International Journal of Computer Sciences and Mathematics Engineering Vol. 4 No. 1 (2025)
Published by: CV. HAWARI

International Journal of Computer Sciences and Mathematics Engineering

Analysis of Trends and Profitability Growth Opportunities of the VTuber Industry Using Exploratory Data Analysis (EDA) Methodology

Ananda Aulia[1], Muhammad Iqbal[2]

Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: anandaaulia30@gmail.com, muhammadiqbal@dosen.pancabudi.ac.id
Email Penulis Korespondensi: anandaaulia30@gmail.com

Article Info

Article history:
Received Feb 28, 2025
Revised Mar 07, 2025
Accepted Mar 17, 2025

Keywords:
Data Mining
VTuber
Super Chat
EDA

ABSTRACT

The VTuber industry has experienced significant growth in recent years, with VTubers as virtual influencers engaging with global audiences through live streaming and content creation. However, the financial performance of this industry remains underexplored. By applying EDA, this research investigates the key factors influencing the profitability of VTubers, including audience engagement, monetization strategies, and content diversity. The analysis is conducted using data collected from various VTuber channels, including viewer statistics, revenue streams (such as Super Chats, merchandise, and sponsorships), and social media metrics.

The EDA methodology proves to be effective in providing valuable insights into patterns and relationships between variables affecting profitability in this industry. Through data visualization and descriptive statistical analysis, EDA identifies factors contributing to VTuber performance, such as audience interaction levels and monetization success. The findings demonstrate that EDA offers a clear overview of trends and growth opportunities within the VTuber industry, helping content creators, investors, and industry stakeholders make more strategic, data-driven decisions.

Corresponding Author:
Ananda Aulia,
Faculty of Computer Science and Information Technology,
Universitas Pembangunan Panca Budi,
Jend. Gatot Subroto Road, 4.5 KM Seikambing, 20122, Medan City, Sumatera Utara, Indonesia,
Email: anandaaulia30@gmail.com

I. INTRODUCTION

Rapid advancements in technology have provided numerous benefits across various fields. The utilization of technology to assist in completing tasks has become essential in everyday life[1].VTubers emerged as an internet trend that gained popularity in 2016. A VTuber, or virtual YouTuber, entertains audiences using a

digital avatar to conceal their real-world appearance. Through computer graphics and motion capture technology, VTuber avatars can interact with viewers [2]. The peak of the virtual YouTuber era was in 2020, when there were over 10,000 VTubers, with many receiving exclusive airtime on television stations in Japan[3].

According to YouTube's revenue rankings based on Super Chats in 2021, 8 out of the 10 highest-earning video creators worldwide were VTubers. These VTubers were affiliated with different agencies, but two major agencies—Hololive and Nijisanji—played a dominant role in the VTuber industry [4]. As the VTuber industry continues to grow, a vast amount of data related to VTuber performance—such as viewership numbers, audience engagement levels, and financial data—has become available, creating opportunities for more systematic data analysis methodologies. One such approach is Exploratory Data Analysis (EDA), which is implemented as a solution for examining data by exploring various aspects of datasets, identifying key features for modeling, and providing recommendations for further stages in a data science cycle [5].

A study conducted by [6] found that after performing EDA, improvements in achieving Key Performance Indicator (KPI) 3 of the Faculty of Engineering at UNG could be made by focusing on the development of new lecturers and those who had not yet participated in external activities. The data showed that the program with the highest contribution to lecturer participation accounted for 83%, while the lowest contributed around 18%. These findings enable policymakers to make informed decisions to enhance the achievement of KPI 3 in the Faculty of Engineering at UNG.

In research conducted by [7], sales data evaluation using RMSE, MAE, the XGBoost algorithm, and Exploratory Data Analysis (EDA) yielded predictions with an error margin of 1.3%, compared to ARIMA at 1.6%, outperforming Linear Regression, Random Forest, and LSTM, which had error rates of 1.81%, 1.97%, and 2.21%, respectively, when compared to actual data.

Another study by [8], found that mobility levels in Jakarta significantly decreased during the implementation of Large-Scale Social Restrictions (PSBB). EDA analysis demonstrated that this decline in mobility contributed to a reduction in the daily number of COVID-19 cases in the region. The decrease in mobility was seen as a crucial step in controlling the spread of COVID-19 in Jakarta.

II. RESEARCH METHOD

1) Research Stages

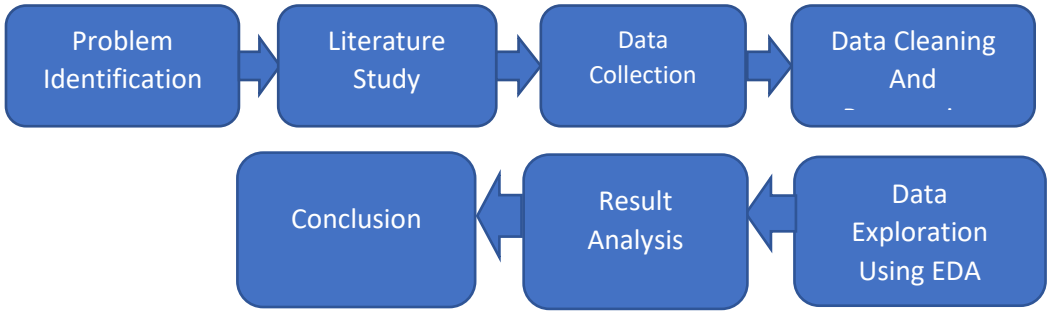

Figure 1. Research Stages

a. Problem Identification

i. Analyzing the profitability of the VTuber industry using the Exploratory Data Analysis (EDA) methodology..
ii. Formulating research questions to address factors affecting VTuber revenue and how audience engagement and monetization strategies contribute to profitability.

b. Literature Review

i. Conducting a literature review to understand the development of the VTuber industry, including the Super Chat phenomenon and the role of VTubers in the digital entertainment world.
ii. Analyzing previous studies related to the VTuber industry, monetization techniques, and relevant data analysis methodologies.

c. Data Collection

i. Data collection is conducted by retrieving datasets from Kaggle.com
ii. The required data for analysis includes viewer statistics, Super Chat revenue, subscriber count, and video uploads.

d. Data Cleaning and Preparation

i. Cleaning the data to ensure that the dataset used is free from missing, duplicate, or irrelevant data.
ii. Transforming the data into an appropriate format for further analysis, such as normalization or data aggregation, to clarify existing patterns.

e. Data Exploration Using EDA

i. Conducting exploratory analysis using data visualization techniques (such as histograms, boxplots, and scatterplots) to understand distribution and relationships between variables.
ii. Analyzing correlations between key variables, such as audience engagement (e.g., view count) and revenue generation.
iii. Identifying patterns or trends that explain the factors influencing VTuber profitability.

f. Results Analisys

i. Analyzing EDA findings to identify key factors affecting VTuber profitability, such as content type, streaming frequency, and diversity of revenue sources.
ii. Summarizing findings that provide insights into how VTubers can maximize their revenue based on patterns found in the data.

g. Conclusion

i. Concluding the analysis results and providing strategic recommendations for VTubers to optimize profitability.
ii. Offering recommendations to content creators, investors, and industry stakeholders regarding factors that may influence their decisions within the industry.

2) VTuber

A VTuber is an online talent entertainer created using a virtual avatar in either 2D (two-dimensional) or 3D (three-dimensional) form [9]. VTubers have become part of Japanese pop culture, attracting audiences due to the positive impacts they bring [10]. The VTuber trend began to rise in 2016 in Japan and has since gained global popularity, with more people enjoying VTuber content and even aspiring to become VTubers themselves [11]. VTubers often use iconic anime-style characters to represent them, primarily attracting viewers from the "Weaboo" or "Otaku" community [12]. The peak of the virtual YouTuber era was in 2020, with over 10,000 VTubers, many of whom received exclusive airtime on television stations in Japan [3].

3) Super Chat

Viewers can interact with streamers through YouTube's Super Chat donation system. Super Chat allows viewers to financially support their favorite content creators [13]. It enables viewers to highlight their messages in the live chat, making them more visible, while also donating a certain amount [14]. Through Super Chat in livestreaming, viewers can engage with each other and directly interact with the streamer. This interactive experience creates a more immersive atmosphere than regular videos, allowing viewers to feel more connected to the streamer [15].

4) Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the process of analyzing and visualizing data to gain a deeper understanding of its insights [16]. EDA is also a technique used in data understanding to explore which data has the quality necessary for the modeling stage [5]. The EDA process involves identifying problems, collecting data, cleaning data, and then conducting exploratory data analysis to derive useful insights [6].

III. RESULTS AND DISCUSSION

Various steps are involved in conducting EDA. The following are common steps taken in performing EDA analysis:

a) Maximizing insights into the dataset,
b) Uncovering the data structure,
c) Extracting important variables,
d) Detecting outliers and anomalies,
e) Performing assumption tests,
f) Developing models, and
g) Identifying optimal factors.

1) The first step involves importing the necessary libraries for data processing, as shown in Figure 2.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.ticker as tick
import seaborn as sns
from dask import dataframe as dd
```

Figure 2. Importing Required Libraries

2) The second step is loading the acquired dataset, followed by testing and analysis. This process is illustrated in Figure 3.

```
from google.colab import drive
drive.mount("/content/channels.csv")
drive.mount("/content/chat.csv")
drive.mount("/content/superchat.csv")
drive.mount("/content/chatLegacy.csv")
drive.mount("/content/markedAsDeleted.csv")
drive.mount("/content/markedAsBanned.csv")
```

Figure 3. Loading Dataset

3) The third step is reading the dataset that was previously loaded, as shown in Figure 4.

```
df_channels = pd.read_csv("/content/channels.csv")
df_superchat = pd.read_csv("/content/superchat.csv")
```

Figure 4. Reading Dataset

4) In the fourth step, the results of the "Top 5 Subscriber Count in VTuber Agencies" analysis reveal a significant difference in subscriber numbers between Hololive and Nijisanji, as shown in Figure 5.

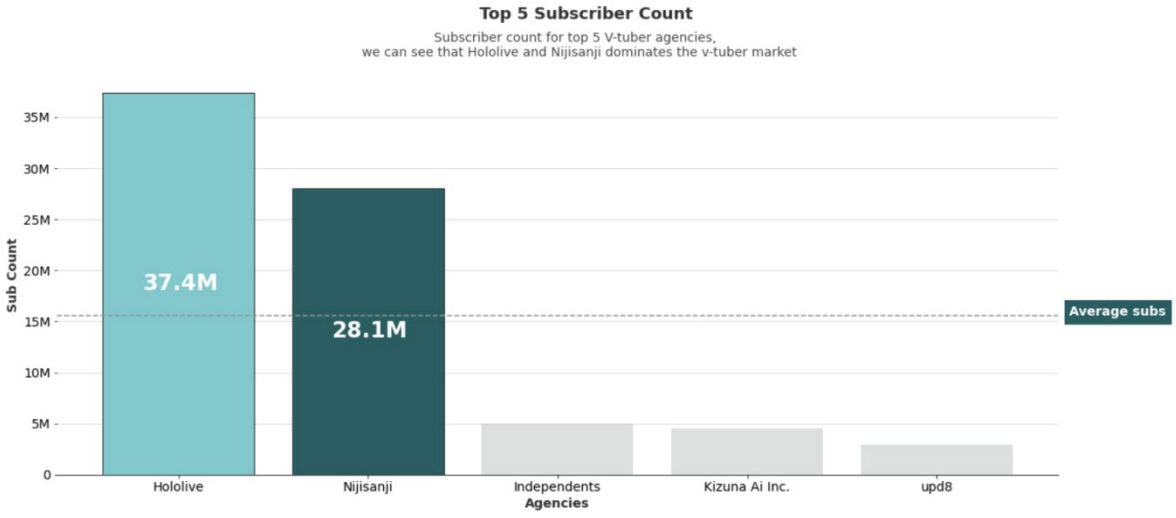

Figure 5. Top 5 Subscriber Count Results

5) In the fifth step, the "Top 5 Video Uploaded Count" results indicate that Nijisanji dominates in video uploads, as shown in Figure 6.

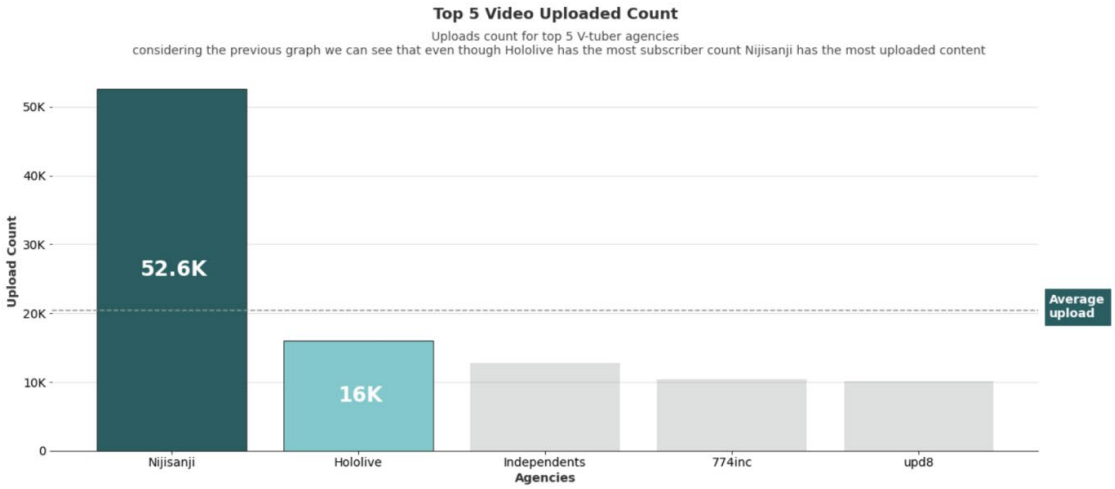

Figure 6. Top 5 Video Uploaded Count Results

6) In the sixth step, the "Top 20 VTubers on Hololive and Nijisanji" analysis reveals the VTubers with the highest upload count and subscriber numbers, as shown in Figure 7.

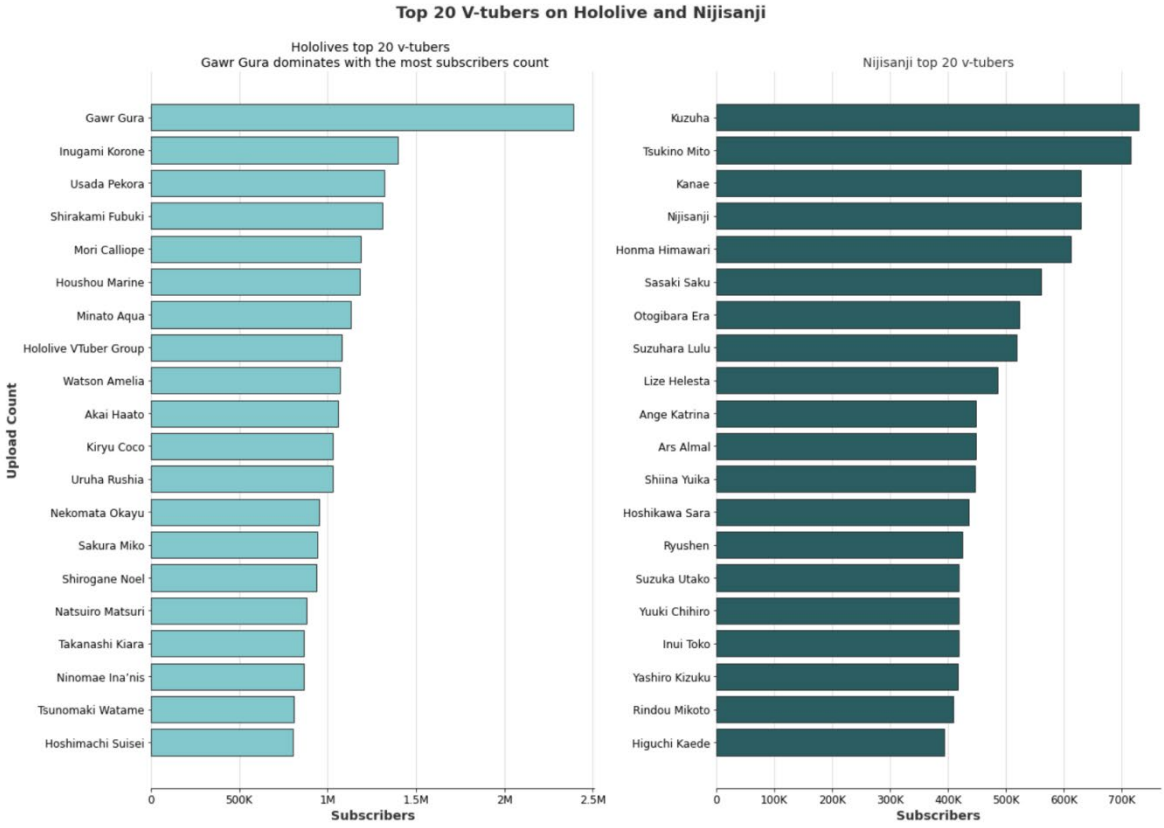

Figure 7. Top 20 VTubers on Hololive and Nijisanji Results

7) In the seventh step, the analysis of viewer Super Chat patterns reveals that Super Chats are most frequently sent on Tuesdays and Saturdays, while they appear less often on Fridays. This is shown in Figure 8.

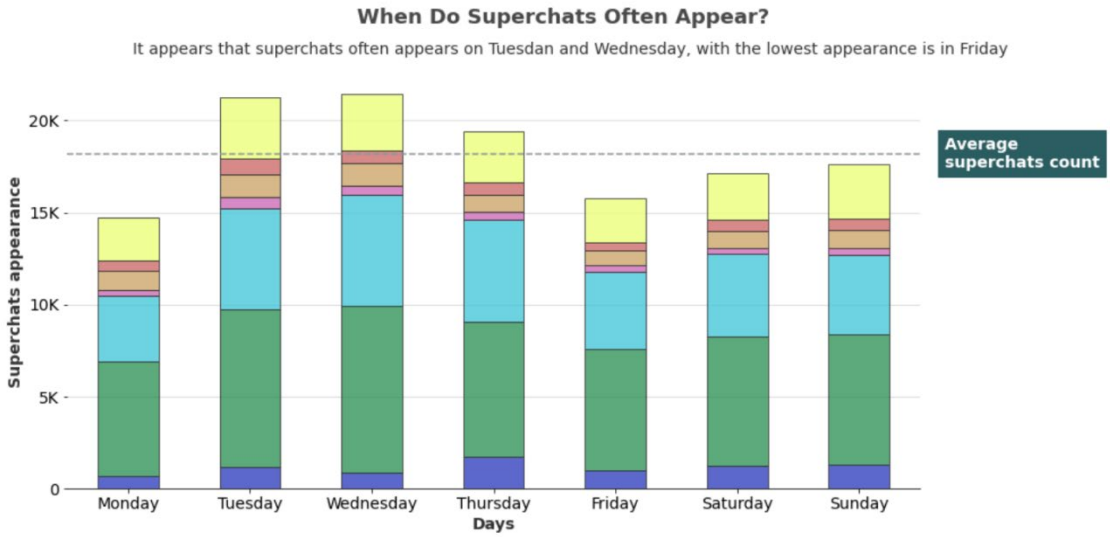

Figure 8. Super Chat Activity Trends

8) In the eighth step, the analysis of Super Chat revenue across VTuber agencies shows that Hololive and Nijisanji generate the highest earnings, as illustrated in Figure 9.

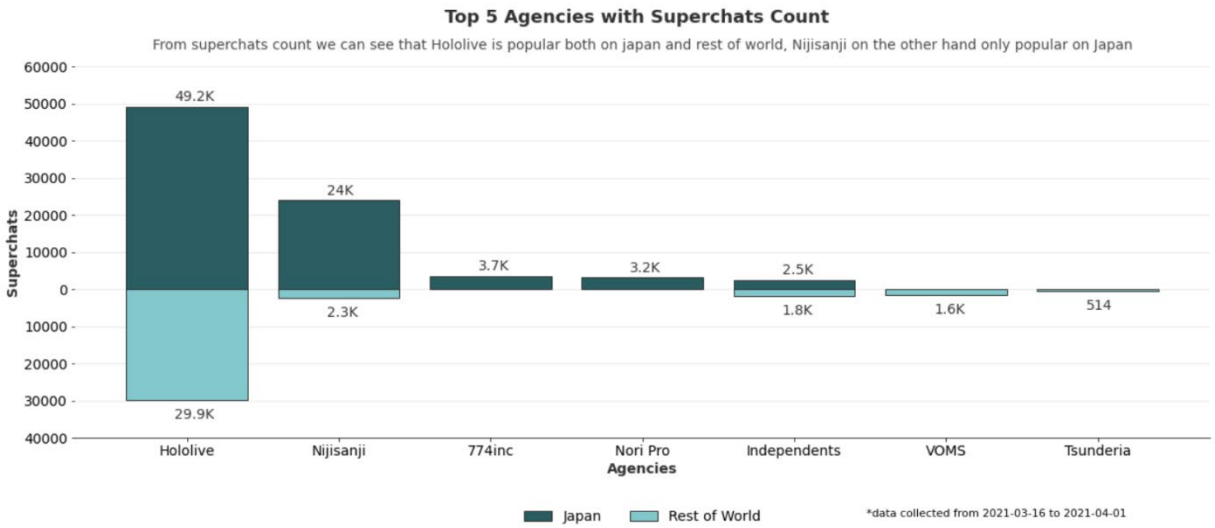

Figure 9. Top 5 Agencies with Super Chat Count

9) In the ninth step, Super Chat revenue from the two largest VTuber agencies, Hololive and Nijisanji, is analyzed. The revenue is converted into Japanese Yen, with Hololive earning 44 million yen and Nijisanji earning 31.7 million yen, as shown in Figure 10.

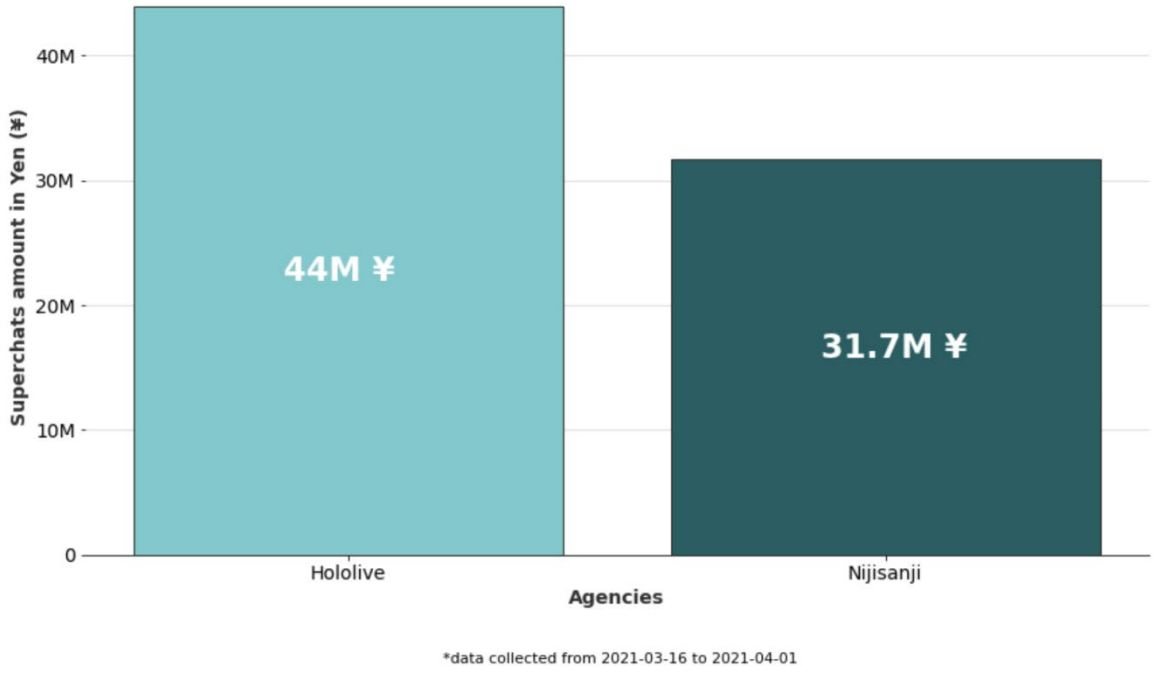

Figure 10. Super Chat Revenue for Hololive and Nijisanji

IV. CONCLUSION

Based on the graphical analysis conducted using Google Colab, this study successfully identified patterns related to subscriber growth, video uploads, Super Chat activity, agency earnings, and total Super Chat revenue. According to Figure 5, there is a significant gap in subscriber numbers between Hololive (37.4 million subscribers) and Nijisanji (28.1 million subscribers). Figure 6 shows that Nijisanji has uploaded 52.6K videos, significantly more than Hololive, which has uploaded only 16K videos. Figure 8 indicates that Super Chats are most frequently sent on Tuesdays and Saturdays but appear less often on Fridays each week. Figure 9 confirms that Hololive and Nijisanji are the top-earning agencies in terms of Super Chat revenue.

Lastly, Figure 10 highlights the difference in Super Chat earnings between Hololive (44 million yen) and Nijisanji (31.7 million yen). Several key factors influence the number of subscribers and Super Chats, including video upload consistency, livestream timing, and content quality presented to viewers.

V. REFERENCES

[1] F. Khatami, "PENGEMBANGAN MEDIA VTUBER DAN IMPLEMENTASI DALAM PEMBELAJARAN BAHASA Indonesia," 2023.

[2] M. Reyus Liman Prodi Perhotelan and S. Yapari Bandung, "Pengaruh Hololive Indonesia dalam Memperkenalkan Bahasa Indonesia Kepada Penonton Internasional," vol. 1, no. 2, 2023.

[3] R. A. Putra, "Analisa Media Siber pada Tayangan Langsung Virtual Youtube Mythia Batford," Jurnal Bisnis dan Komunikasi Digital, vol. 1, no. 1, p. 13, Nov. 2023, doi: 10.47134/jbkd.v1i1.1914.

[4] S. Bermúdez Badia, I. Kuramoto, and H. Sakuma, "YouTubers vs. VTubers: Persuasiveness of human and virtual presenters in promotional videos," Mar. 2023.

[5] I. Griha Tofik Isa, Z. Zulkarnaini, L. Novianti, F. Elfaladonna, and S. Agustri, "Exploratory Data Analysis (EDA) dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ Palembang," Smart Comp: Jurnalnya Orang Pintar Komputer, vol. 12, no. 3, 2023, doi: 10.30591/smartcomp.v12i3.4125.

[6] I. Hidayat, A. I. Tolago, R. D. R. Dako, and J. Ilham, "Analisis Data Eksploratif Capaian Indikator Kinerja Utama 3 Fakultas Teknik," Jambura Journal of Electrical and Electronics Engineering, vol. 5, no. 2, 2023, doi: 10.37905/jjeee.v5i2.18397.

[7] A. M. Husein, F. R. Lubis, and M. K. Harahap, "Analisis Prediktif untuk Keputusan Bisnis : Peramalan Penjualan," Data Sciences Indonesia (DSI), vol. 1, no. 1, 2021, doi: 10.47709/dsi.v1i1.1196.

[8] N. Khairu Nissa, Y. Nugraha, C. F. Finola, A. Ernesto, J. I. Kanggrawan, and A. L. Suherman, "Evaluasi Berbasis Data: Kebijakan Pembatasan Mobilitas Publik dalam Mitigasi Persebaran COVID-19 di Jakarta," Jurnal Sistem Cerdas, vol. 3, no. 2, 2020, doi: 10.37396/jsc.v3i2.77.

[9] D. I. S. Saputra and I. Setyawan, "Virtual YouTuber (VTuber) sebagai Konten Media Pembelajaran Online," Prosiding SISFOTEK, vol. 5, no. 1, 2021.

[10] J. I. Adzania and A. Arianingsih, "DAMPAK MENONTON VTUBER TERHADAP PENGGEMAR BUDAYA POPULER JEPANG DI BANDUNG," MAHADAYA: Jurnal Bahasa, Sastra, dan Budaya, vol. 2, no. 2, 2022, doi: 10.34010/mhd.v2i2.8265.

[11] Akdeas Oktanae Widodo, F. Septiadi, and Nur Aini Rakhmawati, "ANALISIS TREN KONTEN PADA VTUBER INDONESIA MENGGUNAKAN LATENT DIRICHLET ALLOCATION," Jurnal Informatika dan Rekayasa Elektronik, vol. 6, no. 1, 2023, doi: 10.36595/jire.v6i1.718.

[12] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," Jurnal Teknik Komputer AMIK BSI, vol. 8, no. 2, 2022, doi: 10.31294/jtk.v4i2.

[13] W. Mohammad, H. Saifurrahman, J. Rausyan Fiqrussalam, A. Umar Sulthoni, and D. Fansyiari Austi, "Triwikrama: Jurnal Multidisiplin Ilmu Sosial PENGARUH JUMLAH SUBSCRIBERS DAN JUMLAH PENAYANGAN TERHADAP PENDAPATAN SUPER CHAT PADA VIRTUAL YOUTUBER KOBO KANAERU," vol. 01, pp. 30–45, 2023.

[14] M. Rasyid Nafi' et al., "BUDAYA DONASI ONLINE DI YOUTUBE MELALUI FITUR SUPERCHAT," 2023.

[15] F. Z. Rabbani and R. Wati, "KOMUNITAS DAN INTERAKSI PENONTON PADA SIARAN LANGSUNG DI KANAL YOUTUBE NAPLIVE," Jurnal Sosiohumaniora Nusantara, vol. 1, no. 1, pp. 406–415, 2024.

[16] M. Radhi, D. Ryan Hamonangan Sitompul, S. Hamonangan Sinurat, and E. Indra, "ANALISIS BIG DATA DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA) DAN METODE VISUALISASI MENGGUNAKAN JUPYTER NOTEBOOK," Jurnal Sistem Informasi dan Ilmu Komputer Prima, vol. 4, no. 2, 2021.