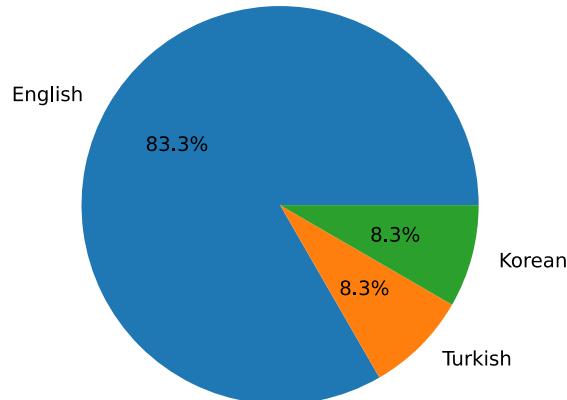
Movies

- Only 12 movies in the dataset
- All accessible only with Premium account, none of them gained huge success
- Almost all in English

YouTube access (manual annotation)



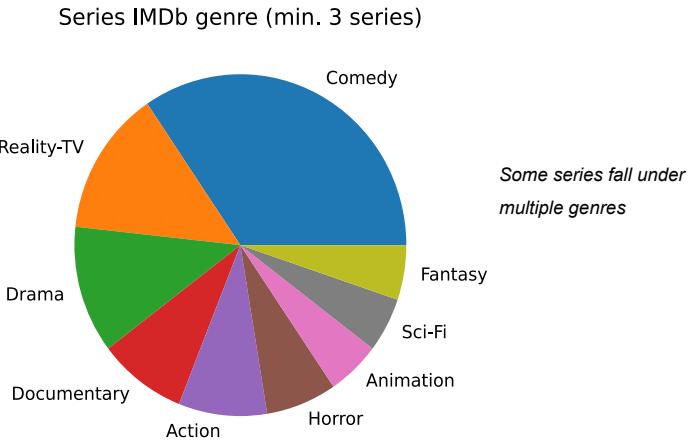
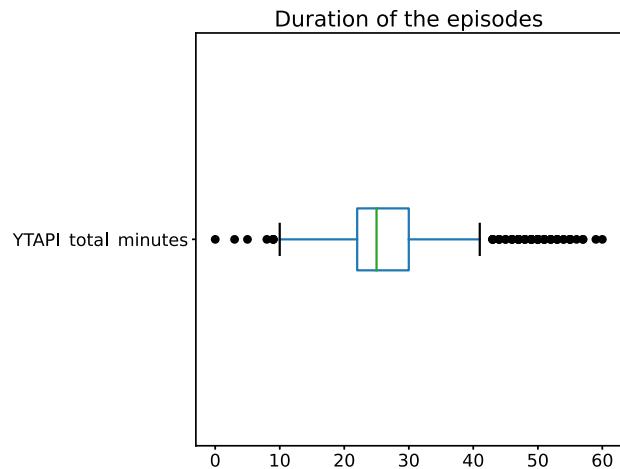
Language of the movies



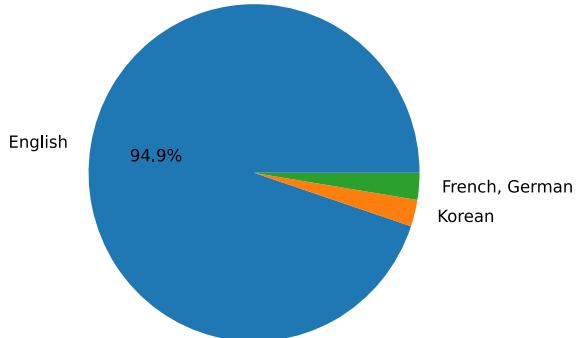
YouTube Data Analysis: Premium+Originals

Series

- 40 series in the dataset (453 episodes)
- ~20-30 min episodes dominate
- Almost all in English



Language of the series



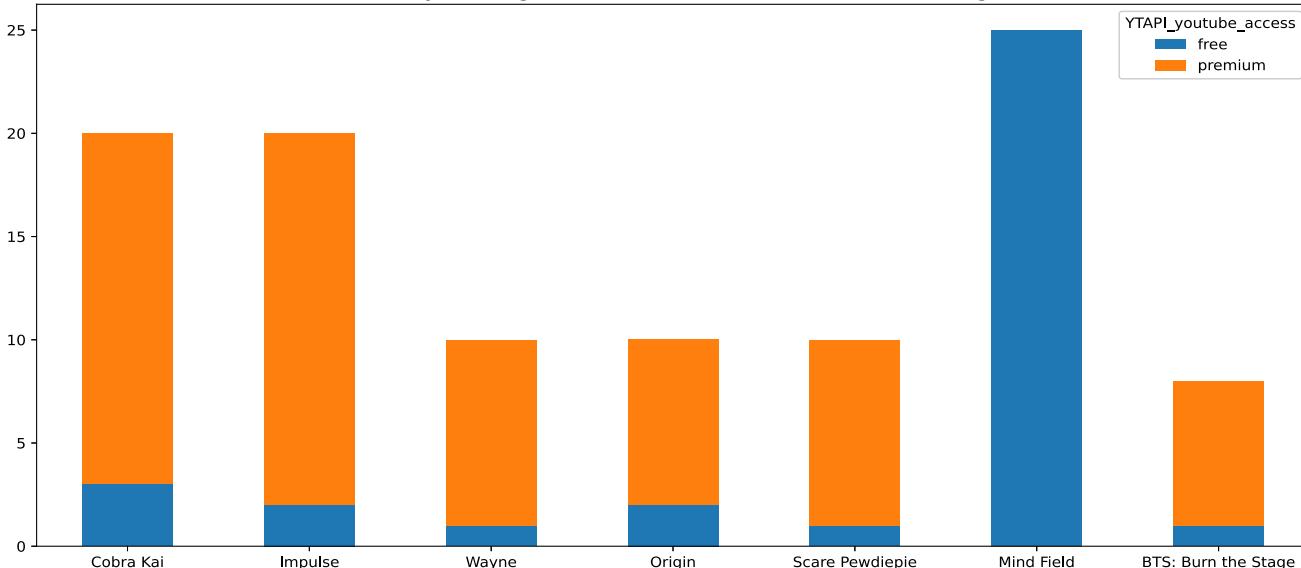
YouTube Data Analysis: Premium+Originals

Series: Amount of free/premium episodes

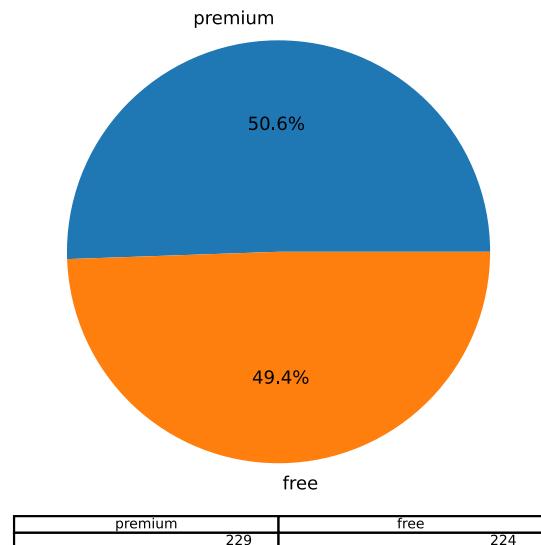
- Almost each premium+originals series have a free episode

Series with at least 1000 IMDb votes: amount of free/premium episodes

Sorted by the highest amount of IMDb votes (left to right)



YouTube access (manual annotation)



YouTube Data Analysis: Channel KPIs

- Key performance indicators of channels using the acquired data (the data contain best YouTube data features that are available to the public)
- These KPIs and their strength/potential:
 - lead to strategy-oriented results and content insights
 - allow to target specific qualities of channels and serve as an analytical basis
 - are applicable to any market and any video category

YouTube Channel KPIs: All qualities measured

General KPIs	absolute totals of views/subs/videos of a channel <i>data source: channel data, top-level, generally available</i>
Growth KPIs	absolute growth (raw numbers) or relative growth (percentage increase) in views or subscribers over time <i>data source: channel data, top-level, computed using time series data</i>
Activity KPIs	number of published videos over time <i>data source: channel data, top-level, computed using time series data</i>
Consistency KPIs	spread of views per video of a channel <i>data source: video data, low-level, generally available</i>
Engagement KPIs	users' engagement with the videos, number of conversions (like/comment) divided by views of a channel <i>data source: video data, low-level, generally available</i>

YouTube Channel KPIs

General

absolute totals of views/subs/videos of a channel

- Total view count
- Total subscriber count
- Total video count
- Ranks of total view count and total subscriber count

YouTube Channel KPIs

Absolute Growth

raw numbers of differences in views or subscribers over time

- MNV = Moving mean New Views [biweekly/monthly]
- MNS = Moving mean New Subscribers [biweekly/monthly]
- Ranks of MNV and MNS
- Combined rank of MNV and MNS = rank MNV + rank MNS

Big channels would dominate

=> combine with a filtering using other metrics/features (for a more focused insight)

YouTube Channel KPIs

Relative Growth

percentage increase of differences in views or subscribers over time

- MPIV = Moving mean Percentage Increase Views [biweekly/monthly]
- MPIS = Moving mean Percentage Increase Subscribers [biweekly/monthly]
- Ranks of MPIV and MPIS
- Combined rank of MPIV and MPIS = rank of MPIV + rank of MPIS

Small channels would dominate

=> combine with a filtering using other metrics/features (for a more focused insight)

YouTube Channel KPIs

Activity

number of published videos over time

- Is channel active?
 - Binary value classified as 1 if the channel's video count increased over a specified time period
- Moving mean new videos [biweekly/monthly]

YouTube Channel KPIs

Consistency

spread of views per video of a channel

- Views per video IQR (interquartile range)

Why IQR?:

- More resistant/robust measure than SD or var
 - Mean absolute deviation can be considered as well, rather than standard deviation where the large outliers could create a huge dispersion
- Covers 50% of the data (from Q1 to Q3), but it gives the gut feeling of dispersion of views in an average video of a channel
- Small channels would dominate (less views) and would have the smallest IQR
=> combine with a filtering using other metrics/features [e.g. views per video on average can be useful in this combination]

YouTube Channel KPIs

Engagement

users' engagement with the videos, number of conversions (like/comment) divided by views of a channel

- Conversion rate on average: likes to views [channel-wise]
- Conversion rate on average: likes to comments [channel-wise]

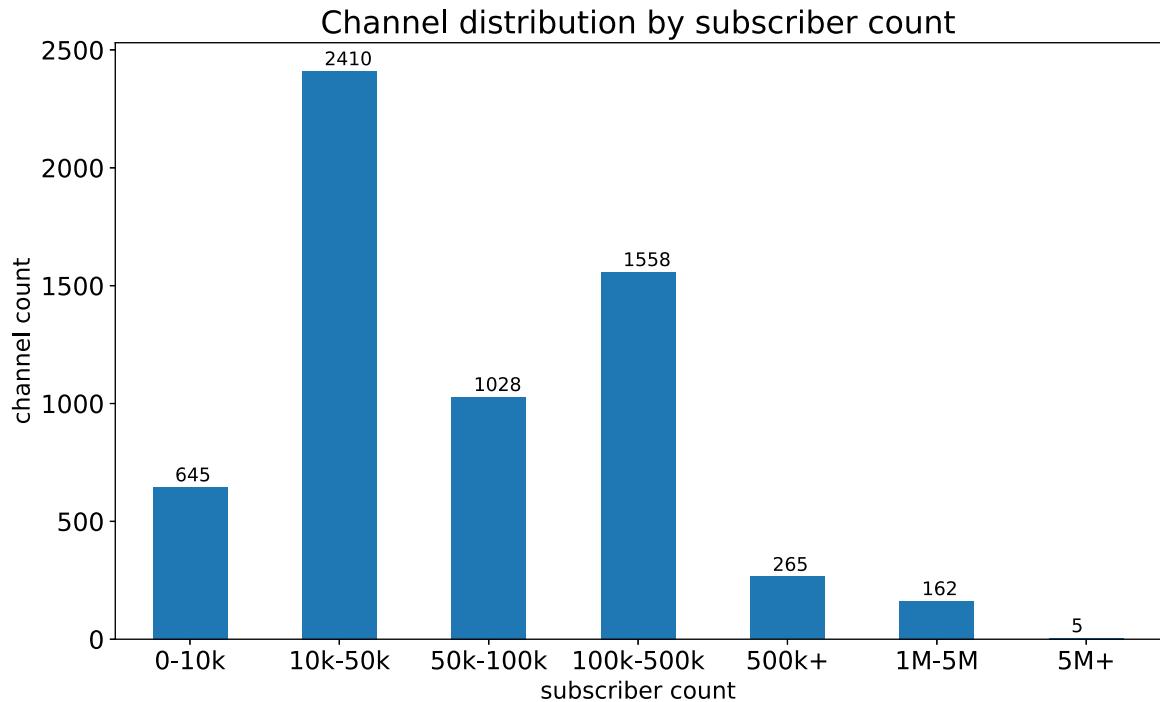
German YouTube Analysis

- Analysis based on datasets that contain data about **Top channels falling under Germany (at Noxinfluencer.com)**
 - Considering only the channels in German language
 - A majority of the channels in open-ended questions from JIM Studie 2020 and our survey are in German language
 - If a channel is in English, it might be a channel with a country set to Germany but in reality a majority of the audience is not German
 - => false positive (for our purpose)
 - Categories at Noxinfluencer.com are the same as the YouTube categories

German YouTube Analysis

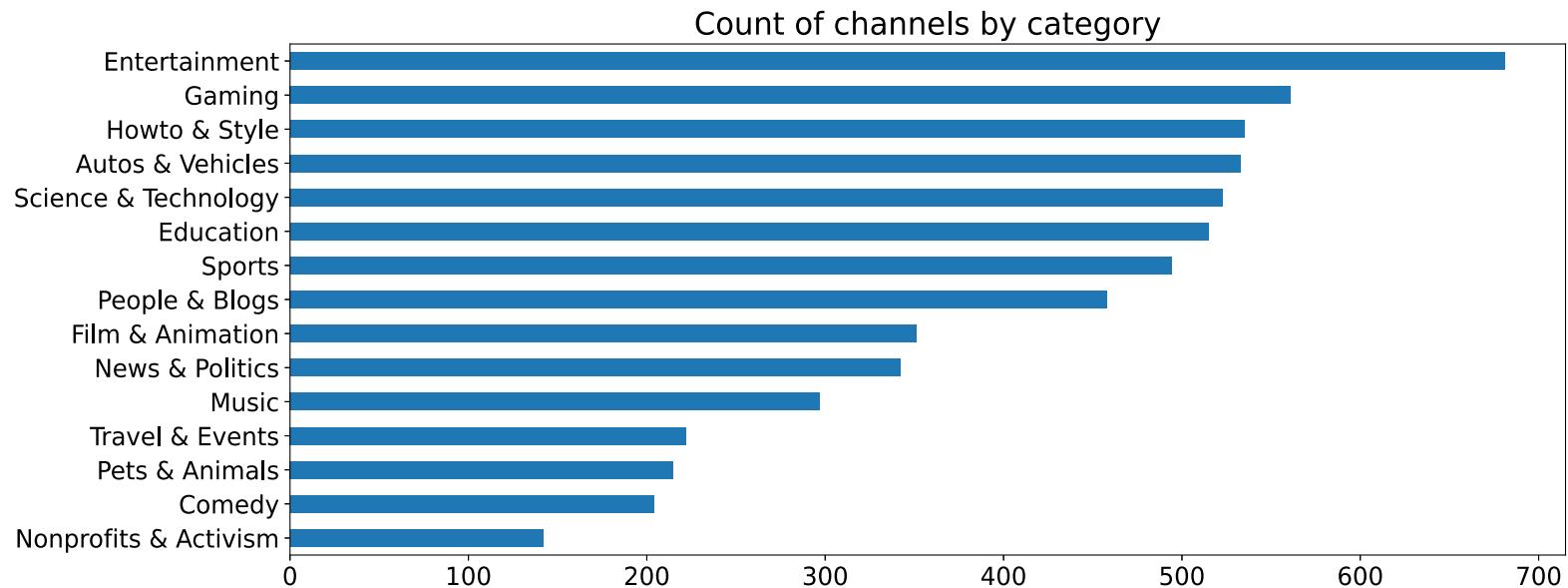
Size of the channels

- Channels of all sizes (in matter of subscribers) are in the dataset



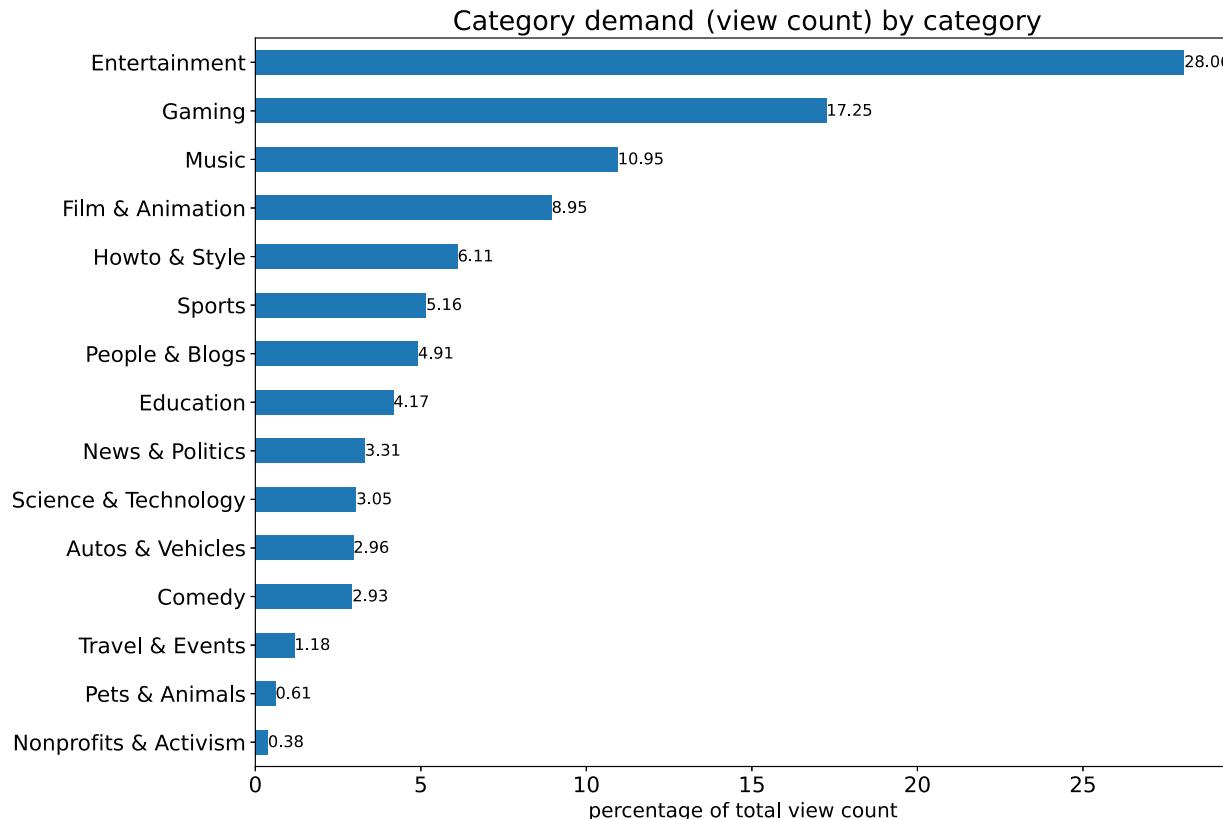
German YouTube Analysis

Category of the channels (data source: Noxinfluencer.com)



German YouTube Analysis: Channels (Top-level)

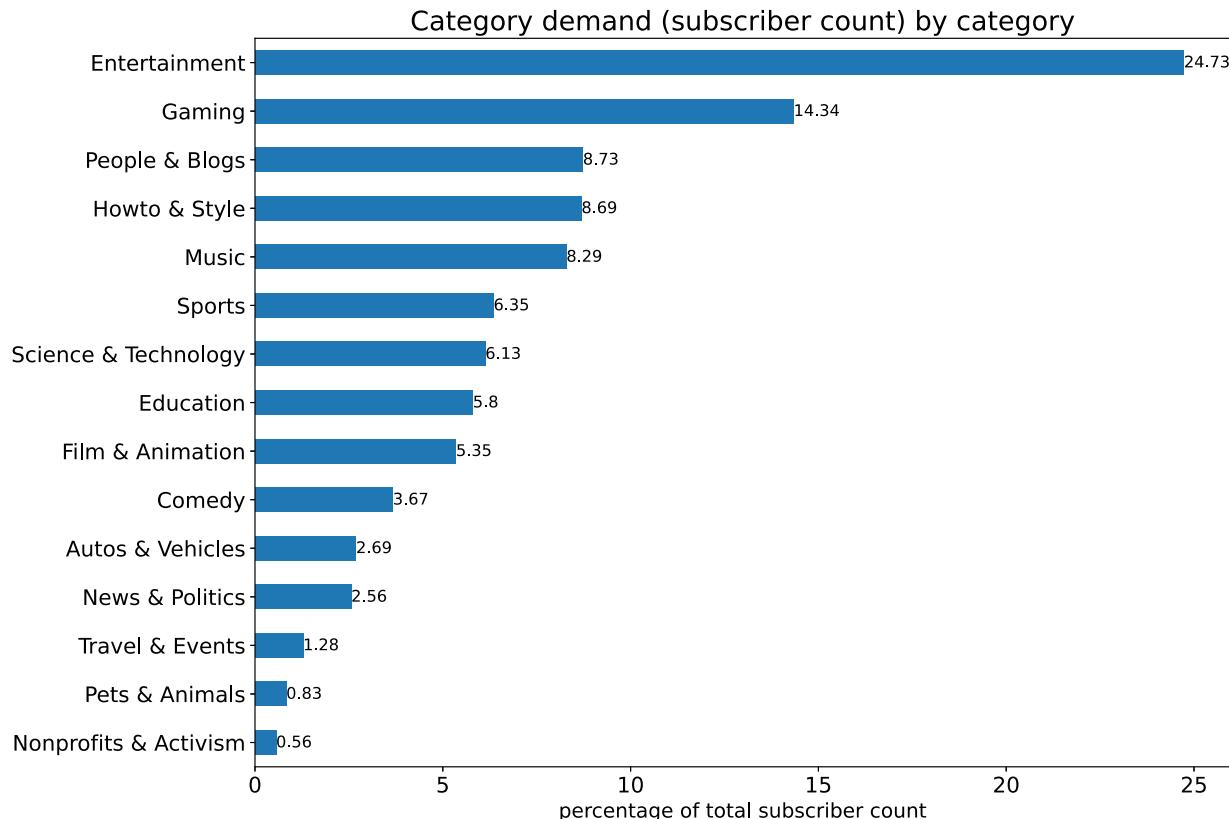
Category demand: Views (top 100 channels for each category)



- **Entertainment and Gaming categories have generated the most views (~45% of total views come from these 2 categories)**

German YouTube Analysis: Channels (Top-level)

Category demand: Subs (top 100 channels for each category)



- **Entertainment and Gaming categories have a lot more subscribers than other categories (~39% of total subscribers come from these 2 categories)**

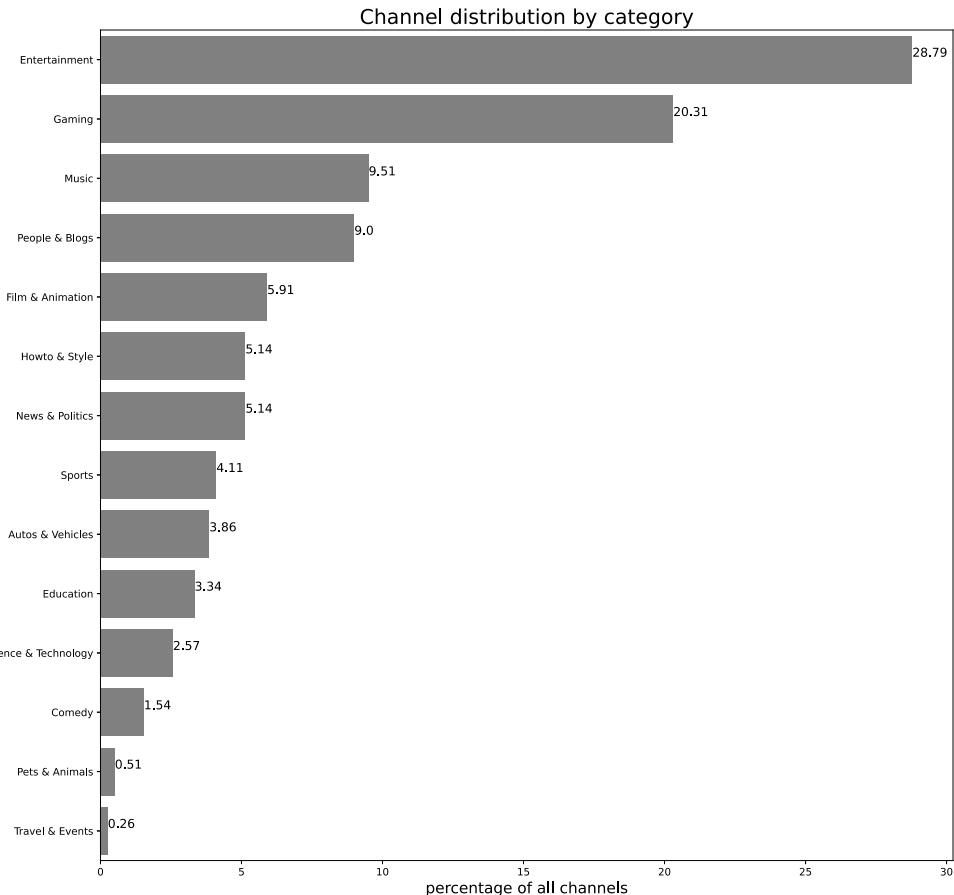
Category distribution considering top 10% ACTIVE channels
using KPI "Combined Rank: Mean New Views + Mean New Subs (Monthly)"

German YouTube Analysis

Channels (Top-level)

Category demand

Active top channel categories



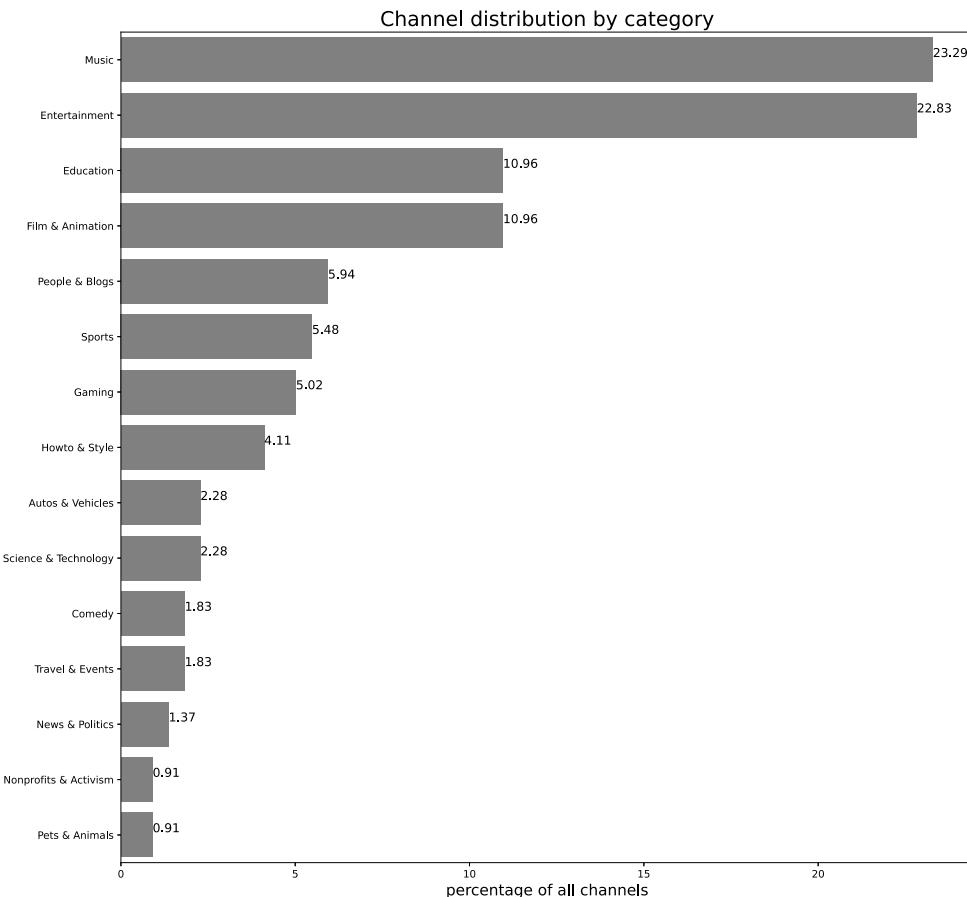
Category distribution considering top 10% INACTIVE channels
using KPI "Combined Rank: Mean New Views + Mean New Subs (Monthly)"

German YouTube Analysis

Channels (Top-level)

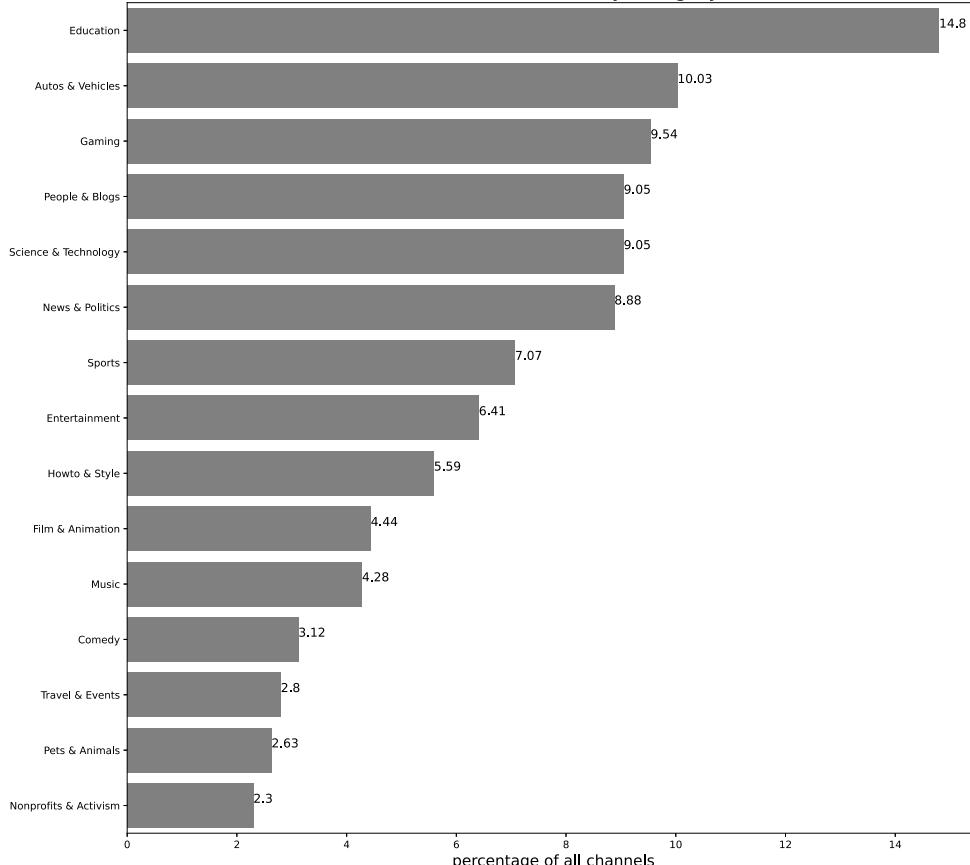
Category demand

Inactive top channel categories



Category distribution considering top 10% channels
using KPI "Combined Rank:
Mean Percentage Increase Views + Mean Percentage Increase Subs (Monthly)"

Channel distribution by category



German YouTube Analysis

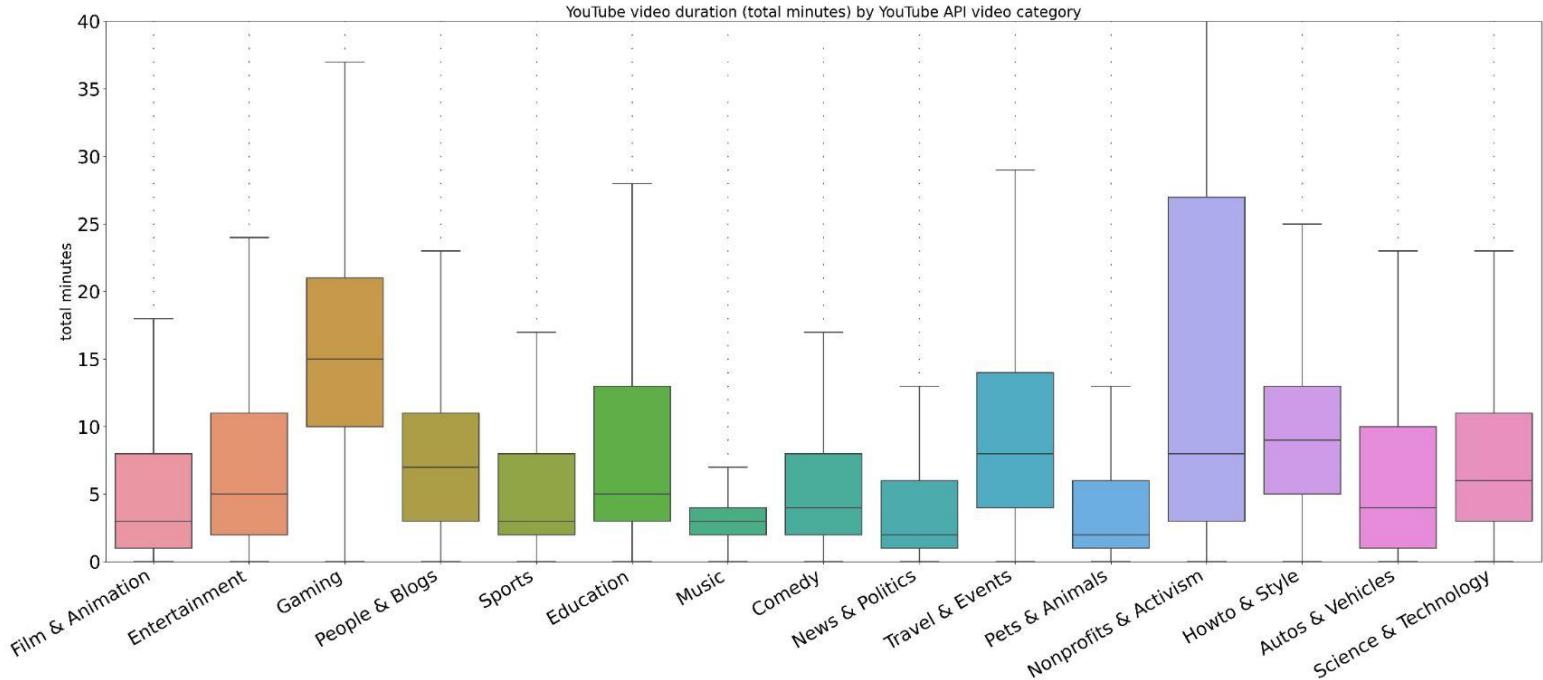
Channels (Top-level)

Category demand

Growing top channel categories

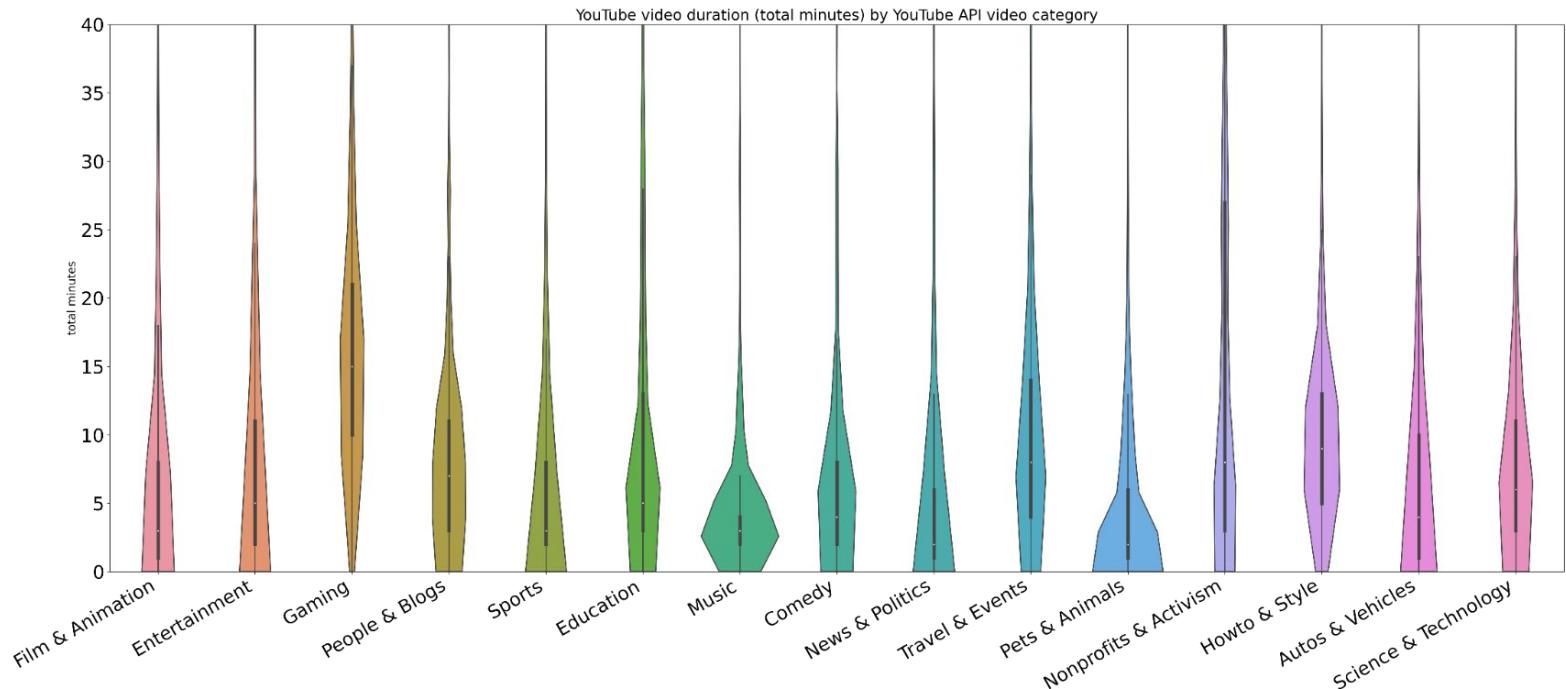
German YouTube Analysis: Videos (Low-level)

Video duration per category



German YouTube Analysis: Videos (Low-level)

Video duration per category



German YouTube Analysis: Videos (Low-level)

Conversion rate by category

- Top 3 median conversion rate of likes to views:
 1. Gaming
 2. Nonprofits & Activism
 3. People & Blogs
- Top 3 median conversion rate of comments to views:
 1. Science & Technology
 2. Gaming
 3. News & Politics

YouTube Case Study: German Comedy

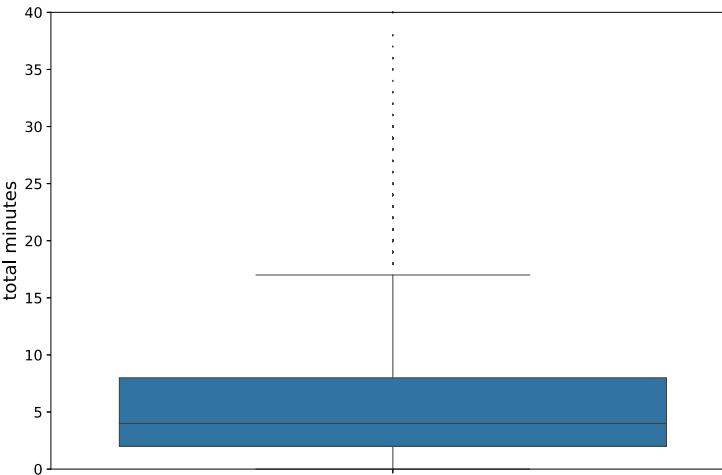
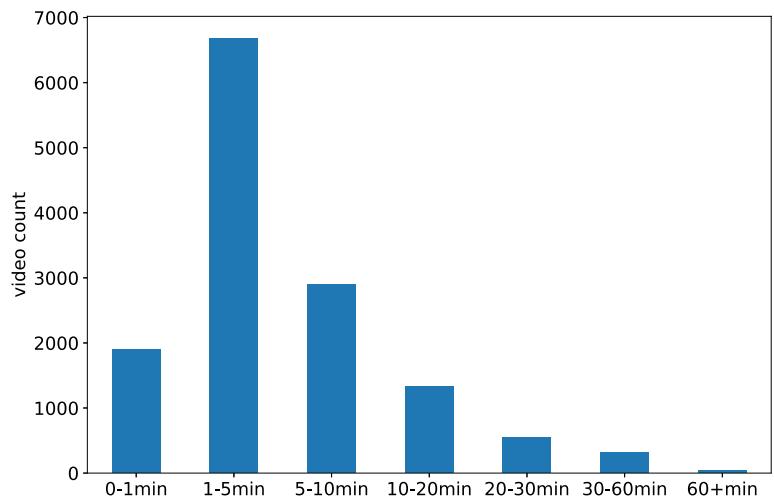
Why Comedy as case study

- Relevant for APV
- There is enough interest for Comedy in our target group
 - according to JIM Studie 2020, 23% (12-19 y.o.) watch Comedy on YouTube
 - according to our survey, 38% (14-18 y.o.) find Comedy somewhat or very interesting

YouTube Case Study: German Comedy

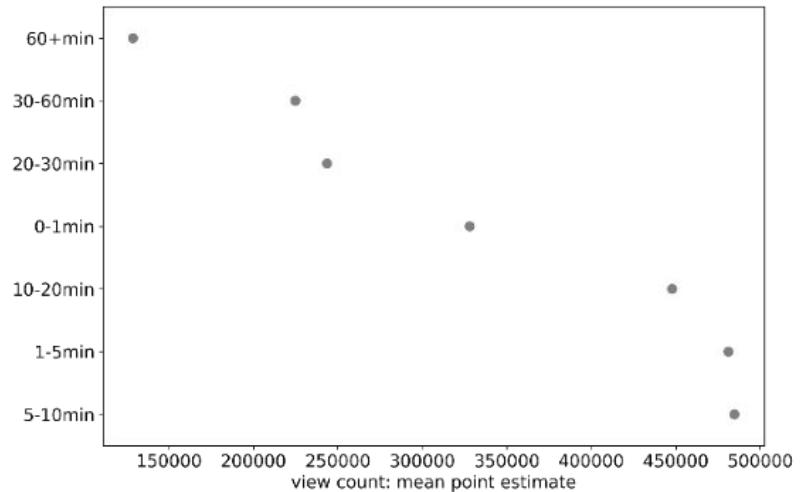
Duration

Comedy video duration



YouTube Case Study: German Comedy

Views by duration category



- Mean view count is ~2x lower for videos longer than 20 min [20-30, 30-60, 60+] than for shorter videos [5-10, 1-5, 10-20]

YouTube Case Study: German Comedy channels with specific qualities

Data from the period January 2021–March 2021, moving means are based on a 2-month window

- Top 10 **active** channels
 - Channels that generate the most new views and subscribers over a time period while being active
 - **10 channels with smallest combined rank of moving mean new views and subs & channel_activity = 1**
 - ['extra 3', 'Varion', 'ZDF heute-show', 'Freshtorge', 'Younes Jones', 'ZDF Comedy', 'KalkTV', 'Der Echte', 'Pumping mnky', 'sinDC']
- Top 10 **inactive** channels
 - Channels that generate the most new views and subscribers over a time period while being inactive
 - **10 channels with smallest combined rank of moving mean new views and subs & channel_activity = 0**
 - ['Die dreisten Drei', 'Mr Bean Deutschland', 'www.nbfotos.de - Bunte Gernsehwelt', 'comedyde', 'RobBubble', 'KNICK KNACK', 'Neues vom Känguru reloaded', 'Witze meisterschaft', 'www.web-tv-produktion.de', 'Tedros Teddy Teclebrhan']
- Top 10 **growing** channels
 - Channels that grow in views and subscribers considerably (when comparing to their previous numbers)
 - **10 channels with smallest combined rank of moving mean percentage increase in views and subs & >10k subscribers**
 - ['Jeyisbaee', 'sinDC', 'Der Echte', 'Matze Knop Entertainment', 'KalkTV', 'Stand-Up 44', 'Tagessau', 'Fabian Köster', 'Schock TV [YTK]', 'No IDea']

YouTube Case Study: German Comedy channels with specific qualities

Data from the period January 2021–March 2021

- Top 10 channels with **consistent success**
 - 10 channels with the lowest dispersion of views per an average video of a channel
 - 10 channels with smallest **video per view IQR & >100k subscribers & >100k views per video on average**
 - ['Comedy Rocket', 'Mr Bean Deutschland', 'Birger', 'NELSON', 'Stand-Up 44', 'Achi der Entertainer', 'ZDF Comedy', 'RobBubble', 'Franky Allstar', 'Fittihollywood']
- Top 10 channels with **best conversion rate**
 - 10 channels with highest **conversion rate on average: likes to views & >100k subscribers & >100k views per video on average**
 - ['RobBubble', 'Fittihollywood', 'JokaH Tululu', 'Birger', 'Varion', 'Franky Allstar', 'NELSON', 'Bodyformus', 'LIONTTV', 'KüsengsTV']

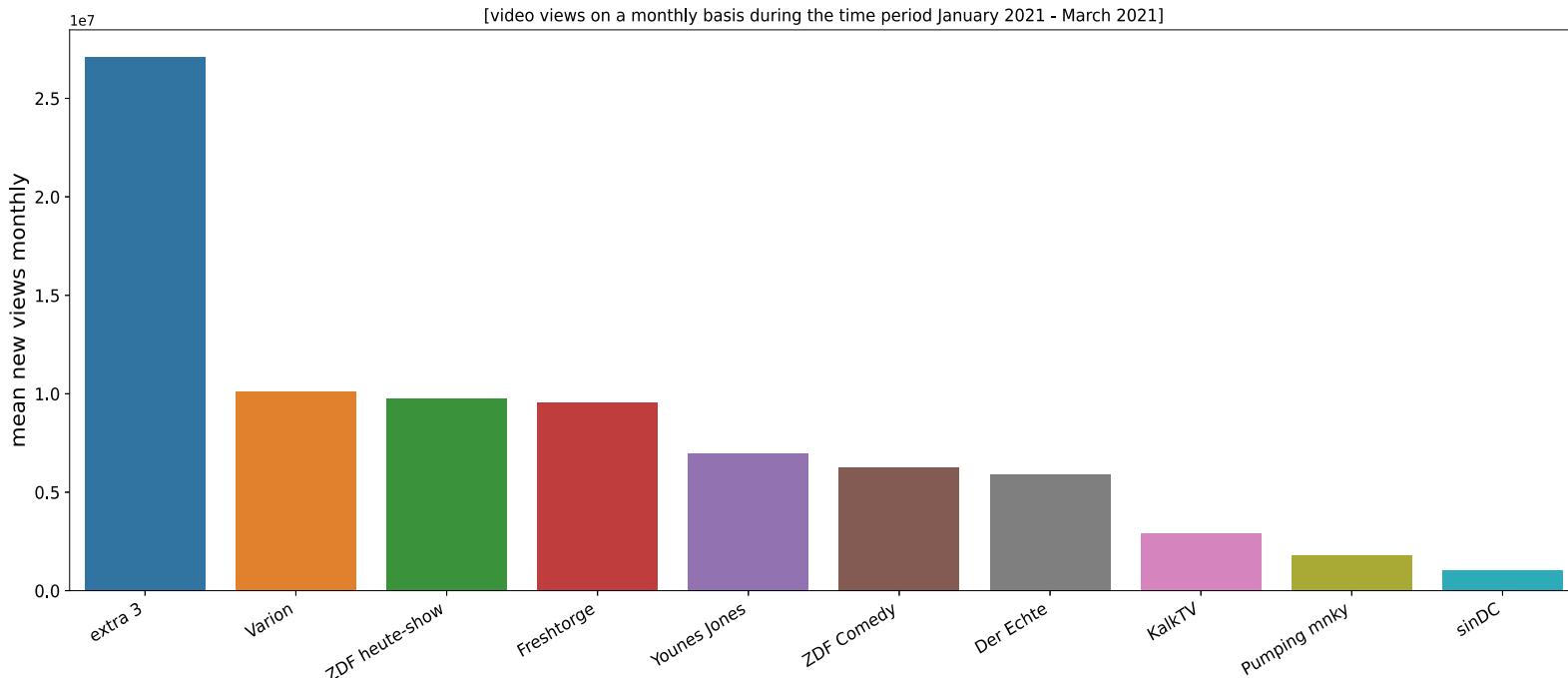
Combination of KPIs/features can be reapplied to any category and tuned.

YouTube Case Study: German Comedy channels with specific qualities

Data from the period January 2021–March 2021

- Example: Overview of Top 10 **active** channels (looking at each channel separately): **Moving mean New Views**

Views of Top 10 active comedy German channels



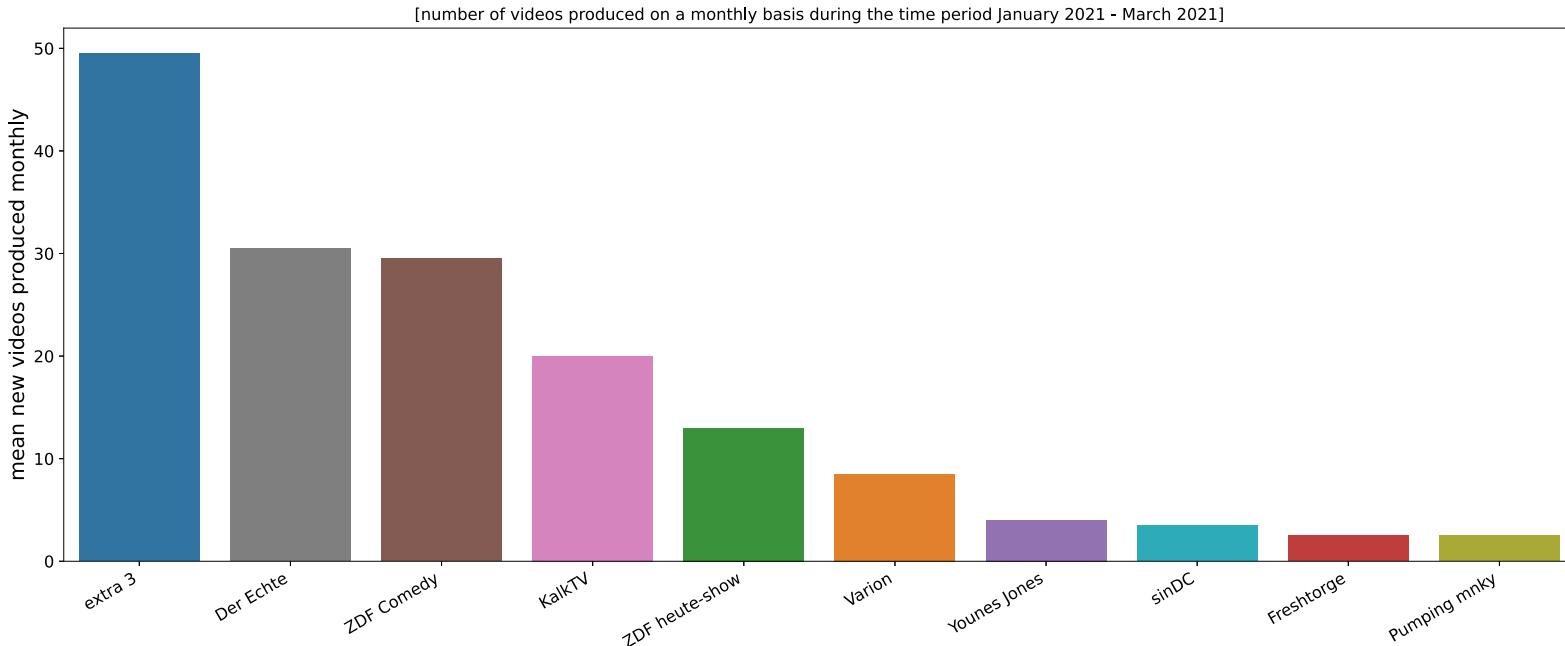
YouTube Case Study: German Comedy channels with specific qualities

Data from the period January 2021–March 2021

- Example: Overview of Top 10 **active** channels (looking at each channel separately):

Publishing rate

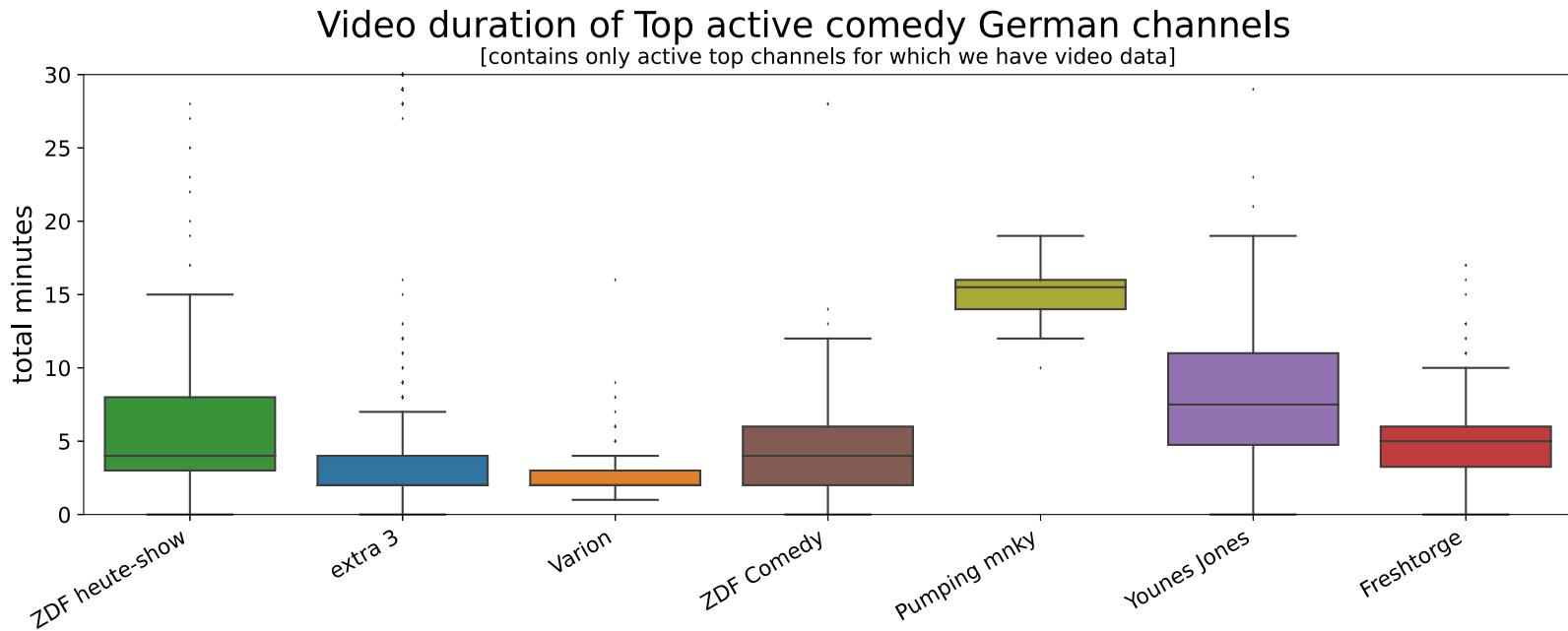
Publishing rate of Top 10 active comedy German channels



YouTube Case Study: German Comedy channels with specific qualities

Data from the period January 2021–March 2021

- Example: Overview of Top 10 **active** channels (looking at each channel separately): **Video duration**





- 1 Project Introduction
- 2 Survey: Design & Results
- 3 IMDb Data
- 4 YouTube Data
- 5 **Summary & Discussion**

Summary: Survey



- APV has high awareness, but around 20% of users do not know that it is included in their Prime Membership
- Teens are an active user group of APV (27% of Teens reported to have their own accounts)
- Teens watch APV slightly more than YA, while YA watch Youtube more than Teens.
- **Top 5 categories:** Teens - Music 71 %, Gaming 53 %, Film & Animation 49 %, Sports 44%, Animals & Pets 40 %. Young adults: Music 64 %, Gaming 46 %, Film & Animation 44%, Comedy 40 %, Sports 38%.
- **Gaming, Cooking (How To subcategory), People & Blogs, Animals & pets, Sports** are significantly more consumed on Youtube & reported to be found on APV less → the categories to increase on APV
- Teens follow most actively influencers in the areas of **Gaming, Music, Film & Animation, Tik tok & Celebrities**.
- Only 10% of Teens are very interested in seeing a **documentary** about their favorite influencer, another 22% are only somewhat interested.

Summary: IMDb Data Analysis



Actors Analysis

- Given IMDb data and Starmeter data, it is not possible to conclude that any specific actor has an impact as an influencer on teenagers engagement / title success.
- It is also not possible to conclude that the age of the cast plays a meaningful role for the under 30 audience.

Teenagers vs Young Adults

- There is not a clear distinction between teenagers and young adults. Even though for some genres, i.e. animation, the difference is wider, in general this is not a clear trend.
- On the other hand the distinction Under30 and Over30 seems to make a big difference.

Survey Validation

- IMDb validates the fact that gaming is the most appealing genre for teenagers, having the highest overindex and high ratings.
- IMDb data suggests a smaller difference between the behaviors of the two age groups when compared to the survey.

Summary: YouTube Data

Content Insights

- **YouTube Premium has a lack of premium content** + the premium content did not gain a huge success. The series that are relevant for YouTube Premium mostly contain some free episodes.
- **Using the duration analysis, we can make important conclusions for content acquisition and relevancy of a category for APV platform.**
 - Audience of categories with a potential towards longer videos (10+min):
 - Gaming (more than 75% of videos is 10+min), Travel & Events, Nonprofits & Activism
- **Entertainment and Gaming have the highest demand** (in views & subs)
- **Gaming has the highest conversion rate** (of likes to views)
- **Ultimate recommendation for APV based on YouTube data:**
 - Analyze Gaming category and the channels on the German market; data & tools for data collection and analysis (from this project) can be easily reused for this purpose
 - Gaming category has the highest potential for user acquisition => Produce content with people from Gaming area or content about a game that is thriving on German YouTube

Strategy KPIs

- **Time series data about YouTube channels enhance the analytical basis and allow us to reach strategic objectives.** [example: using the time series KPIs from this project, we can target selected channel qualities and analyze a specialized segment of the market (e.g. active channels that are recently growing and have a high conversion rate)]

References

1. bpb.de. 2021. *Bevölkerung nach Bundesländern* | bpb. [online] Available at: <<https://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61535/bevoelkerung-nach-laendern>> [Accessed 11 April 2021].
2. Brosius, H., Haas, A. and Koschel, F., (2008). *Methoden der empirischen Kommunikationsforschung*. 4th ed. Wiesbaden: VS Verlag für Sozialwissenschaften, pp.92-120.
3. Budzinski, O., Gaensle, S., & Lindstädt, N. (2020). The Battle of YouTube, TV and Netflix—An Empirical Analysis of Competition in Audio-visual Media Markets. *TV and Netflix—An Empirical Analysis of Competition in Audio-visual Media Markets* (March 31, 2020).
4. Chen, W. H. (2020). Leverage social filtering to enhance viewing experience with Amazon Prime Video.
5. Hwang, Kumju & Zhang, Qi. (2018). Influence of Parasocial Relationship Between Digital Celebrities and Their Followers on Followers' Purchase and Electronic Word-of-Mouth Intentions, and persuasion knowledge. *Computers in Human Behavior*. 87. 10.1016/j.chb.2018.05.029.
6. Hwang, T. G., Park, C. S., Hong, J. H., & Kim, S. K. (2016). An algorithm for movie classification and recommendation using genre correlation. *Multimedia Tools and Applications*, 75(20), 12843-12858.
7. Lad, A., Butala, S., & Bide, P. (2019, November). A Comparative Analysis of Over-the-Top Platforms: Amazon Prime Video and Netflix. In *International Conference on Communication and Intelligent Systems* (pp. 283-299). Springer, Singapore.
8. Loock, K. (2018). American TV series revivals: Introduction. *Television & New Media*, 19(4), 299-309.
9. Martín-Quevedo, J., Fernández-Gómez, E., & Segado-Boj, F. (2019). How to engage with younger users on Instagram: A comparative analysis of HBO and Netflix in the Spanish and US markets. *International journal on media management*, 21(2), 67-87.
10. Mikos, L. (2016). Digital media platforms and the use of TV content: Binge watching and video-on-demand in Germany. *Media and Communication*, 4(3), 154-161.

References

11. Mládková, L. (2017, September). Learning habits of generation Z students. In European Conference on Knowledge Management (pp. 698-703). Academic Conferences International Limited.
12. Perks, L. G. (2015). Media marathoning. Immersions in morality. Lanham, MD: Lexington Books.
13. Rothman, D. (2016). A Tsunami of learners called Generation Z. URL: www.mdle.net/JoumaFA_Tsunami_of_Learners_Called_Generation_Z.pdf.
14. Shetty, p., & Ghate, s. L. (2020). The nostalgic economy: a study on the domination of nostalgic focus in contemporary media and its influence on streaming preferences.
15. Singh, A. P., & Dangmei, J. (2016). Understanding the generation Z: the future workforce. South-Asian Journal of Multidisciplinary Studies, 3(3), 1-5.
16. Tryon, C. (2013). On-demand culture. Digital delivery and the future of movies. New Brunswick, NJ & London: Rutgers University Press.
17. www.statistik-bw.de. 2021. *Demografie: Höchster Anteil unter 18-Jähriger in Baden-Württemberg*. [online] Available at: <<https://www.statistik-bw.de/Presse/Pressemitteilungen/2015085>> [Accessed 11 April 2021].
18. Statista. 2021. *Germany: population by age* | Statista. [online] Available at: <<https://www.statista.com/statistics/454349/population-by-age-group-germany/>> [Accessed 11 April 2021].
19. Ncss-wpengine.netdna-ssl.com. 2021. *Tests for Two Proportions*. [online] Available at: <https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/PASS/Tests_for_Two_Proportions.pdf> [Accessed 11 April 2021].
20. Nielsen.com. 2020. *Which Consumer Attitudes Will Shape the Streaming Wars?*. [online] Available at: <<https://www.nielsen.com/us/en/insights/article/2020/playback-time-which-consumer-attitudes-will-shape-the-streaming-wars/>> [Accessed 11 April 2021].
21. Südwest, M. F. (2020). JIM-Studie. Jugend, Information,(Multi-) Media. [online] Available at: <https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf> [Accessed 11 April 2021].
22. Richtlinie für die Befragung von Minderjährigen