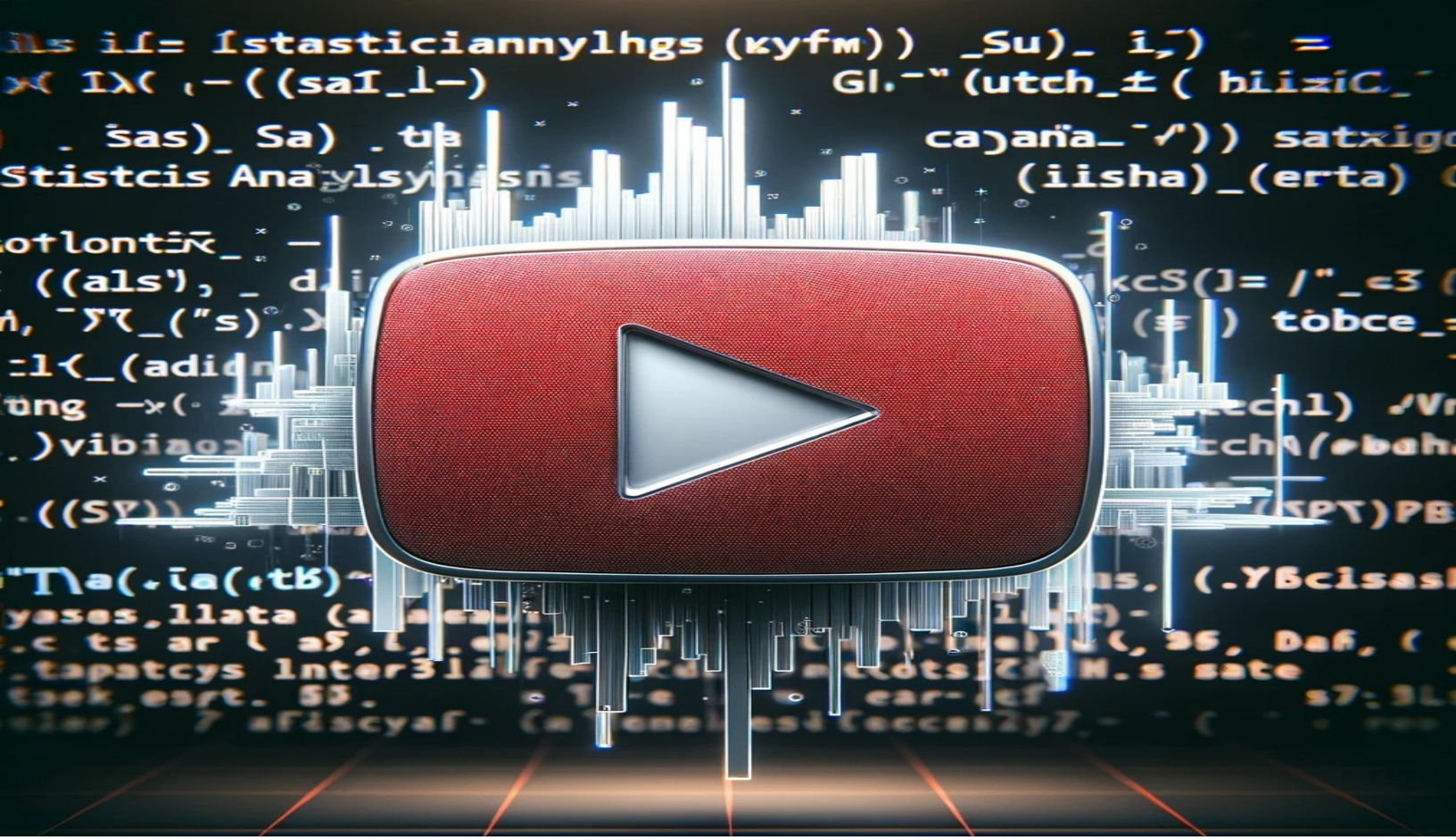


Customer’s preferences regarding YouTube

Prepared by:

Peter Baraka Opiyo,
Margherita Lo Faso,
Simone Assirelli,
Stefano Vaprio

Bologna Business School
Master Data Science and Business Analytics

Table of Contents

Introduction

What we observe in today's society is a growing trend where individuals of various ages dedicate a significant portion of their time to the use of digital platforms such as YouTube, Instagram, TikTok, and Facebook in search for content ranging from music and gaming videos to DIYs and educational clips, to mention a few. According to Forbes, this habit results in the average user devoting nearly four years of their life, out of eighty, to engaging with these digital platforms.

Our study primarily focuses on examining the use of the platform YouTube, which, since its launch in 2005, has grown to account for 50 billion daily views as of February 2023. To put it in perspective "In 2022, YouTube counted over 2.56 billion users accessing its video content worldwide. The platform's user base was composed of more men than women, with around 12% of YouTube's total users being men aged between 25 and 34 years, and approximately 9% being women aged 35 and 44 years. In January 2023, India counted the largest YouTube audience by far - almost 470 million users, followed by the United States with 246 million users on the popular video platform."[1]

Given the wide range of content offered by YouTube, spanning across different categories, our main interest lies in analyzing the types of users frequenting the platform and their preferences regarding the content provided by YouTube. It is noteworthy that YouTube offers a highly personalized experience, as its algorithms work to suggest relevant content based on the specific viewing habits of individual users.

The study opted to investigate eight distinct behavioral characteristics (gender, age, presence of children, occupational status, technological devices usage, YouTube usage, subscriptions, and notification activation) and sixteen opinion questions regarding the platform, of which eight psychographic questions related to respondent's opinion regarding the platform with the question - What is YouTube for you? while the remaining are related to the customers' general opinions regarding several features of the platform (in particular related to ads). The survey queried 521 respondents about the significance of each aspect.

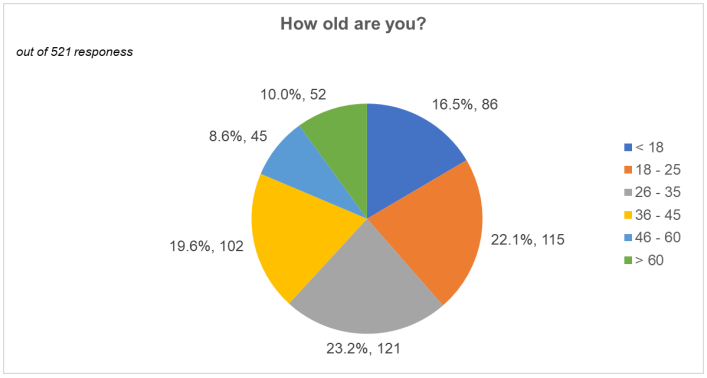
The survey was created using Google Forms in English, distributed digitally, and collected anonymously. Participants were instructed to rate the importance of the personal value of YouTube usage on a scale ranging from 1 to 7, where 1 represented the lowest importance and 7 represented the highest importance for each feature.

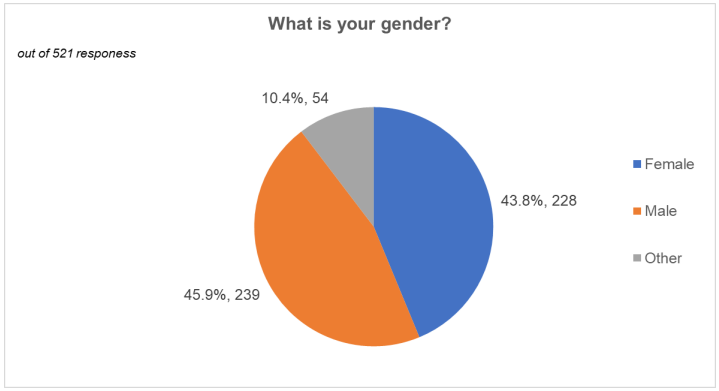
[1] https://www.statista.com/topics/2019/youtube/#topicOverview

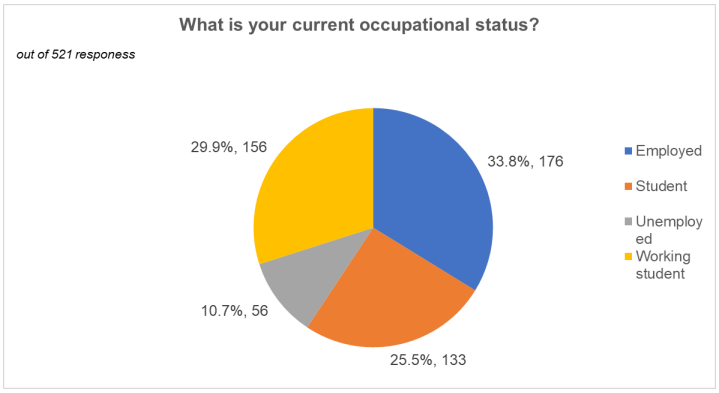
Analysis

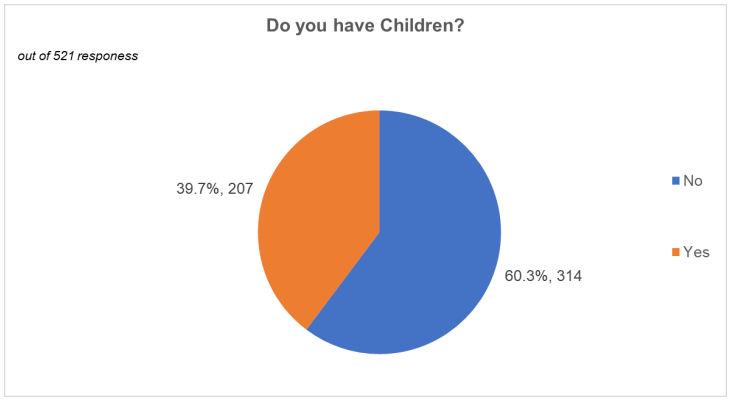
Survey Description

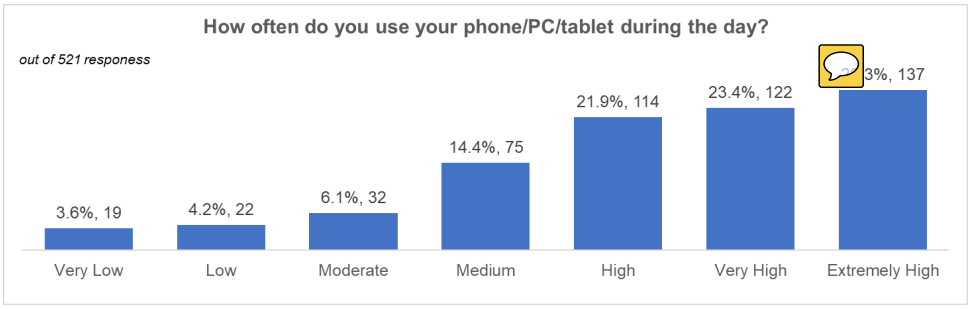
The following graphs serve as a comprehensive summary of our survey findings, providing a visual representation of key insights. Thus, providing a concise overview of the predominant trends within the surveyed categories and offering a whole overview of the available information.

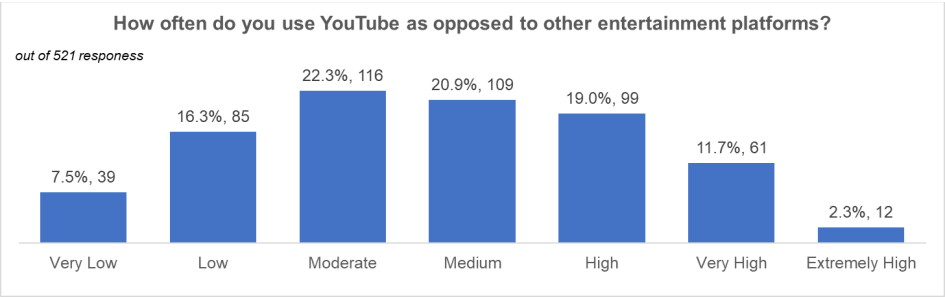


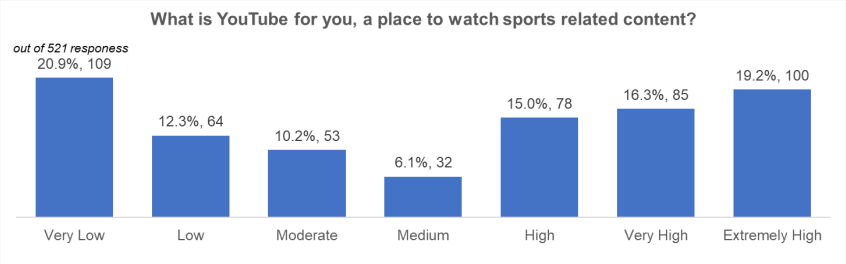


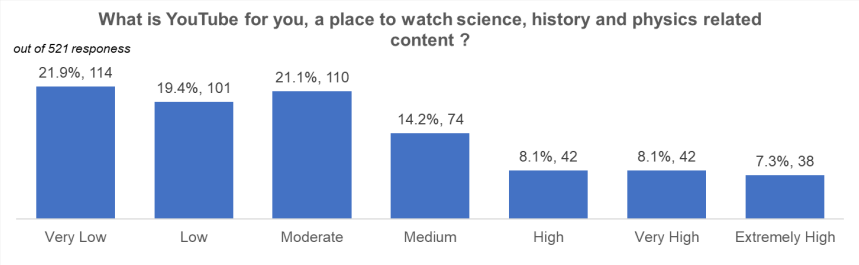


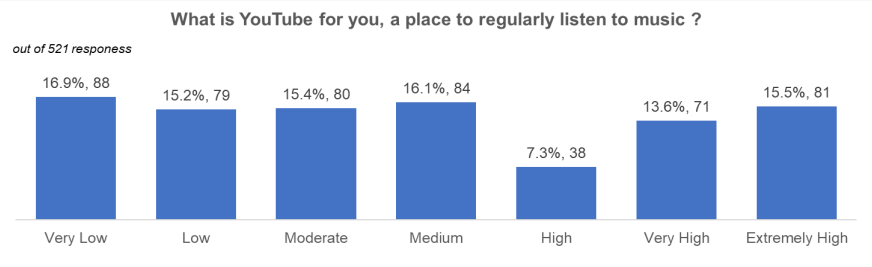


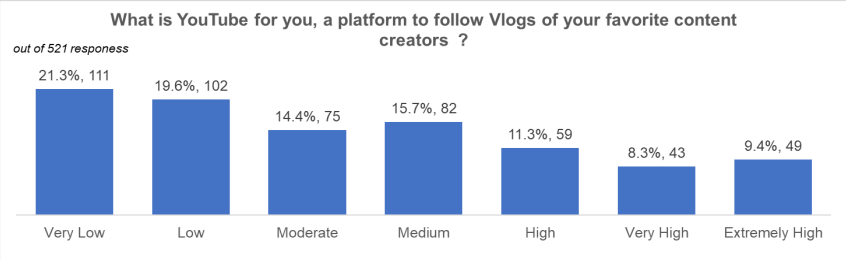


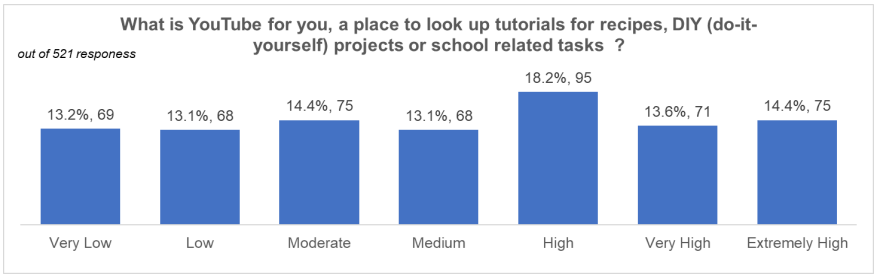

How often do you use YouTube as opposed to other entertainment platforms?
out of 521 responess
7.5%, 39
16.3%, 85
22.3%, 116
20.9%, 109
19.0%, 99
11.7%, 61
2.3%, 12
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

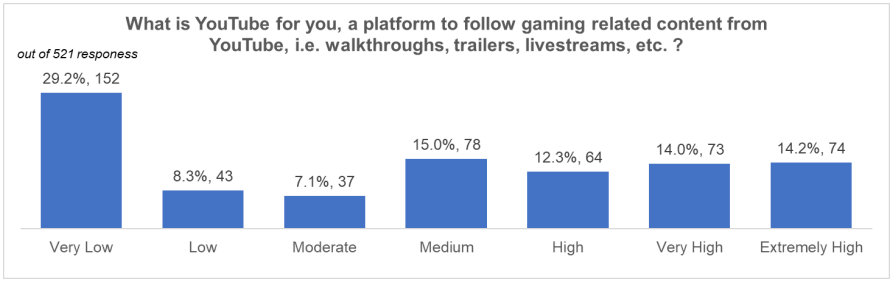

What is YouTube for you, a place to watch sports related content?
out of 521 responess
20.9%, 109
12.3%, 64
10.2%, 53
6.1%, 32
15.0%, 78
16.3%, 85
19.2%, 100
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

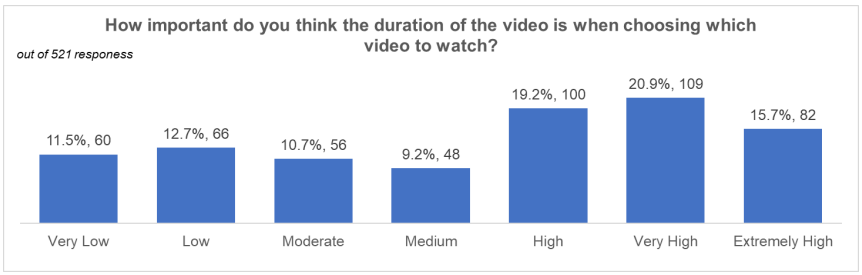

What is YouTube for you, a place to watch science, history and physics related content ?
out of 521 responess
21.9%, 114
19.4%, 101
21.1%, 110
14.2%, 74
8.1%, 42
8.1%, 42
7.3%, 38
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

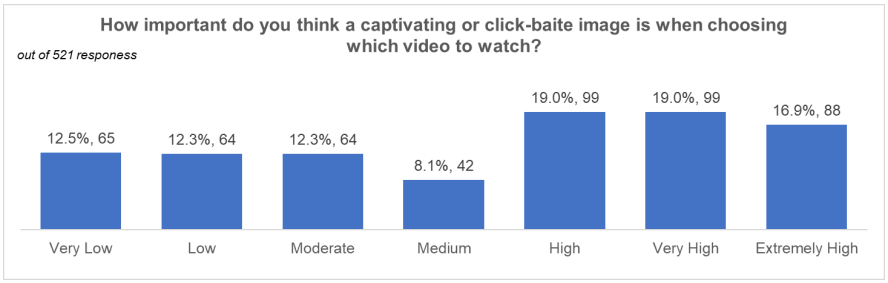

What is YouTube for you, a place to regularly listen to music ?
out of 521 responess
16.9%, 88
15.2%, 79
15.4%, 80
16.1%, 84
7.3%, 38
13.6%, 71
15.5%, 81
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

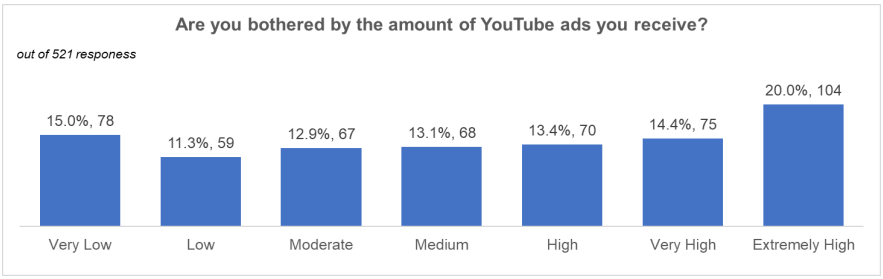

What is YouTube for you, a platform to follow Vlogs of your favorite content creators ?
out of 521 responess
21.3%, 111
19.6%, 102
14.4%, 75
15.7%, 82
11.3%, 59
8.3%, 43
9.4%, 49
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

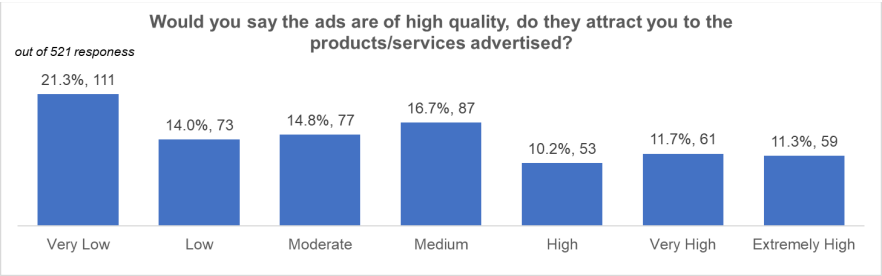

What is YouTube for you, a place to look up tutorials for recipes, DIY (do-it-yourself) projects or school related tasks ?
out of 521 responess
13.2%, 69
13.1%, 68
14.4%, 75
13.1%, 68
18.2%, 95
13.6%, 71
14.4%, 75
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

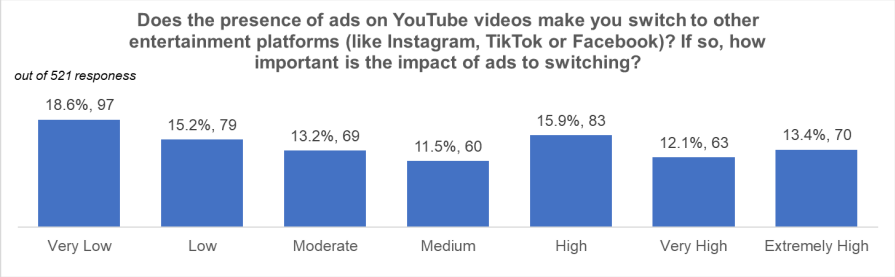

What is YouTube for you, a platform to follow gaming related content from YouTube, i.e. walkthroughs, trailers, livestreams, etc. ?
out of 521 responess
29.2%, 152
8.3%, 43
7.1%, 37
15.0%, 78
12.3%, 64
14.0%, 73
14.2%, 74
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

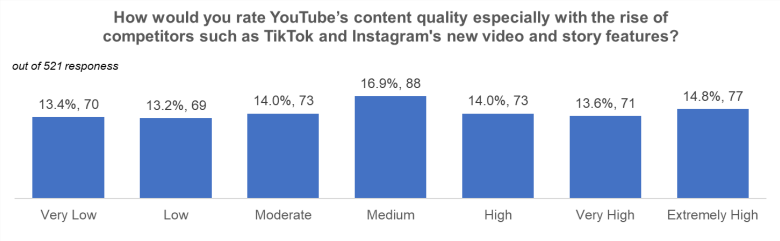

How important do you think the duration of the video is when choosing which video to watch?
out of 521 responess
11.5%, 60
12.7%, 66
10.7%, 56
9.2%, 48
19.2%, 100
20.9%, 109
15.7%, 82
Very Low
Low
Moderate
Medium
High
Very High
Extremely High


How important do you think a captivating or click-baite image is when choosing which video to watch?
out of 521 responess
12.5%, 65
12.3%, 64
12.3%, 64
8.1%, 42
19.0%, 99
19.0%, 99
16.9%, 88
Very Low
Low
Moderate
Medium
High
Very High
Extremely High


Are you bothered by the amount of YouTube ads you receive?
out of 521 responess
15.0%, 78
11.3%, 59
12.9%, 67
13.1%, 68
13.4%, 70
14.4%, 75
20.0%, 104
Very Low
Low
Moderate
Medium
High
Very High
Extremely High


Would you say the ads are of high quality, do they attract you to the products/services advertised?
out of 521 responess
21.3%, 111
14.0%, 73
14.8%, 77
16.7%, 87
10.2%, 53
11.7%, 61
11.3%, 59
Very Low
Low
Moderate
Medium
High
Very High
Extremely High


Does the presence of ads on YouTube videos make you switch to other entertainment platforms (like Instagram, TikTok or Facebook)? If so, how important is the impact of ads to switching?
out of 521 responess
18.6%, 97
15.2%, 79
13.2%, 69
11.5%, 60
15.9%, 83
12.1%, 63
13.4%, 70
Very Low
Low
Moderate
Medium
High
Very High
Extremely High


How would you rate YouTube's content quality especially with the rise of competitors such as TikTok and Instagram's new video and story features?
out of 521 responess
13.4%, 70
13.2%, 69
14.0%, 73
16.9%, 88
14.0%, 73
13.6%, 71
14.8%, 77
Very Low
Low
Moderate
Medium
High
Very High
Extremely High

Principal Component Analysis (PCA) and Size Effect

To identify patterns in our dataset and determine the most important features of our selected variables we carried out Principal Component Analysis (PCA), which in the context of survey data works by identifying patterns that represent the underlying structure of the responses.

It's a statistical technique used in data analysis and machine learning for dimensionality reduction while retaining as much of the variability in the data as possible. The core idea behind utilizing the PCA technique is to transform a dataset composed of possibly correlated variables into a set of orthogonal (uncorrelated) variables, called principal components. The principal components are ordered in such a way that the first few retain most of the variation present in all of the original variables.

The PCA was applied to eight (8) psychometric variables, which represent respondents' opinions regarding YouTube (specifically answering the question "What is YouTube for you?"), plus two (2) questions related to the importance of the duration of the video and its cover image. For this part, the "**proc princomp"** procedure was applied as follows:

```
/* Storing the principle components in a new dataset yt_coord  */
proc princomp data=sas_proj.yt_dataset out=sas_proj.yt_coord;
    var d10_1-d10_8;
run;
```

When survey respondents rate different aspects (like "a place to watch sports", "a place to listen to music" etc…), their responses are not independent in most cases, however, they often show some interrelationships. For this reason, it was checked for independence by correlating the principal components with the average ratings given by each respondent. This step indicates if there's a significant relationship between the patterns identified in the PCA and the respondents' general rating tendencies.

Pearson Correlation Coefficients, N = 521
Prob > |r| under H0: Rho=0

	avgi	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
avgi	1.00000	0.56763 <.0001	0.42273 <.0001	0.43121 <.0001	0.45379 <.0001	0.17168 <.0001	0.00158 0.9713	0.03496 0.4258	0.27662 <.0001

For instance, it was found that the first four(4) principal components (which might represent overall satisfaction) are highly correlated with the average ratings given by each respondent, indicating a "size effect".

It suggests that customers' general rating tendency (whether they are generally high raters or low raters) is influencing their ratings across all categories, not just their specific feelings about each category. This means that the respondents tend to provide their ratings by referring to a personal mean vote, suggesting the need for size effect elimination.

To solve this problem, the data was scaled using a non-linear row standardization of the input matrix. The scaling process is based on centering each user's answer based on the mean perceived by the user. This is the code applied for scaling the data:

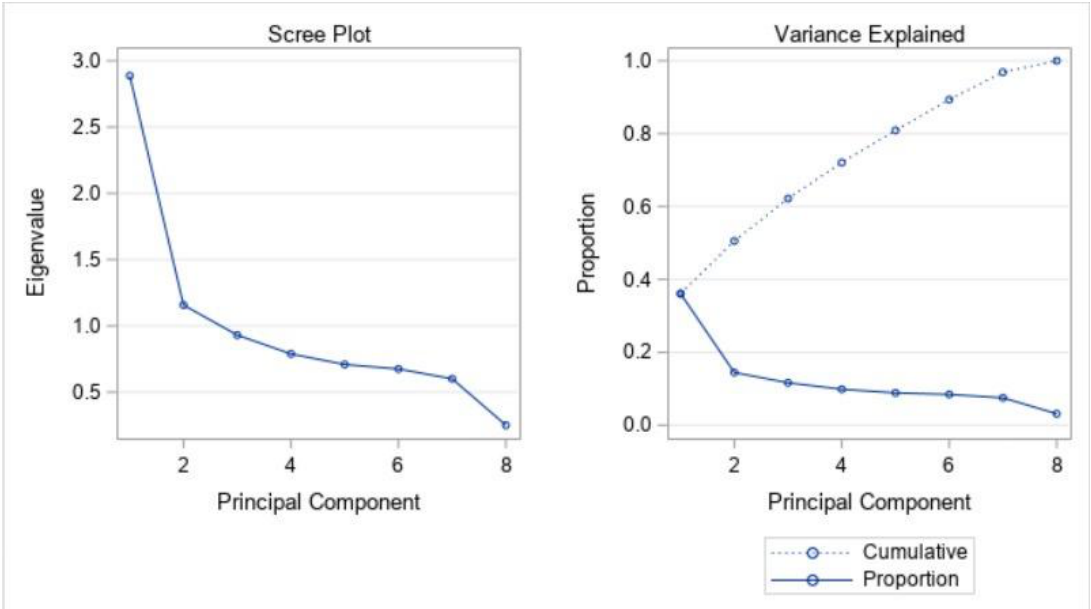
```
data sas_proj.sz_yt; set sas_proj.yt_dataset;
    avgi=mean(of d10_1-d10_8);
    mini=min(of d10_1-d10_8);
    maxi=max(of d10_1-d10_8);
    /* Active variables array */
    array a1 d10_1-d10_8; /* Input array */
    array a2 new_1-new_8; /* Output array */
    /* Conditions to be checked SAS for loop*/
    do over a2;
    if a1 > avgi then a2=(a1-avgi)/(maxi-avgi);
    if a1 < avgi then a2=(a1-avgi)/(avgi-mini);
    if a1 = avgi then a2=0;
    if a1 =. then a2=0;
    end;
    /* Adding labels */
    label new_1='w_sports';
    label new_2='w_educative';
    label new_3='l_music';
    label new_4='w_vlogs';
    label new_5='w_diy';
    label new_6='w_gaming';
    label new_7='imp_duration';
    label new_8='imp_thumbnail';
run;
```

After implementing the process for size effect removal, an additional PCA was conducted over the scaled dataset. The “proc princomp” command in SAS consists of the Eigen Decomposition of the covariance matrix of the selected variables, which results in a set of eigenvectors and corresponding eigenvalues. The eigenvectors represent the directions of maximum variance in the data, while eigenvalues indicate the magnitude of the variance in those directions. The corresponding eigenvalues rank the eigenvectors in decreasing order.

The four principal components capturing the most variance in the data (with the highest eigenvalues) were chosen, shown in the table below, explaining the large 72.05% of the variability. We decided not to strictly apply the Kaiser heuristic rule, as we believed 50.56% (the cumulative variance that only the first two principal components would be able to explain) not to be a sufficiently high variance explained by the principal components.

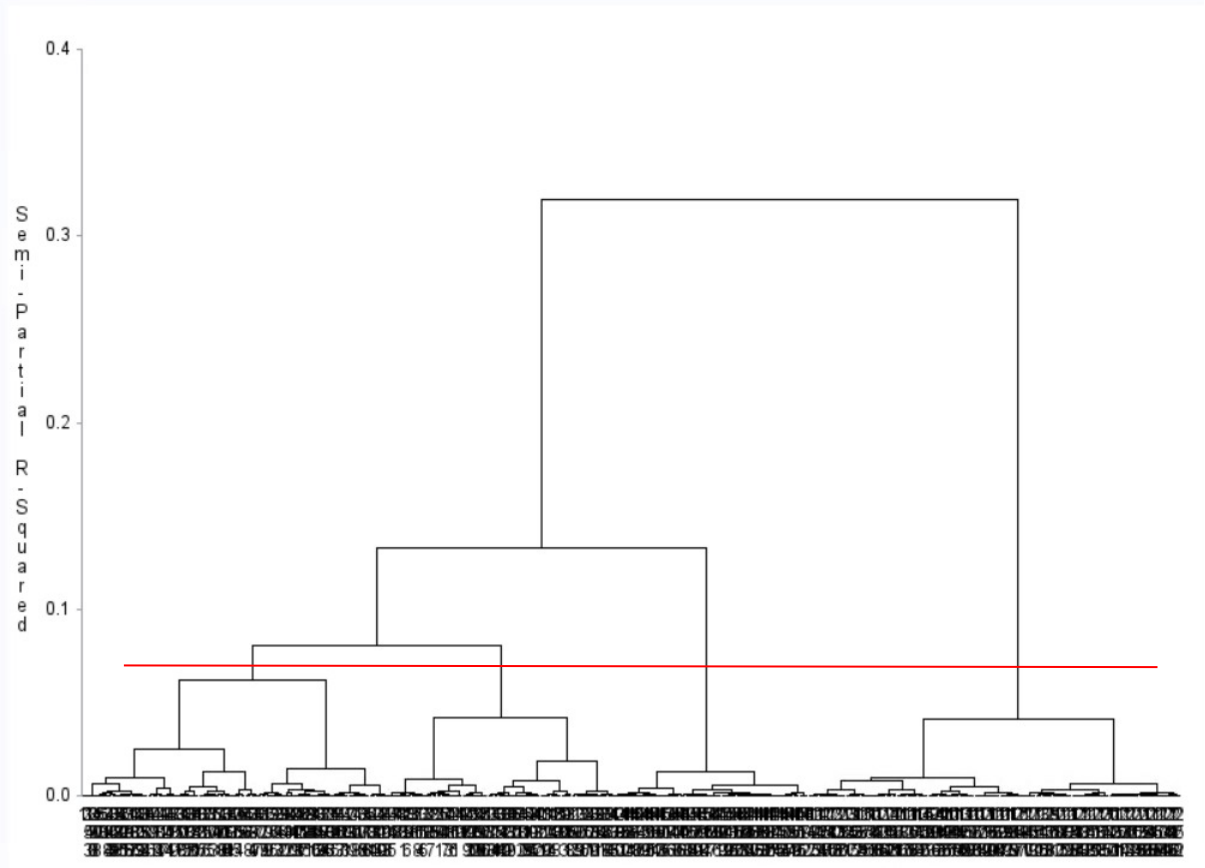
Eigenvectors									
		Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
new_1	w_sports	0.413634	0.210383	0.039485	-.433337	-.220222	-.344457	-.487866	0.436056
new_2	w_educative	-.142962	0.724540	0.461650	0.261658	0.191841	-.090023	0.223227	0.279788
new_3	l_music	-.406618	-.077234	-.533880	-.130270	0.424370	-.199219	0.125784	0.539536
new_4	w_vlogs	-.404809	0.069021	-.081381	0.071804	-.671541	0.475877	-.089252	0.366326
new_5	w_diy	-.194403	-.581507	0.687485	-.097129	0.172712	0.030745	-.085152	0.322432
new_6	w_gaming	0.378747	0.080844	0.006506	-.489181	0.192395	0.610542	0.417908	0.162019
new_7	imp_duration	0.393212	-.265754	-.042551	0.334948	-.348454	-.313423	0.582152	0.319694
new_8	imp_thumbnail	0.378687	-.068851	-.138497	0.600588	0.308831	0.367638	-.409354	0.271777

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.88725251	1.73006315	0.3609	0.3609
2	1.15718936	0.22644755	0.1446	0.5056
3	0.93074181	0.14196736	0.1163	0.6219
4	0.78877444	0.07977447	0.0986	0.7205
5	0.70899998	0.03373961	0.0886	0.8091
6	0.67526037	0.07446700	0.0844	0.8935
7	0.60079337	0.34980520	0.0751	0.9686
8	0.25098817		0.0314	1.0000



Clustering

Using the variables treated for size effect and the first 4 principal components, the Ward method was used to construct 4 clusters that minimize the inter-cluster variance. In the figure below, the dendrogram and the cut-off point identifying the 4 clusters are visible. The accompanying figures show the dendrogram we obtained and the frequency of observations in each cluster respectively.



CLUSTER	Frequency	Percent	Cumulative frequency	Cumulative percentage
1	145	27.83%	145	27.83%
2	177	33.97%	322	61.8%
3	91	17.47%	413	79.27%
4	108	20.73%	521	100%

T-test

After segmenting the sample into four individual clusters, the mean of each attribute within each cluster was compared against a synthetic cluster made up of all the observations. This step was performed to identify the fundamental characteristics of each cluster and is based on the assumption that those characteristics, the "partial" averages, should be distant from the "general" average of the fake cluster composed of all the data.

description	tvalue_ clus1	tvalue_ clus2	tvalue_ clus3	tvalue_ clus4	pvalue_ clus1	pvalue_ clus2	pvalue_ clus3	pvalue_ clus4
W_sports	1.3	14.772	-20.2	-5.57	0.19	<.0001	<.0001	<.0001
W_education	11.16	-9.55	-2.16	-1.97	<.0001	<.0001	0.032	0.05
L_music	0.22	-11.79	28.39	-0.11	0.82	<.0001	<.0001	0.92
W_vlogs	3.05	-14.04	16.24	1.92	0.002	<.0001	<.0001	0.26
W_diy	-4.55	-2.46	3.77	8.34	<.0001	0.014	0.0002	<.0001
W_gaming	-0.06	12.88	-21.96	-3.3	0.95	<.0001	<.0001	0.001
Importance_duration	-0.14	11.26	-12.78	3.38	<.0001	<.0001	<.0001	0.001
Importance_thumbnail	-3.77	10.03	-11.04	1.93	0.0002	<.0001	<.0001	0.06
Device_usage	1.66	7.82	-11.52	2.11	0.11	<.0001	<.0001	0.04
Youtube_usage	1.57	-4.02	5.25	-0.55	0.12	<.0001	<.0001	0.58

Legend:

- the green rows relate to behavioral attributes
- the red rows relate to opinion attributes
- the values highlighted in yellow refer to statistically significant values
- the values highlighted in light blue refer to borderline values

The next step was to perform a t-test over all the scalable values within each of the clusters to determine if the individual active attributes were significant to each of the clusters represented by the data. The t-value was used to determine whether a statistically significant difference exists between the mean of an individual cluster related to the entire dataset. Here the null hypothesis is given as no significant difference between the sample means. We decided to create a table summing up the results of the "**proc ttest"** procedures that were automatically produced using a SAS macro. For each variable under analysis, you can find its corresponding t-values and p-values for each of the four clusters.

Chi-Squared Test

After carrying out the t-tests, we also wished to check for the relevance of some categorical variables (such as gender, age, and employment status) in determining the clusters. For this purpose, a Chi-squared test was implemented. The objective of this test is to check whether or not there is a significant deviation between the expected frequencies of the categorical variables concerning the observed frequency.

FREQ procedure for **gender**

The FREQ Procedure

Table of gender by CLUSTER

Frequency / Expected / Percent / Row Pct / Col Pct	CLUSTER 1	2	3	4	Total
Female	68 63.455 13.05 29.82 46.90	62 77.459 11.90 27.19 35.03	45 39.823 8.64 19.74 49.45	53 47.263 10.17 23.25 49.07	228 43.76
Male	77 66.516 14.78 32.22 53.10	65 81.196 12.48 27.20 36.72	46 41.745 8.83 19.25 50.55	51 49.543 9.79 21.34 47.22	239 45.87
Other	0 15.029 0.00 0.00 0.00	50 18.345 9.60 92.59 28.25	0 9.4319 0.00 0.00 0.00	4 11.194 0.77 7.41 3.70	54 10.36
Total	145 27.83	177 33.97	91 17.47	108 20.73	521 100.00

Statistics for Table of gender by CLUSTER

Statistic	DF	Value	Prob
Chi-Square	6	93.8421	<.0001
Likelihood Ratio Chi-Square	6	102.4821	<.0001
Mantel-Haenszel Chi-Square	1	1.6420	0.2000
Phi Coefficient		0.4244	
Contingency Coefficient		0.3907	
Cramer's V		0.3001	

FREQ procedure for **age**

Table of age by CLUSTER

Frequency / Expected / Percent / Row Pct / Col Pct	CLUSTER 1	2	3	4	Total
18 - 25	19 32.006 3.65 16.52 13.10	74 39.069 14.20 64.35 41.81	3 20.086 0.58 2.61 3.30	19 23.839 3.65 16.52 17.59	115 22.07
26 - 35	60 33.676 11.52 49.59 41.38	11 41.107 2.11 9.09 6.21	3 21.134 0.58 2.48 3.30	47 25.083 9.02 38.84 43.52	121 23.22
36 - 45	55 28.388 10.56 53.92 37.93	8 34.653 1.54 7.84 4.52	7 17.816 1.34 6.86 7.69	32 21.144 6.14 31.37 29.63	102 19.58
46 - 60	6 12.524 1.15 13.33 4.14	1 15.288 0.19 2.22 0.56	35 7.8599 6.72 77.78 38.46	3 9.3282 0.58 6.67 2.78	45 8.64
< 18	1 23.935 0.19 1.16 0.69	82 29.217 15.74 95.35 46.33	0 15.021 0.00 0.00 0.00	3 17.827 0.58 3.49 2.78	86 16.51
> 60	4 14.472 0.77 7.69 2.76	1 17.666 0.19 1.92 0.56	43 9.0825 8.25 82.69 47.25	4 10.779 0.77 7.69 3.70	52 9.98
Total	145 27.83	177 33.97	91 17.47	108 20.73	521 100.00

Statistics for Table of age by CLUSTER

Statistic	DF	Value	Prob
Chi-Square	15	600.6318	<.0001
Likelihood Ratio Chi-Square	15	550.4616	<.0001
Mantel-Haenszel Chi-Square	1	5.6450	0.0175
Phi Coefficient		1.0737	
Contingency Coefficient		0.7318	
Cramer's V		0.6199	

FREQ procedure for **occupation**

The FREQ Procedure

Frequency
Expected
Percent
Row Pct
Col Pct

Table of occupation by CLUSTER

occupation	CLUSTER 1	2	3	4	Total
Employed	48 48.983 9.21 27.27 33.10	10 59.793 1.92 5.68 5.65	80 30.741 15.36 45.45 87.91	38 36.484 7.29 21.59 35.19	176 33.78
Student	20 37.015 3.84 15.04 13.79	84 45.184 16.12 63.16 47.46	3 23.23 0.58 2.26 3.30	26 27.57 4.99 19.55 24.07	133 25.53
Unemployed	30 15.585 5.76 53.57 20.69	4 19.025 0.77 7.14 2.26	2 9.7812 0.38 3.57 2.20	20 11.608 3.84 35.71 18.52	56 10.75
Working student	47 43.417 9.02 30.13 32.41	79 52.998 15.16 50.64 44.63	6 27.248 1.15 3.85 6.59	24 32.338 4.61 15.38 22.22	156 29.94
Total	145 27.83	177 33.97	91 17.47	108 20.73	521 100.00

Statistics for Table of occupation by CLUSTER

Statistic	DF	Value	Prob
Chi-Square	9	248.5796	<.0001
Likelihood Ratio Chi-Square	9	262.9127	<.0001
Mantel-Haenszel Chi-Square	1	21.5286	<.0001
Phi Coefficient		0.6907	
Contingency Coefficient		0.5683	
Cramer's V		0.3988	

From these tests we can determine that all three categorical variables (gender, age, and occupation) are statistically significant for cluster determination, meaning that they are all relevant for determining the clusters. As you can see, the p-values of the three Chi-squared tests that we ran are very small (close to zero), suggesting the significance of those variables. This will be the basis for describing the clusters in the next section.

Cluster Description

We can now proceed with the description of the four clusters that have emerged in our cluster analysis.

Cluster 1 – "The Busy Workers"

The first cluster is characterized by a population aged between 28 and 45, with 41% falling within the 28-35 age range, and 36.45% representing a population between 34 and 45 years old. The population appears to have an even distribution between male and female genders and is predominantly composed of workers (33%) and student-workers (32.4%) who favor YouTube as a platform for viewing educational videos and tutorials of limited duration. The most influential variables explaining their behaviors are related to the search for educational videos, which are particularly appreciated by student-workers (t-test = 11.16), while the younger individuals within this category still show interest in vlogs (t-value = 3.05). However, there is a notable lack of enthusiasm for "Do-It-Yourself" videos (t-value = -4.55). Additionally, respondents seem to attach less importance to video thumbnails (or click-bait videos), especially among those who are still studying, as they prioritize the content of the chosen video.

These individuals are characterized by a rather busy lifestyle. They use YouTube primarily as a platform for learning rather than entertainment, driven by their limited available time since the majority of users are either employed or simultaneously working and pursuing their studies; for that reason, YouTube is still used, but not in a statistically significant way (t-value = 1.57), as the majority of users are in the early part of their busy working life and might be more prone to prefer other entertainment platforms.

Cluster 2 – "The Young and Lazy Kids"

This cluster is characterized by a very young generation. The majority of the population consists of individuals under 18 years old (46.33%), while the remaining portion (41.81%) falls within the 18-25 age group. The gender distribution is very similar, with also other gender identities. The occupation is primarily divided into two major categories, namely students (47.48%) and working students (44.63%). Users appear to have diverse content preferences, characterized by well-defined values, especially among young individuals where there is a significant interest in sports-related videos (t-value = 14.772) and gaming videos (t-value = 12.88). However, the high number of responses by young participants results in limited interest in educational videos. Vlogs are steadily losing their appeal (t-test = -14.04), partly because Instagram is preferred over them, and do-it-yourself tutorials do not seem to be of interest yet. Music (t-value = -11.79) is predominantly listened to on other platforms, such as Spotify, which is very popular among the younger generation. Once again, the user's attention remains focused on videos with visually appealing thumbnails (t-value = 10.03) and limited duration (t-value = 11.26).

From this cluster, we can confidently define the interests of the youth, who spend a significant amount of time on their phones (t-value for device usage = 7.82) and understand how to make the best use of various platforms. For that reason, we can conclude that this generation has been replacing YouTube with other social media platforms and its usage is decreasing considerably.

Cluster 3 – "The YouTube Gramps"

The third cluster is predominantly (47.25%) characterized by more mature individuals (older than 60), including those aged between 46 and 60 (38.46%). The large majority of responses come from employed individuals (87.91%) and the population is evenly divided between men (50.55%) and women (49.45%). These users do not pay particular attention to the length of videos (t-value = -12.78), as it can be one of the few moments of leisure. The population, significantly more mature, shows no interest in gaming videos (t-value = -21.96) and sports videos (t-value = -20.2) because they are likely to prefer watching them on television platforms such as Sky, Dazn, etc. On the other hand, music is widely appreciated on YouTube (t-value = 28.93), as users are unlikely to subscribe to Spotify, which is used mainly by younger people. The components of this cluster show interest in vlogs (t-value= 16.24) and DIY videos are also quite popular (t-value = 3.77). However, despite their lack of interest in video length, the use of technological devices is low (t-value = -11.54), and video selection is not particularly influenced by the cover (t-value = -11.04).

From the analysis of this cluster, we can infer that slightly more mature users, especially men, prefer not to use YouTube for watching sports. As a result, its usage has been replaced by other television platforms capable of catering to users with broader and more detailed content. However, YouTube continues to be widely used for music listening, especially among women. These female users, on the other hand, do not deem it essential to subscribe to other platforms like Spotify. Moreover, we can deduce that more mature individuals still appreciate watching videos as a moment of relaxation from family and work duties. Video length is not considered relevant, mainly because it primarily involves the playback of music that does not hinder concurrent activities.

Cluster 4 – "The Independent Folks"

The last cluster is again predominantly characterized by young individuals (43.52%) aged between 26 and 35 years. This group is evenly divided among workers (35.19%), students (24.07%), and working students (22.22%). The young population is further divided into slightly more females (49.07%) than males (47.22%) within the previously mentioned age range. These individuals are gaining increased independence by living on their own or cohabiting with others, thus distancing themselves from their families. Consequently, we observe a strong interest in 'DIY' videos (t-value = 8.34), while less practical content, such as gaming videos (t-value = -3.3), and academic content such as educational videos (t-value = -1.97), garner less attention. The fourth cluster demonstrates that a segment of young adults prefers using YouTube as a tool to enhance their practical skills. Regarding other platform content, we lack significant variables, especially concerning listening to music, likely conducted on alternative platforms, as well as the consumption of vlogs.

Conclusion

In our analysis, we observed that young individuals' primary interest revolves around sports and gaming videos. On the other hand, girls may prefer to replace YouTube with other platforms like Instagram, where they can find content more aligned with their tastes. Regarding music, young people generally prefer to use Spotify and engage in subscriptions, completely replacing YouTube.

We determined that the usage of YouTube in the population between the ages of 28 and 45 changes compared to the younger demographic due to shifting needs. In this cluster, we find working students who use the platform as a space for gathering information, while users balancing family and work responsibilities may not be interested in DIY videos, recipes, or tutorials, preferring potentially quicker alternatives to meet their needs. Generally, this segment of the population stands out for its lack of free time, and consequently, their usage of the platform is limited.
Another style of YouTube usage is evident among young individuals who are preparing to live independently or away from their families. The independence they gain leads to increased usage of YouTube, particularly for DIY videos, replacing purely recreational content.
The older population, especially females, seems to use YouTube more as a platform for listening to music. On the other hand, males prefer to watch sports on television or more specialized platforms, which results in reduced usage of YouTube. However, both categories of users appreciate DIY videos. This population does not particularly pay attention to video length, a factor that can be explained by the presence of homemakers or retirees in the population.

The analysis employed statistical techniques such as Principal Component Analysis (PCA), T-tests, and Chi-Squared tests to extract meaningful patterns from the survey data, providing a robust basis for understanding user behaviors. The insights gained underscore the necessity for content creators and platform administrators to acknowledge and adapt to the diverse needs and preferences of YouTube users across various life stages. As YouTube continues to evolve, these findings can inform tailored content strategies, ensuring relevance and engagement across its expansive and varied user base.