Privilege and Representation: Analyzing Bias in TikTok College Admission Consultant Influencers

Ashley Cong
ac135@wellesley.edu
Wellesley College
Wellesley, MA, USA

Tiffany Lin
tl108@wellesley.edu
Wellesley College
Wellesley, MA, USA

Han Nguyen
hn103@wellesley.edu
Wellesley College
Wellesley, MA, USA

Claire Xue
yx105@wellesley.edu
Wellesley College
Wellesley, MA, USA

Abstract

As social media platforms become more integral to students' college application journeys, influencers specializing in college admissions counseling have emerged as key figures shaping perceptions of academic success. Yet little is known about how these influencers represent different demographic groups in the student profiles they share. This study examines TikTok content created by college admissions consulting influencers to investigate patterns of demographic representation and targeting. We focus on two research questions: (1) What factors in influencer content can be used to infer the demographics of their intended audience? (2) How do the topics and persuasion techniques used in videos differ across target demographics? Drawing on a content analysis of TikTok videos from independent counselors, we analyze indicators such as student gender, race, socioeconomic background, and influencers' language use. Utilizing machine learning techniques such as zero-shot classification and topic modeling, we systematically coded demographic indicators and thematic content. By comparing these portrayals against known demographic trends in college admissions, we aim to identify disparities in visibility and framing across student groups. Our findings show that the content of college consulting influencers is predominantly targeted towards white and/or male students, with a stronger focus on the sciences. Meanwhile, videos targeted towards females and lower socioeconomic statuses seem to focus more on emotional encouragement. Future research could focus on incorporating visual elements of videos and adopting more inclusive, non-binary definitions of demographic groups to capture the full spectrum of representation in online educational content.

Keywords

college consulting, tiktok influencers, college admission, college student demographics,

CS 315, Wellesley, MA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Ashley Cong, Tiffany Lin, Han Nguyen, and Claire Xue. 2025. Privilege and Representation: Analyzing Bias in TikTok College Admission Consultant Influencers. In *Proceedings of Data Science for the Web (CS 315).* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

While society has made significant progress in expanding access to higher education, persistent disparities remain across various demographic groups. Differences in college admissions outcomes, fields of study, and postsecondary achievement are often shaped by a student's gender, socioeconomic status, race, first-generation college status, and citizenship. Research shows that colleges sometimes favor less-qualified male applicants over more-qualified female counterparts in an effort to maintain gender balance on campus[20]. Gendered disparities also extend to academic interests and achievements: women remain underrepresented in STEM fields, and the proportion of women completing degrees in computer and information sciences has declined by 10% since 1990, as of 2021[41]. Other researchers also critique the holistic evaluations focused on extracurricular activities, revealing how definitions of "exceptional performance" systematically advantage higher-income, predominantly White students due to disparities in opportunity, specialization, and support[12].

In the age of technology, these issues gradually manifest themselves in digital spaces beyond real-life occurrences. Social media has evolved into a multi-functional platform that goes well beyond the purpose of entertainment. According to The 2017 Social Admissions Report, 63% of students use social media to research colleges they are interested in [35]. As a result, these platforms play a pivotal role in shaping prospective students' decisions about colleges. In a study conducted by the National Association for College Admission Counseling [35], 47% of the participants claimed that social media was an important factor that influenced their college decision. With its rise in visibility, social media has led to the rise of a new category of influencers: college admissions counseling influencers. These influencers create content related to the college admission process for students to refer to as guidance in completing their own college applications.

TikTok, the most downloaded app globally from 2020 to 2022 [15–17], plays an especially important role in this shift. Its short-form, algorithm-driven structure allows influencers to amplify narratives about successful applicants to vast audiences. Yet, questions remain about whose stories are most often shared—and which demographic profiles are amplified, idealized, or potentially underrepresented. Given that TikTok's recommendation algorithms personalize content based on user engagement, influencers may strategically showcase profiles they believe will resonate most with their target audiences. This dynamic raises concerns about the potential reinforcement of preexisting disparities in educational narratives.

This study focuses on college admissions counselors, specifically in a social media context. College admission counselors advise prospective college students during their college application process, in which they help students find the right colleges for them. Counselors are either independent or employed by colleges. Independent college counselors often receive payment when working with students and may operate solo or as part of an organization. In our study, we focus on independent college admissions counselors who promote their content on TikTok[7].

This study is guided by four assumptions: (1) Students from all demographic groups deserve equitable and accurate representation in educational narratives; (2) TikTok's scale and influence mean that its content can shape college aspirations, perceptions, and opportunities; (3) Admissions influencers, seeking to grow their audiences and monetize their content, may feature student profiles that align with their perceived audience demographics; (4) High school students, as a key audience for admissions influencers, may be particularly impressionable due to ongoing development of media literacy and critical evaluation skills[9] Based on these assumptions, we examine the demographic characteristics of student profiles highlighted by college admissions consultant influencers on TikTok. We focus on understanding patterns in representation, the demographic targeting strategies that may underlie influencer content, and how different student backgrounds are framed and discussed.

We hope to answer the following research questions:

RQ1: What factors in the content can be used to infer the demographics of their target audience?

RQ2: How do the topics and persuasion techniques used in videos differ across target demographics?

Our first research question will examine the representation of different demographics in content created by college admissions consultant influencers on TikTok. We hope to uncover potential disparities in visibility across demographics. Through understanding how certain groups are over- or underrepresented, we aim to relate these findings to the broader societal biases in college admission narratives.

Our second research question aims to analyze the word choice and sentiment of the content that TikTok college admissions consultant influencers use when sharing information about their students. We hope to discover insights into how these influencers use certain topics or persuasion techniques to target different demographics.

Through this research, we aim to uncover whether certain demographic groups—such as students of specific gender identities, racial backgrounds, or socioeconomic statuses—are portrayed differently or are more frequently featured in influencers' content. By examining frequency, framing, and linguistic patterns, we seek to understand how influencer narratives may reflect, amplify, or challenge broader patterns of inequality in the college admissions process.

The demographics of students that are talked about in the content of the videos provide indicators of what groups the college consulting influencers may be targeting. The assumption is that by mentioning specific demographics of students, the college consulting influencer is hoping to relate their content to a specific group so that it is effective in helping students of that group. If the influencer's content is well received by their target audience, then they will most likely see an increase in engagement and have a greater impact in the college consulting ecosystem.

The intended target audience, so the group(s) of students we believed that the influencers were targeting in their content, was specifically focused on in this research because there was no feasible way for us to get the demographic information of the students/users watching the content. A lot of users prefer to keep their information private and it wouldn't have provided the kind of information we were hoping to investigate, specifically how the influencers impacted the ecosystem. Additionally, even though we essentially used the demographics of the students mentioned in the video, they wouldn't have provided the kind of reasoning we were getting by focusing on the intended target audience.

Persuasion techniques are techniques that are used to convince the audience to agree with the person using them. The three main persuasion techniques that we focus on are: ethos, logos, pathos. Ethos being an appeal to credibility, logos being an appeal to logic and reasoning, and pathos being an appeal to emotion [33]. We decided to study the persuasion techniques and their presence in college consulting influencer content because we wanted to see if we could connect them back to how certain demographics of students were portrayed.

2 Literature Review

2.1 Social Media's Influence on College Admissions

Social media's influence on college decisions is a relatively new phenomenon that is being observed as society continues to innovate and evolve. However, that doesn't mean no one has been studying this situation. This section focuses on the influence social media has on prospective college students, and how influencers making college consulting content may be greatly affecting the users who watch their content.

Research on social media's influence on prospective students has shown that students often use it to collect information and ask questions about the universities they are researching and connect with students and alumni [1]. This makes it all the more important that all students have access to the information needed to make educated decisions about higher education. Additionally, research has shown that when influencers make authentic content and allow for more personal connection with their viewers, the impact the influencers have on students is greater. Utilizing current college students has been found to be especially effective in doing outreach with prospective students [19]. Current college students are the best resources to learn more about college, and a lot of them that make content really know how to incorporate their personalities into their content. Moreover, research has shown that students of every gender and ethnicity are equally influenced by social media when making their decisions [27]. It seems there's no denying the huge impact social media can have on students' college decisions and how effective and can be for college institutions. That being said, it's important to look out for any potential biases influencers might have that negatively affect students.

In addition to general research that has been done on social media's influence on prospective students, more focused research of social media's influence on students of underrepresented communities has been studied. One study has been done on how different types of social capital affected students' confidence regarding the college application process and how to succeed in college. In this study, the researchers found that social media plays an especially significant role in first-generation students as it allowed them to gain more knowledge about how to apply to college and succeed. These students were able to access information that didn't require too many resources such as money, and they were able to join a network of people who would answer their questions [38]. In another study focused on underrepresented students in Detroit, researchers were able to find out how social media influencers were able to help students develop aspirations to attend college, navigate the admissions process, and engage with a supportive community, whilst building a community of people who were able to resist stereotypes about them [4]. The influence of social media on underrepresented students is an important note to keep in mind as it brings more context regarding how students may react to influencers who have biases running around in their content. This also emphasizes how important social media can be in helping underrepresented students gain more opportunities and learn how to build a community.

Additionally, since our research topic includes a focus on possible gender biases in college admissions influencers' content we can reference research that delves into the roles of mentors in female students' college decision process. The study's findings showed that female students are heavily influenced when it comes to their choice of college. Specifically, female students were influenced by people such as their friends and families, as well as third-parties [37]. This knowledge is especially useful considering it may be helpful when interpreting our data analysis. It highlights how female students may be more influenced by college admissions influencers than others.

2.2 Existing Disparities in College Admissions

Race-conscious admissions remain central to contemporary debates. Past research underscores significant racial disparities in acceptance rates at students' first-choice colleges, highlighting that Black and Asian students experience substantially lower odds of acceptance compared to White students. Although studies show that such disparities diminish at highly competitive institutions, which may indicate that selectivity level potentially mediates racial disparities, nevertheless systemic inequities persist broadly in admissions outcomes. Other studies found disadvantages faced by Asian Americans students in elite college admissions, suggesting that despite comparable academic and extracurricular achievements, Asian American applicants, particularly those of South Asian descent, have significantly lower probabilities of acceptance at highly selective Ivy League institutions. Legacy admissions policies and geographic biases may further compound these disparities, systematically favoring White students who disproportionately benefit from alumni relationships and favorable regional policies. These concerns surrounding legacy admissions have drawn scrutiny for many years. A study specifically highlights ongoing legal challenges, such as the Chica Project complaint against Harvard, underscoring that legacy and donor admissions policies potentially violate civil rights legislation by disproportionately disadvantaging racial minority applicants.

The intersectionality of race and socioeconomic status further complicates admission dynamics. Studies have found that the socioeconomic status (SES) of students can predict differential use of admission-enhancing strategies. Researchers claim that inequality is further exacerbated and perpetuated through the student's academic and extracurricular activities on their college applications. This stratification indicates that high-SES students consistently capitalize on privileged access to resources and strategic advantages, exacerbating existing inequalities in college admissions. Holistic admissions practices is another aspect of debate in the college application process. Researchers found that individualized holistic reviews through evaluations of letters of recommendation could effectively identify and uplift students from disadvantaged contexts.

Studies have demonstrated that gender disparities exist in college admissions in subtle ways that shape access to higher education. For example, empirical data from Nigerian tertiary institutions reveal persistent gender gaps in admission rates, highlighting that equal access remains an ongoing challenge [26]. Similarly, research on Uganda's Affirmative Action found that while it increased access for some women, it primarily benefited specific demographics rather than the most historically disadvantaged groups [25]. These findings suggest that policies aimed at gender equity in admissions do not always achieve uniform benefits across gender identities and socio-economic backgrounds. Gender inequities in standardized testing also contribute to disparities in college admissions. The National Merit Scholarship Competition has been criticized for favoring male students due to gendered differences in PSAT scoring patterns [23]. These biases are further reflected in SAT performance, where male students consistently outperform female students on the quantitative section, despite evidence that these differences do not correlate with actual college performance [22]. This discrepancy suggests that standardized testing mechanisms may undervalue female students' potential, reinforcing gender disparities in admissions.

2.3 Persuasion Techniques Across Demographics

Understanding how persuasion operates in social media environments is central to our investigation of how college admissions influencers tailor content to specific demographic audiences. Our research asks not only which groups are being addressed but also how influencers communicate differently depending on the presumed identity of their audience. Recent scholarship across communication theory, psychology, and machine learning provides valuable insight into the mechanisms of influence that may be present in influencer content and how they interact with demographic traits.

Persuasion strategies used by influencers—such as appeals to emotion (pathos), credibility (ethos), or logic (logos)—are known to shape audience responses. However, the effectiveness of these techniques can differ based on demographic factors such as age, gender, SES, or cultural background. Recent studies highlight that persuasion is not one-size-fits-all—different audiences respond to different rhetorical strategies depending on their psychological traits, belief systems, and social identities [3]. They found that emotional appeals and rhetorical devices were more persuasive than authority-based strategies, especially when matched to individuals' psychometric profiles. This is directly relevant to our research, as influencers may unconsciously use pathos-driven storytelling for first-generation or low-income viewers, while using logos or ethos-based appeals for high-SES audiences. Other researchers also showed that influencers build trust and action through tactics like social proof and community-building [30]. This further supports the idea that persuasive techniques are often strategically aligned with the presumed demographic identity of the viewer.

Other studies show that persuasion in the digital era operates through repetition and perceived consensus rather than factual accuracy [10]. In this sense, influencers who repeatedly promote certain narratives may shape beliefs indirectly—especially when algorithmic amplification favors widely shared, "general"-seeming content. This may help explain the high rate of "general" labels in our zero-shot classification results. Another study shows that adolescents form mental associations between influencer traits and their own decisions, suggesting that persuasion is embedded in identity [18]. For our study, this means that demographic targeting often occurs through tone and self-presentation, not just through explicit labels—limiting the effectiveness of literal methods like regex.

Our study extends this literature by analyzing persuasion at scale using both linguistic rules and machine learning, showing how different demographics are not only addressed but differently persuaded in college admissions content on TikTok.

2.4 Extending on Existing Literature

Our research extends the literature by analyzing how college admissions influencers use targeted persuasive strategies that may implicitly favor or exclude certain demographic groups. We go beyond previous work by examining how different demographics are not just passively influenced by social media, but how they may be differently addressed by influencer content through strategic rhetoric, tone, and topic emphasis. We challenge the assumption that social media influence is demographically neutral by using machine learning tools to analyze demographic targeting patterns in content. Additionally, our project complements existing work on admissions disparities by identifying a new layer of digital influence: persuasive messaging on platforms like TikTok, which may reinforce existing inequalities or provide new forms of access and inspiration. This highlights the role of influencers as unofficial gatekeepers in the admissions landscape and suggests that digital spaces are not merely neutral arenas of outreach, but active sites of stratification and strategic messaging.

3 Data and Methods

3.1 Data Collection

Hashtag Collection. In line with previous research about studying content [2] and influencers on Tiktok, we first developed an initial set of seed hashtags related to college admission for data collection. We used the TikTok Search Engine to search for lingos that are most associated with college admissions listed on College Board's college admission glossary [5]. Some of the seed hashtags we used include "college application essay", "admission process", "financial aid", "admission tests", and "common application". Using this set of seed hashtags, we conducted iterative searches on TikTok using Selenium to collect additional hashtags related to college admission counseling. We then manually verified the relevance of these hashtags, narrowing it down to a final set of 52. We manually removed some noisy hashtags such as "college" as it is too broad, and "collegedecision" since it is more focused on reaction to college admission results. From these hashtags, we then collected data for 9,574 algorithmically determined videos. From these videos, we then extracted 3,144 unique accounts.

Influencer Filtering. To reduce noise from potentially irrelevant content, we aimed to only include accounts of independent college admission counselors. Since a Tiktok bio allows influencers to introduce who they are and announce what followers can expect from them, we examine each account's bio to verify whether they are considered college admission counseling influencers [31]. Specifically, similar to prior research on social media users, we used account bios as the users' personal identifier [28]. To achieve the largest number of accounts with highest accuracy, we tested 2 different methods to filter out accounts by bio: rule-based regular expression and zero shot learning.

Similar to past research on identifying social media accounts using regex [39], we created a set of patterns and phrases that are commonly found in college admission counselors influencers' bios. Some of the patterns used include "college application", "former admissions", "common app", and test scores like "SAT" or "ACT". We then searched for these patterns in each account's bio and flag as a potential college admissions influencer if a bio matched at least 2 of these patterns. After narrowing the set to 178 accounts, we manually reviewed each one to resolve any ambiguity in the bios. We achieve a 68% accuracy with using regular expressions to filter out accounts.

Next, we applied zero-shot learning, an NLP inference technique that classifies data points without labeled pre-labelled data to extract semantic meaning [40]. We first designed and fed label descriptions like "college counselor" and "education influencer" into a pre-trained language model to determine whether a Tiktok account fits our target profile. This resulting refined dataset contains 380 influencers. Similar to the regex-based approach, we manually reviewed these accounts and found that 66.6%, or 253 of these accounts satisfy our definition of college admission counseling influencers. Since both techniques yielded comparable accuracy, we chose to use the set of accounts filtered by zero-shot learning, as it captured a broader range of relevant accounts.

Video Collection. From the set of 253 relevant accounts, we then collected all videos that these accounts posted after 01/01/2021 to

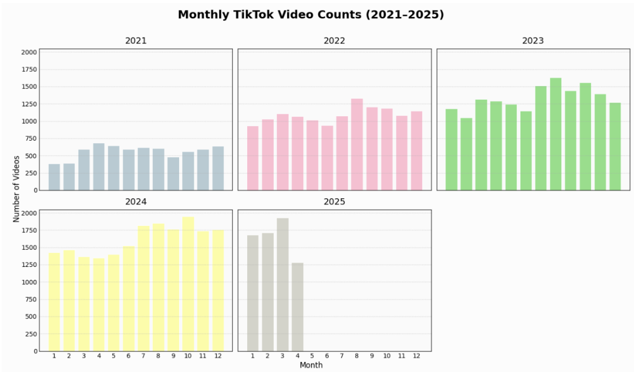

Figure 1: Monthly TikTok Video Counts (2021-2025)

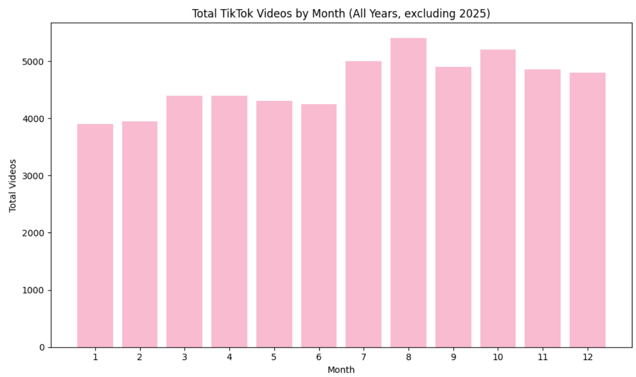

Figure 2: Total TikTok Videos by Month (All Years, excluding 2025)

ensure that our analysis reflects the significant shifts in college admissions practices following COVID-19 pandemic [6]. To further ensure that our analysis focuses on relevant content, we use stratified random sampling to choose videos from only the most relevant months based on post volume and context understanding by month. As shown in Figure 1, video activity in aggregated data across all years (excluding 2025 as we don't have enough data for all months in 2025), activity peaks in March, April, August, September, October, November, and December. This trend holds across posting distribution across individual years as well, as seen in Figure 2. The selection of these months also aligns with key phases in the U.S. college admission cycle. August to December are critical months for application preparation for early decision and early action application submissions, test preparation, recommendation gathering while March-April is the time when admission decisions are out with major enrollment choices, financial aid considerations [14]. By focusing on these months, we capture the peaks in the admission process when students are most actively seeking guidance and counselors are most involved.

Once we established the months from which we would be collecting videos, we needed to determine how many videos each account would contribute per month. Since we're looking to assess

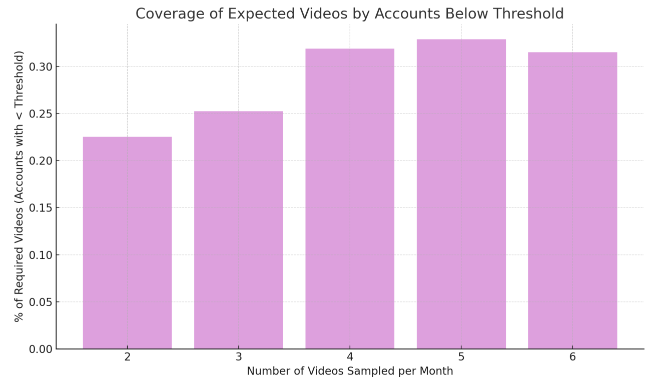

Figure 3: Coverage of Expected Videos by Accounts Below Threshold

Total number of accounts	209
Total number of videos	4344
Number of months collected	7
Number of videos collected per month	5
Number of videos expected per influencer	35

Figure 4: Statistics of final set of accounts and videos

the college counseling content target audience ecosystem, it is necessary that each account apportions a relative amount of videos so that the independent creator dataset will be generalizable. Yet some accounts do not produce enough videos to fulfill our desired high quotas. As seen in Figure 3, 5 videos per month is the compromise where accounts falling short of required contribution still effectively contribute an average of over 33% of their expected value. In other words, those accounts that post low amounts will not be eliminated from the study and their content will be substantially accounted for in the larger dataset.

We then randomly select 5 videos per month for all the chosen months. For months that an influencer doesn't have enough 5 videos, we will randomly get more videos from other months. This means that for accounts that have less than the required amount of videos overall, we will select all of their videos posted to ensure representation. We then download these videos for further transcription, which narrows down our dataset to 209 influencers and 4344 videos as some influencers disabled their downloading option.

3.2 Transcribing Videos

After collecting all the video data from each influencer, we used WhisperAI to transcribe the videos in order to extract the content/text from each of the videos. WhisperAI is an open-source AI model designed for automatic speech recognition. Whisper is able to transcribe speech into text for many languages and can also recognize what language the speech is in.

In addition to using WhisperAI to transcribe audio from videos we can use pytesseract to get text embedded in images/frames of videos. Pytesseract is a python library under Tesseract OCR (optical character recognition). The library helps extract text from images into machine readable data.

Previous research that has utilized Whisper has shown that it is effective in documenting meetings and relatively good at detecting the correct and incorrect words. However, it does struggle with words that are used in rare contexts or are non-words [13]. Additionally, previous research has shown that pytesseract is extremely useful for real world applications [32].

Transcribing the text from the Tiktok videos we collected is necessary for us to work with our data in a meaningful way. Tools such as Whisper and pytesseract make the process of extracting and processing data more efficient. Instead of manually transcribing everything that the videos say, one can just feed the video into Whisper and have a transcript ready to use. Once the data is transcribed, we can then conduct further analysis on the content of the videos. We aim to find meaningful contributions from the transcriptions of the video and if we can make any meaningful insights about the ecosystem of college consulting influencers on Tiktok.

3.3 Extracting Demographic Indicators

To investigate how college admissions influencers on TikTok communicate with and potentially target different demographic audiences, we employed a mixed-methods approach combining machine learning-based natural language processing (NLP) and rule-based text analysis. We developed two primary computational approaches to infer the intended audience of each video based on transcript data: regular expressions (regex) and zero-shot classification. For each method, we implemented validation strategies including accuracy checks, human spot-checks, and error analysis to assess reliability and identify limitations.

Regular Expression. We applied a regular expression (regex) approach designed to detect explicit mentions of identity markers in transcript text. We created hand-crafted regex patterns that captured mentions of demographic traits such as gender (e.g., "girl," "she/her"), race (e.g., "Asian," "Black"), income level (e.g., "low-income," "full pay"), and parental education status (e.g., "first-gen"). For each video, we applied these patterns to the transcript to count matches and extract the specific terms used. This yielded binary indicators for each demographic category as well as frequency counts. If one or more matches were found, the corresponding demographic label was assigned to the video as part of its inferred audience profile. The strength of the regex approach lies in its transparency and interpretability. Each label assignment can be directly traced to the matched terms, enabling straightforward audits and corrections. However, the method is limited in scope. Because it depends on exact keyword matching, it often fails to detect relevant content expressed in more indirect or varied language. For example, phrases like "my parents didn't go to college" or "I had to work part-time to support my family" may indicate first-generation or low-income status, but would be missed unless explicitly encoded in the regex patterns. This leads to a low recall rate, where many demographically-relevant videos go undetected.

Zero-Shot Classification. To overcome the limitations of literal keyword detection, we implemented a zero-shot classification pipeline using the facebook/bart-large-mnli transformer model. This approach allowed us to classify transcripts according to their inferred audience without needing labeled training data. We used a set of candidate labels representing possible target demographics, including "low income," "international students," "people of color," "first generation," "high socioeconomic status," "male," "female," and "general." For each transcript, the model returned a ranked list of labels with associated confidence scores, allowing us to tag the most likely demographic groups each video was addressing. Because this model is trained to understand natural language inference, it can assign labels even when the transcript uses nuanced or indirect language.

The zero-shot approach enabled broader coverage compared to regex. It was able to infer audience intent in videos that did not contain explicit demographic keywords but whose tone, examples, or narrative implied relevance to a particular group. However, the model sometimes returned false positives, especially in cases where content was generic or where the model made overconfident assumptions. For instance, transcripts discussing general financial aid were sometimes incorrectly labeled as "low income" even when no such group was clearly addressed.

3.4 Identifying Variations in Video Content

Non-negative Matrix Factorization for Identifying Content Types. To identify topics from the collected video transcripts, we employed a topic modelling pipeline. We cleaned and normalized all video transcripts, and vectorized the processed texts to construct a TF-IDF matrix. We then tested three unsupervised topic modelling methods, Latent Dirichlet Allocation (LDA), KMeans Clustering, and Non-negative Matrix Factorization (NMF), on the constructed sparse matrix respectively for comparisons for optimal results. To identify the optimal number of topics, we used silhouette score to evaluate the KMeans model, perplexity to evaluate the LDA model, and reconstruction error to evaluate the NMF model. We proceeded with the optimal model, NMF, to assign each video transcript to a topic out of 15 optimal topics as evaluated through the reconstruction error for the model. We then extracted ten top keywords for each identified topic and included them as a part of the topics definition.

To have meaningful topics, in addition to identifying latent topics, we created one topic for all video transcripts that have less than 100 characters to filter out short or mis-transcribed texts. Because the NMF model identified both stylistic (topics identified with keywords such as "know", "don't", and "probably") and thematic topics (topics identified with keywords such as "college", "admissions", and "common") yet we interested in the variations and types of thematic topics, we grouped all stylistic topics into one topic. Therefore, we have one topic for all short texts (<100 characters), one topic for styles and presentation of language, and twelve topics for college admissions consulting contents among videos on TikTok. For results, we printed all the resulting topics, each of which included a topic label defined by the three top terms and a complementary definition of the ten top keywords associated with the topic, except for the short text topic. We fed the resulting topics define by

3.5 Identifying Persuasive Appeals

Lexicon-based Analysis. To identify persuasive appeals— ethos, logos and pathos—, we first used a lexicon based approach, relying upon dictionaries for each appeal through regular expressions (regex). The compiled words assessed distinction across a comprehensive dictionary. Ethos was found with relative words of "authority," "expert," "credentials"; logos was found with "logic," "evidence," "statistics"; and pathos was found with "passion," "feelings," "emotion." Regular expressions were generated so that they found these words in the text body and thus detection was made for overt patterning of language associated with each appeal.

Such lexicon based approaches are highly successful due to their interpretability and application ease. Studies that apply lexicon based rhetorical analysis benefit the findings. For instance, in a 2022 study by Hairul Azhar Mohamad which sought to explore rhetorical appeals from research abstracts in English as a Native Language (ENL) or as a Second Language (ESL), he too used a lexicon based method through finding instances of ethos, logos and pathos for the corresponding appeals were evaluated on linguistic features and rhetorical strategies at different levels [21]. However, because these are hard-coded dictionaries, they may omit certain words/phrases.

Established contextual nuances as well as rhetorical appeals rendered in an implied fashion could become disbanded through lexicon based efforts. Thus, since context is what shifts meanings of words, a reliance upon a lexicon does not account for such shifts properly. This is supported by the assertion that supplemental to a lexicon based approach, machine learning is required to grasp contextualized manifestations.

Multi-Label Text Classification. Alongside the lexicon approach, a machine learning approach was trained via Term Frequency–Inverse Document Frequency (TF-IDF) vectorization along with a One-vs-Rest (OvR) logistic regression. TF-IDF creates sparse vectors from documents by weighting the word frequency against its inverse document frequency; frequently used words are down weighted while words that distinguish documents across the corpus are weighted more [29]. Therefore, this captures nuances of word frequency and distinction at the document level advantageously, rendering it suitable for linear application against text over time.

Since we use lexicon-based labelled data where each row can have multiple appeal labelings, we utilize a One-vs-Rest approach where a binary logistic regression classifier is trained per rhetorical appeal—ethos, logos, pathos. Thus, each model builds and evaluates regression independent from one another to either add or remove the class in question. This is critical because classes can overlap in persuasive works as ethos can be conveyed in addition to pathos and logos, all at once [34]. We use logistic regression due to its capacity to convey probabilistic results in an understandable fashion while also being robust in extensive feature space commonly found with TF-IDF usage. This approach strengthens our intersection of hidden rhetorical moves that are not always reflected by keyword patterns.

4 Results

4.1 RQ1: What factors in the content can be used to infer the demographics of their target audience?

To analyze the inferred target demographics of college admissions influencer content, we applied both regular expression (regex) and zero-shot classification methods to a dataset of TikTok video transcripts.

Zero-Shot Classification Results. Figure 5 shows that the most frequently assigned demographic label was "general," appearing in nearly 1,500 videos, followed by "talking about female" and "talking about male." The large number of "general" labels suggests that the model often defaulted to a non-specific demographic, potentially due to ambiguous or vague language in transcripts. This may also be an artifact of transcription noise from Whisper AI, which, despite its strengths, can introduce background speech, music, or filler content that makes it harder for the model to detect clear audience targeting. All videos were labeled with at least one demographic category, although some of the additional labels assigned have very low confidence scores.

Interestingly, female-targeted content appeared more frequently than male-targeted content, though both were prominent. This pattern could reflect influencer strategies that engage explicitly with gender identities, often tailoring advice or storytelling to resonate with gender-specific experiences in college admissions. Other demographic labels, including "high socioeconomic status," "international students," and "people of color," appeared far less frequently in the output, which may suggest either underrepresentation in the content itself or the model's limitations in detecting more nuanced identity targeting. Human spot-checks of the labels indicated that many "general" classifications reflected cases where the content addressed broad student audiences without demographic specificity—though in some cases, more specific targeting may have been missed due to vague wording. The discrepancy between the "general" label and specific demographic labels raises questions about the zero-shot classifier's scoring system. We conducted manual inspections that revealed most videos assigned with "general" labels often received high confidence scores in other demographic categories, such as "male" or "female." Despite the absence of targeted audience cues, since many of the "general" labeled videos contain transcripts from background music or unrelated commentary, the model still assigned demographic labels to these videos, indicating a potential overfitting issue where the classifier attempts to categorize noise as meaningful content. These misclassifications suggest that the model may be overly sensitive to superficial language patterns or non-contextual cues, leading to demographic assignments where no genuine targeting exists.

Regex-Based Inference Results. Unlike the zero-shot model, regex requires exact term matches, leading to a more conservative set of demographic inferences. Out of 209 influencers, regex identified that 92.3% of them mention at least 1 demographic and 26.01% of all the videos posted target at least 1 demographic. Figure 6 shows the proportion of videos that explicitly mention each demographic group. Male-related terms (e.g., "he," "guy," "young man") appeared

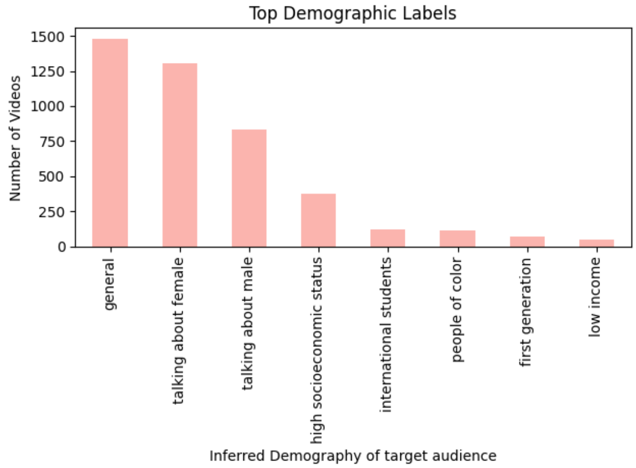

Figure 5: Top Demographic Labels

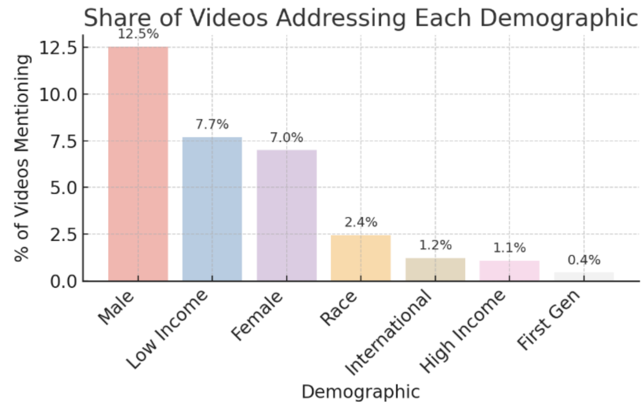

Figure 6: Share of Videos Addressing Each Demographic

in 12.5% of all videos, the highest proportion of any demographic. This aligns with our earlier finding that regex is particularly effective at capturing explicit and straightforward identity terms. Low-income references also stood out, representing 7.7% of videos and over 1,500 total mentions. These included phrases such as “financial aid,” “pell grant,” or “need-based,” which often appear in influencer content related to affordability or scholarships.

Mentions of female-related terms, race, and international status followed, though with significantly lower frequency. Notably, first-generation and high-income groups were underrepresented in both share and total mentions, possibly indicating either a genuine lack of references or the limitations of regex in identifying paraphrased or implied descriptors like “my family didn’t go to college” or “I paid full tuition.” In total, regex matched at least one demographic keyword in approximately 35% of videos, meaning a large portion of content had no detectable identity markers under this method.

Qualitatively, regex yielded high precision in labeling—over 90% of matches were judged accurate during spot-checks—but recall was limited, estimated below 60%, as the approach often failed to capture nuanced or informal phrasing. By contrast, zero-shot classification had wider coverage but occasionally introduced false positives, particularly when ambiguous or general statements were interpreted as targeting a demographic.

Comparing the two methods reveals complementary strengths. Regex is better in detecting overt identity references, making it useful for surfacing concrete examples of demographic targeting. However, it underestimates the true extent of demographic content due to its rigidity. Zero-shot classification is more effective at identifying implicit or nuanced targeting, such as advice tailored toward first-gen students without naming them directly. Yet, its broad interpretation sometimes misfires in noisy transcripts, leading to inflated counts of “general” or incorrect demographic labels.

The presence of “male” as the top regex label and “general” as the top zero-shot label underscores this difference: regex detects what is said, while zero-shot infers what is implied—even at the risk of error. Together, these results suggest that influencer content frequently targets gender identities and socioeconomic issues, with male-related references being more common than female under regex but reversed under zero-shot. Other identities, such as race and first-generation status, appear less frequently and may require more sensitive detection methods or richer contextual analysis to identify effectively.

To further understand the relationship between demographic labelings from regex, we analyze the distribution of explicit words mentioned and the co-occurrence patterns of demographics. Figure 7 reinforces that gendered terms are the most apparent demographic cues, with the highest number of indicators stemming from "he," "she," and guys". "Financial aid" and "low income" are frequently used as well, meaning that financial/funding related discussions serve as a strong indicator of content directed to lower socioeconomic status populations.

Figure 8 provides further meaning of where these demographics frequently overlap. For example: - Mentions of male co-occurs with race (890), suggesting that content directed to men will also call upon racial identity. - Mentions of low income strongly co-occurs with race (941) and high income (103), meaning that class distinctions are made within discussions of financial opportunity. - International students co-occur the most with low-income (312) and female (78), implying that some of these influencers render international applicants as a specific, less resourced population—females, in particular.

Ultimately, these trends suggest that gender and income are the most explicitly targeted demographic features called out in the language of the influencers. Regex easily finds such explicit commentary but does not fare as well finding more subtle, implied targeting—especially for populations like first-gen status or intersectional combinations. However, when combined with zero-shot findings which better find implied targeting, a more comprehensive picture of audience segmentation emerges.

4.2 RQ2: How do the topics and persuasion techniques used in videos differ across target demographics?

Topic Modeling (NMF). Figure 9 shows 15 optimal latent topics among the contents of all video transcripts collected. Each topic is

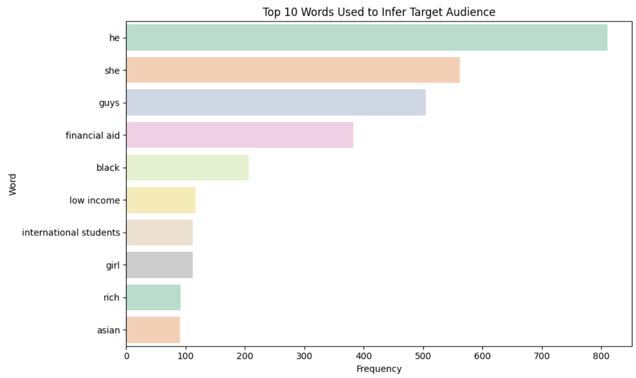

Figure 7: Top 10 Words Used to Infer Target Audience

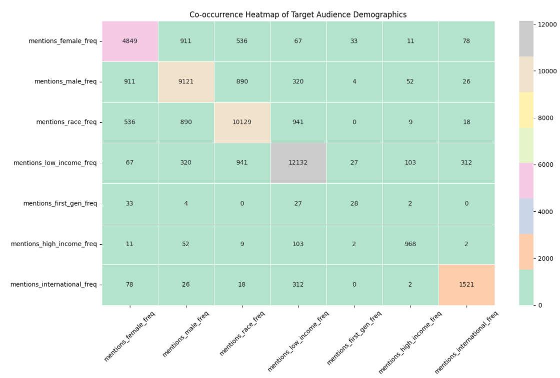

Figure 8: Co-occurrence Heatmap of Target Audience Demographics

defined by the top three terms of the topic, with a complementary list of top ten words of the topic. To make the topics more intuitive, we created a label for each topic manually with the results from the NMF model. After relabeling with ChatGPT, the 15 identified latent topics (presented as the relabeled topics here) among all videos are: Elite & Public Universities, Essay Writing Tips, Application Guidance & Counseling, SAT / ACT Prep, Grad School Admissions (MBA, Law), Common App & Activities Section, Spanish-Language College Guidance, High School & Enrichment Programs, Transfer & Admissions Coaching, Early vs. Regular Decision Strategy, Financial Aid & Scholarships, College Rankings & Waitlists, Essay Dos & Don'ts, Style & Presentation (Stylistic), and Short Text (<100 characters).

Figure 10 shows the top three video topics by their amounts of videos are 1. Elite & Public Universities (1061 videos), 2. Essay Writing Tips (622 videos), 3. Application Guidance & Counseling (546 videos). There are exceptionally many videos that talk about universities, both elite and private, as the number of videos that fall under the topic Elite & Public Universities (1061) is almost double the amount of the videos that fall under the second largest topic, Essay Writing Tips (622). After the top three video topics by amount, there are six topics with median appearances (200 to 500 videos per topic) and six topics with rather rare appearances (under 200 videos per topic). The six topics with median appearances (200 to 500 videos per topic) are: SAT / ACT Prep, Grad School Admissions (MBA, Law), Common App & Activities Section, Spanish-Language College Guidance, High School & Enrichment Programs, Transfer & Admissions Coaching; and the six topics with rather rare appearances (under 200 videos per topic) are Early vs. Regular Decision Strategy, Financial Aid & Scholarships, College Rankings & Waitlists, Essay Dos & Don'ts, Style & Presentation (Stylistic), and Short Text (<100 characters). The numbers show that the kind of videos with the most appearances on TikTok within the realm of college admissions consulting is about introducing all kinds of universities. Moreover, the appearances of videos for topics decrease as the specificity of the topic increases.

Across all topics, males are the topic one target audience by the amount of videos. After males, the second largest target audience body across all topics are low income groups, and the third largest target audience across all topics are females. For an inter-group comparison, there are more videos that target males rather than females (figure 12) and more videos that target low income groups rather than high income groups (figure 13). The inter-groups disparities between males and females as target audiences indicate that there may be implicit gender bias among college admission consulting content creators in tailoring topics towards males rather than females while females consist of the larger shares of college attendances. The implicit bias would be persisted through the circulation of the unproportional emphasis on genders in all the college admission consulting videos on TikTok. The other inter-groups disparities between low income and high income groups could infer a conjecture that TikTok as a platform free of charge assumed the target audience to be people who seek free advice. The high income body, who could afford finding college admission consulting service offline, are less of the target audience for free advice, while low income groups, who would look for free advice, in this case conveniently on TikTok, naturally became the target audience for whom to tailor contents. Yet, as it can be seen on figure 6, males and females have a larger intergroups disparity than low income and high income groups, as the ranks of these four target audiences by the amounts of videos are 1. males, 2. low income groups, 3. high income groups, and 4. females.

Persuasion Appeals. Our analysis reveals patterns in the usage of persuasion appeals across different demographics groups. We undertook two approaches including a lexicon-based detection approach via regular expressions and a multi-label One-vs-Rest logistic regression classifier.

Our classifier yielded a strong performance for logo appeal (F1 = 0.84) but considerably lower scores for both ethos F1 = 0.64) and pathos (F1 = 0.66). We suggest that this underperformance is due to our training set being unbalanced—over 51% of all detected appeals were logos. Given the performance gap and inconsistency of data with across, we decided to go with lexicon-based detections for our final evaluations as it seems less ambitious yet more clear and consistent across the entire video sample.

As shown in figure 14, only 46.5% of videos had any detectable form of persuasive appeal and while some of these groups possessed great amounts—mainly in terms of logos—many others did not.

Topic ID	Topic Label Resulted from the NMF Model	Relabeled Topic Name
0	(university & state & california - Thematic): university, state, california, harvard, florida, texas, michigan, thanks, columbia, watching	Elite & Public Universities
1	(essay & essays & write - Thematic): essay, essays, write, writing, personal, prompt, prompts, tips, story, supplemental	Essay Writing Tips
2	(college & help & application - Thematic): college, help, application, admissions, community, apps, colleges, process, counselor, essays	Application Guidance & Counseling
3	(sat & test & score - Thematic): sat, test, score, act, scores, questions, math, prep, practice, question	SAT / ACT Prep
4	(admissions & director & mba - Thematic): admissions, director, mba, wharton, upenn, law, legatt, aviva, talk, consultant	Grad School Admissions (MBA, Law)
5	(app & common & application - Thematic): app, common, application, activities, section, list, start, thank, personal, colleges	Common App & Activities Section
6	(que & universidad & para - Thematic): que, universidad, para, las, los, con, una, más, como, del	Spanish-Language College Guidance
7	(school & high & students - Thematic): school, high, students, classes, year, program, schools, business, student, research	High School & Enrichment Programs
8	(coach & advice & transfer - Thematic): coach, advice, transfer, admissions, essay, penn, graduate, year, freshman, college	Transfer & Admissions Coaching
9	(early & decision & action - Thematic): early, decision, action, regular, apply, application, deadline, admission, november, schools	Early vs. Regular Decision Strategy
10	(scholarship & scholarships & aid - Thematic): scholarship, scholarships, aid, financial, students, apply, international, money, need, fafsa	Financial Aid & Scholarships
11	(colleges & waitlist & ranking - Thematic): colleges, waitlist, ranking, stanford, acceptance, rate, list, wait, likely, harvard	College Rankings & Waitlists
12	(sentence & avoid & essay - Thematic): sentence, avoid, essay, college, undergrad, grad, school, specificity, campus, graduate	Essay Dos & Don'ts
13	(Style & Presentation - Combined Stylistic): like, youre, dont, know, really	Style & Presentation (Stylistic)
14	(Short Text (<100 chars)): [No keywords - texts too short]	Short Text (<100 characters)

Figure 9: Labels of all identified latent topics through the NMF model. The figure includes the topic ID of each topic, the topic labels defined by the top three terms associated with each topic and a complementary list of the top ten terms associated with each topic resulting from the NMF model, and the manually created labels for better readability with the top terms from the NMF model.

Using lexicon-based annotations, we examined how persuasive strategies varied by target demographics (figure **??**):

- Gender: For videos targeting females and males, logos was the most significant for both; however, for content targeting females, levels of pathos was even higher. This suggests that content for females is more emotionally driven, even when clear guidelines are provided while content for males is more practically driven with lowered emotional appeal in the process.

- Race: For those videos that had identity as a central focus through the lens of race, there was a concentration of logos and ethos, implying that statistical data or policy references were used to present the concept of racial identity through a systemic level conversation. This shows that when discussing race, influencers take a more factual/credible approach to maintain integrity.

- Income Level: High income audiences were more likely to receive content with logos whereas low income audiences received

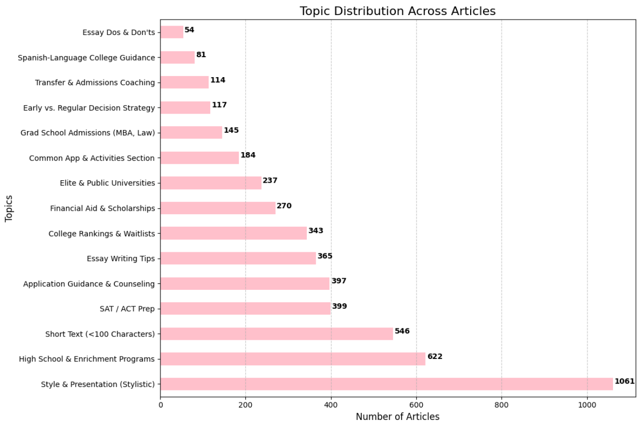

Figure 10: Topic Distribution across Articles

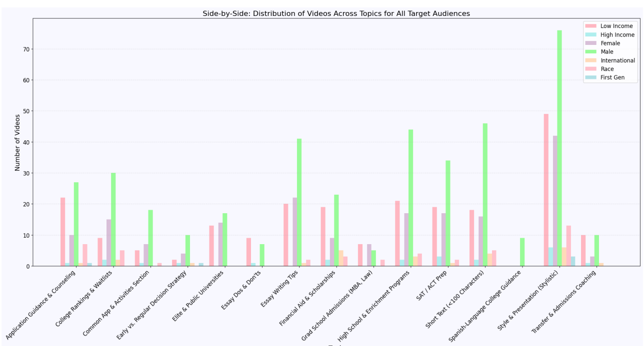

Figure 11: Distribution of Articles across Topic for All Target Audience

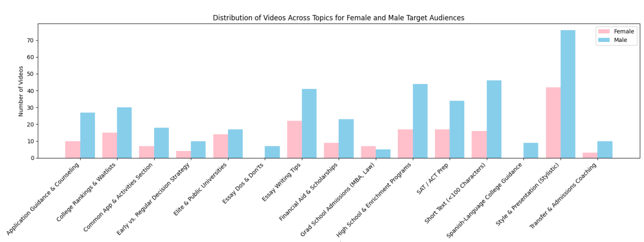

Figure 12: Male vs. Female Inter-Group Comparison

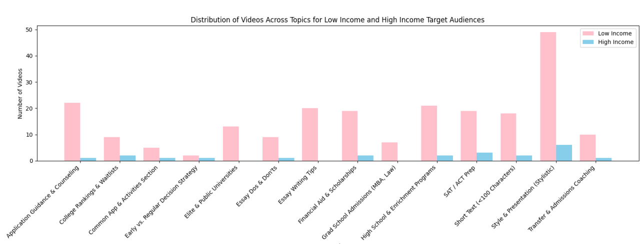

Figure 13: Low Income vs. High Income Inter-Group Comparison

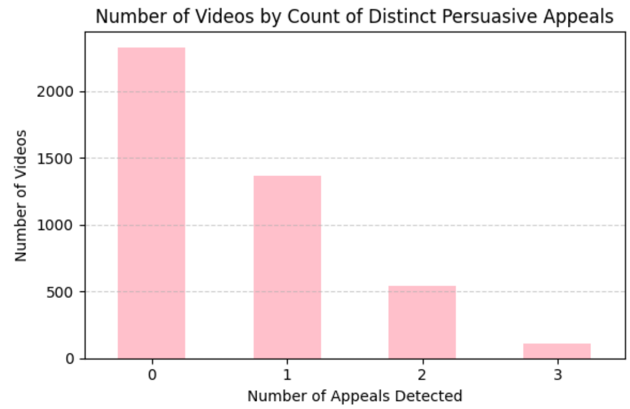

Figure 14: Number of Videos by Count of Distinct Persuasive Appeals

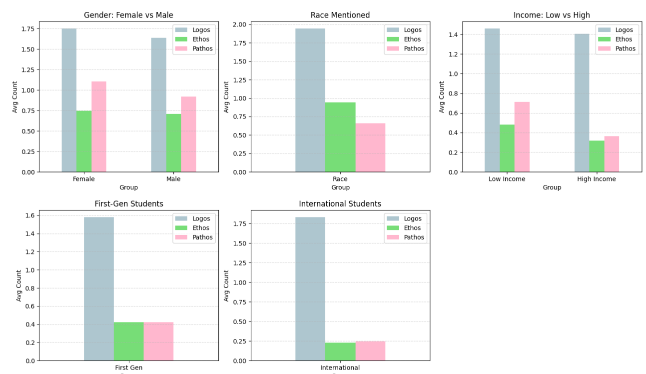

Figure 15: Appeals across Demographics

more messages geared toward pathos. This is a calculated, tailored persuasive approach based on financial identity and presumed emotional disposition. High income audiences are often expected to have more resources and know more, so factual and procedural content are more appreciative. Meanwhile, low income audiences would appreciate words of encouragement when receiving advice.

- First Generation and International Students. These two groups received content focused on logos and practicality, step-by-step informational clarity. Yet the first-gen students received more ethos and pathos to establish credibility and provide motivation as they are more unfamiliar with the admission systems.

Overall, these trends assess how these influencers not only change persuasive appeals to get across content but also recognize the specific challenges relative to their audiences' identities. The calculated use of rhetorical appeals shows that college admissions advice from TikTok across demographics are not the same but rather tailored to audience-specific intention.

5 Discussion

5.1 Demographics of Intended Targeted Audience

To further elaborate on our findings, when we examine the percentage of videos explicitly referencing a particular demographic, we observe that 12.5% of our videos mention male students, followed by 7.7% mentioning low-income students. This trend supports our hypothesis that male students are more frequently targeted than female students—despite the reality that women have been enrolling in and graduating from college at higher rates than men for over a decade. According to the National Center for Education Statistics, women made up 59% of college students in the U.S. as of 2021, a share that has steadily grown since the early 2000s [24]. Furthermore, women consistently outpace men in college completion rates [11]. Yet the content ecosystem appears skewed toward advising male students. By continuing to prioritize male-targeted content, influencers may reinforce existing gender disparities in access to educational resources where women are already underrepresented.

We initially thought the reference to high-income students would be higher based on the presumption that only high-income families can pay for college admissions consulting. Yet low-income students were referenced more suggesting a change in sentiment from the industry—a conscious decision by creators to attempt to help those students with no private consultation or school resources at least have some guidance. It also highlights the potential of social media as a more inclusive platform, where influencers focus more on students historically excluded from elite educational services.

5.2 Topics and Persuasion Techniques

The most frequent topic discussed for all demographics is admissions tips, which makes sense because the college consulting influencers want to have general content to ensure they reach as wide of a target as possible. However, when we analyze and delve deeper into topics of videos across demographics, it appears that videos discussing sciences or include stylistic advice are directed to male audiences. This supports the perception that college consulting content believes that men are more focused on particular careers like STEM fields [36]. Yet videos tailored towards females and lower socioeconomic statuses seem to focus more on emotional support and validation. This is concerning, as it may insinuate that even college consulting content may reinforce gendered expectations for education and eventually gender roles in career paths.

When it comes to persuasive appeals, 46.5% of the videos employed at least one identifiable appeal. The most significant appeal was logos (logical appeals). Yet for underrepresented demographics (female, low-income, racial minorities), these videos have a more diverse mix of appeals. More specifically, appeals to credibility (ethos) and emotions (pathos) were used extensively. Meanwhile, for overrepresented demographic appeals (male, high-income, international), videos are more factual with mere procedural details. This suggests that influencers recognize the emotional and credibility concerns that underrepresented people have who struggle more during the process and simultaneously need guidance and reassurance. Since social media is considerably affordable (or free most of the time), social media is the best platform to reach these people and supplement information not offered through standard counseling options [8].

5.3 Considerations

Our findings reveal that there are biases towards creating and promoting content that is targeted towards white and/or male students (RQ1). Additionally, our findings show that low income students are targeted significantly and students either from a non-white, low income, or non-male demographic tend to have a richer mix of persuasive techniques used when talking about them (RQ1). This is significant because it provides evidence for our hypothesis that white and/or male students were targeted more than others. These findings further the need to investigate the college consulting ecosystem as it is so influential on the college application process for prospective students. The kind of advice prospective students receive greatly shapes their college application and admissions process.

Our findings do confirm findings that are in existing literature. Specifically literature that supports the disparities amongst certain gender and racial groups. Non-male and non-white students tend to have lower acceptance rates compared to their counterparts and non-male students tend to be disadvantaged when it comes to scoring for standardized testing.

Our study does not include any harmful use of personal data, as TikTok data is public. When users signed the terms of agreement, they should've understood that their data would be public to a certain extent. Additionally, we are working with locally collected data and will not share it publicly on the web. TikTok has tens of millions of videos posted per day, so scraping a few thousand videos is okay. As far as we know, as long as we are not trying to profit off of any copyrighted material in the TikTok data. We also do not plan to deanonymize the data in our research, in hopes that this helps protect the privacy of our subjects. We tried to make sure we aren't misrepresenting individuals in our study and making unfounded assumptions/generalizations. Lastly, we understand that social media algorithms are different for each and every user, and recognize the built in biases that are connected to the algorithms.

Limitations. There are limitations to our study. When collecting our dataset, we downloaded videos from TikTok, transcribed the audios of each video, and saved all the transcripts as the subjects of analysis. However, for one video, its content may not be transmitted only through audio as the narration of the influencers or voicevers, but also through visual aids such as the video cover and words and graphics on the background of the video. While our data collection pipeline focused on the audible communication, it neglected inaudible communications that passed on as important information to the audience visually.

Second, our demographic groups are rather binary. We had male versus female as the inter-group comparisons in considering genders, and we had low-income versus high income as the inter-group comparisons in considering economic backgrounds. Nevertheless, we did not take into consideration other genders that do not fall under the binary genders as well as other economic possibilities that are located at the middle ground of low and high incomes – some students from families who would not necessarily be considered high income but are making sustainably enough to be considered not low income as well. Moreover, in line with the clear-cut economic backgrounds, we considered financial aids as an indicator of low income groups in defining the low income target audience. However, while financial aids are usually need-based that do not have anything to do with merits, financial aids can vary for different amounts, ranging from full-ride to a couple of thousands. Because colleges in the U.S. are generally very expensive, not only students from low income backgrounds, but also students from families who have sustainably good amounts of income would be interested in financial aid. While not interested in financial aids could be an indicator for high income, having interests in financial aids is not necessarily an indicator for low income. Yet our binary categorization of economic backgrounds and lack of specificities in defining levels of needs of financial aid failed to account for the ambiguity in applicants' interests for financial aid.

Third, we used probabilistic machine learning methods like topic modelling for processing data for analysis. Probabilistic methods predict results to the best of their ability, yet they are subject to a certain degree of ambiguity and do not guarantee the generation of hundred percent accurate predictions.

A step to take in our study is to refine our current pipeline to account for limitations described above. Ideally, we would want to collect more comprehensive data for the contents of each video in order to analyze the aggregated video contents of all college admission consulting related videos to the best of our ability and accuracy. We would want to collect inaudible communications into our dataset as well. Moreover, Our definitions of indicators for demographic groups currently look for apparent symbols. Having financial aids as indicators for low income is one example of such a case. We would want to conduct a further refinement of our indicators that breaks down apparent indicators for more nuanced analysis of demographic backgrounds.

6 Conclusion

The main impact we focused on is that more research needs to be done into the college consulting ecosystem. We've already found signs of biases in the content college consultant influencers produce, and we need to emphasize the importance of perhaps targeting marginalized groups of students since they are less likely to be accepted into college.

The major insight we need to focus on is that college consultant influencers have a major impact on the decision process for prospective students, and our findings suggests that influencers tend to create content targeted towards white and/or male students, with a lot of the content also being targeted towards low-income students using persuasive techniques such as ethos, logos, and pathos. With there being a heavier focus on ethos and pathos in content targeted towards non-white, non-male, and low-income students.

References

[1] M. Borges. 2023. *Social Media's Influence on Out-of-State Students During Their College Decision.* Ph.D. Dissertation. Unknown. https://www.proquest.com/dissertations-theses/social-media-s-influence-on-out-state-students/docview/2903788384/se-2 Order No. 30815835.

[2] L. Bouzoubaa, M. Guo, J. Trybala, A. Razi, and R. Rezapour. 2025. From Dr00gtiktok to harmreduction: Exploring Substance Use Hashtags on TikTok. *arXiv preprint arXiv:2501.16123* (2025). https://arxiv.org/abs/2501.16123

[3] A. Braca and P. Dondio. 2023. Persuasive communication systems: a machine learning approach to predict the effect of linguistic styles and persuasion techniques. *Journal of Systems and Information Technology* 25, 2 (2023), 160–191. doi:10.1108/JSIT-07-2022-0166

[4] M. Brown, C. Pyle, and N. B. Ellison. 2022. "On my head about it": College aspirations, social media participation, and community cultural wealth. *Social Media + Society* 8, 2 (2022). doi:10.1177/20563051221091545

[5] College Board. [n. d.]. College Admission Glossary: Learn the Lingo. BigFuture. https://bigfuture.collegeboard.org/plan-for-college/apply-to-college/college-admission-glossary-learn-the-lingo Accessed 9 Apr. 2025.

[6] College Board Research. 2022. How Did Pandemic Disruptions Impact Applications, Enrollment, and Outcomes at Selective Institutions? All Access | College Board. https://allaccess.collegeboard.org/how-did-pandemic-disruptions-impact-applications-enrollment-and-outcomes-selective-institutions

[7] J. de Raadt. 2022. What is a college admissions counselor, and what do they do? Empowerly. https://empowerly.com/applications/what-is-a-college-admissions-counselor-and-what-do-they-do/

[8] EAB. 2020. Recruiting the Digital Native: Actionable Insights from Our 2019 Student Communication Preferences Survey. https://eab.com.

[9] S. M. Edelson, V. F. Reyna, A. Singh, and J. E. Roue. 2024. The psychology of misinformation across the lifespan. *Annual Review of Developmental Psychology* (2024). doi:10.1146/annurev-devpsych-010923-093547

[10] C. L. E. Foster. 2023. Truth as social practice in a digital era: iteration as persuasion. *AI & Society* 38, 5 (2023), 2009. doi:10.1007/s00146-021-01306-w

[11] Kiley Hurst. 2024. U.S. Women Are Outpacing Men in College Completion, Including in Every Major Racial and Ethnic Group. *Pew Research Center* (Nov. 2024). https://www.pewresearch.org/short-reads/2024/11/18/us-women-are-outpacing-men-in-college-completion-including-in-every-major-racial-and-ethnic-group

[12] U. M. Jayakumar and S. E. Page. 2021. Cultural Capital and Opportunities for Exceptionalism: Bias in University Admissions. *Journal of Higher Education* 92, 7 (2021), 1109–1139. doi:10.1080/00221546.2021.1912554

[13] J. Johansson. 2024. *Automatic Word Decoding Assessment Using Whisper and Machine Learning Techniques.* Master's thesis. KTH Royal Institute of Technology. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-359755

[14] Khan Academy. 2025. Master Timeline: College Admission. Khan Academy. https://www.khanacademy.org/college-careers-more/college-admissions/get-started/introduction-ca/a/master-timeline-college-admissions Accessed 4 May 2025.

[15] J. Koetsier. 2021. Here are the 10 most downloaded apps of 2020. Forbes. https://www.forbes.com/sites/johnkoetsier/2021/01/07/here-are-the-10-most-downloaded-apps-of-2020/

[16] J. Koetsier. 2023. 10 most downloaded apps of 2022: Facebook down, Spotify up, TikTok stable, CapCut keeps growing. Forbes. https://www.forbes.com/sites/johnkoetsier/2023/01/04/top-10-most-downloaded-apps-of-2022-facebook-down-spotify-up-tiktok-stable-capcut-keeps-growing/

[17] J. Koetsier. 2024. Top 10 most downloaded apps and games of 2021: TikTok, Telegram big winners. Forbes. https://www.forbes.com/sites/johnkoetsier/2021/12/27/top-10-most-downloaded-apps-and-games-of-2021-tiktok-telegram-big-winners/

[18] K. Lajnef. 2023. The effect of social media influencers' on teenagers Behavior: an empirical study using cognitive map technique. *Current Psychology* 42, 22 (2023), 19364. doi:10.1007/s12144-023-04273-1

[19] A. D. Leitao Leonelli. 2023. *What's the vibe: Exploring the impact of peer influence through social media on college-going decisions.* Ph. D. Dissertation. Northeastern University. Publication No. 30638829.

[20] S. Medley. 2016. "Gender balancing" as sex discrimination in college admissions. *Harvard Civil Rights-Civil Liberties Law Review* 51 (2016), 537.

[21] H. A. Mohamad. 2022. Analysis of rhetorical appeals to logos, ethos and pathos in ENL and ESL research abstracts. *Malaysian Journal of Social Sciences and Humanities (MJSSH)* 7, 3 (2022), e001314–e001314.

[22] B. Nankervis. 2011. Gender Inequities in University Admission Due to the Differential Validity of the SAT. *Journal of College Admission* 213 (2011), 24–30.

[23] B. Nankervis. 2013. Gender Inequity in the National Merit Scholarship Program. *Journal of College Admission* 219 (2013), 20–25.

[24] National Center for Education Statistics, United States. Office of Educational Research and Improvement, Center for Education Statistics, and Institute of Education Sciences (US). 2009. *The Condition of Education.* U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, Washington, D.C.

[25] G. Odaga. 2022. Affirmative Action and Women in Uganda's Public University Education. *Higher Education Policy* 35, 1 (2022), 1–18. doi:10.1057/s41307-020-00189-8

[26] Olumuyiwa A. Oludayo, Segun I. Popoola, Comfort O. Akanbi, and Aderemi A. Atayero. 2019. Gender disparity in admissions into tertiary institutions: Empirical evidence from Nigerian data (2010–2015). *Data in Brief* 22 (2019), 920–933. doi:10.1016/j.dib.2019.01.031

[27] Y. P. Pandeya. 2023. Social media's influence on students' college selection processes. *Shanti Journal: A Multidisciplinary Peer-Reviewed Journal* 3, 2 (2023), 142–159. https://nepjol.info/index.php/shantij/article/view/60867

[28] A. Pathak, N. Madani, and K. Joseph. 2021. A method to analyze multiple social identities in twitter bios. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–35.

[29] J. Ramos. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning*, Vol. 242. 29–48.

[30] P. C. Rothschild. 2014. Authors influencing others to follow: An analysis of a social media platform through the framework of persuasion theory. *MEIEA Journal: Journal of the Music & Entertainment Industry Educators Association* 14, 1 (2014), 251–278.

[31] Shopify. 2024. Tiktok Bio Tips: Define Your Brand and Engage New Followers. Shopify. https://www.shopify.com/blog/tiktok-bio-ideas

[32] P. Sisodia and S. W. A. Rizvi. 2023. Optical Character Recognition Development Using Python. *Journal of Informatics Electrical and Electronics Engineering* 4, 3 (2023), 1–13. doi:10.54060/jieee.2023.75

[33] A. H. Snyman. 1992. Modes of Persuasion in Galatians 6:7-10. *Neotestamentica* 26, 2 (1992), 475–484.

[34] G. Tsoumakas and I. Katakis. 2008. Multi-label classification: An overview. In *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications.* 64–74.

[35] M. L. Turner. 2017. Like, love, delete: Social media's influence on college choice. *Journal of College Admission* 237 (2017), 31–33.

[36] Ming-Te Wang and Jennifer L. Degol. 2017. Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. *Educational Psychology Review* 29 (2017), 119–140. doi:10.1007/s10648-015-9355-x

[37] K. L. Williamson. 2024. *Understanding the role a mentor plays in female students' college choice: A qualitative study.* Ph. D. Dissertation. Troy University. Publication No. 31762408.

[38] D. Y. Wohn, N. B. Ellison, M. L. Khan, R. Fewins-Bliss, and R. Gray. 2013. The role of social media in shaping first-generation high school students' college aspirations: A social capital lens. *Computers & Education* 63 (2013), 424–436. doi:10.1016/j.compedu.2013.01.004

[39] J. Xu, T. C. Lu, R. Compton, and D. Allen. 2014. Quantifying cross-platform engagement through large-scale user alignment. In *Proceedings of the 2014 ACM conference on Web science.* 281–282.

[40] W. Yin, J. Hay, and D. Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *arXiv preprint arXiv:1909.00161* (2019). https://arxiv.org/abs/1909.00161

[41] A. Zenker. 2023. Women in STEM USA statistics. STEM Women. https://www.stemwomen.com/women-in-stem-usa-statistics