Introduction to Digital Humanities and Social Analytics

Is there a correlation between higher education levels and negative attitudes towards mainstream ideology as presented in short video content on Douyin?

Team 7

2779252 | Oussama El Mojahid

2775341 | Sihao Fu

2768409 | Yuli Wang

2780957 | Tomash Oosterbaan

2859503 | Stefana Catinca Pop

Coordinators: Lorella Viola & Erika Kuijpers

18/10/2024

Table of content

Project Overview

This article was edited by an interdisciplinary research team of five students from Media Studies, Linguistics, Computer Science, Artificial Intelligence, and Business Management. We were assigned the task of writing a research article on the topic of mainland Chinese users' perceptions of the dominant ideology on the Douyin platform. After reading the article and reviewing all the data, the research question that emerged from our group was:

Is there a correlation between higher education and negative attitudes toward mainstream ideologies presented in short video content?

Our dataset analysis showed that some respondents indicated that traditional factors/mainstream ideologies were offensive to them. In the dataset, such groups appeared to be positively correlated with those who were highly educated and from more economically developed regions. Therefore, we identified the direction of our research, which was to investigate correlations based on demographic information.

Data Acquisition

In the process of data collection, the researchers (*Xinyu Li & Sabariah Mohammed Salleh*) mainly used questionnaires. The main target of the survey was the adolescent group (between 18 and 30 years old) in mainland China. The questionnaire included 14 questions; the types of questions being single-choice, multiple-choice, question and answer, seven-dimensional scale questions, screening questions, etc., which covered demographic information, media consumption behavior, content preference, attitudes towards mainstream culture, and so on.

Before the official launch, the researcher conducted a small-scale test for the respondents in advance, and a total of 80 people participated in this test. Based on the feedback from the pre-test participants, the researcher modified some questions and re-edited the questionnaire using more precise and easier-to-understand terms and distributed it on the intelligent research platform Credamo China.

In order to ensure the comprehensiveness and diversity of the data, the researcher used a simple random sampling strategy, i.e., n units were arbitrarily selected as samples from a total of N units, so that each possible sample had the same probability of being selected.

The sampled respondents mainly had the following character portraits: they were teenagers who grew up in mainland China, 18-30 years old, and often used *Douyin*[1].

The investigator rewarded eligible participants with 2 RMB (equal to 0.25 EUR) each in return for completing the survey.

A total of 500 questionnaires were collected after the release of the questionnaire. Of all the results, 5 were rejected by manual screening and 164 by automated screening, mainly to eliminate incomplete, irregular, or false questionnaires. In the end, out of the 500 questionnaires from 135 cities, the researcher retained 331 valid questionnaires to be analyzed.

---

[1] *Douyin: Douyin is a Chinese short-form video platform launched in 2016 by ByteDance, operating separately from its global counterpart, TikTok. It thrives in China due to the "Great Firewall," which restricts international platforms and allows it to comply with local regulations. Douyin features live commerce, educational content, and integrated mini programs, making it a versatile app. Unlike TikTok, its algorithm favours high-quality content from major brands and influencers. It is tailored to China's cultural and regulatory environment, offering a unique digital experience for Chinese users.*

Methodology

The responses were analyzed using a combination of statistical tests and regression models to investigate the relationship between education, gender, and aversion toward mainstream content. To make trends and relationships clearer, we used visualizations alongside statistical analysis.

We processed the data using the Python library pandas. The dataset was processed by removing any missing or incomplete responses, focusing on three variables: education level, gender, and the respondents' aversion to mainstream content.

Education levels were grouped into six categories: Primary school, Junior High School, High School, Undergraduate, Post Graduate, and PhD student, while gender was divided into three categories: Man, Woman, and Third Gender. This step was essential for both the statistical analysis and the visualization, ensuring that the data was easy to interpret and analyze.

We used two types of regression models to better understand the relationship between education and aversion to mainstream content. We used a Logistic Regression model to look at whether people with higher education levels are more likely to have a strong negative reaction to mainstream content. This model allowed us to better understand the chances of someone having high aversion based on their education and gender.

To explore the relationship between education level and aversion to mainstream content, we used several correlation tests. First, we ran the Pearson Correlation test to see how strongly education level and aversion are related in a linear way. This gave us an idea of whether there was a significant link between these two factors.

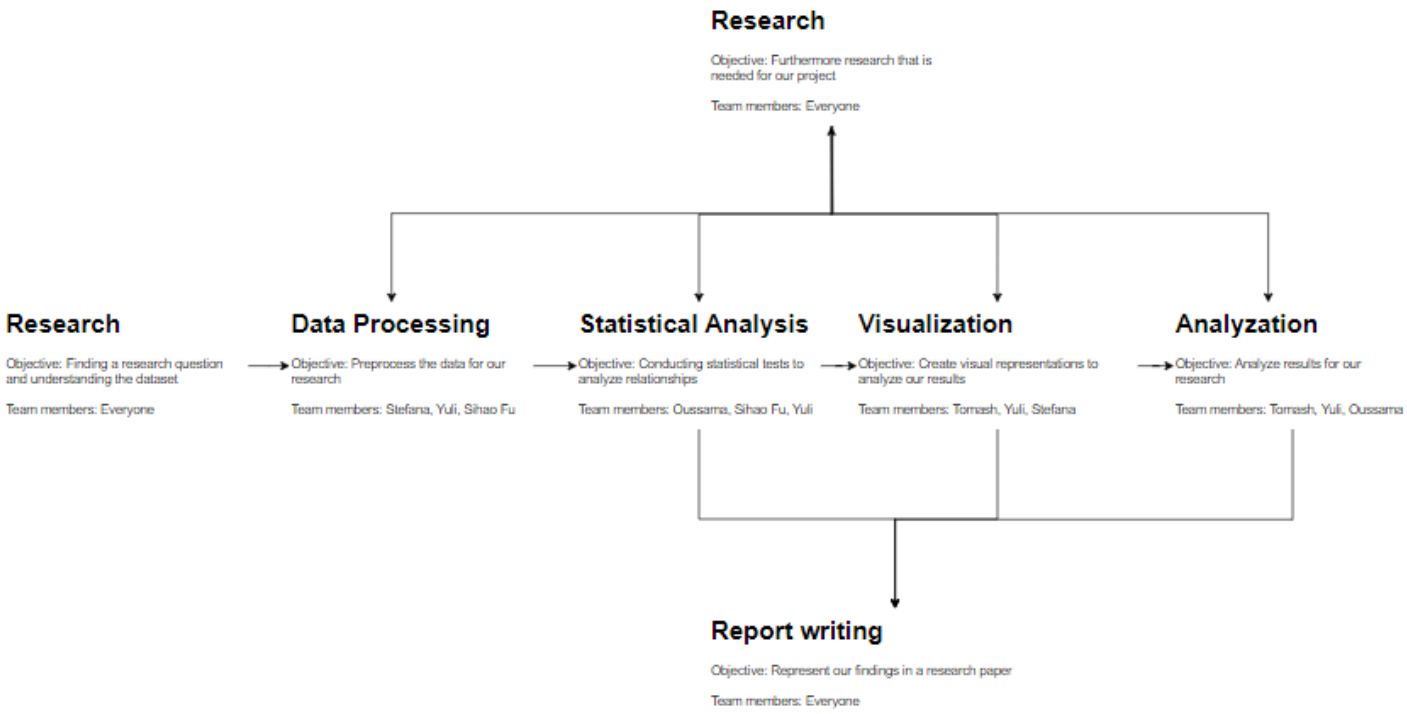
Then, because our dataset appeared to be better suited for non-parametric tests due to the ordinal nature of education levels and aversion scores, we applied multiple non-parametric tests. We used the Spearman Rank Correlation and Kendall's Tau tests, which are better for situations where the relationship isn't strictly linear. These tests helped us better understand the strength and direction of the connection between education and aversion.

We also used the Point-biserial Correlation test to look at the link between education level and whether someone had a high aversion to mainstream content (where we defined high aversion as a rating of 1-3).

Visualisations were an important part of our methodology, making it easier to identify trends and relationships within the data. Using the Seaborn and Matplotlib libraries, we created several graphs to present our findings. First, we plotted the distribution of education levels across all respondents, as well as within the group that expressed the strongest negative feelings (aversion ratings of 1-3). This allowed us to see how education levels varied among those who felt most offended by the mainstream content.

We also created boxplots to visualize how aversion levels differed by gender and for all respondents. These plots illustrated whether there were significant differences in the way contestants reacted to mainstream content. Additionally, we generated a pie chart showing the overall distribution of aversion scores, which provided a clear overview of how offended respondents were by the content on Douyin.

Workflow Steps



An Overview: Initial Assumptions

A workflow is a structured series of steps or tasks designed to accomplish a specific goal. It serves as a blueprint that organizes tasks, assigns responsibilities, and facilitates the flow of information between individuals or teams. A well-designed workflow ensures that tasks are performed in the correct sequence, helping to standardize processes, reduce errors, and enhance collaboration.

Workflows can be automated using software tools that apply business rules to determine when one task is complete and the next should begin (a method we tried to implement in our own data analysis process). Some of the software also manages dependencies between tasks through workflow orchestration, ensuring a smooth transition between interconnected activities. Besides, workflow management software often includes templates for documenting processes and modeling them.

Our workflow could be considered effective if it provided benefits such as improving project success by providing clear steps and reducing ambiguity, enhancing productivity by clarifying task responsibilities, optimizing resource use, and preventing bottlenecks.

The first step to ensure the success of our project realization began with a clearly defined objective and scope. We tried implementing the SMART goals (Specific, Measurable, Achievable, Relevant, and Time-bound), and a project charter that would help in aligning our team efforts and set a clear direction, milestones, and deadlines. A well-defined scope would prevent scope creep, ensuring the project stayed on track without unnecessary delays or costs.

We wanted to use automation in order to streamline the scope definition process, with useful project management tools like PM3 (aiding in creating detailed plans and setting task dependencies, offering real-time updates, and ensuring timely delivery). Workflow diagrams or flowcharts would also be useful by helping in the visualization of the task sequence and the overall process.

We were aware that breaking the project into smaller, manageable tasks would be crucial for understanding the full scope of our project. Tools that could aid this undertaking included Gantt charts (that provide a visual timeline, mapping start and end dates of each task), and workflow templates (that supply a standardized format for outlining the project plan, making it easier for replication and helping in maintaining consistency).

We needed to assign our available assets to the specific tasks/activities within the project (resource allocation). To do this, we identified all the necessary resources for the project, matched them to the defined tasks based on skills, availability, and workload, and managed resource usage to avoid overloading or under-utilizing team members. During the whole process, we decided to plan for contingencies in case of unexpected changes or shortages (as seen in the team chart).

This leads us to risk management: spotting, assessing, and mitigating potential risks that could negatively impact our project (data deficiency, lack of skill, lack of cooperation & communication…)

We tried to identify these risks through techniques like brainstorming and SWOT analysis, creating a management plan that would be continuously monitored and updated to ensure the risks were managed throughout the project.

For the duration of our collaboration, effective communication was decided to be our most important intent. With a communication plan we would be able to outline who needed to be informed, what information had to be shared, and how/when updates occurred. We tried to have regular check-ins and status meetings (Zoom, GoogleMeets, Whatsapp) to facilitate team alignment. Open dialogue fostered teamwork, while integration with task management software (PM3) improved productivity.

We attempted to handle project alterations (adjustments in scope, timelines, or requirements) smoothly without disrupting progress (change management), and to ensure that project deliverables met the required standards and there was continuous improvement through testing (quality management: PDCA-> Plan-Do-Check-Act).

Finally, we tried to include a post-evaluation of our work using mainly self-reports that summarised/criticized outcomes and insights.

A step-by-step review

This step-by-step workflow describes how our team approached the project, from forming the research question to analyzing the dataset and finally drawing conclusions. Throughout the process, we used various software tools, decided on possible approaches, and iteratively reviewed and adjusted our outlooks to address challenges. Our workflow was not linear, as we encountered difficulties and revised our steps multiple times before reaching a final result.

1. **Team formation and task allocation:** the first step involved establishing a team structure and assigning roles based on each member's strengths, availability, and expertise. We created a team chart to outline responsibilities and track progress (Google Sheets/Google Drive).

By dividing tasks based on strengths we were able to maximize efficiency and ensure that each member contributed effectively. Having a clear schedule for regular meetings helped us stay on track and resolve any issues early.

A more flexible role-sharing system was considered, but we opted for task-specific roles to avoid confusion and allow each member to specialize in their area.

2. **Developing the research question and thesis:** the next step was to develop a clear research question. After reviewing the dataset - Young People's Perception on Dissemination of Government Propaganda & Socialist Culture in China - we formulated the question: "Do Chinese individuals with higher education exhibit a stronger aversion to mainstream political ideology on the *Douyin* platform?"

The RQ was chosen because it was directly linked to several key variables in the dataset (such as education level, gender, and aversion to political content). It addresses a timely social issue while allowing us to leverage the dataset's existing structure, which was relatively well-suited to this type of sociopolitical analysis.

We also considered investigating different social behaviors on *Douyin* (such as age-based reactions to content) but eventually found that focusing on education levels would provide a more meaningful insight given the available data.

3. **Data review, shortcomings, and preprocessing:** after our RQ was defined, we appraised the dataset to identify its structure, variables, and blemishes (missing, incomplete, contradictory data). We also assessed the potential biases in the dataset (e.g. underrepresentation of certain gender identities or education levels).

The focus was placed on three key variables: education, gender, and aversion to mainstream political content. For preprocessing we used tools such as Python Pandas.

We considered supplementing the dataset with additional demographic data (other literature and datasets), but ultimately chose to concentrate on the existing dataset to avoid introducing inconsistencies or new biases.

4. **Analyzing the dataset and statistical methods:** after preprocessing, we proceeded to the analysis stage (applying statistical tests and correlation tests).

Using a combination of parametric and non-parametric tests allowed us to explore different aspects of the data, ensuring a more comprehensive understanding of the relationships between variables.

5. **Visualization and interpretation:** to make trends and relationships clearer, we created visualizations to support our analysis. They helped us better understand the distribution of education levels and aversion scores across different genders and demographics (using Seaborn/Matplotlib).

We opted for simple, clear visuals to avoid overwhelming viewers with unnecessary details. They made the data easier to interpret, allowing us to communicate complex relationships in a more digestible format.

6. **Ethical considerations and bias review:** throughout the project, we took care to address ethical concerns, particularly around data bias and potential misinterpretations of politically sensitive data (both in the dataset and our own process). We sought to mitigate these issues by maintaining transparency in how we handled the data.

Team discussions and reviews helped us label these ethical concerns collaboratively.

7. **Iterative problem-solving and workflow adjustments:** our research was not linear, encountered several challenges and difficulties that we had to work around (by holding additional meetings, consulting literature, and going back in the workflow for re-assessment).
8. **Final review, documentation, and drawing conclusions:** we compared the outcomes to our original hypothesis and finalized our documentation of every step (for transparency and replicability).

Furthermore, we discussed how this research could inform future investigations into political ideology and social media behavior, particularly in the Chinese context.

Challenges and Solutions

As the project progressed, we encountered one challenge after another, but fortunately, we gradually mastered ways to address them. The main challenges we faced were in two areas: first, the differences in communication and thought processes caused by our diverse academic backgrounds, and second, the questions and conflicts we had regarding the dataset.

For the first aspect, our team consisted of students from linguistics, communication, business, AI, and CS backgrounds. We had vastly different perspectives on how to approach problems and work together. Students from CS and AI backgrounds tended to focus on clear requirements and issues, relying on pure statistical/data-driven analysis for solutions. In contrast, students from other fields were better at literature studies, often discovering interesting and meaningful insights & ideas.

Initially, these differences led to conflicts among members when presenting ideas, and a lack of mutual trust on course-related issues further caused communication breakdowns. Before the first presentation, even for the proposal, our working state was almost like working independently and then piecing the results together, which naturally prevented us from achieving an ideal outcome. Fortunately, as the course progressed and our understanding of digital humanities deepened, we gained a better understanding of our research question.

We began exploring ways to work more harmoniously and efficiently, to find common ground, while respecting differences: Yuli first analyzed and visualized the dataset, sharing the algorithmic process with Tomash, who had a similar background. They collaborated to construct the framework for the methodology and jointly created preliminary visualizations. These visualizations made it easier for other team members without a technical background to understand what was being done, how it was done, and what conclusions were drawn. On this basis, Stefi, Sihao, and Oussama were able to understand the unfamiliar parts more easily and then incorporate and expand these findings within their literature research. They critically analyzed different literature and our research question, providing Yuli and Tomash with guiding directions. This approach allowed us to avoid unfamiliar areas and fully leverage our strengths, showcasing each member's best abilities.

As for the dataset itself, we encountered our first mistake during preliminary research: placing too much trust in the dataset. We took for granted that the dataset was reliable and accurate, leading us to propose a research question that we later found to be irresponsible due to the unreliability of the dataset. As the project progressed, we gradually discovered issues such as the lack of diverse samples and insufficient data volume. With two of our members being from the dataset's source country, we knew that, in such a large-population country, even a small survey conducted at an average university could receive over 2,000 to 3,000 responses. A dataset with only 300 valid responses seemed unreasonable.

However, through continued learning in the course, we concluded that simply dwelling on the dataset itself or forcefully using different testing methods to prove or disprove our research question would be irresponsible. After discussion, we decided to respect the results we obtained and conduct more detailed testing and analysis of the dataset to obtain more convincing answers—why we got such answers and how the dataset impacted our results across different dimensions. After extensive discussion and revision, we finally reached a consensus and successfully completed the project.

Lastly, in terms of communication and time management, I think it’s a common question that not only our group but other groups as well may have encountered similar challenges. Team members come from different faculties, and differences in daily routines made synchronization difficult. In our team, some members had numerous lectures to attend almost every day, while someone went to sleep at 6 p.m. and woke up at 2 a.m., causing him to miss all messages, and some often shared their new findings at 3 or 4 in the morning. This naturally created communication difficulties.

Despite having a well-established team charter, practical implementation always comes with certain challenges. Fortunately, all team members were friendly and responsible, and we gradually found a communication style that worked best for us. Although it was difficult to ensure everyone was available at the same time, whenever any member had a new idea or result, they would update it in our WhatsApp group immediately, and others would respond as soon as possible. We also used Google Drive for collaborative editing, making each member's

contributions visible to the others and allowing us to monitor each other's progress. This resulted in an impressively high level of efficiency during the final stages of the project.

Ethical Considerations

There are some very important ethical issues that we need to be aware of in the process of data use. The first is the legality and reasonableness of data use. All data must be collected openly and by lawful, non-coercive, non-misleading means, in a way that makes it clear to respondents that they are being surveyed and that they know exactly what points they are being surveyed on. Data is handled with particular care when it relates to sensitive content, such as personal information, home addresses, and other data that could be directly linked to an individual. In such cases, to ensure that the data is not used for purposes other than those for which it was originally intended, the scope of data use is strictly limited and the use of each dataset is documented in detail to ensure that it is consistent with its original purpose.

From a specific point of view, when it comes to information about sensitive content, we should ensure that everything possible is done to protect the privacy of the respondents, such as anonymization, obfuscation of key information, and disclosure of as few datasets as possible that are potentially reverse-identifiable (e.g., the respondents' mobile phone model, ip address, and location) in the final report. In the context of academic instruments, the Regulations to be complied with include, but are not limited to, relevant privacy regulations including the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA).

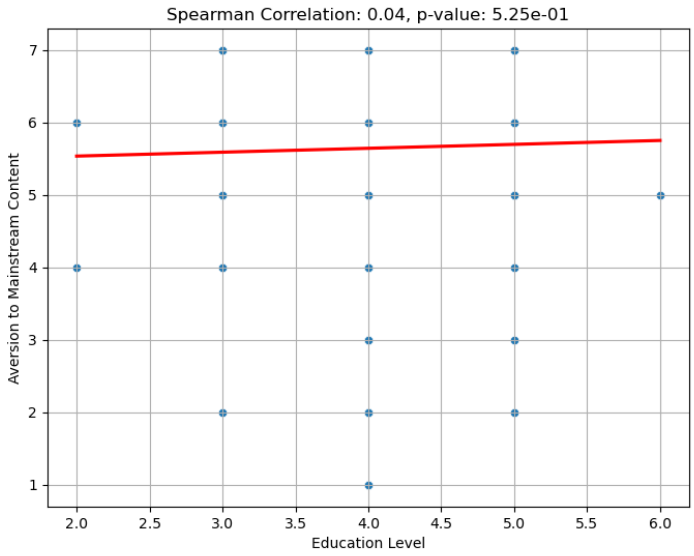
Another point of concern is data bias, as the person collecting the data cannot be completely accurate and objective, and the source and method of data during the collection process may also cause bias to occur. In our case, two-thirds of the respondents were female, which is a good example of data bias, as a reasonable gender distribution is 50 percent to 50 percent. Another even bigger bias is that all the interviewers were undergraduates, and the respondents were not grouped into surveys with different academic qualifications, which led to a lot of trouble interpreting the data afterward. The key to avoiding this is to try to diversify the sources of the data and rely as much as possible on real-world ethnicity, gender, geographic location, and so on.

Finally, it is important to emphasize that social research in countries that are not free needs to pay more attention to the protection of respondents' privacy. This is because, unlike in Europe and the United States, some of the politically sensitive content involved in these countries can expose respondents not only to data leakage but also to the possibility of retaliation from the government. Therefore, humanistic social surveys conducted in authoritative institutions need to pay more attention to the protection of respondents' data privacy.

Result

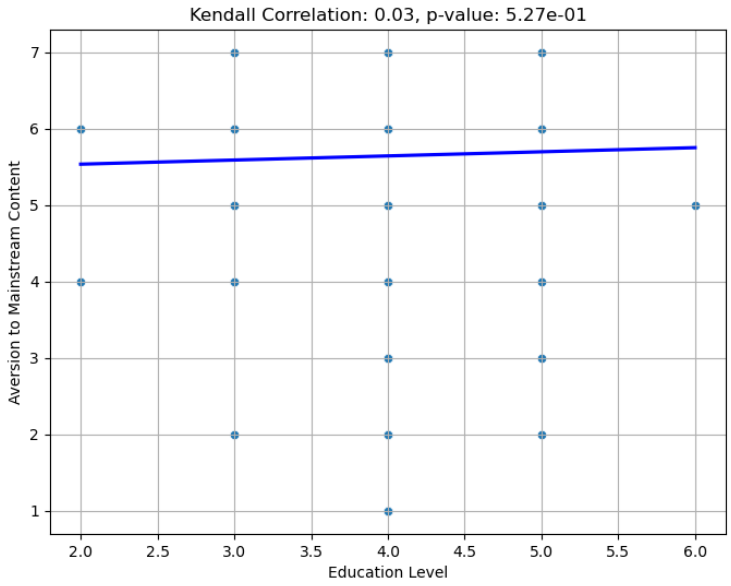
We began by analyzing the results of the correlation analysis, and to explore the potential linear and non-linear relationships between educational level and mainstream content aversion, we conducted four different correlation tests:

Spearman Correlation:



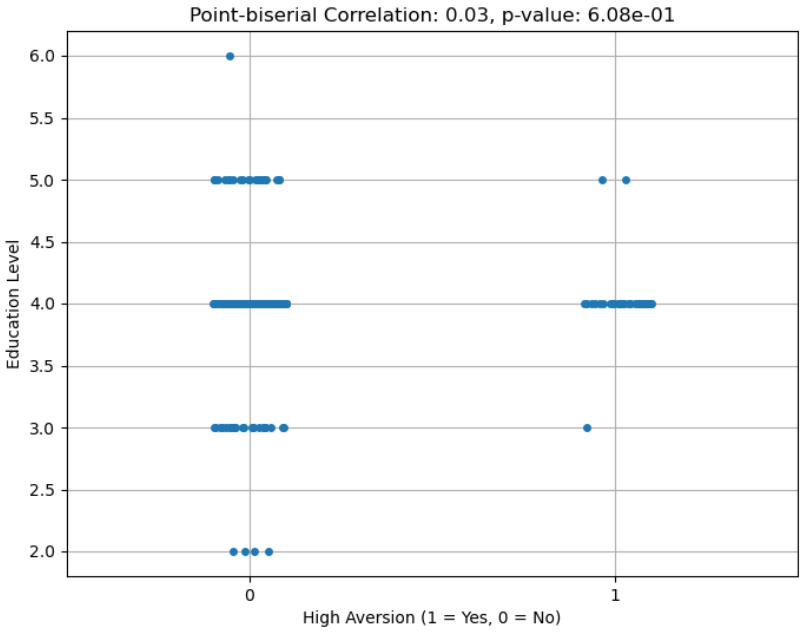
The Spearman correlation, which checks for a monotonic relationship, was 0.0351, with a p-value of 0.525. Like the Pearson correlation, this test shows no meaningful relationship between education level and aversion, even when the relationship is non-linear.

Kendall Tau Correlation:



The Kendall correlation, also a non-parametric test, yielded a correlation of 0.0310 with a p-value of 0.527, further supporting the conclusion that education level does not have a significant association with aversion levels.

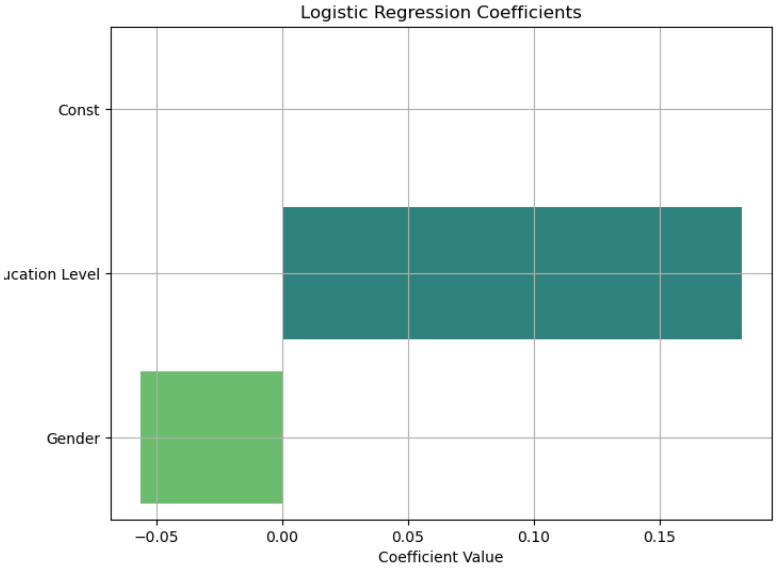
Point-biserial Correlation:



A point-biserial correlation was used to assess the relationship between High Aversion (binary variable: 1 for high aversion, 0 for low) and Education Level. The correlation was 0.0283, with a p-value of 0.608, indicating no significant relationship.

All correlation analyses consistently suggest that education level is not significantly associated with aversion to mainstream content. This confirms that education level is not a key factor in determining how individuals perceive and react to this content. Subsequently, we performed the logistic regression

Logistic regression



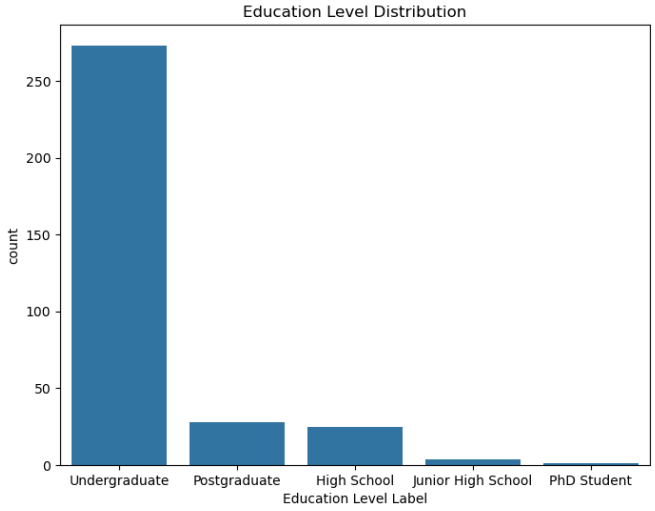
The logistic regression analysis aimed to determine whether education level and gender can predict the likelihood of individuals falling into the high aversion group (aversion levels 1-3) toward mainstream Chinese cultural content on Douyin. The findings indicated that both education level and gender had minimal effects on predicting high aversion.

The coefficient for education level was 0.1826, suggesting a slight increase in the likelihood of individuals with higher education being highly averse to mainstream content. However, the effect size was minimal, indicating that education is not a strong predictor. Similarly, the coefficient for gender was -0.0562, implying a slight trend where individuals in the reference gender group (likely males) were more likely to be highly averse compared to the other group (likely females). Again, this effect was small and not practically significant.

Overall, while there were slight trends, the effects of both education level and gender were weak, and their statistical significance was likely minimal. Therefore, neither demographic factor strongly predicts a high aversion to the content.

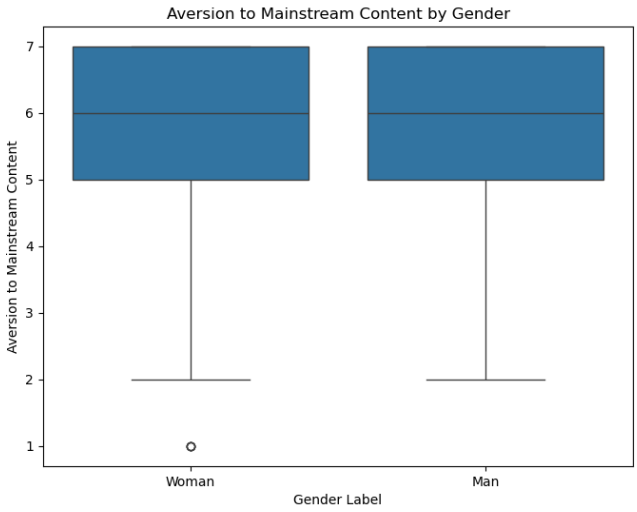
To better illustrate the distribution of the data, we created visualizations:

Education Level Distribution:



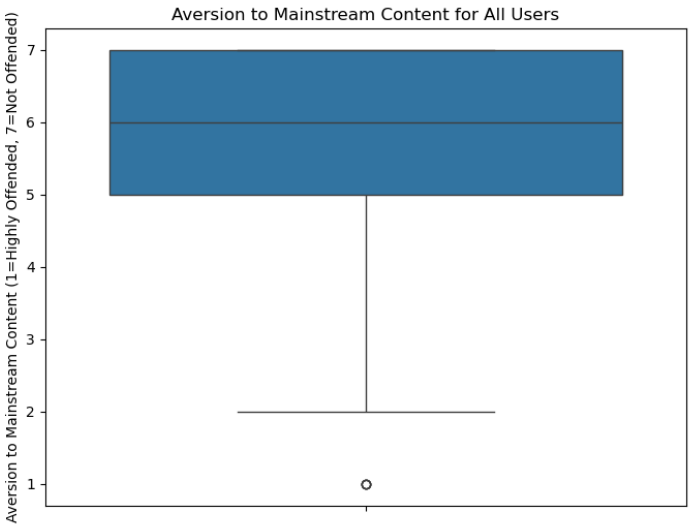
The majority of participants are at the undergraduate level, with smaller proportions from postgraduate, PhD, and high school levels

Aversion to Mainstream Content by Gender:



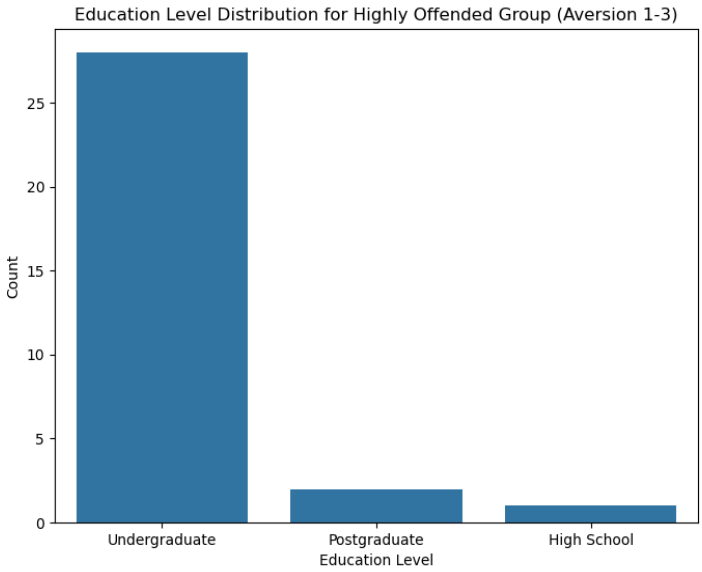
The boxplot shows that both men and women exhibit similar distributions of aversion levels, with no significant difference in their median scores or variability. This aligns with the logistic regression results, confirming that gender is not a significant factor.

Aversion to Mainstream Content for All Users:



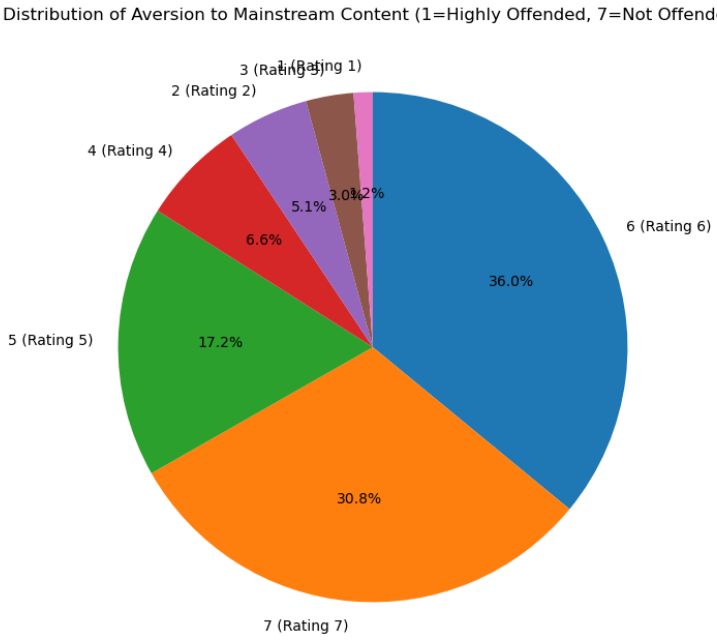
The majority of respondents reported aversion levels between 5 and 7, indicating that most are not highly offended by mainstream content. Only a small subset of users reported high aversion (scores between 1 and 3).

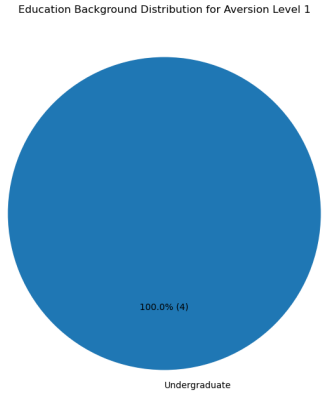
Education Level Distribution for Highly Offended Group:

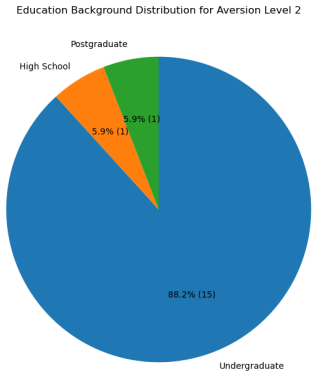


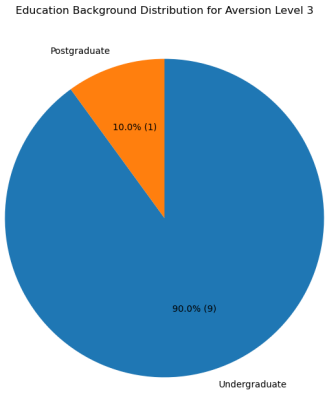
For users with high aversion (scores 1-3), the majority still come from the undergraduate level. However, this does not imply a causal relationship, as this group also forms the largest portion of the overall sample.

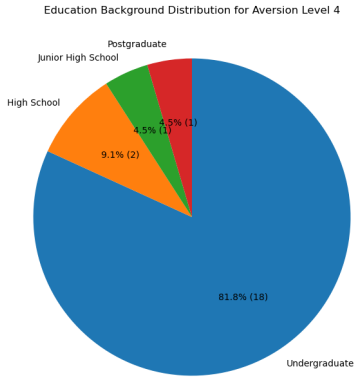
Pie Charts for Aversion Levels and pie charts for different Aversion Levels:

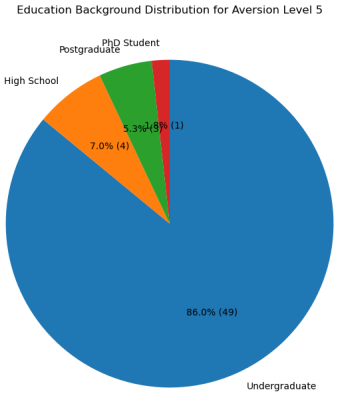


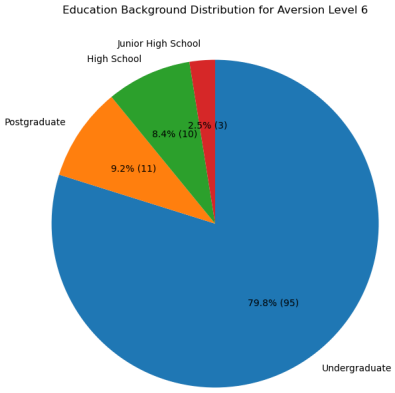

Education Background Distribution for Aversion Level 1
100.0% (4)
Undergraduate

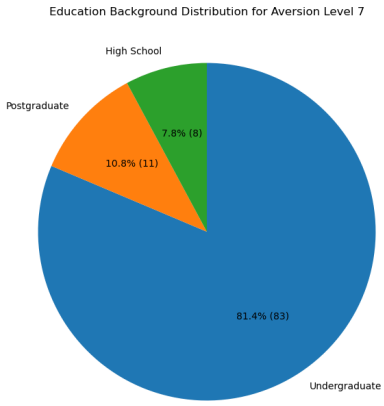

Education Background Distribution for Aversion Level 2
Postgraduate
High School
5.9% (1)
5.9% (1)
88.2% (15)
Undergraduate


Education Background Distribution for Aversion Level 3
Postgraduate
10.0% (1)
90.0% (9)
Undergraduate


Education Background Distribution for Aversion Level 4
Postgraduate
Junior High School
High School
4.5% (1)
4.5% (1)
9.1% (2)
81.8% (18)
Undergraduate


Education Background Distribution for Aversion Level 5
PhD Student
Postgraduate
High School
1.8% (1)
5.3% (3)
7.0% (4)
86.0% (49)
Undergraduate


Education Background Distribution for Aversion Level 6
Junior High School
High School
Postgraduate
2.5% (3)
8.4% (10)
9.2% (11)
79.8% (95)
Undergraduate



The pie chart illustrates that the largest proportion of respondents fall in the less offended category (6-7), confirming that most respondents are not highly offended by the mainstream content on Douyin.

The results of correlation tests, trend analyses, and regression models consistently showed that education level and gender did not significantly predict aversion to mainstream Chinese cultural content on Douyin. Despite the presence of a small correlation, none of the statistical tests indicated a meaningful or significant relationship. Also, most participants were undergraduates, and the group of undergraduates dominated all aversion levels, suggesting a lack of educational diversity in our sample. Therefore, our study concludes that based on the given dataset, education level and gender do not explain variation in aversion to mainstream content, thus the research question is not proven. While it is widely understood that higher education is generally associated with the development of critical thinking skills, which would ostensibly lead to increased scepticism towards mainstream narratives. Therefore, these findings are somewhat counterintuitive for us.

When interpreting these results, several limitations should be taken into account. First, the term "mainstream" has been politicised due to the political context, with the government and the official media of the Communist Party specifying it to mean "the socialist official propaganda of

the Communist Party." Throughout the life of every Chinese individual, who has been immersed in Chinese political propaganda, such a definition undoubtedly becomes a cognitive fixation, making the concept of "mainstream" simultaneously clear and ambiguous. When any Chinese person hears the term "mainstream," they invariably associate it with "the socialist official propaganda of the Communist Party." However, when discussing mainstream culture in the broader societal sense, the definition of "mainstream" suddenly becomes quite vague for them. This complexity was neither mentioned nor warned about in the dataset, and such an issue which is common knowledge, or even conventional wisdom for the Chinese people, is evidently difficult for researchers from other backgrounds to recognize. Differences in understanding of "mainstream" among researchers from different backgrounds could evidently result in inaccuracies in the final findings.

Secondly, the majority of participants in the dataset held undergraduate degrees, which limits educational diversity and the ability to detect significant differences in aversion levels across educational backgrounds. The sample also had a gender imbalance, with two-thirds of respondents being female, which may have weakened the ability to identify gender-based differences in media aversion. Additionally, focusing solely on Douyin restricted the scope of the study, as other major platforms like Weibo, Bilibili, and Kuaishou have different audiences and could provide valuable insights into diverse group interactions with mainstream content. Another limitation to consider is the possibility of self-censorship, especially among more educated respondents, due to the political sensitivity of the topic, which may have led to an underestimation of critical attitudes toward official content.

Sustainability

Access Project Data and Code

- GitHub Repository: Yulivu/Introduction-to-Digital-Humanities-and-Social-Analytics-Group7: Group 7's project repositoty (github.com)
- Information: A Dataset of Young People's Perception on Dissemination of Government Propaganda and Socialist Culture Content in China | Journal of Open Humanities Data (metajnl.com)
- Dataset DOI: A Dataset of Young People's Perception on Dissemination of Government Propaganda and Socialist Culture Content in China. - Mendeley Data
- Dataset Location in Repository: Introduction-to-Digital-Humanities-and-Social-Analytics-Group7/project/dataset at master · Yulivu/Introduction-to-Digital-Humanities-and-Social-Analytics-Group7 (github.com)

For instructions on how to use the code, please refer to the README file in this repository.

To ensure that the project's outcomes can be maintained and reused by future researchers, we have taken the following steps:

- Open Access: The GitHub repository has been made open source, ensuring open access to the code and dataset. We also regularly update the documentation based on new findings to maintain the accuracy of the information.
- Version Control: We use GitHub's version control functionality to record modifications and updates to the project, ensuring that every change is tracked and can be traced back to any previous version.
- Portability: The code includes clear portability instructions through detailed comments.

- Data Sources and Documentation: We have documented our data sources in the README file, and the final project documentation can be found in the repository, providing future researchers with a clear path for reproducing the research.

Reflection

As mentioned in the Challenges & Solutions section, initially we were unable to smoothly follow the workflow, which led to various doubts about our proposed workflow. However, as we gained more knowledge about digital humanities, we gradually reached a consensus during the middle stage of the project: although we all have solid foundations and understanding in our respective fields, we shouldn't approach the challenges within the project and collaboration with rigid thinking in such a new domain. It was crucial to let go of the stubbornness from our own disciplines and to respect the new perspectives brought by other fields. This helped reduce conflicts and enhance collaboration. With this shared understanding, we gradually adapted to the workflow and achieved high efficiency.

For future research, researchers should recognize that due to China's unique political nature and its complex socio-political context, certain terms or behaviors in China may have interpretations that are completely different from those in other parts of the world. They should caution against the risks arising from these differences and provide detailed explanations to help future researchers understand such potential risks. For more in-depth and accurate research on similar topics, expanding the sample to include participants from more diverse educational, geographic, and socioeconomic backgrounds, as well as users of different mainstream platforms, would help address these limitations. Incorporating data from a broader range of social media platforms could provide a more comprehensive understanding of media engagement within China's digital landscape. Additionally, including variables such as political orientation, media habits, and digital literacy could help capture the complex factors driving media skepticism. Longitudinal studies that track changes in attitudes over time, particularly in response to political or social shifts, would also offer deeper insights into the evolving nature of media engagement in politically controlled environments.

References & Literature

Li, Xinyu; Mohamed Salleh, Sabariah (2023), “A Dataset of Young People’s Perception on Dissemination of Government Propaganda and Socialist Culture Content in China.”, Mendeley Data, V1, doi: 10.17632/mzptp5cmr7.1

Chow Yiu Fai, Chow Yiu Fai, Living with Their Own Images, Caring in Times of Precarity, 10.1007/978-3-319-76898-4_2, (51-89), (2018).

Huang, H. Z. (2021). From Confrontation to Dialogue: Development and Reflection on Mobile Short Video in the Context of Youth Subculture. *Journal of Hengyang Normal University*, 42(01), 106–112. DOI: https://doi.org/10.13914/j.cnki.cn43-1453/z.2021.01.017

Yang Lei . Research on the image shaping of mainstream culture in contemporary China[D].2017.

A study on the content, characteristics and influencing factors of youth online political participation: Based on text analysis of relevant materials from seven Chinese forums [J]. Qi Guanghong, Wang Jianying, Yang Zhiqiang. Chinese Youth Research. 2012(10)

Ma, Z. H. (2013). 2012 Research Strategy on Chinese Youth Subculture. *Youth Exploration*, 06, 5–12. DOI: https://doi.org/10.13583/j.cnki.issn1004-3780.2013.06.003

Wang, T. Q., Tian, Y., & Xu, Z. H. (2023). Study on the Impact and Development of the Construction of Contemporary Mainstream and Subcultural Accommodation. *PR Magazine*, 14, 22–24. DOI: https://doi.org/10.16645/j.cnki.cn11-5281/c.2023.14.056

Zhang, G. Y. (2007). The impact of postmodernist thinking on youth subcultures. *Contemporary Youth Research*, 01, 35–47. DOI: https://doi.org/10.3969/j.issn.1006-1789.2007.01.008

Bourdieu, Pierre. *Distinction: A Social Critique of the Judgment of Taste*. Translated by Richard Nice, Harvard University Press, 1984.

Livingstone, Sonia, et al. "Critical Digital Literacy: Evaluating Media Education and Its Impact."

Journal of Media Literacy Education, vol. 11, no. 2, 2019, pp. 1-15.

Pew Research Center. "Public Perceptions of Media Bias and Views of News Outlets." *Pew Research Center*, 10 Jan. 2020, www.pewresearch.org. (The Pew Research Center provides data on how political beliefs and education levels influence trust and perceptions of bias in mainstream media).

James G. Webster, Thomas B. Ksiazek, The Dynamics of Audience Fragmentation: Public Attention in an Age of Digital Media, *Journal of Communication*, Volume 62, Issue 1, February 2012, Pages 39–56, https://doi.org/10.1111/j.1460-2466.2011.01616.x

Hallin, Daniel C., and Paolo Mancini. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge University Press, 2004.

Yang, Guobin. *The Red Guard Generation and Political Activism in China*. Columbia University Press, 2016.

Yin, Yiyi, and Anthony Fung. "Youth Online Cultural Participation and Bilibili: An Alternative Form of Democracy in China?." *Digital Media Integration for Participatory Democracy,* edited by Rocci Luppicini and Rachel Baarda, IGI Global, 2017, pp. 130-154. https://doi.org/10.4018/978-1-5225-2463-2.ch007

China Youth & Children Research Center. "The Ideological Trends of Chinese Youth." *Journal of Youth Studies*, 2021.

Zhuang Chen, Qian He, Zhifei Mao, Hwei-Ming Chung, and Sabita Maharjan. 2019. A study on the characteristics of douyin short videos and implications for edge caching. In Proceedings of the ACM Turing Celebration Conference - China (ACM TURC '19). Association for Computing Machinery, New York, NY, USA, Article 13, 1–6. https://doi.org/10.1145/3321408.3323082

Zhou, Wei, et al. "Douyin as an Educational Tool: How Short Videos Promote Knowledge Sharing." *Journal of Educational Technology*, 2020.

Zhao, J., & Zhang, D. (2024). Visual propaganda in Chinese central and local news agencies: a Douyin case study. Humanities And Social Sciences Communications, 11. doi:10.1057/s41599-

024-03059-5