

Facultat Internacional de Comerç i Economia Digital La Salle

Trabajo Final de Máster

Máster Universitario en Ciencia de los Datos / Data Science

TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA ESTABLECER LA RELACIÓN ENTRE INFLUENCERS, SUS AUDIENCIAS Y LOS VALORES PERCIBIDOS

Alumno

Xiang Liu

Profesor/a Ponente

Xavier Vilasís Cardona

Sue Aran Ramspott

ACTA DEL EXAMEN DEL TRABAJO FINAL DE MÁSTER

Reunido el Tribunal calificador en el día de la fecha, el alumno

Xiang Liu

Expuso su Trabajo de Final de Máster, el cual trató sobre el tema siguiente:

Técnicas de inteligencia artificial para establecer la relación entre influencers, sus audiencias y los valores percibidos

Acabada la exposición y contestadas por parte del alumno las objeciones formuladas por los miembros del tribunal, este valoró el mencionado Trabajo con la calificación de

Barcelona,

VOCAL DEL TRIBUNAL

VOCAL DEL TRIBUNAL

PRESIDENTE DEL TRIBUNAL

ÍNDICE DE CONTENIDO

RESUMEN.....	- 1 -
ABSTRACT	- 1 -
AGRADECIMIENTO	- 2 -
INTRODUCCIÓN.....	- 3 -
MOTIVACIÓN Y OBJETIVOS.....	- 3 -
MOTIVACIÓN PERSONAL	- 3 -
OBJETIVOS DEL PROYECTO.....	- 4 -
MARCO TEÓRICO.....	- 4 -
UNA MIRADA HACIA REDES SOCIALES.....	- 4 -
<i>Conceptualización y Trayectoria Histórica de las RRSS</i>	- 4 -
<i>Evolución Histórica</i>	- 4 -
<i>Características Intrínsecas de las RRSS</i>	- 5 -
<i>Panorama de las RRSS en España: Datos y Plataformas Predominantes</i>	- 5 -
<i>Repercusiones Socioculturales de las RRSS en el Contexto Español</i>	- 6 -
LA VIDA MEDIÁTICA DEL ADOLESCENTE EN ESPAÑA.....	- 6 -
LA EMERGENCIA Y CONSOLIDACIÓN DE LA PROFESIÓN "INFLUENCER"	- 7 -
<i>Origen y Nacimiento del Influencer</i>	- 7 -
<i>Clasificación de Influencers</i>	- 8 -
<i>Relación con la Audiencia y Construcción de Confianza: La Interacción Parasocial</i>	- 8 -
<i>Impacto Sociocultural y Económico</i>	- 8 -
<i>El Influencer en el Contexto Español</i>	- 9 -
EL PANORAMA DE YOUTUBE Y SUS PROTAGONISTAS: LOS YOUTUBERS	- 9 -
BREVE INTRODUCCIÓN DE INTELIGENCIA ARTIFICIAL	- 10 -
<i>El Procesamiento de Lenguaje Natural</i>	- 10 -
<i>Corpus</i>	- 11 -
<i>Tokens</i>	- 12 -
<i>TF-IDF</i>	- 12 -
<i>N-grams</i>	- 13 -
<i>Lematización y Stemming</i>	- 14 -
<i>Topic Modeling</i>	- 14 -
<i>NER</i>	- 15 -
SOCIAL NETWORK ANALYSIS	- 16 -
<i>Componentes Principales</i>	- 17 -
<i>Métricas Clave</i>	- 17 -
DISEÑO DEL PROYECTO	- 18 -
DESCRIPCIÓN Y EXPLICACIÓN DE LOS YOUTUBERS ESCOGIDOS	- 18 -
RECOPILACIÓN DE DATOS DESDE YOUTUBE API	- 19 -
<i>Configuración Inicial</i>	- 20 -
<i>Estructura del Código y Funciones</i>	- 20 -
<i>Consideraciones Adicionales</i>	- 26 -
ANÁLISIS EXPLORATORIO DE DATOS	- 27 -
<i>Librerías necesarias para el EDA</i>	- 27 -
<i>Descripción de variables y creación de nuevas variables</i>	- 28 -
<i>Visualización</i>	- 28 -
PREPROCESAMIENTO DE TEXTOS	- 29 -
<i>Librerías extras necesarias</i>	- 29 -
<i>Estructura de preprocesamiento</i>	- 29 -
MINERÍA DE TEXTOS	- 32 -
<i>Librerías Extras</i>	- 32 -
<i>Frecuencia de palabras</i>	- 33 -

<i>Nube de palabras</i>	- 34 -
<i>N-grams</i>	- 35 -
<i>Topic Modeling</i>	- 36 -
<i>Identificación términos únicos mediante TF-IDF</i>	- 36 -
<i>NER</i>	- 37 -
<i>Revisión y Mapping de NER</i>	- 38 -
SOCIAL NETWORK ANALYSIS	- 39 -
<i>Librerías Extras</i>	- 39 -
<i>Estructuras de Código</i>	- 40 -
ANÁLISIS DE LOS RESULTADOS.....	- 40 -
IBAI	- 40 -
<i>EDA</i>	- 41 -
<i>Minería de Textos</i>	- 49 -
<i>Netwrok</i>	- 55 -
AURONPLAY	- 56 -
<i>EDA</i>	- 56 -
<i>Minería de Textos</i>	- 62 -
<i>Network</i>	- 68 -
MARTA DÍAZ	- 69 -
<i>EDA</i>	- 69 -
<i>Minería de Textos</i>	- 74 -
<i>Network</i>	- 79 -
DJMARIO.....	- 81 -
<i>EDA</i>	- 81 -
<i>Minería de Textos</i>	- 85 -
<i>Network</i>	- 91 -
CLAKOVI	- 92 -
<i>EDA</i>	- 92 -
<i>Minería de Textos</i>	- 97 -
<i>Network</i>	- 101 -
ELRUBIUS.....	- 101 -
<i>EDA</i>	- 101 -
<i>Minería de Textos</i>	- 107 -
<i>Network</i>	- 111 -
EMMA CHAMBERLAIN	- 112 -
<i>EDA</i>	- 112 -
<i>Minería de Textos</i>	- 116 -
<i>Network</i>	- 122 -
MALBERT	- 122 -
<i>EDA</i>	- 123 -
<i>Minería de Textos</i>	- 127 -
<i>Network</i>	- 132 -
MOSTOPAPI	- 133 -
<i>EDA</i>	- 133 -
<i>Minería de Textos</i>	- 138 -
<i>Network</i>	- 143 -
ORSLOK.....	- 144 -
<i>EDA</i>	- 144 -
<i>Minería de Textos</i>	- 148 -
<i>Network</i>	- 154 -
VEGETTA	- 154 -
<i>EDA</i>	- 154 -
<i>Minería de Textos</i>	- 160 -
<i>Network</i>	- 163 -
CONCLUSIÓN DE LOS RESULTADOS	- 164 -
NETWORK DE LA LISTA PREDEFINIDA DE INFLUENCERS	- 165 -

CONSIDERACIONES ÉTICAS Y SOCIALES	- 166 -
CONCLUSIÓN	- 166 -
COSTE ECONÓMICO	- 167 -
LIMITACIONES Y LÍNEAS DE FUTURO.....	- 168 -
BIBLIOGRAFÍA.....	- 170 -
ANEXO.....	- 173 -

ÍNDICE DE GRÁFICOS

Captura de código 1 Librerías recopilación de datos	- 20 -
Captura de código 2 get channel id by name.....	- 20 -
Captura de código 3 get channel videos	- 21 -
Captura de código 4 get video details.....	- 23 -
Captura de código 5 Proceso ETL.....	- 24 -
Captura de código 6 Main Script Recopilación de datos	- 25 -
Captura de código 7 Creación de nuevas variables de fecha.....	- 28 -
Captura de código 8 Creación de variable 'engagement_rate'	- 28 -
Captura de código 9 Pre-Limpieza de textos.....	- 30 -
Captura de código 10 Función preprocess.....	- 31 -
Captura de código 11 Frecuencia de palabras	- 34 -
Captura de código 12 Nube de palabras.....	- 35 -
Captura de código 13 N-grams.....	- 35 -
Captura de código 14 Topic modeling.....	- 36 -
Captura de código 15 TF-IDF	- 37 -
Captura de código 16 NER.....	- 38 -
Captura de código 17 Mapping NER	- 39 -

Figura 1 Estadística Ibai	- 41 -
Figura 2 Boxplot Ibai	- 42 -
Figura 3 Histograma Ibai	- 43 -
Figura 4 Sactterplot Ibai	- 44 -
Figura 5 Matriz de correlación Ibai	- 45 -
Figura 6 Serie Temporal mes Ibai	- 46 -
Figura 7 Serie Temporal semana Ibai	- 47 -
Figura 8 Serie Temporal hora Ibai	- 48 -
Figura 9 Wordcloud título Ibai	- 50 -
Figura 10 Wordcloud descripción Ibai.....	- 51 -
Figura 11 Wordcloud transcripción Ibai	- 52 -
Figura 12 Frases communes Ibai.....	- 53 -
Figura 13 Topic Modeling Ibai	- 54 -
Figura 14 Términos únicos Ibai.....	- 55 -
Figura 15 Network Ibai.....	- 56 -
Figura 16 Estadística Auronplay	- 57 -
Figura 17 Boxplot Auronplay.....	- 58 -
Figura 18 Histograma Auronplay.....	- 58 -
Figura 19 Scatterplot Auronplay	- 59 -
Figura 20 Matriz Correlación Auronplay	- 60 -
Figura 21 Serie Temporal mes Auronplay	- 60 -
Figura 22 Serie Temporal semana Auronplay	- 61 -
Figura 23 Serie Temporal hora Auronplay.....	- 62 -
Figura 24 Wordcloud título Auronplay.....	- 63 -
Figura 25 Wordcloud descripción Auronplay	- 64 -
Figura 26 Wordcloud transcripción Auronplay	- 65 -
Figura 27 Frases comunes Auronplay	- 65 -
Figura 28 Topic Modeling Auronplay	- 66 -
Figura 29 Términos únicos Auronplay	- 67 -
Figura 30 Network Auronplay	- 68 -
Figura 31 Estadística Marta Díaz	- 69 -
Figura 32 Boxplot Marta Díaz.....	- 70 -
Figura 33 Histograma Marta Díaz.....	- 70 -
Figura 34 Scatterplot Marta Díaz	- 71 -
Figura 35 Matriz Correlación Marta Díaz	- 72 -
Figura 36 Serie Temporal mes Marta Díaz	- 72 -
Figura 37 Serie Temporal semana Marta Díaz	- 73 -
Figura 38 Serie Temporal hora Marta Díaz.....	- 73 -
Figura 39 Wordcloud título Marta Díaz.....	- 74 -
Figura 40 Wordcloud descripción Marta Díaz	- 75 -
Figura 41 Wordcloud transcripción Marta Díaz	- 76 -
Figura 42 Frases comunes Marta Díaz	- 77 -
Figura 43 Topic Modeling Marta Díaz	- 78 -
Figura 44 Términos únicos Marta Díaz	- 78 -
Figura 45 Network Marta Díaz	- 80 -
Figura 46 Estadística DjMario.....	- 81 -

Figura 47 Boxplot DjMario	- 82 -
Figura 48 Histograma DjMario	- 82 -
Figura 49 Sactterplot DjMario.....	- 83 -
Figura 50 Matriz de correlación DjMario	- 84 -
Figura 51 Serie Temporal mes DjMario.....	- 84 -
Figura 52 Serie Temporal semana DjMario.....	- 85 -
Figura 53 Serie Temporal hora DjMario	- 85 -
Figura 54 Wordcloud título DjMario	- 86 -
Figura 55 Wordcloud descripción DjMario	- 87 -
Figura 56 Wordcloud transcripción DjMario.....	- 88 -
Figura 57 Frases comunes DjMario.....	- 89 -
Figura 58 Topic Modeling DjMario	- 89 -
Figura 59 Términos únicos DjMario	- 90 -
Figura 60 Network DjMario.....	- 91 -
Figura 61 Estadística Clakovi	- 92 -
Figura 62 Boxplot Clakovi	- 93 -
Figura 63 Histograma Clakovi.....	- 93 -
Figura 64 Scatterplot Clakovi.....	- 94 -
Figura 65 Matriz correlación Clakovi.....	- 95 -
Figura 66 Serie Temporal mes Clakovi.....	- 95 -
Figura 67 Serie Temporal semana Clakovi.....	- 96 -
Figura 68 Serie Temporal hora Clakovi.....	- 96 -
Figura 69 Wordcloud título Clakovi	- 97 -
Figura 70 Wordcloud descripción Clakovi	- 98 -
Figura 71 Wordcloud transcripción Clakovi	- 99 -
Figura 72 Frases comunes Clakovi.....	- 99 -
Figura 73 Topic Modeling Clakovi.....	- 100 -
Figura 74 Términos únicos Clakovi	- 100 -
Figura 75 Network Clakovi	- 101 -
Figura 76 Estadística ElRubius	- 102 -
Figura 77 Boxplot ElRubius.....	- 103 -
Figura 78 Histograma ElRubius	- 103 -
Figura 79 Scatterplot ElRubius	- 104 -
Figura 80 Matriz correlación ElRubius	- 104 -
Figura 81 Serie Temporal mes ElRubius	- 105 -
Figura 82 Serie Temporal semana ElRubius	- 106 -
Figura 83 Serie Temporal hora ElRubius	- 106 -
Figura 84 Wordcloud título ElRubius.....	- 107 -
Figura 85 Wordcloud descripción ElRubius.....	- 108 -
Figura 86 Wordcloud transcripción ElRubius	- 109 -
Figura 87 Frases comunes ElRubius	- 109 -
Figura 88 Topic modeling ElRubius	- 110 -
Figura 89 Términos únicos ElRubius	- 111 -
Figura 90 Network ElRubius	- 111 -
Figura 91 Estadística Emma Chamberlain	- 112 -
Figura 92 Boxplot Emma Chamberlain.....	- 113 -
Figura 93 Histograma Emma Chamberlain	- 113 -

Figura 94 Scatterplot Emma Chamberlain	- 114 -
Figura 95 Matriz correlación Emma Chamberlain	- 114 -
Figura 96 Serie Temporal mes Emma Chamberlain	- 115 -
Figura 97 Serie Temporal semana Emma Chamberlain	- 116 -
Figura 98 Serie Temporal hora Emma Chamberlain	- 116 -
Figura 99 Wordcloud título Emma Chamberlain	- 117 -
Figura 100 Wordcloud descripción Emma Chamberlain.....	- 118 -
Figura 101 Wordcloud transcripción Emma Chamberlain	- 119 -
Figura 102 Frases comunes Emma Chamberlain	- 120 -
Figura 103 Topic Modeling Emma Chamberlain	- 120 -
Figura 104 Términos únicos Emma Chamberlain.....	- 121 -
Figura 105 Network Emma Chamberlain	- 122 -
Figura 106 Estadística Malbert.....	- 123 -
Figura 107 Boxplot Malbert	- 124 -
Figura 108 Histograma Malbert	- 124 -
Figura 109 Scatterplot Malbert	- 125 -
Figura 110 Matriz correlación Malbert	- 125 -
Figura 111 Serie Temporal mes Malbert	- 126 -
Figura 112 Serie Temporal semana Malbert	- 126 -
Figura 113 Serie Temporal hora Malbert	- 127 -
Figura 114 Wordcloud título Malbert	- 127 -
Figura 115 Wordcloud descripción Malbert	- 128 -
Figura 116 Wordcloud transcripción Malbert	- 129 -
Figura 117 Frases comunes Malbert	- 130 -
Figura 118 Topic Modeling	- 130 -
Figura 119 Términos únicos Malbert	- 131 -
Figura 120 Netwrok Malbert.....	- 132 -
Figura 121 Estadística Mostopapi	- 133 -
Figura 122 Boxplot Mostopapi	- 134 -
Figura 123 Histograma Mostopapi.....	- 134 -
Figura 124 Sactterplot Mostopapi	- 135 -
Figura 125 Matriz correlación Mostopapi.....	- 135 -
Figura 126 Serie Temporal mes Mostopapi	- 136 -
Figura 127 Serie Temporal semana Mostopapi.....	- 137 -
Figura 128 Serie Temporal hora Mostopapi	- 137 -
Figura 129 Wordcloud título Mostopapi.....	- 138 -
Figura 130 Wordcloud descripción Mostopapi	- 139 -
Figura 131 Wordcloud transcripción Mostopapi	- 140 -
Figura 132 Frases comunes Mostopapi	- 141 -
Figura 133 Topic Modeling Mostopapi.....	- 141 -
Figura 134 Términos únicos Mostopapi	- 142 -
Figura 135 Estadística Orslok	- 144 -
Figura 136 Boxplot Orslok	- 145 -
Figura 137 Histograma Orslok.....	- 145 -
Figura 138 Sactterplot Orslok.....	- 146 -
Figura 139 Matriz correlación Orslok.....	- 147 -
Figura 140 Serie Temporal mes Orslok.....	- 147 -

Figura 141 Serie Temporal semana Orslok.....	- 148 -
Figura 142 Serie Temporal hora Orslok.....	- 148 -
Figura 143 Wordcloud título Orslok	- 149 -
Figura 144 Wordcloud descripción Orslok	- 150 -
Figura 145 Wordcloud transcripción Orslok	- 151 -
Figura 146 Frases comunes Orlsok.....	- 152 -
Figura 147 Topic Modeling	- 152 -
Figura 148 Términos únicos Orslok	- 153 -
Figura 149 Network Orslok	- 154 -
Figura 150 Estadística Vegetta	- 155 -
Figura 151 Boxplot Vegetta	- 156 -
Figura 152 Histograma Vegetta	- 156 -
Figura 153 Scatterplot Vegetta.....	- 157 -
Figura 154 Matriz correlación Vegetta	- 157 -
Figura 155 Serie Temporal mes Vegetta.....	- 158 -
Figura 156 Serie Temporal semana Vegetta.....	- 159 -
Figura 157 Serie Temporal hora Vegetta	- 159 -
Figura 158 Wordclooud título Vegetta	- 160 -
Figura 159 Wordcloud descripción Vegetta	- 160 -
Figura 160 Wordcloud transcripción Vegetta.....	- 161 -
Figura 161 Frases comunes Vegetta.....	- 162 -
Figura 162 Topic Modeling Vegetta.....	- 162 -
Figura 163 Términos únicos Vegetta	- 163 -
Figura 164 Network Vegetta	- 164 -
Figura 165 Network all influencers	- 165 -
Figura 166 Horas de trabajo.....	- 168 -

Resumen

El estudio profundiza en la interacción entre influencers y su audiencia en YouTube, la plataforma más destacada de España, poniendo especial énfasis en los propios influencers. El objetivo es descubrir su marca personal única y los valores esenciales que transmiten a través de su contenido. Integrando las disciplinas de Ciencia de Datos y Comunicación, se utilizaron técnicas de Inteligencia Artificial (IA), en particular el Procesamiento de Lenguaje Natural (NLP), para analizar datos de YouTube relacionados con una lista seleccionada de 11 influencers prominentes en España, elegidos por el equipo de la Dra. Sue Aran-Ramspott. Esta metodología centrada en la IA permitió un análisis eficiente y automatizado, minando grandes volúmenes de texto de los títulos, descripciones y transcripciones de videos de estos youtubers. Como resultado, se crearon redes utilizando los resultados del Reconocimiento de Entidades Nombradas (NER) con etiquetas de personas, ofreciendo una representación visual de las posibles relaciones entre estos influencers.

Palabras Clave: Adolescentes, youtubers, influencers, NER, IA, NLP y SNA

Abstract

The study delves into the interaction between influencers and their audience on Spain's premier platform YouTube, with a primary focus on the influencers themselves. The aim is to unravel their unique personal brand and the core values they impart through their content. By integrating Data Science and Communication disciplines, Artificial Intelligence (AI) techniques, notably Natural Language Processing (NLP), were utilized to scrutinize data from YouTube pertaining to a curated list of 11 standout influencers in Spain, handpicked by Dr. Sue Aran-Ramspott's team. This AI-centric approach facilitated an efficient and automated analysis, mining vast volumes of text from the titles, descriptions, and video transcriptions of these YouTubers. As an outcome, networks were crafted using the results from Name Entity Recognition (NER) with person tags, offering a visual representation of potential affiliations among these influencers.

Keywords: Teenagers, youtubers, influencers, NER, AI, NLP and SNA

Agradecimiento

En primer lugar, deseo expresar mi más profundo agradecimiento a mis tutores, el Dr. Xavier Vilasís Cardona, Profesor Titular de la Facultad de Ingeniería de La Salle de la Universidad Ramon Llull de Barcelona, y la Dra. Sue Aran-Ramspott, Profesora Titular de la Facultad de Comunicación y Relaciones Internacionales Blanquerna de la misma universidad. Ambos han sido pilares fundamentales en mi formación, brindándome no solo guía académica, sino también apoyo espiritual. Agradezco especialmente al Dr. Xavier Vilasís Cardona por las sesiones quincenales de seguimiento, en las cuales tuve la oportunidad de compartir mis ideas y resolver dudas, siempre con su inagotable paciencia y sabiduría.

Adicionalmente quisiera expresar mi sincero agradecimiento a Micaela Paredes Cusme, exalumna, quien ha sido de gran ayuda en este proyecto, asistiendo con mis inquietudes a través de correos electrónicos. Asimismo, estoy profundamente agradecido con la profesora Virginia Jiménez Ruano, cuyos valiosos comentarios y sugerencias enriquecieron la propuesta inicial del proyecto y guiaron su desarrollo.

Introducción

La transformación digital ha dejado una huella indeleble en la sociedad contemporánea. Esta metamorfosis, impulsada por avances tecnológicos y cambios en los patrones de consumo de información, ha reconfigurado la manera en que interactuamos, nos comunicamos y construimos nuestra identidad.[1] En este marco, las redes sociales han dejado de ser meras herramientas tecnológicas para convertirse en espacios vitales de interacción, debate y construcción de identidad.

La Generación Z, aquellos individuos nacidos entre mediados de los años 90 y principios del 2000, se encuentra en la vanguardia de esta revolución digital. Criados en un entorno saturado de tecnología, estos jóvenes no solo son consumidores de contenido digital, sino también creadores y propagadores activos[2]. En el caso de España, las estadísticas reflejan una alta penetración de redes sociales entre los jóvenes españoles, con plataformas como YouTube e Instagram no solo como medios de entretenimiento, sino también como canales de expresión y aprendizaje[3].

En este vasto universo digital, la figura del influencer ha cobrado una relevancia sin precedentes. Estos individuos, armados con carisma y habilidades de comunicación, han trascendido la tradicional noción de celebridad para convertirse en líderes de opinión, formadores de comunidades y, en muchos casos, en auténticos agentes de cambio. La relación que establecen con sus seguidores, denominada "parasocial", redefine las dinámicas tradicionales de interacción, creando un vínculo que, aunque unidireccional, genera un profundo sentido de pertenencia y lealtad[4].

Los adolescentes, en su etapa de formación y búsqueda de identidad, ven en estos influencers no solo a referentes estilísticos, sino también a guías ideológicos[5]. Las preferencias de contenido de esta generación, como se evidencia en el estudio de la Dra. Sue Aran-Ramspott, reflejan una combinación de entretenimiento, educación y exploración de identidad, con una marcada inclinación hacia contenidos lúdicos y entretenidos[6].

En este escenario multifacético, la Inteligencia Artificial (IA) emerge como una herramienta crucial. Su capacidad para procesar vastas cantidades de datos y discernir patrones complejos la posiciona como una aliada esencial en la comprensión de las dinámicas de las redes sociales[7]. Esta investigación se sumerge en el entramado digital con un enfoque académico riguroso, buscando desvelar las interacciones entre influencers y seguidores, y cómo estas dinámicas influyen en la construcción de la identidad y valores de la juventud española. Con el apoyo de la IA, se aspira a ofrecer una visión holística y detallada de la comunicación digital en la España contemporánea, proporcionando insights valiosos para académicos, profesionales y, en última instancia, para la sociedad en su conjunto.

Motivación y Objetivos

Motivación Personal

El autor opta por este tema a partir de una confluencia de experiencias académicas y un profundo interés personal. Durante su formación en el Máster Universitario de Marketing y Comunicación de la Universidad Autónoma de Barcelona, el autor adquirió competencias esenciales en el ámbito de las Redes Sociales. Esta base académica se ve complementada por la influencia de investigaciones previas, en particular, el trabajo desarrollado por la exalumna Micaela Paredes Cusme. Asimismo, se incorpora a su marco teórico el estudio dirigido por el equipo de la Dra. Sue Aran-Ramspott, centrado en las preferencias de los adolescentes españoles hacia YouTubers e Instagramers. Más allá

de la formación académica, el autor se sumerge en el análisis de los contenidos que los influencers producen en YouTube, aprovechando herramientas avanzadas de Inteligencia Artificial. Esta amalgama de formación, referencias y herramientas metodológicas le proporciona una perspectiva integral para abordar la investigación propuesta.

Objetivos del Proyecto

Objetivo Principal:

El propósito central de este proyecto es utilizar técnicas avanzadas de procesamiento de lenguaje natural para profundizar en el análisis y comprensión del contenido que se crea y difunde en la plataforma de YouTube.

Objetivos Secundarios:

Aparte del objetivo principal, se encuentran también tales objetivos secundarios para este estudio a lo siguiente:

1. Explorar en detalle las características y patrones lingüísticos que definen a los influencers de YouTube. Esto permitirá reconocer y entender la identidad única y el estilo distintivo que cada uno de ellos manifiesta en sus publicaciones.
2. Profundizar en los valores y temas que se destacan en los títulos, descripciones y transcripciones de los videos. Para ello, se emplearán técnicas especializadas de minería de texto que faciliten la identificación de tendencias y patrones recurrentes.
3. Investigar las dinámicas y relaciones que se establecen entre los influencers de YouTube. A través de herramientas como Name Entity Recognition (NER) y el análisis de redes sociales (Social Network Analysis, SNA), se busca descifrar las interconexiones y colaboraciones entre estos creadores de contenido.
4. A partir de los hallazgos obtenidos, se pretende elaborar propuestas estratégicas que no solo optimicen la comunicación en la plataforma, sino que también fomenten una influencia constructiva y positiva. El objetivo es fortalecer el vínculo entre la comunidad de creadores de contenido y su audiencia virtual en YouTube, promoviendo un espacio de interacción enriquecedor y respetuoso.

Marco Teórico

Una mirada hacia Redes Sociales

Conceptualización y Trayectoria Histórica de las RRSS

Las redes sociales (RRSS) se conceptualizan como plataformas digitales diseñadas para mediar y facilitar la interacción social en el ciberespacio. Estas herramientas permiten a los individuos construir perfiles digitales, establecer conexiones interpersonales y diseminar contenidos multimedia[8].

Evolución Histórica

- **Fase inicial (2000-2005):** Con plataformas pioneras como Friendster y MySpace, las RRSS se introdujeron como espacios primordialmente orientados a la socialización y compartición de intereses.
- **Fase de consolidación (2006-2010):** Con la irrupción y consolidación de plataformas como Facebook, Twitter y LinkedIn, las RRSS se posicionaron como herramientas esenciales tanto para la comunicación personal como profesional.

- **Fase multimedia (2011-2015):** La aparición de Instagram y Snapchat marcó una transición hacia un enfoque más visual y multimedia en la compartición de historias y experiencias.
- **Fase de diversificación (2016-presente):** Plataformas emergentes como TikTok y Clubhouse han diversificado el panorama de las RRSS, cada una con propuestas y públicos objetivos distintivos.[3]

Características Intrínsecas de las RRSS

- **Ubicuidad:** Las RRSS, accesibles desde cualquier dispositivo con conexión a Internet, trascienden las barreras geográficas y temporales, permitiendo la interacción en tiempo real.
- **Construcción de redes:** Estas plataformas posibilitan la formación y consolidación de redes de contactos, desde conexiones amistosas hasta vínculos profesionales.
- **Algoritmos personalizados:** Las RRSS implementan algoritmos sofisticados para curar y personalizar el contenido que los usuarios visualizan, basándose en patrones de interacción previa.
- **Economía de la atención:** En un ecosistema saturado de información, las RRSS compiten intensamente por capturar y retener la atención del usuario, incentivando la creación de contenidos virales y atractivos.[9]

Panorama de las RRSS en España: Datos y Plataformas Predominantes

Según los datos más recientes de 2023[3], España cuenta con aproximadamente 40,7 millones de usuarios activos en redes sociales, lo que equivale al 85,6% de su población total. En promedio, cada usuario español está presente en al menos 6 plataformas de redes sociales distintas, pasa 5 horas y 45 minutos al día en Internet, de las cuales 1 hora y 55 minutos están dedicadas a las redes sociales.

Plataformas más populares

- **Facebook:** A pesar de su consolidación global, esta plataforma ha experimentado una disminución en su popularidad entre los segmentos más jóvenes de la población española. Aunque sigue siendo relevante para ciertos grupos demográficos, ahora con un total de 21,1 millones usuarios, su atractivo ha disminuido frente a otras redes emergentes.
- **Instagram:** Esta red social ha logrado una penetración significativa en el mercado español. Su interfaz visual y las múltiples funcionalidades que ofrece la han posicionado como una de las redes sociales favoritas, especialmente entre las generaciones más jóvenes. Su capacidad para adaptarse a las tendencias y necesidades de los usuarios ha sido clave para su éxito en el territorio español. (23,8 millones usuarios)
- **Twitter:** Utilizada por un segmento considerable de la población internauta española, Twitter, con 4,4 millones de usuarios en España, se destaca por ser una plataforma centrada en la actualidad, debates públicos y la interacción en tiempo real. Su naturaleza dinámica y su enfoque en la brevedad del contenido la hacen única en su tipo.
- **YouTube:** Es una de las plataformas más populares y consolidadas no solo en España, sino a nivel global. Es la red social con más usuarios en España, contando con 40,7 millones. Esta plataforma ha permitido que creadores de contenido, desde vloggers hasta educadores, alcancen a audiencias masivas y ha servido como plataforma de lanzamiento para numerosos influencers y artistas.
- **Twitch:** Originalmente centrada en la transmisión de videojuegos, Twitch ha evolucionado para albergar una variedad de contenidos en vivo. En el contexto español, ha ganado una popularidad notable, especialmente entre los segmentos jóvenes y la comunidad gamer. Datos de 2021 indican que Twitch ha experimentado un crecimiento significativo en España, registrando un aumento del 67% en las horas de visualización en comparación con el año anterior.

- **TikTok:** 4,4 millones de usuarios. Esta red social ha experimentado un crecimiento exponencial en un corto período de tiempo, capturando principalmente la atención de la Generación Z. Su formato innovador, basado en videos cortos y creativos, junto con su algoritmo de recomendación, la han consolidado rápidamente en el panorama digital español. TikTok no solo ha transformado la forma en que se consume contenido, sino que también ha influenciado tendencias culturales y estilos de vida.[3]

Repercusiones Socioculturales de las RRSS en el Contexto Español

- **Dimensión Política:** Las RRSS han democratizado el acceso a la información política, sirviendo como catalizadores de movimientos sociales y plataformas de protesta.
- **Dimensión Cultural:** Artistas y creadores han adoptado estas plataformas como medios primordiales para la promoción de sus obras y la interacción con su público.
- **Dimensión Psicosocial:** Aunque las RRSS ofrecen espacios de conexión y socialización, también han sido objeto de estudio debido a sus posibles repercusiones en la salud mental, particularmente entre los jóvenes.
- **Dimensión Educativa:** El ámbito educativo ha integrado las RRSS como herramientas complementarias para el aprendizaje, promoviendo la compartición de recursos y la discusión académica.
- **Dimensión Económica:** El fenómeno de los influencers ha propiciado la emergencia de una nueva economía digital basada en la promoción y publicidad en línea, generando impactos significativos en el mercado publicitario español.[5]

En síntesis, las RRSS han instaurado una serie de transformaciones profundas en la sociedad española contemporánea, reconfigurando dinámicas comunicativas, culturales y económicas. Su influencia, en constante evolución, promete seguir modelando el tejido social en los años venideros.

La Vida Mediática del Adolescente en España

La era digital ha reconfigurado profundamente las interacciones y comportamientos de los adolescentes con su entorno. En países como España, las redes sociales se han erigido como elementos centrales en la cotidianidad de los jóvenes, desempeñando roles cruciales en áreas como la comunicación, el entretenimiento y la construcción de la identidad personal y colectiva.[5]

El Instituto Nacional de Estadística de España (INE) ha destacado que, en términos de participación en redes sociales, los estudiantes son particularmente activos, con un 94,4% de ellos involucrados en estas plataformas. De manera similar, los jóvenes entre 16 y 24 años muestran una alta tasa de participación, alcanzando el 92,6%. Estas cifras no solo subrayan la relevancia de las redes sociales en la vida de los jóvenes españoles, sino que también evidencian la necesidad de comprender a fondo las dinámicas y tendencias asociadas a su uso.[10]

Este análisis se propone examinar la trayectoria del uso de redes sociales entre los adolescentes en España, segmentando el estudio en tres fases clave: el periodo previo a la pandemia de COVID-19, la etapa de confinamiento y restricciones asociadas a la pandemia, y la fase posterior o de adaptación a la "nueva normalidad". Estas etapas, marcadas por contextos socioculturales y sanitarios distintos, ofrecen una panorámica comprensiva de cómo la vida mediática de los adolescentes ha evolucionado en respuesta a desafíos globales sin precedentes.[11]

1. Antes de la Pandemia:

Antes de la irrupción del COVID-19, las redes sociales ya eran populares entre los adolescentes españoles, pero su uso variaba según la plataforma:

- Facebook: Era utilizado regularmente por el 45% de los adolescentes.
- Instagram: Se destacaba como la plataforma favorita, con un 78% de uso regular.
- TikTok: Aunque estaba ganando popularidad, solo el 22% de los adolescentes lo usaban con frecuencia.
- Twitter: Mantenía una presencia constante con un 35% de adolescentes usándolo regularmente.

2. Durante la Pandemia:

La pandemia trajo consigo confinamientos y restricciones, lo que llevó a un aumento en el consumo digital. Las cifras reflejan este cambio en el comportamiento de los adolescentes:

- Facebook: A pesar de ser una de las redes más antiguas, su uso disminuyó ligeramente al 40%.
- Instagram: Su popularidad creció, alcanzando un 85% de uso regular entre los adolescentes.
- TikTok: Experimentó un auge significativo, con un 55% de adolescentes incorporándose a la plataforma.
- Twitter: Vio un ligero aumento, con un 38% de adolescentes usándolo regularmente.

3. Después de la Pandemia:

A medida que las restricciones comenzaron a levantarse y la vida empezó a regresar a la "normalidad", el uso de redes sociales mostró nuevas tendencias:

- Facebook: Continuó su tendencia decreciente, con solo el 38% de los adolescentes usándolo regularmente.
- Instagram: A pesar de su popularidad durante la pandemia, mostró una ligera disminución al 82%.
- TikTok: Siguió ganando terreno, con un 60% de adolescentes usándolo con frecuencia.
- Twitter: Experimentó una ligera disminución, con un 36% de uso regular.

La vida mediática de los adolescentes en España ha experimentado cambios significativos en los últimos años, influenciada en gran medida por factores externos como la pandemia de COVID-19. Las cifras reflejan una clara preferencia por plataformas visuales como Instagram y TikTok, mientras que redes más tradicionales como Facebook han visto una disminución en su uso. Es esencial entender estas tendencias para comprender mejor cómo los jóvenes españoles interactúan en el espacio digital y cómo esto afecta su desarrollo y bienestar.

La Emergencia y Consolidación de la Profesión "Influencer"

La era digital ha transformado no solo la forma en que nos comunicamos, sino también la manera en que consumimos información y nos relacionamos con figuras públicas. En este contexto, ha surgido una nueva figura profesional: el influencer.[12] A continuación, se presenta una introducción detallada sobre su emergencia, clasificación, relación con la audiencia, impacto en la sociedad y su situación actual en España.

Origen y Nacimiento del Influencer

La influencia social ha sido una constante en la historia de la humanidad. Sin embargo, el término "influencer" como lo conocemos hoy tiene sus raíces en la popularización de las redes sociales a mediados de la década de 2000[12]. Originalmente, los bloggers y vloggers, que compartían sus opiniones y experiencias en plataformas como Blogspot o YouTube, fueron los pioneros en ejercer

influencia digital. Con la evolución de las redes sociales y la aparición de plataformas como Instagram, Snapchat y TikTok, la figura del influencer se consolidó, diversificando su presencia y alcance.

Clasificación de Influencers

La categorización de los influencers se realiza generalmente en función de su número de seguidores y su nicho de especialización. Los micro-influencers, con entre 1.000 y 100.000 seguidores, suelen tener una relación más estrecha y personalizada con su audiencia, centrada en nichos específicos. Los macro-influencers, con seguidores que oscilan entre 100.000 y 1 millón, tienen un alcance más amplio, a menudo a nivel nacional. Por último, los mega-influencers superan el millón de seguidores y su influencia puede tener un alcance global, siendo en muchos casos figuras públicas en otros ámbitos antes de su incursión en las redes[4].

Relación con la Audiencia y Construcción de Confianza: La Interacción Parasocial

La relación entre influencers y su audiencia es única y se distingue claramente de las relaciones tradicionales que los fanáticos tienen con las celebridades. Esta relación se puede entender mejor a través del concepto de "interacción parasocial".

La interacción parasocial es un término acuñado por los psicólogos Donald Horton y Richard Wohl en 1956 para describir una especie de relación unilateral que se establece entre una audiencia y una figura mediática. En esta relación, los individuos (espectadores o seguidores) sienten que tienen una conexión personal y directa con la figura mediática, a pesar de que esta relación no es recíproca en el sentido tradicional. La figura mediática no conoce personalmente a cada uno de sus seguidores, pero estos últimos sienten que tienen un conocimiento íntimo y una conexión con ella.

En el contexto de los influencers, esta interacción parasocial se intensifica. A diferencia de las celebridades tradicionales, que pueden mantener una distancia y un aura de inaccesibilidad, los influencers prosperan en la apertura y la accesibilidad. Comparten detalles diarios de su vida, responden directamente a los comentarios y, en muchos casos, narran sus experiencias personales, desafíos y logros. Esta transparencia crea una ilusión de amistad y cercanía[13].

La tecnología y las plataformas de redes sociales amplifican esta sensación. Las "historias" en Instagram, los "directos" en varias plataformas y las encuestas permiten una interacción en tiempo real, lo que refuerza la sensación de una relación genuina y directa.

Sin embargo, es esencial reconocer que, aunque estas relaciones parasociales pueden ofrecer consuelo, conexión y comunidad a muchos seguidores, también pueden llevar a expectativas poco realistas y a una inversión emocional desproporcionada. La línea entre la realidad y la representación mediática puede volverse borrosa, y es crucial que tanto influencers como seguidores sean conscientes de las dinámicas y limitaciones de estas relaciones[4].

Impacto Sociocultural y Económico

La influencia de estas figuras ha trascendido el ámbito digital, impactando en la cultura, la economía y la sociedad en general. A continuación, se detallan algunos de estos impactos:

- **Reconfiguración del Marketing y la Publicidad:** Los influencers han revolucionado las tácticas tradicionales de marketing. Las marcas ya no dependen exclusivamente de los medios convencionales para promocionar sus productos o servicios. En lugar de ello, colaboran con

influencers que, gracias a su cercanía con el público y su capacidad para generar confianza, promueven de manera más efectiva y orgánica. Esta nueva forma de publicidad, denominada "marketing de influencia", ha demostrado ser especialmente efectiva en segmentos jóvenes de la población[14].

- **Debates sobre Representación y Autenticidad:** La visibilidad y el alcance de los influencers han llevado a discusiones críticas sobre quién tiene voz en la esfera digital. Se cuestiona la autenticidad de sus contenidos, la representación de minorías y la perpetuación de ciertos estándares de belleza, éxito y bienestar. Estos debates han llevado a una mayor conciencia sobre la responsabilidad que tienen estas figuras al influir en las percepciones y aspiraciones de sus seguidores[15].
- **Activismo y Concienciación Social:** Muchos influencers han trascendido la mera promoción de productos para convertirse en voceros de causas sociales, políticas y medioambientales. Su capacidad para movilizar a grandes audiencias les ha permitido generar conciencia sobre temas de relevancia global, desde el cambio climático hasta los derechos humanos. Esta influencia ha sido evidente en movimientos sociales recientes, donde influencers han jugado un papel crucial en la movilización y difusión de información[16].
- **Impacto Económico Directo:** La influencia digital ha dado lugar a nuevas oportunidades económicas. Los influencers generan ingresos a través de patrocinios, publicidad y la venta de productos propios. Además, han impulsado sectores económicos relacionados, como la fotografía, la edición de video y la gestión de redes sociales[12].
- **Educación y Formación:** Algunos influencers se han especializado en áreas educativas, ofreciendo tutoriales, cursos y consejos en áreas tan diversas como la ciencia, la tecnología o las humanidades. Esta democratización del conocimiento ha permitido que muchas personas accedan a información y formación de calidad de manera gratuita o a bajo costo[12].

En resumen, la figura del influencer ha dejado una huella indeleble en la sociedad contemporánea, reconfigurando la forma en que consumimos información, nos relacionamos con las marcas y percibimos la realidad. Su impacto, tanto positivo como negativo, seguirá siendo objeto de estudio y debate en los años venideros.

El Influencer en el Contexto Español

España no ha sido ajena a esta tendencia global. El país ha visto surgir influencers reconocidos a nivel internacional en diversas áreas, desde la moda hasta la gastronomía. Plataformas como Instagram y YouTube dominan el panorama, aunque TikTok ha ganado terreno rápidamente entre los más jóvenes. La profesionalización del sector ha llevado a la necesidad de una regulación, especialmente en lo que respecta a la publicidad y las colaboraciones comerciales, garantizando transparencia y protección al consumidor[17].

El Panorama de YouTube y sus Protagonistas: Los YouTubers

En el vasto universo de las redes sociales, YouTube destaca como la plataforma líder[3], convirtiéndose en el epicentro de la creación de contenido y la interacción digital. En este contexto, los YouTubers emergen como figuras centrales, siendo el objeto de análisis principal de nuestro proyecto.

Un estudio dirigido por la Dra. Sue Aran-Ramspott revela patrones interesantes en el comportamiento de los usuarios de YouTube. Los encuestados señalaron que recurren a YouTube principalmente para "entretenerte", "distraerse" y "adquirir conocimientos". Aunque también existen motivaciones secundarias, como "seguir a celebridades" y "ver lo que otros están viendo".[6]

Antes de sumergirnos en técnicas avanzadas de Inteligencia Artificial, es esencial comprender las características intrínsecas de los YouTubers y el valor que aportan a su joven audiencia:

Características de los YouTubers

- **Espontaneidad:** Aunque los YouTubers proyectan una imagen de autenticidad y frescura, esto no garantiza la originalidad en su contenido.
- **Apropiación de Contenido:** Es común que los YouTubers se inspiren o incluso reproduzcan contenido de otros, estableciendo una especie de identidad colectiva. A pesar de las críticas, esta práctica no parece mermar su autenticidad ante su audiencia.
- **Humor:** Predomina en sus creaciones, variando desde un tono ligero hasta uno más irreverente, construyendo así una identidad accesible y cercana.
- **Lenguaje y Comunicación:** Utilizan un lenguaje coloquial, a menudo salpicado de jerga juvenil y expresiones audaces. Además, recurren a herramientas como parodias, efectos visuales, hashtags y emojis, reflejando la dinámica comunicativa de la juventud digital.[6]

Valor que Aportan a la Audiencia Joven

- **Conexión y Comunidad:** Fomentan un sentido de pertenencia, creando espacios donde los jóvenes se sienten comprendidos y representados.
- **Modelos a Seguir:** Representan ideales de autenticidad y estilos de vida que resuenan en las generaciones emergentes.
- **Entretenimiento y Cercanía:** Los jóvenes gravitan hacia YouTubers que abordan temas de su interés de manera amena y directa. Aunque aspectos como la inteligencia y la apariencia son valorados, no son determinantes.
- **Contribución Cultural y Social:** Durante la crisis de la Covid-19, los YouTubers se convirtieron en puentes culturales, introduciendo a los jóvenes a nuevas perspectivas, estilos de vida y conocimientos esenciales. Además, ofrecieron un refugio emocional, reafirmando a los jóvenes en sus creencias y emociones.[6]

Estos puntos resaltan la influencia y el impacto que tienen los YouTubers en la audiencia joven, así como las características que los hacen atractivos para este grupo demográfico.

Breve Introducción de Inteligencia Artificial

En la confluencia de la tecnología y la sociedad, la Inteligencia Artificial (IA) se erige como un pilar transformador, redefiniendo cómo interactuamos, comprendemos y evolucionamos en el mundo digital. Su capacidad para automatizar procesos y desentrañar patrones complejos en vastos conjuntos de datos la posiciona más allá de las capacidades analíticas tradicionales. En el marco de este proyecto, que se adentra en el universo de los YouTubers y su resonancia en la audiencia, la IA se convierte en una herramienta indispensable. Nos brinda la posibilidad de descifrar tendencias de contenido, dinámicas de interacción y comportamientos de usuarios con una precisión y escala sin precedentes[7]. Al incorporar la IA en nuestro enfoque, aspiramos no solo a capturar la esencia del ecosistema actual de YouTube, sino también a anticipar las corrientes emergentes en este espacio digital en constante metamorfosis. Esta sección introductoria establece el escenario para una exploración más profunda de cómo la sinergia entre la IA y el análisis sociocultural puede desvelar perspectivas enriquecedoras en el vibrante ámbito de los YouTubers. A continuación, se detallarán las teorías y técnicas específicas de IA que se han implementado en este proyecto.

El Procesamiento de Lenguaje Natural

El Procesamiento de Lenguaje Natural (NLP) es una especialidad dentro de la Inteligencia Artificial que busca simular la capacidad humana de interpretar y generar lenguaje a través de máquinas. Mediante técnicas avanzadas y algoritmos de vanguardia, el NLP ha propiciado innovaciones en áreas como la traducción automática, generación de texto, análisis de sentimientos y extracción de información relevante.[18]

Dentro del ámbito de nuestro proyecto, que se enfoca en el fenómeno de los YouTubers, el NLP desempeña un papel crucial. Estos creadores de contenido producen una vasta cantidad de material textual, ya sea en títulos, descripciones o transcripciones de sus videos. Al aplicar técnicas de NLP a este contenido, es posible obtener percepciones profundas sobre:

- **Tendencias Temáticas:** Permitiendo identificar patrones en los temas abordados, ya sean recurrentes o novedosos.
- **Patrones Lingüísticos:** Revelando estilos de comunicación particulares, jergas distintivas o frases que se vuelven icónicas en ciertos géneros o entre YouTubers específicos.

La integración del NLP en nuestra investigación nos brinda la posibilidad de ir más allá del contenido explícito, adentrándonos en las complejidades y matices del lenguaje que caracterizan el universo de YouTube. Esta aproximación nos ofrece una ventana hacia el núcleo lingüístico y emocional que subyace en la creación de contenido, permitiéndonos comprender de manera más detallada las dinámicas que animan este ecosistema digital en constante evolución.[19]

Adicionalmente, la minería de textos, una faceta especializada del NLP, se dedica a la extracción de información significativa de conjuntos extensos de datos textuales. A través de diversas técnicas y herramientas, se pueden discernir patrones y relaciones que, a simple vista, permanecerían ocultas. En relación con los YouTubers, la minería de textos se convierte en una herramienta primordial para desglosar y entender el inmenso volumen de contenido que generan. En las siguientes secciones, exploraremos en detalle las técnicas y conceptos esenciales de la minería de textos aplicados a este contexto.

Corpus

En el contexto del Procesamiento de Lenguaje Natural (NLP) y la lingüística, un corpus se refiere a un conjunto estructurado de textos escritos, orales o multimodales que se recopilan con un propósito específico. Estos textos pueden abarcar desde transcripciones de conversaciones, artículos de periódicos, libros, publicaciones en redes sociales, hasta grabaciones de voz, entre otros. La teoría del corpus se centra en el estudio y análisis de estos conjuntos de datos lingüísticos para obtener insights sobre el uso y estructura del lenguaje[20].

Características Principales del Corpus

- **Representatividad:** Un corpus debe ser representativo del tipo de lenguaje o variedad lingüística que se pretende estudiar. Por ejemplo, un corpus diseñado para estudiar el lenguaje coloquial debería contener ejemplos auténticos de conversaciones cotidianas.
- **Autenticidad:** Los textos incluidos en un corpus deben ser auténticos, es decir, no deben ser alterados o fabricados para el estudio. Deben reflejar el uso real del lenguaje en contextos naturales.
- **Equilibrio y Diversidad:** Idealmente, un corpus debe tener un equilibrio en términos de géneros, estilos, contextos y participantes. Esto asegura una amplia cobertura y diversidad del lenguaje.

Tipos de Corpus

- **Corpus Monolingües:** Contienen textos de un solo idioma.

- **Corpus Paralelos:** Contienen textos en dos o más idiomas, donde cada texto en un idioma tiene su correspondiente traducción en otro(s) idioma(s).
- **Corpus Multimodales:** Combinan diferentes modos de comunicación, como texto, imagen y sonido.
- **Corpus Dinámicos:** Se actualizan y expanden continuamente con nuevos textos.
- **Corpus Anotados:** Contienen información adicional sobre los textos, como anotaciones gramaticales, semánticas o fonéticas.

Tokens

Un "token" es una unidad básica en el procesamiento de texto. En la mayoría de los contextos, un token se refiere a una palabra, pero dependiendo del nivel de análisis, un token también puede referirse a un carácter, una oración o incluso un párrafo. La acción de dividir un texto en tokens se denomina "tokenización".[21]

Características y Aspectos Clave de los Tokens

- **Unidad Básica:** En el contexto del NLP, un token es generalmente la unidad básica de procesamiento. Por ejemplo, al analizar la frecuencia de palabras en un texto, cada palabra se considera un token.
- **Tokenización:** Es el proceso de convertir un texto en una lista de tokens. Por ejemplo, la frase "El gato duerme" se tokenizaría en ["El", "gato", "duerme"].
- **Delimitadores:** Durante la tokenización, se utilizan delimitadores para determinar dónde termina un token y comienza otro. Los delimitadores comunes incluyen espacios, puntuación y saltos de línea.
- **Variedad de Tokens:** Dependiendo del análisis, un token puede ser una palabra, una oración, un párrafo o incluso un carácter individual. Por ejemplo, al analizar la estructura de las oraciones, cada oración podría considerarse un token.
- **Normalización:** Después de la tokenización, es común realizar una normalización de los tokens. Esto puede incluir convertir todo el texto a minúsculas, eliminar puntuación o aplicar técnicas como la lematización o el stemming para reducir las palabras a su forma base.

La tokenización es un paso fundamental en muchos procesos de NLP. Al dividir un texto en tokens, se facilita el análisis y procesamiento del mismo. Los tokens se utilizan en una variedad de tareas, como:

- **Análisis de Frecuencia:** Determinar cuántas veces aparece cada palabra en un texto.
- **Búsqueda de Texto:** Encontrar y recuperar información específica de un conjunto de datos.
- **Traducción Automática:** Traducir texto de un idioma a otro.
- **Reconocimiento de Entidades Nombradas:** Identificar y clasificar nombres propios, como nombres de personas, lugares o empresas.
- **Modelos de Lenguaje:** Predecir la siguiente palabra en una secuencia.

TF-IDF

TF-IDF es un acrónimo que representa "Frecuencia de Término - Frecuencia Inversa de Documento". Es una técnica estadística utilizada para evaluar la importancia de una palabra en un documento en relación con un conjunto de documentos o corpus. Es ampliamente utilizado en tareas de recuperación de información y minería de textos.[20]

- **TF (Frecuencia de Término):** Mide la frecuencia con la que aparece un término en un documento. Se calcula como:

$$TF(t) = \frac{\text{Número de veces que el término } t \text{ aparece en el documento}}{\text{Número total de términos en el documento}}$$

- **IDF (Frecuencia Inversa de Documento):** Evalúa la importancia del término en el corpus completo. Se calcula como:

$$IDF(t) = \log\left(\frac{\text{Número total de documentos}}{1 + \text{Número de documentos con el término } t}\right)$$

La puntuación TF-IDF para un término en un documento específico se calcula multiplicando su TF y su IDF:

$$TF-IDF(t) = TF(t) \times IDF(t)$$

Aplicaciones:

- **Resaltar Palabras Clave:** TF-IDF puede identificar palabras que son cruciales para el significado de un documento dentro de un corpus. Las palabras que aparecen frecuentemente en un documento, pero no en muchos documentos del corpus, tendrán una alta puntuación TF-IDF.
- **Recuperación de Información:** Es fundamental en sistemas de búsqueda para clasificar y recuperar documentos relevantes basados en términos de consulta.
- **Extracción de Características:** En tareas de aprendizaje automático, TF-IDF se utiliza para convertir textos en vectores numéricos que pueden ser usados como entrada para modelos.
- **Filtrado de Palabras Comunes:** Palabras que aparecen en muchos documentos del corpus (como "y", "el", "la") tendrán una baja puntuación IDF, lo que reduce su relevancia en el análisis.[22]

N-grams

Un N-gram es una secuencia contigua de n ítems (pueden ser sílabas, caracteres o palabras) de un texto o discurso dado. Los N-grams son ampliamente utilizados en lingüística y procesamiento de lenguaje natural para analizar la estructura y contenido de los textos.[20]

Definición y Ejemplos:

- **Unigram:** Cuando n=1, se trata de un único ítem. Por ejemplo, en la frase "Me gusta el sol", los unigrams serían "Me", "gusta", "el", "sol".
- **Bigram:** Cuando n=2, son dos ítems consecutivos. Usando la misma frase, los bigrams serían: "Me gusta", "gusta el", "el sol".
- **Trigram:** Con n=3, son tres ítems en secuencia. En este caso, los trigrams serían: "Me gusta el", "gusta el sol".

Y así sucesivamente. A medida que n aumenta, los N-grams capturan más contexto, pero también se vuelven más escasos en el texto.

Aplicaciones:

- **Modelos de Lenguaje:** Los N-grams son esenciales para construir modelos de lenguaje probabilísticos, que predicen la siguiente palabra en una secuencia basándose en las n-1 palabras anteriores.
- **Corrección Ortográfica:** Se utilizan para detectar y corregir errores en textos basándose en la probabilidad de aparición de ciertas secuencias de palabras.
- **Detección de Plagio:** Al comparar la frecuencia y estructura de N-grams entre diferentes textos, es posible identificar posibles casos de copia.
- **Análisis de Sentimiento:** Los N-grams pueden ayudar a identificar frases o combinaciones de palabras que tienen connotaciones positivas o negativas.

Consideraciones:

- **Esparsidad:** A medida que n aumenta, los N-grams se vuelven más raros, lo que puede dificultar su análisis estadístico. Por ejemplo, un 7-gram probablemente aparecerá muy pocas veces en un texto.
- **Almacenamiento:** Los modelos basados en N-grams, especialmente con valores altos de n, pueden requerir una gran cantidad de memoria para almacenar todas las combinaciones posibles.
- **Contexto vs. Ruido:** Mientras que los N-grams capturan más contexto a medida que n aumenta, también pueden introducir ruido en el análisis, especialmente si las combinaciones de palabras son poco comunes o irrelevantes.[23]

Lematización y Stemming

El NLP a menudo requiere reducir las palabras a una forma base para facilitar el análisis y la comparación. Dos técnicas comunes para lograr esto son la lematización y el stemming. Ambas técnicas buscan reducir las palabras a su forma raíz, pero lo hacen de maneras diferentes y tienen diferentes aplicaciones y resultados.

Lematización

- **Definición:** La lematización implica reducir una palabra a su forma base o lema. Esta forma base es la que se encontraría en un diccionario o en un léxico.
- **Proceso:** La lematización considera el contexto y utiliza el análisis morfológico para reducir la palabra. Por ejemplo, el verbo "corriendo" se lematizaría como "correr".
- **Herramientas:** Las herramientas de lematización, como el lematizador de WordNet en la biblioteca NLTK, a menudo requieren información adicional sobre una palabra, como su parte de discurso.
- **Ventajas:** Produce resultados más precisos y coherentes al considerar el contexto y la morfología de la palabra.

Stemming

- **Definición:** El stemming implica reducir una palabra a su raíz o base mediante la eliminación de afijos (como prefijos o sufijos) sin tener en cuenta el significado real de la palabra.
- **Proceso:** El stemming utiliza algoritmos heurísticos que cortan los extremos de las palabras. Por ejemplo, "corriendo", "corredor" y "corre" podrían reducirse al mismo stem "corr-".
- **Herramientas:** Algunos algoritmos populares de stemming incluyen el algoritmo de Porter y el algoritmo de Snowball.
- **Ventajas:** Es más rápido y menos complejo que la lematización. Sin embargo, puede producir resultados que no son palabras reales.[21]

Comparación:

- **Precisión vs. Velocidad:** La lematización es más precisa ya que considera el contexto y la morfología de la palabra, pero es computacionalmente más costosa. El stemming es más rápido, pero puede producir resultados menos precisos.
- **Aplicación:** Ambas técnicas son útiles en tareas como la recuperación de información y el análisis de texto. La elección entre lematización y stemming a menudo depende de la aplicación específica y de los recursos disponibles.[24]

Topic Modeling

El modelado de temas es una técnica de procesamiento de lenguaje natural (NLP) y aprendizaje automático utilizada para descubrir los temas latentes que se distribuyen en una colección de

documentos. Es especialmente útil para organizar, entender y resumir grandes conjuntos de datos textuales.

Métodos Principales:

- **LDA (Latent Dirichlet Allocation):** Es el algoritmo más popular para el modelado de temas. LDA asume que cada documento es una mezcla de temas y un tema es una mezcla de palabras. A través de iteraciones, LDA ajusta las asignaciones de palabras a temas y temas a documentos para maximizar la coherencia y la distribución.
- **NMF (Non-negative Matrix Factorization):** Es una técnica de factorización que puede ser utilizada para el modelado de temas. Descompone la matriz de documentos y términos en dos matrices más pequeñas: una matriz de documentos y temas y otra matriz de temas y términos.
- **LSA (Latent Semantic Analysis):** Es una técnica que utiliza la descomposición de valores singulares (SVD) para reducir la dimensionalidad de la matriz de término-documento, lo que permite identificar patrones y temas en los documentos.

Aplicaciones:

- **Organización de Información:** Ayuda a categorizar y etiquetar grandes conjuntos de datos textuales, como artículos de noticias o entradas de blogs.
- **Recomendación de Contenido:** Al identificar temas en documentos, se pueden hacer recomendaciones basadas en la similitud temática.
- **Descubrimiento de Conocimiento:** Permite descubrir tendencias, patrones y temas emergentes en grandes conjuntos de datos.

Proceso Básico:

- **Preprocesamiento:** Incluye la eliminación de palabras vacías, lematización o stemming, y la conversión de documentos a una representación vectorial (como TF-IDF).
- **Elección del Número de Temas:** A menudo se requiere especificar el número de temas que el modelo debe encontrar. Esto puede hacerse a través de conocimiento experto o técnicas como la optimización de la coherencia.
- **Modelado:** Aplicación de algoritmos como LDA, NMF o LSA para descubrir temas.
- **Evaluación:** Una vez que se han identificado los temas, se pueden evaluar utilizando métricas como la coherencia de temas para determinar la calidad y relevancia de los temas descubiertos.
- **Interpretación:** Los temas identificados se interpretan y etiquetan según las palabras clave asociadas a cada tema.

Consideraciones:

- **Interpretabilidad:** No todos los temas identificados por los algoritmos serán interpretables o útiles. La interpretación y validación a menudo requieren intervención humana.
- **Elección del Número de Temas:** La elección del número correcto de temas es crucial y puede requerir experimentación.[25]

NER

El Reconocimiento de Entidades Nombradas (NER) es una sub-tarea del NLP que se centra en identificar y clasificar entidades nombradas en texto en categorías predefinidas como nombres de personas, organizaciones, lugares, fechas, porcentajes, entre otros.

Componentes Principales:

- **Entidad:** Cualquier palabra o secuencia de palabras que represente un objeto específico o nombre propio. Por ejemplo: "Barcelona", "Leonardo da Vinci", "2023".
- **Categoría:** La clase o tipo al que pertenece una entidad. Las categorías comunes incluyen: **PERSON** (persona), **ORG** (organización), **GPE** (entidades geopolíticas), **DATE** (fecha), entre otras.

Aplicaciones:

- **Extracción de Información:** Extraer detalles específicos de grandes volúmenes de texto, como noticias o artículos académicos.
- **Búsqueda Mejorada:** Mejorar la precisión de los motores de búsqueda al identificar y dar prioridad a las entidades en los documentos.
- **Análisis de Sentimiento:** Determinar el sentimiento hacia entidades específicas en el texto.
- **Automatización y Chatbots:** Ayudar a los chatbots y asistentes virtuales a entender y responder a las consultas de los usuarios al identificar entidades clave en el input del usuario.

Proceso Básico:

- **Preprocesamiento:** El texto se limpia y se tokeniza para prepararlo para el análisis.
- **Identificación:** Las palabras o secuencias de palabras que se parecen a entidades conocidas se marcan. Por ejemplo, las palabras que comienzan con mayúsculas en medio de una oración podrían ser candidatas para ser entidades.
- **Clasificación:** Las entidades identificadas se clasifican en categorías predefinidas.
- **Post-procesamiento:** Las entidades y categorías identificadas se refinan y ajustan según sea necesario.

Herramientas y Técnicas:

- **Modelos Estadísticos:** Utilizan grandes conjuntos de datos anotados para entrenar modelos que pueden identificar y clasificar entidades en texto nuevo.
- **Reglas Basadas en Gramática:** Utilizan reglas lingüísticas y patrones para identificar entidades.
- **Deep Learning:** Modelos como las redes neuronales recurrentes (RNN) y las redes neuronales de transformadores (como BERT) han demostrado ser efectivos en tareas de NER.

Consideraciones:

- **Ambigüedad:** Una misma palabra puede ser una entidad en un contexto y no serlo en otro. Por ejemplo, "Júpiter" podría referirse al planeta o a la deidad romana.
- **Dominio Específico:** Los modelos de NER entrenados en un dominio (por ejemplo, noticias) pueden no funcionar bien en otro dominio (por ejemplo, literatura médica). Es posible que se requiera entrenamiento específico del dominio.

Al aplicar estas técnicas y conceptos de NLP al mundo de YouTube, no solo obtenemos una comprensión más profunda de los creadores y su contenido, sino que también podemos identificar tendencias emergentes, patrones de interacción con la audiencia y la evolución del lenguaje y la cultura en esta plataforma. En resumen, el NLP proporciona las herramientas necesarias para navegar y analizar el vasto y dinámico ecosistema de YouTube, ofreciendo insights valiosos para investigadores, anunciantes y, por supuesto, los propios creadores.[26]

Social Network Analysis

Mientras que la Inteligencia Artificial (IA) se centra en emular capacidades humanas de aprendizaje y razonamiento en máquinas, existe otra técnica complementaria pero distinta, conocida como Análisis de Redes Sociales (SNA). Aunque el SNA no pertenece al dominio de la IA, es una

herramienta poderosa que se aplicará en nuestro estudio para examinar las interacciones y relaciones dentro de la red de influencers. El SNA se enfoca en desentrañar la estructura intrínseca de las redes, identificando actores clave, las conexiones entre ellos y la configuración global de la red. A través de este análisis, podemos descifrar patrones de interacción, determinar influencias predominantes y comprender los flujos de información que circulan en la red. En el contexto de nuestro proyecto, el SNA nos proporcionará una visión detallada de cómo los influencers se relacionan entre sí y con su audiencia, ofreciendo insights valiosos sobre la dinámica y estructura de esta comunidad digital.

Componentes Principales

- **Nodos:** Representan las entidades individuales dentro de la red. Pueden ser personas, organizaciones, países, etc.
- **Enlaces o Aristas:** Representan las relaciones o interacciones entre los nodos. Pueden ser amistades, comunicaciones, transacciones comerciales, entre otros.[27]

Métricas Clave

- **Grado:** Número de conexiones directas que tiene un nodo.

$$\text{Grado}(v) = \sum_{u \in V} a_{uv}$$

Donde V es el conjunto de nodos y a_{uv} es 1 si hay un enlace entre u y v , y 0 en caso contrario.

- **Centralidad de Cercanía:** Mide cuán cerca está un nodo de todos los demás nodos en la red.

$$C(u) = \frac{n - 1}{\sum_{v=1}^n d(u, v)}$$

Donde n es el número total de nodos y $d(u, v)$ es la distancia más corta entre los nodos u y v .

- **Centralidad de Intermediación:** Mide la influencia de un nodo en términos de controlar la comunicación entre otros nodos.

$$B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Donde σ_{st} es el número total de caminos más cortos de s a t y $\sigma_{st}(v)$ es el número de esos caminos que pasan por v .

- **Densidad:** Proporción de conexiones directas en una red en relación con el total de conexiones posibles.

$$D = \frac{2L}{n(n - 1)}$$

Donde L es el número total de enlaces y n es el número total de nodos.

- **Diámetro:** La distancia más larga entre dos nodos en la red.
- **Coeficiente de Clustering:** Mide la tendencia de los nodos a agruparse o formar clusters.

$$C(v) = \frac{2e_v}{k_v(k_v - 1)}$$

Donde e_v es el número de enlaces entre los vecinos de v y k_v es el grado de v .[28]

Aplicaciones:

- **Sociología:** Estudiar patrones de amistad, estructuras organizativas o dinámicas de grupos.
- **Biología:** Analizar redes de interacción entre proteínas o genes.
- **Informática:** Estudiar la estructura de la web, redes de computadoras o sistemas distribuidos.
- **Economía:** Analizar redes de transacciones comerciales o flujos financieros.
- **Estudios de Comunicación:** Analizar patrones de comunicación, difusión de información o influencia en redes sociales.

Proceso Básico:

- **Recolección de Datos:** Obtener datos sobre los nodos y las relaciones. Esto puede hacerse a través de encuestas, observación, bases de datos, entre otros.
- **Construcción de la Red:** Usar los datos para construir una representación gráfica o matemática de la red.
- **Análisis:** Aplicar métricas y técnicas para analizar la estructura y las propiedades de la red.
- **Interpretación:** Derivar insights y conclusiones basadas en el análisis.[29]

Herramientas y Técnicas:

- **Software Específico:** Herramientas como Gephi, UCINet o Pajek son ampliamente utilizadas para el análisis y visualización de redes.
- **Modelos Matemáticos:** Los grafos son una representación matemática común de las redes, y se utilizan para describir y analizar su estructura.[30]

Consideraciones:

- **Tamaño y Complejidad:** Las redes grandes y complejas pueden ser difíciles de analizar y visualizar.
- **Dinámica:** Las redes pueden cambiar con el tiempo, lo que requiere análisis dinámicos.
- **Calidad de los Datos:** Las conclusiones derivadas del SNA son tan buenas como los datos en los que se basan.[29]

El Análisis de Redes Sociales proporciona un marco poderoso para estudiar sistemas complejos y sus interacciones. Al centrarse en las relaciones y la estructura, el SNA puede revelar patrones, influencias y flujos de información que de otro modo serían invisibles. Ya sea en sociología, biología, economía o tecnología, el SNA ofrece insights valiosos y una comprensión más profunda de la naturaleza interconectada de nuestro mundo.

Diseño del Proyecto

Descripción y explicación de los youtubers escogidos

En este trabajo, se aplicará las técnicas de ciencia de datos, en su mayoría, las de IA, con el fin de analizar los contenidos generados y publicados en YouTube de una lista de ciertos influencers, la cual viene proporcionada por el grupo de investigación DIGILAB de la Facultad de comunicación y relaciones internacionales de Blanquerna de la Universidad Ramon Llull, y cuenta con una totalidad de 590 influncers mencionados por los jóvenes[6]. De la totalidad de youtubers, para hacer este análisis se escogerá una lista reducida de 11 influencers representativos teniendo en cuenta la variedad de género, contenido y suscriptores o ranking de importancia, entre ellos son:

- Ibai: Es un comentarista de deportes electrónicos y streamer que se hizo famoso por narrar partidas de League of Legends. En su canal de YouTube, sube vídeos de humor, entrevistas y reacciones.
- Auronplay: Es uno de los youtubers más veteranos y populares de España. Se dedica a hacer críticas de otros vídeos, bromas telefónicas y gameplays de diversos juegos.
- Marta Díaz: Es una influencer y youtuber que se hizo conocida por ser la novia de TheGrefg, otro famoso youtuber. En su canal, sube vídeos de moda, belleza, retos y vlogs.
- DjMario: Es un youtuber especializado en el juego FIFA, donde muestra sus habilidades, trucos y consejos. También hace vídeos de otros deportes, como fútbol o baloncesto.
- Clakovi: Es una youtuber que se dedica a hacer vídeos de reacción a otros contenidos virales, como memes, tiktoks o canciones. Tiene un estilo divertido y sarcástico.

- El Rubius: Es el youtuber más seguido de España y uno de los más famosos del mundo. Su contenido es muy variado, desde gameplays de todo tipo de juegos, hasta sketches, animaciones o challenges.
- Emma Chamberlain: Es una youtuber estadounidense que se hizo famosa por su estilo de edición rápido y cómico. En su canal, sube vídeos de su vida cotidiana, sus gustos, sus viajes y sus opiniones.
- Malbert: Es un youtuber mexicano que se caracteriza por su humor absurdo y surrealista. En su canal, sube vídeos de parodias, sketches, reacciones y experimentos sociales.
- Mostopapi: Es un youtuber colombiano que se dedica a hacer vídeos de rap, freestyle y música urbana. También hace colaboraciones con otros artistas del género y reacciona a sus canciones.
- Orslok: Es un youtuber español que se enfoca en el contenido de terror, misterio y creepypastas. En su canal, sube vídeos de historias escalofriantes, análisis de casos paranormales y gameplays de juegos de terror.
- Vegetta: Es un youtuber español que se hizo famoso por sus gameplays de Minecraft. En su canal, sube vídeos de este juego y otros similares, como Roblox o GTA. También hace series con otros youtubers como Willyrex o Luzu.

La elección de esta lista de influencers, basada en el estudio previo de la Dra. Sue Aran-Ramspott, refleja claramente la preferencia identificada de diversión y entretenimiento por parte de la audiencia adolescente en YouTube.

La lista abarca una amplia gama de estilos de entretenimiento, desde el humor absurdo de Malbert hasta los escalofriantes relatos de terror de Orslok. Esta variedad asegura que se cubran múltiples facetas del entretenimiento que atraen a la audiencia adolescente. Además, muchos de los influencers seleccionados, como Ibai o El Rubius, han demostrado una habilidad especial para conectar con la generación más joven, ya sea a través de su estilo de comunicación, temáticas o simplemente su personalidad.

Por otro lado, influencers como Mostopapi y Marta Díaz reflejan y moldean tendencias culturales que son particularmente relevantes para los adolescentes, ya sea en el ámbito de la música urbana o la moda y belleza. La interacción directa que estos influencers tienen con su audiencia, ya sea a través de comentarios, encuestas o streams en vivo, fomenta un sentido de comunidad y pertenencia. Esto es especialmente atractivo para los adolescentes que buscan espacios donde sentirse escuchados y validados.

Adicionalmente, la lista incluye creadores que han experimentado con formatos y estilos innovadores, como la edición rápida y cómica de Emma Chamberlain o las parodias y sketches de Malbert. Estos formatos frescos y dinámicos resuenan con una audiencia joven ávida de contenido novedoso y diferente. Por último, influencers como DjMario o Vegetta ofrecen modelos de éxito y pasión por lo que hacen, inspirando a los jóvenes a seguir sus propias pasiones.

En resumen, esta selección de influencers representa adecuadamente las preferencias de diversión y entretenimiento de los jóvenes y ofrece una base sólida para un análisis más profundo de las dinámicas de influencia y consumo en YouTube.

Recopilación de datos desde YouTube API

La recopilación de datos es una piedra angular en cualquier proyecto de análisis. Cuando se trata de YouTube, una plataforma rica en contenido y metadatos, es imperativo tener un enfoque sistemático y detallado. La API de YouTube es una herramienta invaluable en este contexto. A continuación, se

desglosa con precisión cómo se ha estructurado la recopilación de datos de videos de YouTube de varios influencers utilizando la API:[31]

Configuración Inicial

a. Acceso a la API de YouTube:

- Google Cloud Console: Primero, es necesario dirigirse a la Google Cloud Console (<https://console.cloud.google.com/>) y crear un nuevo proyecto.
- Activación de la API: Dentro del proyecto, buscar y habilitar la API de YouTube Data v3.
- Generación de Credenciales: Posteriormente, en la sección de credenciales, optar por crear una clave API. Esta clave será esencial para autenticarse y hacer solicitudes a la API.

b. Preparación del Entorno de Desarrollo:

- Instalación de Librerías: Antes de escribir cualquier código, es necesario instalar las bibliotecas requeridas.

Estructura del Código y Funciones

a. Importaciones y Configuraciones Iniciales:

- Las bibliotecas necesarias se importan al inicio. Esto incluye herramientas para trabajar con archivos CSV, interactuar con la API de YouTube y obtener transcripciones de videos.

```
import csv
import os
from googleapiclient.discovery import build
from googleapiclient.errors import HttpError
from youtube_transcript_api import YouTubeTranscriptApi,TranscriptsDisabled,NoTranscriptFound
Captura de código 1 Librerías recopilación de datos
```

- La clave API generada anteriormente se configura como `API_KEY`.

b. Funciones Auxiliares:

1. Función: `get_channel_id_by_name(channel_name)`

- `get_channel_id_by_name`: Esta función toma un nombre de canal y devuelve su ID correspondiente. Es esencial porque la mayoría de las operaciones de la API requieren el ID del canal.

```
def get_channel_id_by_name(channel_name):
    response = youtube.search().list(
        part='snippet',
        q=channel_name,
        type='channel',
        maxResults=1
    ).execute()

    if response['items']:
        return response['items'][0]['snippet'][' channelId']
    else:
        return None
Captura de código 2 get channel id by name
```

- `youtube.search().list()`: Esta es una llamada a la API de YouTube que busca en la plataforma basándose en los criterios proporcionados.

Parámetros de la llamada a la API:

- `part='snippet'`: Indica qué partes de los datos del canal se quieren recuperar. "Snippet" es un conjunto de datos que incluye información básica del canal, como el título, la descripción y el ID.
- `q=channel_name`: Es el criterio de búsqueda. Aquí, se busca el nombre del canal proporcionado en el parámetro de la función.
- `type='channel'`: Especifica que se está buscando un canal. YouTube tiene diferentes tipos de contenidos, como videos, canales y listas de reproducción. Al establecer este parámetro en 'channel', se asegura de que la búsqueda se limite solo a canales.
- `maxResults=1`: Limita los resultados de la búsqueda al primer canal que coincide con el criterio. Esto es útil porque, en la mayoría de los casos, se quiere el canal más relevante basado en el nombre proporcionado.

Evaluación de la respuesta:

- `if response['items']:`: Verifica si la respuesta de la API contiene algún ítem. Si la búsqueda fue exitosa y encontró un canal que coincide con el nombre proporcionado, `response['items']` contendrá información sobre ese canal.
- `return response['items'][0]['snippet']['channelId']`: Si se encontró un canal, esta línea devuelve el ID del canal. Aquí, se accede al primer ítem (el canal más relevante) y se extrae el ID del canal del campo 'snippet'.
- `else: return None`: Si no se encontró ningún canal que coincida con el nombre proporcionado, la función devuelve 'None'.

En resumen, la función `get_channel_id_by_name` realiza una búsqueda en YouTube utilizando el nombre del canal proporcionado y devuelve el ID de ese canal si existe. Si no se encuentra ningún canal con ese nombre, devuelve 'None'.

2. Función: `get_channel_videos(channel_id, max_results=50, published_before=None)`

- `get_channel_videos`: Dado un ID de canal, esta función recopila una lista de videos asociados con ese canal.

```
def get_channel_videos(channel_id, max_results=50, published_before=None):
    videos = []
    page_token = None
    while True:
        response = youtube.search().list(
            part='id,snippet',
            channelId=channel_id,
            maxResults=max_results,
            type='video',
            order='date',
            publishedBefore=published_before,
            pageToken=page_token
        ).execute()
        videos.extend(response['items'])
        page_token = response.get('nextPageToken')
        if not page_token:
            break
    return videos
```

Captura de código 3 get channel videos

Parámetros:

- `channel_id`: Es el ID del canal de YouTube del que se desean obtener los videos.

- `max_results`=50: Especifica el número máximo de resultados que se desean obtener en una sola llamada a la API. Por defecto, está configurado para 50, que es el máximo permitido por la API de YouTube.
- `published_before=None`: Es una fecha y hora optionales en formato ISO 8601 (por ejemplo, '2023-04-01T00:00:00Z'). Si se proporciona, la función solo recuperará videos publicados antes de esta fecha y hora. En el contexto de este proyecto, se busca obtener datos de videos publicados antes del 1 de abril de 2023. Esto se hace para asegurar que los videos más recientes, que podrían no tener aún un volumen representativo de visualizaciones, no sesguen el análisis.

Dentro de la función:

Inicialización:

- `videos = []`: Una lista vacía que almacenará los videos recuperados.
- `page_token = None`: Un token que permite paginar a través de los resultados si hay más videos de los que se pueden recuperar en una sola llamada a la API.

Bucle `while True`: Este bucle se ejecuta hasta que se hayan recuperado todos los videos del canal o hasta que se alcance el límite especificado por `max_results`.

Llamada a la API: `youtube.search().list()`

Parámetros de la llamada a la API:

- `part='id,snippet'`: Indica qué partes de los datos del video se quieren recuperar.
- `channelId=channel_id`: Especifica el ID del canal del que se desean obtener los videos.
- `maxResults=max_results`: Limita el número de resultados por llamada a la API.
- `type='video'`: Indica que se está buscando videos.
- `order='date'`: Ordena los videos por fecha de publicación, con los más recientes primero.
- `publishedBefore=published_before`: Si se proporciona, limita la búsqueda a videos publicados antes de esta fecha y hora.
- `pageToken=page_token`: Si se proporciona, indica a la API que se desea la siguiente página de resultados.

Extensión de la lista de videos: `videos.extend(response['items'])`

Añade los videos recuperados en la llamada actual a la API a la lista total de videos.

Actualización del `page_token`: `page_token = response.get('nextPageToken')`

Si hay más páginas de resultados, la API devolverá un nextPageToken que se utiliza para obtener la siguiente página de resultados en la próxima iteración del bucle.

Condición de salida del bucle: `if not page_token: break`

Si no hay un `nextPageToken` en la respuesta, significa que se han recuperado todos los videos disponibles y se sale del bucle.

Retorno: `return videos`

Finalmente, la función devuelve la lista completa de videos recuperados.

En resumen, la función `get_channel_videos` utiliza la API de YouTube para recuperar una lista de videos de un canal específico, posiblemente limitando los resultados a videos publicados antes de una fecha y hora específicas.

3. Función: `get_video_details(video_id)`

`get_video_details`: Una vez que se tiene el ID de un video, esta función extrae detalles como título, descripción, estadísticas y transcripciones.

```

def get_video_details(video_id):
    response = youtube.videos().list(
        part='statistics,snippet',
        id=video_id
    ).execute()
    video = response['items'][0]
    languages = ['en', 'es']

    try:
        srt = YouTubeTranscriptApi.get_transcript(video_id, languages=languages)
        transcript_text = '\n'.join([entry['text'] for entry in srt])
    except (NoTranscriptFound, TranscriptsDisabled):
        transcript_text = 'Subtitle not available'

    return {
        'video_id': video_id,
        'title': video['snippet']['title'],
        'description': video['snippet']['description'],
        'publish_date': video['snippet']['publishedAt'],
        'likes': video['statistics'].get('likeCount', 'Not available'),
        'views': video['statistics'].get('viewCount', 'Not available'),
        'comments': video['statistics'].get('commentCount', 'Not available'),
        'transcript': transcript_text
    }
}

```

Captura de código 4 get video details

Parámetro:

- `video_id`: Es el ID único del video de YouTube del que se desean obtener los detalles.

Dentro de la función:

Llamada a la API: `youtube.videos().list()`

Parámetros de la llamada a la API:

- `part='statistics,snippet'`: Indica qué partes de los datos del video se quieren recuperar. En este caso, se solicitan las estadísticas del video y el fragmento (snippet) que contiene información básica como el título, la descripción y la fecha de publicación.
- `id=video_id`: Especifica el ID del video del que se desean obtener los detalles.

Extracción del video: `video = response['items'][0]`

Una vez que se ha obtenido la respuesta de la API, se extrae el video de la lista de items.

Definición de idiomas para la transcripción: `languages = ['en', 'es']`

Se define una lista de idiomas preferidos para la transcripción. En este caso, se buscan transcripciones en inglés ('en') y español ('es').

Intento de obtener la transcripción:

Se utiliza la API YouTubeTranscriptApi para intentar obtener la transcripción del video en los idiomas especificados.

- Si se encuentra una transcripción, se extrae el texto de la transcripción y se concatena en una única cadena.
- Si no se encuentra una transcripción o si los subtítulos están desactivados para el video, se establece un mensaje indicando que no está disponible.

Construcción del diccionario de detalles del video:

Se crea un diccionario que contiene:

- `video_id`: ID del video.
- `title`: Título del video.
- `description`: Descripción del video.
- `publish_date`: Fecha y hora de publicación del video.
- `likes`: Número de "me gusta" del video. Si no está disponible, se establece como "No disponible".
- `views`: Número de visualizaciones del video. Si no está disponible, se establece como "No disponible".
- `comments`: Número de comentarios del video. Si no está disponible, se establece como "No disponible".
- `transcript`: Texto de la transcripción del video o el mensaje indicando que no está disponible.

Retorno: `return {...}`

Finalmente, la función devuelve el diccionario con los detalles del video.

En resumen, la función `get_video_details` utiliza la API de YouTube y la API `YouTubeTranscriptApi` para recuperar información detallada y la transcripción de un video específico de YouTube utilizando su ID único.

c. Proceso ETL (Extract, Transform, Load):

```
def etl(channel_name, output_csv, published_before='2023-04-01T00:00:00Z'):
    channel_id = get_channel_id_by_name(channel_name)

    if channel_id:
        # Extract
        videos = get_channel_videos(channel_id, published_before=published_before)
        video_details = [get_video_details(video['id']['videoId']) for video in videos]

        # Transform
        # The data is already in a suitable format

        # Load
        with open(output_csv, 'w', newline='', encoding='utf-8') as csvfile:
            fieldnames = ['video_id', 'title', 'description', 'publish_date', 'likes', 'views', 'comments', 'transcript']
            writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
            writer.writeheader()
            for video in video_details:
                writer.writerow(video)
        print('Data successfully extracted and saved to {}'.format(output_csv))
    else:
        print(f'Channel not found for the name: {channel_name}')

```

Captura de código 5 Proceso ETL

Función: `etl(channel_name, output_csv, published_before='2023-04-01T00:00:00Z')`

Parámetros:

- `channel_name`: Nombre del canal de YouTube del que se desean extraer los datos.
- `output_csv`: Ruta y nombre del archivo CSV donde se guardarán los datos extraídos.
- `published_before`: Fecha límite opcional para filtrar videos. Solo se extraerán videos publicados antes de esta fecha.

Dentro de la función:

Obtención del ID del canal: `channel_id = get_channel_id_by_name(channel_name)`

Se utiliza la función auxiliar `get_channel_id_by_name` para obtener el ID del canal basándose en su nombre.

Verificación del ID del canal: `if channel_id`:

Se verifica si se ha obtenido un ID válido para el canal. Si no se encuentra un ID, se muestra un mensaje indicando que el canal no fue encontrado.

Extracción de datos:

- `videos = get_channel_videos(channel_id, published_before=published_before)` : Se utiliza la función auxiliar `get_channel_videos` para obtener una lista de videos del canal, filtrados por la fecha proporcionada.
- `video_details = [get_video_details(video['id']['videoId']) for video in videos]` : Para cada video en la lista, se utiliza la función auxiliar `get_video_details` para obtener detalles específicos del video.

Transformación de datos:

En este caso, se menciona que los datos ya están en un formato adecuado, por lo que no se realiza ninguna transformación adicional.

Carga de datos:

- Se abre (o crea) un archivo CSV en modo escritura.
- Se define una lista de nombres de campos que representan las columnas del archivo CSV.
- Se utiliza `csv.DictWriter` para escribir los datos en el archivo CSV en formato de diccionario.
- Se escribe la cabecera del archivo CSV.
- Se recorre la lista de detalles de videos y se escribe cada video en el archivo CSV.
- Finalmente, se muestra un mensaje indicando que los datos se extrajeron y guardaron con éxito.

En resumen, la función `etl` implementa un proceso ETL para extraer datos de videos de un canal de YouTube específico, transformarlos (en este caso no necesario) y cargarlos en un archivo CSV. Utiliza varias funciones auxiliares para realizar estas tareas y garantizar que los datos se manejen de manera eficiente y precisa.

d. Script Principal:

```
if __name__ == '__main__':
    influencers = [
        'Mostopapi',
        'Orslok',
        'Vegetta'
    ]

    for influencer in influencers:
        try:
            output_file = f'{influencer}_videos.csv'
            etl(influencer, output_file)
            print(f'Data for {influencer} successfully extracted and saved to {output_file}')
        except HttpError as e:
            print(f'An error occurred while processing {influencer}: {e}')
```

Captura de código 6 Main Script Recopilación de datos

`if __name__ == '__main__':`

Esta línea verifica si el script se está ejecutando como un programa principal. Si es así, el código dentro de este bloque se ejecutará. Esta es una práctica común en Python para permitir que un script se use tanto de forma independiente como importado como un módulo en otro script sin ejecutar automáticamente ciertas partes del código.

`influencers = ['Mostopapi', 'Orslok', 'Vegetta']`

Se define una lista llamada influencers que contiene los nombres de tres influencers (o canales de YouTube) para los cuales se desea extraer datos. En nuestro caso, teniendo en cuenta la limitación diaria de cuotas de YouTube API V3, se realizó la llamada en días separados.

Bucle `for`:

`for influencer in influencers`: Este bucle recorre cada influencer en la lista influencers.

Dentro del bucle:

`output_file = f'{influencer}_videos.csv'`

Se define el nombre del archivo de salida para cada influencer. Por ejemplo, para el influencer 'Mostopapi', el archivo de salida será 'Mostopapi_videos.csv'.

Bloque `try`:

Se utiliza para intentar ejecutar ciertas operaciones que podrían generar errores.

- **`etl(influencer, output_file)`:**
Se llama a la función `etl` (que hemos discutido anteriormente) para cada influencer, pasando el nombre del influencer y el nombre del archivo de salida como argumentos.
- **`print(f'Data for {influencer} successfully extracted and saved to {output_file}')`:**
Si la función `etl` se ejecuta con éxito, se muestra un mensaje indicando que los datos para ese influencer se trajeron y guardaron con éxito.

Bloque `except`:

`except HttpError as e:`

Si ocurre un error del tipo `HttpError` durante la ejecución del bloque `try`, se ejecutará este bloque. Específicamente, se mostrará un mensaje indicando que ocurrió un error mientras se procesaba el influencer y se proporcionará detalles del error.

Finalmente, este fragmento de código define una lista de influencers, luego intenta extraer y guardar datos para cada influencer usando la función `etl`. Si todo va bien, se muestra un mensaje de éxito; si ocurre un error, se muestra un mensaje de error.

Consideraciones Adicionales

- **Manejo de Excepciones:** Durante la recopilación de datos, es común encontrar errores, como videos sin transcripciones o cuotas de API excedidas. Es crucial manejar estas excepciones para garantizar que el proceso continúe sin interrupciones.
- **Cuotas de la API:** La API de YouTube establece restricciones en el número de peticiones permitidas dentro de un intervalo específico. Es esencial supervisar activamente este consumo para prevenir bloqueos o posibles costes adicionales. Para el desarrollo de nuestro proyecto, se ha optado por distribuir las solicitudes a lo largo de varios días, garantizando así la adecuada recolección de los datos necesarios en las fechas estipuladas.
- **Seguridad:** La clave API es sensible y no debe ser compartida ni expuesta públicamente. Es recomendable almacenarla en un archivo de configuración o utilizar variables de entorno.

En conclusión, la recopilación de datos de YouTube mediante su API es un proceso detallado que requiere una combinación de configuraciones previas, codificación estratégica y consideraciones prácticas. Al seguir este enfoque exhaustivo, se garantiza una recopilación de datos efectiva y eficiente, sentando una base sólida para cualquier análisis posterior.

Análisis Exploratorio de Datos

Para profundizar en el entendimiento de los datos obtenidos de YouTube, se llevará a cabo un análisis exploratorio de datos (EDA) de cada influencer de manera individual. Este proceso se estructura en varias etapas, que incluyen la carga de las librerías esenciales y de los datos, la detección de valores faltantes o atípicos, el cálculo de estadísticas descriptivas y la representación gráfica de la información.

Librerías necesarias para el EDA

Pandas:

- **Descripción:** Pandas es una librería de Python que proporciona estructuras de datos y herramientas de análisis de datos flexibles y eficientes. Es fundamental para la manipulación y el análisis de datos en Python.
- **Funcionalidades principales:**
 - Creación y manipulación de DataFrames y Series.
 - Lectura y escritura de datos en diversos formatos (CSV, Excel, SQL, entre otros).
 - Limpieza y preprocesamiento de datos (gestión de valores faltantes, filtrado, agrupación, etc.).
 - Realización de operaciones estadísticas básicas.

Matplotlib:

- **Descripción:** Matplotlib es una librería de visualización en Python que permite crear gráficos estáticos, animados e interactivos en múltiples plataformas.
- **Funcionalidades principales:**
 - Creación de una amplia variedad de gráficos, como líneas, barras, dispersión, histogramas, entre otros.
 - Personalización detallada de gráficos, incluyendo colores, leyendas, ejes y títulos.
 - Integración con Pandas para visualizar directamente DataFrames y Series.

Seaborn:

- **Descripción:** Seaborn es una librería de visualización basada en Matplotlib que proporciona una interfaz de alto nivel para la creación de gráficos estadísticos atractivos y descriptivos.
- **Funcionalidades principales:**
 - Ofrece una variedad de estilos y paletas de colores para mejorar la estética de los gráficos.
 - Facilita la creación de gráficos complejos, como mapas de calor, gráficos de violín y gráficos de parejas.
 - Integración con Pandas, permitiendo visualizar directamente DataFrames.
 - Proporciona herramientas para visualizar la estructura de datos y las relaciones entre variables.

Al combinar estas librerías, se puede llevar a cabo un EDA completo, desde la carga y limpieza de datos con Pandas, hasta la visualización detallada de las características y relaciones de los datos mediante Matplotlib y Seaborn.

Descripción de variables y creación de nuevas variables

A partir de los datos obtenidos mediante la API de YouTube, contamos con variables como 'video_id', 'title', 'description', 'publish_date', 'likes', 'views', 'comments' y 'transcript'. Para enriquecer nuestro análisis, es posible derivar nuevas variables de las ya existentes. Por ejemplo, de la variable 'publish_date', se pueden extraer 'year', 'month', 'day_of_week' y 'hour', lo que facilitaría un análisis de series temporales. Adicionalmente, se puede calcular una variable 'engagement_rate' al dividir la suma de 'likes' y 'comments' entre 'views', proporcionando una métrica de interacción del público con el contenido.

```
1 # Convert 'publish_date' to datetime format
2 df['publish_date'] = pd.to_datetime(df['publish_date'])
```

```
1 # Extract features from 'publish_date'
2 df['year'] = df['publish_date'].dt.year
3 df['month'] = df['publish_date'].dt.month
4 df['day_of_week'] = df['publish_date'].dt.dayofweek
5 df['hour'] = df['publish_date'].dt.hour
```

Captura de código 7 Creación de nuevas variables de fecha

```
# Create 'engagement rate'
df['engagement_rate'] = (df['likes'] + df['comments']) / df['views']
```

Captura de código 8 Creación de variable 'engagement_rate'

Visualización

- **Boxplot:** Este gráfico muestra la distribución de una variable numérica, identificando la mediana, cuartiles y posibles valores atípicos. Al aplicarlo a variables como 'likes', 'views' o 'comments' de los videos de YouTube, se puede identificar rápidamente qué videos son excepcionalmente populares o impopulares y detectar outliers que podrían ser el resultado de eventos virales o problemas con el video.
- **Histograma:** Representa la frecuencia de diferentes rangos de valores de una variable. Al visualizar la frecuencia de 'likes' o 'views' en YouTube, se puede obtener una imagen clara de la popularidad general de los videos de un influencer. Si la mayoría de los videos tienen un rango similar de 'likes', indica una consistencia en el engagement de la audiencia.
- **Scatterplot Matrix:** Muestra relaciones bivariadas entre múltiples variables. Al observar las relaciones entre variables como 'likes' y 'views', o 'comments' y 'views' en YouTube, se puede deducir, por ejemplo, que los videos con más vistas tienden a tener más likes, lo que indica una relación positiva entre la visibilidad y la aprobación del contenido.
- **Correlation Matrix:** Al mostrar el coeficiente de correlación entre pares de variables, se puede identificar qué variables están fuertemente relacionadas. Una alta correlación entre 'likes' y 'views' en YouTube sugiere que los videos más vistos también son los más apreciados, lo que puede influir en las decisiones sobre qué tipo de contenido producir en el futuro.
- **Gráfico de Serie Temporal:** Representa cómo una variable cambia con el tiempo. Al trazar 'views' o 'likes' en función del tiempo para un influencer de YouTube, se pueden identificar tendencias o patrones estacionales. Si un influencer experimenta un aumento en las vistas

durante las vacaciones, podría ser beneficioso planificar contenido especial durante esos períodos.

Estos gráficos, al proporcionar tanto información descriptiva como insights interpretativos, se convierten en herramientas poderosas para analizar y entender el comportamiento y las preferencias de la audiencia en YouTube.

Preprocesamiento de textos

El preprocesamiento de textos es una etapa esencial en cualquier proyecto de procesamiento de lenguaje natural. Esta fase implica la preparación y limpieza de datos textuales para su posterior análisis y modelado. A continuación, se describen las librerías mencionadas y su relevancia en el preprocesamiento de textos:

Librerías extras necesarias

Spacy:

- Spacy es una de las bibliotecas más populares para el procesamiento de lenguaje natural en Python. Ofrece modelos preentrenados para diversas tareas de NLP y es conocida por su eficiencia y precisión. Con Spacy, se pueden realizar tareas como tokenización, lematización, detección de entidades nombradas y etiquetado gramatical. Además, permite la identificación y eliminación de palabras vacías (stop words), que suelen ser irrelevantes para el análisis.

Emoji:

- La biblioteca Emoji en Python facilita el manejo y análisis de emojis en textos. Dado que los emojis son una forma común de expresión en plataformas digitales, es crucial poder identificar, extraer o incluso traducir estos símbolos a texto claro. La librería Emoji permite estas operaciones, ayudando a convertir emojis en palabras descriptivas o eliminarlos si no son relevantes para el análisis.

Re:

- Re es el módulo de expresiones regulares de Python. Permite buscar, coincidir y manipular cadenas de texto basadas en patrones específicos. Las expresiones regulares son esenciales para tareas como la eliminación de URLs, etiquetas HTML, caracteres especiales y otros patrones no deseados en los datos textuales. Con Re, se puede limpiar y estandarizar el texto de manera eficiente.

Collections.counter:

- Counter es una subclase de diccionario dentro del módulo Collections. Se utiliza para contar elementos en objetos iterables. En el contexto del preprocesamiento de textos, Counter puede ser útil para identificar y contar la frecuencia de palabras o tokens en un corpus. Esto puede ayudar a identificar términos comunes, palabras vacías o incluso términos raros que podrían ser candidatos para la eliminación.

Estructura de preprocesamiento

Antes de sumergirse en el preprocesamiento de textos, es esencial realizar una limpieza preliminar de los datos. Esta etapa inicial implica examinar los valores nulos y decidir cómo manejárlas, ya que pueden afectar la calidad y precisión del análisis posterior. En muchos conjuntos de datos, especialmente aquellos derivados de transcripciones automáticas como las de YouTube, es común encontrar casillas blancas representadas como '[__]'. Estas casillas suelen indicar palabras que han sido censuradas o filtradas, como insultos o términos ofensivos.

```
df = pd.read_csv('Auronplay_videos.csv')
df = df[df['transcript'] != 'Subtitle not available']
df['transcript'] = df['transcript'].str.replace('\xa0_\xa0', 'xxxx')
```

Captura de código 9 Pre-Limpieza de textos

Para garantizar la integridad del análisis, es crucial abordar estas casillas de manera adecuada. En este proyecto, se optó por reemplazar las casillas blancas en la variable 'transcript' con el término 'xxxx'. Esta decisión se tomó para evitar confusiones durante la tokenización y para asegurarse de que estas palabras no se eliminan o se confundan con otros términos durante el conteo de frecuencias.

Es importante destacar que, aunque se pueden implementar técnicas de preprocesamiento sin abordar estos problemas iniciales, es altamente recomendable revisar y corregir estos detalles antes de finalizar el proceso. Al hacerlo, se garantiza que los datos estén en el mejor estado posible para el análisis y modelado subsiguientes.

La función principal para el preprocesamiento de textos utiliza el modelo de lenguaje español de Spacy denominado 'es_core_news_sm'. Se eligió este modelo por su equilibrio entre velocidad y precisión. Al trabajar con grandes volúmenes de datos, el uso de un modelo más grande podría resultar en tiempos de procesamiento prolongados. Sin embargo, para el caso del inglés, específicamente para Emma Chamberlain, se optará por un modelo de tamaño medio con el objetivo de obtener una mayor precisión en el análisis.

```

# Load Spanish tokenizer, POS tagger, parser, NER and word vectors
nlp = spacy.load('es_core_news_sm')

def preprocess(text):
    # Lowercase the text
    text = text.lower()

    # Remove URLs
    text = re.sub(r'http\S+|www\S+|https\S+', '', text, flags=re.MULTILINE)

    # Remove emojis
    text = emoji.replace_emoji(text, replace='')

    # Remove punctuation
    text = re.sub(r'[\^w\s]', '', text)

    # Remove numbers
    text = re.sub(r'\d+', '', text)

    # Remove extra whitespace
    text = re.sub(r'\s+', ' ', text).strip()

    # Tokenize, remove stop words, and lemmatize using SpaCy's Spanish model
    doc = nlp(text)
    tokens = [token.lemma_ for token in doc if not token.is_stop and token.is_alpha]

    return tokens

# Apply preprocessing to each column
df['title'] = df['title'].apply(preprocess)
df['description'] = df['description'].apply(preprocess)
df['transcript'] = df['transcript'].apply(preprocess)

```

Captura de código 10 Función preprocess

Función preprocess:

- Convertir a minúsculas: Se convierte todo el texto a minúsculas para mantener la uniformidad y evitar la duplicidad basada en el caso.
- Eliminar URLs: Se utilizan expresiones regulares para identificar y eliminar cualquier URL presente en el texto.
- Eliminar emojis: Se utiliza la biblioteca `emoji` para identificar y eliminar cualquier emoji del texto.
- Eliminar puntuación: Se eliminan todos los signos de puntuación del texto para simplificar el análisis y centrarse solo en las palabras.
- Eliminar números: Se eliminan todos los números del texto, ya que, en este contexto, no aportan información relevante.
- Eliminar espacios en blanco adicionales: Se eliminan espacios en blanco adicionales y se asegura que el texto no tenga espacios al principio o al final.
- Tokenización, eliminación de palabras vacías y lematización:
 - o Se tokeniza el texto, es decir, se divide en palabras individuales o "tokens".
 - o Se eliminan las palabras vacías (stop words), que son palabras comunes que generalmente no aportan significado al texto (por ejemplo, "y", "el", "la").
 - o Se realiza la lematización, que implica convertir cada palabra a su forma base o raíz. Por ejemplo, "corriendo" se convierte en "correr".

- Resultado: La función devuelve una lista de tokens procesados que están listos para análisis posteriores o para ser utilizados en modelos de aprendizaje automático.

Al terminar el proceso de preprocessamiento, en un archivo en formato CSV llamado 'preprocessed_data.csv' sin incluir el índice del DataFrame para análisis posteriores.

Minería de Textos

Con los datos ya preprocessados, nos adentramos en la fase de minería de textos para descifrar las características y patrones lingüísticos empleados por los youtubers. Esta etapa nos permite desentrañar cómo estos creadores de contenido se comunican y qué elementos lingüísticos utilizan para conectar con su audiencia. Mediante técnicas como el análisis de frecuencia de palabras, visualizaciones con nubes de palabras y el estudio de n-gramas, se busca identificar las palabras y frases más recurrentes en sus discursos. El topic modeling nos permite descubrir los temas predominantes en su contenido, mientras que el TF-IDF nos ayuda a resaltar aquellos términos que son únicos y distintivos para cada youtuber. Finalmente, mediante el Reconocimiento de Entidades Nombradas (NER), nos centramos en identificar y categorizar las personas mencionadas, proporcionando una perspectiva más clara sobre con quiénes interactúan o a quiénes hacen referencia en sus videos. En conjunto, estas técnicas nos ofrecen una visión holística y detallada de la narrativa y el estilo lingüístico de los youtubers, permitiéndonos entender mejor su impacto y resonancia en la audiencia.

Librerías Extras

ast:

La biblioteca ‘**ast**’ (Abstract Syntax Trees) permite a los programadores de Python procesar árboles de sintaxis abstracta de código fuente Python. Una de las funciones más utilizadas de esta biblioteca es ‘`literal_eval`’.

from nltk import ngrams:

la función **ngrams** de **nltk**. Los n-gramas son secuencias contiguas de n palabras de un texto dado. Por ejemplo, para el texto "Ibai es genial", los bigramas (2-gramas) serían: ["Ibai es", "es genial"]. Esta función es útil para analizar la estructura y las secuencias de palabras en un texto, lo que puede ser valioso para diversas tareas, como la construcción de modelos de lenguaje o la identificación de frases comunes.

from gensim.corpora import Dictionary:

- ‘gensim’ es una biblioteca especializada en modelado de tópicos y procesamiento de vectores de palabras (como Word2Vec).
- ‘Dictionary’ es una clase de ‘gensim’ que permite crear un diccionario de palabras únicas a partir de un conjunto de textos. Este diccionario se utiliza a menudo para convertir palabras en identificadores numéricos, lo que facilita la construcción de representaciones vectoriales de documentos.

from gensim.models import LdaMulticore:

- ‘LdaMulticore’ es una implementación paralelizada de LDA (Latent Dirichlet Allocation) en gensim. LDA es un modelo de tópicos que se utiliza para descubrir los tópicos latentes presentes en un conjunto de documentos. La versión multicore aprovecha múltiples núcleos de CPU para acelerar el entrenamiento.

from sklearn.feature_extraction.text import TfidfVectorizer:

- ‘scikit-learn’ (sklearn) es una biblioteca de aprendizaje automático en Python que proporciona herramientas simples y eficientes para análisis de datos.

- ‘TfidfVectorizer’ es una clase de ‘sklearn’ que convierte una colección de documentos en bruto en una matriz de características TF-IDF. TF-IDF (Term Frequency-Inverse Document Frequency) es una medida que evalúa la importancia de una palabra en un documento en relación con un conjunto de documentos. Es una técnica comúnmente utilizada para la extracción de características en tareas de procesamiento de texto.

from wordcloud import WordCloud:

- La biblioteca **wordcloud** es una herramienta popular en Python diseñada para visualizar palabras en forma de nubes de palabras. Una nube de palabras es una representación gráfica donde las palabras se muestran en diferentes tamaños según su frecuencia o importancia en un conjunto de texto. Las palabras que aparecen con más frecuencia se muestran en un tamaño más grande, mientras que las palabras menos frecuentes se muestran en un tamaño más pequeño. Esta representación visual permite identificar rápidamente las palabras o términos más destacados en un conjunto de datos.

from flair.models import SequenceTagger

from flair.data import Sentence:

- ‘flair’ es una biblioteca de NLP basada en PyTorch que se destaca por ofrecer incrustaciones contextuales avanzadas. A diferencia de las representaciones estáticas tradicionales, las incrustaciones de ‘flair’ varían según el contexto en el que aparece una palabra. La biblioteca proporciona modelos preentrenados para tareas como la identificación de entidades nombradas y el etiquetado de partes del discurso, y es conocida por su flexibilidad y soporte multilingüe. En esencia, ‘flair’ combina facilidad de uso con potentes capacidades de aprendizaje profundo, siendo una herramienta esencial para proyectos modernos de NLP.

Frecuencia de palabras

Re-Tokenización

En este fragmento de código, se construye una lista denominada ‘all_tokens’, la cual almacena todos los tokens (palabras individuales) encontrados en transcripciones del DataFrame ‘df’. Esta lista se compila utilizando una técnica de comprensión de lista anidada, que itera sobre cada transcripción y, posteriormente, sobre cada token dentro de esa transcripción.

Sin embargo, al recuperar textos previamente tokenizados desde un archivo y cargarlos en un DataFrame, estos tokens se interpretan como cadenas de texto simples en lugar de estructuras de datos de Python. Para revertir estos textos a su forma original de estructuras de datos (como listas o diccionarios), es necesario emplear la función ‘literal_eval’ de la biblioteca ‘ast’. Al aplicar esta función a la columna ‘transcript’, se convierten las representaciones en cadena de las estructuras de datos de Python de nuevo a sus respectivas formas originales como objetos Python.

```

import ast

processed_transcripts = df['transcript'].apply(ast.literal_eval)
all_tokens = [token for transcript in processed_transcripts for token in transcript]

# Count the occurrences of each token
word_freq = Counter(all_tokens)

# Display the 15 most common words and their frequencies
most_common_words = word_freq.most_common(15)
print(most_common_words)

# Output in a DataFrame
df_word_freq = pd.DataFrame(most_common_words, columns=['Word', 'Frequency'])
print(df_word_freq)

```

Captura de código 11 Frecuencia de palabras

Conteo de Frecuencias

Se utiliza la clase ‘Counter’ del módulo ‘collections’ para contar la frecuencia de cada token en la lista ‘all_tokens’. El resultado es un diccionario donde las claves son los tokens y los valores son sus respectivas frecuencias.

Mostrar las Palabras Más Comunes

Con el método ‘most_common()’, se obtienen las 15 palabras más frecuentes junto con sus respectivas frecuencias. Luego, se imprimen estos resultados.

Conversión a DataFrame

Para una visualización más estructurada, se convierten las 15 palabras más comunes y sus frecuencias en un DataFrame de pandas. Las columnas del DataFrame se etiquetan como ‘Word’ y ‘Frequency’. Finalmente, se imprime este DataFrame.

Nube de palabras

En el código abajo, se crea una visualización conocida como "nube de palabras" a partir de transcripciones procesadas. Inicialmente, se consolidan las listas de tokens de cada transcripción (ya sea títulos, descripciones, etc.) en una cadena continua denominada ‘all_text’. Posteriormente, con la ayuda de la biblioteca ‘WordCloud’, se genera la nube de palabras, definiendo ciertos parámetros como el color de fondo, las dimensiones y el esquema de color, y asegurando que no haya repetición de combinaciones de palabras. Una vez creada la nube, se visualiza usando ‘matplotlib’, donde las palabras que aparecen con mayor frecuencia se muestran en un tamaño más grande, destacando su relevancia. Al finalizar, esta visualización se almacena en un archivo de imagen con el nombre “transcript_wordcloud.png” y se presenta al usuario. Esta representación gráfica permite una rápida

comprensión de los términos predominantes en las transcripciones, facilitando la identificación de temas o palabras clave en el contenido analizado.

```
# Convert the list of tokens back into a single string for each transcript
all_text = ' '.join([' '.join(transcript) for transcript in processed_transcripts])

# Generate the word cloud.
wordcloud = WordCloud(background_color="white", # Set the background color to white
                      width=600,           # Width of the canvas
                      height=400,          # Height of the canvas
                      collocations=False, # Avoid duplicate words
                      colormap='viridis'  # Color scheme
                     ).generate(all_text)

# Plot the word cloud.
plt.figure(figsize=(10, 8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")

# Save the visualization to a file.
plt.savefig("transcript_wordcloud.png")

plt.show()
```

Captura de código 12 Nube de palabras

N- grams

Este permite identificar y visualizar las frases de tres palabras que se repiten con mayor frecuencia en las transcripciones procesadas, ofreciendo insights sobre los temas o patrones lingüísticos más comunes en el contenido analizado.

Proceso:

- Se crea una lista ‘all_ngrams’ que contendrá todos los trigramas generados a partir de las transcripciones procesadas. La comprensión de lista anidada recorre cada transcripción y, para cada una, genera trigramas utilizando la función ‘ngrams’ de la biblioteca ‘nltk’.
- Luego, se utiliza la clase ‘Counter’ para contar la frecuencia de cada trígrama en ‘all_ngrams’.
- La función ‘most_common(20)’ se utiliza para obtener los 20 trigramas más comunes junto con sus frecuencias.

Salida:

- Finalmente, se imprime la lista de las 20 frases (trigramas) más comunes junto con su frecuencia de aparición. Esto proporciona una visión rápida de las combinaciones de tres palabras que aparecen con mayor frecuencia en las transcripciones, lo que puede ser útil para identificar temas recurrentes o frases clave en el contenido.

```
# 1. Identify commonly used phrases using N-grams
all_ngrams = [Ingram for transcript in processed_transcripts for ngram in ngrams(transcript, 3)] # trigrams
common_phrases = Counter(all_ngrams).most_common(20)
print("Common Phrases:")
for phrase, freq in common_phrases:
    print(f'{phrase}: {freq}'")
```

Captura de código 13 N-grams

Topic Modeling

Se utiliza el modelo LDA para identificar y visualizar los principales temas presentes en las transcripciones procesadas. Estos temas proporcionan una visión general de los tópicos o áreas de discusión más recurrentes en el contenido analizado.

Preparación de datos:

- ‘Dictionary(processed_transcripts)’: Se crea un diccionario a partir de las transcripciones procesadas. Este diccionario mapea cada palabra única con un identificador numérico.
- ‘doc2bow(transcript)’: Convierte cada transcripción en una representación de "bolsa de palabras" (BoW). Esta representación cuenta la frecuencia de cada palabra en el documento.

Modelo LDA:

- Se inicializa y entrena un modelo LDA utilizando ‘LdaMulticore’. Los parámetros clave son:
- ‘corpus’: El corpus en formato BoW.
- ‘id2word’: El diccionario que mapea los identificadores con las palabras.
- ‘num_topics’: El número de temas que se desea identificar. En este caso, se han elegido 5 temas.
- ‘passes’: El número de veces que el modelo recorre todo el corpus durante el entrenamiento. Un mayor número de pasadas puede mejorar la calidad del modelo a expensas del tiempo de entrenamiento.
- ‘workers’: El número de núcleos de CPU que se utilizarán para el entrenamiento. Esto acelera el proceso de entrenamiento.

Salida:

- Una vez entrenado el modelo, se imprimen los temas identificados con ‘print_topics(-1)’. Cada tema se representa como una combinación de palabras junto con su relevancia en ese tema.

```
# 2. Identify topics using LDA
dictionary = Dictionary(processed_transcripts)
corpus = [dictionary.doc2bow(transcript) for transcript in processed_transcripts]
lda_model = LdaMulticore(corpus, id2word=dictionary, num_topics=5, passes=15, workers=4)
print("\nTopics:")
for idx, topic in lda_model.print_topics(-1):
    print(f"Topic: {idx}, Words: {topic}")
```

Captura de código 14 Topic modeling

Identificación términos únicos mediante TF-IDF

TF-IDF Vectorizer:

- Se inicializa un ‘TfidfVectorizer’ con un parámetro ‘max_df’ de 0.85. Esto significa que cualquier término que aparezca en más del 85% de los documentos (transcripciones) será ignorado, ya que se considera demasiado común y, por lo tanto, menos distintivo.
- Se ajusta y transforma el vectorizador utilizando las transcripciones procesadas. Cada transcripción, que es una lista de tokens, se convierte en una única cadena de texto para esta operación.

Extracción de términos únicos:

- Una vez obtenida la matriz TF-IDF, se suman los valores TF-IDF de cada término a lo largo de todos los documentos para obtener una puntuación acumulada para cada término.
- Se ordenan estos términos basados en sus puntuaciones acumuladas en orden descendente.
- Se extraen los nombres de las características (términos) utilizando ‘get_feature_names_out()’.

Salida:

- Se imprimen los 10 términos más distintivos (con las puntuaciones TF-IDF más altas) junto con sus puntuaciones acumuladas.

```
# 3. Identify unique terms/values using TF-IDF
tfidf_vectorizer = TfidfVectorizer(max_df=0.85)
tfidf_matrix = tfidf_vectorizer.fit_transform([' '.join(transcript) for transcript in processed_transcripts])
feature_names = tfidf_vectorizer.get_feature_names_out()
sorted_items = sorted(list(enumerate(tfidf_matrix.sum(axis=0).tolist()[0])), key=lambda x: x[1], reverse=True)
print("\nUnique Terms:")
for idx, score in sorted_items[:10]:
    print(f"{feature_names[idx]}: {score}")
```

Captura de código 15 TF-IDF

En esencia, este código identifica y muestra los términos que son más distintivos o únicos en las transcripciones procesadas. Estos términos, al tener altas puntuaciones TF-IDF, son aquellos que son particularmente relevantes o característicos en el conjunto de datos, pero no necesariamente los más frecuentes.

NER

Este fragmento se centra en la identificación y recuento de entidades nombradas, específicamente nombres de personas, dentro de un conjunto de datos de videos de YouTube. A continuación, se desglosa el proceso:

- Carga de Datos: Se carga un archivo CSV que contiene datos de videos de un YouTuber llamado 'Auronplay'.
- Combinación de Textos: Se combinan las columnas 'title', 'description' y 'transcript' en una única lista llamada `texts` para procesar todo el contenido textual de manera conjunta.
- Modelo NER de Flair: Se carga un modelo preentrenado de Flair para la identificación de entidades nombradas en español. Este modelo se especializa en detectar nombres de personas.
- Lista de Influencers: Se crea una lista de nombres de influencers para detectar manualmente en el texto, ya que algunos nombres pueden no ser reconocidos por el modelo NER.
- Extracción de Entidades: Se recorre cada texto y se utiliza el modelo NER para identificar y extraer nombres de personas. Además, se realiza una búsqueda manual de los nombres de influencers en la lista para garantizar que no se pasen por alto.
- Conteo de Entidades: Se utiliza la clase 'Counter' para contar la frecuencia de cada entidad nombrada detectada.
- Exportación a CSV: Se crea un nuevo DataFrame que contiene las entidades detectadas y sus frecuencias. Luego, este DataFrame se exporta a un archivo CSV llamado 'per_data.csv' para que luego poder hacer análisis de redes sociales (SNA).

```

# Load the preprocessed data
df = pd.read_csv('Auronplay_videos.csv')

# Combine all text columns into a single list
texts = df['title'].astype(str).tolist() + df['description'].astype(str).tolist() + df['transcript'].astype(str).tolist()

# Load the Flair NER model for Spanish
tagger = SequenceTagger.load("flair/ner-spanish-large")

# List of names to recognize persons
influencer_names = [
    'ibai',
    'auronplay',
    'auron',
    'marta diáz',
    'marta',
    'djmario',
    'djmario',
    'clakovi',
    'elrubius',
    'rubius',
    'emma chamberlain',
    'emma',
    'chamberlain',
    'malbert',
    'malbet',
    'mostopapi',
    'orslok',
    'vegetta',
    'vegeta'
]

%%time
# Perform NER on text data and extract named entities
person_entities = []
for text in texts:
    sentence = Sentence(text)
    tagger.predict(sentence)
    for entity in sentence.get_spans('ner'):
        if entity.labels[0].value == 'PER':
            person_entities.append(entity.text)

    # Adding custom names detection
    for name in influencer_names:
        if name.lower() in text.lower():
            person_entities.append(name)

# Count the frequency of person entities
entity_frequencies = Counter(person_entities)

# Create a DataFrame and export to CSV
df_export = pd.DataFrame({
    'Source': ['YouTuber'] * len(entity_frequencies),
    'Target': list(entity_frequencies.keys()),
    'Weight': list(entity_frequencies.values())
})

df_export.to_csv('per_data.csv', index=False)

```

Captura de código 16 NER

Revisión y Mapping de NER

En el proceso posterior a la identificación de entidades de personas, es esencial realizar una revisión y limpieza de los datos. Al examinar el conjunto de datos, es probable encontrar variaciones en la representación de nombres, así como entidades que no corresponden a personas. Para abordar las variaciones en los nombres, se recomienda convertir todos los nombres a minúsculas y luego utilizar

un mapeo para consolidar diferentes variantes de un nombre en una forma estándar. Por ejemplo, las variaciones "AuronPlay" y "auron" podrían mapearse a "auronplay". Una vez realizado este mapeo, se pueden sumar las frecuencias de las variantes para obtener un conteo consolidado. En cuanto a las entidades que no son personas, aunque el modelo NER tiene una precisión relativamente alta, siempre hay margen de error. Sin embargo, si estas instancias incorrectas son mínimas, es probable que no afecten significativamente los resultados finales ni el análisis de redes sociales (SNA) que se pueda realizar posteriormente. Por lo tanto, es crucial abordar estos problemas para garantizar la calidad y precisión del análisis.

```

df = pd.read_csv('per_data.csv')
df['Target'] = df['Target'].str.lower()

# Group all variations of a name into a standard form
name_mapping = {
    'auron': 'auronplay',
    'auro': 'auronplay',
    'djmario': 'djmario',
    'dj mario': 'djmario',
    'mario': 'djmario',
    'el rubius': 'elrubis',
    'rubius': 'elrubius',
    'vegeta': 'vegetta',
    'josé josé': 'josé',
    'dani': 'daniel',
    'danny': 'daniel',
    'juanito': 'juan',
    'balvin': 'j balvin',
    'juanito juanito': 'juan',
    'juan juan': 'juan',
    'pablito': 'pablo',
    'malbet': 'malbert'
}

df['Target'] = df['Target'].replace(name_mapping)

# Sum the weights for each standard form
final_df = df.groupby(['Source', 'Target']).sum().reset_index()

final_df.to_csv('refined_per_data.csv', index=False)

```

Captura de código 17 Mapping NER

Social Network Analysis

Utilizando los nombres correctamente exportados, se implementará un Análisis de Redes Sociales (SNA) para cada youtuber con el objetivo de descifrar las interacciones y relaciones entre los influencers y las personas o entidades que mencionan. Adicionalmente, se construirá una red que conecte a todos los influencers de la lista para visualizar las posibles conexiones y dinámicas entre ellos.

Librerías Extras

from pyvis.network import Network

- La librería ‘pyvis’ es una biblioteca de Python diseñada para la visualización interactiva de redes. Específicamente, el módulo Network de ‘pyvis’ permite crear visualizaciones de redes que son interactivas, lo que significa que puedes hacer zoom, arrastrar nodos, ver etiquetas y otras interacciones directamente en la visualización.
- La ventaja de usar ‘pyvis’ es que combina la potencia de ‘networkx’ (una biblioteca popular de Python para la creación, manipulación y estudio de la estructura, dinámica y funciones de redes complejas) con la interactividad de ‘vis.js’ (una biblioteca de JavaScript para visualizar datos en forma de gráficos de red).

```
import numpy as np
```

- NumPy es una biblioteca fundamental para la computación científica en Python. Proporciona soporte para trabajar con arrays (vectores y matrices) de gran tamaño y ofrece una amplia colección de funciones matemáticas para operar con estos arrays.

Estructuras de Código

El código proporcionado realizará la visualización de una red de relaciones entre diferentes nodos, en este caso, posiblemente youtubers o influencers y sus conexiones.

- Importación de bibliotecas: Se importan las bibliotecas necesarias: ‘Network’ de `pyvis` para la visualización de redes y `numpy` para operaciones matemáticas.
- Carga de datos: Se carga un archivo CSV llamado 'refined_per_data.csv' en un DataFrame de pandas.
- Filtrado de datos: Se filtran las conexiones (o aristas) con un peso menor o igual a 10 (o cuanto sea según el volumen del total de mencionados), pero se conservan aquellas que involucran a ciertos influencers específicos, independientemente de su peso.
- Determinación de percentiles: Se calculan los percentiles 10 y 90 de los pesos de las conexiones para normalizar el ancho de las aristas en la visualización.
- Cálculo de pesos acumulativos: Se suman los pesos de todas las conexiones que involucran a un nodo específico para determinar su "importancia" o "peso" (medido por frecuencia) en la red.
- Normalización de tamaños de nodos: Se ajustan los tamaños de los nodos según su peso acumulativo para que los nodos más "importantes" sean más grandes en la visualización.
- Creación de la red: Se inicializa una red utilizando `pyvis` .
- Adición de nodos: Se agregan todos los nodos únicos a la red con tamaños basados en sus pesos acumulativos.
- Adición de aristas: Se agregan las conexiones entre los nodos a la red con anchos basados en sus pesos individuales. Estos anchos se normalizan utilizando los percentiles 10 y 90 calculados anteriormente.
- Visualización: Finalmente, se muestra la red en un archivo HTML llamado "network.html"

Análisis de los Resultados

En este segmento, profundizaremos en los hallazgos derivados del EDA), la minería de textos y el SNA. Examinaremos detalladamente a cada influencer de nuestra lista preestablecida en YouTube, desentrañando patrones individuales y características distintivas. Posteriormente, contrastaremos estos hallazgos para identificar similitudes y diferencias entre ellos. Finalizaremos con un análisis de las interconexiones y relaciones que existen entre estos influencers, proporcionando una visión holística de su dinámica en la plataforma.

Ibai

EDA

Empezaremos el análisis observando los datos del youtuber Ibai. Para este youtuber, la API de Youtube ha extraído 371 vídeos antes de la fecha predicha, más relevantes según los parámetros fijados en el código.

	likes	views	comments	year	month	\
count	3.710000e+02	3.710000e+02	371.000000	371.000000	371.000000	
mean	2.028117e+05	3.696588e+06	2996.617251	2021.854447	5.978437	
std	2.020211e+05	3.415008e+06	4363.101852	0.955977	3.656844	
min	4.690000e+03	8.154300e+04	70.000000	2019.000000	1.000000	
25%	7.957400e+04	1.423352e+06	864.000000	2021.000000	3.000000	
50%	1.417550e+05	2.585160e+06	1765.000000	2022.000000	5.000000	
75%	2.610485e+05	4.970159e+06	3532.500000	2023.000000	9.000000	
max	2.337261e+06	2.677771e+07	46865.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	371.000000	371.000000	371.000000			
mean	3.126685	18.547170	0.057288			
std	1.951469	4.441793	0.013393			
min	0.000000	0.000000	0.023467			
25%	2.000000	18.000000	0.048693			
50%	3.000000	20.000000	0.055932			
75%	5.000000	21.000000	0.064845			
max	6.000000	23.000000	0.113055			

Figura 1 Estadística Ibai

Una vez tenemos los datos bien cargados, empezamos por una descripción estadística.

Como se puede observar en la figura arriba:

- **Volumen de Datos:** Hay un total de 371 registros, lo que indica que se han analizado 371 videos de Ibai.
- **Interacción:**
 - o **Likes:** En promedio, los videos de Ibai reciben aproximadamente 202,812 likes, con un video que ha alcanzado un máximo de 2,337,261 likes y un mínimo de 4,690 likes.
 - o **Vistas:** Los videos tienen un promedio de 3,696,588 vistas, con un video alcanzando hasta 26,777,710 vistas y otro tan bajo como 81,543 vistas.
 - o **Comentarios:** En promedio, los videos tienen 2,996 comentarios, con un video que ha alcanzado un máximo de 46,865 comentarios y un mínimo de 70 comentarios.
- **Tiempo:**
 - o **Año:** Los videos se extienden desde 2019 hasta 2023.
 - o **Mes:** El mes promedio en el que se publican los videos es alrededor de junio (5.978).
 - o **Día de la semana:** En promedio, los videos tienden a publicarse a mediados de semana, alrededor del jueves (3.126).
 - o **Hora:** La mayoría de los videos se publican alrededor de las 6:30 p.m. (18.547 horas).
- **Tasa de Engagement :** La tasa de engagement promedio es del 5.73%, lo que indica que, en promedio, el 5.73% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 11.31%, mientras que la más baja es del 2.35%.
- **Distribución de Datos:** Las desviaciones estándar (std) para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos.

Esto sugiere que, aunque Ibai tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

En resumen, Ibai tiene una fuerte presencia en YouTube con una alta interacción en sus videos. Sin embargo, hay una variabilidad en la interacción, lo que sugiere que algunos videos son más populares que otros. La mayoría de los videos se publican en la tarde o temprano en la noche, y la tasa de engagement promedio es saludable, lo que indica una audiencia activa y comprometida.

Ahora vamos a observar algunas visualizaciones que probablemente nos ofrecerán perspectivas distintas.

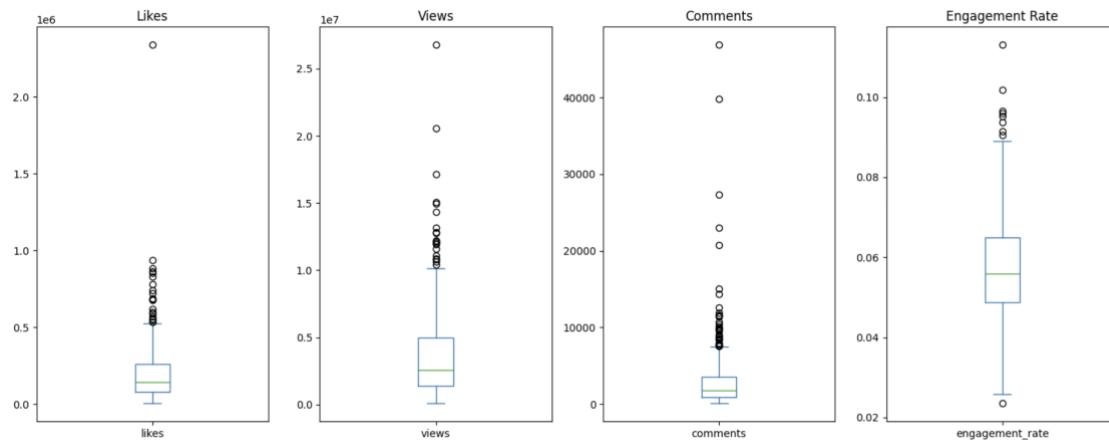


Figura 2 Boxplot Ibai

Al examinar los boxplots (referencia anterior) relacionados con Likes, Views, Comments y Engagement Rate de Ibai, es evidente que, si bien mantiene una interacción constante en la mayoría de sus videos, existen ciertas publicaciones que sobresalen y capturan una atención excepcionalmente alta. A continuación, se detallan algunas observaciones clave:

- **Likes:** La mediana de Ibai se sitúa en torno a los 142k likes por video. No obstante, hay contenidos que logran capturar más de 2.3 millones de likes, evidenciando la existencia de videos que resuenan de manera especial con su audiencia.
- **Views:** Con una mediana establecida en aproximadamente 2.6 millones de vistas, hay videos que rompen esta norma y llegan a impresionantes 26.8 millones de reproducciones. Esto resalta la capacidad de Ibai para crear contenidos virales en ocasiones.
- **Comments:** A pesar de que la mediana se sitúa cerca de los 1.8k comentarios, existen videos que provocan hasta 46.9k interacciones en esta categoría, reflejando la diversidad en la respuesta del público.
- **Engagement Rate:** La tasa de interacción promedio se encuentra en el 5.6%. Sin embargo, hay videos que logran una tasa del 11.3%, indicando que ciertos temas o formatos logran conectar de manera más profunda con los espectadores.

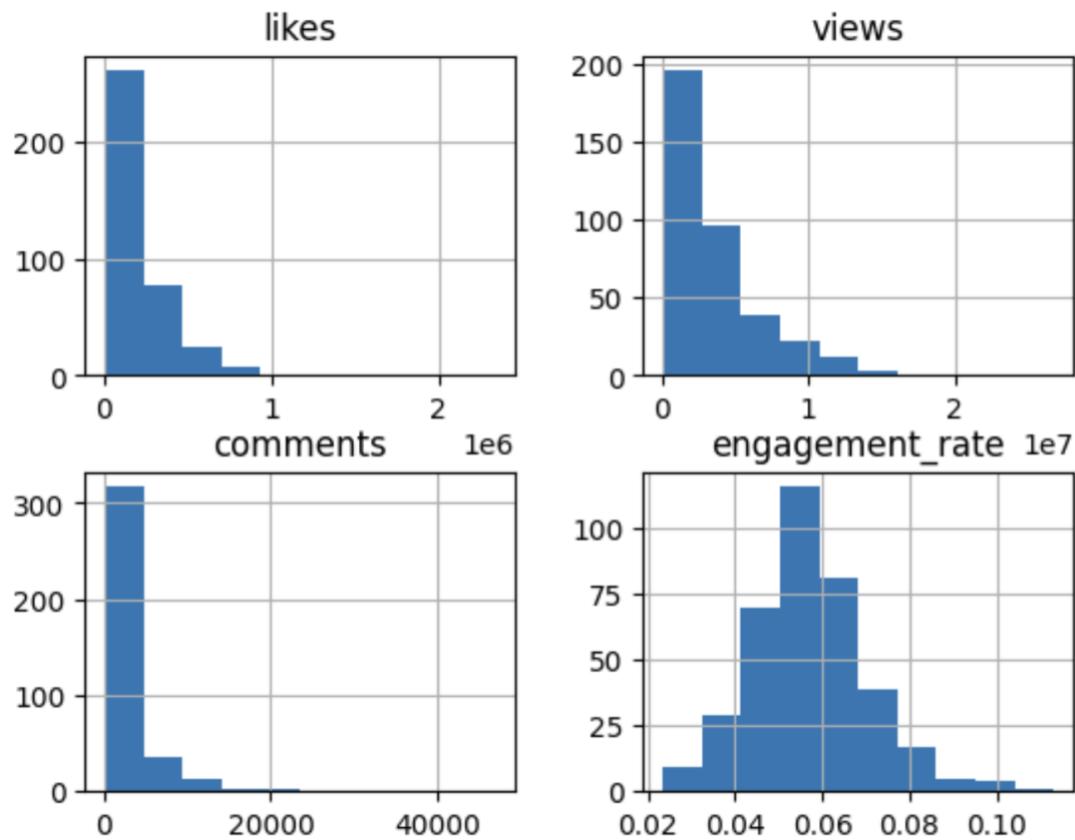


Figura 3 Histograma Ibai

Entre los gráficos (véase arriba), observamos que la mayoría de los videos de Ibai tienden a tener un número moderado de "likes", "views" y "comments", con la mayoría concentrada en el extremo izquierdo, indicando valores más bajos. Sin embargo, la presencia de colas hacia la derecha en estas distribuciones sugiere que hay ciertos videos que se desempeñan excepcionalmente bien en términos de interacción. En cuanto a la tasa de engagement, la variabilidad en su distribución sugiere que, si bien muchos videos tienen una tasa de engagement promedio, hay algunos que logran captar una interacción significativamente mayor de la audiencia.

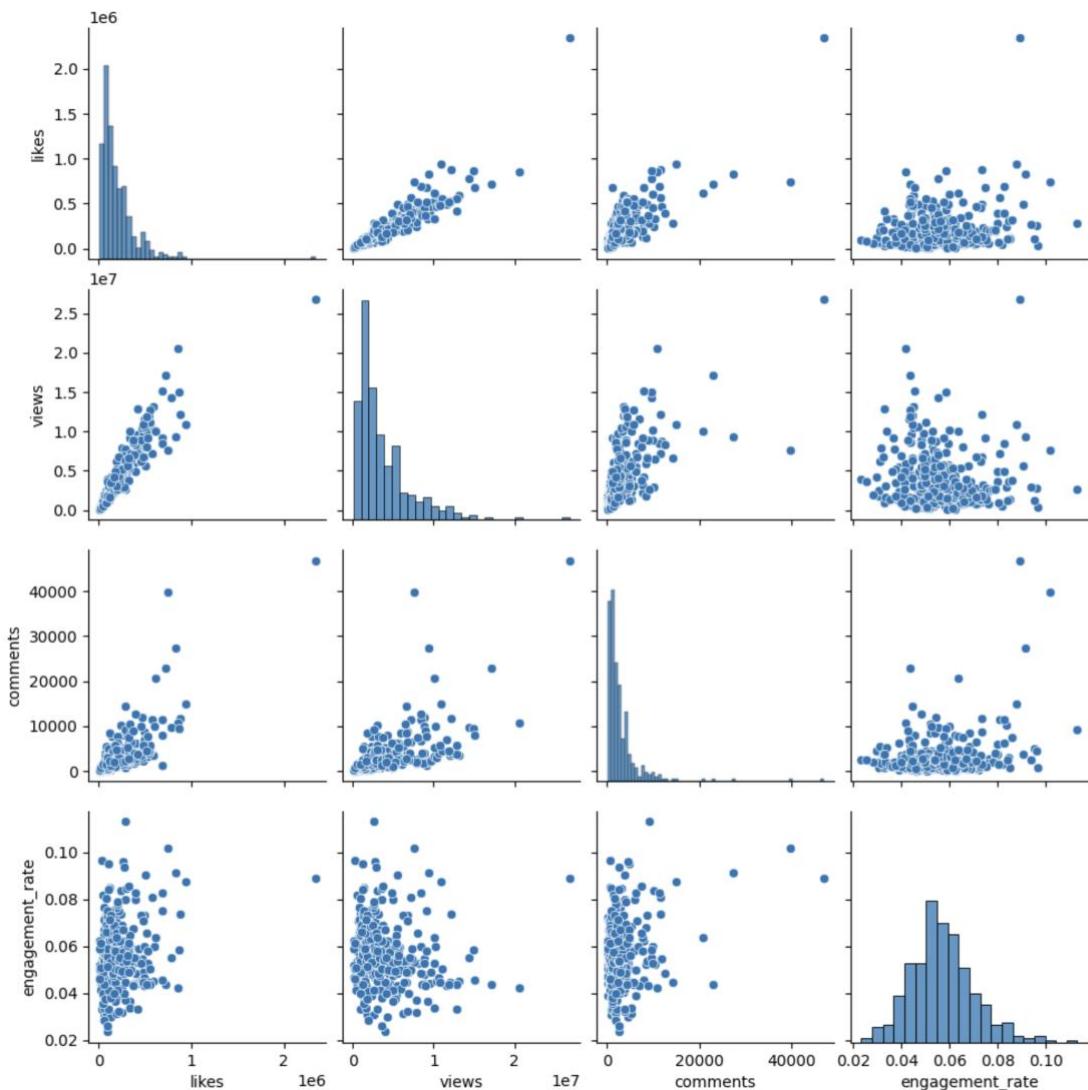


Figura 4 Sactterplot Ibai

En el scatterplot matrix de Ibai, podríamos observar lo siguiente:

- **Relaciones Likes y Views:** Al observar la relación entre "Likes" y "Views", podríamos notar, por ejemplo, que a medida que aumentan los "Likes", también lo hacen las "Views". Esto indicaría que los videos más populares de Ibai no solo reciben más "me gusta", sino también más visualizaciones.
- **Comments y Engagement:** Si hay una concentración densa de puntos en la parte inferior del gráfico de "Comments" vs. "Engagement Rate", podría sugerir que, aunque algunos videos tienen altas tasas de interacción, no necesariamente generan una gran cantidad de comentarios.
- **Identificación de Outliers:** Podríamos identificar videos específicos que, a pesar de tener un alto número de "Views", tienen "Likes" sorprendentemente bajos. Estos videos podrían ser outliers y merecerían un análisis más detallado para entender por qué difieren del patrón general.
- **Relaciones No Lineales:** Observamos patrones curvos o agrupaciones específicas en los gráficos, esto podría indicar relaciones no lineales entre variables, lo que sugiere que ciertos factores o eventos pueden haber influido en el desempeño de esos videos específicos.

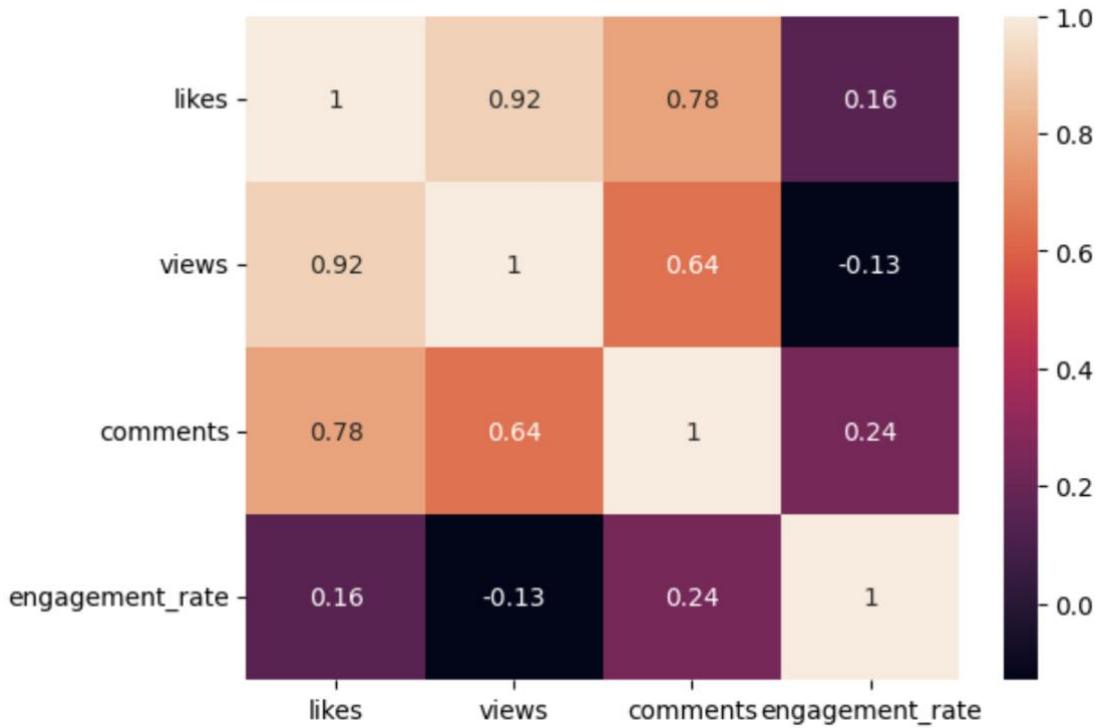


Figura 5 Matriz de correlación Ibai

En la matriz de correlación, se puede observar que los videos que reciben más "likes" tienden a tener también más "views" y "comments", mostrando una fuerte correlación entre estas tres métricas. Sin embargo, la tasa de interacción no muestra una relación tan directa con las visualizaciones o "likes". Específicamente, aunque un video pueda tener muchas visualizaciones, no garantiza una alta tasa de interacción. Por otro lado, los videos con más comentarios tienden a tener una tasa de interacción ligeramente mayor, pero la relación no es tan pronunciada.

En cuanto a las visualizaciones de los análisis de series temporales, nos basamos en meses, días por semana y horas por día.

Gráfico de meses:

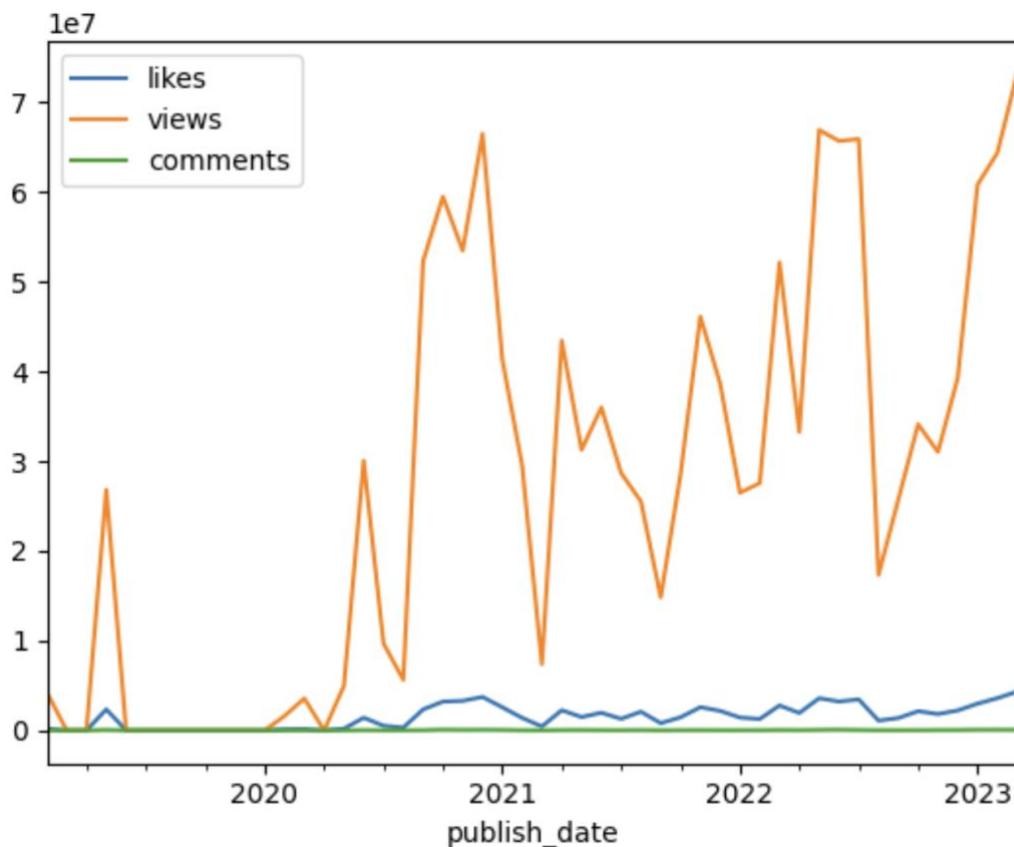


Figura 6 Serie Temporal mes Ibai

El análisis mensual muestra que, aunque los "likes" y "comments" han tenido una tendencia bastante estable a lo largo del tiempo, las "views" han experimentado fluctuaciones más notables. Es interesante observar que durante los meses de verano, las visualizaciones tienden a disminuir, lo que podría deberse a que la audiencia está más ocupada con actividades al aire libre o vacaciones. Sin embargo, a pesar de estas fluctuaciones estacionales, el canal ha experimentado un crecimiento general en los últimos años. Es notable que antes del 2020, el volumen de interacción era significativamente menor, lo que sugiere que Ibai pudo haber ganado popularidad o aumentado la frecuencia de publicación después de ese año.

Gráfico de días por semana:

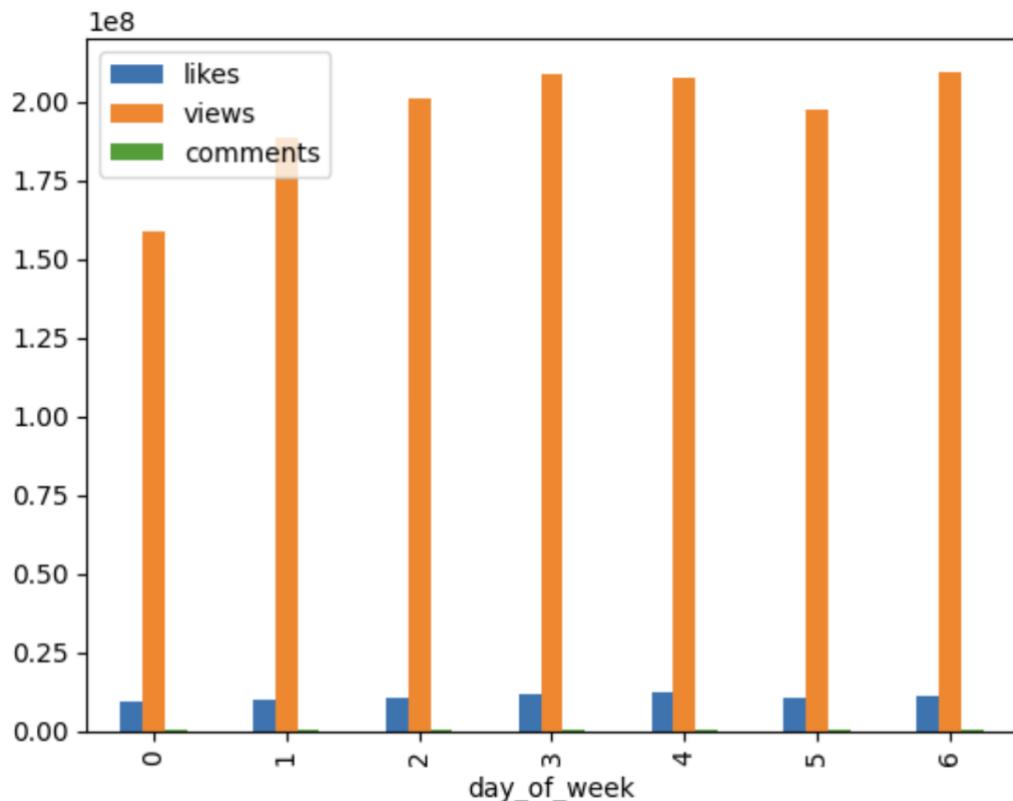


Figura 7 Serie Temporal semana Ibai

El análisis semanal revela que el primer día de la semana (el lunes) es cuando Ibai tiende a publicar menos videos. Esto podría ser una estrategia deliberada o simplemente una preferencia personal. Para el resto de la semana, la actividad y la interacción son bastante consistentes, lo que indica que la audiencia de Ibai está activa y comprometida durante toda la semana.

Gráfico de horas por día:

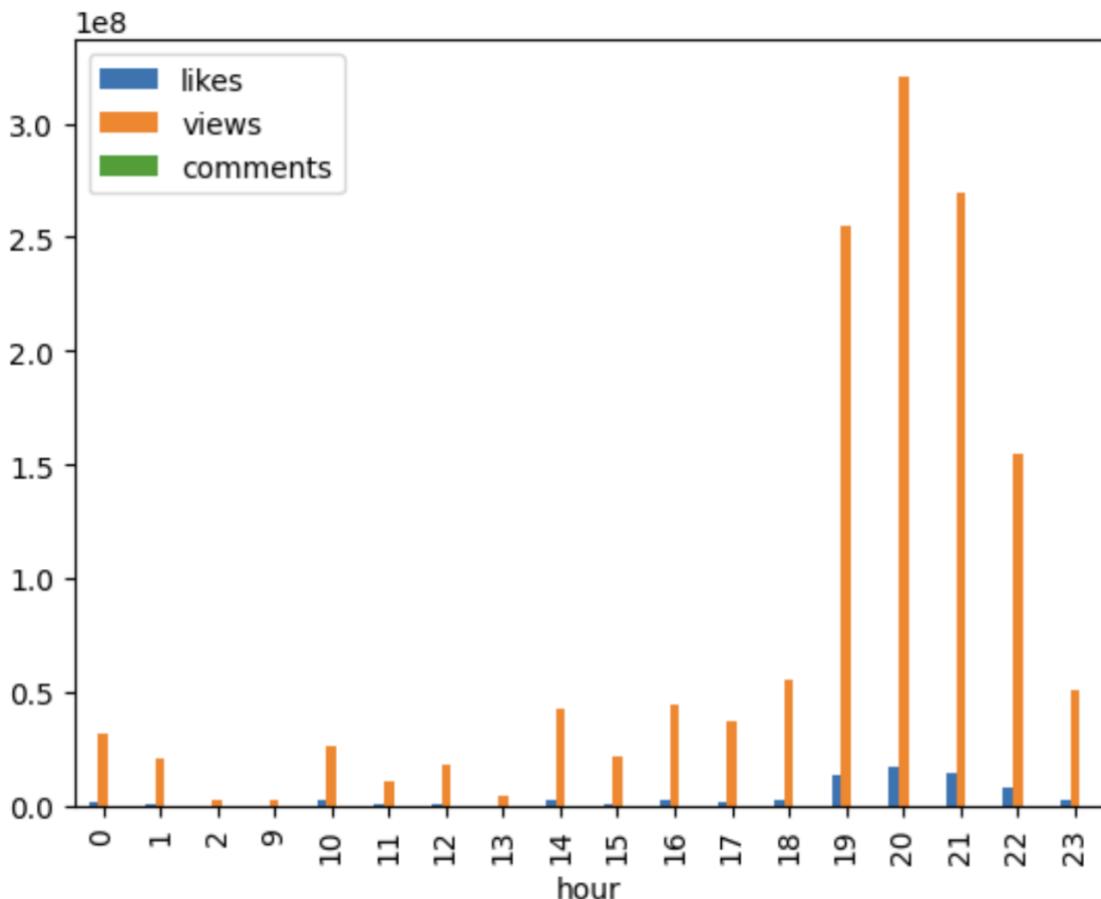


Figura 8 Serie Temporal hora Ibai

El análisis por hora muestra un pico claro de actividad entre las 19 y las 22 horas. Esto sugiere que esta franja horaria es cuando Ibai suele publicar sus videos o cuando su contenido recibe la mayor interacción. Es una ventana estratégica, ya que muchas personas suelen estar en línea durante esas horas, después del trabajo o la escuela, buscando entretenimiento.

Estos análisis proporcionan una visión valiosa de los patrones de publicación y la interacción de la audiencia en el canal de Ibai, lo que puede ser útil para futuras estrategias de contenido y marketing.

Al final de EDA, encontramos el video más popular ‘**El vídeo que necesitas para aprobar selectivid...**’, lo cual nos proporciona información valiosa sobre su audiencia y el tipo de contenido que resuena con ellos. Ibai, conocido por su carisma y estilo de comunicación directo, pudo haber abordado un tema que es crucial para muchos jóvenes en España: la selectividad, que es un examen que determina el acceso a la universidad.

El título del video es intrigante y promete valor, lo que podría haber atraído a muchos estudiantes ansiosos por consejos o información que les ayude en sus exámenes. Además, el hecho de que este video en particular haya sobresalido entre los demás sugiere que Ibai logró combinar su estilo de entretenimiento con un tema de gran relevancia para su audiencia. Esto no solo refuerza su conexión con sus seguidores, sino que también le posiciona como una figura influyente en temas más allá del entretenimiento, llegando a aspectos educativos o de orientación para jóvenes.

Minería de Textos

Título

Top 15 palabras más repetidas:

- **Reaccionar (58 veces):** Ibai podría estar haciendo muchos videos de reacciones, un formato popular en YouTube donde los creadores responden a otros videos, eventos o noticias.
- **Vs (37 veces):** Indica comparaciones o enfrentamientos, quizás en juegos o debates.
- **Reír (32 veces):** Sugiere que el contenido tiene un tono humorístico o que busca entretener.
- **FC (29 veces):** Podría estar relacionado con el fútbol, como "Football Club".
- **Porcino (27 veces):** "Porcino" se refiere a "Porcinos FC", es probable que "Porcinos FC" sea un equipo o grupo relacionado con Ibai, ya sea en un juego, en un contexto deportivo o incluso como un nombre humorístico para un equipo o grupo de amigos.
- **Kings (23 veces):** Podría referirse a un equipo, juego o simplemente usar la palabra "reyes" en un contexto específico.
- **Leagar (22 veces):** Parece ser un error tipográfico o una palabra específica de algún contexto particular.
- **Dar (19 veces):** Verbo común que puede usarse en diversos contextos.
- **Pierd (17 veces):** Parece ser una forma truncada de "pierde" o "perder", indicando fallos o derrotas.
- **Año (16 veces):** Referencia temporal, posiblemente revisando eventos anuales o haciendo resúmenes.
- **Cara (16 veces):** Puede referirse a enfrentamientos cara a cara o a expresiones faciales.
- **Velada (16 veces):** Podría indicar eventos nocturnos o reuniones.
- **Among (15 veces):** Es probable que se refiera al popular juego "Among Us".
- **Hacer (14 veces):** Verbo común que puede usarse en diversos contextos.
- **Kun (12 veces):** Podría ser una referencia a "Kun Agüero", el famoso futbolista con el que Ibai ha colaborado en varias ocasiones.

Estas palabras clave ofrecen una visión general del tipo de contenido que Ibai podría estar produciendo y los temas que aborda con mayor frecuencia.



Figura 9 Wordcloud título Ibai

En la visualización de la nube de palabras asociada a Ibai, se destacan términos que nos dan pistas sobre el tipo de contenido que suele producir. La palabra "reaccionar" resalta, lo que sugiere que Ibai frecuentemente comenta o reacciona a diversos temas o videos. Términos como "streamer", "among us", "tortillaland" y "twitch" indican una fuerte inclinación hacia el mundo de los videojuegos y las transmisiones en vivo. Además, palabras como "messi", "piqué", "barça", "porcinos fc" y "king leagar" (probablemente referenciando a "kings league") nos hablan de su interés y contenido relacionado con el fútbol. Finalmente, nombres como "kun", "bizarrap", "willyrex" y "djmarioo" sugieren colaboraciones o interacciones con otros influencers reconocidos. En conjunto, estos términos nos ofrecen una imagen clara de los intereses y temas predominantes en el contenido de Ibai.

Descripción

Las frecuencias de palabras en las descripciones de los videos de Ibai nos ofrecen una visión clara de los temas recurrentes y las plataformas en las que está activo:

- **canal (618 veces)**: Es probable que Ibai haga referencia frecuentemente a su propio canal o a otros canales relacionados en YouTube.
 - **twitter (333 veces), instagram (331 veces), discord (329 veces), y twitch (328 veces)**: Estas son plataformas populares de redes sociales y streaming. La alta frecuencia de estas palabras indica que Ibai promociona activamente sus perfiles en estas plataformas en las descripciones de sus videos.
 - **contacto (326 veces)**: Es probable que proporcione información de contacto o enlace a una página de contacto para colaboraciones o asuntos comerciales.
 - **podcast (293 veces)**: Ibai podría tener un podcast o hacer referencia a episodios de podcasts en sus descripciones.
 - **tik (292 veces) y tok (292 veces)**: Estas palabras, cuando se combinan, forman "TikTok", lo que sugiere que Ibai también podría estar promocionando o haciendo referencia a su perfil de TikTok.
 - **extraído (285 veces) y editor (285 veces)**: Estas palabras sugieren que Ibai podría estar mencionando de dónde se extrajo cierto contenido o quién editó el video.

- **reddit (251 veces)**: Ibai podría estar haciendo referencia a una comunidad de Reddit o a discusiones relacionadas con su contenido.
 - **ibaikoisquadcom (228 veces) y ibaivizzagencycom (98 veces)**: Estos parecen ser dominios web relacionados con Ibai, posiblemente su sitio web oficial o una agencia con la que esté asociado.
 - **ronda (71 veces)**: Podría referirse a una serie de videos, un tema recurrente o incluso a un patrocinador o colaborador.

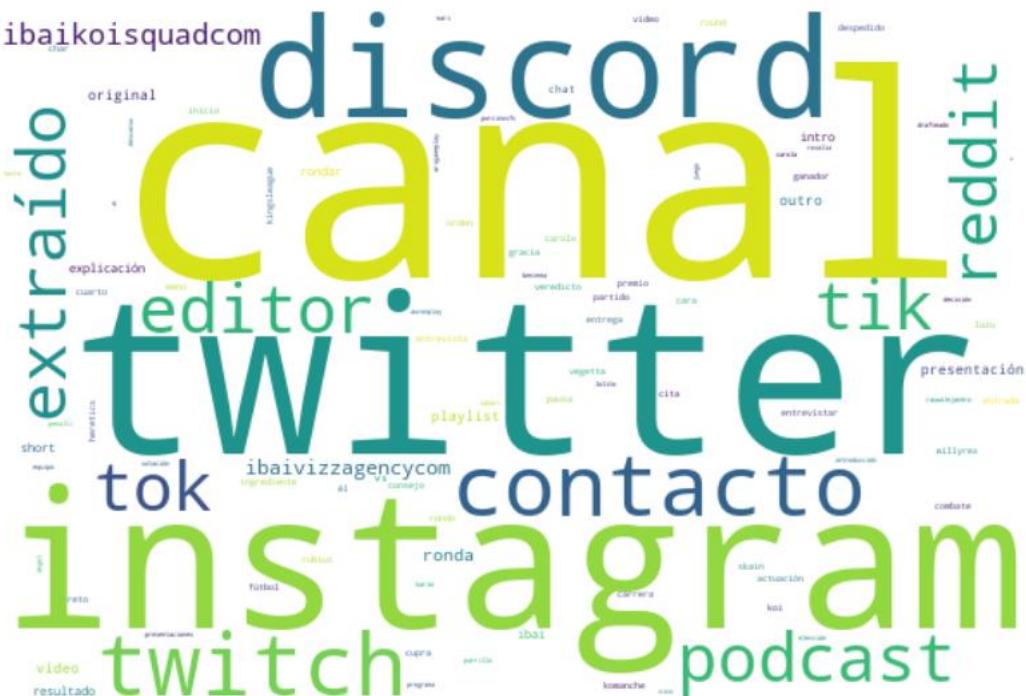


Figura 10 Wordcloud descripción Ibai

En resumen, las descripciones de los videos de Ibai parecen estar centradas en promocionar sus perfiles en diversas plataformas, proporcionar créditos y enlaces relevantes, y mencionar colaboraciones o temas recurrentes.

Transcripción

Las frecuencias de palabras en las transcripciones de los videos de Ibai nos dan una idea de los temas y expresiones recurrentes que utiliza en su contenido:

- **ver (3342 veces):** Ibai utiliza frecuentemente esta palabra, lo que sugiere que a menudo comenta sobre algo que está viendo o pide a su audiencia que observe algo específico.
 - **eh (3203 veces):** Es una interjección común en el habla española, utilizada para llamar la atención o enfatizar algo.
 - **valer (2650 veces):** Esta palabra puede ser usada en el contexto de "vale" como una afirmación o para indicar acuerdo.
 - **tío (2530 veces):** En España, "tío" es una expresión coloquial utilizada para referirse a alguien de manera amigable, similar a "amigo" o "bro" en inglés.
 - **decir (2259 veces):** Indica que Ibai a menudo relata o cita lo que alguien ha dicho.
 - **venir (2249 veces) y pasar (2213 veces):** Estas palabras sugieren que Ibai a menudo habla sobre eventos o situaciones que han ocurrido o están por ocurrir.
 - **querer (2097 veces):** Puede referirse a deseos, intenciones o afecto hacia algo o alguien.
 - **música (2094 veces):** En su mayoría se refiere a música del ambiente que usa en sus videos.
 - **haber (1940 veces):** Es un verbo auxiliar común en español, pero también puede referirse a la existencia de algo.

- **gente (1842 veces):** Ibai a menudo se refiere a las personas o a su audiencia en general.
 - **estar (1758 veces):** Otro verbo común que indica estado o posición.
 - **ibai (1713 veces):** Es interesante que su propio nombre aparezca con frecuencia, lo que sugiere que a menudo habla de sí mismo o que otros lo mencionan en sus videos.
 - **cosa (1694 veces):** Una palabra general que puede referirse a cualquier objeto, situación o tema.
 - **gustar (1691 veces):** Ibai podría estar expresando sus preferencias o preguntando a su audiencia sobre las suyas.



Figura 11 Wordcloud transcripción Ibai

En general, estas palabras reflejan el estilo conversacional y amigable de Ibai. Las transcripciones sugieren que sus videos tienen un tono relajado y cercano, donde a menudo interactúa con su audiencia, comenta sobre diversos temas y comparte sus opiniones y experiencias.

N-grams

Common Phrases:

ay ay ay: 462
 caer puta madre: 239
 eh eh eh: 171
 uy uy uy: 159
 gol gol gol: 152
 valer valer valer: 121
 mira mira mira: 120
 polla polla polla: 51
 cuidado cuidado cuidado: 48
 elegir categoría puntuación: 41
 oy oy oy: 40
 hola hola hola: 40
 allez allez allez: 39
 pol pol pol: 38
 música música música: 37
 pa pa pa: 32
 aplauso música aplauso: 30
 combate velada año: 28
 josé luis blanco: 28
 ey ey ey: 27

Figura 12 Frases communes Ibai

Mediante la técnica 3-grams, podemos ver las frases que usa Ibai en común:

- **ay ay ay (462 veces)**: Esta es una expresión común en español que denota sorpresa, asombro o incluso preocupación. La alta frecuencia sugiere que Ibai a menudo reacciona con sorpresa o emoción a los eventos o contenidos que está comentando o viendo.
- **caer puta madre (239 veces)**: Esta es una expresión coloquial en español que indica que algo es genial o excelente. Ibai podría estar usando esta frase para expresar su aprobación o admiración por algo.
- **eh eh eh (171 veces) y oy oy oy (40 veces)**: Son interjecciones que pueden ser usadas para llamar la atención, expresar duda o sorpresa.
- **uy uy uy (159 veces)**: Similar a "ay ay ay", esta expresión denota sorpresa o preocupación.
- **gol gol gol (152 veces)**: Ibai podría estar comentando un momento en un juego de fútbol o algún otro deporte donde se anota un gol.
- **valer valer valer (121 veces)**: Esta repetición de "vale" (que significa "ok" o "entendido") podría ser una forma de Ibai de enfatizar un punto o mostrar acuerdo.
- **mira mira mira (120 veces)**: Ibai podría estar pidiendo a su audiencia o a alguien más que preste atención a algo específico.
- **polla polla polla (51 veces)**: Es una expresión coloquial que puede tener diferentes connotaciones dependiendo del contexto, desde sorpresa hasta frustración.
- **cuidado cuidado cuidado (48 veces)**: Ibai podría estar advirtiendo o expresando preocupación sobre algo.
- **elegir categoría puntuación (41 veces)**: Esto sugiere que Ibai podría estar refiriéndose a algún tipo de juego o actividad donde se elige una categoría y se otorgan puntos.
- **hola hola hola (40 veces)**: Una forma entusiasta de saludar a su audiencia al inicio de un video o segmento.
- **allez allez allez (39 veces)**: Es una expresión en francés que significa "vamos". Podría estar animando o mostrando apoyo.
- **pol pol pol (38 veces) y ey ey ey (27 veces)**: Estas son interjecciones que pueden ser usadas para llamar la atención o expresar sorpresa.
- **música música música (37 veces)**: Dado que "música" se refiere a la música de ambiente, Ibai podría estar haciendo referencia a un cambio en la música de fondo o reaccionando a ella.

- **pa pa pa (32 veces)**: Podría ser una forma de Ibai de imitar un sonido o ritmo, o simplemente una interjección para llamar la atención.
- **aplauso música aplauso (30 veces)**: Ibai podría estar aplaudiendo al ritmo de la música o en respuesta a un evento en el video.
- **combate velada año (28 veces) y josé luis blanco (28 veces)**: Estas frases sugieren que Ibai podría estar comentando un evento de combate o boxeo, y "José Luis Blanco" podría ser una figura relevante en ese contexto.

En general, estas frases comunes nos dan una idea del estilo comunicativo de Ibai, su entusiasmo, y los temas o eventos que a menudo comenta en sus videos.

Topic Modeling

```
Topics:
Topic: 0, Words: 0.010*"valer" + 0.008*"eh" + 0.007*"ver" + 0.006*"música" + 0.006*"venir" + 0.006*"haber" + 0.006*"decir" + 0.005*"pasar" + 0.005*"tío" + 0.005*"gustar"
Topic: 1, Words: 0.010*"ver" + 0.009*"decir" + 0.008*"gente" + 0.008*"querer" + 0.006*"año" + 0.006*"pasar" + 0.006*"combate" + 0.006*"haber" + 0.005*"cosa" + 0.005*"venir"
Topic: 2, Words: 0.010*"ver" + 0.008*"ander" + 0.008*"eh" + 0.007*"vuelta" + 0.007*"salir" + 0.007*"equipo" + 0.006*"carrera" + 0.006*"karchez" + 0.006*"quedar" + 0.006*"ven".
Topic: 3, Words: 0.014*"eh" + 0.010*"tío" + 0.010*"valer" + 0.010*"rie" + 0.010*"puta" + 0.009*"ibai" + 0.008*"ay" + 0.007*"reír" + 0.007*"ver" + 0.006*"venir"
Topic: 4, Words: 0.009*"ver" + 0.009*"the" + 0.008*"música" + 0.007*"gustar" + 0.006*"decir" + 0.005*"querer" + 0.005*"cosa" + 0.005*"to" + 0.005*"gente" + 0.005*"año"
```

Figura 13 Topic Modeling Ibai

Los tópicos identificados a través del modelo LDA nos ofrecen una visión de los temas y palabras clave recurrentes en el contenido de Ibai. Aquí está una interpretación de los tópicos:

- Tópico 0: Este tópico parece centrarse en reacciones y observaciones generales. Palabras como "valer", "eh", "ver", "música", "venir", "haber", "decir", "pasar", "tío" y "gustar" sugieren comentarios generales sobre lo que está sucediendo, posiblemente en un videojuego o en una situación que está comentando.
- Tópico 1: Este tópico podría estar relacionado con opiniones y reflexiones sobre eventos o situaciones. Las palabras "ver", "decir", "gente", "querer", "año", "pasar", "combate", "haber", "cosa" y "venir" indican una discusión más profunda o una reacción a eventos específicos, posiblemente combates o eventos anuales.
- Tópico 2: Aquí, el tópico parece centrarse en eventos deportivos o competencias. Palabras como "ander", "vuelta", "salir", "equipo", "carrera", "karchez", "quedar" y "venir" sugieren que Ibai podría estar comentando sobre carreras, equipos deportivos o competencias.
- Tópico 3: Este tópico destaca las reacciones emocionales y expresiones coloquiales. Con palabras como "eh", "tío", "valer", "rie", "puta", "ibai", "ay", "reír" y "ver", se puede inferir que Ibai está expresando sorpresa, risa o emoción en respuesta a algo que está viendo o comentando.
- Tópico 4: Este tópico parece tener un enfoque en contenido internacional o música. Las palabras "ver", "the", "música", "gustar", "decir", "querer", "cosa", "to", "gente" y "año" sugieren que podría estar comentando sobre música o contenidos en inglés, o dando su opinión sobre temas generales.

En resumen, estos tópicos nos dan una idea de la variedad de contenidos y temas que Ibai aborda en sus videos, desde reacciones emocionales y comentarios generales hasta discusiones sobre eventos deportivos y música.

Términos Únicos

Unique Terms:

- eh: 15.369802854616294
- tío: 14.410832245053212
- música: 13.55182353825893
- rie: 12.329938487346627
- valer: 11.739517901292501
- xxxx: 10.656477330283549
- ibai: 10.156026667928147
- pasar: 10.048462351217696
- venir: 9.938527887696239
- ay: 9.815273965738031

Figura 14 Términos únicos Ibai

Los términos únicos de Ibai nos ofrecen una visión de las palabras o frases que él utiliza con mayor frecuencia y que lo caracterizan en su contenido:

- eh: Con un valor de 15.37, esta interjección es una de las marcas registradas de Ibai. Es una expresión que utiliza para llamar la atención, enfatizar un punto o expresar sorpresa.
- tío: Con un valor de 14.41, es una expresión coloquial que Ibai utiliza a menudo para dirigirse a su audiencia o a sus amigos. Refleja su estilo cercano y amigable.
- música: Con un valor de 13.55, sugiere que la música, ya sea de fondo o como tema de discusión, juega un papel importante en su contenido.
- rie: Con un valor de 12.33, indica que el humor es esencial en los videos de Ibai. Es probable que se ría con frecuencia, creando un ambiente relajado y entretenido.
- valer: Con un valor de 11.74, es una expresión que utiliza para confirmar o aceptar algo, similar a "vale" o "está bien".
- xxxx: Con un valor de 10.66, es un marcador de posición para palabras que podrían ser inapropiadas o insultos. Indica que Ibai tiene un estilo directo y sin filtros, pero también es consciente de las sensibilidades de su audiencia.
- ibai: Con un valor de 10.16, es natural que mencione su propio nombre con frecuencia, ya sea al presentarse o al referirse a sí mismo en tercera persona.
- pasar: Con un valor de 10.05, sugiere que Ibai a menudo habla de eventos o situaciones, preguntando qué está sucediendo o comentando sobre ellos.
- venir: Con un valor de 9.94, indica que discute sobre cosas que están por venir o que se acercan, ya sea en términos de eventos futuros o de tendencias.
- ay: Con un valor de 9.82, es una interjección que expresa sorpresa, emoción o dolor. Refleja el estilo expresivo y emocional de Ibai.

En resumen, estos términos únicos destacan el estilo comunicativo de Ibai: coloquial, humorístico y directo. También muestran su habilidad para conectar con su audiencia a través de un lenguaje relatable y su capacidad para entretener y mantener a su público comprometido.

Netwrok

Utilizando el Reconocimiento de Entidades Nombradas (NER), hemos identificado las personas que Ibai menciona con frecuencia en sus videos. Estas menciones no solo nos dan una idea de las colaboraciones y relaciones profesionales de Ibai, sino que también destacan las figuras públicas y otros creadores de contenido con los que interactúa o comenta. Al construir una red basada en estas menciones, podemos visualizar y analizar las conexiones y relaciones de Ibai en el mundo del entretenimiento y los medios digitales. Esta red nos permite entender mejor la dinámica y el alcance de su influencia, así como identificar posibles colaboraciones o tendencias en su contenido.

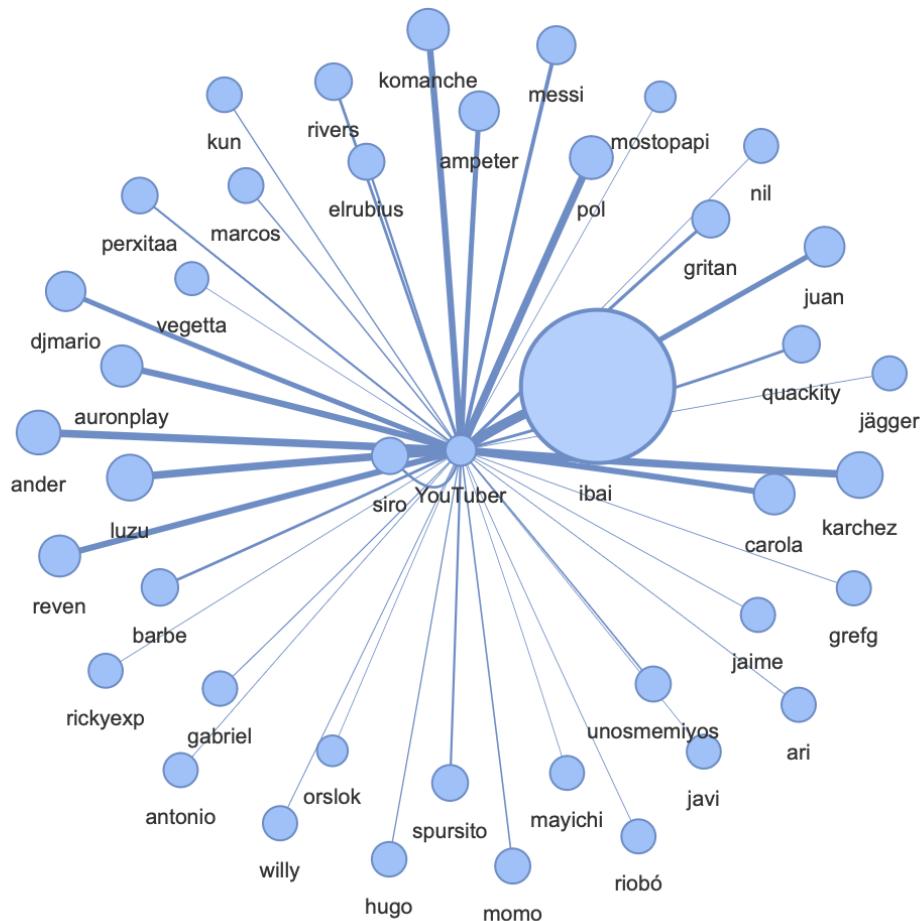


Figura 15 Network Ibai

Según el grafo, podríamos interpretar que Ibai, reconocido creador de contenido, ha establecido conexiones significativas con una variedad de YouTubers y personalidades en la plataforma. Es evidente que Ibai tiene una interacción predominante consigo mismo, lo que podría reflejar temas auto-referenciales o la prominencia de su propia marca en sus contenidos. Además, tiene fuertes interacciones con figuras como "luzu", "karchez" y "ander", lo que indica colaboraciones frecuentes o discusiones sobre estos creadores. Otros nombres prominentes como "auronplay", "elrubius" y "djmario" también figuran en el grafo, lo que subraya la extensa red de relaciones de Ibai en la comunidad de YouTube. En conjunto, estos datos resaltan la influencia y el alcance de Ibai, así como su capacidad para interactuar y colaborar con una amplia gama de creadores en la plataforma.

Auronplay

EDA

	likes	views	comments	year	month	\
count	3.510000e+02	3.510000e+02	351.000000	351.000000	351.000000	
mean	2.328499e+05	4.427076e+06	5029.860399	2021.586895	7.028490	
std	2.330183e+05	4.093930e+06	10328.090258	0.951840	3.651111	
min	4.113000e+04	5.647880e+05	461.000000	2019.000000	1.000000	
25%	1.117720e+05	2.061535e+06	1520.000000	2021.000000	3.000000	
50%	1.701750e+05	3.077356e+06	2103.000000	2022.000000	8.000000	
75%	2.386080e+05	5.011849e+06	4391.500000	2022.000000	10.000000	
max	2.466775e+06	3.080988e+07	135150.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	351.000000	351.000000	351.000000			
mean	2.951567	17.720798	0.054778			
std	1.995836	3.975151	0.012048			
min	0.000000	0.000000	0.020489			
25%	1.000000	16.000000	0.047793			
50%	3.000000	19.000000	0.053940			
75%	5.000000	20.000000	0.059313			
max	6.000000	23.000000	0.180863			

Figura 16 Estadística Auronplay

- **Volumen de Datos:** Hay un total de 351 registros, lo que indica que se han analizado 351 videos de Auronplay.
- **Interacción:**
 - o **Likes:** En promedio, los videos de Auronplay reciben aproximadamente 232,850 likes, con un video que ha alcanzado un máximo de 2,466,775 likes y un mínimo de 41,130 likes.
 - o **Vistas:** Los videos tienen un promedio de 4,427,076 vistas, con un video alcanzando hasta 30,809,880 vistas y otro tan bajo como 564,788 vistas.
 - o **Comentarios:** En promedio, los videos tienen 5,029 comentarios, con un video que ha alcanzado un máximo de 135,150 comentarios y un mínimo de 461 comentarios.
- **Tiempo:**
 - o **Año:** Los videos se extienden desde 2019 hasta 2023.
 - o **Mes:** El mes promedio en el que se publican los videos es alrededor de julio (7.028).
 - o **Día de la semana:** En promedio, los videos tienden a publicarse a principios de semana, alrededor del miércoles (2.952).
 - o **Hora:** La mayoría de los videos se publican alrededor de las 5:45 p.m. (17.721 horas).
- **Tasa de Engagement:** La tasa de engagement promedio es del 5.48%, lo que indica que, en promedio, el 5.48% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 18.09%, mientras que la más baja es del 2.05%.
- **Distribución de Datos:** Las desviaciones estándar (std) para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Auronplay tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

En resumen, Auronplay es un creador de contenido muy popular con una fuerte interacción de la audiencia. Sus videos no solo obtienen un alto número de vistas y likes, sino que también generan una cantidad significativa de comentarios, lo que indica un alto nivel de compromiso de su comunidad.

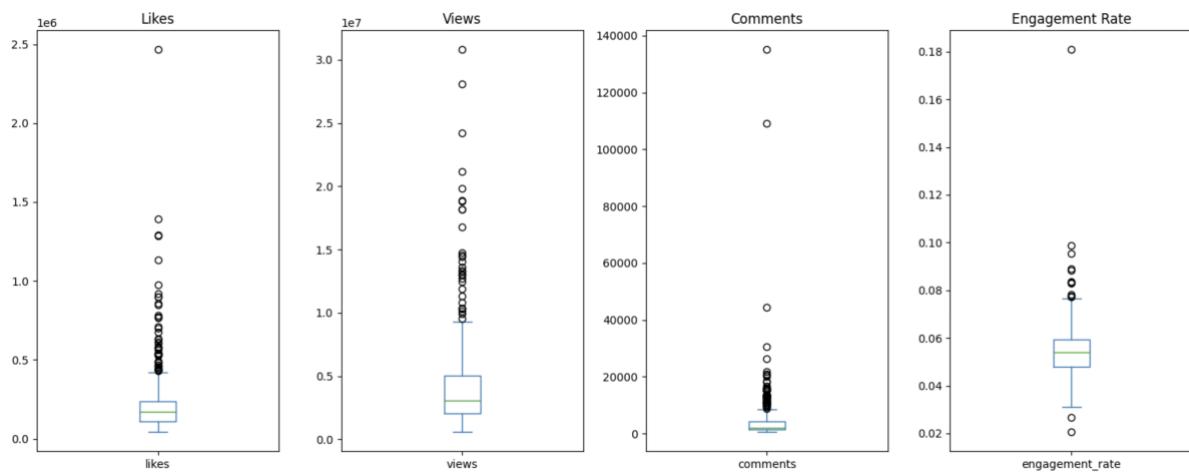


Figura 17 Boxplot Auronplay

Los boxplots de Auronplay muestran que, si bien tiene una base sólida de interacción en la mayoría de sus videos, hay ciertos contenidos que destacan y generan una interacción y visualización significativamente mayor que otros. Los outliers en cada boxplot indican videos que han tenido un rendimiento excepcionalmente bueno en términos de likes, vistas, comentarios y engagement.

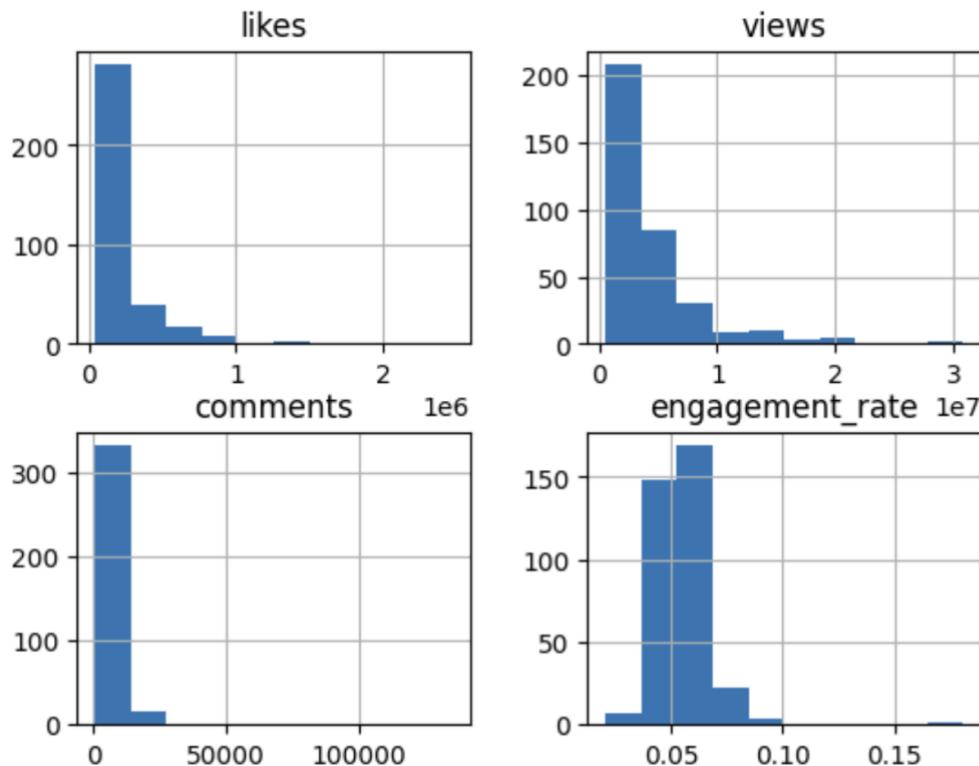


Figura 18 Histograma Auronplay

Entre los gráficos de histograma, se puede observar que la mayoría de los videos de Auronplay tienden a acumular una cantidad moderada de "likes", "views" y "comments". Estos valores se concentran principalmente en el extremo izquierdo del histograma, lo que indica que la mayoría de sus videos obtienen una interacción moderada. Sin embargo, la presencia de colas largas hacia la derecha en estas distribuciones revela que hay ciertos videos que sobresalen y alcanzan niveles de interacción excepcionalmente altos. En relación con la tasa de engagement, la distribución muestra una variabilidad interesante. Aunque muchos videos de Auronplay mantienen una tasa de

engagement estándar, hay otros que logran capturar una atención y participación mucho mayor por parte de la audiencia, lo que se refleja en las colas del histograma. En resumen, mientras que Auronplay tiene una base sólida y constante de interacción en la mayoría de sus contenidos, ciertos videos destacan por su popularidad y capacidad para generar una respuesta más fuerte de la comunidad.

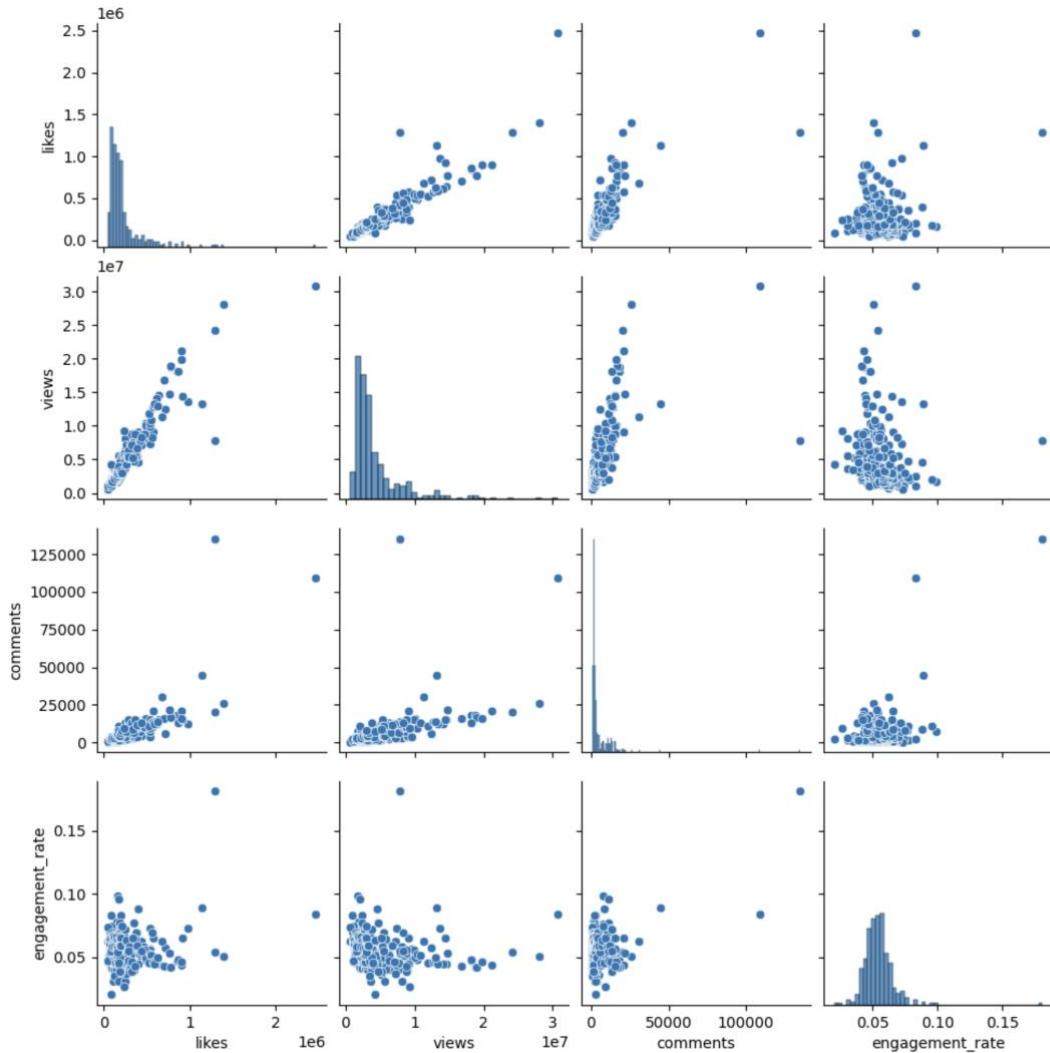


Figura 19 Scatterplot Auronplay

En la visualización scatterplot de Auronplay, cabe mencionar la presencia de ciertos puntos que se desvían notablemente de la concentración principal de datos. Estos valores atípicos representan videos que han alcanzado un rendimiento excepcional en términos de popularidad e interacción. Mientras que la mayoría de sus videos se agrupan en una región específica del gráfico, indicando un rendimiento consistente y predecible, hay ciertos contenidos que han resonado de manera especial con la audiencia, llevándolos a alcanzar cifras de interacción mucho más altas. Estos outliers no solo destacan la capacidad de Auronplay para crear contenido viral ocasionalmente, sino que también subrayan la importancia de analizar y entender qué hace que estos videos en particular sean tan atractivos para su audiencia.

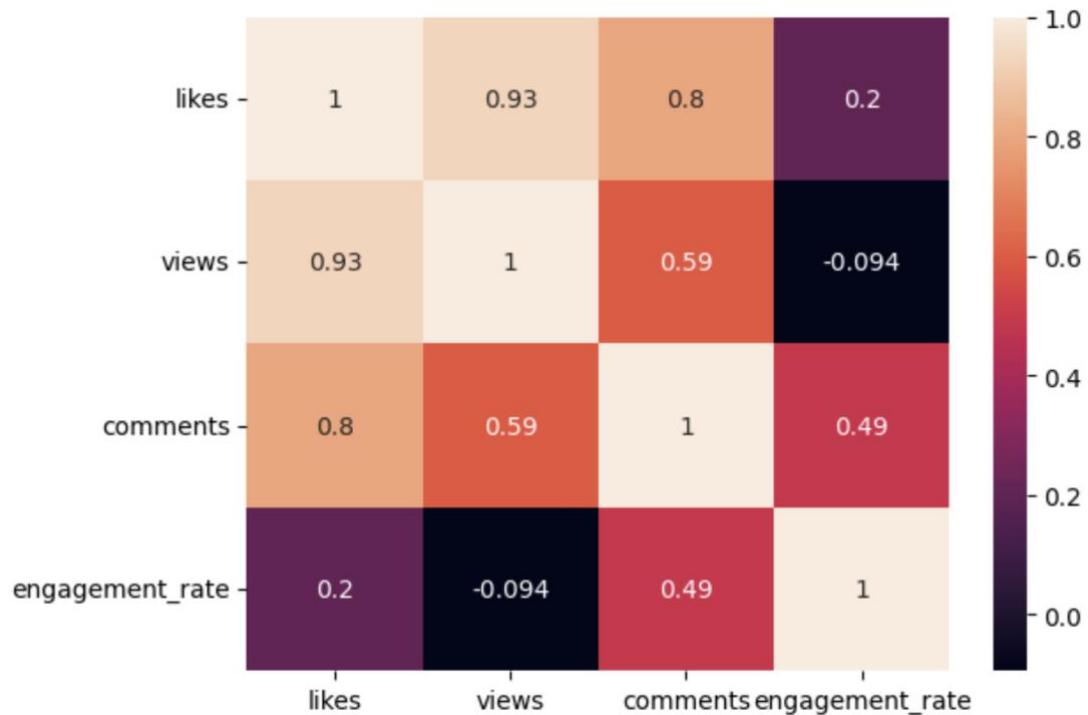


Figura 20 Matriz Correlación Auronplay

En la matriz de correlación, mientras que "likes", "views" y "comments" tienden a tener relaciones positivas entre sí, la tasa de engagement no siempre sigue esta tendencia, especialmente con respecto a las "views".

Serie Temporal

El análisis temporal de Auronplay revela patrones interesantes en su actividad y engagement:

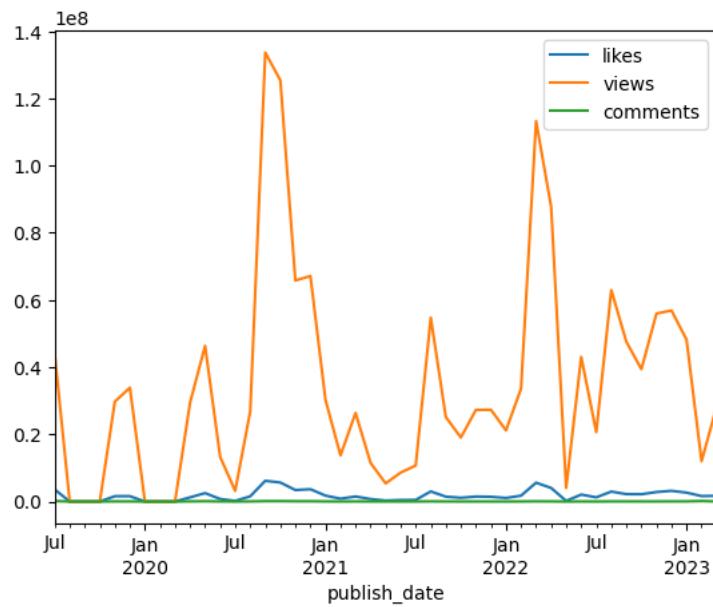


Figura 21 Serie Temporal mes Auronplay

Análisis Mensual: A lo largo de los meses, mientras que los "likes" y "comments" se mantienen relativamente estables, las "views" presentan fluctuaciones más marcadas. Específicamente, hay picos notables en octubre de 2020 y marzo de 2022. Estos picos en las vistas podrían estar asociados

a la publicación de videos que resonaron particularmente bien con la audiencia o que abordaron temas de gran actualidad o interés en esos momentos.

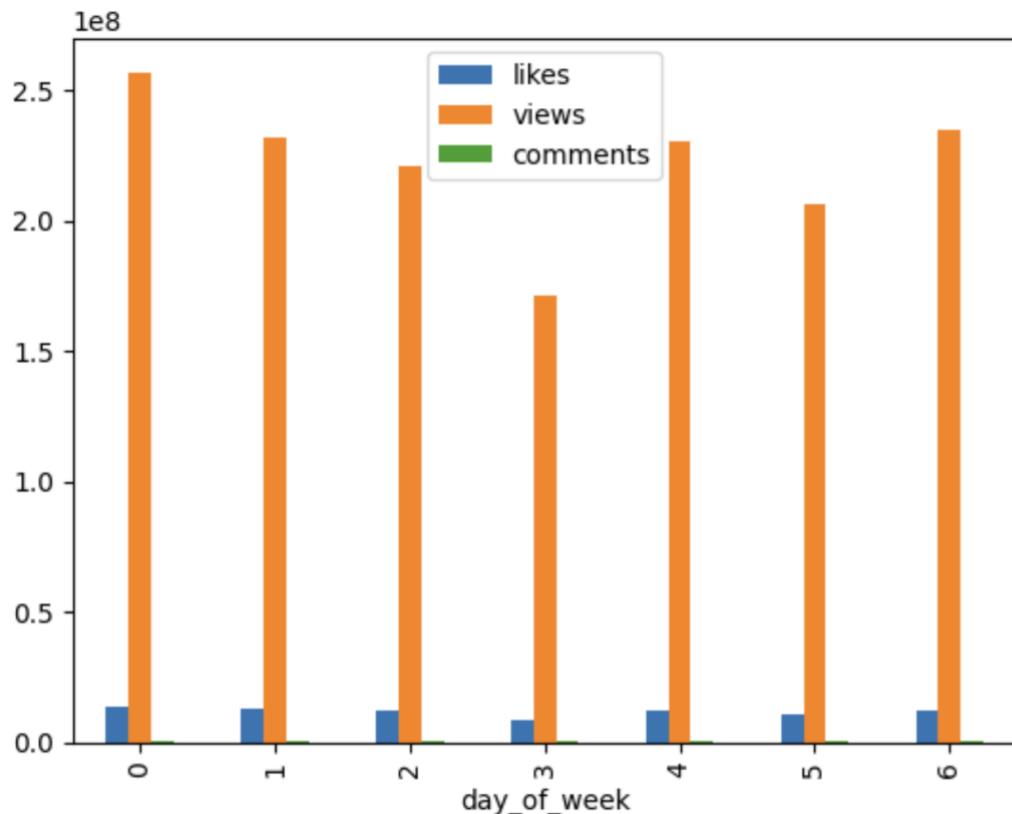


Figura 22 Serie Temporal semana Auronplay

Análisis Semanal: A lo largo de la semana, se observa una tendencia peculiar: Auronplay tiende a no publicar videos los jueves. Esto podría ser una estrategia deliberada o simplemente una coincidencia basada en su ritmo de trabajo y producción.

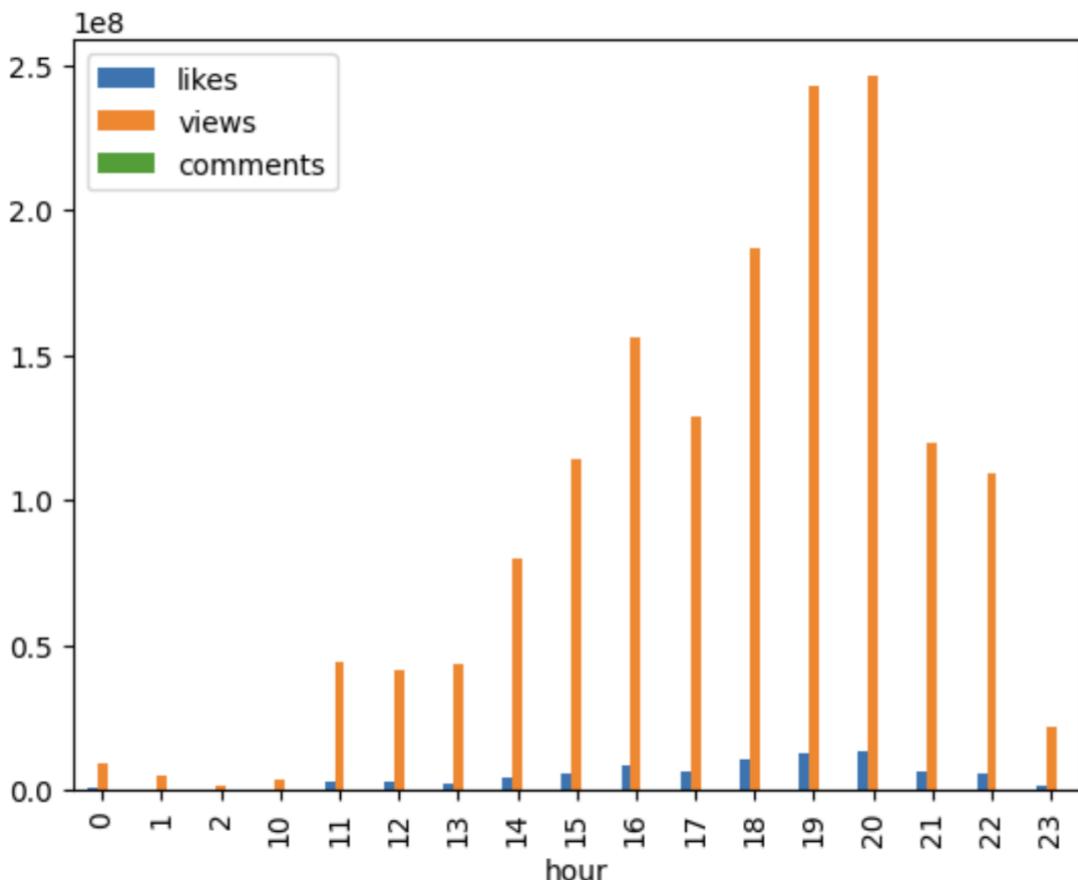


Figura 23 Serie Temporal hora Auronplay

Análisis Horario: Durante el día, hay un claro aumento en la actividad y engagement entre las 16:00 y las 20:00 horas. Esta ventana podría ser el horario preferido de Auronplay para publicar contenido, o simplemente el momento en que su audiencia está más activa y receptiva.

Al final, cabe mencionar el video más popular de Auronplay, "mi primer ASMR espero k os guste", lo cual refleja su habilidad para aprovechar tendencias actuales, en este caso, el ASMR. Su éxito con un género diferente al usual muestra su versatilidad como creador y la lealtad de su audiencia, que está dispuesta a seguirlo en nuevas aventuras de contenido. El título humilde y directo del video también podría haber jugado un papel en atraer a los espectadores. En esencia, este video destaca la adaptabilidad de Auronplay y su conexión con su audiencia.

Minería de Textos

Basándonos en las frecuencias de palabras en los títulos de los videos de Auronplay, podemos inferir ciertos temas y tendencias en su contenido:

- Tortillaland (90 veces): Este parece ser un tema o serie recurrente en su canal, posiblemente relacionado con algún juego o concepto específico que ha ganado popularidad entre sus seguidores.
- EP (56 veces): La abreviatura "EP" sugiere que Auronplay tiene series de videos o episodios en su canal.
- Among Us (44 y 37 veces respectivamente): "Among Us" es un juego multijugador en línea que se volvió viral en 2020. Esto indica que Auronplay ha creado contenido relacionado con este juego, lo que podría haber atraído a una audiencia interesada en dicho juego.

- Cara (28 veces): Podría referirse a desafíos o reacciones "cara a cara" o tener otro significado específico en el contexto de su contenido.
- Carrera y Carrerita (21 y 11 veces): Estas palabras sugieren que Auronplay podría estar haciendo desafíos de carreras o jugando juegos que involucran carreras.
- Roleplay (20 veces): Indica que Auronplay podría estar participando en juegos de rol o creando escenarios ficticios para entretenimiento.
- Aurelio (16 veces): Podría ser un personaje o tema recurrente en su canal.
- Minecraft y GTA (15 veces cada uno): Son dos juegos populares, lo que sugiere que Auronplay ha creado contenido relacionado con estos juegos.
- Volver (14 veces): Podría indicar secuelas o seguimientos de videos anteriores.
- Extremo (11 veces): Sugiere desafíos o actividades que son más intensas o llevadas al límite.



Figura 24 Wordcloud título Auronplay

En general, los títulos reflejan una combinación de contenido de juegos, desafíos y series recurrentes que han resonado con su audiencia.

Basándonos en las frecuencias de palabras en las descripciones de los videos de Auronplay, podemos inferir ciertos temas y tendencias en su contenido:

- Participar (20 veces): Auronplay podría estar organizando eventos, desafíos o sorteos en los que invita a sus seguidores a participar.
- Perxitaa, Bjean, Reborn, Momo, Violeta, Pol, Lolito, Djmario, Axozer, Pato (19, 18, 18, 18, 18, 18, 16, 16, 16, 16 veces respectivamente): Estos parecen ser nombres de otros creadores de contenido o personajes recurrentes en su canal. Es probable que Auronplay colabore frecuentemente con ellos o los mencione en sus videos.
- Mando, Top, Aimcontroller (19 veces cada uno): Estas palabras sugieren que Auronplay podría estar revisando o promocionando ciertos productos relacionados con videojuegos, en particular, controladores o mandos. "Aimcontroller" podría ser una marca o tipo específico de mando.
- Vídeo (10 veces): Aunque es una palabra general, su frecuencia indica que Auronplay podría estar haciendo referencia a videos anteriores, recomendando otros videos o hablando sobre la creación de contenido en sí.



Figura 25 Wordcloud descripción Auronplay

En resumen, las descripciones de Auronplay reflejan colaboraciones con otros creadores, promociones o revisiones de productos relacionados con videojuegos y posiblemente eventos o desafíos en los que invita a sus seguidores a participar.

Basandonos en las frecuencias de palabras en las transcripciones de los videos de Auronplay, podemos inferir ciertos patrones y temas en su contenido:

- xxxx (12,529 veces): Esta palabra, que representa un insulto, sugiere que Auronplay utiliza un lenguaje coloquial y, a veces, puede ser contundente en sus expresiones. Es probable que este estilo de comunicación sea parte de su marca personal y lo que lo hace atractivo para su audiencia.
 - Valer (8,525 veces): Es una palabra común en el español coloquial y puede ser usada en diversos contextos, desde afirmaciones hasta interrogaciones.
 - Música (5,253 veces): Dado que "música" se refiere a la música de fondo, Auronplay podría estar haciendo comentarios o referencias a la música que se está reproduciendo en sus videos, o podría ser una indicación de cambios en la atmósfera o el tono de sus contenidos.
 - Ver, Pasar, Venir, Querer, Mirar (4,957, 4,047, 3,844, 3,762, 3,619 veces respectivamente): Estos verbos sugieren una narrativa activa en sus videos. Auronplay podría estar describiendo situaciones, contando historias o reaccionando a contenido.
 - Eh (3,337 veces): Esta interjección refleja un estilo de comunicación informal y coloquial, y puede ser utilizado para enfatizar puntos o hacer pausas mientras habla.
 - Cosa, Casa, Tío (2,873, 2,796, 2,707 veces respectivamente): "Cosa" y "Casa" son palabras generales que pueden referirse a múltiples temas, mientras que "Tío" es una expresión coloquial en España que se utiliza para referirse a una persona, similar a "amigo" o "compañero".
 - Decir, Poner, Tener (2,684, 2,649, 2,404 veces respectivamente): Estos verbos son comunes en conversaciones y narrativas, y su frecuencia sugiere que Auronplay podría estar compartiendo opiniones, estableciendo contextos o describiendo situaciones en sus videos.



Figura 26 Wordcloud transcripción Auronplay

En resumen, las transcripciones de Auronplay reflejan un estilo de comunicación coloquial y directo, con frecuentes referencias a la música de fondo y un uso activo de verbos que sugieren narración y reacción. Su lenguaje y estilo son probablemente factores clave en su popularidad y conexión con su audiencia.

Common Phrases:

ay ay ay: 709
valer valer valer: 556
mira mira mira: 384
caer xxxx madre: 188
hola hola hola: 163
ver ver ver: 121
correr correr correr: 120
música música música: 99
xxxx xxxx xxxx: 93
dalar dalar dalar: 93
xxxx hijo xxxx: 87
cuidado cuidado cuidado: 83
uy uy uy: 77
mira mira mirar: 72
venir venir venir: 66
ah valer valer: 51
vale valer valer: 51
bienvenido queridísimo amigo
eh eh eh: 51
queridísimo amigo calvo: 50

Figura 27 Frases comunes Auronplay

Las frases comunes en las transcripciones de Auronplay nos ofrecen una visión más detallada de su estilo de comunicación y los temas recurrentes en sus videos:

- ay ay ay (709 veces): Esta expresión puede indicar sorpresa, asombro o énfasis en algo que está comentando.

- valer valer valer (556 veces): "Valer" es una palabra de afirmación en español, similar a "okay" en inglés. Su repetición sugiere que Auronplay podría estar enfatizando un punto o dando su aprobación a algo.
- mira mira mira (384 veces): Esta frase sugiere que está señalando algo específico, pidiendo a su audiencia que preste atención a un detalle particular.
- caer xxxx madre (188 veces): Es una expresión coloquial que denota sorpresa o énfasis, aunque el uso de "xxxx" indica que podría tener un tono más fuerte o vulgar.
- hola hola hola (163 veces): Es probable que esta sea una de sus frases de introducción en sus videos.
- música música música (99 veces): Dado que "música" se refiere a la música de fondo, esta repetición podría indicar momentos en los que la música juega un papel importante en el video, ya sea para enfatizar un punto o cambiar el tono de la narración.
- xxxx xxxx xxxx (93 veces) y xxxx hijo xxxx (87 veces): Estas frases con "xxxx" sugieren que Auronplay utiliza un lenguaje fuerte o coloquial en ciertos momentos para enfatizar sus puntos o reacciones.
- bienvenido queridísimo amigo (51 veces) y queridísimo amigo calvo (50 veces): Estas frases sugieren que Auronplay tiene ciertas expresiones cariñosas o humorísticas para referirse a sus amigos o colaboradores en sus videos.
- eh eh eh (51 veces): Al igual que "ay ay ay", esta interjección puede ser utilizada para enfatizar o señalar algo importante.
- correr correr correr (120 veces) y dalar dalar dalar (93 veces): Estas frases sugieren que hay momentos de acción o emoción en sus videos, donde algo o alguien está en movimiento o hay un sentido de urgencia.

En general, las frases comunes reflejan el estilo energético y coloquial de Auronplay, con un énfasis en la reacción y la interacción con su audiencia. Las repetidas menciones de ciertas palabras y frases sugieren que tiene un estilo distintivo y reconocible que es parte de su marca personal.

```
Topics:
Topic: 0, Words: 0.004*"lluvia" + 0.003*"anunciar" + 0.003*"amigo" + 0.002*"xxxx" + 0.002*"raptor" + 0.002*"música" + 0.002*"bikini" + 0.002*"tomar" + 0.002*"martini"
Topic: 1, Words: 0.019*"xxxx" + 0.013*"valer" + 0.010*"ver" + 0.010*"casa" + 0.009*"querer" + 0.009*"mirar" + 0.008*"música" + 0.007*"poner" + 0.007*"venir" + 0.007*"cosa"
Topic: 2, Words: 0.017*"xxxx" + 0.010*"ver" + 0.009*"valer" + 0.009*"querer" + 0.008*"decir" + 0.007*"coche" + 0.007*"venir" + 0.006*"pasar" + 0.006*"cosa" + 0.006*"tener"
Topic: 3, Words: 0.026*"xxxx" + 0.019*"música" + 0.012*"ver" + 0.012*"valer" + 0.012*"pasar" + 0.008*"favor" + 0.007*"pato" + 0.007*"venir" + 0.007*"matar" + 0.006*"querer"
Topic: 4, Words: 0.035*"xxxx" + 0.029*"valer" + 0.015*"eh" + 0.013*"música" + 0.011*"tío" + 0.009*"venir" + 0.009*"mirar" + 0.009*"pasar" + 0.008*"ver" + 0.006*"mira"
```

Figura 28 Topic Modeling Auronplay

Los tópicos generados a partir de las transcripciones de Auronplay nos ofrecen una visión sobre los temas y palabras clave que predominan en sus videos:

- Tópico 0: Este tópico parece centrarse en situaciones relajadas o momentos de ocio. Palabras como "lluvia", "anunciar", "amigo", "raptor", "bikini", "tomar" y "martini" sugieren escenarios de relajación, posiblemente en un ambiente de playa o vacaciones. La mención de "música" podría indicar que hay un fondo musical en estos momentos.
- Tópico 1: Este tópico parece ser más general y podría relacionarse con las interacciones cotidianas de Auronplay. Las palabras "xxxx", "valer", "ver", "casa", "querer", "mirar", "música", "poner", "venir" y "cosa" sugieren conversaciones y reacciones comunes en sus videos.
- Tópico 2: Aquí, las palabras como "coche" y "tener" junto con "xxxx", "ver", "valer", "querer", "decir", "venir", "pasar" y "cosa" podrían indicar situaciones relacionadas con viajes o desplazamientos, posiblemente aventuras o anécdotas en las que Auronplay comparte experiencias relacionadas con vehículos o desplazamientos.
- Tópico 3: Este tópico tiene un fuerte énfasis en la música, con palabras como "música", "favor", "pato", "matar" y "querer". Podría relacionarse con momentos en los que Auronplay reacciona a canciones o comparte sus opiniones sobre ciertos géneros o artistas musicales.
- Tópico 4: Este tópico parece centrarse en las reacciones y emociones de Auronplay. Las palabras "xxxx", "valer", "eh", "música", "tío", "venir", "mirar", "pasar", "ver" y "mira" sugieren que

Auronplay está comentando o reaccionando a algo, posiblemente compartiendo su opinión o sorpresa sobre un tema en particular.

En general, estos tópicos reflejan la variedad de contenidos y temas que Auronplay aborda en sus videos, desde momentos relajados y de ocio hasta reacciones y opiniones sobre temas cotidianos y música. La repetida mención de "xxxx" sugiere que Auronplay utiliza un lenguaje coloquial y, en ocasiones, fuerte para enfatizar sus puntos o expresar sus emociones.

Unique Terms:
eh: 29.398376421715543
casa: 16.514659438424477
matar: 13.96370609514036
coche: 13.655124376505592
ay: 13.407560474594115
mira: 12.709510892488797
juan: 12.060997497761033
pato: 11.77154180213312
cuidado: 11.361960957424364
risa: 9.375211131644821

Figura 29 Términos únicos Auronplay

Los términos únicos nos ofrecen una perspectiva sobre las palabras o frases que Auronplay utiliza con mayor frecuencia y que lo distinguen en su contenido:

- eh: Con un valor de 29.40, es el término más distintivo de Auronplay. Es una interjección que puede usarse para llamar la atención, expresar sorpresa o enfatizar un punto. Su uso frecuente sugiere que Ibai tiene un estilo de comunicación directo y coloquial.
- casa: Con un valor de 16.51, podría indicar que Auronplay habla a menudo de su hogar, ya sea refiriéndose a anécdotas que ocurren en su casa o a su vida personal.
- matar: Con un valor de 13.96, es probable que esta palabra se use en un contexto humorístico o exagerado, refiriéndose a situaciones en videojuegos o metáforas.
- coche: Con un valor de 13.65, sugiere que Auronplay podría hablar sobre coches, ya sea en relación con anécdotas personales, videojuegos de carreras o temas actuales.
- ay: Con un valor de 13.41, es una interjección que expresa sorpresa, dolor o emoción. Refleja el estilo expresivo de Ibai.
- mira: Con un valor de 12.71, es un verbo que Auronplay podría usar para llamar la atención sobre algo específico o para enfatizar un punto.
- juan: Con un valor de 12.06, podría ser una referencia a una persona específica, quizás un amigo o colaborador frecuente en sus videos.
- pato: Con un valor de 11.77, es incierto sin más contexto, pero podría referirse a un apodo, un personaje o una anécdota específica.
- cuidado: Con un valor de 11.36, sugiere que Auronplay a menudo advierte o enfatiza la precaución en ciertas situaciones, ya sea en un juego o al comentar sobre un tema.
- risa: Con un valor de 9.38, indica que el humor es una parte esencial del contenido de Ibai. Es probable que haga reír a su audiencia con frecuencia o comente sobre situaciones divertidas.

En resumen, estos términos únicos reflejan el estilo comunicativo de Auronplay, que es directo, humorístico y coloquial. También sugieren que aborda una variedad de temas, desde anécdotas personales hasta videojuegos y situaciones cotidianas.

Network

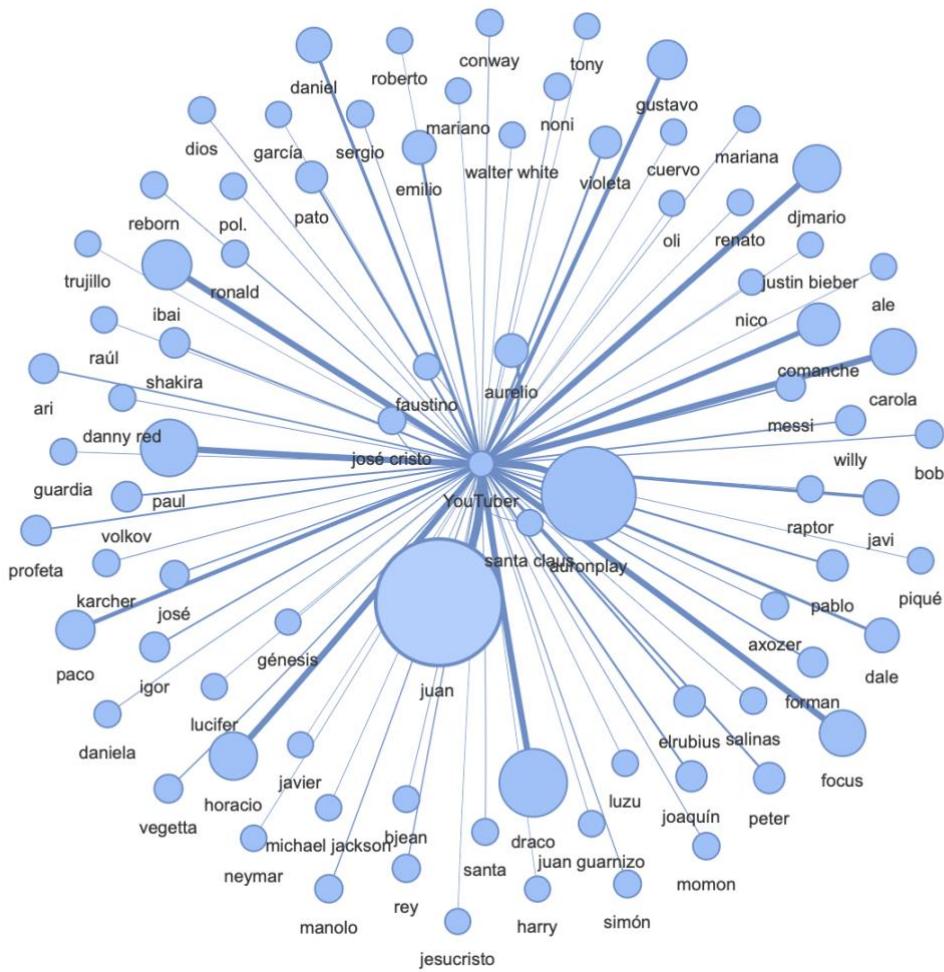


Figura 30 Network Auronplay

El análisis de la red de Auronplay nos proporciona una visión de las interacciones y relaciones que tiene con otros YouTubers y personalidades:

- juan: Con 562 menciones, Juan parece ser una figura prominente en el contenido de Auronplay, lo que sugiere una relación cercana o colaboraciones frecuentes entre ellos.
 - auronplay: Con 386 menciones, es natural que Auronplay hable de sí mismo o se refiera a sí mismo en tercera persona en su contenido.
 - draco: Draco, con 242 menciones, es otro colaborador o amigo cercano de Auronplay, indicando una relación significativa.
 - paul: Con 183 menciones, Paul es otro YouTuber o figura con la que Auronplay interactúa regularmente.
 - ibai: Con 142 menciones, indica que Auronplay e Ibai tienen interacciones o colaboraciones, lo que es coherente dado que ambos son figuras prominentes en la comunidad de YouTube en España.
 - horacio, djmario, focus, carola, comanche: Todos estos nombres tienen más de 100 menciones, lo que sugiere que son colaboradores frecuentes o amigos cercanos de Auronplay.
 - gustavo, paco, javi, daniel, dale, emilio, aurelio, violeta, pato, elrubius, pablo, peter, joaquín: Estos nombres, con menciones que varían entre 25 y 86, indican interacciones ocasionales o colaboraciones con Auronplay.

- volkov, forman, profeta, shakira, igor, willy, ari, messi, josé, vegetta, bob, manolo, rey, josé cristo: Con menos de 40 menciones, estos nombres sugieren interacciones menos frecuentes o menciones ocasionales en el contenido de Auronplay.

Para concluir, la red de Auronplay muestra una amplia gama de interacciones con otros YouTubers y figuras públicas. Algunas de estas relaciones parecen ser más estrechas y frecuentes, mientras que otras son más ocasionales. Estas interacciones y menciones reflejan la naturaleza colaborativa de la comunidad de YouTube y cómo los creadores a menudo trabajan juntos o se mencionan mutuamente en su contenido.

Marta Díaz

EDA

	likes	views	comments	year	month	\
count	233.000000	2.330000e+02	233.000000	233.000000	233.000000	
mean	52802.090129	8.611988e+05	1561.995708	2020.210300	6.618026	
std	49539.453766	8.626668e+05	2541.571457	1.260636	3.477145	
min	3605.000000	7.814500e+04	46.000000	2018.000000	1.000000	
25%	16427.000000	3.035050e+05	298.000000	2019.000000	4.000000	
50%	41068.000000	5.990930e+05	937.000000	2020.000000	7.000000	
75%	69080.000000	1.166748e+06	1883.000000	2021.000000	10.000000	
max	315242.000000	7.682161e+06	30504.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	233.000000	233.000000	233.000000			
mean	4.866953	18.596567	0.063448			
std	1.264358	1.958709	0.017894			
min	0.000000	0.000000	0.028855			
25%	5.000000	18.000000	0.049598			
50%	5.000000	19.000000	0.059975			
75%	5.000000	19.000000	0.074670			
max	6.000000	23.000000	0.121057			

Figura 31 Estadística Marta Díaz

- **Volumen de Datos:** Se han analizado un total de 233 videos de Marta Díaz.
- **Interacción:**
 - **Likes:** En promedio, los videos de Marta Díaz reciben aproximadamente 52,802 likes. El video más popular ha alcanzado 315,242 likes, mientras que el menos popular ha obtenido 3,605 likes.
 - **Vistas:** Los videos tienen un promedio de 861,198 vistas, con un video alcanzando un máximo de 7,682,161 vistas y un mínimo de 78,145 vistas.
 - **Comentarios:** En promedio, los videos reciben 1,562 comentarios, con un video que ha alcanzado un máximo de 30,504 comentarios y un mínimo de 46 comentarios.
- **Tiempo:**
 - **Año:** Los videos se han publicado desde 2018 hasta 2023.
 - **Mes:** El mes promedio de publicación es alrededor de junio.
 - **Día de la semana:** La mayoría de los videos se publican los viernes, con un promedio de 4.87 (donde 0 es lunes y 6 es domingo).
 - **Hora:** La mayoría de los videos se publican alrededor de las 6:36 p.m.

- **Tasa de Engagement:** La tasa de engagement promedio es del 6.34%. Esto indica que, en promedio, el 6.34% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 12.11%, mientras que la más baja es del 2.89%.
- **Distribución de Datos:** Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Marta Díaz tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

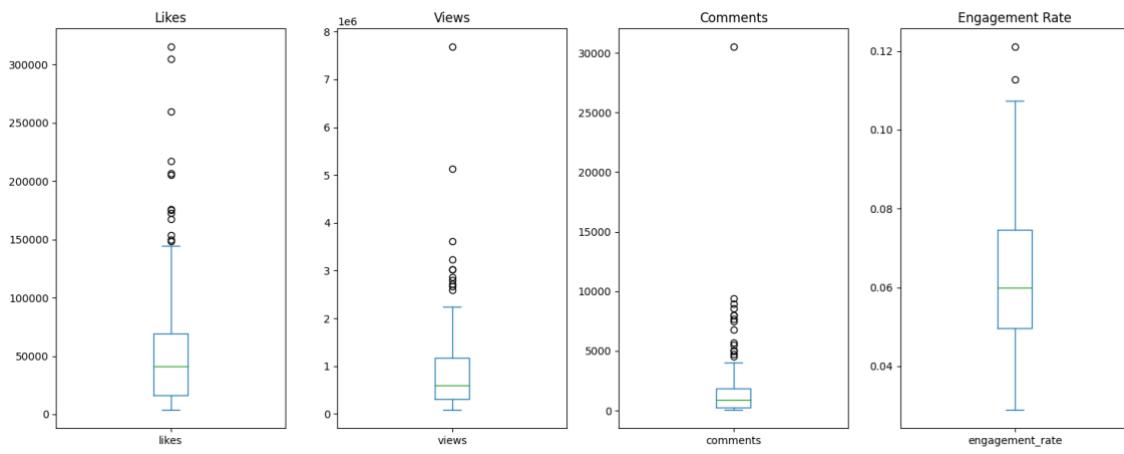


Figura 32 Boxplot Marta Díaz

Los boxplots de Marta Díaz revelan una variabilidad significativa en la interacción de los videos, con ciertos videos que se desempeñan excepcionalmente bien en términos de likes, vistas y comentarios. Estos valores atípicos pueden ser el resultado de contenido particularmente popular o eventos especiales.

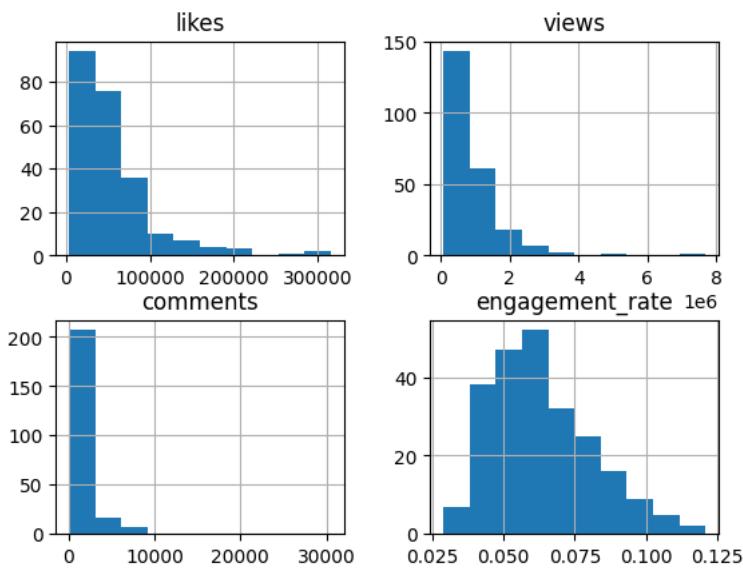


Figura 33 Histograma Marta Díaz

Entre los gráficos de histograma de Marta Díaz, se destaca que la mayoría de sus videos tienden a recibir una cantidad moderada de "likes", "views" y "comments". Estos valores se agrupan en su mayoría en el extremo izquierdo del histograma, sugiriendo que la gran mayoría de sus videos logran una interacción estándar. No obstante, las colas extendidas hacia la derecha en estas distribuciones indican que hay ciertos videos que logran una interacción excepcionalmente alta. En cuanto a la tasa de engagement, la distribución es más variada, mostrando que, si bien muchos videos de Marta Díaz

alcanzan una tasa de engagement promedio, hay algunos que logran capturar una interacción significativamente mayor de la audiencia. En resumen, Marta Díaz mantiene una interacción constante y sólida en la mayoría de sus contenidos, pero ciertos videos sobresalen, demostrando su capacidad para resonar y generar una respuesta más intensa de su comunidad.

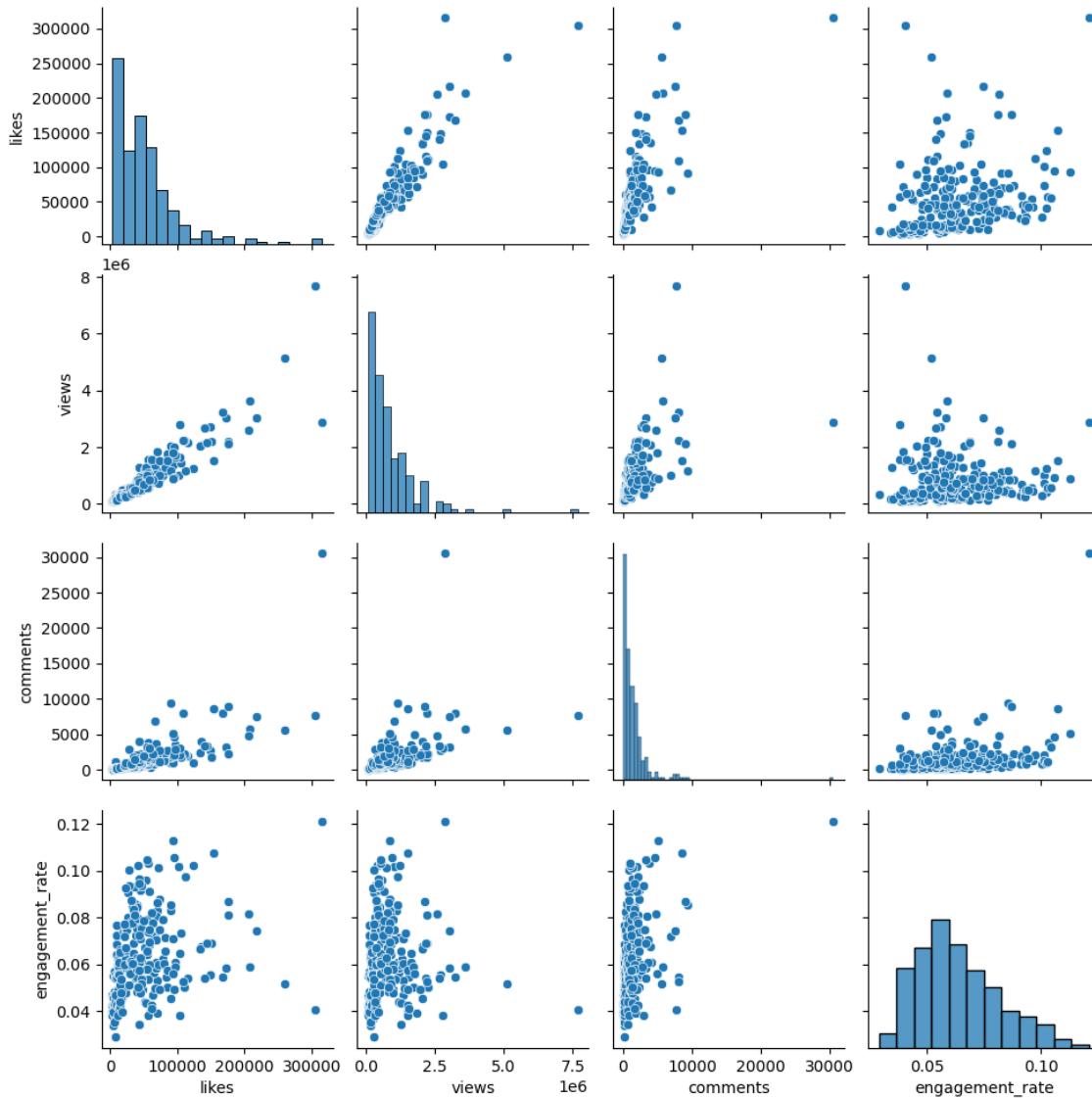


Figura 34 Scatterplot Marta Díaz

En el gráfico de scatterplot, se observa una relación lineal entre "likes" y "views", lo que sugiere que, en general, a medida que un video recibe más "likes", también tiende a acumular más "views". Sin embargo, lo que es particularmente interesante son los valores atípicos presentes en el gráfico. Estos puntos, que se desvían del patrón general, indican videos que, aunque han recibido un número inusualmente alto o bajo de "likes", no han seguido la tendencia general en términos de "views". Estos outliers podrían ser el resultado de videos particularmente virales o controversiales, o incluso de campañas promocionales específicas. Es esencial analizar estos valores atípicos en detalle para comprender mejor las dinámicas únicas que pueden estar influyendo en la interacción del público con esos videos específicos.



Figura 35 Matriz Correlación Marta Díaz

En cuanto a las correlaciones, mientras que "likes", "views" y "comments" están fuertemente correlacionados entre sí, la tasa de engagement tiene una relación más débil con estas métricas, lo que sugiere que otros factores pueden influir en la tasa de engagement de los videos de Marta Díaz.

Serie Temporal

Análisis Mensual: A lo largo de los últimos años, se observa un decrecimiento en la actividad y la interacción en el canal de Marta Díaz en YouTube. Esto puede indicar que Marta ya no está tan activa en la plataforma o que ha perdido parte del apoyo de su audiencia. Es posible que haya factores externos que hayan influido en esta tendencia, como la aparición de nuevas plataformas o cambios en sus intereses y contenidos.

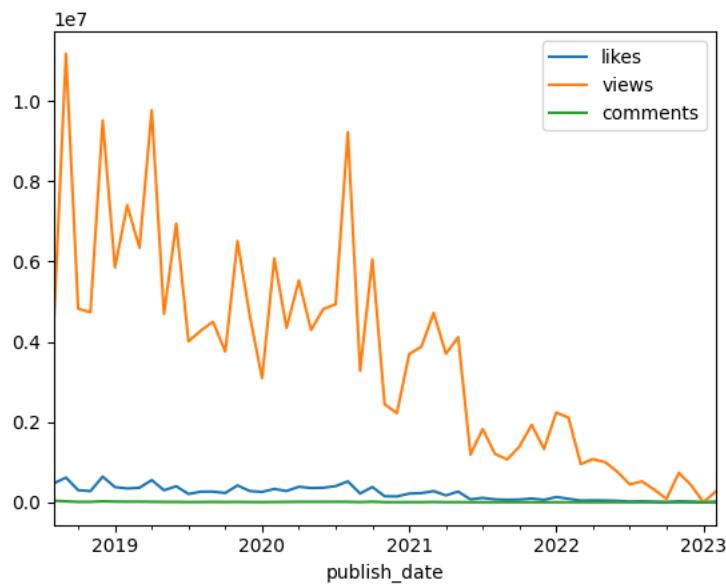


Figura 36 Serie Temporal mes Marta Díaz

Análisis Semanal: El sábado destaca como el día preferido de Marta para publicar contenidos en YouTube. La cantidad de vistas que recibe los sábados es significativamente mayor en comparación con otros días de la semana. Esto sugiere que Marta ha identificado que su audiencia está más activa o dispuesta a consumir contenido durante el fin de semana, específicamente los sábados.

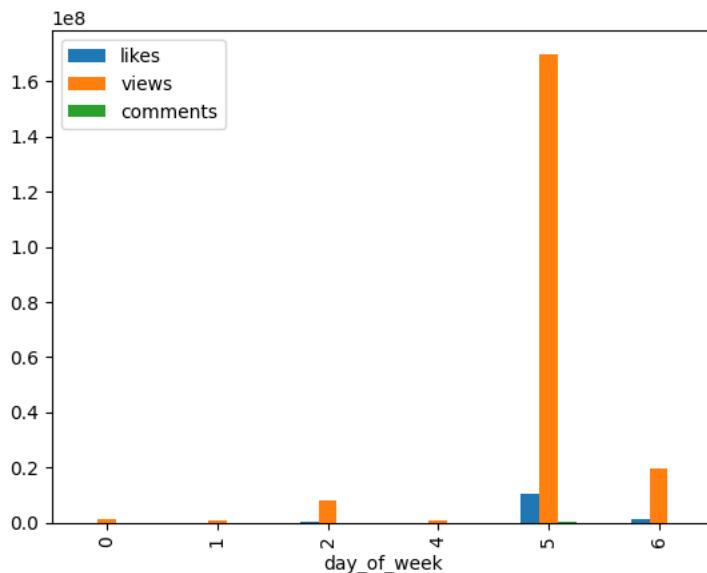


Figura 37 Serie Temporal semana Marta Díaz

Análisis Horario: Se observa un pico de actividad entre las 18 y 19 horas. Esto indica que este es el horario preferido de Marta para publicar sus videos. Es probable que haya identificado que su audiencia está más activa durante estas horas, lo que le permite maximizar la visibilidad y la interacción en sus publicaciones.

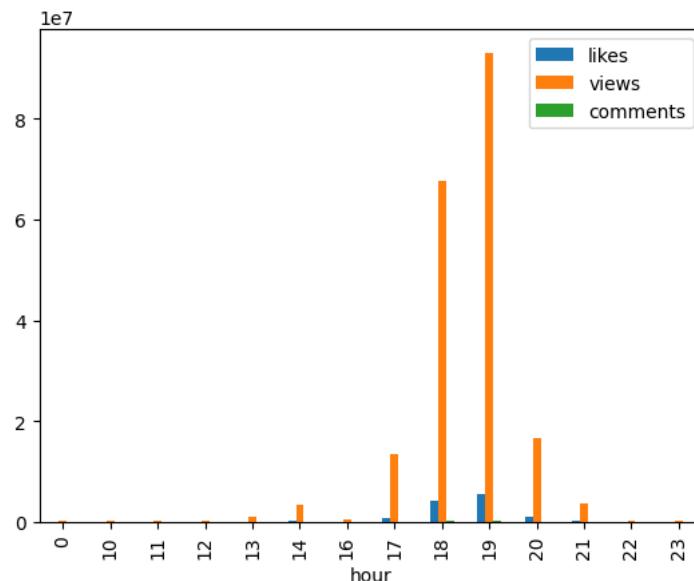


Figura 38 Serie Temporal hora Marta Díaz

Aunque Marta Díaz ha experimentado una disminución en la interacción en los últimos años, todavía sigue una estrategia clara en cuanto a los días y horarios de publicación, buscando optimizar la interacción con su audiencia.

El video titulado "MANERAS de MOLESTAR a MI HERMANO LA VENGANZA" de Marta Díaz es su video más popular. A partir del título, podemos inferir que el contenido del video gira en torno a bromas o travesuras que Marta juega a su hermano como una forma de "venganza". Es probable que este tipo de contenido, que involucra interacciones familiares y bromas ligeras, sea muy atractivo para la audiencia, ya que ofrece entretenimiento y permite a los espectadores relacionarse con situaciones cotidianas entre hermanos. Además, el uso de palabras como "venganza" en el título

puede haber despertado la curiosidad de los espectadores, llevándolos a hacer clic en el video para ver qué sucede. En general, el éxito de este video podría atribuirse a una combinación de un título intrigante, contenido relativo y la dinámica entre Marta y su hermano.

Minería de Textos

Basándonos en las frecuencias de las palabras en los títulos de los videos de Marta Díaz, podemos deducir lo siguiente:

- Marta/Martar/Marto (190 veces en total): Estas palabras, que parecen ser variaciones o errores tipográficos del nombre "Marta", son las más predominantes en los títulos. Esto sugiere que Marta Díaz a menudo personaliza sus videos o los centra en su vida personal y experiencias.
 - Londres (13 veces): Marta podría haber realizado varios videos relacionados con viajes o experiencias en Londres.
 - Hermano (12 veces): Marta a menudo incluye o menciona a su hermano en sus videos, lo que podría ser una dinámica familiar que atrae a la audiencia.
 - Viaje y H (10 veces cada uno): Marta comparte sus experiencias de viaje con su audiencia. La "H" podría ser una abreviatura o inicial relacionada con un tema o persona específica.
 - Reaccionar (10 veces): Marta podría hacer videos de reacciones a diferentes contenidos.
 - Novio (9 veces): Marta podría compartir aspectos de su vida amorosa o experiencias con su pareja en algunos de sus videos.
 - Tik (8 veces): Esto podría referirse a "TikTok", lo que indica que Marta podría estar comentando, reaccionando o haciendo referencia a contenido de TikTok en sus videos.
 - Rutina (7 veces): Marta comparte sus rutinas diarias o actividades específicas con su audiencia, como una "rutina de mañana".



Figura 39 Wordcloud título Marta Díaz

En resumen, los títulos y sus frecuencias reflejan que Marta Díaz crea contenido variado que abarca desde su vida personal hasta reacciones y viajes, lo que la hace atractiva para una amplia gama de espectadores.

Basándonos en las frecuencias de las palabras en las descripciones de los videos de Marta Díaz, podemos deducir lo siguiente:

- Instagram (146 veces): Marta hace referencia frecuente a su perfil de Instagram, lo que sugiere que promueve activamente su cuenta en esta plataforma a través de sus videos de YouTube.
 - Marta/Martar (184 veces en total): Al igual que en los títulos, Marta a menudo personaliza las descripciones de sus videos, reforzando su marca personal.
 - Tik/Tok (203 veces en total): Estas palabras, que claramente se refieren a "TikTok", indican que Marta podría estar compartiendo contenido relacionado con esta plataforma, ya sea promocionando su cuenta de TikTok o haciendo referencia a tendencias populares de TikTok.
 - Ropa (83 veces): Marta podría estar involucrada en la moda o hacer videos relacionados con hauls, outfits o recomendaciones de ropa.
 - Código y Dtco (130 veces en total): Estas palabras sugieren que Marta podría estar ofreciendo códigos de descuento a sus seguidores, posiblemente en colaboración con marcas o tiendas.
 - Especial (59 veces): Marta podría tener videos que destaqueen ocasiones o temas especiales.
 - Sígueme/Suígueme (90 veces en total): Marta anima activamente a sus seguidores a seguirla en otras plataformas o a suscribirse a su canal.
 - Libro (44 veces): Marta podría haber escrito un libro o estar promocionando uno. También podría estar haciendo reseñas o mencionando libros en sus videos.
 - Esperar y Encantar (84 veces en total): Estas palabras sugieren un tono positivo y amigable en las descripciones de Marta, mostrando entusiasmo y agradecimiento hacia su audiencia.
 - Link (17 veces): Marta proporciona enlaces en sus descripciones, posiblemente a productos, colaboraciones o a sus otras redes sociales.



Figura 40 Wordcloud descripción Marta Díaz

En resumen, las descripciones de Marta Díaz reflejan una combinación de promoción personal, colaboraciones con marcas y un enfoque en la interacción con su audiencia. También muestra una diversidad de intereses, desde la moda hasta las redes sociales y la literatura.

Mediante el análisis de las frecuencias de las palabras en las transcripciones de los videos de Marta Díaz, podemos deducir lo siguiente:

- Música (1573 veces): Es evidente que la música juega un papel importante en los videos de Marta. Podría estar haciendo videos relacionados con la música, bailando, o simplemente usando música de fondo en sus videos.
 - Vídeo (1131 veces): Marta hace referencia frecuente a sus propios videos, posiblemente dando contexto, introduciendo o concluyendo temas.
 - Súper (966 veces): Esta palabra sugiere que Marta tiene un tono entusiasta y positivo en sus videos. Probablemente use "súper" para describir cosas que le gustan o experiencias que ha disfrutado.
 - Gustar (958 veces): Marta comparte frecuentemente su opinión sobre diferentes temas, indicando lo que le gusta o no.
 - Ver (890 veces): Podría estar refiriéndose a mostrar algo a su audiencia o hablar sobre algo que ha visto.
 - Cosa (872 veces): Marta podría estar discutiendo una variedad de temas y "cosa" es una palabra genérica que podría usar para referirse a diferentes temas o ítems.
 - Querer (750 veces): Marta comparte sus deseos o cosas que le gustaría hacer o tener.
 - Valer (704 veces): Podría ser una expresión que usa con frecuencia, similar a "vale" en español, para afirmar o concluir un punto.
 - Chico (698 veces): Marta podría estar hablando sobre relaciones, amigos o simplemente usando "chico" como una palabra genérica para referirse a las personas.
 - Poquito (638 veces): Sugiere que Marta a menudo modera o minimiza ciertas afirmaciones, quizás para ser humilde o para dar énfasis a algo.
 - Quedar, Decir, Pasar, Venir, Estar: Estos verbos comunes sugieren que Marta narra historias o experiencias en sus videos, compartiendo eventos pasados, presentes o futuros y dando su opinión o reacción a ellos.



Figura 41 Wordcloud transcripción Marta Díaz

En general, las transcripciones de Marta Díaz reflejan una personalidad entusiasta y comunicativa que comparte sus experiencias, opiniones y emociones con su audiencia. Su contenido es variado, y aborda temas desde música hasta relaciones y eventos cotidianos.

Common Phrases:

esperar gustar muchísimo:	78
estar bienvenido vídeo:	77
ver semana venir:	69
chico estar bienvenido:	56
vídeo esperar gustar:	55
música aplauso música:	47
bienvenido vídeo sábado:	43
vídeo sábado canal:	39
esperar gustar montón:	36
semana venir vídeo:	28
chica estar bienvenido:	26
música música música:	24
chico vídeo esperar:	24
gustar muchísimo vídeo:	24
súper súper súper:	21
chico esperar gustar:	19
sábado canal sábado:	18
aplauso música aplauso:	17
chico acabar llegar:	15
semana venir música:	14

Figura 42 Frases comunes Marta Díaz

Basándonos en las frases comunes de las transcripciones de los videos de Marta Díaz, podemos deducir lo siguiente:

- Introducciones y Saludos: Frases como "esperar gustar muchísimo", "estar bienvenido vídeo", "chico estar bienvenido", "bienvenido vídeo sábado" y "chica estar bienvenido" sugieren que Marta tiene una forma característica de saludar a su audiencia al comienzo de sus videos. Estas frases reflejan una cálida bienvenida y un deseo de que a su audiencia le guste el contenido que presenta
- Referencias Temporales: Las frases "ver semana venir", "vídeo sábado canal" y "semana venir vídeo" indican que Marta podría estar haciendo referencia a la frecuencia con la que publica o a eventos específicos que ocurrirán en el futuro.
- Música: Las repetidas menciones de "música aplauso música", "música música música" y "semana venir música" sugieren que la música juega un papel importante en sus videos, ya sea como fondo o como tema principal de discusión.
- Énfasis en el Contenido: Las frases "vídeo esperar gustar", "esperar gustar montón", "gustar muchísimo vídeo" y "chico vídeo esperar" reflejan el deseo de Marta de que su contenido sea bien recibido y apreciado por su audiencia.
- Repetición para Énfasis: La repetición en frases como "súper súper súper" y "aplauso música aplauso" sugiere que Marta utiliza la repetición como una técnica para enfatizar ciertos puntos o emociones.
- Interacción con la Audiencia: Frases como "chico acabar llegar" indican que Marta podría estar narrando o describiendo eventos en tiempo real, creando una sensación de interacción inmediata con su audiencia.

En resumen, las frases comunes en las transcripciones de Marta Díaz reflejan una personalidad entusiasta y acogedora que valora la interacción con su audiencia. Marta utiliza frases repetitivas para enfatizar puntos y crear un tono familiar y reconocible en sus videos. Su contenido parece estar

centrado en compartir experiencias, música y eventos con sus seguidores, siempre con una actitud positiva y esperanzada.

```

Topics:
Topic: 0, Words: 0.020*música" + 0.011*video" + 0.009*chico" + 0.008*querer" + 0.007*ver" + 0.006*hermano" + 0.006*valer" + 0.006*gustar" + 0.006*comer" + 0.005*foto"
Topic: 1, Words: 0.013*súper" + 0.010*música" + 0.010*quedarn" + 0.009*gustar" + 0.009*querer" + 0.009*casa" + 0.008*ver" + 0.008*video" + 0.008*poquito" + 0.007*poner"
Topic: 2, Words: 0.025*música" + 0.007*súper" + 0.006*gustar" + 0.006*querer" + 0.005*gente" + 0.005*chico" + 0.005*video" + 0.005*encantar" + 0.005*casa"
Topic: 3, Words: 0.012*video" + 0.012*casa" + 0.010*gustar" + 0.009*valer" + 0.009*ver" + 0.008*decir" + 0.008*musical" + 0.008*súper" + 0.007*chico" + 0.007*pasar"
Topic: 4, Words: 0.020*música" + 0.013*video" + 0.011*gustar" + 0.010*súper" + 0.010*ver" + 0.008*casa" + 0.007*chico" + 0.007*querer" + 0.007*pasar" + 0.007*venir"

```

Figura 43 Topic Modeling Marta Díaz

Basándonos en los tópicos extraídos de las transcripciones de los videos de Marta Díaz, podemos deducir lo siguiente:

- Tópico 0 - Estilo de Vida y Experiencias Personales: Las palabras clave como "música", "vídeo", "chico", "hermano", "comer" y "foto" sugieren que este tópico podría estar centrado en compartir experiencias personales, momentos con su familia, especialmente con su hermano, y su estilo de vida diario. También podría hablar sobre sus gustos musicales y compartir fotos o recuerdos.
- Tópico 1 - Opiniones y Sentimientos: Las palabras "súper", "gustar", "querer", "cosa", "poquito" y "poner" indican que Marta comparte sus opiniones y sentimientos sobre diferentes temas. Podría estar describiendo algo que le gusta mucho o algo que quiere hacer o tener.
- Tópico 2 - Música y Emociones: Con palabras como "música", "súper", "poquito", "gente", "encantar" y "casa", este tópico parece estar centrado en la música y cómo esta se relaciona con sus emociones y experiencias. Podría hablar sobre canciones que le encantan o cómo se siente en ciertos momentos.
- Tópico 3 - Reflexiones y Anécdotas: Las palabras "vídeo", "cosa", "decir", "súper", "pasar" y "valer" sugieren que Marta podría estar reflexionando sobre eventos recientes o compartiendo anécdotas y historias con su audiencia.
- Tópico 4 - Actividades y Encuentros: Las palabras "música", "vídeo", "ver", "cosa", "chico", "venir" y "pasar" indican que este tópico podría estar relacionado con actividades que realiza, lugares que visita y encuentros con amigos o conocidos.

Los tópicos reflejan que Marta Díaz comparte una amplia variedad de contenidos con su audiencia, desde experiencias personales y familiares hasta sus gustos musicales, opiniones y reflexiones. Su contenido es diverso y ofrece una mezcla de entretenimiento, emociones y anécdotas que la conectan con su comunidad.

Unique Terms:

valer:	8.747709055982027
poquito:	8.457345663718362
quedar:	7.800952435747889
pasar:	7.653497351006587
poner:	7.485549034322666
foto:	7.228858623618623
sabéis:	6.868433854212611
muchísimo:	6.16590693166335
encantar:	6.155779947522807
yo:	5.883596465289366

Figura 44 Términos únicos Marta Díaz

Basándonos en los términos únicos extraídos de las transcripciones de los videos de Marta Díaz, podemos inferir lo siguiente:

- valer: Es una palabra que Marta utiliza con frecuencia, posiblemente como una expresión para afirmar o dar énfasis a algo, similar a "vale" en español.
- poquito: Esta palabra sugiere que Marta a menudo habla de cosas en pequeñas cantidades o de momentos breves. Podría referirse a hacer algo "un poquito" o a describir algo que le gusta "un poquito".
- quedar: Marta podría hablar sobre cómo se siente con respecto a algo, cómo algo queda en un lugar específico o cómo algo se ve.
- pasar: Este término sugiere que Marta comparte eventos o experiencias que le han sucedido. Podría estar contando historias o anécdotas sobre lo que le ha "pasado".
- poner: Marta podría hablar sobre colocar algo en un lugar específico, o sobre cómo se "pone" algo, como maquillaje o ropa.
- foto: Es probable que Marta comparta fotos con su audiencia o hable sobre fotografías, ya sea mostrando recuerdos o discutiendo sobre fotografía en general.
- sabéis: Esta palabra indica que Marta se dirige directamente a su audiencia, buscando su validación o compartiendo algo que cree que podrían saber.
- muchísimo: Marta utiliza esta palabra para dar énfasis, indicando que algo le gusta o le importa "muchísimo".
- encantar: Es evidente que Marta comparte con su audiencia cosas que le "encantan", ya sean actividades, lugares, personas o experiencias.
- yo: Marta habla desde una perspectiva personal, compartiendo sus propias experiencias, pensamientos y sentimientos con su audiencia.

En conclusión, estos términos únicos ofrecen una visión de los temas y el estilo de comunicación de Marta Díaz. Ella comparte experiencias personales, se comunica directamente con su audiencia y enfatiza sus emociones y opiniones sobre diversos temas.

Network

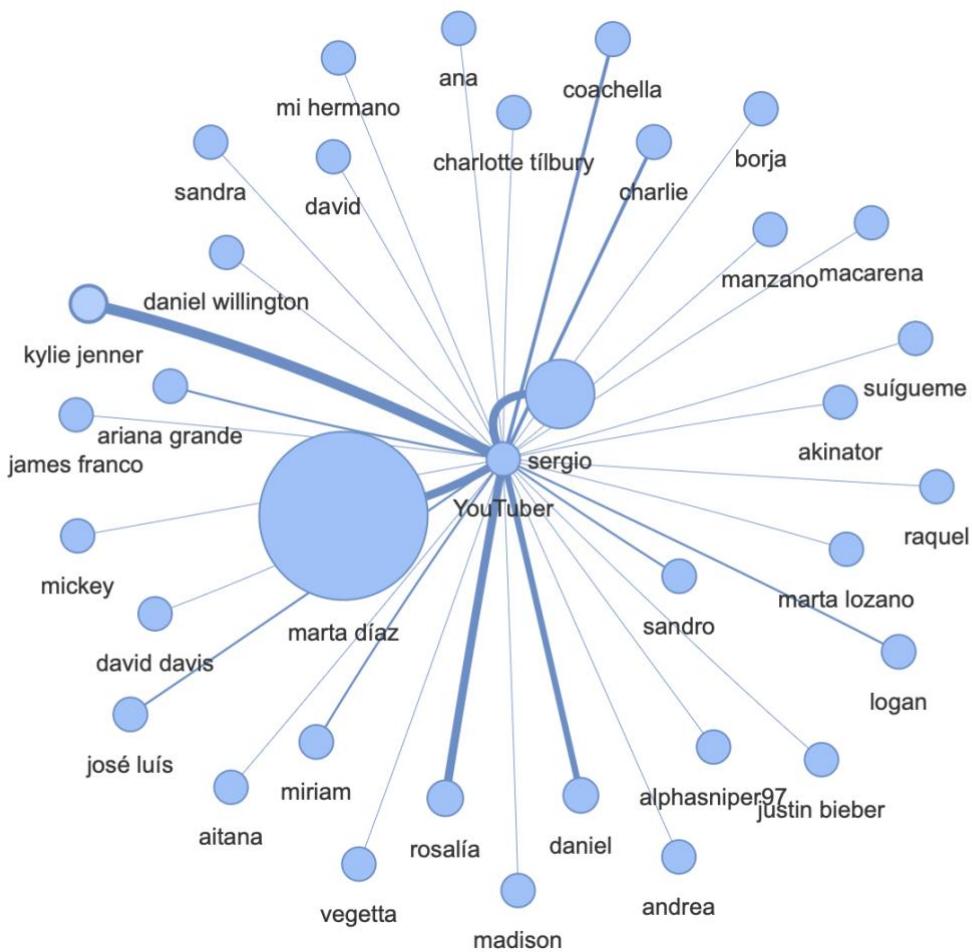


Figura 45 Network Marta Díaz

El "network" proporcionado muestra las conexiones o interacciones entre Marta Díaz y otros YouTubers o personalidades. A partir de la información proporcionada, podemos inferir lo siguiente:

- Marta Díaz: Es la principal figura de este network, y todas las conexiones se relacionan con ella.
- Sergio: Es la segunda figura más prominente en este network con 101 menciones, lo que sugiere que tiene una relación o interacción significativa con Marta Díaz. Podría ser un colaborador frecuente, un amigo cercano o incluso un familiar.
- Kylie Jenner y Rosalía: Aunque son figuras muy conocidas a nivel mundial, tienen menos menciones en comparación con Sergio. Esto podría indicar que Marta ha mencionado o interactuado con ellas en algunas ocasiones, pero no tan frecuentemente.
- Daniel, Charlie, Coachella: Estas figuras y eventos tienen un número moderado de menciones, lo que sugiere que han tenido alguna relevancia en el contenido de Marta Díaz, pero no son figuras centrales.
- Ariana Grande, José Luis, Logan, Miriam, Sandro: Estas figuras tienen un número menor de menciones, lo que indica interacciones o menciones ocasionales en el contenido de Marta Díaz.
- Otros YouTubers y personalidades: La lista continúa con muchas otras figuras con solo 2 menciones, lo que sugiere que han sido mencionadas o han interactuado con Marta Díaz en contadas ocasiones.

En conclusión, este "network" nos da una idea de las interacciones y menciones de Marta Díaz con otros YouTubers y personalidades. Mientras que algunas figuras como Sergio tienen una presencia significativa, otras son mencionadas de manera más esporádica. Esto puede reflejar las

colaboraciones, amistades, influencias o simplemente temas de discusión en el contenido de Marta Díaz.

DjMario

EDA

	likes	views	comments	year	month	\
count	4.680000e+02	4.680000e+02	468.000000	468.000000	468.000000	
mean	7.135504e+04	1.376513e+06	1875.709402	2021.549145	6.425214	
std	9.147707e+04	2.035061e+06	3855.257294	1.044891	3.714196	
min	1.038900e+04	2.928160e+05	57.000000	2017.000000	1.000000	
25%	3.241650e+04	7.767550e+05	215.500000	2021.000000	3.000000	
50%	5.193250e+04	9.990520e+05	874.500000	2022.000000	6.000000	
75%	8.160425e+04	1.350598e+06	2136.000000	2022.000000	10.000000	
max	1.493243e+06	3.527614e+07	49689.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	468.000000	468.000000	468.000000			
mean	2.634615	16.151709	0.054715			
std	2.002418	7.231005	0.022851			
min	0.000000	0.000000	0.014718			
25%	1.000000	15.000000	0.038089			
50%	3.000000	19.000000	0.047676			
75%	4.000000	21.000000	0.069470			
max	6.000000	23.000000	0.129443			

Figura 46 Estadística DjMario

- Volumen de Datos: Se han analizado un total de 468 videos de DjMario.
- Interacción:
 - o Likes: En promedio, los videos de DjMario reciben aproximadamente 71,355 likes. El video más popular ha alcanzado 1,493,243 likes, mientras que el menos popular ha obtenido 10,389 likes.
 - o Vistas: Los videos tienen un promedio de 1,376,513 vistas, con un video alcanzando un máximo de 35,276,140 vistas y un mínimo de 292,816 vistas.
 - o Comentarios: En promedio, los videos reciben 1,875 comentarios, con un video que ha alcanzado un máximo de 49,689 comentarios y un mínimo de 57 comentarios.
- Tiempo:
 - o Año: Los videos se han publicado desde 2017 hasta 2023.
 - o Mes: El mes promedio de publicación es alrededor de junio.
 - o Día de la semana: La mayoría de los videos se publican a mediados de semana, con un promedio de 2.63 (donde 0 es lunes y 6 es domingo).
 - o Hora: La mayoría de los videos se publican alrededor de las 4:09 p.m.
- Tasa de Engagement: La tasa de engagement promedio es del 5.47%. Esto indica que, en promedio, el 5.47% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 12.94%, mientras que la más baja es del 1.47%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque DjMario tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

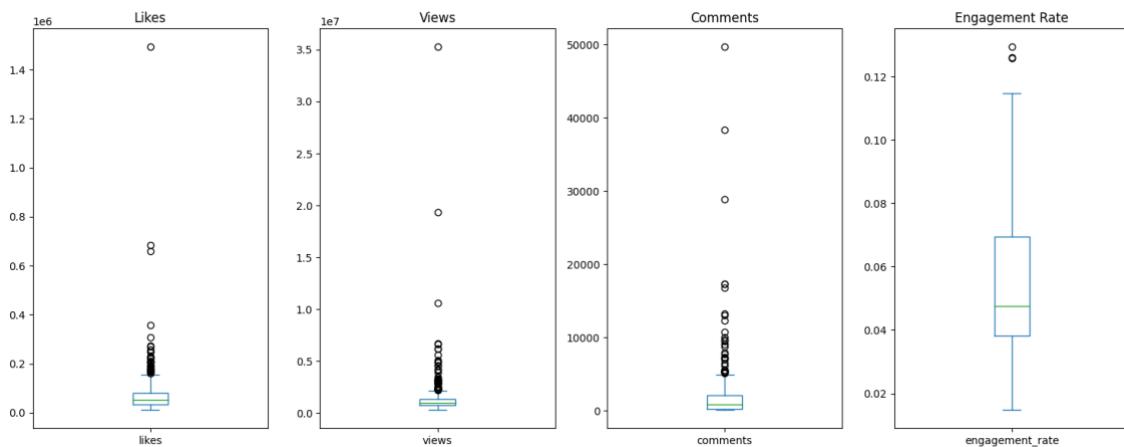


Figura 47 Boxplot DjMario

Los boxplots de DjMario revelan que, aunque mantiene una interacción constante en la mayoría de sus contenidos, existen ciertos videos que sobresalen y alcanzan niveles de interacción y visualización notablemente superiores al resto. Los valores atípicos presentes en cada boxplot señalan aquellos videos que han logrado un desempeño particularmente destacado en términos de likes, vistas, comentarios y tasa de engagement.

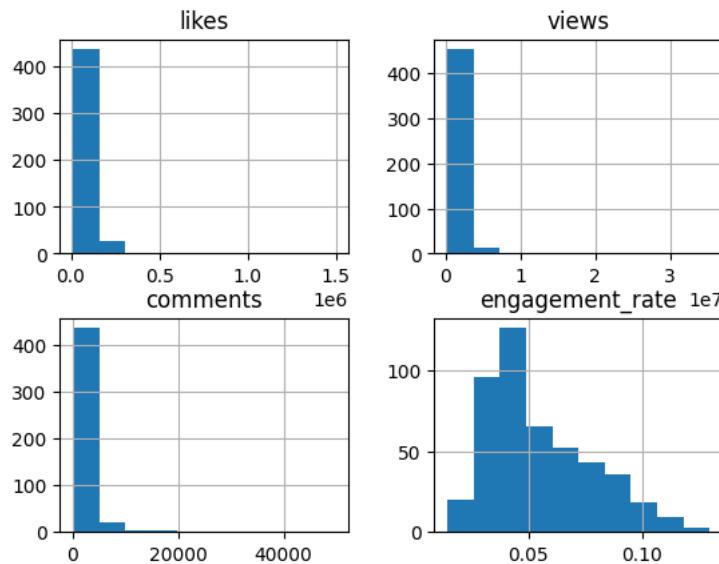


Figura 48 Histograma DjMario

Los histogramas de DjMario indican que la mayoría de sus videos tienden a acumular una cantidad moderada de "likes", "views" y "comments". Estos valores se concentran principalmente en el extremo izquierdo del histograma, lo que sugiere que la mayoría de sus videos obtienen una interacción estándar. Sin embargo, la presencia de colas largas hacia la derecha en estas distribuciones revela que hay ciertos videos que sobresalen y alcanzan niveles de interacción excepcionalmente altos. En cuanto a la tasa de engagement, la distribución muestra una variabilidad interesante, con algunos videos logrando capturar una atención y participación mucho mayor por parte de la audiencia.

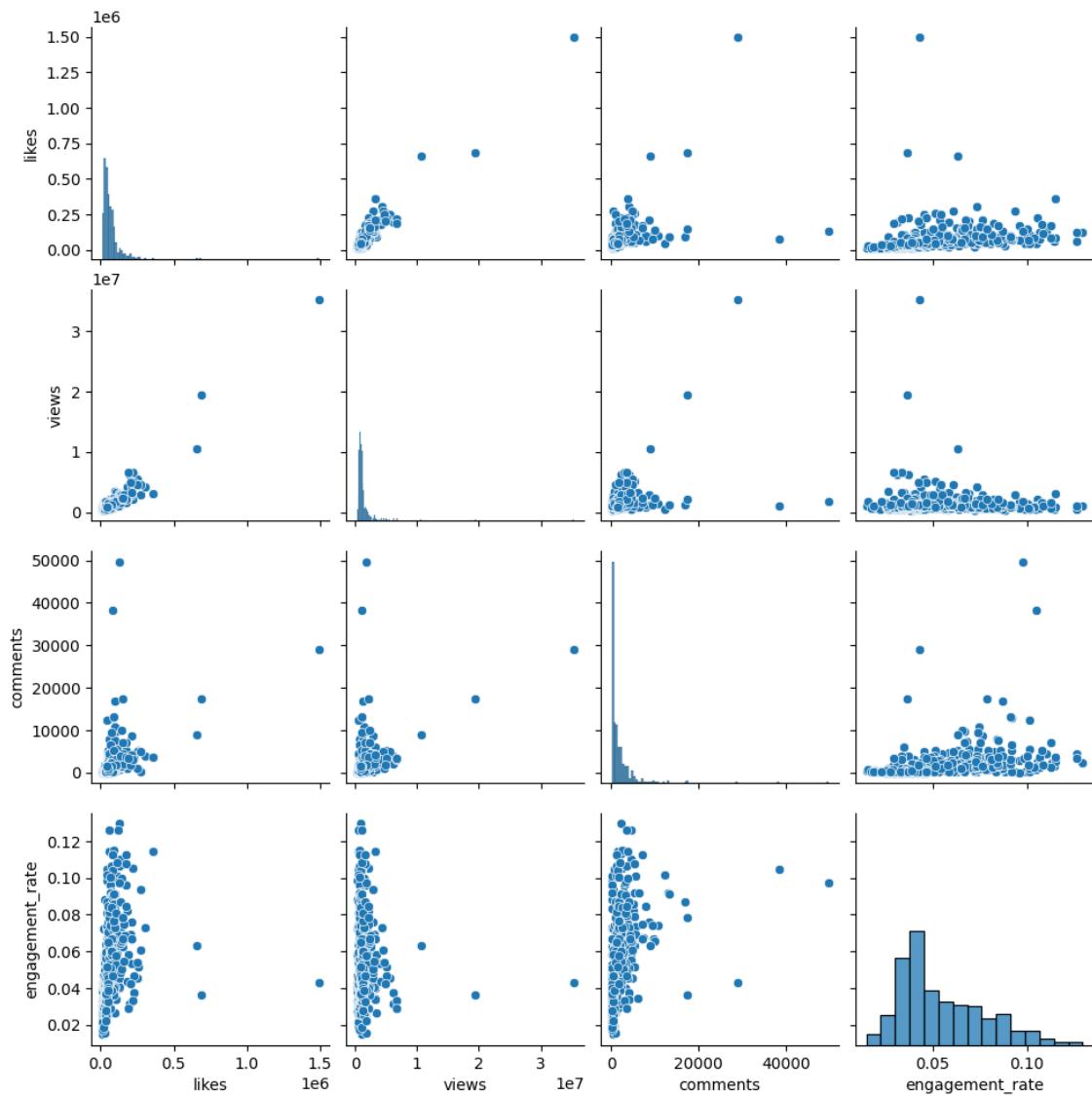


Figura 49 Sactterplot DjMario

En el scatterplot de DjMario, se observa una correlación positiva entre "likes" y "views", lo que sugiere que a medida que un video recibe más vistas, también tiende a recibir más "likes". Esta relación lineal indica que la popularidad de sus videos, en términos de visualizaciones, está estrechamente vinculada con la cantidad de interacciones positivas que recibe. Además, se pueden identificar varios valores atípicos que se desvían de esta tendencia general. Estos puntos representan videos que, aunque han obtenido un número inusualmente alto o bajo de vistas, no han seguido la tendencia general en términos de "likes". Estos outliers podrían ser indicativos de videos que generaron reacciones mixtas o controversiales entre la audiencia.

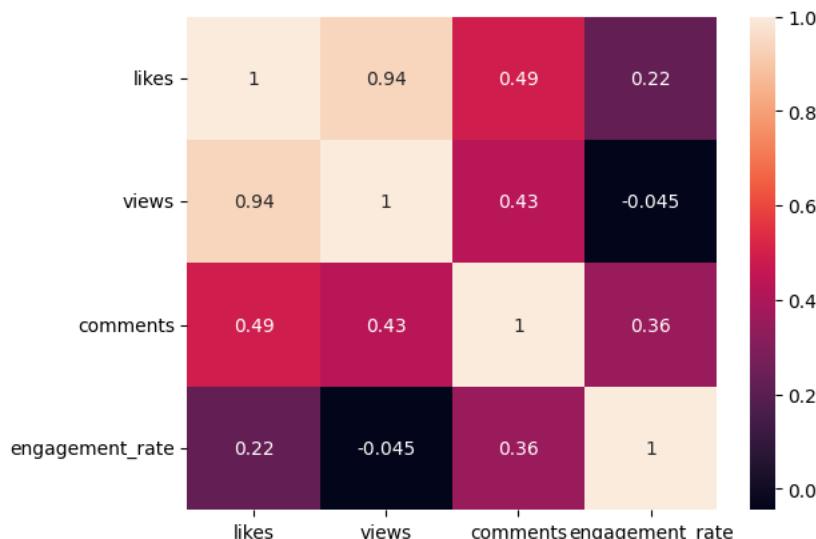


Figura 50 Matriz de correlación DjMario

En función de la matriz de correlación de DjMario, mientras que "likes" y "views" están estrechamente relacionados, la tasa de engagement no necesariamente sigue la misma tendencia, lo que podría indicar diferencias en cómo la audiencia interactúa con diferentes tipos de contenido.

Serie Temporal

- Análisis Mensual: Se observa que DjMario tuvo su mayor actividad antes de 2018, marcando un pico en su producción de contenido. Posteriormente, hubo un notable descenso en su actividad hasta 2019, año que representa el punto más bajo en términos de publicaciones. Sin embargo, desde ese momento, se evidencia una tendencia ascendente en su actividad, mostrando un crecimiento sostenido a lo largo de los años subsiguientes.

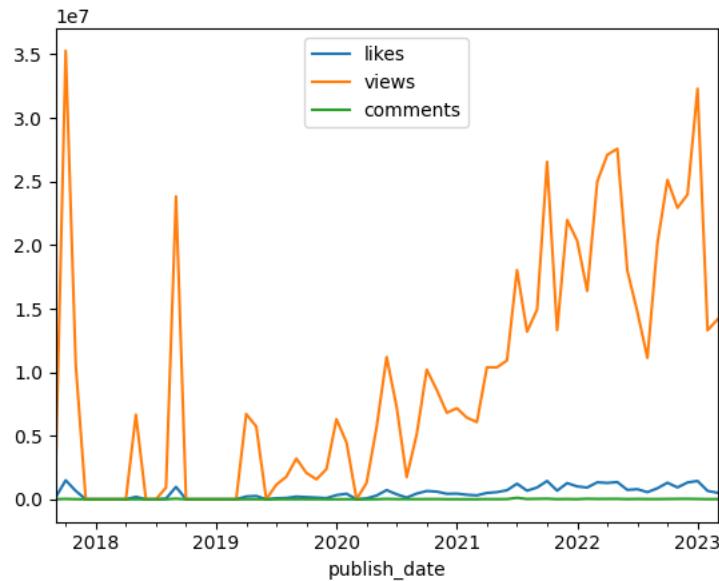


Figura 51 Serie Temporal mes DjMario

- Análisis Semanal: DjMario mantiene una frecuencia de publicación constante a lo largo de la semana. Aunque el lunes podría ser el día en que publica con más frecuencia, la diferencia no es significativamente mayor en comparación con otros días, lo que indica una distribución uniforme de su contenido a lo largo de los días de la semana.

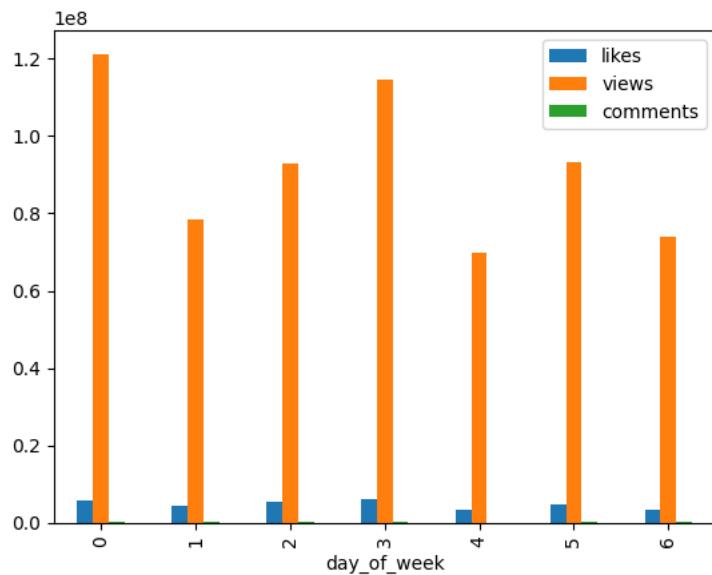


Figura 52 Serie Temporal semana DjMario

- Análisis Horario: En cuanto a las horas de publicación, DjMario muestra una clara preferencia por el horario vespertino y nocturno, específicamente entre las 17:00 y las 23:00 horas. La actividad disminuye considerablemente durante las horas diurnas, lo que sugiere que prefiere publicar su contenido cuando su audiencia está potencialmente más activa en la plataforma.

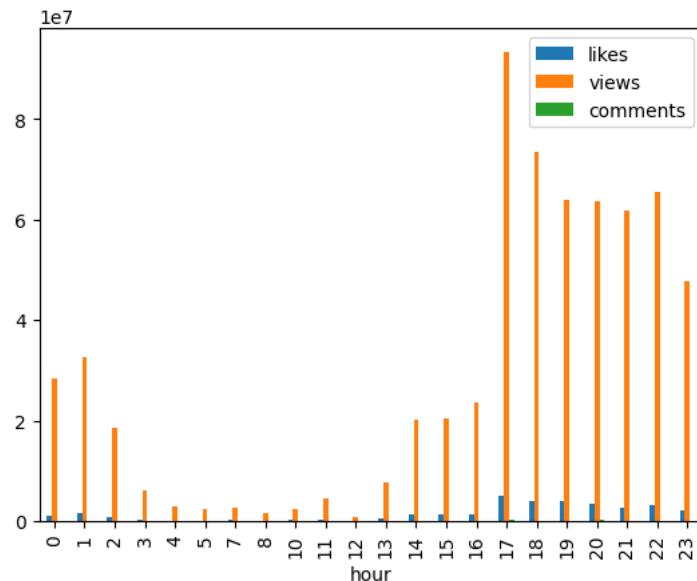


Figura 53 Serie Temporal hora DjMario

El video "DjMaRiiO vs CRISTIANO RONALDO" siendo el más popular, sugiere que la colaboración o interacción con figuras de renombre, en este caso Cristiano Ronaldo, uno de los futbolistas más famosos del mundo, puede generar un alto nivel de interés y engagement por parte de la audiencia. Este tipo de colaboraciones puede ser una estrategia efectiva para atraer a una audiencia más amplia, combinar seguidores de ambos participantes y potenciar la visibilidad del contenido. Es un claro ejemplo de cómo la combinación de personalidades influyentes en un video puede resultar en un éxito significativo en términos de visualizaciones y interacción.

Minería de Textos



Figura 54 Wordcloud título DjMario

La nube de palabras para DjMario revela sus temas y enfoques principales en su contenido. La palabra "djmariio" siendo la más frecuente, indica que su marca personal es fuertemente promocionada en sus videos. "FIFA" aparece como el segundo término más frecuente, lo que sugiere que gran parte de su contenido está relacionado con este popular videojuego de fútbol. Las palabras "vs", "carrera", "kings" y "episodio" podrían indicar que realiza enfrentamientos o competencias, quizás ligas o torneos dentro del juego. "Real Madrid" destaca como un equipo de fútbol de interés, lo que podría sugerir que es su equipo favorito o que realiza contenido específico relacionado con ellos. Finalmente, términos como "historia", "directo" y "champions" sugieren que también podría estar haciendo transmisiones en vivo o discutiendo eventos actuales en el mundo del fútbol, especialmente en la Liga de Campeones. En general, la nube de palabras pinta una imagen de DjMario como un creador de contenido centrado en el fútbol y, en particular, en el videojuego FIFA.



Figura 55 Wordcloud descripción DjMario

La nube de palabras de la descripción de DjMario proporciona una visión clara de los temas y enfoques predominantes en su canal. "FIFA" es el término más destacado, lo que confirma que el videojuego de fútbol es central en su contenido. Su marca personal, "djmarioo", también es prominente, subrayando la importancia de su identidad en la plataforma. Las menciones de "canal", "twitter" e "instagram" sugieren que promociona activamente sus otros medios y plataformas sociales.

El término "vs" indica que los enfrentamientos o comparaciones son un formato recurrente en su contenido. Las referencias a "Madrid", "Real", "Messi", "Ronaldo" y "Cristiano" demuestran un fuerte enfoque en el fútbol profesional, especialmente en los jugadores de élite y en el Real Madrid. La presencia de "Messi" y "Ronaldo" juntas, junto con "Cristiano", sugiere que puede haber debates o comparaciones entre estos dos titanes del fútbol en su canal. Además, la palabra "mejor" puede estar relacionada con discusiones sobre quién es el mejor jugador o equipo. En resumen, la nube de palabras revela que DjMario es un apasionado del fútbol y del videojuego FIFA, y utiliza su plataforma para discutir y debatir sobre los aspectos más destacados y controvertidos del deporte.



Figura 56 Wordcloud transcripción DjMario

En función de las transcripciones, nos ofrecen una visión más clara de su estilo de contenido. La frecuente aparición de "música" y "aplauso" indica que DjMario utiliza estos elementos sonoros para mejorar la experiencia del espectador, ya sea para enfatizar momentos destacados o para crear una atmósfera específica en sus videos.

El término "chico" sigue siendo prominente, lo que sugiere que DjMario se comunica directamente con su audiencia o hace referencia a otros jugadores o personajes. La presencia de "xxxx" como representación de insultos muestra que, aunque puede haber momentos de intensidad o frustración durante los juegos, estos términos son censurados, lo que podría ser un esfuerzo por mantener el contenido adecuado para una audiencia más amplia.

Los términos relacionados con el fútbol y los videojuegos, como "partido", "jugar" y "equipo", confirman el enfoque temático de su canal. "Tío", siendo una expresión coloquial, refleja un tono amigable e informal.

En resumen, las transcripciones revelan que DjMario crea un ambiente enriquecido con sonidos ambientales para sumergir a su audiencia en la experiencia del juego, mientras mantiene un tono directo y amigable. A pesar de los momentos de intensidad, se esfuerza por ofrecer un contenido adecuado al censurar los insultos.

Las frases comunes en las transcripciones de DjMario proporcionan una visión detallada de su estilo de comunicación y las emociones que transmite en sus videos.

```

Common Phrases:
uy uy uy: 1441
cara sudar cara: 1007
sudar cara sudar: 976
ay ay ay: 794
música música música: 530
mira mira mira: 471
valer valer valer: 376
aplauso aplauso aplauso: 342
aplauso música aplauso: 238
gol gol gol: 225
cuidado cuidado cuidado: 223
parís saint germain: 217
chico chico chico: 215
música aplauso música: 202
cara payaso cara: 185
cara llorar risa: 184
payaso cara payaso: 184
risa cara llorar: 152
sonriente cara sonriente: 151
xxxx xxxx xxxx: 148

```

Figura 57 Frases comunes DjMario

- Repetición para énfasis: Frases como "uy uy uy", "ay ay ay", "mira mira mira", y "valer valer valer" sugieren momentos de sorpresa, emoción o énfasis. Estas repeticiones pueden ser reacciones espontáneas a eventos inesperados o emocionantes en el juego.
- Emojis y expresiones faciales: Las referencias a "cara sudar", "cara payaso", y "cara llorar risa" indican que DjMario utiliza emojis o describe expresiones faciales para transmitir emociones y reacciones a sus espectadores. Estas menciones pueden ser una forma de conectar con la audiencia joven que está familiarizada con el lenguaje de los emojis.
- Sonidos ambientales: Las repetidas menciones de "música música música" y "aplauso aplauso aplauso" reflejan la incorporación de efectos sonoros y música para mejorar la experiencia del espectador.
- Reacciones al juego: Frases como "gol gol gol" y "cuidado cuidado cuidado" sugieren momentos emocionantes o tensos en el juego, probablemente relacionados con el fútbol dada la temática del canal.
- Equipos y jugadores: La mención de "parís saint germain" indica que DjMario comenta o juega partidos relacionados con este equipo, lo que refleja su interés en ligas y equipos de fútbol internacionales.
- Censura: La frase "xxxx xxxx xxxx" confirma la censura de ciertas palabras, posiblemente insultos, para mantener el contenido adecuado para una audiencia más amplia.

En conjunto, estas frases comunes pintan un retrato de DjMario como un creador de contenido apasionado y expresivo, que utiliza una variedad de herramientas lingüísticas y sonoras para involucrar y entretener a su audiencia.

Los tópicos derivados de las transcripciones de DjMario ofrecen una comprensión profunda de los temas y el contenido que aborda en sus videos:

```

Topics:
Topic 0, Words: 0.013*"aplauso" + 0.009*"chico" + 0.007*"jugar" + 0.007*"partido" + 0.007*"ver" + 0.006*"música" + 0.005*"pasar" + 0.005*"madrid" + 0.005*"valer" + 0.004*"poner"
Topic 1, Words: 0.016*"partido" + 0.010*"jugar" + 0.009*"gol" + 0.009*"balón" + 0.009*"ver" + 0.008*"chico" + 0.008*"equipo" + 0.008*"ganar" + 0.007*"madrid" + 0.007*"música"
Topic 2, Words: 0.011*"jugador" + 0.011*"ver" + 0.010*"decir" + 0.010*"equipo" + 0.008*"jugar" + 0.007*"valer" + 0.007*"querer" + 0.006*"eh" + 0.006*"xxxx"
Topic 3, Words: 0.018*"chico" + 0.014*"xxxx" + 0.014*"aplauso" + 0.012*"música" + 0.011*"salir" + 0.009*"vale" + 0.008*"ver" + 0.008*"poner" + 0.007*"partido"
Topic 4, Words: 0.010*"música" + 0.010*"xxxx" + 0.009*"chico" + 0.009*"ver" + 0.007*"valer" + 0.007*"decir" + 0.006*"pasar" + 0.006*"salir" + 0.006*"venir" + 0.006*"madrid"

```

Figura 58 Topic Modeling DjMario

- Tópico 0 - Reacciones y Comentarios: Las palabras "aplauso", "chico", "jugar", "partido" y "ver" sugieren que este tema se centra en las reacciones y comentarios generales de DjMario durante el juego. La mención de "madrid" podría indicar discusiones o partidas relacionadas con el equipo Real Madrid.

- Tópico 1 - Análisis de Juego: Este tópico parece centrarse en el análisis detallado del juego con palabras como "partido", "jugar", "gol", "balón" y "equipo". DjMario podría estar comentando sobre la dinámica del juego, los goles marcados y la estrategia del equipo.
- Tópico 2 - Discusión sobre Jugadores: Las palabras "jugador", "equipo", "decir" y "querer" sugieren que DjMario podría estar discutiendo sobre jugadores específicos, sus habilidades y posiblemente rumores o noticias relacionadas con ellos. La palabra "xxxx" indica censura, lo que sugiere que podría estar expresando opiniones fuertes o controvertidas.
- Tópico 3 - Interacciones y Emociones: Con palabras como "chico", "tío", "xxxx", "aplauso" y "salir", este tópico refleja las emociones y reacciones de DjMario durante el juego. Podría estar expresando sorpresa, frustración o emoción en diferentes momentos.
- Tópico 4 - Comentarios Generales: Este tópico parece ser una mezcla de comentarios generales y observaciones con palabras como "música", "xxxx", "ver", "valer" y "madrid". DjMario podría estar compartiendo sus pensamientos sobre el juego, la música de fondo y otros aspectos del video.

En conclusión, estos tópicos reflejan la diversidad de contenido en los videos de DjMario, desde análisis detallados y discusiones sobre jugadores hasta reacciones emocionales y comentarios generales sobre el juego.

Los términos únicos proporcionan una visión valiosa de las palabras que distinguen a DjMario de otros creadores de contenido:

```
Unique Terms:
aplauso: 35.193068630356606
xxxx: 34.95071398165938
tío: 25.882557814203096
gol: 24.109064414052934
eh: 18.81053914916156
madrid: 18.19795509585576
directo: 16.988244961850068
gente: 16.8232268853536
año: 16.62127612244177
mario: 16.284309636157957
```

Figura 59 Términos únicos DjMario

- aplauso: Este término tiene la puntuación más alta, lo que sugiere que DjMario utiliza frecuentemente efectos de sonido de aplausos o hace referencia a ellos en sus videos, posiblemente para enfatizar momentos destacados o logros.
- xxxx: Este término, que representa palabras censuradas o insultos, indica que DjMario podría tener un estilo de comunicación directo y sin filtros, lo que podría ser atractivo para una audiencia que valora la autenticidad.
- tío: Es una palabra coloquial en español que se utiliza a menudo para referirse a alguien de manera amigable. Su uso frecuente sugiere un tono informal y cercano en sus videos.
- gol: La presencia de esta palabra refleja el enfoque de DjMario en los videojuegos relacionados con el fútbol o sus comentarios sobre partidos reales.
- eh: Esta interjección podría ser utilizada por DjMario para llamar la atención o enfatizar un punto.
- madrid: La mención frecuente de "Madrid" podría indicar una afinidad o enfoque en el equipo de fútbol Real Madrid o simplemente una referencia a la ciudad.
- directo: Este término sugiere que DjMario realiza transmisiones en vivo o "streamings" para interactuar con su audiencia en tiempo real.
- gente: La referencia a "gente" indica que DjMario se dirige a menudo a su audiencia o habla sobre grupos de personas en sus videos.
- año: La mención de "año" podría estar relacionada con recuentos anuales, aniversarios de canal o eventos significativos que ocurrieron en un año específico.

- mario: Aunque es similar a su nombre de usuario "DjMaRiiO", la mención de "mario" podría estar relacionada con el videojuego "Super Mario" o simplemente una variante de su nombre.

Network

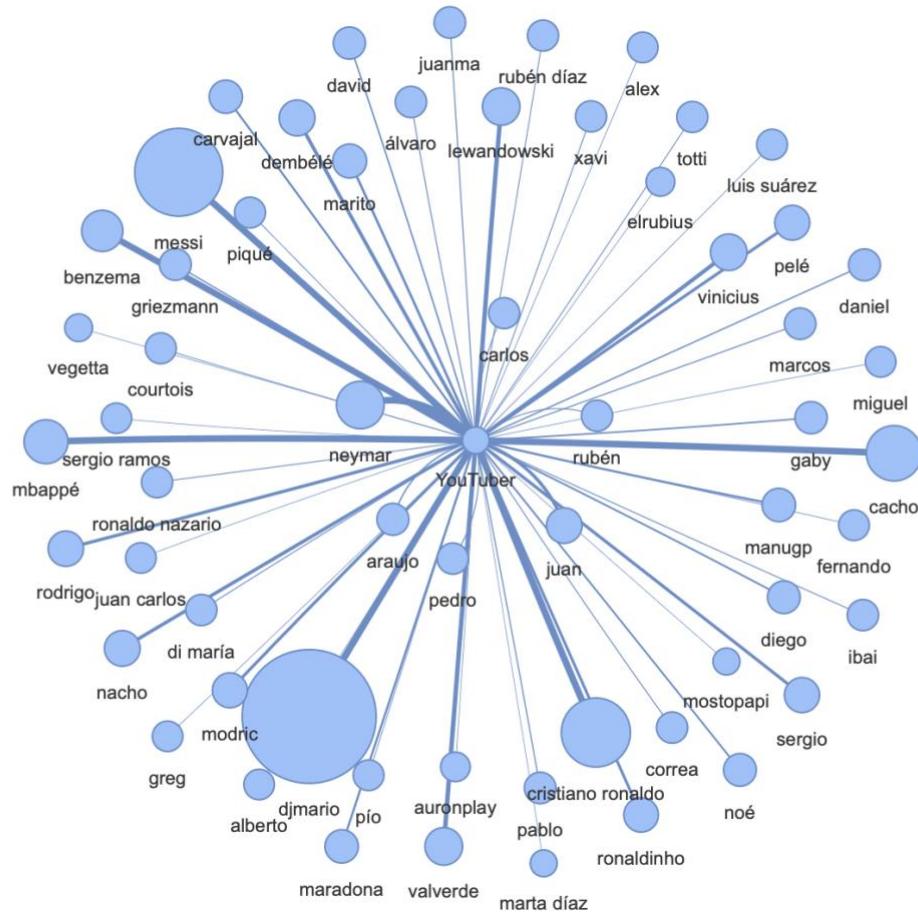


Figura 60 Network DjMario

En el análisis de la red de interacciones relacionadas con DjMario, se observan patrones interesantes que reflejan tanto su enfoque en el contenido como las conexiones y menciones que establece:

- Predominio del Fútbol: La mayoría de los nombres en la lista son de futbolistas famosos como Messi, Cristiano Ronaldo, Neymar, Mbappé, Benzema, entre otros. Esto sugiere que DjMario tiene un fuerte enfoque en el fútbol, ya sea comentando partidos, jugando videojuegos relacionados con el fútbol o discutiendo sobre jugadores y equipos.
- DjMario en el Centro: Con 6122 menciones, DjMario es, con diferencia, el más mencionado, lo que indica que se refiere a sí mismo con frecuencia, reforzando su marca personal y presencia en sus contenidos.
- Otros Creadores de Contenido: Además de los futbolistas, hay nombres como Cacho, Ibai, y Marito que podrían ser otros creadores de contenido o personalidades con las que DjMario ha colaborado o mencionado en sus videos.
- Diversidad de Nombres: Aunque hay una clara predominancia de futbolistas, también hay una variedad de otros nombres, lo que sugiere que DjMario no se limita únicamente al fútbol, sino que aborda otros temas o interactúa con una variedad de personalidades.
- Menciones de Clubes y Jugadores: La presencia de nombres como Valverde, Vinicius, Dembelé, y Sergio Ramos sugiere que DjMario podría estar comentando sobre clubes

específicos, especialmente el Real Madrid y el FC Barcelona, dada la prominencia de estos jugadores en dichos equipos.

En resumen, la red de interacciones de DjMario refleja su enfoque en el fútbol y su interacción con otras personalidades, ya sean futbolistas o creadores de contenido. Su contenido parece ser una mezcla de comentarios sobre partidos, jugadores, y colaboraciones con otros youtubers.

Clakovi

EDA

	likes	views	comments	year	month	\
count	95.000000	9.500000e+01	95.000000	95.000000	95.000000	
mean	9829.031579	2.980446e+05	309.378947	2019.115789	7.000000	
std	8167.668481	4.423096e+05	302.564155	1.335785	3.566988	
min	841.000000	3.083800e+04	0.000000	2017.000000	1.000000	
25%	5084.500000	1.172515e+05	141.000000	2018.000000	4.000000	
50%	6850.000000	1.754350e+05	225.000000	2019.000000	7.000000	
75%	12084.000000	2.818605e+05	374.000000	2020.000000	10.000000	
max	51841.000000	3.648953e+06	2230.000000	2021.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	95.000000	95.000000	95.000000			
mean	3.978947	15.926316	0.042466			
std	2.365438	3.265689	0.011130			
min	0.000000	9.000000	0.009735			
25%	2.000000	13.000000	0.037297			
50%	6.000000	15.000000	0.042359			
75%	6.000000	19.000000	0.048582			
max	6.000000	22.000000	0.091413			

Figura 61 Estadística Clakovi

En el caso de Clakovi, basándonos en la estadística de sus datos:

- Volumen de Datos: Se han analizado un total de 95 videos de Clakovi, en comparación con otros youtubers de la lista no hay tantos registros, pero se debe analizar por su amplitud de categorías de influencer.
- Interacción:
 - o Likes: En promedio, los videos de Clakovi reciben aproximadamente 9,829 likes. El video más popular ha alcanzado 51,841 likes, mientras que el menos popular ha obtenido 841 likes.
 - o Vistas: Los videos tienen un promedio de 298,044 vistas, con un video alcanzando un máximo de 3,648,953 vistas y un mínimo de 30,838 vistas.
 - o Comentarios: En promedio, los videos reciben 309 comentarios, con un video que ha alcanzado un máximo de 2,230 comentarios y un mínimo de 0 comentarios.
- Tiempo:
 - o Año: Los videos se han publicado desde 2017 hasta 2021.
 - o Mes: El mes promedio de publicación es alrededor de julio.
 - o Día de la semana: La mayoría de los videos se publican los domingos, con un promedio de 3.97 (donde 0 es lunes y 6 es domingo).
 - o Hora: La mayoría de los videos se publican alrededor de las 3:55 p.m.

- Tasa de Engagement: La tasa de engagement promedio es del 4.25%. Esto indica que, en promedio, el 4.25% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 9.14%, mientras que la más baja es del 0.97%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Clakovi tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

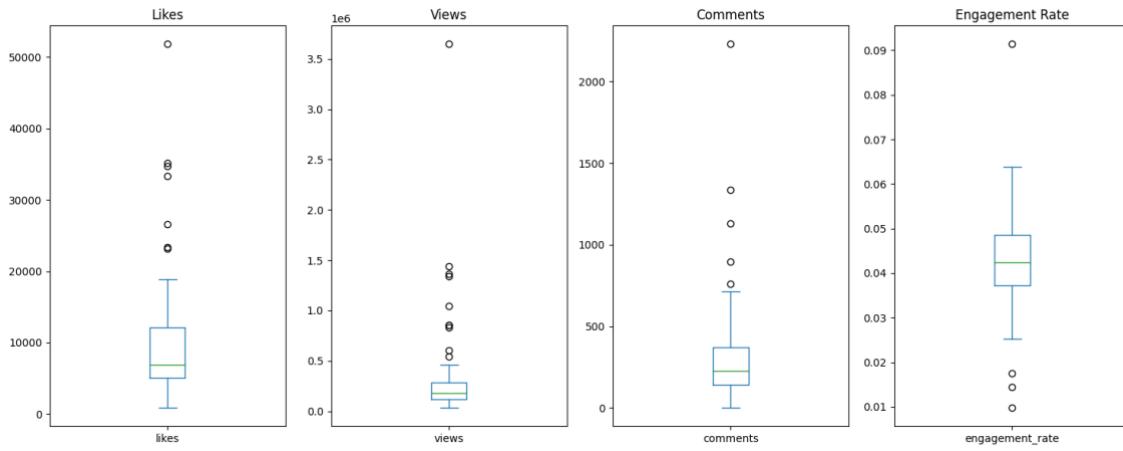


Figura 62 Boxplot Clakovi

Los boxplots de Clakovi revelan que, aunque tiene una base consistente de interacción en la mayoría de sus videos, existen ciertos contenidos que sobresalen y generan una interacción y visualización notablemente superior al resto. Los valores atípicos en cada boxplot señalan videos que han tenido un rendimiento excepcionalmente alto en términos de likes, vistas y comentarios. Además, la variabilidad en los datos, reflejada en la amplitud de las cajas y las colas de los boxplots, sugiere que, si bien Clakovi tiene un nivel medio de interacción, hay videos que son particularmente populares y otros que, en comparación, no han captado tanto la atención de la audiencia.

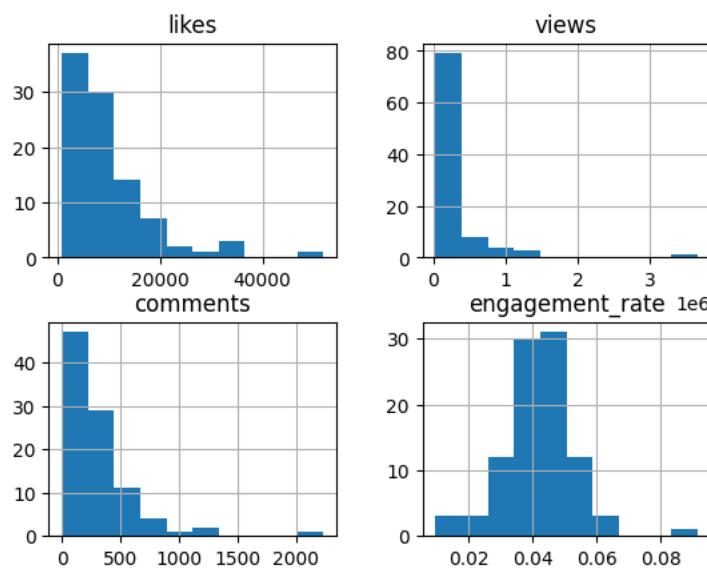


Figura 63 Histograma Clakovi

Los histogramas de Clakovi revelan una distribución sesgada hacia la derecha en las métricas de "likes", "vistas" y "comentarios". Esto sugiere que la mayoría de sus videos obtienen una cantidad moderada de interacción, pero hay un conjunto selecto de videos que logran una interacción

excepcionalmente alta. En cuanto a la tasa de compromiso, la distribución parece ser más uniforme, aunque con un ligero sesgo hacia la derecha, indicando que la mayoría de los videos de Clakovi mantienen tasas de compromiso similares, pero algunos sobresalen con tasas más altas. En conjunto, los histogramas reflejan una variabilidad en la interacción de la audiencia con el contenido de Clakovi, con ciertos videos que claramente resuenan más con los espectadores.

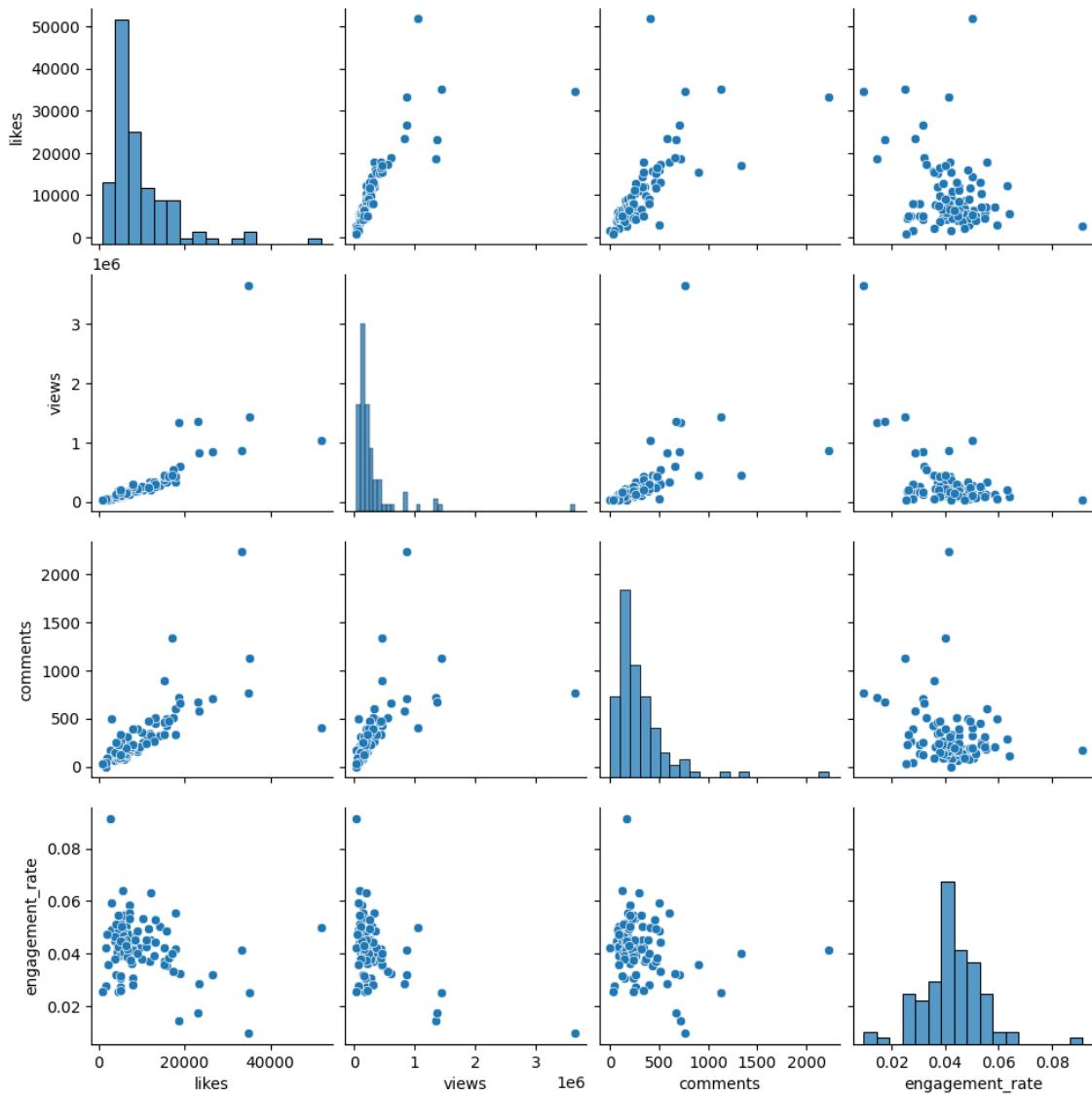


Figura 64 Scatterplot Clakovi



Figura 65 Matriz correlación Clakovi

Analizando el scatterplot y sus correlaciones entre dichas variables, mientras que "likes", "vistas" y "comentarios" tienden a moverse juntos en la misma dirección, la tasa de engagement muestra una tendencia opuesta, especialmente en relación con las "vistas". Esto sugiere que aunque los videos de Clakovi pueden ser populares en términos de vistas, no todos los espectadores están interactuando activamente con el contenido.

Serie Temporal

- Análisis Mensual:** La actividad de Clakovi muestra fluctuaciones a lo largo de los meses, con un pico notable en el verano de 2019. Esto podría indicar una campaña o serie de videos que tuvieron especial relevancia en ese período. Sin embargo, es notable que no hay datos después de 2021, lo que podría indicar un hiato en la producción de contenido o un cambio en la plataforma de publicación.

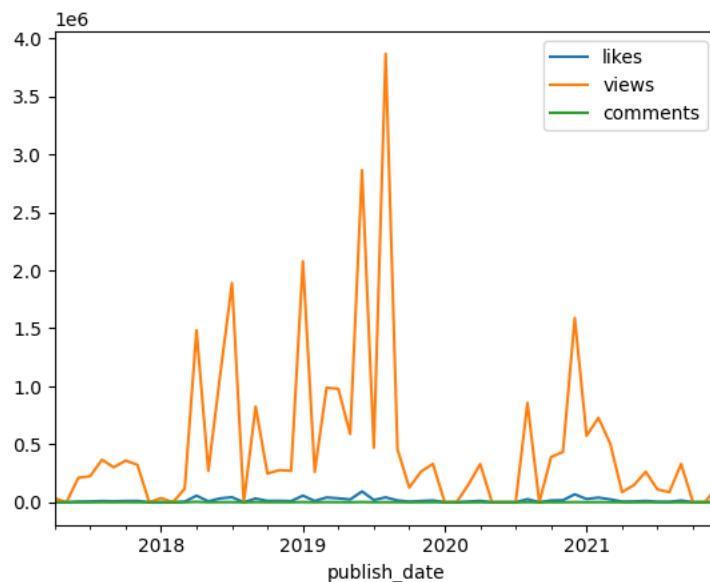


Figura 66 Serie Temporal mes Clakovi

- Análisis Semanal:** La tendencia de Clakovi a publicar más los domingos podría estar relacionada con patrones de consumo de su audiencia. Los domingos, las personas suelen tener más tiempo

libre, lo que podría traducirse en mayores vistas y engagement. Es esencial considerar cómo esta tendencia semanal se relaciona con las métricas de engagement a lo largo del tiempo.

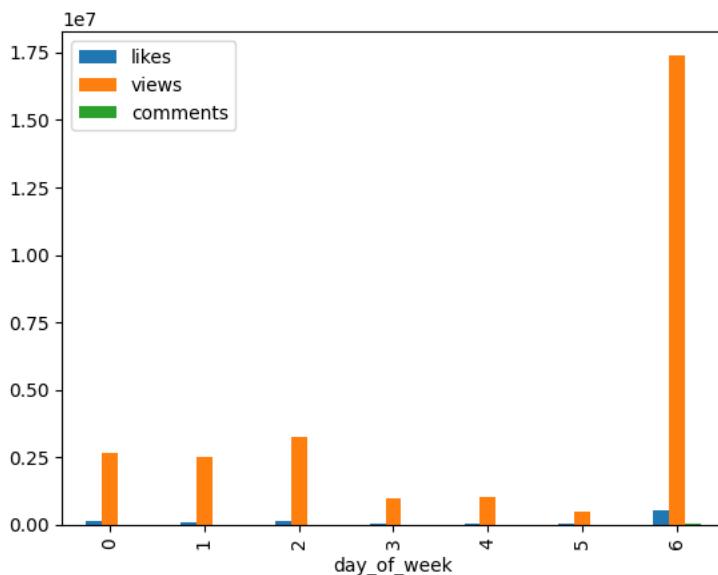


Figura 67 Serie Temporal semana Clakovi

- **Análisis Horario:** La preferencia por publicar alrededor de las 13 horas podría estar alineada con los hábitos de consumo de su audiencia. Publicar a esta hora podría coincidir con momentos de descanso o pausas laborales, maximizando así la visibilidad.

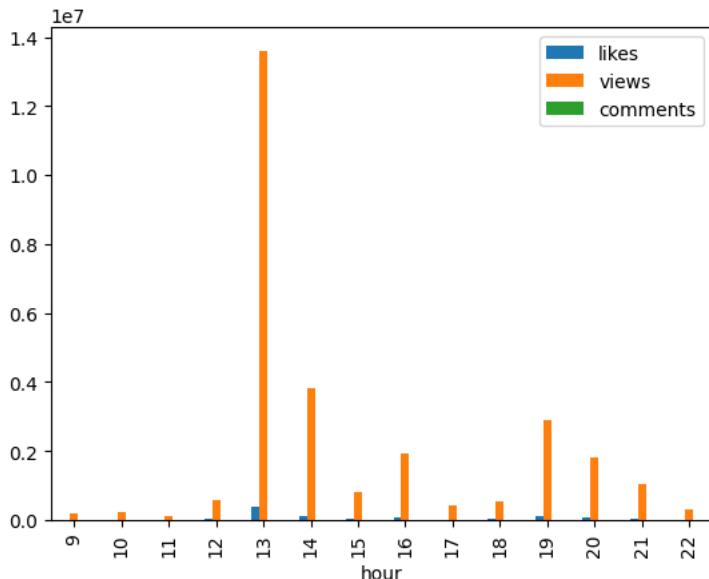


Figura 68 Serie Temporal hora Clakovi

El video más popular de Clakovi, "summer morning routine | mi rutina de mañana...", parece ser un contenido relacionado con las actividades diarias que Clakovi realiza durante las mañanas de verano. Estos tipos de videos, conocidos como "routine videos", son populares en plataformas como YouTube porque ofrecen una visión íntima de la vida diaria de los creadores, permitiendo a los espectadores conectarse de manera más personal con ellos. Además, las rutinas de mañana suelen incluir hábitos, consejos de belleza, alimentación, ejercicio, entre otros, que pueden ser de interés para una amplia audiencia.

El hecho de que este video sea el más popular sugiere que los seguidores de Clakovi están particularmente interesados en su vida personal y en cómo se prepara para comenzar su día,

especialmente durante la temporada de verano. Es posible que el video contenga consejos útiles, productos recomendados o simplemente una narrativa relajante que haya resonado con su audiencia. Es una excelente oportunidad para Clakovi de considerar crear más contenido similar o series estacionales, dada la respuesta positiva de su comunidad.

Minería de Textos



Figura 69 Wordcloud título Clakovi

La nube de palabras de los títulos de Clakovi revela que su contenido se centra en su marca personal, siendo "clakovi" el término más recurrente. Además, se destaca su interés por la moda, con términos como "haul", "ropa", "brandy", "melville" y "subdued", sugiriendo que realiza videos mostrando y comentando sus compras recientes. También se observa una inclinación hacia el contenido personal y relacional, con palabras como "vlog", "pregunta", "respuesta" y "amiga". La mención de "madrid" podría indicar que comparte experiencias o lugares específicos de esta ciudad, mientras que "verano" señala una adaptación de su contenido según la estacionalidad.

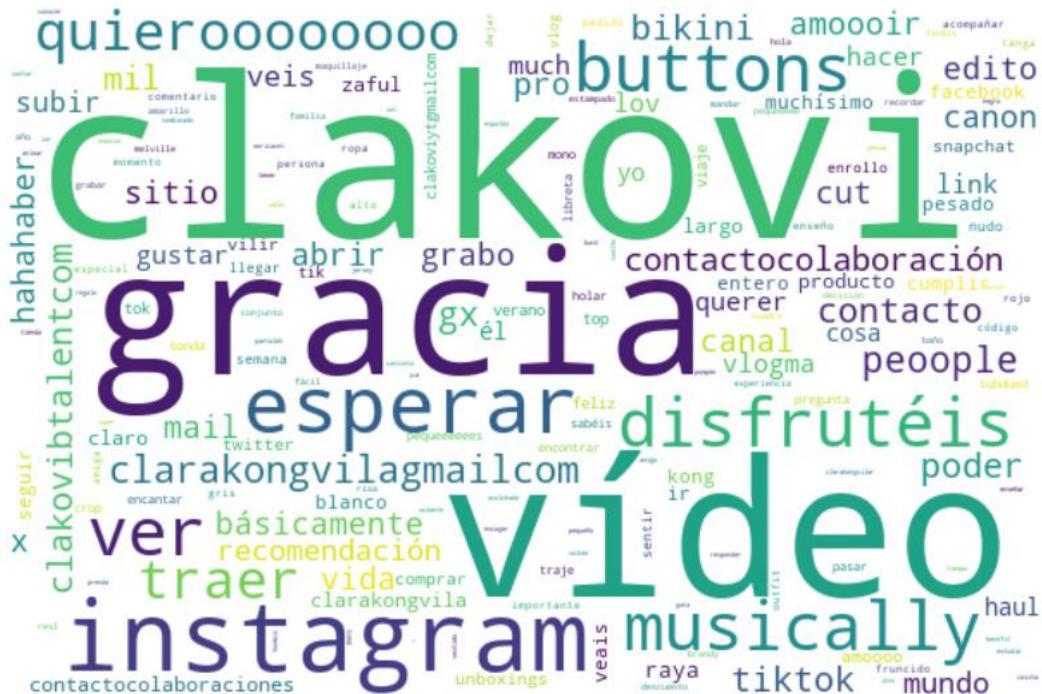


Figura 70 Wordcloud descripción Clakovi

La nube de palabras de las descripciones de Clakovi muestra una fuerte presencia de su marca personal, con "clakovi" como el término más frecuente. Esto indica que ella enfatiza su identidad en sus descripciones. La palabra "vídeo" sugiere que a menudo se refiere o menciona otros contenidos que ha creado. Las palabras "gracias" y "gracia" reflejan una actitud de agradecimiento hacia su audiencia, lo que puede ser una estrategia para fomentar la lealtad y la conexión con sus seguidores. La presencia de "instagram", "musically" y "tiktok" indica que Clakovi está activa en múltiples plataformas y probablemente promociona o comparte contenido entre ellas. "Disfrutéis" y "quieroxxxxxx" sugieren un tono amigable y cercano en sus comunicaciones, mientras que "buttons" y "peopple" podrían estar relacionados con colaboraciones o promociones específicas. La mención de "clarakongvilagmailcom" indica que proporciona un medio de contacto directo para sus seguidores o posibles colaboraciones comerciales.



Figura 71 Wordcloud transcripción Clakovi

En función de las transcripciones, es evidente que la música juega un papel importante en sus contenidos, siendo "música" la palabra más frecuente. Esto puede indicar que utiliza música de fondo o discute sobre temas musicales en sus videos. Palabras como "cosa", "vídeo", "gustar", "súper" y "querer" sugieren que Clakovi tiene un enfoque conversacional y expresivo en sus videos, compartiendo sus opiniones y sentimientos sobre diversos temas. La frecuencia de palabras como "encantar", "valer" y "muchísimo" refleja un tono positivo y entusiasta. Además, términos como "comprar", "poner" y "quedan" pueden indicar que a menudo habla sobre compras, moda o cómo ciertos estilos o productos le quedan o le sientan. La mención del "año" sugiere que también podría hacer referencias temporales o hablar sobre eventos y experiencias pasadas. En general, las transcripciones revelan que Clakovi tiene un enfoque personal y positivo en sus videos, compartiendo sus experiencias y opiniones con su audiencia.

Common Phrases:
música música música: 38
traer vídeo canal: 30
esperar gustar muchísimo: 20
música aplauso música: 16
música ah música: 14
muchísimo muchísimo muchísimo: 14
música i música: 13
súper súper súper: 11
esperar gustar vídeo: 11
vídeo canal veis: 10
encantar encantar encantar: 10
vídeo canal traer: 9
gustar muchísimo muchísimo: 9
música músico música: 8
gustar muchísimo vídeo: 8
music music music: 8
querer muchísimo muchísimo: 7
ah música música: 6
habéis visto vídeo: 6
música música ah: 6

Figura 72 Frases comunes Clakovi

Las frases comunes en las transcripciones de Clakovi reflejan una combinación de expresiones relacionadas con la música y comentarios personales sobre sus videos. La repetición de la palabra

"música" en varias combinaciones sugiere que Clakovi hace pausas frecuentes para la música o comenta sobre ella en sus videos. Las frases "traer vídeo canal" y "vídeo canal veis" indican que a menudo se refiere a sus contenidos anteriores o futuros, incentivando a sus espectadores a ver más de su canal. Las expresiones "esperar gustar muchísimo" y "gustar muchísimo vídeo" reflejan su deseo de que a su audiencia le agrade lo que presenta. Además, la repetición de palabras como "súper", "muchísimo" y "encantar" en triples secuencias enfatiza su entusiasmo y pasión por lo que comparte. En general, estas frases comunes destacan el estilo entusiasta y musical de Clakovi, así como su interacción constante con su audiencia.

Los tópicos extraídos de las transcripciones de Clakovi ofrecen una visión detallada de los temas recurrentes en sus videos:

```
Topics:
Topic: 0, Words: 0.028*música" + 0.016*cosa" + 0.010*vídeo" + 0.009*querer" + 0.008*súper" + 0.007*año" + 0.006*gustar" + 0.006*ver" + 0.006*encantar"
Topic: 1, Words: 0.014*gustar" + 0.013*cosa" + 0.012*querer" + 0.011*vídeo" + 0.010*estudiar" + 0.009*tener" + 0.009*vivir" + 0.009*vida" + 0.008*persona" + 0.008*arte"
Topic: 2, Words: 0.017*súper" + 0.013*gustar" + 0.012*comprar" + 0.012*muchísimo" + 0.011*encantar" + 0.011*blanco" + 0.011*quedar" + 0.010*bikini" + 0.009*vídeo" + 0.008*negro"
Topic: 3, Words: 0.016*foto" + 0.008*vídeo" + 0.007*gente" + 0.007*music" + 0.005*poner" + 0.005*poder" + 0.005*valer" + 0.005*quedar" + 0.005*ejemplo" + 0.005*cámara"
Topic: 4, Words: 0.011*vídeo" + 0.010*cosa" + 0.008*aplicación" + 0.008*música" + 0.007*foto" + 0.007*valer" + 0.007*poner" + 0.006*súper" + 0.006*quedar"
```

Figura 73 Topic Modeling Clakovi

- Tópico 0: Este tópico se centra en la música y las reacciones generales hacia diferentes contenidos. Las palabras como "música", "cosa", "vídeo" y "gustar" sugieren que Clakovi comparte y comenta sobre diferentes temas musicales y da su opinión sobre ellos.
- Tópico 1: Aquí, el enfoque parece ser más introspectivo y personal. Palabras como "estudiar", "vivir", "vida" y "arte" indican que Clakovi podría estar compartiendo experiencias personales, reflexiones sobre la vida y posiblemente su relación con el arte.
- Tópico 2: Este tópico sugiere un enfoque en la moda y la estética. Las palabras "súper", "comprar", "muchísimo", "encantar", "blanco", "bikini" y "negro" indican que Clakovi podría estar hablando de compras, outfits y tendencias de moda.
- Tópico 3: Parece estar relacionado con la fotografía y la creación de contenido. Las palabras "foto", "gente", "music" y "cámara" sugieren que Clakovi comparte consejos o experiencias relacionadas con la fotografía y la producción de videos.
- Tópico 4: Este tópico combina elementos de los tópicos anteriores, con un enfoque en aplicaciones, música y estética. Las palabras "aplicación", "foto" y "súper" sugieren que Clakovi podría estar revisando o recomendando aplicaciones, posiblemente relacionadas con la edición de fotos o música.

En general, estos tópicos reflejan la diversidad de contenidos que Clakovi ofrece a su audiencia, abarcando desde música y moda hasta reflexiones personales y consejos sobre creación de contenido.

Unique Terms:

gustar:	5.160647713553408
comprar:	4.073495799647887
poner:	4.049386840246329
muchísimo:	4.017574403983856
encantar:	3.9891803398811363
valer:	3.9453099182079345
quedar:	3.362827261149701
foto:	3.13180088256254
año:	3.130966511028955
tener:	3.111960728365843

Figura 74 Términos únicos Clakovi

Los términos distintivos de Clakovi revelan que a menudo comparte sus preferencias y opiniones, especialmente en relación con compras y productos. Palabras como "gustar", "comprar" y "encantar" sugieren que ofrece recomendaciones y muestra sus adquisiciones. Además, términos como "poner", "quedar" y "foto" indican que podría estar discutiendo temas relacionados con la moda, el estilo y la

fotografía. Estas palabras clave proporcionan una visión de su enfoque en compartir experiencias personales y consejos con su audiencia.

Network

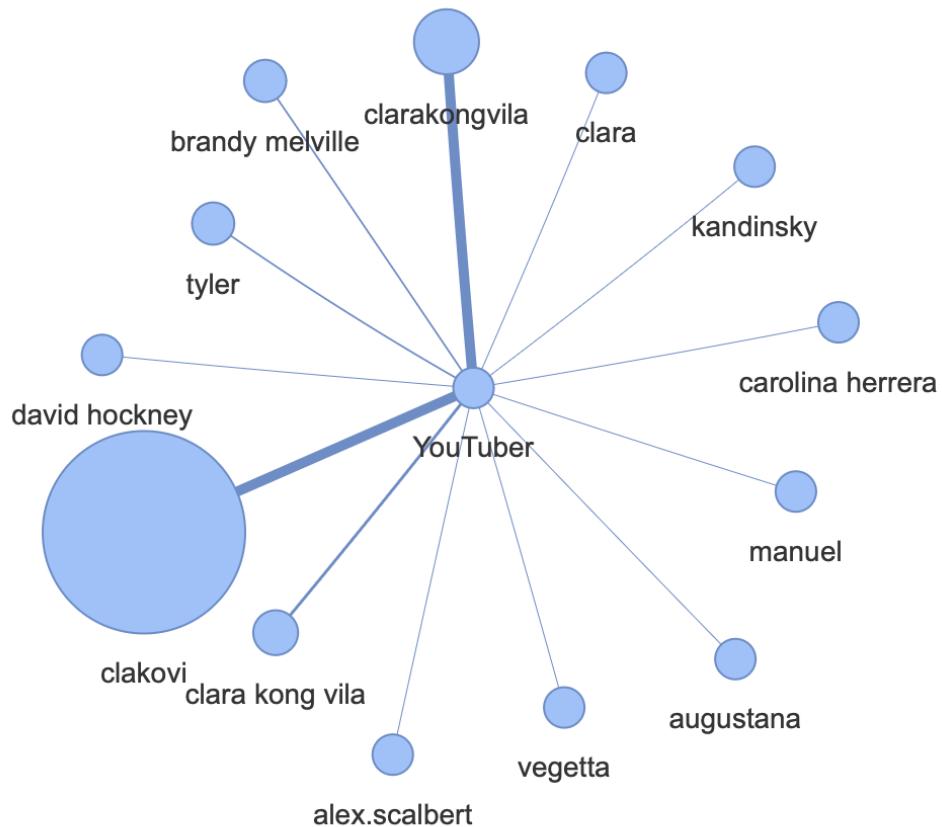


Figura 75 Network Clakovi

El análisis de la red de Clakovi muestra que, sin duda, es la figura central y más mencionada en su contenido, seguida de cerca por variantes de su nombre como "clarakongvila" y "clara kong vila". Esto sugiere que su marca personal es fuerte y que a menudo se refiere a sí misma en tercera persona o utiliza diferentes variaciones de su nombre. Además, hay menciones de marcas y personalidades como "brandy melville", "tyler" y "carolina herrera", lo que podría indicar colaboraciones, influencias o simplemente preferencias personales. También es interesante notar la presencia de nombres de artistas y figuras culturales como "david hockney", "kandinsky" y "dalí", lo que podría sugerir un interés en el arte y la cultura. En general, la red refleja una combinación de su vida personal, sus intereses y posiblemente las marcas o figuras con las que ha interactuado o se ha inspirado.

ElRubius

EDA

	likes	views	comments	year	month	\
count	4.280000e+02	4.280000e+02	428.000000	428.000000	428.000000	
mean	3.062946e+05	3.965870e+06	7556.623832	2020.668224	6.649533	
std	1.889936e+05	2.633231e+06	9625.305887	1.945925	3.501565	
min	5.271000e+03	1.359100e+05	309.000000	2006.000000	1.000000	
25%	1.662560e+05	2.113056e+06	2194.250000	2020.000000	3.000000	
50%	2.831310e+05	3.534673e+06	5771.500000	2021.000000	7.000000	
75%	3.930728e+05	5.023261e+06	9986.250000	2022.000000	10.000000	
max	1.236915e+06	2.376212e+07	123795.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	428.000000	428.000000	428.000000			
mean	2.950935	18.785047	0.080800			
std	1.983521	2.041848	0.020809			
min	0.000000	0.000000	0.011111			
25%	1.000000	18.000000	0.071399			
50%	3.000000	19.000000	0.078356			
75%	5.000000	20.000000	0.086572			
max	6.000000	23.000000	0.187760			

Figura 76 Estadística ElRubius

- **Volumen de Datos:** Se han analizado un total de 428 videos de ElRubius.
- **Interacción:**
 - o **Likes:** En promedio, los videos de ElRubius reciben aproximadamente 306,294 likes. El video más popular ha alcanzado 1,236,915 likes, mientras que el menos popular ha obtenido 5,271 likes.
 - o **Vistas:** Los videos tienen un promedio de 3,965,870 vistas, con un video alcanzando un máximo de 23,762,120 vistas y un mínimo de 135,910 vistas.
 - o **Comentarios:** En promedio, los videos reciben 7,556 comentarios, con un video que ha alcanzado un máximo de 123,795 comentarios y un mínimo de 309 comentarios.
- **Tiempo:**
 - o **Año:** Los videos se han publicado desde 2006 hasta 2023.
 - o **Mes:** El mes promedio de publicación es alrededor de julio.
 - o **Día de la semana:** La mayoría de los videos se publican a mediados de semana, con un promedio de 2.95 (donde 0 es lunes y 6 es domingo).
 - o **Hora:** La mayoría de los videos se publican alrededor de las 6:47 p.m.
- **Tasa de Engagement:** La tasa de engagement promedio es del 8.08%. Esto indica que, en promedio, el 8.08% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 18.776%, mientras que la más baja es del 1.111%.
- **Distribución de Datos:** Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque ElRubius tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

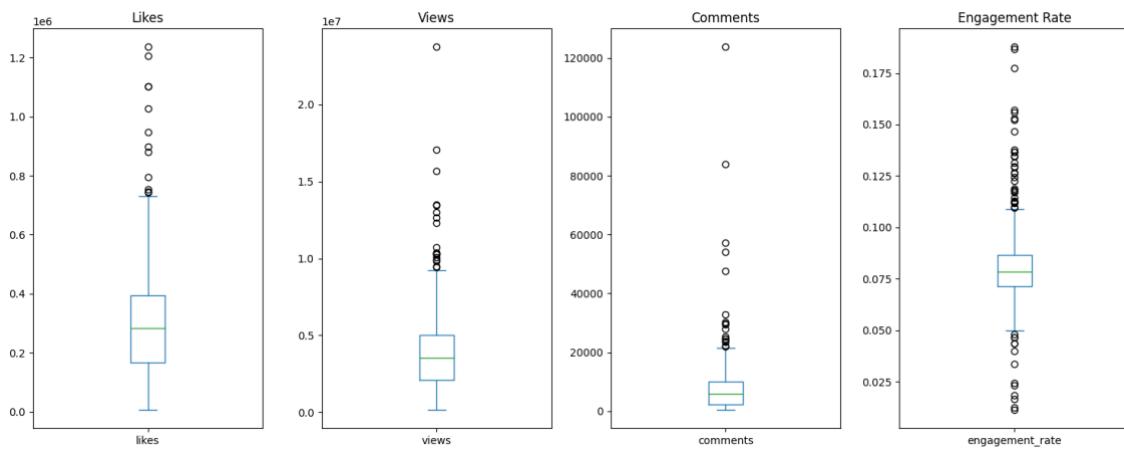


Figura 77 Boxplot ElRubius

Los boxplots de ElRubius revelan valores atípicos altos en likes, vistas y comentarios, indicando videos que tuvieron un rendimiento excepcionalmente bueno. Además, la tasa de engagement muestra outliers en ambos extremos, sugiriendo que hay videos que generan mucha interacción en relación a sus vistas y otros que no.

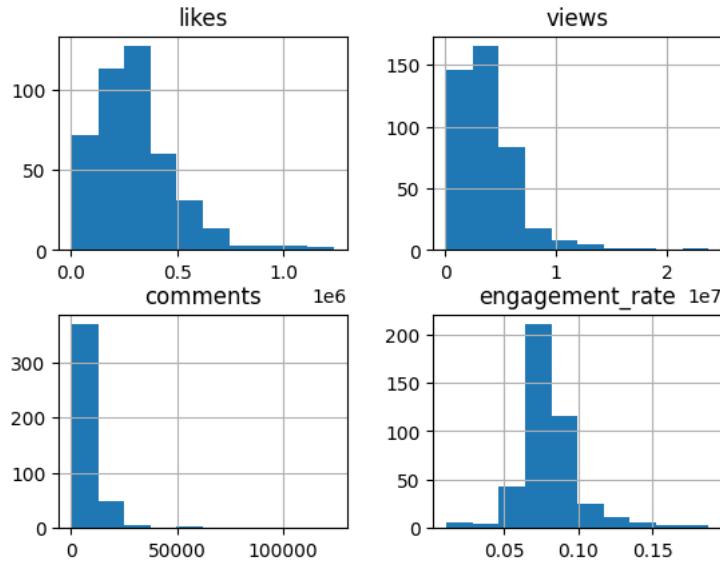


Figura 78 Histograma ElRubius

Los histogramas de ElRubius muestran una distribución con una tendencia central en términos de likes, vistas y comentarios, lo que indica que la mayoría de sus videos reciben una cantidad promedio de interacción. Sin embargo, también hay barras en los extremos, especialmente en el lado derecho, lo que sugiere que hay videos que superan significativamente el promedio en términos de interacción. Esto podría indicar que, aunque la mayoría de sus videos tienen un rendimiento constante, hay ciertos contenidos que resuenan excepcionalmente bien con su audiencia y generan una interacción mucho mayor.

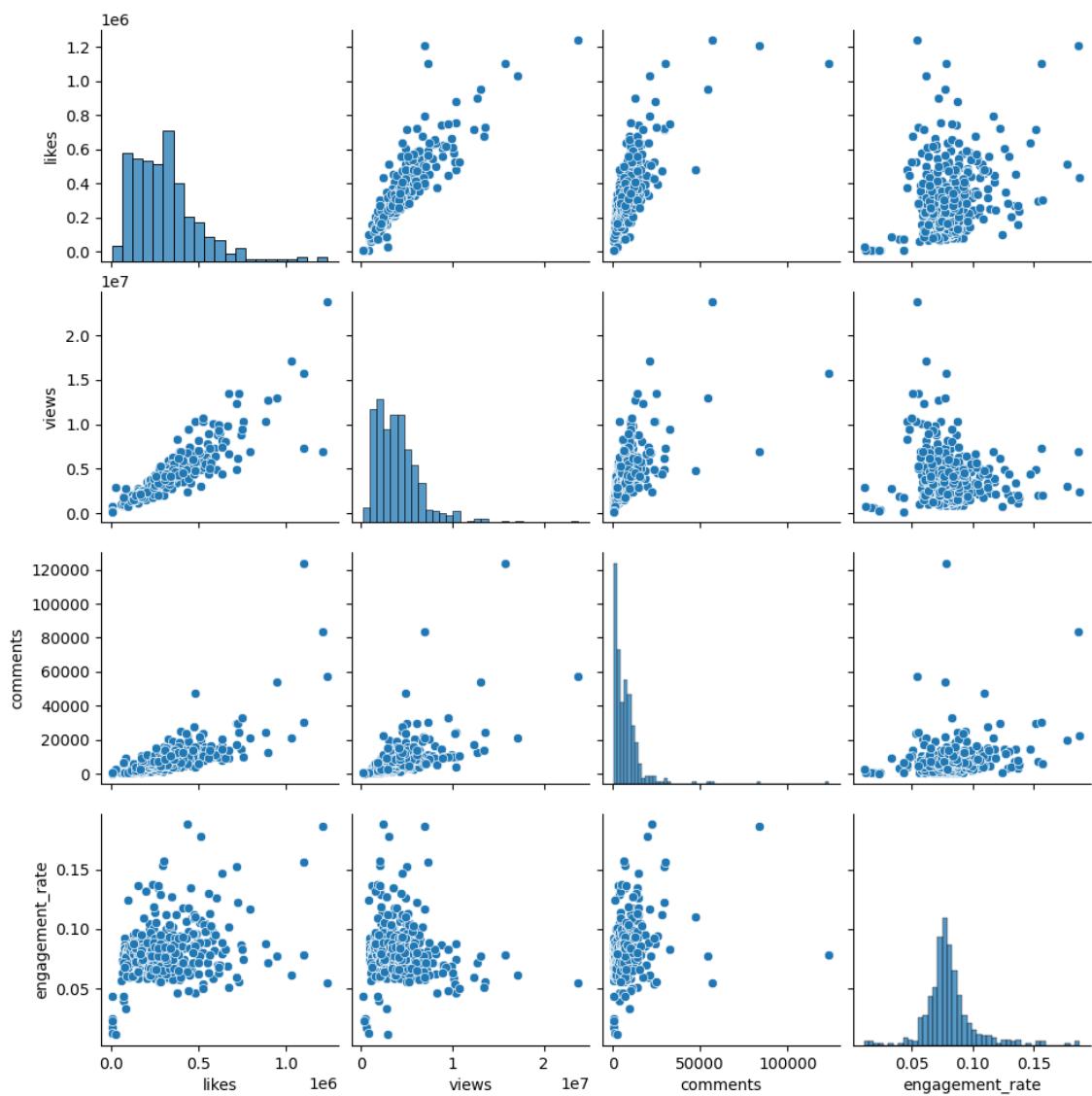


Figura 79 Scatterplot ElRubius

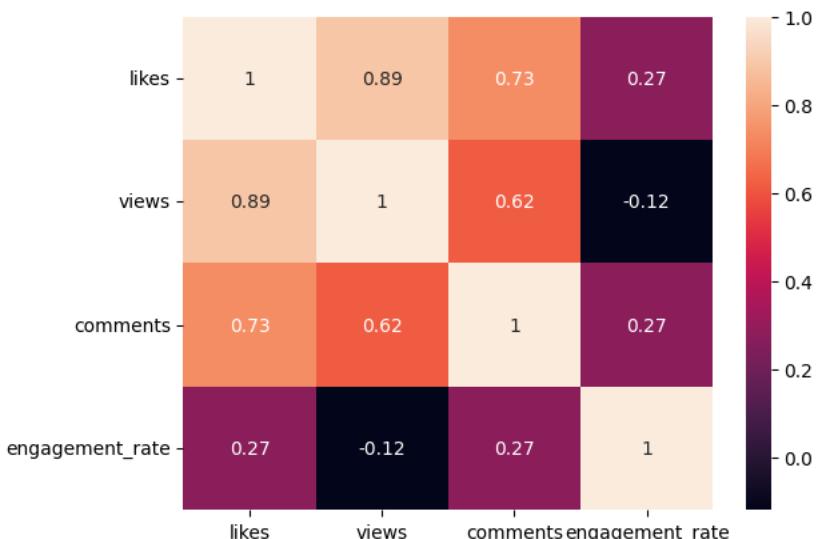


Figura 80 Matriz correlación ElRubius

Según el scatterplot y matriz de correlaciones, los videos de ElRubius muestran una fuerte correlación entre likes y vistas, aproximadamente del 89%. Esto sugiere que los videos más vistos suelen ser también los más apreciados. Además, hay una correlación positiva moderada del 73% entre likes y comentarios, indicando que los videos más populares generan más interacción. Sin embargo, es interesante observar que, aunque las vistas y la tasa de engagement están correlacionadas negativamente en un 12%, esto sugiere que un mayor número de vistas no siempre se traduce en una mayor proporción de interacciones en relación con el total de vistas.

Serie Temporal

- Análisis Mensual:

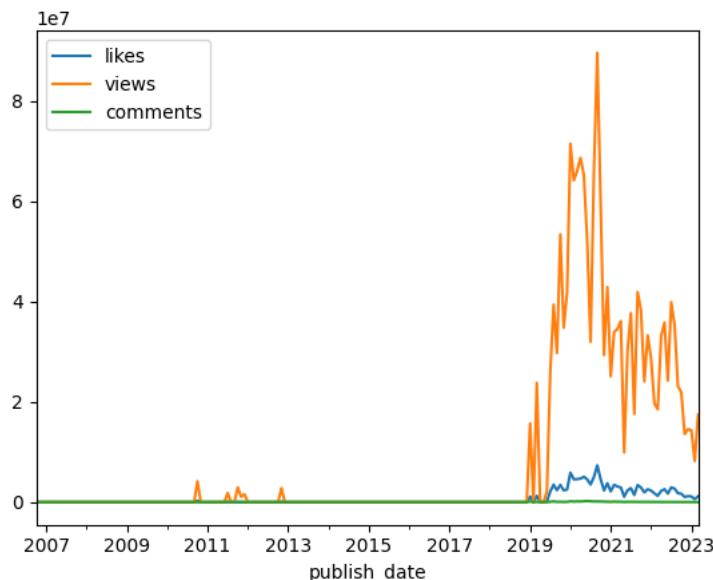


Figura 81 Serie Temporal mes ElRubius

- La actividad de ElRubius antes de 2019 es mínima, lo que sugiere que su canal estaba en sus etapas iniciales o no estaba tan activo.
- A partir de 2019, se observa un crecimiento en la actividad, alcanzando su pico a mediados de 2020. Esto podría estar relacionado con eventos específicos, colaboraciones o tendencias virales que impulsaron su popularidad.
- Después de su pico en 2020, la actividad disminuye, lo que podría indicar una saturación del contenido, cambios en las preferencias de la audiencia o una estrategia de contenido diferente.
- Análisis Semanal: ElRubius muestra una distribución uniforme en la publicación de videos a lo largo de la semana. Esto indica que no tiene un día preferido para publicar y busca mantener una presencia constante en la plataforma.

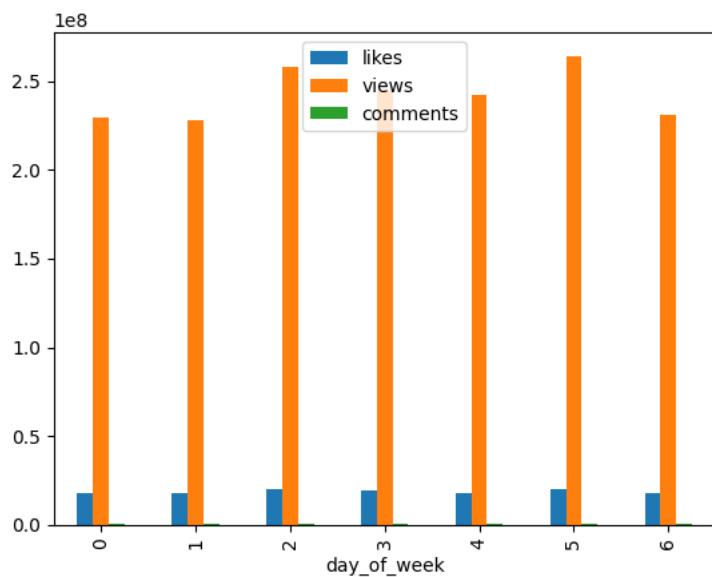


Figura 82 Serie Temporal semana ElRubius

- Análisis Horario: Se observa una preferencia clara de ElRubius por publicar videos entre las 18 y 21 horas. Esta tendencia podría estar relacionada con patrones de consumo de su audiencia, aprovechando las horas pico para maximizar la visibilidad y el engagement. Durante el día, la actividad es notablemente menor, lo que sugiere que este no es su horario objetivo para la publicación de contenido.

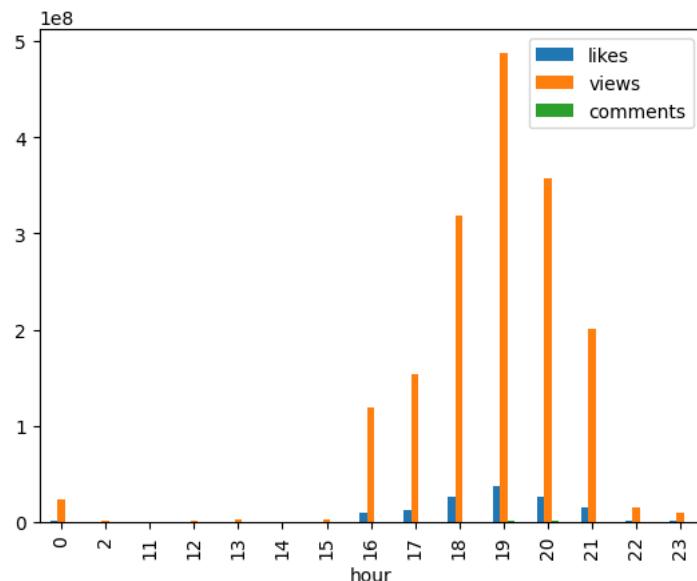


Figura 83 Serie Temporal hora ElRubius

El video más popular es “NIÑO DE 13 AÑOS ME DONA 4300\$ EN DIRECTO”, este título sugiere un evento inesperado y significativo durante una transmisión en vivo, lo que podría haber atraído a una gran audiencia debido a su naturaleza impactante y única. La mención de una donación tan grande de un niño de 13 años genera curiosidad y podría haber impulsado a los espectadores a compartir el video, aumentando su popularidad. Es esencial considerar cómo este video se compara en términos de engagement y vistas con otros contenidos del canal para entender su impacto total.

Minería de Textos



Figura 84 Wordcloud título *ElRubius*

En función de nube de palabras de los títulos, "Karmaland" es el término más frecuente en los títulos de los videos de ElRubius, con una frecuencia de 140. Esto sugiere que es una serie o temática recurrente en su canal, posiblemente relacionada con el juego "Minecraft", que también aparece en la lista. Otros términos destacados incluyen "Minecraft", "Rubius", "Among Us" y "Detroit: Become Human". Estos términos indican los juegos y temas que ElRubius ha cubierto con más frecuencia en su canal. La presencia de palabras como "volver", "casa" y "vegetta" podrían indicar colaboraciones o temáticas específicas relacionadas con otros creadores o eventos personales. Es esencial considerar cómo estos temas se relacionan con las métricas de engagement y vistas para entender su relevancia en el canal.



Figura 85 Wordcloud descripción *ElRubius*

En cuanto a las descripciones, se observa que ElRubius a menudo hace referencia a "like" en sus descripciones, posiblemente incentivando a sus seguidores a interactuar con sus videos. Además, menciona con frecuencia "discord" y "streams", lo que sugiere que tiene una comunidad activa en Discord y realiza transmisiones en vivo regularmente. La "tienda" es otro término recurrente, lo que podría indicar que promociona productos o mercancía relacionada con su marca. Aunque "directo" aparece con menos frecuencia, esto refuerza la idea de que realiza transmisiones en vivo. El término "karmaland" vuelve a aparecer, confirmando su relevancia en el contenido del canal. Además, menciones como "juego", "twitch", "casa" y "server" ofrecen una visión de las plataformas que utiliza, los temas que aborda y posiblemente lugares donde juega o realiza sus streams. Por último, palabras como "querer", "mañana" y "likir" podrían estar relacionadas con anuncios, eventos próximos o simplemente errores tipográficos en las descripciones.

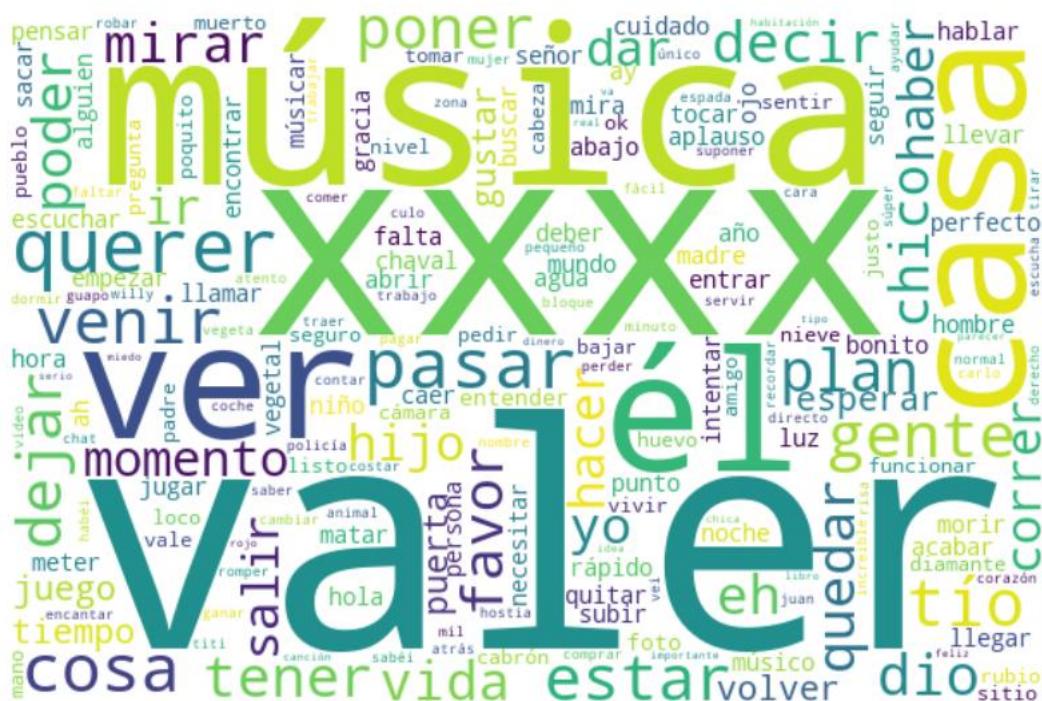


Figura 86 Wordcloud transcripción ElRubius

En las transcripciones de los videos de ElRubius, la palabra "valer" es la más repetida, lo que podría indicar que es una expresión que utiliza con frecuencia, posiblemente como una muletilla o para enfatizar algo. La palabra "xxxx" aparece en segundo lugar, lo que sugiere que hay momentos en los que utiliza lenguaje fuerte o insultos, que han sido censurados o reemplazados en la transcripción. Las menciones de "música" y "casa" podrían estar relacionadas con el ambiente o el contexto en el que se graban los videos, siendo la música un elemento de fondo y la casa el lugar principal de grabación. Las palabras "ver", "querer", "pasar" y "cosa" son verbos y sustantivos comunes que podrían estar relacionados con las actividades o temas que aborda en sus videos. "Tío" es una expresión coloquial que utiliza para referirse a las personas o para enfatizar algo. "Venir", "gente", "estar" y "plan" son términos que podrían estar relacionados con las interacciones, planes o eventos que menciona. Finalmente, "dejar" y "favor" podrían estar asociados a peticiones o acciones que realiza o comenta durante sus videos.

Common Phrases:

correr	correr	correr:	343
valer	valer	valer:	333
ay	ay	ay:	238
música	aplauso	música:	143
xxxx	xxxx	xxxx:	109
hola	hola	hola:	98
mira	mira	mira:	85
música	música	música:	79
ojo	ojo	ojo:	74
pam	pam	pam:	67
rápido	rápido	rápido:	63
dio	dio	dio:	57
uy	uy	uy:	55
suscríbete	dale	like:	53
dale	like	querer:	45
favor	favor	favor:	45
ey	ey	ey:	44
pa	pa	pa:	43
go	go	go:	41
ver	ver	ver:	40

Figura 87 Frases comunes ElRubius

Las frases comunes en las transcripciones de ElRubius revelan ciertos patrones y muletillas que utiliza con frecuencia en sus videos:

- Repetición para énfasis: Muchas de las frases se repiten tres veces, como "correr correr correr", "valer valer valer", y "ay ay ay". Esta repetición puede ser una forma de enfatizar una acción, emoción o reacción durante sus gameplays o interacciones.
- Expresiones de sorpresa o énfasis: Frases como "ay ay ay", "ojito ojito ojito", "uy uy uy", y "ey ey ey" indican momentos de sorpresa, atención o énfasis en algo que está sucediendo en el video.
- Ambiente y sonidos de fondo: Las menciones de "música aplauso música" y "música música música" sugieren que a menudo hay música o sonidos de fondo en sus videos, y estos podrían ser usados para crear ambiente o destacar momentos específicos.
- Lenguaje fuerte o censurado: La frase "xxxx xxxx xxxx" indica que hay momentos en los que utiliza lenguaje fuerte o palabras que han sido censuradas o reemplazadas en la transcripción.
- Llamados a la acción: Las frases "suscríbete dale like" y "dale like querer" son llamados a la acción para que sus seguidores interactúen con el video, ya sea suscribiéndose a su canal o dando "like" al contenido.
- Reacciones rápidas o movimientos: Las frases "correr correr correr", "rápido rápido rápido", "go go go", y "ver ver ver" sugieren momentos de acción rápida o reacciones inmediatas durante sus gameplays.

Estas frases comunes ofrecen una visión de la dinámica y el estilo de ElRubius en sus videos, mostrando su energía, énfasis y la forma en que interactúa con su audiencia.

```
Topics:
Topic: 0, Words: 0.027*"xxxx" + 0.020*"valer" + 0.012*"música" + 0.009*"casa" + 0.008*"ver" + 0.006*"gente" + 0.006*"pasar" + 0.006*"querer" + 0.006*"tío" + 0.006*"estar"
Topic: 1, Words: 0.050*"música" + 0.033*"valer" + 0.011*"tío" + 0.008*"ver" + 0.007*"xxxx" + 0.006*"ah" + 0.006*"eh" + 0.005*"chico" + 0.005*"querer"
Topic: 2, Words: 0.019*"valer" + 0.014*"casa" + 0.010*"xxxx" + 0.010*"música" + 0.010*"querer" + 0.009*"ver" + 0.009*"cosa" + 0.007*"pasar" + 0.007*"estar" + 0.006*"plan"
Topic: 3, Words: 0.006*"the" + 0.006*"matar" + 0.005*"you" + 0.005*"misión" + 0.005*"ver" + 0.005*"jesús" + 0.004*"i" + 0.004*"cadáver" + 0.004*"cámara" + 0.004*"abajo"
Topic: 4, Words: 0.031*"valer" + 0.029*"xxxx" + 0.010*"música" + 0.008*"ver" + 0.008*"querer" + 0.008*"casa" + 0.006*"venir" + 0.006*"pasar" + 0.006*"tío" + 0.006*"gente"
```

Figura 88 Topic modeling ElRubius

Los temas extraídos de las transcripciones de ElRubius revelan ciertos patrones y temas recurrentes en su contenido:

- Tópico 0: Este tema parece centrarse en las reacciones y emociones de ElRubius. Las palabras clave son "xxxx" (posiblemente un lenguaje fuerte o censurado), "valer", "música", "casa", "ver", "gente", "pasar", "querer", "tío" y "estar". Esto sugiere momentos de sorpresa, énfasis y reacciones a eventos específicos, posiblemente en sus gameplays.
- Tópico 1: La música juega un papel prominente en este tema, junto con palabras como "valer", "tío", "ver", "xxxx", "ah", "eh", y "chico". Esto podría indicar momentos de diversión, sorpresa y énfasis musical en sus videos.
- Tópico 2: Aquí, las palabras clave son "valer", "casa", "xxxx", "música", "querer", "ver", "cosa", "pasar", "estar" y "plan". Este tema podría centrarse en las actividades y planes de ElRubius, posiblemente relacionados con sus gameplays o vlogs.
- Tópico 3: Este tema parece tener un enfoque más internacional o diverso, con palabras en inglés como "the" y "you". También incluye palabras como "matar", "misión", "jesús", "i", "cadáver", "cámara" y "abajo". Esto sugiere que podría estar relacionado con misiones o tareas específicas en juegos, o incluso con momentos más intensos o dramáticos.
- Tópico 4: Al igual que los otros temas, este también se centra en palabras como "valer", "xxxx", "música", "ver", "querer", "casa", "venir", "pasar", "tío" y "gente". Sin embargo, la combinación específica de estas palabras podría indicar momentos de interacción y reacción con su audiencia o con otros jugadores en sus gameplays.

Unique Terms:

- casa: 26.61243696907685
- eh: 15.863475893452524
- tío: 13.997984836387701
- plan: 13.744236610707103
- poner: 12.651005466641855
- dejar: 12.516552018492394
- favor: 11.894675664401024
- correr: 11.632063962641805
- decir: 11.180523685358974
- chico: 11.140901132811893

Figura 89 Términos únicos ElRubius

ElRubius, en función de términos únicos, muestra una tendencia a mencionar ciertas palabras que reflejan su estilo y contenido. Habla con frecuencia sobre su "casa", lo que sugiere que a menudo comparte eventos o situaciones domésticas con su audiencia. La interjección "eh" es recurrente, posiblemente usada para enfatizar puntos o captar la atención. Se refiere a menudo a sus seguidores o a otros de manera amigable usando "tío". Las palabras "plan", "poner" y "dejar" indican acciones o estrategias que comenta, ya sea en el contexto de un juego o en su vida diaria. Además, "favor" podría señalar peticiones que hace a su comunidad, mientras que "correr" podría estar relacionado con acciones en videojuegos. Las menciones a "decir" y "chico" reflejan su estilo coloquial y cercano al narrar o interactuar en sus videos.

Network

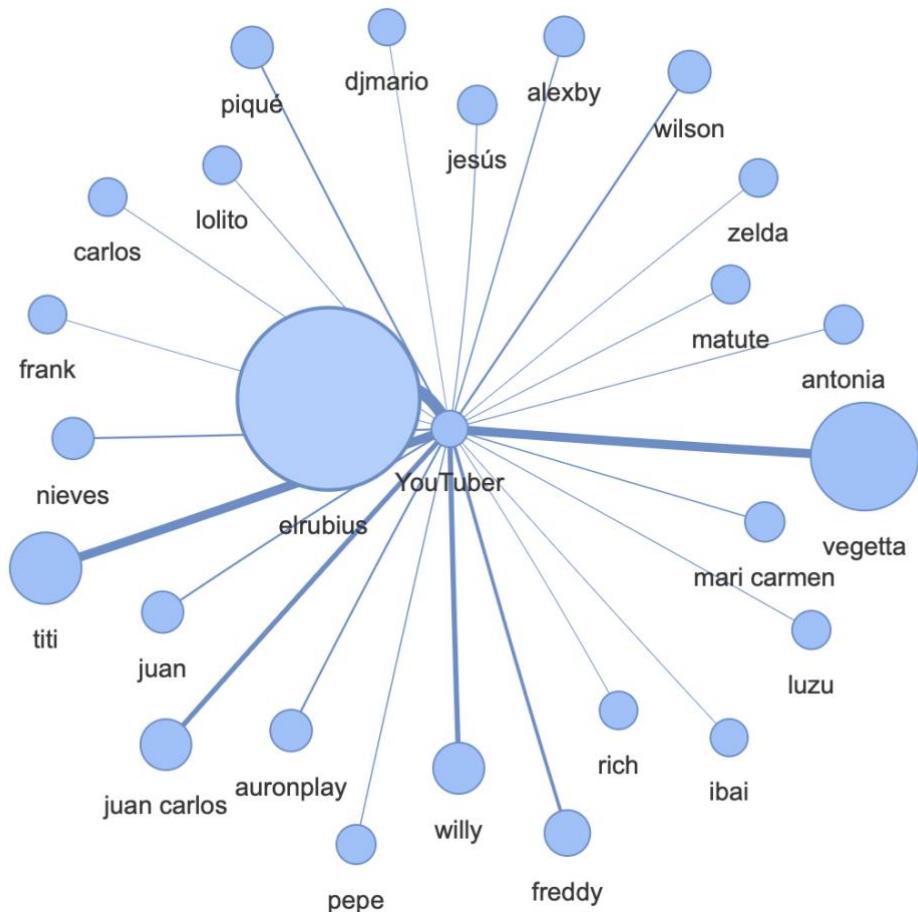


Figura 90 Network ElRubius

En el análisis de red relacionado con ElRubius, se observa que tiene conexiones significativas con varios otros YouTubers y personalidades. "Vegetta" es el segundo más mencionado después de ElRubius, lo que sugiere una colaboración o relación cercana entre ellos. Otros nombres como "titi", "willy", y "juan carlos" también tienen un alto número de menciones, indicando posibles colaboraciones o interacciones frecuentes. Además, hay nombres que podrían estar relacionados con personajes o referencias dentro de sus videos, como "zelda" o "freddy". También es interesante notar la presencia de "auronplay", otro YouTuber popular, y "cristiano ronaldo", el famoso futbolista. Estas conexiones reflejan la diversidad de contenidos y colaboraciones en el canal de ElRubius, así como las múltiples interacciones y referencias que hace en sus videos.

Emma Chamberlain

EDA

	likes	views	comments	year	month	\
count	2.820000e+02	2.820000e+02	282.000000	282.000000	282.000000	
mean	2.346354e+05	5.593459e+06	10249.992908	2018.695035	7.187943	
std	1.517226e+05	4.230503e+06	9602.106079	1.571287	3.240666	
min	1.894200e+04	5.447670e+05	0.000000	2017.000000	1.000000	
25%	1.073982e+05	2.991707e+06	3935.750000	2017.000000	5.000000	
50%	2.371690e+05	4.913082e+06	7713.500000	2018.000000	7.500000	
75%	3.203258e+05	7.002393e+06	13336.250000	2020.000000	10.000000	
max	1.042064e+06	3.772295e+07	69228.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	282.000000	282.000000	282.000000			
mean	3.787234	15.964539	0.044915			
std	2.229880	6.750154	0.012884			
min	0.000000	0.000000	0.016232			
25%	2.000000	15.000000	0.034877			
50%	4.000000	18.000000	0.042685			
75%	6.000000	20.000000	0.052815			
max	6.000000	23.000000	0.111455			

Figura 91 Estadística Emma Chamberlain

- **Volumen de Datos:** Se han analizado un total de 282 videos de Emma Chamberlain.
- **Interacción:**
 - **Likes:** En promedio, los videos de Emma Chamberlain reciben aproximadamente 234,635 likes. El video más popular ha alcanzado 1,042,064 likes, mientras que el menos popular ha obtenido 18,942 likes.
 - **Vistas:** Los videos tienen un promedio de 5,593,459 vistas, con un video alcanzando un máximo de 37,722,950 vistas y un mínimo de 544,767 vistas.
 - **Comentarios:** En promedio, los videos reciben 10,249 comentarios, con un video que ha alcanzado un máximo de 69,228 comentarios y un mínimo de 0 comentarios.
- **Tiempo:**
 - **Año:** Los videos se han publicado desde 2017 hasta 2023.
 - **Mes:** El mes promedio de publicación es alrededor de julio.
 - **Día de la semana:** La mayoría de los videos se publican a mediados de semana, con un promedio de 3.78 (donde 0 es lunes y 6 es domingo).
 - **Hora:** La mayoría de los videos se publican alrededor de las 3:57 p.m.

- **Tasa de Engagement:** La tasa de engagement promedio es del 4.49%. Esto indica que, en promedio, el 4.49% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 11.15%, mientras que la más baja es del 1.62%.
- **Distribución de Datos:** Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Emma Chamberlain tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

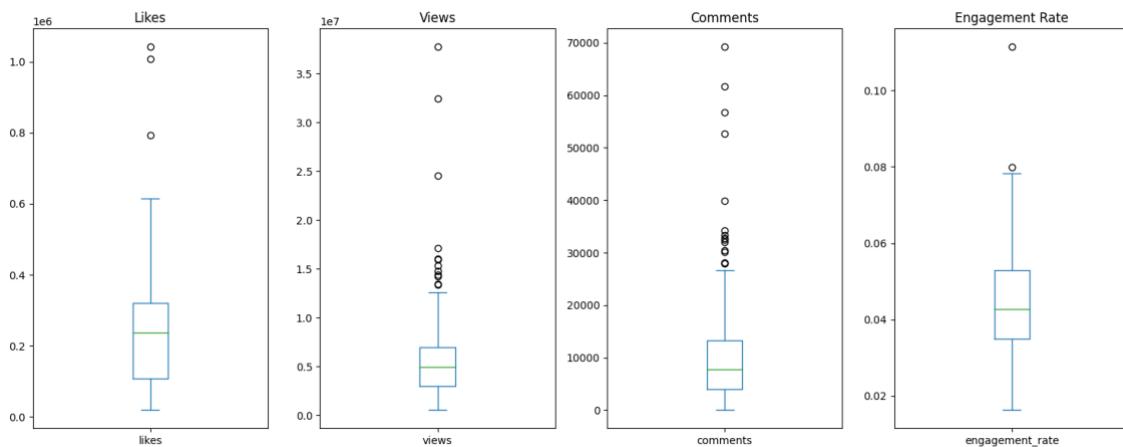


Figura 92 Boxplot Emma Chamberlain

En los boxplots correspondientes a Emma Chamberlain, se observa una notable presencia de valores atípicos elevados en las categorías de likes, vistas y comentarios. Sin embargo, la distribución de la tasa de engagement parece ser más uniforme y no presenta anomalías significativas.

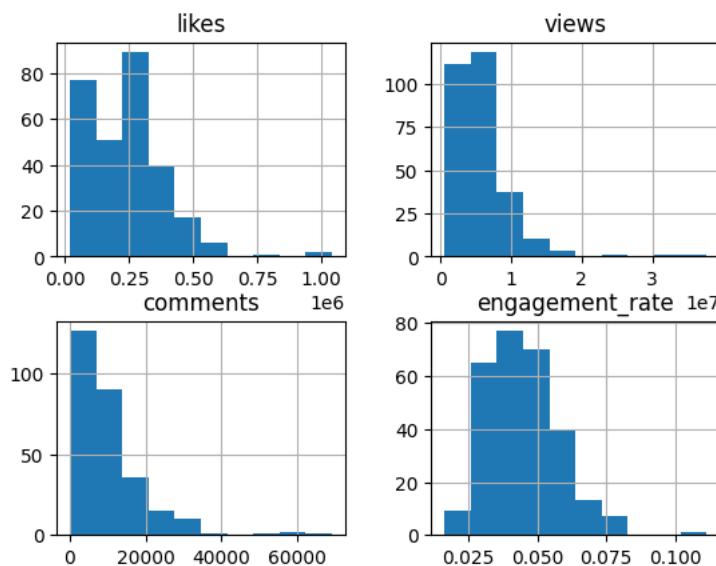


Figura 93 Histograma Emma Chamberlain

En los histogramas, se observa que tanto los "likes", "vistas" como "comentarios" muestran una tendencia sesgada hacia la derecha, indicando que la mayoría de los videos tienen cifras moderadas en estas categorías, pero hay algunos que destacan con números excepcionalmente altos. Por otro lado, la tasa de engagement parece distribuirse de manera más uniforme, sugiriendo que la interacción relativa de los espectadores es constante, independientemente de la popularidad del video.

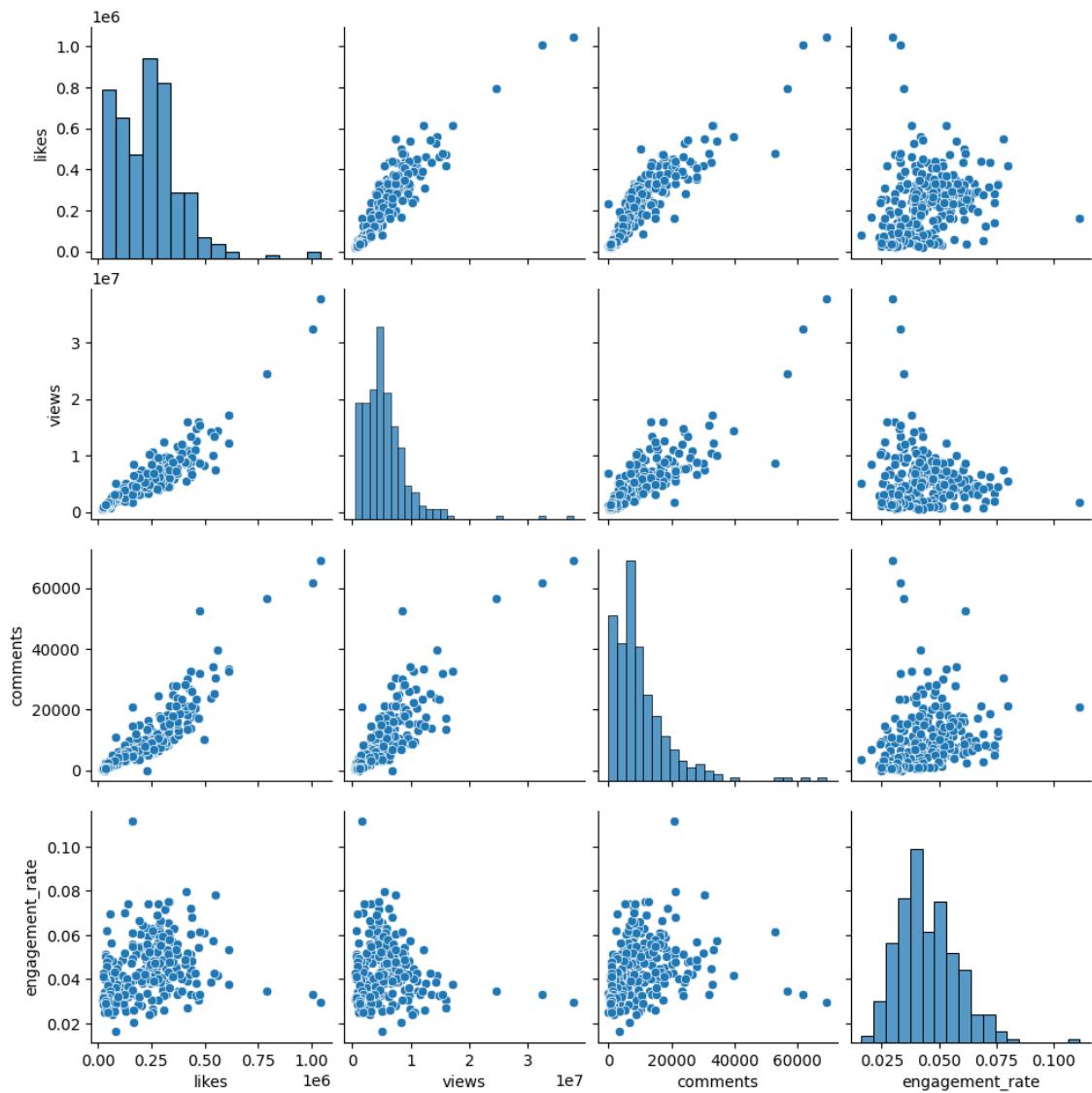


Figura 94 Scatterplot Emma Chamberlain

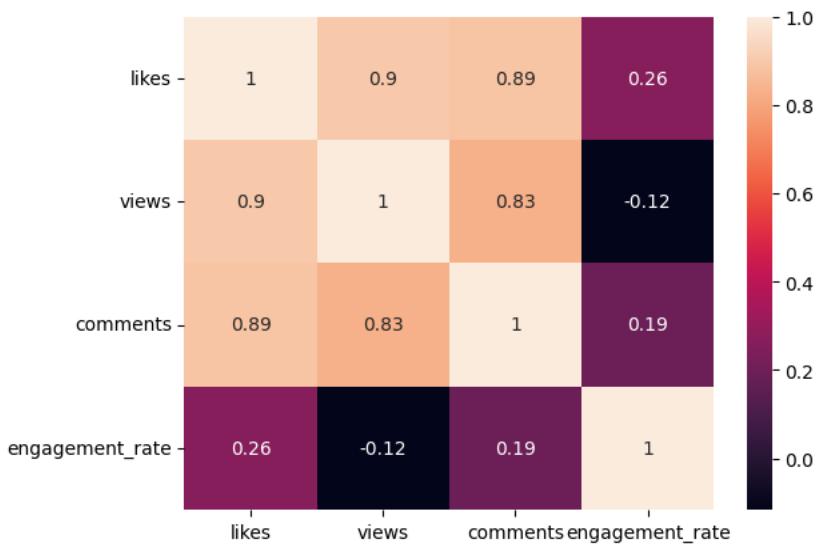


Figura 95 Matriz correlación Emma Chamberlain

Observando el scatterplot y matriz de correlaciones, "likes" y "views" tienen una correlación positiva muy fuerte de 0.898, lo que indica que cuando un video recibe más likes, generalmente también tiene más vistas. "likes" y "comments" también muestran una correlación positiva fuerte de 0.885, lo que sugiere que los videos con más likes tienden a recibir más comentarios. La tasa de "engagement_rate" tiene una correlación positiva moderada con "likes" (0.26) y "comments" (0.19), lo que indica que a medida que aumentan los likes y comentarios, la tasa de engagement también tiende a aumentar. Sin embargo, hay una correlación negativa leve entre "views" y "engagement_rate" (-0.12), lo que sugiere que no necesariamente los videos con más vistas tienen una mayor tasa de engagement.

Serie Temporal

- Análisis Mensual: A lo largo de los años, se ha notado una disminución en la actividad, especialmente antes de 2020. Sin embargo, después de junio de 2020, hay una leve recuperación en las vistas, aunque esta tendencia no se mantiene, ya que la actividad parece disminuir nuevamente después de 2022.

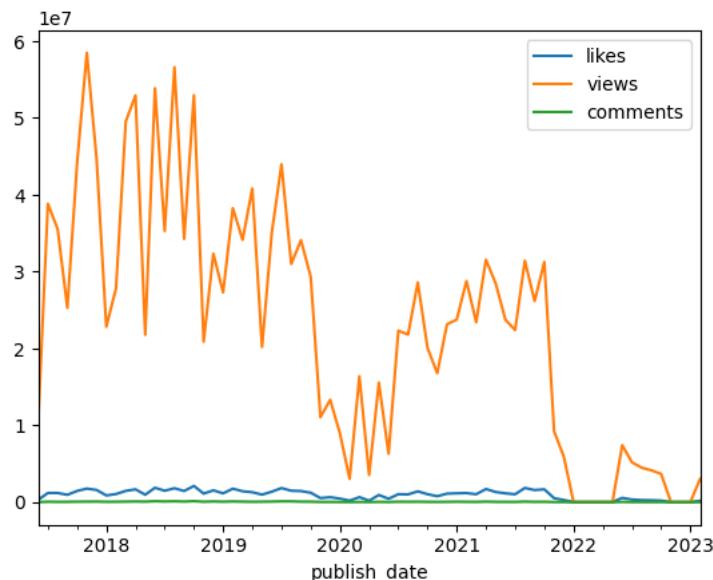


Figura 96 Serie Temporal mes Emma Chamberlain

- Análisis Semanal: Los domingos se destacan como el día con la mayor actividad, lo que sugiere que es el día preferido para publicar contenido. Esta tendencia podría estar relacionada con patrones de consumo de la audiencia, donde los domingos, las personas podrían tener más tiempo libre para ver videos.

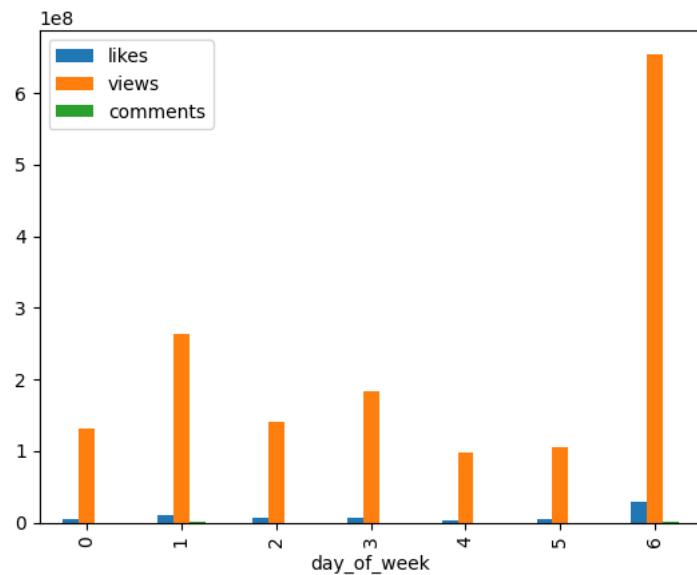


Figura 97 Serie Temporal semana Emma Chamberlain

- Análisis Horario: La franja horaria entre las 17:00 y las 23:00 horas es cuando se registra la mayor actividad, siendo las 20:00 horas el momento más frecuente de publicación. Esto podría indicar que se busca coincidir con las horas en que la audiencia está más activa o disponible para consumir contenido.

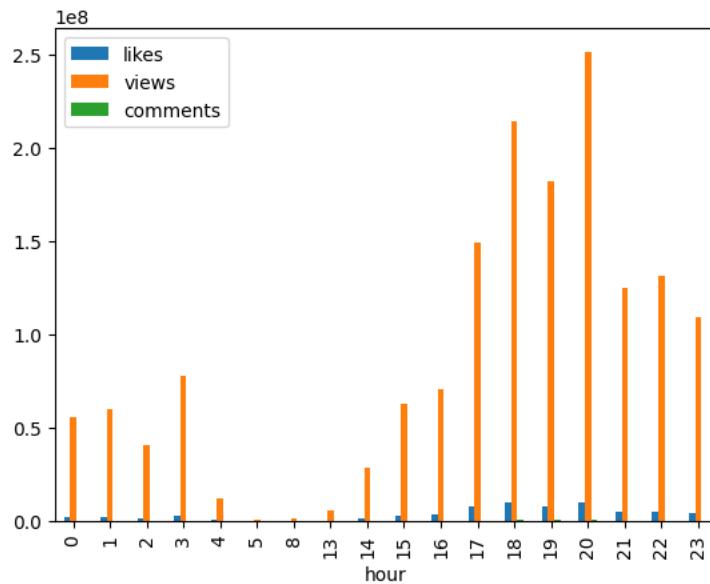


Figura 98 Serie Temporal hora Emma Chamberlain

El video titulado "ROADTRIP TO VEGAS FT DOLAN TWINS & JAMES CHARLES" se destaca como el más popular. La inclusión de personalidades conocidas como los Dolan Twins y James Charles probablemente haya contribuido a su alta popularidad, ya que estos influencers tienen una gran base de seguidores. Además, el tema del video, un viaje por carretera a Las Vegas, es atractivo y promete aventura y entretenimiento, lo que podría haber atraído a una amplia audiencia. Es esencial considerar cómo este video en particular se compara en términos de interacción y vistas con otros videos del canal para entender mejor su impacto y éxito.

Minería de Textos



Figura 99 Wordcloud título *Emma Chamberlain*

Basándonos en la nube de palabras del título, se pueden identificar varios temas y tendencias en los videos:

- Tiempo y Rutina: Las palabras "day", "time", y "hour" sugieren que muchos de los videos pueden estar centrados en actividades diarias, rutinas o desafíos relacionados con el tiempo.
 - Actividades: Las palabras "go", "trip", "vlog", "try", "make", y "cook" indican una variedad de actividades y aventuras que se documentan en los videos. Esto podría incluir viajes, experimentos o intentos de hacer algo nuevo, y momentos de cocina.
 - Estilo de Vida: Las palabras "coffee" y "fashion" sugieren que hay un enfoque en el estilo de vida, desde la moda hasta los momentos cotidianos como disfrutar de un café.
 - Colaboraciones y Personal Branding: La palabra "ft" (abreviatura de "featuring") indica colaboraciones con otros creadores o personalidades. Además, la mención de "emma" sugiere que la personalidad detrás del canal, presumiblemente Emma, es una parte central de la marca y el contenido.
 - Tono Ligero: La inclusión de "lol" sugiere un tono ligero y humorístico en algunos de los videos. En conjunto, estos títulos pintan la imagen de un canal que combina vlogs diarios, desafíos, estilo de vida, moda, y momentos auténticos, todo presentado con un toque personal y, a menudo, con un tono humorístico.

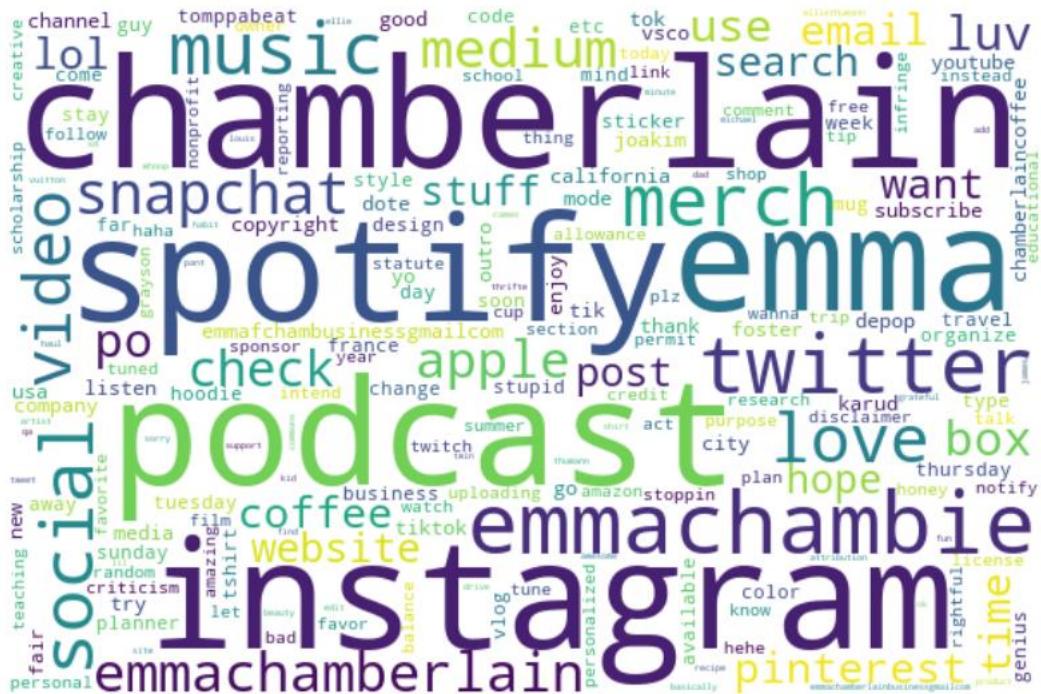


Figura 100 Wordcloud descripción Emma Chamberlain

En función de las descripciones:

- Branding Personal: Las palabras "chamberlain", "emma", "emmachambie" y "emmachamberlain" destacan la marca personal de Emma Chamberlain. Es evidente que su nombre y su marca son centrales en su contenido y descripciones.
 - Plataformas Sociales: Las menciones de "instagram", "podcast", "spotify", "twitter", "snapchat" y "medium" sugieren que Emma tiene una presencia activa en múltiples plataformas sociales y de medios. Esto indica que promociona activamente sus otros canales y contenido en diferentes plataformas.
 - Contenido Multimedia: La presencia de palabras como "podcast", "music" y "video" indica que Emma produce y promociona una variedad de contenido multimedia, desde videos hasta podcasts y música.
 - Merchandising: La palabra "merch" sugiere que Emma tiene su propia línea de mercancía que promociona a través de sus descripciones.
 - Interacción y Comunidad: La palabra "love" puede indicar un tono positivo y afectuoso en sus descripciones, posiblemente reflejando su relación con sus seguidores. Además, "social" y "medium" podrían referirse a su interacción en las redes sociales y su enfoque en compartir a través de diferentes medios.

En resumen, las descripciones pintan la imagen de Emma Chamberlain como una creadora de contenido multifacética con una fuerte marca personal, activa en múltiples plataformas y con una relación cercana y positiva con su comunidad.

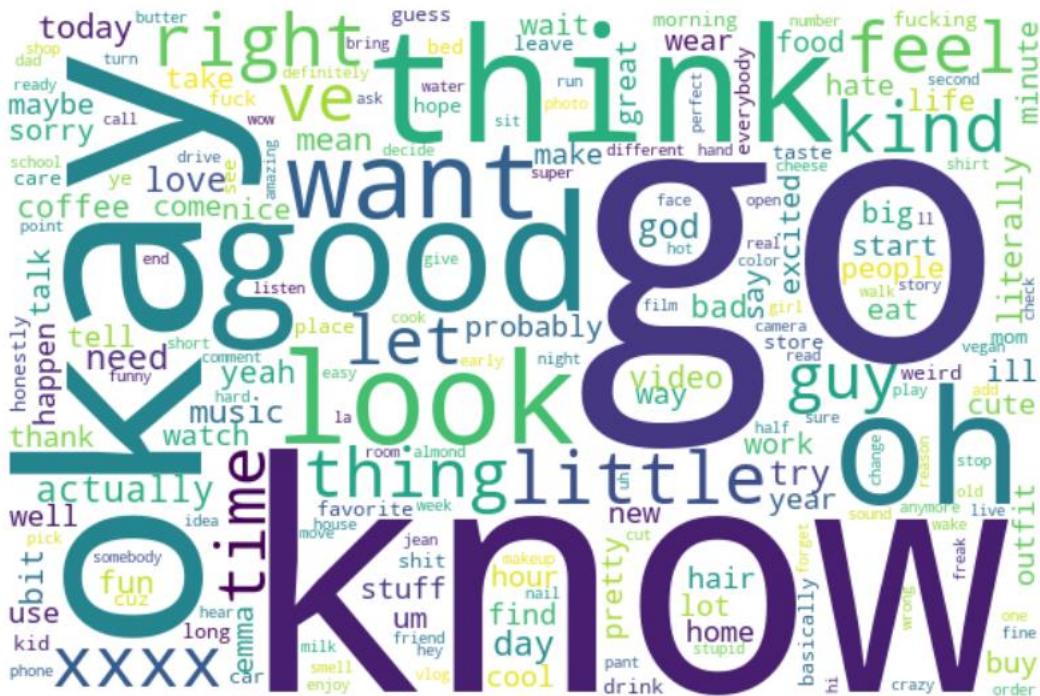


Figura 101 Wordcloud transcripción Emma Chamberlain

Basándonos en la nube de palabras de las transcripciones:

- Reacciones y Emociones: Palabras como "like", "okay", "oh", y "good" sugieren que Emma utiliza un lenguaje coloquial y expresivo, típico de las conversaciones cotidianas y las reacciones genuinas. Estas palabras pueden indicar momentos de sorpresa, acuerdo o reflexión.
 - Reflexiones y Opiniones: Las palabras "know", "think", "want", y "kind" indican que Emma comparte sus pensamientos y opiniones con su audiencia. Es probable que hable sobre sus experiencias personales, reflexiones y deseos.
 - Acciones y Descripciones: Términos como "go", "look", "right", y "get" sugieren que Emma describe acciones, situaciones o eventos. Puede estar narrando lo que está haciendo, describiendo algo que ve o contando una historia.
 - Detalles y Enfoque: La palabra "little" indica que Emma presta atención a los detalles y puede mencionar aspectos específicos o menores de una situación o tema.
 - Negación y Contraste: La palabra "not" es prominente, lo que sugiere que Emma a menudo contrasta sus opiniones o describe situaciones en términos de lo que no son o no ocurren.

Las transcripciones revelan que Emma Chamberlain tiene un estilo de comunicación relatable y auténtico, combinando reacciones emocionales con reflexiones personales y descripciones detalladas. Su lenguaje es coloquial, lo que la hace accesible y cercana a su audiencia.

```

Common Phrases:
like not know: 306
know go to: 255
think go to: 241
like go to: 221
not know like: 208
go to like: 186
not know not: 174
not know go: 157
okay go to: 155
go to try: 154
like feel like: 136
probably go to: 116
know not know: 106
go to go: 96
like like like: 95
go to lie: 88
today go to: 82
like not like: 81
to go to: 75
like not want: 75

```

Figura 102 Frases comunes Emma Chamberlain

Las frases comunes en las transcripciones de Emma Chamberlain revelan patrones recurrentes en su forma de hablar y los temas que aborda:

- Indecisión y Reflexión: Las combinaciones "like not know", "not know like", "know not know", y "like not like" sugieren momentos de indecisión o reflexión. Emma parece compartir con su audiencia sus pensamientos en tiempo real, mostrando autenticidad y vulnerabilidad.
- Acciones y Planes: Las frases "know go to", "think go to", "like go to", "go to like", "not know go", "okay go to", "go to try", "probably go to", "go to go", "go to lie", y "today go to" indican que Emma a menudo habla sobre sus planes, acciones o intenciones. Puede estar describiendo lo que va a hacer, lo que piensa hacer o lo que le gustaría hacer.
- Sentimientos y Emociones: "like feel like" y "like not want" sugieren que Emma comparte sus sentimientos y emociones con su audiencia. Estas frases pueden indicar momentos en los que está describiendo cómo se siente acerca de una situación o experiencia.
- Reiteración y Énfasis: Las frases "like like like" y "to go to" muestran que Emma a veces repite palabras o frases para enfatizar un punto o para expresar una emoción intensa.

```

Topics:
Topic: 0, Words: 0.048*like" + 0.033*"not" + 0.021*"know" + 0.021*"go" + 0.014*"okay" + 0.014*"to" + 0.012*"oh" + 0.010*"want" + 0.010*"right" + 0.010*"get"
Topic: 1, Words: 0.026*"music" + 0.011*"not" + 0.010*"like" + 0.008*"foreign" + 0.008*"know" + 0.007*"took" + 0.005*"photo" + 0.004*"actually" + 0.004*"david" + 0.004*"right"
Topic: 2, Words: 0.073*"like" + 0.037*"not" + 0.028*"know" + 0.017*"go" + 0.011*"okay" + 0.011*"think" + 0.011*"to" + 0.010*"good" + 0.009*"kind" + 0.008*"look"
Topic: 3, Words: 0.019*"amanda" + 0.018*"fry" + 0.014*"uh" + 0.014*"emma" + 0.009*"yo" + 0.007*"wendys" + 0.006*"yeah" + 0.005*"chickfila" + 0.005*"alright" + 0.005*"mcdonald"
Topic: 4, Words: 0.065*"like" + 0.032*"not" + 0.028*"go" + 0.028*"know" + 0.018*"to" + 0.013*"okay" + 0.012*"good" + 0.012*"xxxx" + 0.011*"oh" + 0.011*"think"

```

Figura 103 Topic Modeling Emma Chamberlain

En función de analizar los resultados de Topic Modeling, observamos que:

- Tópico 0: Este tópico se centra en las reacciones y emociones de Emma Chamberlain. Las palabras clave son "like", "not", "know", "go", "okay", "to", "oh", "want", "right" y "get". Esto sugiere momentos de reflexión, indecisión, sorpresa y reacciones espontáneas en sus vlogs o situaciones cotidianas.
- Tópico 1: La música y la moda parecen ser el foco principal de este tópico. Las palabras clave son "music", "foreign", "look", "photo", "actually", "david" y "right". Esto indica que Emma podría estar discutiendo sobre estilo, fotografía con la música del fondo y posiblemente sobre amigos o eventos específicos.
- Tópico 2: Aquí, las palabras clave son "like", "not", "know", "go", "okay", "think", "to", "good", "kind" y "look". Este tópico sugiere una introspección profunda, donde Emma comparte sus pensamientos, sentimientos y opiniones sobre diversos temas o situaciones.
- Tópico 3: Este tópico parece tener un enfoque en la vida cotidiana y las interacciones sociales de Emma. Las palabras clave son "amanda", "fry", "uh", "emma", "yo", "wendys", "chickfila", entre otras. Esto sugiere discusiones sobre comida, salidas, amistades y experiencias diarias.

- Tópico 4: Al igual que el primer tópico, este se centra en las emociones y reacciones de Emma. Las palabras clave son "like", "not", "go", "know", "to", "okay", "good", "xxxx", "oh" y "think". Sin embargo, la presencia de "xxxx" podría indicar momentos de lenguaje fuerte o censurado, mostrando momentos más intensos o apasionados en sus videos.

Unique Terms:

xxxx:	26.332007580757598
music:	14.102860022733692
coffee:	12.469524423348867
video:	9.561000850719333
wear:	9.402278005420253
eat:	9.04063016416169
um:	8.632648696181159
outfit:	7.752552094567262
wait:	7.1703667429815425
maybe:	6.917504011195274

Figura 104 Términos únicos Emma Chamberlain

En el contenido de Emma Chamberlain, la palabra "xxxx" (insulto) es la más única y prominente, lo que sugiere que puede usar lenguaje fuerte con frecuencia. La música juega un papel significativo en su contenido, seguido de temas relacionados con el café, lo que indica su posible afición por esta bebida. Además, palabras como "video", "wear" (usar/vestir) y "outfit" (atuendo) sugieren que a menudo discute moda o lo que lleva puesto. Las palabras "eat" (comer), "um" (una muletilla) y "maybe" (quizás) reflejan su estilo casual y espontáneo de comunicación, mientras que "wait" (esperar) puede indicar momentos de anticipación o pausa en sus videos.

Network

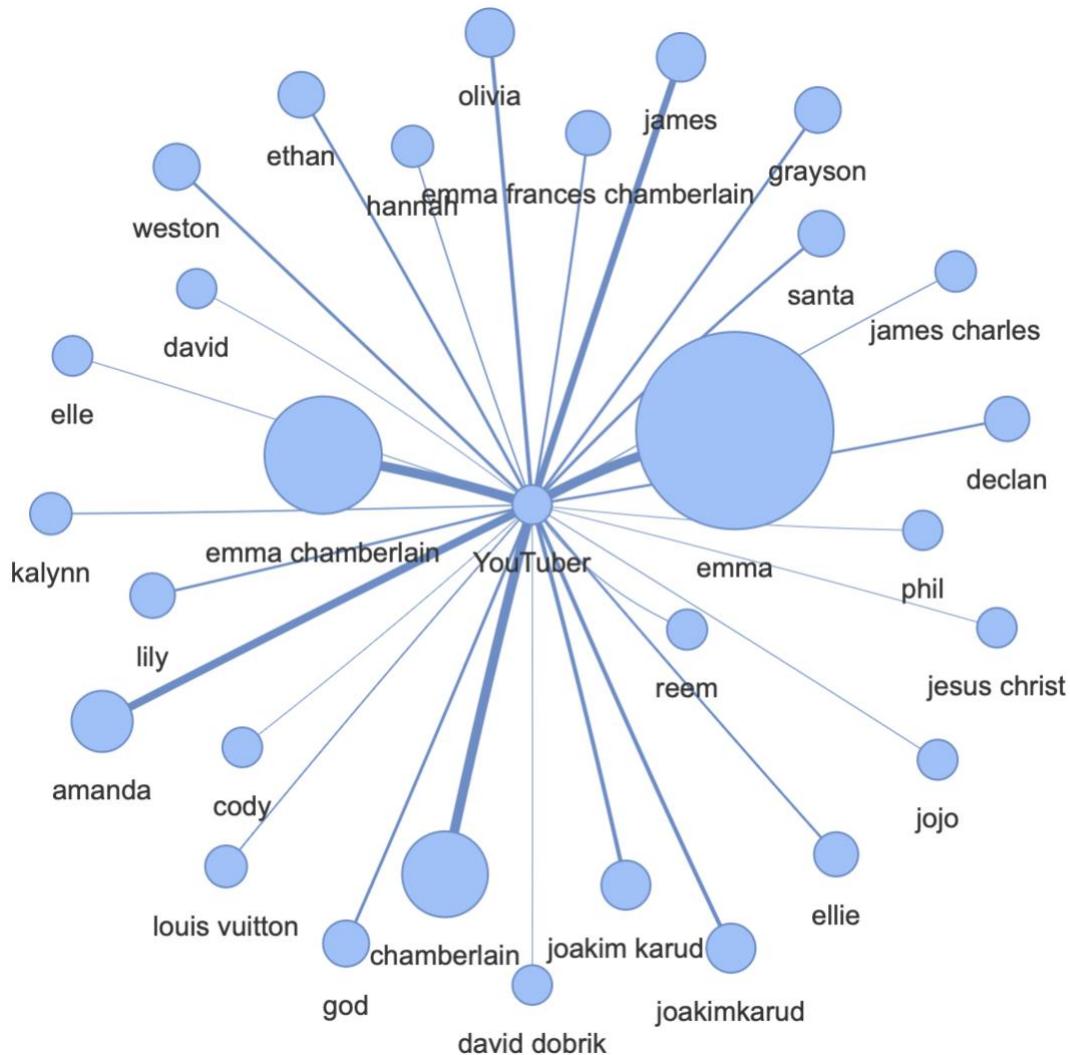


Figura 105 Network Emma Chamberlain

En la red de conexiones de Emma Chamberlain, "emma" es el término más frecuentemente asociado, seguido de "emma chamberlain" y "chamberlain". Esto indica que su nombre completo y su apellido son comúnmente mencionados en el contenido. Además, hay conexiones con otros nombres y personalidades, como "amanda", "joakim karud", "james", "olivia", y "weston". También se mencionan figuras religiosas y festivas como "god" (Dios) y "santa". Es interesante notar que hay conexiones con marcas y figuras de la moda como "louis vuitton" y "harry styles", lo que sugiere que Emma podría estar discutiendo moda o colaborando con estas personalidades. Además, nombres como "james charles" y "david dobrik" indican posibles colaboraciones o discusiones sobre otros YouTubers populares. Otros términos como "louie vuitton", "bigfoot", y "mac demarco" reflejan una variedad de temas y posibles intereses o colaboraciones en su contenido.

Malbert

EDA

	likes	views	comments	year	month	\
count	319.000000	3.190000e+02	319.000000	319.000000	319.000000	
mean	13575.956113	2.910587e+05	924.852665	2018.962382	6.316614	
std	10819.748862	2.207228e+05	620.643722	1.613181	3.767213	
min	294.000000	9.448000e+03	16.000000	2016.000000	1.000000	
25%	5612.000000	1.243565e+05	499.500000	2018.000000	3.000000	
50%	11560.000000	2.487010e+05	769.000000	2019.000000	5.000000	
75%	18590.500000	4.121800e+05	1207.000000	2020.000000	10.000000	
max	58854.000000	1.400073e+06	4176.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	319.000000	319.000000	319.000000			
mean	2.855799	16.673981	0.050525			
std	1.800914	1.477392	0.015385			
min	0.000000	11.000000	0.019510			
25%	2.000000	16.000000	0.040405			
50%	3.000000	17.000000	0.049399			
75%	4.000000	18.000000	0.058712			
max	6.000000	20.000000	0.112639			

Figura 106 Estadística Malbert

- Volumen de Datos: Se han analizado un total de 319 videos de Malbert.
- Interacción:
 - o Likes: En promedio, los videos de Malbert reciben aproximadamente 13,576 likes. El video más popular ha alcanzado 58,854 likes, mientras que el menos popular ha obtenido 294 likes.
 - o Vistas: Los videos tienen un promedio de 291,058 vistas, con un video alcanzando un máximo de 1,400,073 vistas y un mínimo de 9,448 vistas.
 - o Comentarios: En promedio, los videos reciben 925 comentarios, con un video que ha alcanzado un máximo de 4,176 comentarios y un mínimo de 16 comentarios.
- Tiempo:
 - o Año: Los videos se han publicado desde 2016 hasta 2023.
 - o Mes: El mes promedio de publicación es alrededor de junio.
 - o Día de la semana: La mayoría de los videos se publican a principios de semana, con un promedio de 2.86 (donde 0 es lunes y 6 es domingo).
 - o Hora: La mayoría de los videos se publican alrededor de las 4:40 p.m.
- Tasa de Engagement: La tasa de engagement promedio es del 5.05%. Esto indica que, en promedio, el 5.05% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 11.26%, mientras que la más baja es del 1.95%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Malbert tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

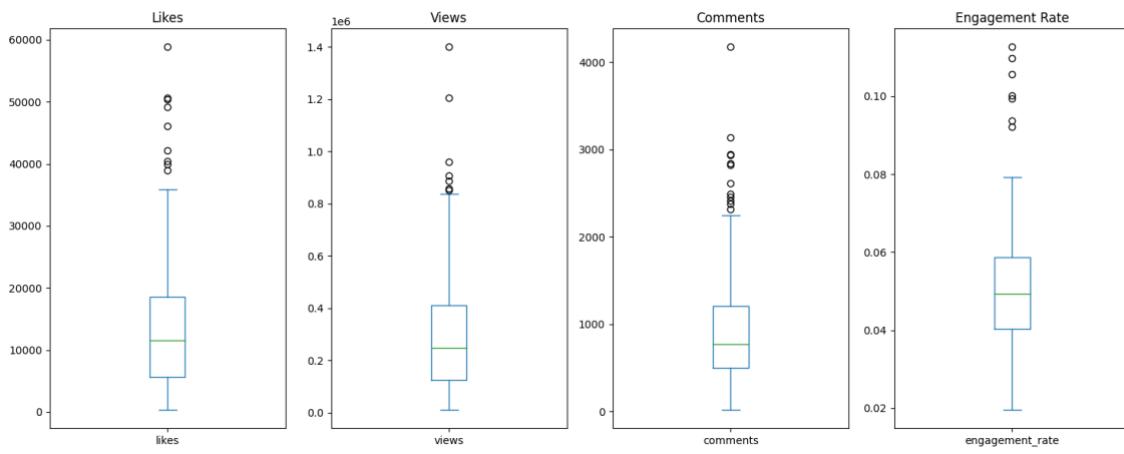


Figura 107 Boxplot Malbert

En los boxplots, se observan ciertos valores que se desvían significativamente del rango intercuartil, lo que indica la presencia de valores atípicos. Estos valores atípicos son notablemente altos en comparación con el resto de los datos. Esto sugiere que, aunque la mayoría de los datos se agrupan dentro de un rango específico, hay ciertas observaciones que tienen valores excepcionalmente altos y no siguen la tendencia general del conjunto de datos. Estos valores atípicos pueden ser el resultado de eventos o circunstancias inusuales.

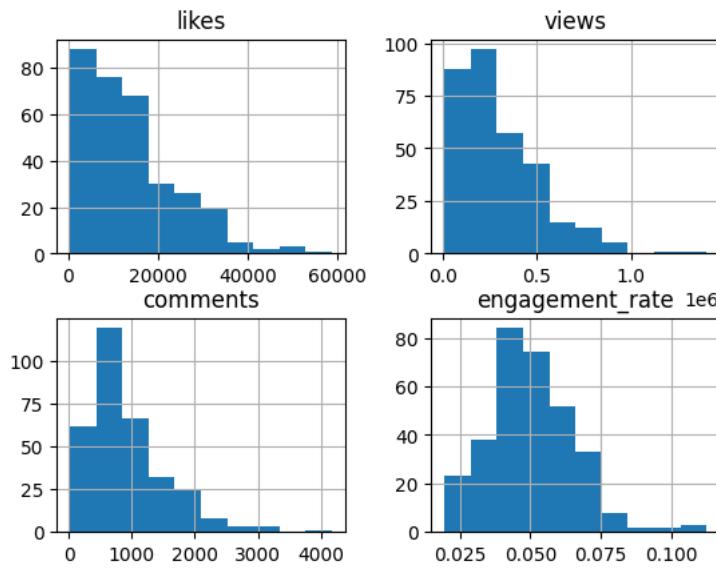


Figura 108 Histograma Malbert

Observando los histogramas, la distribución de likes, vistas y comentarios en sus videos muestra un sesgo hacia la derecha, indicando que la mayoría de sus videos tienen interacciones más bajas, pero hay unos pocos que destacan con números significativamente altos. Estos valores atípicos sugieren que ocasionalmente publica contenido que resuena de manera excepcional con su audiencia. Por otro lado, la tasa de engagement presenta una distribución más uniforme, lo que sugiere una interacción constante de la audiencia en términos de proporción, independientemente del número total de vistas.

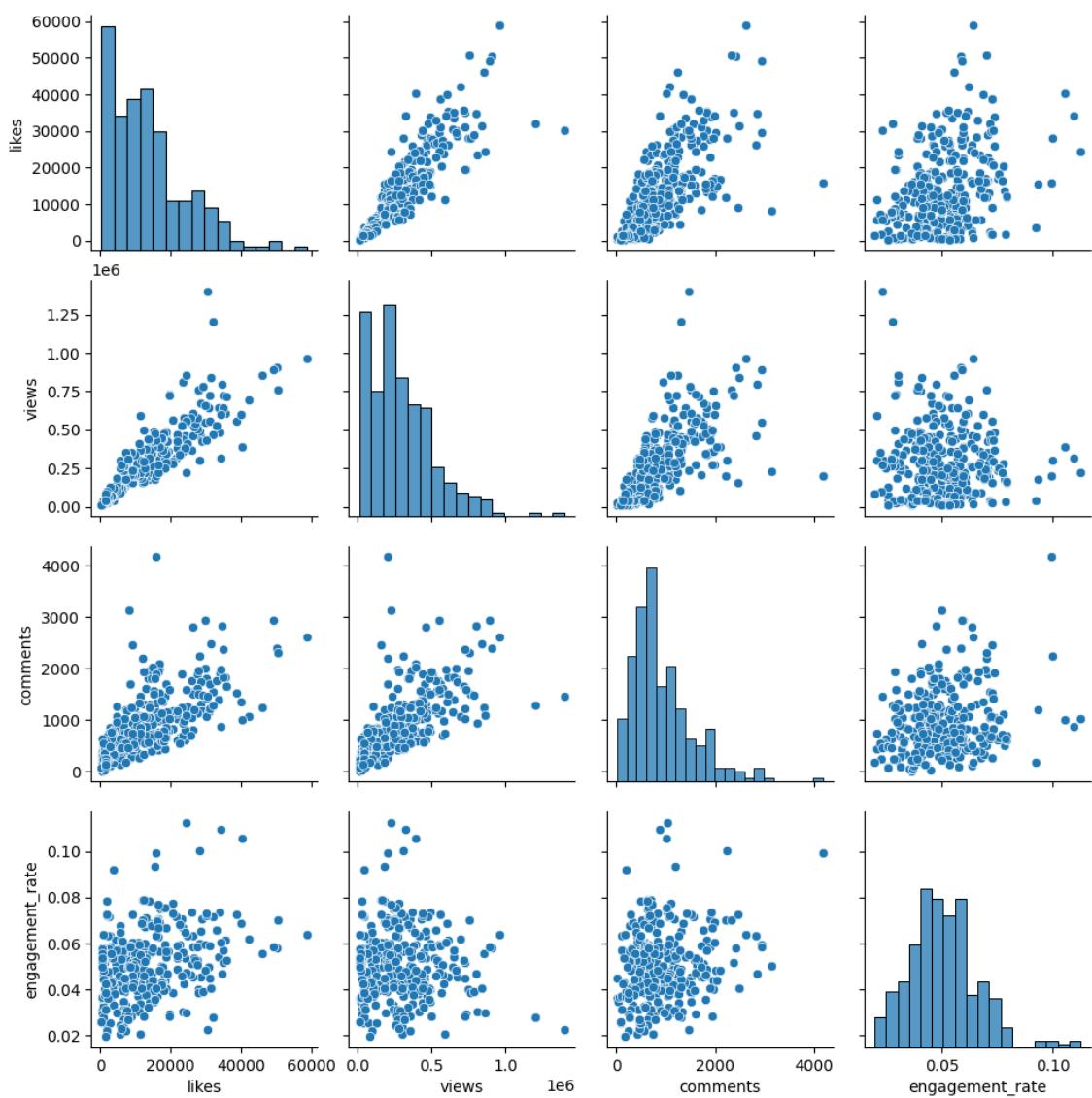


Figura 109 Scatterplot Malbert

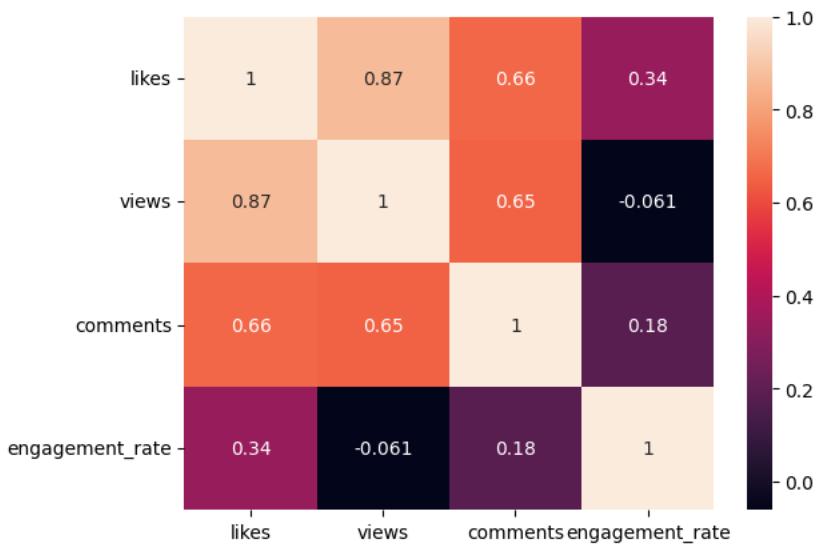


Figura 110 Matriz correlación Malbert

Después de mirar el scatterplot y matriz de correlaciones, podemos concluir que mientras que likes, vistas y comentarios tienden a moverse juntos en términos de correlación, la tasa de engagement no siempre sigue esta tendencia, especialmente en relación con las vistas. Esto sugiere que el simple hecho de tener muchas vistas no garantiza una alta interacción por parte de la audiencia.

Serie Temporal

- Análisis Mensual: La actividad muestra fluctuaciones a lo largo del tiempo, con dos picos notables. El primero se observa en el invierno de 2018 y el segundo en los primeros dos meses de 2020. Estos picos pueden indicar momentos de mayor interacción o publicación de contenido relevante.

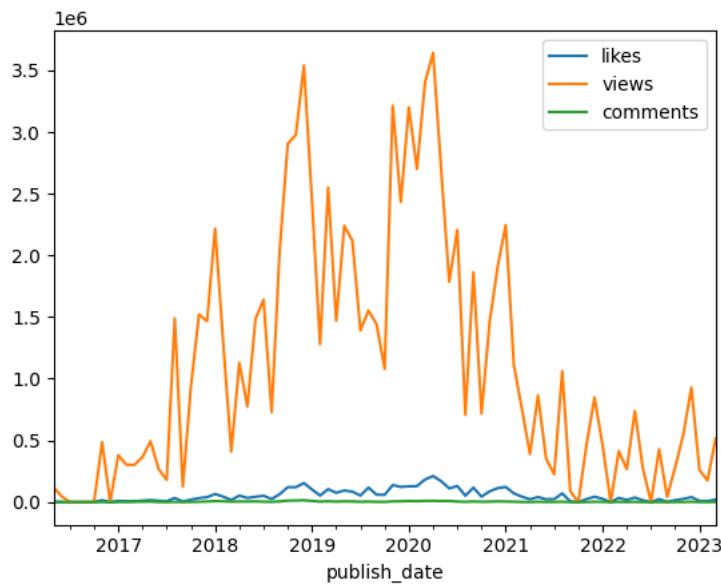


Figura 111 Serie Temporal mes Malbert

- Análisis Semanal: Malbert muestra una clara preferencia por publicar videos al inicio y a mitad de semana, especialmente los lunes, martes y jueves. Sin embargo, es notable que mantiene una presencia constante durante toda la semana, publicando contenido todos los días.

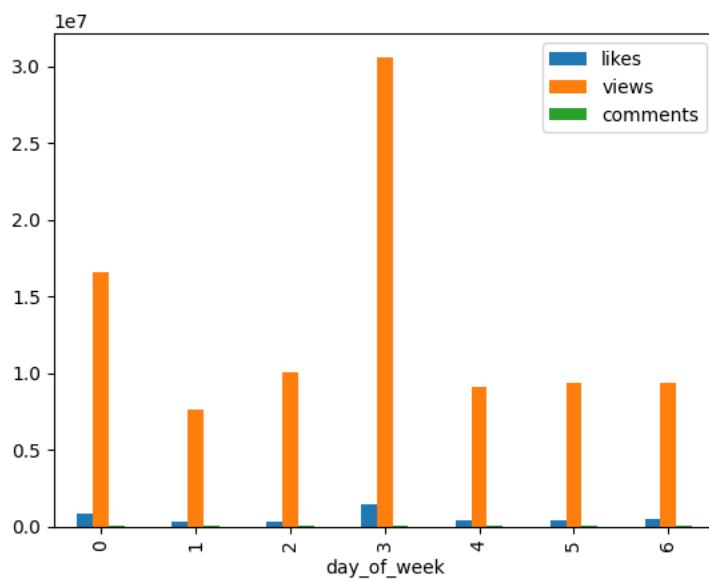


Figura 112 Serie Temporal semana Malbert

- Análisis Horario: La ventana horaria preferida para publicar sus videos es entre las 14:00 y las 19:00 horas, siendo las 17:00 horas el momento en el que más frecuentemente lanza nuevo contenido. Esta elección puede estar orientada a captar una mayor audiencia en horas pico de visualización.

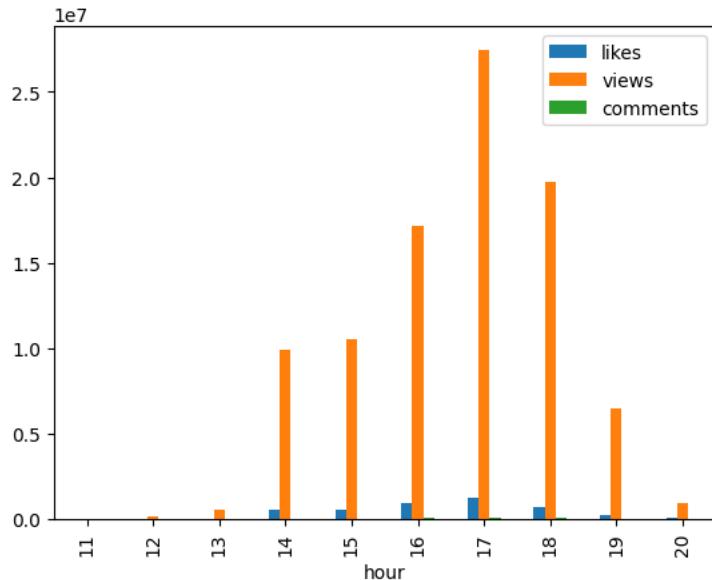


Figura 113 Serie Temporal hora Malbert

El video más destacado de Malbert lleva por título "COMIENDO UNA SEMANA CON 1€ AL DÍA | MALBERT". Este título sugiere un desafío o experimento en el que Malbert intenta subsistir alimentándose con un presupuesto extremadamente limitado de 1€ al día durante una semana. Es probable que este tipo de contenido, que combina entretenimiento con una perspectiva realista de la vida con recursos limitados, haya capturado la atención y el interés de una amplia audiencia, llevándolo a ser el video más popular de su canal.

Minería de Textos

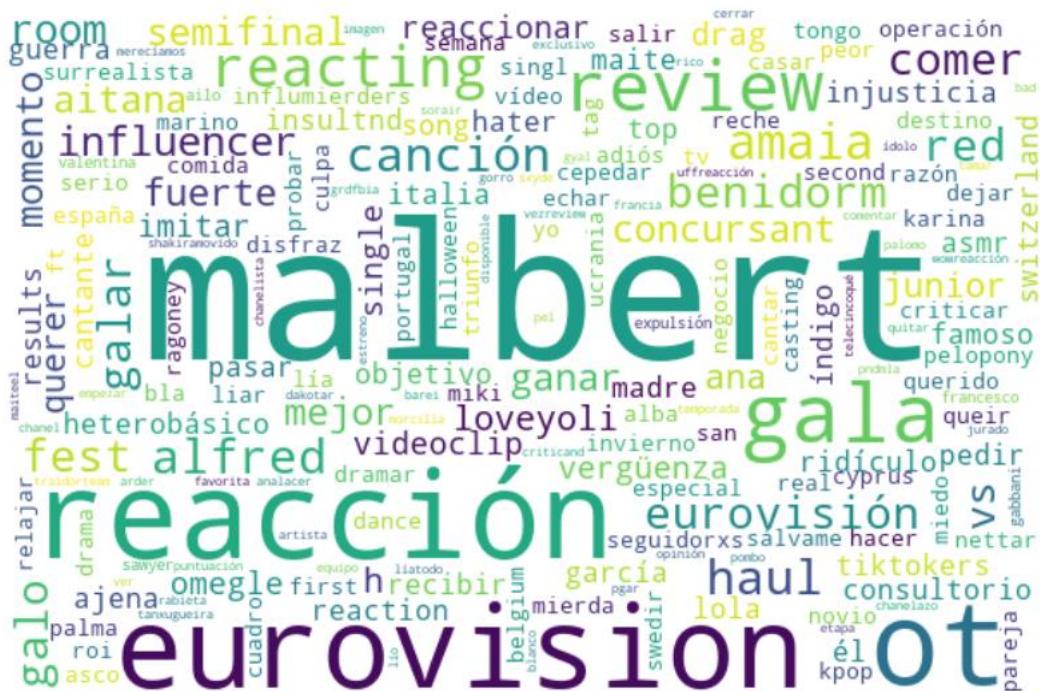


Figura 114 Wordcloud título Malbert

En la nube de palabras relacionada con los títulos de los videos de Malbert, el término "malbert" es el más frecuente, lo que indica que es una marca personal fuerte y recurrente en sus contenidos. Las palabras "reacción", "ot" (probablemente refiriéndose a "Operación Triunfo", un conocido programa de talentos en España) y "eurovision" sugieren que muchos de sus videos están centrados en reacciones y comentarios sobre programas de música y eventos relacionados. Las palabras "gala", "review", "amaia", "canción", "alfred", y "benidorm fest" refuerzan esta idea, indicando que Malbert podría estar especializado en comentar y reaccionar a eventos musicales y figuras populares en el ámbito musical español. Por otro lado, términos como "haul" y "red" podrían indicar una diversidad en sus contenidos, abordando temas de moda o tendencias actuales.



Figura 115 Wordcloud descripción Malbert

En la nube de palabras relacionada con las descripciones de los videos de Malbert, el término "itsmalbert" destaca como el más frecuente, lo que sugiere que es su identificador o marca personal en las redes sociales. Las palabras "contacto", "red", "instagram", "twitter", "facebook" y "holamalbertgmailcom" indican que Malbert proporciona múltiples medios para que sus seguidores y posibles colaboradores se pongan en contacto con él. Además, términos como "vídeo", "malbert", "canal", "hacer" y "gustar" reflejan que las descripciones suelen hablar sobre el contenido de sus videos, su canal y posiblemente invitan a la audiencia a interactuar, ya sea dando "me gusta" o participando de alguna otra manera. Las palabras "personal" y "business" podrían indicar que

Malbert diferencia entre consultas personales y profesionales en sus descripciones.



Figura 116 Wordcloud transcripción Malbert

En la nube de palabras basada en las transcripciones de los videos de Malbert, la palabra "música" es la más frecuente, lo que sugiere que Malbert a menudo comenta o hace referencia a la música de fondo en sus videos. La palabra "xxxx" aparece en segundo lugar, indicando que Malbert utiliza palabras de insulto o términos fuertes con cierta frecuencia. Las palabras "decir", "canción", "gustar", "ver" y "vídeo" reflejan que Malbert habla sobre lo que está diciendo o mostrando en sus videos, y posiblemente da su opinión sobre ciertas canciones o temas. Las palabras "querer", "venir", "pasar", "encantar" y "gente" sugieren que Malbert a menudo expresa sus deseos, opiniones y reacciones a eventos o situaciones, y también se refiere a la gente en general o a su audiencia. Las palabras "dejar", "año" y "favor" podrían indicar que hace peticiones o referencias a eventos pasados o futuros en sus videos.

Common Phrases:

dejar botón maravilloso: 201
 vídeo esperar gustar: 186
 botón maravilloso suscribir: 179
 maravilloso suscribir canal: 176
 suscribir canal vídeo: 161
 vídeo ver vídeo: 156
 canal vídeo ver: 129
 ver vídeo música: 123
 chico vídeo esperar: 107
 esperar gustar dar él: 80
 gustar dar él gustar: 80
 música aplauso música: 76
 dar él gustar gustar: 76
 poder dejar botón: 66
 gusto poder dejar: 63
 gente vídeo esperar: 58
 gustar gusto poder: 50
 aplauso música aplauso: 49
 ay ay ay: 36
 dalar gustar gustar: 35

Figura 117 Frases comunes Malbert

En las frases comunes de Malbert, se observa una fuerte tendencia hacia la interacción con su audiencia y la promoción de su contenido. Frases como "dejar botón maravilloso", "vídeo esperar gustar" y "botón maravilloso suscribir" sugieren que Malbert a menudo anima a sus espectadores a interactuar con sus videos, ya sea dando "me gusta" o suscribiéndose a su canal. La repetición de palabras como "vídeo", "maravilloso", "suscribir" y "canal" indica la importancia que da a la promoción y al engagement con su contenido. Además, la mención frecuente de "música" y frases como "música aplauso música" y "aplauso música aplauso" sugieren que la música juega un papel importante en sus videos, ya sea como música de fondo o como tema de discusión. Las frases "ay ay ay" y "dalar gustar gustar" podrían ser expresiones o muletillas características de Malbert, que utiliza para enfatizar o reaccionar a ciertos momentos en sus videos. En general, estas frases reflejan el estilo interactivo y animado de Malbert y su enfoque en la promoción y el engagement con su contenido.

```
Topics:
Topic: 0, Words: 0.000="decir" + 0.007="xxxx" + 0.006="ver" + 0.005="querer" + 0.005="venir" + 0.005="libro" + 0.004="persona" + 0.004="estar" + 0.004="pasar" + 0.004="gente"
Topic: 1, Words: 0.014="i" + 0.013="and" + 0.012="you" + 0.010="the" + 0.009="to" + 0.006="it" + 0.005="that" + 0.005="my" + 0.005="like" + 0.005="all"
Topic: 2, Words: 0.015="música" + 0.013="xxxx" + 0.012="canción" + 0.011="decir" + 0.010="gustar" + 0.009="ver" + 0.008="video" + 0.006="actuación" + 0.006="pasar" + 0.005="eurovisión"
Topic: 3, Words: 0.016="música" + 0.011="xxxx" + 0.010="video" + 0.009="ver" + 0.009="decir" + 0.007="gustar" + 0.006="querer" + 0.006="venir" + 0.005="gente" + 0.005="encantar"
Topic: 4, Words: 0.021="música" + 0.015="canción" + 0.011="xxxx" + 0.010="gustar" + 0.008="decir" + 0.008="ver" + 0.006="video" + 0.005="querer" + 0.005="encantar" + 0.005="aplauso"
```

Figura 118 Topic Modeling

En función de los tópicos top 5 de los resultados de Topic Modeling, se observa que:

- Tópico 0: Este tópico parece centrarse en las interacciones y opiniones de Malbert. Las palabras clave son "decir", "xxxx" (palabra de insulto), "ver", "querer", "venir", "libro", "persona", "estar", "pasar" y "gente". Esto sugiere discusiones generales, opiniones y posiblemente reseñas de libros o eventos.
- Tópico 1: Este tópico tiene un enfoque en el inglés, con palabras como "i", "and", "you", "the", "to", entre otras. Esto podría indicar segmentos de sus videos donde Malbert habla o comenta en inglés, o posiblemente reacciona a contenido en inglés.
- Tópico 2: Aquí, la música y las canciones son prominentes. Las palabras clave son "música", "xxxx", "canción", "decir", "gustar", "ver", "vídeo", "actuación" y "eurovisión". Esto sugiere que Malbert podría estar comentando o reaccionando a actuaciones musicales, especialmente relacionadas con Eurovisión.

- Tópico 3: Este tópico también se centra en la música y las reacciones de Malbert. Las palabras clave son "música", "xxxx", "vídeo", "ver", "decir", "gustar", "querer", "venir" y "gente". Esto indica discusiones y opiniones sobre música y videos, así como interacciones con su audiencia.
- Tópico 4: Al igual que el Tópico 2, este tópico se centra en la música, pero con un enfoque más en las reacciones y opiniones. Las palabras clave son "música", "canción", "xxxx", "gustar", "decir", "ver", "vídeo", "querer", "encantar" y "aplauso". Esto sugiere momentos de disfrute y apreciación musical, así como reacciones positivas a ciertas canciones o actuaciones.

Unique Terms:

canción:	23.147119943936353
xxxx:	22.279390618344234
actuación:	11.312854025647377
eurovisión:	10.474716292455563
encantar:	9.64823517329264
cantar:	9.10539504310818
año:	8.89702820928666
aplauso:	8.805097617319893
gente:	8.464589518882391
favor:	8.326469222881107

Figura 119 Términos únicos Malbert

Para encontrar los términos únicos, se observa que:

- canción: Es el término más frecuente, lo que indica que Malbert discute o menciona canciones con regularidad en sus contenidos.
- xxxx: Esta palabra, que representa un insulto, es también muy mencionada, lo que sugiere que Malbert podría usar lenguaje fuerte o reaccionar de manera intensa en sus videos.
- actuación y eurovisión: Estos términos sugieren que Malbert podría estar comentando o reaccionando a actuaciones musicales, especialmente relacionadas con el festival de Eurovisión.
- encantar y aplauso: Estas palabras indican una reacción positiva hacia ciertos contenidos, mostrando aprecio o admiración.
- cantar: Este término, junto con "canción", sugiere que la música y el acto de cantar son temas recurrentes en sus videos.
- año: Malbert podría hacer referencias temporales a menudo, tal vez recordando eventos pasados o comentando sobre acontecimientos anuales.
- gente: Indica que Malbert habla sobre personas o reacciona a la gente en sus videos, ya sea mencionando a otros creadores, artistas o a su audiencia.
- favor: Este término podría ser usado en el contexto de pedir algo a su audiencia, como "por favor suscríbete" o "dale like por favor".

Network



Figura 120 Netwrok Malbert

Basándonos en la red generada, es evidente que Malbert ocupa una posición central y destacada en el ecosistema de YouTubers analizado. Su extenso número de conexiones señala claramente que es el creador principal en torno al cual se organiza esta red. Las conexiones que tiene con otros YouTubers sugieren la posibilidad de colaboraciones, menciones o incluso reacciones a los contenidos que produce.

En un análisis más detallado, se observa que después de Malbert, los YouTubers cepeda y aitana son los más mencionados. Esta recurrente mención de ambos podría indicar que podrían ser colaboradores frecuentes o temas de discusión habituales en el contenido de Malbert. Asimismo, figuras como alfred, natalia, miriam, maría y amaya también tienen una cantidad significativa de menciones, lo que señala su relevancia en el contexto de Malbert. Esto podría ser debido a colaboraciones pasadas o a la discusión de temas específicos relacionados con estos YouTubers.

Resulta interesante observar que entre los nombres mencionados en la red, se encuentran diversas personalidades del entretenimiento en España, tales como anabel pantoja, kiko matamoros, mónica naranjo, leticia sabater, ana guerra, blas canto, ana torroja y shakira. Esto podría sugerir que Malbert se involucra en comentarios, reacciones o discusiones acerca de los contenidos de estas figuras o de las noticias relacionadas con ellas.

No obstante, también se perciben nombres como nevera o jolie, que podrían no hacer referencia directa a otros YouTubers. En lugar de ello, es posible que estos términos se relacionen con temas o tendencias específicas que Malbert ha abordado en su contenido.

Mostopapi

EDA

	likes	views	comments	year	month	\
count	189.000000	1.890000e+02	189.000000	189.000000	189.000000	
mean	97360.095238	1.284889e+06	2751.248677	2019.624339	6.100529	
std	95461.972907	1.140183e+06	3348.611579	1.902305	3.721051	
min	512.000000	1.047000e+04	34.000000	2016.000000	1.000000	
25%	14091.000000	2.795520e+05	524.000000	2018.000000	3.000000	
50%	74841.000000	1.047367e+06	1678.000000	2020.000000	6.000000	
75%	138016.000000	1.851073e+06	3514.000000	2021.000000	10.000000	
max	432912.000000	6.271345e+06	21227.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	189.000000	189.000000	189.000000			
mean	4.894180	18.063492	0.071895			
std	1.937705	1.556110	0.031510			
min	0.000000	11.000000	0.012521			
25%	3.000000	17.000000	0.049962			
50%	6.000000	18.000000	0.065514			
75%	6.000000	19.000000	0.088260			
max	6.000000	23.000000	0.200387			

Figura 121 Estadística Mostopapi

- Volumen de Datos: Se han analizado un total de 189 videos de Mostopapi.
- Interacción:
 - o Likes: En promedio, los videos de Mostopapi reciben alrededor de 97,360 likes. El video más popular ha obtenido 432,912 likes, mientras que el menos popular ha tenido 512 likes.
 - o Vistas: Los videos tienen un promedio de aproximadamente 1,284,889 vistas. Un video ha alcanzado un máximo de 6,271,345 vistas, mientras que el mínimo fue de 10,470 vistas.
 - o Comentarios: En promedio, los videos reciben alrededor de 2,751 comentarios. Un video ha llegado a tener un máximo de 21,227 comentarios, mientras que el mínimo fue de 34 comentarios.
- Tiempo:
 - o Año: Los videos se han publicado desde el año 2016 hasta el año 2023.
 - o Mes: El mes promedio de publicación es aproximadamente junio.
 - o Día de la semana: La mayoría de los videos se publican a principios de semana, con un promedio de 4.89 (donde 0 es lunes y 6 es domingo).
 - o Hora: La mayoría de los videos se publican alrededor de las 6:00 p.m.
- Tasa de Engagement: La tasa de engagement promedio es del 7.19%. Esto indica que, en promedio, alrededor del 7.19% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 20.04%, mientras que la más baja es del 1.25%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos.

Esto sugiere que, aunque Mostopapi tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

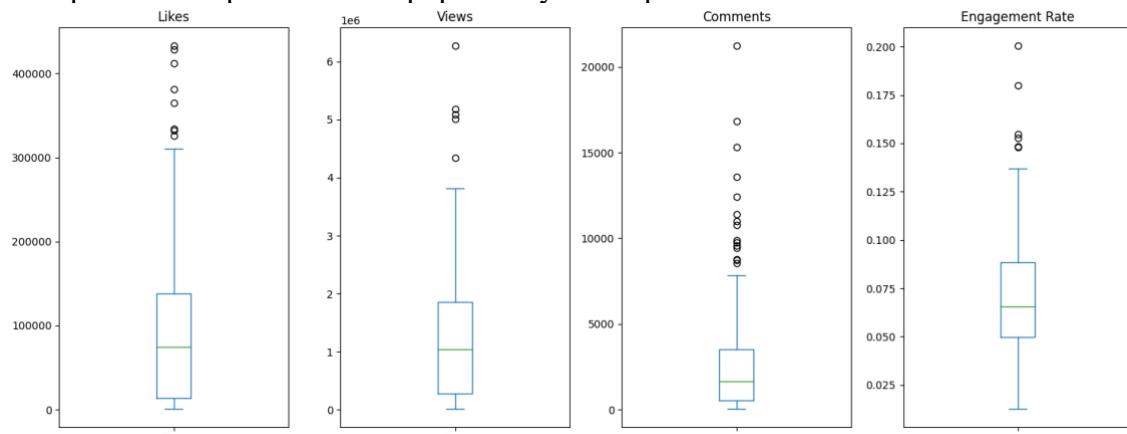


Figura 122 Boxplot Mostopapi

Los boxplots resaltan la variabilidad en las métricas de likes, vistas y comentarios para los videos de Mostopapi. Algunos videos logran un alto nivel de interacción en forma de likes, vistas y comentarios, lo cual se puede ver por los outliers en el lado superior del bigote, mientras que otros tienen menos interacción. La tasa de engagement, aunque más consistente, también presenta algunas variaciones.

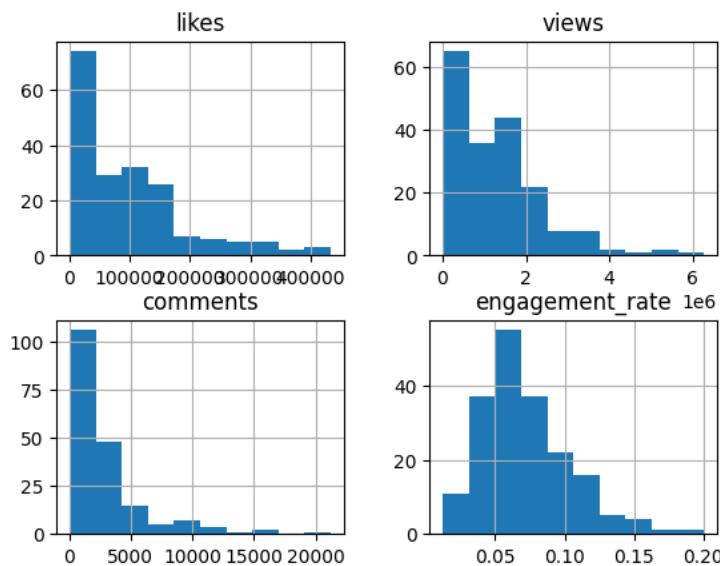


Figura 123 Histograma Mostopapi

La visualización de histogramas proporciona una comprensión rápida y clara de cómo se distribuyen las métricas en los videos de Mostopapi. En el caso de likes y vistas, los histogramas exhiben una inclinación hacia la derecha, con la mayoría de los videos acumulando valores moderados y unos pocos casos excepcionales que generan una notable cantidad de interacción. En el caso de comentarios, se observa una distribución similar, aunque con menos variabilidad en los extremos. Por otro lado, la tasa de engagement muestra una distribución más concentrada, indicando una mayor consistencia en la interacción de los videos. Los histogramas resaltan de manera eficaz cómo algunos videos se destacan con niveles significativamente más altos de interacción, mientras que la mayoría mantiene valores más moderados, proporcionando una instantánea visual de la diversidad en la respuesta de la audiencia a los contenidos de Mostopapi.

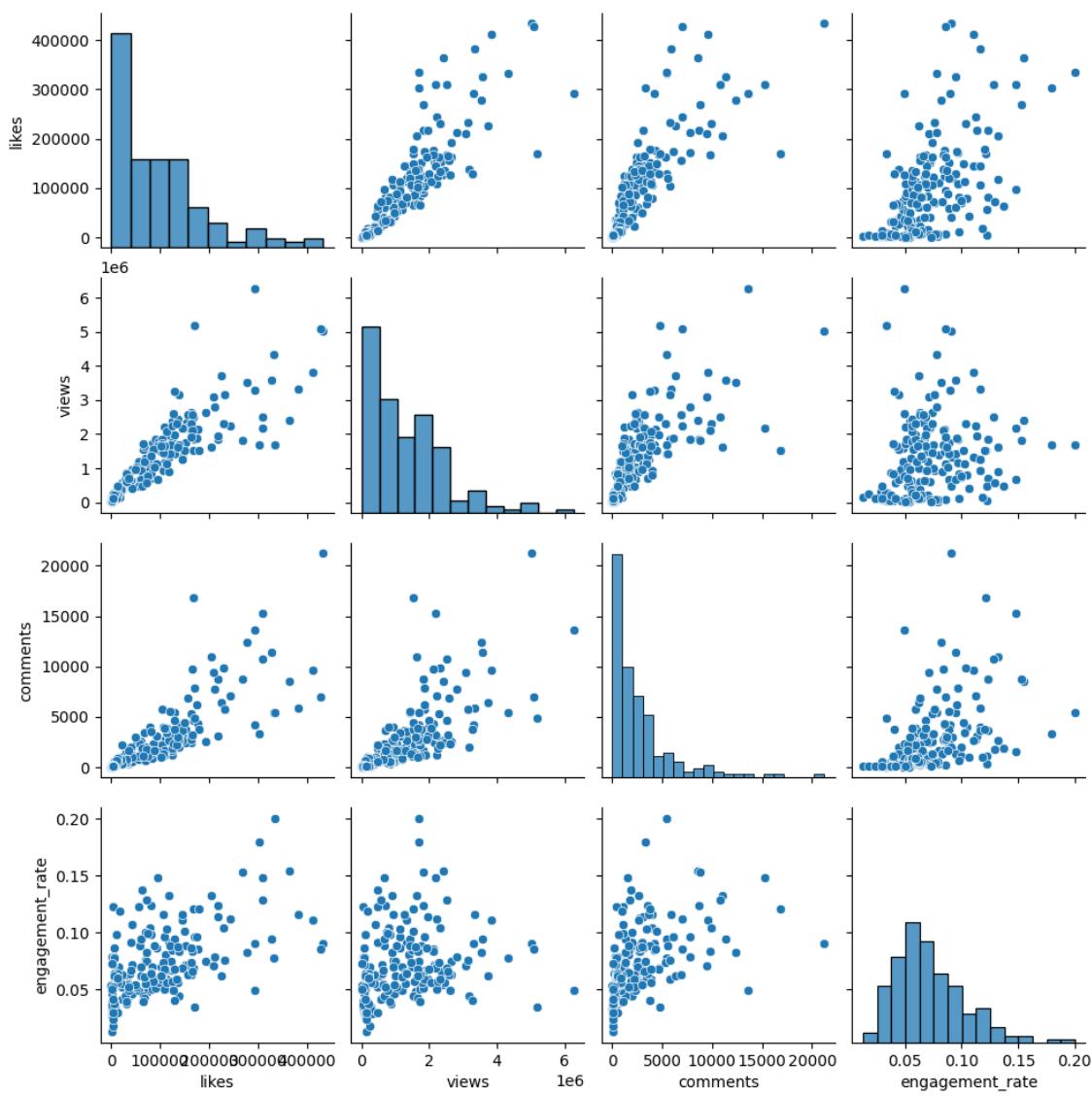


Figura 124 Sactterplot Mostopapi

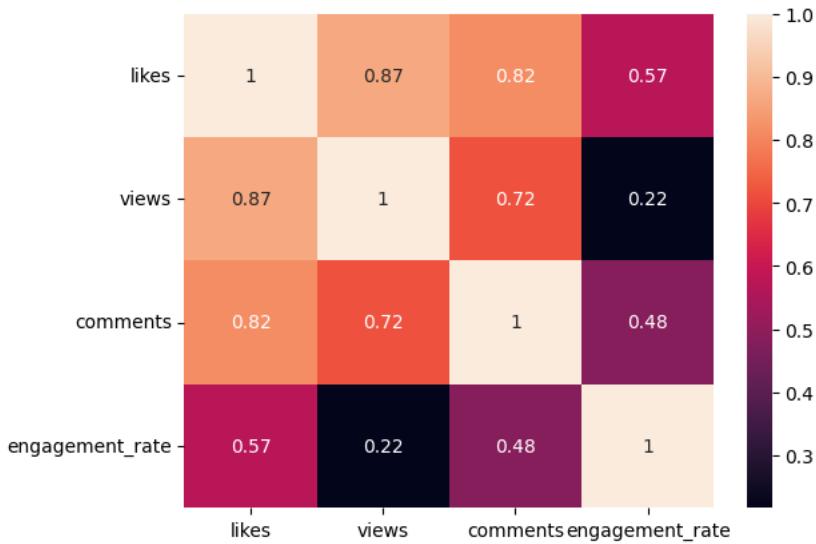


Figura 125 Matriz correlación Mostopapi

Según los dos gráficos arriba, las correlaciones entre las métricas en los videos de Mostopapi revelan patrones interesantes en términos de interacción y audiencia. Se puede observar una fuerte correlación positiva entre likes y views, así como entre likes y comments, lo que sugiere que los videos que reciben más likes también tienden a tener más vistas y comentarios. Además, la correlación entre comments y views también es significativa, indicando que los videos con más vistas también suelen generar más comentarios. La tasa de engagement muestra correlaciones más modestas con las otras métricas, sugiriendo que, aunque existe cierta relación, no es tan fuerte como las observadas entre likes, views y comments.

Serie Temporal

- Análisis Mensual: Mostopapi experimentó picos significativos de actividad o interacción en enero y otoño de 2020, así como en enero de 2022. Estos picos podrían indicar la publicación de contenido muy popular o eventos significativos relacionados con el YouTuber en esos meses. Es notable que antes del año 2020, su actividad era relativamente baja, lo que sugiere que su canal pudo haber ganado popularidad o haber cambiado su estrategia de contenido alrededor de ese año. Sin embargo, en los últimos meses, ha habido una disminución en la actividad, lo que podría indicar una pausa en la creación de contenido o una disminución en la interacción del público.

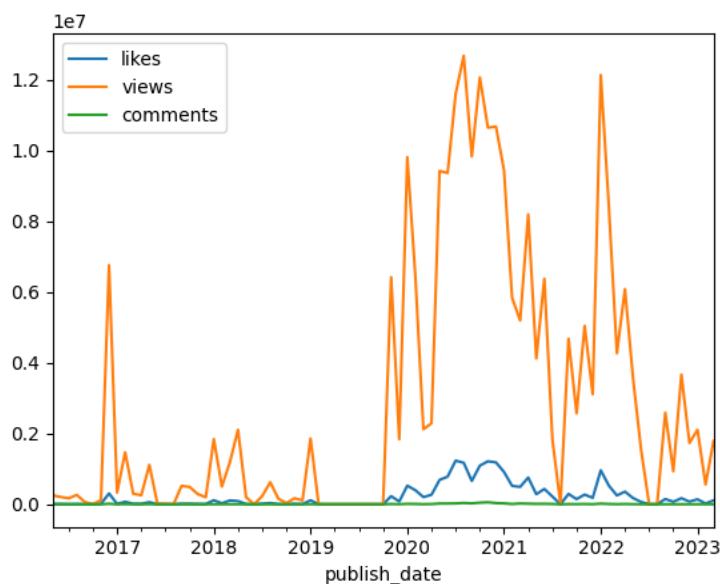


Figura 126 Serie Temporal mes Mostopapi

- Análisis Semanal: Mostopapi muestra una clara preferencia por publicar o interactuar con su audiencia los domingos. Esto podría ser una estrategia para captar a una audiencia más amplia que está más disponible ese día, o simplemente una rutina personal de publicación.

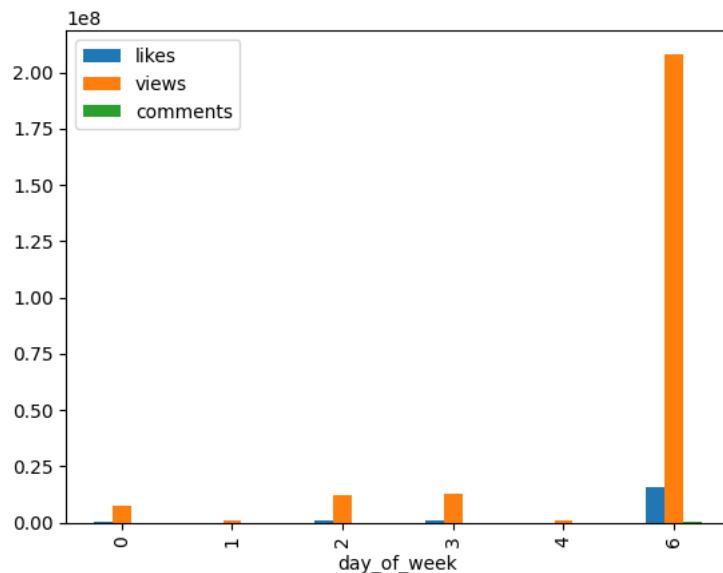


Figura 127 Serie Temporal semana Mostopapi

- Análisis Horario: En cuanto a la hora del día, Mostopapi tiende a publicar o estar más activo entre las 17 y 19 horas. Esta franja horaria es comúnmente conocida como "prime time" en muchos países, lo que sugiere que está buscando maximizar la visibilidad y el alcance de su contenido durante las horas en que su audiencia está más activa o disponible para consumir contenido.

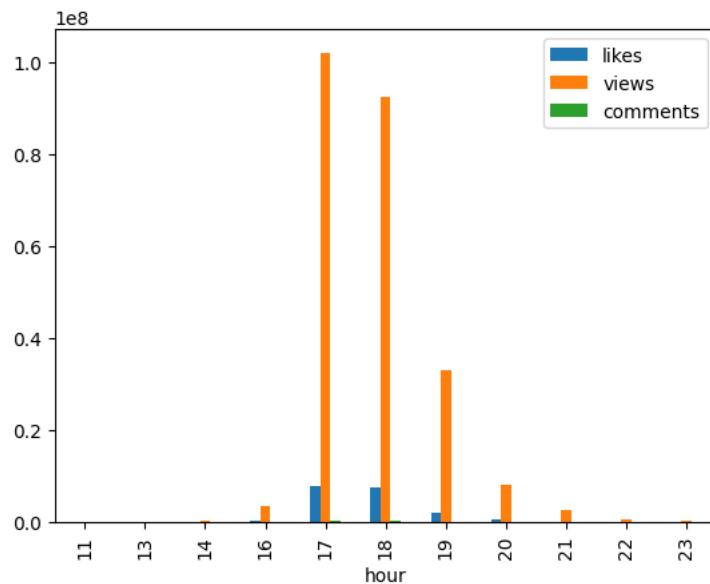


Figura 128 Serie Temporal hora Mostopapi

El video más destacado de Mostopapi lleva por título "CAMBIO LA VIDA DE UNA PROSTITUTA POR NAVIDAD". Este título sugiere un contenido que aborda un tema social y posiblemente emotivo, en el que Mostopapi podría estar realizando un acto de bondad o caridad durante la temporada navideña. Es probable que este video haya capturado la atención de la audiencia debido a su naturaleza conmovedora y al contraste que presenta entre la festividad de la Navidad y la dura realidad de la prostitución. La elección de este tema podría haber generado una amplia gama de reacciones y discusiones entre los espectadores, contribuyendo a su popularidad. Es esencial abordar este tipo de contenidos con sensibilidad y respeto, dada la naturaleza delicada del tema.

Minería de Textos



Figura 129 Wordcloud título Mostopapi

Basándonos en la nube de palabras del título de Mostopapi, se puede inferir:

- El término "hábito" es el más frecuente, lo que sugiere que Mostopapi podría estar abordando temas relacionados con rutinas, comportamientos o prácticas recurrentes en sus videos.
 - Palabras como "joven", "negro", "sexual", "porno", "racista" y "blanco" indican que aborda temas sociales, raciales y sexuales, lo que sugiere que no teme tratar temas controvertidos o sensibles.
 - Términos como "consultorio", "cama", "video", "novia" y "pene" sugieren una variedad de temas, desde relaciones y sexualidad hasta reflexiones personales y análisis de contenido multimedia.
 - La inclusión de su propio nombre, "mostopapi", podría indicar videos donde se centra en experiencias personales o reflexiones propias.

En general, Mostopapi parece abordar una amplia gama de temas en sus títulos, desde cuestiones sociales y raciales hasta temas más ligeros y personales. Es probable que esta variedad de contenidos sea una de las razones de su popularidad, ya que puede atraer a una amplia audiencia con intereses diversos.



Figura 130 Wordcloud descripción Mostopapi

Basándonos en la nube de palabras de las descripciones de Mostopapi, se puede deducir como:

- "Mostopapi" es el término más frecuente, lo que indica que la marca personal y el branding son esenciales en sus descripciones, reforzando su identidad en línea.
 - Las palabras "red", "social", "instagram", "twitter" y "facebook" sugieren que Mostopapi tiene una presencia activa en múltiples plataformas de redes sociales y que promociona activamente sus otros canales en sus descripciones.
 - "Hurra" podría ser una expresión recurrente o un término relacionado con su contenido, quizás un eslogan o una referencia a algo específico de su canal.
 - Términos como "tocar", "entrevistar", "joven", "hábito", "sexual", "negro" y "pregunta" indican la variedad de temas que aborda. Podría estar realizando entrevistas, discutiendo hábitos y comportamientos, y abordando temas raciales y sexuales.
 - La presencia de palabras como "entrevistar" y "pregunta" sugiere que podría estar interactuando con sus seguidores o invitados, realizando entrevistas o respondiendo a preguntas frecuentes.

En resumen, las descripciones de Mostopapi reflejan una combinación de promoción personal, interacción con la comunidad y discusión de temas variados y a menudo profundos. Su enfoque parece ser tanto informativo como interactivo, buscando conectar con su audiencia en múltiples niveles.



Figura 131 Wordcloud transcripción Mostopapi

Por lo último en la nube de palabras de las transcripciones de Mostopapi, se puede interpretar:

- "Valer" es el término más utilizado, lo que sugiere que Mostopapi a menudo hace referencia al valor o la importancia de algo, o podría ser una expresión recurrente que utiliza.
 - Palabras como "gustar", "querer" y "decir" indican que Mostopapi a menudo comparte sus opiniones y sentimientos, y habla sobre lo que le gusta o quiere.
 - "Ver" y "vídeo" sugieren que a menudo comenta sobre lo que está viendo o haciendo referencia a otros videos o contenidos.
 - "xxxx" indica que hay ciertas palabras o frases que han sido censuradas o excluidas, posiblemente lenguaje fuerte o inapropiado.
 - "Música" sugiere que a menudo comenta o discute sobre música en sus videos.
 - Términos como "pasar", "cosa", "persona", "estar" y "tener" reflejan una narrativa descriptiva y conversacional, donde comparte eventos, experiencias y observaciones con su audiencia.
 - "Negro" podría indicar discusiones sobre temas raciales o simplemente una descripción de algo. Dado que también aparece en las descripciones, podría ser un tema recurrente en su contenido.
 - "Gente" sugiere que a menudo habla sobre otras personas, ya sea en general o refiriéndose a individuos específicos.

En resumen, las transcripciones de Mostopapi reflejan un estilo conversacional y descriptivo, donde comparte sus opiniones, experiencias y observaciones sobre una variedad de temas, desde música hasta cuestiones raciales y personales. Su enfoque parece ser auténtico y directo, buscando conectar con su audiencia a través de discusiones abiertas y honestas.

Common Phrases:

- valer valer valer: 92
- esperar gustar vídeo: 88
- negro subir youtube: 69
- papi tve negro: 63
- ver rapear ocurrir: 51
- tve negro subir: 48
- negro manzán podrido: 47
- oveja negro manzán: 45
- vestí piensa nube: 38
- estar preparado hábito: 37
- youtube vestí piensa: 36
- edad perdistir virginidad: 35
- subir youtube vestí: 34
- papi tv negro: 33
- decir oveja negro: 32
- suscribiro dar él like: 31
- mostrar papi tv: 30
- persona haber sexo: 29
- máximo persona haber: 28
- tv negro subir: 28

Figura 132 Frases comunes Mostopapi

Las frases comunes en el contenido de Mostopapi nos brindan una visión más detallada de los temas y estilos recurrentes en sus videos:

- "Valer valer valer" sugiere que Mostopapi a menudo enfatiza o reitera ciertos puntos, posiblemente como una expresión característica.
- "Esperar gustar vídeo" indica que a menudo solicita a su audiencia que muestre aprecio por su contenido, una práctica común entre los creadores de contenido.
- Las referencias a "negro" en combinación con palabras como "subir youtube", "tve", "manzán podrido" y "oveja" sugieren que Mostopapi podría estar abordando temas raciales o identitarios, o podría ser una forma metafórica de describir algo atípico o diferente.
- "Ver rapear ocurrir" y "estar preparado hábito" indican que Mostopapi podría estar comentando o reaccionando a ciertos eventos o tendencias.
- "Vestí piensa nube" y "edad perdistir virginidad" sugieren que aborda temas relacionados con la moda y las experiencias personales o tabúes.
- "Suscribiro dar él like" es otra indicación de que anima a su audiencia a interactuar con su contenido.
- "Mostrar papi tv" y "persona haber sexo" sugieren que discute o reacciona a programas de televisión o a temas relacionados con relaciones y experiencias personales.

Estas frases comunes reflejan que Mostopapi tiene un enfoque variado en su contenido, abordando desde temas personales y experiencias hasta reacciones a eventos actuales y tendencias. Su estilo parece ser directo y auténtico, y a menudo busca la interacción y el feedback de su audiencia.

```
Topics:
Topic: 0, Words: 0.014*"valer" + 0.010*"gustar" + 0.009*"querer" + 0.009*"ver" + 0.009*"xxxx" + 0.008*"video" + 0.008*"decir" + 0.007*"cosa" + 0.007*"pasar" + 0.007*"persona"
Topic: 1, Words: 0.017*"gustar" + 0.015*"valer" + 0.010*"querer" + 0.010*"decir" + 0.009*"ver" + 0.008*"cosa" + 0.008*"xxxx" + 0.008*"música" + 0.008*"persona" + 0.007*"video"
Topic: 2, Words: 0.018*"valer" + 0.014*"gustar" + 0.011*"querer" + 0.011*"decir" + 0.009*"ver" + 0.009*"persona" + 0.009*"pasar" + 0.008*"tener" + 0.008*"cosa" + 0.007*"xxxx"
Topic: 3, Words: 0.019*"valer" + 0.016*"negro" + 0.010*"video" + 0.010*"ver" + 0.010*"querer" + 0.009*"decir" + 0.008*"gustar" + 0.008*"pasar" + 0.007*"gente"
Topic: 4, Words: 0.011*"valer" + 0.010*"ver" + 0.010*"xxxx" + 0.009*"gustar" + 0.009*"poner" + 0.009*"decir" + 0.008*"música" + 0.007*"video" + 0.007*"pasar" + 0.007*"querer"
```

Figura 133 Topic Modeling Mostopapi

Mostopapi parece tener una variedad de temas recurrentes en su contenido:

- Tópico 0: Se centra en las reacciones y opiniones. Las palabras clave son "valer", "gustar", "querer", "ver", "xxxx" (probablemente un lenguaje fuerte o censurado), "vídeo", "decir", "cosa", "pasar" y "persona". Esto sugiere que Mostopapi a menudo comparte sus opiniones y reacciones a diferentes situaciones o contenidos.

- Tópico 1: Este tópico es similar al anterior pero con un énfasis más fuerte en la música y las opiniones personales. Las palabras clave son "gustar", "valer", "querer", "decir", "ver", "cosa", "xxxx", "música", "persona" y "vídeo".
- Tópico 2: Aquí, Mostopapi parece discutir sobre relaciones o interacciones personales. Las palabras clave son "valer", "gustar", "querer", "decir", "ver", "persona", "pasar", "tener", "cosa" y "xxxx".
- Tópico 3: Este tópico tiene un enfoque en la identidad y la música. Las palabras clave son "valer", "negro", "vídeo", "música", "ver", "querer", "decir", "gustar", "pasar" y "gente". Esto sugiere que Mostopapi podría estar discutiendo temas de identidad racial o cultural, además de compartir su amor por la música.
- Tópico 4: En este tópico, Mostopapi parece centrarse en compartir experiencias y música. Las palabras clave son "valer", "ver", "xxxx", "gustar", "poner", "decir", "música", "vídeo", "pasar" y "querer".

En general, Mostopapi aborda una variedad de temas en su contenido, desde compartir opiniones y reacciones hasta discutir temas de identidad y cultura, todo mientras incorpora música y entretenimiento en sus videos.

Unique Terms:
xxxx: 8.03250383123848
persona: 6.6694799477331435
poner: 6.495886954177328
sexo: 5.811788055301893
plan: 5.581753470358822
haber: 5.019619145080023
hacer: 4.557803604672745
hablar: 4.467863290344192
venir: 4.341755378950012
vida: 4.240239828156006

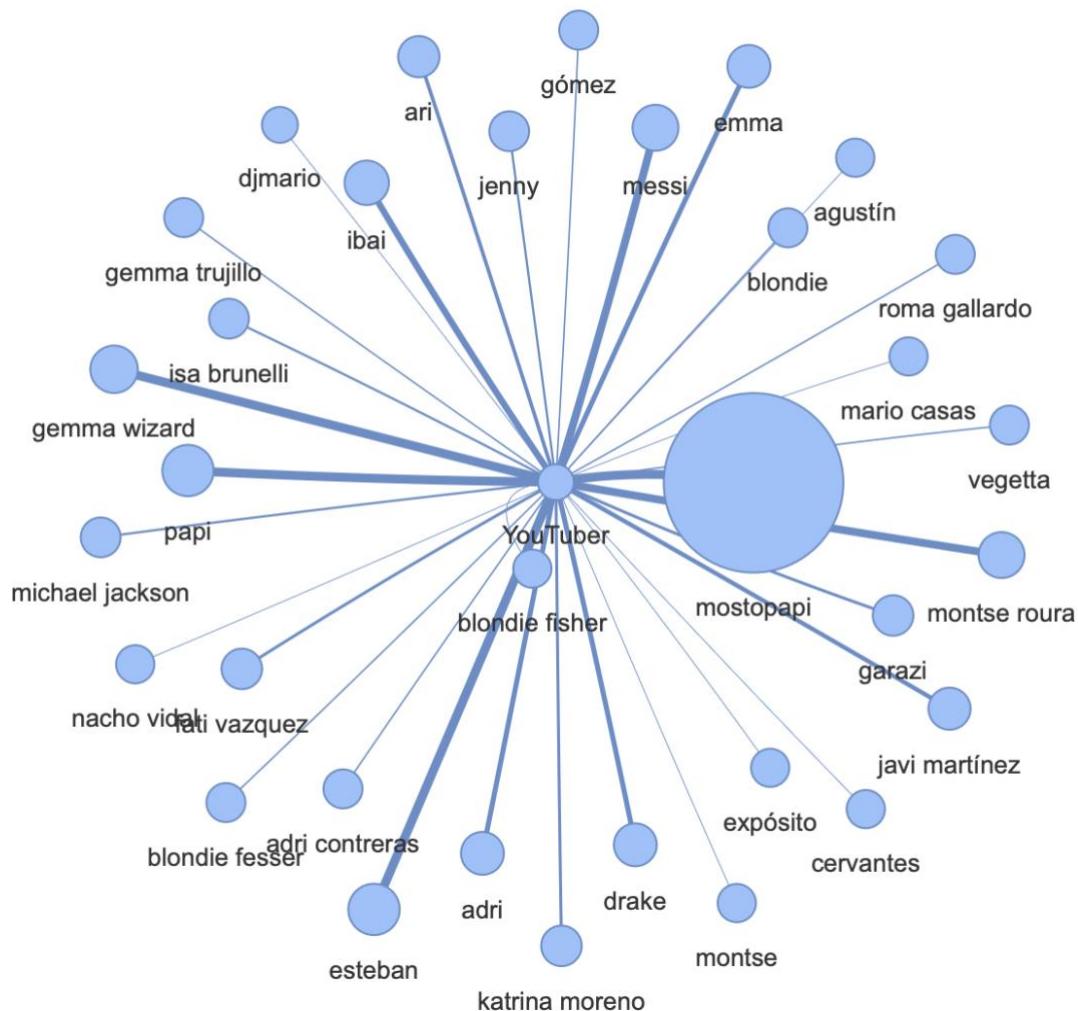
Figura 134 Términos únicos Mostopapi

Basándonos en los términos únicos:

- xxxx: Es el término más frecuente y, como se mencionó anteriormente, probablemente se refiere a un lenguaje fuerte o censurado.
- persona: Es un término general que sugiere que Mostopapi habla a menudo sobre individuos o se refiere a las personas en general en sus videos.
- poner: Este verbo puede indicar acciones o situaciones en las que algo se coloca o se presenta, quizás en el contexto de compartir opiniones o experiencias.
- sexo: Es un tema que Mostopapi aborda, lo que sugiere que podría discutir temas relacionados con la sexualidad, las relaciones o las experiencias personales.
- plan: Puede referirse a planes futuros, estrategias o intenciones que Mostopapi comparte o discute con su audiencia.
- haber: Es un verbo auxiliar en español, lo que indica que Mostopapi a menudo habla sobre experiencias pasadas o situaciones que han ocurrido.
- hacer: Otro verbo que indica acción, sugiriendo que Mostopapi a menudo discute actividades o acciones específicas.
- hablar: Indica que Mostopapi a menudo discute o comparte conversaciones, o se refiere a la acción de hablar sobre un tema en particular.
- venir: Puede referirse a eventos o situaciones que están por venir o a experiencias pasadas que tienen relevancia en el presente.
- vida: Este término sugiere que Mostopapi a menudo comparte o discute aspectos de su vida personal o reflexiona sobre la vida en general.

En resumen, estos términos únicos ofrecen una visión de los temas y enfoques recurrentes en el contenido de Mostopapi, que abarcan desde reflexiones personales hasta discusiones sobre relaciones, experiencias pasadas y planes futuros.

Network



Basándonos en la red de Mostopapi:

- Mostopapi es claramente el YouTuber principal en esta red con 189 menciones, lo que indica que es el foco principal de análisis o discusión.
- Esteban y Papi tienen 23 menciones cada uno, lo que sugiere que son figuras relevantes en relación con Mostopapi o en el mismo nicho de contenido.
- Gemma Wizard, Messi, y Montse Roura tienen menciones significativas, lo que indica que podrían haber colaborado con Mostopapi o ser mencionados en sus videos.
- Ibai, un conocido streamer y YouTuber español, también aparece en la lista, lo que sugiere posibles colaboraciones o discusiones relacionadas con él.
- Hay varios nombres que podrían ser de celebridades o figuras públicas, como Drake, Michael Jackson, y Messi, lo que indica que Mostopapi podría discutir o referirse a estas figuras en su contenido.

- Algunos nombres, como Blondie, Blondie Fesser, y Blondie Fisher, sugieren que podrían estar relacionados o ser variantes del mismo individuo o canal.
- Otros nombres en la lista, como Vegetta, Nacho Vidal, y Toni Kroos, sugieren una variedad de temas y discusiones en el contenido de Mostopapi, desde otros creadores de contenido hasta deportistas y celebridades.

En resumen, la red de YouTubers relacionados con Mostopapi muestra una variedad de figuras, desde otros creadores de contenido hasta celebridades y deportistas, lo que indica una amplia gama de temas y discusiones en su canal.

Orslok

EDA

	likes	views	comments	year	month	\
count	229.000000	2.290000e+02	229.000000	229.000000	229.000000	
mean	35220.179039	5.896466e+05	909.951965	2020.165939	5.820961	
std	19009.689531	3.723835e+05	693.721449	0.427633	2.826612	
min	4051.000000	8.383800e+04	122.000000	2019.000000	1.000000	
25%	21983.000000	3.312370e+05	470.000000	2020.000000	4.000000	
50%	31634.000000	5.018000e+05	730.000000	2020.000000	5.000000	
75%	44673.000000	7.833230e+05	1089.000000	2020.000000	8.000000	
max	126032.000000	2.715860e+06	5634.000000	2021.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	229.000000	229.000000	229.000000			
mean	2.978166	19.004367	0.066056			
std	1.925009	1.831733	0.019091			
min	0.000000	13.000000	0.034568			
25%	1.000000	19.000000	0.051391			
50%	3.000000	19.000000	0.063937			
75%	5.000000	20.000000	0.078051			
max	6.000000	23.000000	0.134890			

Figura 135 Estadística Orslok

- Volumen de Datos: Se han analizado un total de 229 videos de Orslok.
- Interacción:
 - Likes: En promedio, los videos de Orslok reciben aproximadamente 35,220 likes. El video más popular ha alcanzado 126,032 likes, mientras que el menos popular ha obtenido 4,051 likes.
 - Vistas: Los videos tienen un promedio de 589,646 vistas, con un video alcanzando un máximo de 2,715,860 vistas y un mínimo de 83,838 vistas.
 - Comentarios: En promedio, los videos reciben 910 comentarios, con un video que ha alcanzado un máximo de 5,634 comentarios y un mínimo de 122 comentarios.
- Tiempo:
 - Año: Los videos se han publicado desde 2019 hasta 2021.
 - Mes: El mes promedio de publicación es alrededor de junio.
 - Día de la semana: Orslok tiende a publicar más a mediados de semana, con el miércoles siendo el día más común, con un promedio de 2.98 (donde 0 es lunes y 6 es domingo).
 - Hora: La mayoría de los videos se publican alrededor de las 7:00 p.m.

- Tasa de Engagement: La tasa de engagement promedio es del 6.61%. Esto indica que, en promedio, el 6.61% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 13.49%, mientras que la más baja es del 3.46%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Orslok tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

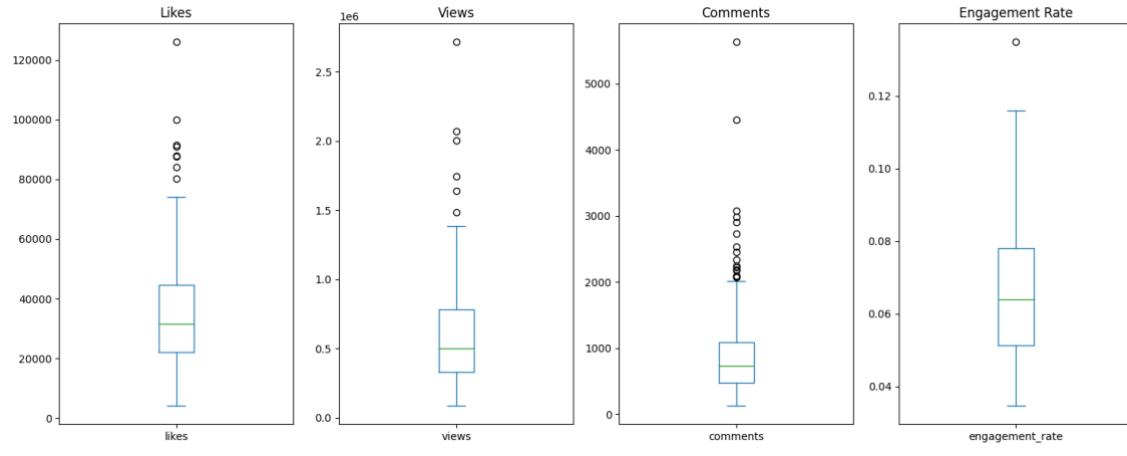


Figura 136 Boxplot Orslok

Basándonos en los boxplots de Orslok, es evidente que hay valores atípicos destacados en likes, vistas y comentarios, señalando que ciertos videos han tenido un impacto excepcionalmente alto en la audiencia. Aunque hay una variabilidad en la tasa de engagement, su distribución es más uniforme, indicando una consistencia en cómo la audiencia interactúa con el contenido. Esta consistencia en el engagement, junto con los valores atípicos en otras métricas, sugiere que mientras Orslok tiene videos que resonan ampliamente, mantiene un nivel estable de interacción en general, independientemente de la popularidad del video. La variabilidad podría estar influenciada por factores como el tema del video, su relevancia en el momento o la calidad de la presentación.

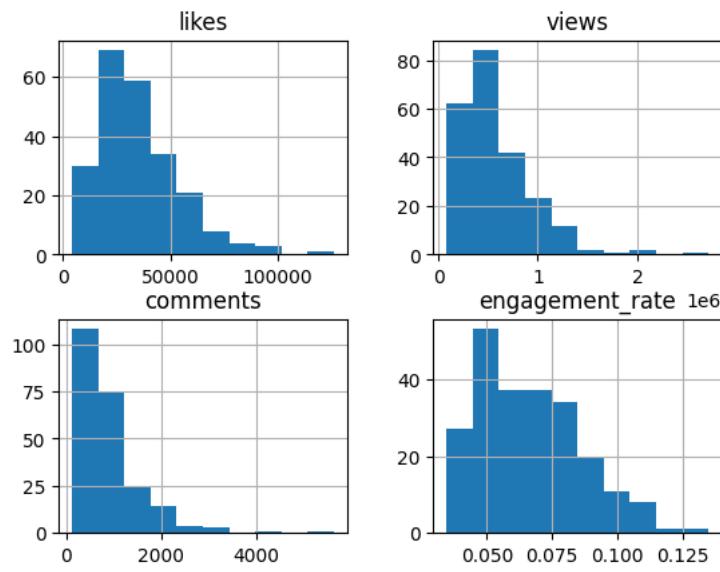


Figura 137 Histograma Orslok

En los histogramas, se observa una distribución ligeramente sesgada hacia la izquierda para métricas como likes, vistas y comentarios, lo que indica que la mayoría de los videos de Orslok tienden a tener valores más bajos en estas métricas, con solo unos pocos alcanzando cifras excepcionalmente

altas. Sin embargo, en cuanto a la tasa de engagement, la distribución parece ser más simétrica y centrada, lo que sugiere que, independientemente de la popularidad del video en términos de vistas o likes, la proporción de espectadores que interactúan activamente con el contenido es consistentemente estable.

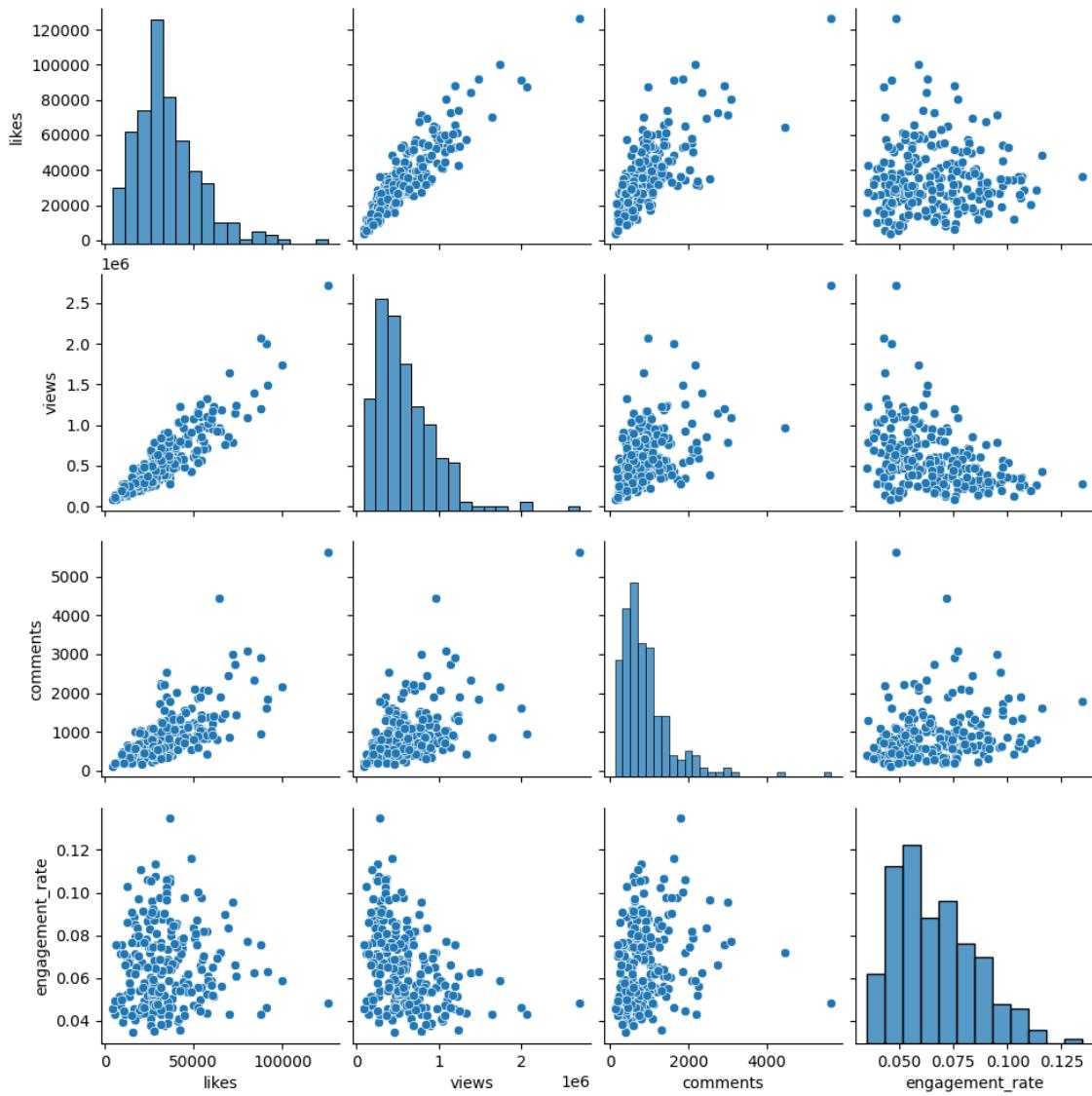


Figura 138 Sactterplot Orslok

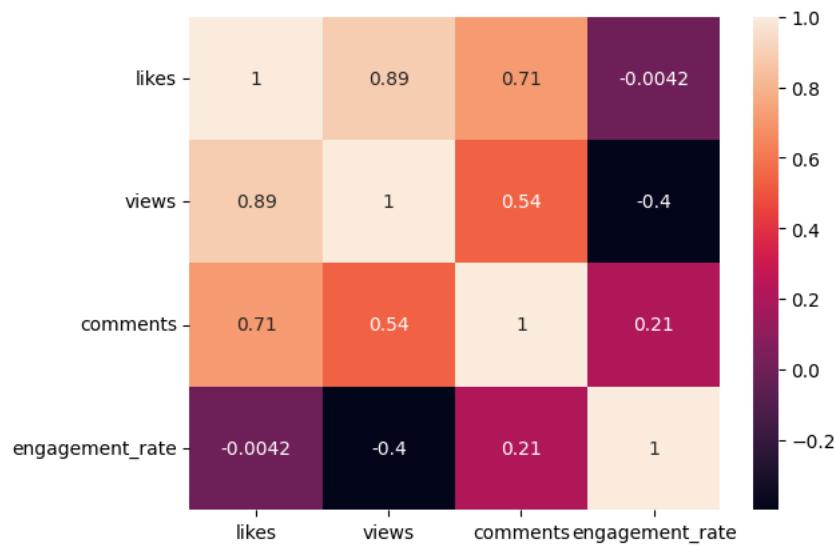


Figura 139 Matriz correlación Orslok

Según los graficos de scatterplot y matriz de correlación, se observa que los videos que acumulan más likes tienden a tener también más vistas y comentarios, evidenciando una fuerte relación entre estas métricas. Sin embargo, a medida que las vistas aumentan, la tasa de engagement tiende a disminuir, lo que sugiere que no todos los espectadores interactúan activamente con el contenido. A pesar de la alta correlación entre likes y vistas, sorprendentemente, los likes no parecen influir significativamente en la tasa de engagement.

Serie Temporal

- Análisis Mensual: Se observa una tendencia con altibajos en la publicación de videos. Aunque hay fluctuaciones, un pico notable se manifiesta entre abril y agosto de 2020. Curiosamente, no se registran datos después de 2022, lo que podría indicar una pausa o cese en sus publicaciones.

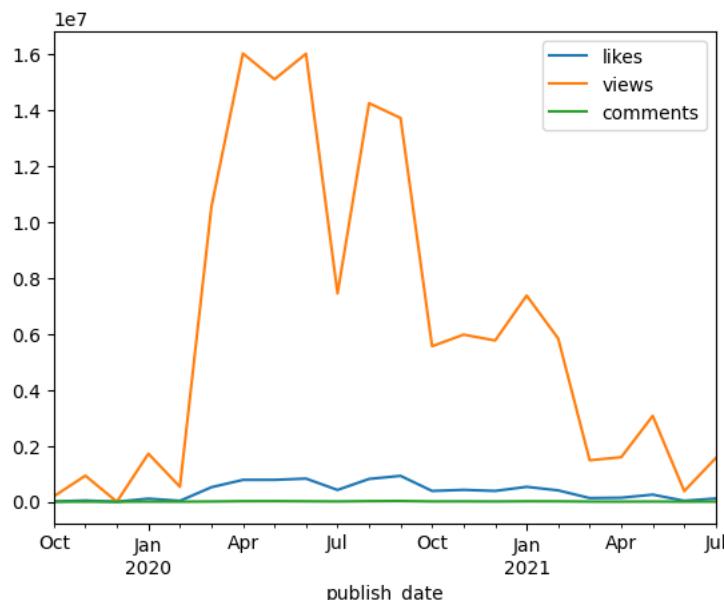


Figura 140 Serie Temporal mes Orslok

- Análisis Semanal: Aunque publica videos todos los días, los jueves parecen ser los más populares en términos de vistas, lo que sugiere que su audiencia podría estar más activa o receptiva ese día.

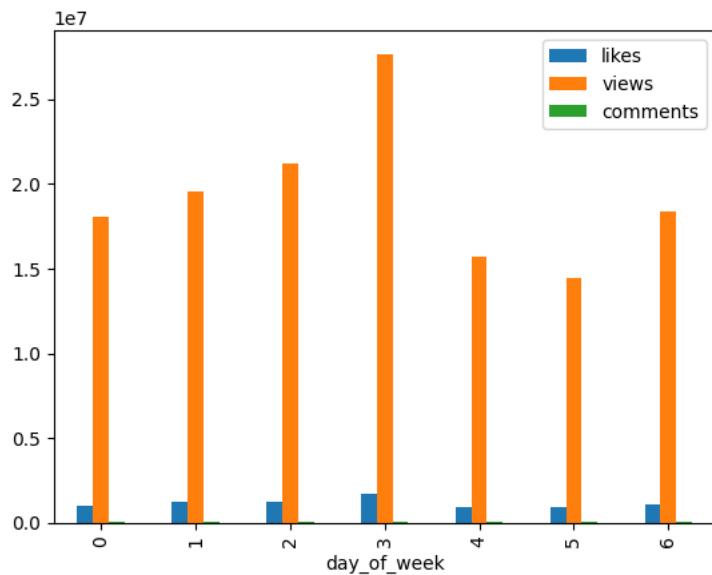


Figura 141 Serie Temporal semana Orslok

- Análisis Horario: Prefiere el intervalo de tiempo entre las 19 y 21 horas para publicar sus videos, lo que podría coincidir con momentos en que su audiencia está más disponible o dispuesta a consumir contenido.

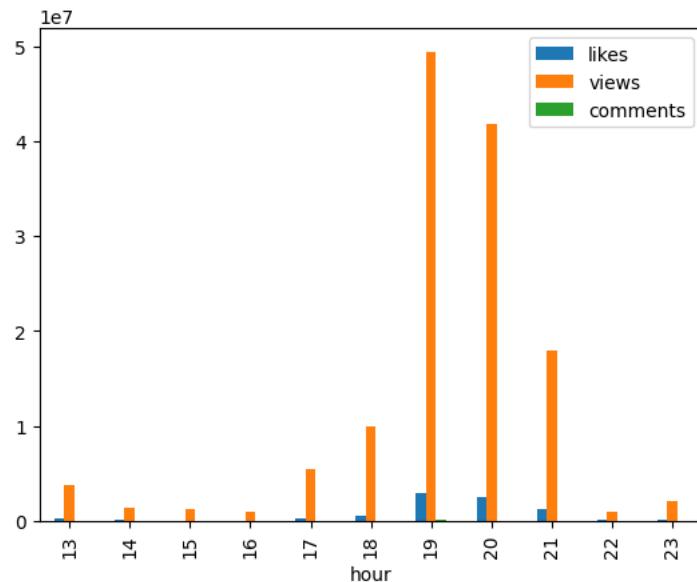


Figura 142 Serie Temporal hora Orslok

El video titulado "reacción añanin" es el más popular en términos de vistas para Orslok. Esto sugiere que el contenido de reacción, especialmente hacia "añanin", ha resonado particularmente bien con su audiencia. Es posible que el tema o el contenido específico de "añanin" haya sido de gran interés o relevancia en ese momento, o que la forma en que Orslok abordó el tema fue particularmente atractiva para sus seguidores. Sería interesante analizar más a fondo el contenido de este video y ver qué lo distingue de otros videos en su canal.

Minería de Textos



Figura 143 Wordcloud título Orsløk

Basándonos en la nube de palabras del título de los videos de Orslok, podemos identificar las siguientes tendencias y temas populares:

- Orslok: Es el término más frecuente, lo que indica que Orslok a menudo incluye su propio nombre en los títulos de sus videos, probablemente para reforzar su marca personal.
 - Reacción/Reaccionar: Estos términos sugieren que una parte significativa de su contenido está centrada en videos de reacción, donde Orslok responde o reacciona a diferentes temas, eventos o contenidos.
 - Viewers: Esto podría indicar que Orslok involucra a sus espectadores en su contenido, posiblemente reaccionando a sus comentarios, sugerencias o participando en actividades interactivas con ellos.
 - Valorant y Egoland: Estos términos sugieren que Orslok podría estar creando contenido relacionado con el juego "Valorant" y algo relacionado con "Egoland", que podría ser un evento, juego o tema específico.
 - Piba, Pib, Pana: Estas palabras podrían ser jerga o términos coloquiales que Orslok utiliza en sus títulos para atraer a su audiencia o referirse a temas específicos.
 - Streamer, Stream: Estos términos indican que Orslok podría estar haciendo streams en vivo o refiriéndose a otros streamers en su contenido.

Los títulos de Orslok parecen centrarse en reacciones, juegos, interacción con los espectadores y temas actuales o populares.



Figura 144 Wordcloud descripción Orslok

A partir de la nube de palabras de las descripciones de los videos de Orslok, se pueden identificar las siguientes tendencias y características:

- Redes Sociales: Las palabras Twitter, IG (Instagram) y Twitch son las más frecuentes, lo que indica que Orslok promociona activamente sus otras plataformas sociales en las descripciones de sus videos. Esto sugiere que tiene una presencia activa en estas plataformas y busca dirigir a sus espectadores hacia ellas.
 - Editor: La frecuencia de la palabra "editor" sugiere que Orslok a menudo acredita o menciona a la persona o equipo que edita sus videos, lo que indica profesionalismo y reconocimiento hacia quienes colaboran en su contenido.
 - Reaccionar y Jugar: Estas palabras indican que una parte significativa de su contenido está centrada en videos de reacción y en gameplays o sesiones de juego.
 - Colaboraciones o Referencias: Las palabras Maximus, Alexby, Rubius y Cheeto podrían ser otros creadores de contenido o personalidades con las que Orslok ha colaborado o a las que hace referencia en sus videos.
 - Reddit y OrslokReddit: Estos términos sugieren que Orslok podría tener una comunidad activa en Reddit o que a menudo reacciona o discute temas que provienen de esta plataforma.
 - OrslokClips: Podría ser un canal secundario o una serie de videos donde Orslok comparte clips cortos o momentos destacados.

Las descripciones de Orslok reflejan una fuerte presencia en las redes sociales, colaboraciones con otros creadores, y una variedad de contenidos que van desde reacciones hasta gameplays.



Figura 145 Wordcloud transcripción Orslok

En las transcripciones de los videos de Orslok, se destacan las siguientes palabras y tendencias:

- Música: La alta frecuencia de esta palabra indica que en los videos de Orslok hay una presencia constante de música de fondo. Esto sugiere que la música es un elemento esencial en la producción y ambientación de su contenido.
 - xxxx: Al ser un marcador para insultos, indica que Orslok utiliza un lenguaje coloquial y, en ocasiones, puede expresarse de manera fuerte o directa en sus videos.
 - Ver, Querer, Decir: Estas palabras reflejan acciones y deseos, lo que sugiere que Orslok a menudo comparte sus opiniones, deseos y observaciones con su audiencia.
 - Tío: Esta palabra es una expresión coloquial en español que se utiliza a menudo como un término amistoso o para referirse a alguien de manera informal. Su frecuencia sugiere que Orslok tiene un estilo de comunicación cercano y amigable.
 - Gente, Cosa, Gustar, Pasar, Vídeo, Vida, Año: Estas palabras reflejan temas comunes en los videos de Orslok, como discutir sobre la gente, compartir opiniones sobre diferentes temas, comentar sobre videos y reflexionar sobre la vida y el paso del tiempo.
 - Valer: Esta palabra, que puede ser una expresión de acuerdo o énfasis, sugiere que Orslok a menudo utiliza esta expresión para reforzar sus puntos o compartir su acuerdo con algo.

En resumen, las transcripciones muestran que Orslok tiene un estilo de comunicación coloquial y directo, y que a menudo discute y reacciona a una variedad de temas, utilizando música de fondo como elemento esencial en sus videos. Su lenguaje puede ser fuerte en ocasiones, reflejando una personalidad auténtica y sin filtros.

Common Phrases:

mamá mamá mamá: 72
bla bla bla: 63
música música música: 25
pacer pacer pacer: 24
música aplauso música: 20
ok ok ok: 17
xxxx xxxx xxxx: 16
alto alto alto: 14
ja ja ja: 13
porro porro porro: 12
sexo sexo sexo: 12
truf truf truf: 10
aplauso música aplauso: 10
hola hola hola: 9
na na na: 9
música i música: 8
valer valer valer: 8
ay ay ay: 8
cuidado cuidado cuidado: 7
ver ver ver: 7

Figura 146 Frases comunes Orsok

Las frases comunes en los videos de Orsok nos dan una idea de su estilo y los temas recurrentes en su contenido:

- Repetición de palabras: La repetición de palabras como "mamá mamá mamá", "bla bla bla", "ok ok ok", entre otras, sugiere que Orsok utiliza la repetición como una técnica estilística para enfatizar o hacer humor.
- Música: Las frases "música música música" y "música aplauso música" indican que la música es un elemento importante en sus videos, ya sea como fondo o como tema de discusión.
- xxxx: La repetición de "xxxx" confirma que Orsok utiliza un lenguaje fuerte o directo en sus videos, como ya se mencionó anteriormente.
- Emociones y Reacciones: Frases como "ja ja ja" y "ay ay ay" reflejan las emociones y reacciones de Orsok durante sus videos, mostrando un lado más humano y relatable.
- Temas específicos: Palabras como "porro", "sexo" y "truf" sugieren que Orsok aborda temas variados, algunos de los cuales pueden ser más controversiales o adultos.
- Interjecciones y expresiones: Frases como "valer valer valer", "alto alto alto" y "cuidado cuidado cuidado" indican que Orsok utiliza interjecciones y expresiones comunes para comunicar sus pensamientos y sentimientos.

```
Topics:
Topic: 0, Words: 0.016*"xxxx" + 0.008*"decir" + 0.007*"cosa" + 0.007*"gente" + 0.006*"año" + 0.006*"música" + 0.006*"ver" + 0.006*"tío" + 0.005*"gustar" + 0.005*"video"
Topic: 1, Words: 0.019*"música" + 0.015*"xxxx" + 0.009*"ver" + 0.008*"querer" + 0.007*"bla" + 0.007*"casa" + 0.006*"decir" + 0.005*"aplauso"
Topic: 2, Words: 0.029*"música" + 0.021*"xxxx" + 0.011*"querer" + 0.018*"gente" + 0.007*"tío" + 0.006*"valer" + 0.006*"decir" + 0.005*"gustar" + 0.005*"vida"
Topic: 3, Words: 0.015*"música" + 0.014*"xxxx" + 0.009*"mamá" + 0.008*"ver" + 0.007*"querer" + 0.005*"tío" + 0.005*"año" + 0.005*"video" + 0.004*"casa" + 0.004*"valer"
Topic: 4, Words: 0.011*"xxxx" + 0.009*"música" + 0.008*"tío" + 0.008*"querer" + 0.008*"pasar" + 0.008*"ver" + 0.005*"decir" + 0.005*"cosa" + 0.005*"tener" + 0.004*"video"
```

Figura 147 Topic Modeling

El modelado de temas (Topic Modeling) nos proporciona una visión general de los temas recurrentes en el contenido de Orsok:

- Tópico 0: Se centra en las interacciones y opiniones. Las palabras clave son "xxxx" (insulto), "decir", "cosa", "gente", "año", "música", "ver", "tío" y "vídeo". Este tema podría estar relacionado con las opiniones y reacciones de Orsok sobre eventos o tendencias actuales.
- Tópico 1: Está dominado por la "música", seguido de "xxxx", "ver", "querer" y "favor". También destaca la palabra "bla", que podría indicar momentos de charla o discusión. Este tema podría estar relacionado con reacciones a la música o discusiones sobre temas musicales.
- Tópico 2: Nuevamente, la "música" y "xxxx" son dominantes. Las palabras "gente", "tío", "valer" y "vida" sugieren que este tema podría estar relacionado con experiencias personales o anécdotas relacionadas con la música.

- Tópico 3: Este tema tiene un enfoque en la familia o relaciones cercanas con palabras como "mamá" y "tío". También se mencionan "música", "xxxx", "ver" y "año", lo que sugiere que podría tratarse de recuerdos o historias familiares con un trasfondo musical.
- Tópico 4: Se centra en experiencias y reflexiones. Las palabras clave son "xxxx", "música", "tío", "querer", "pasar", "ver", "decir" y "cosa". Este tema podría estar relacionado con las reflexiones y pensamientos de Orslok sobre diversos temas, desde música hasta experiencias personales.

En resumen, el contenido de Orslok parece girar en torno a la música, las opiniones y reacciones, las experiencias personales y las relaciones familiares o cercanas. La palabra "xxxx" aparece en todos los temas, lo que indica que Orslok utiliza un lenguaje fuerte o directo en sus videos.

Unique Terms:
 querer: 5.490107482055668
 decir: 4.5573822269256485
 tío: 4.510018706649683
 gente: 4.271171650855942
 risa: 3.836827403112801
 valer: 3.690208382491735
 cosa: 3.611962887168945
 gustar: 3.4727044494499992
 favor: 3.4681417729351876
 pasar: 3.4118088991494324

Figura 148 Términos únicos Orslok

Las palabras únicas más destacadas en el contenido de Orslok son emocionales o relacionadas con opiniones y reacciones. Palabras como "querer", "decir", "tío" y "gente" sugieren discusiones sobre deseos, expresiones, relaciones y observaciones sobre la sociedad o grupos. "Risa" indica momentos humorísticos o divertidos, mientras que "valer", "cosa", "gustar", "favor" y "pasar" reflejan valoraciones, preferencias y eventos o situaciones discutidas en el contenido.

Network

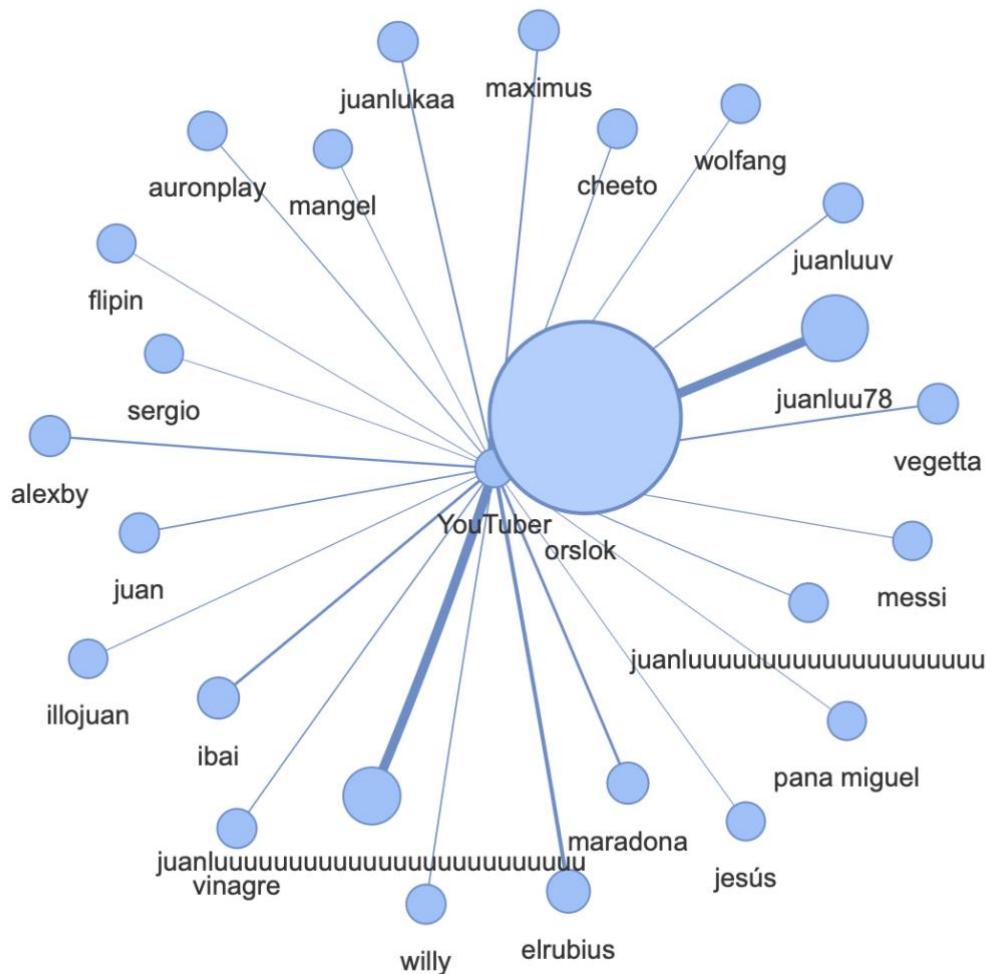


Figura 149 Network Orslok

En cuanto a la red de Orslok, se puede observar que tiene una fuerte conexión con otros YouTubers. La mayor cantidad de menciones es hacia él mismo, "orslok", con 953 menciones. Le sigue "juanluu78" con 176 menciones y una variante del mismo nombre "juanluuuuuuuuuuuuuuuuuuuuuuuuu" con 121 menciones. Otros YouTubers populares como "elrubius", "ibai", y "alexby" también son mencionados, aunque en menor medida. Es interesante notar las múltiples variantes del nombre "juanluu", lo que podría indicar bromas recurrentes o temas específicos relacionados con ese nombre. También hay menciones a figuras no relacionadas directamente con YouTube, como "maradona" y "messi", lo que sugiere una variedad de temas discutidos en su contenido.

Vegetta

EDA

	likes	views	comments	year	month	\
count	466.000000	4.660000e+02	466.000000	466.000000	466.000000	
mean	117841.622318	3.055316e+06	2840.892704	2020.675966	6.500000	
std	125097.474359	6.087836e+06	5618.770949	2.327060	3.531273	
min	4026.000000	6.000400e+04	117.000000	2012.000000	1.000000	
25%	33706.000000	4.218090e+05	291.500000	2020.250000	3.250000	
50%	63945.500000	8.132865e+05	549.500000	2022.000000	7.000000	
75%	165312.000000	3.140720e+06	2288.750000	2022.000000	10.000000	
max	738918.000000	7.251540e+07	50972.000000	2023.000000	12.000000	
	day_of_week	hour	engagement_rate			
count	466.000000	466.000000	466.000000			
mean	2.912017	18.504292	0.074724			
std	1.935182	3.236383	0.032066			
min	0.000000	0.000000	0.007702			
25%	1.000000	18.000000	0.053748			
50%	3.000000	19.000000	0.075961			
75%	4.000000	20.000000	0.092314			
max	6.000000	23.000000	0.204725			

Figura 150 Estadística Vegetta

- Volumen de Datos: Se han analizado un total de 466 videos de Vegetta.
- Interacción:
 - o Likes: En promedio, los videos de Vegetta reciben aproximadamente 117,842 likes. El video más popular ha alcanzado 738,918 likes, mientras que el menos popular ha obtenido 4,026 likes.
 - o Vistas: Los videos tienen un promedio de 3,055,316 vistas, con un video alcanzando un máximo de 72,515,400 vistas y un mínimo de 60,004 vistas.
 - o Comentarios: En promedio, los videos reciben 2,840 comentarios, con un video que ha alcanzado un máximo de 50,972 comentarios y un mínimo de 117 comentarios.
- Tiempo:
 - o Año: Los videos se han publicado desde 2012 hasta 2023.
 - o Mes: El mes promedio de publicación es alrededor de junio.
 - o Día de la semana: La mayoría de los videos se publican a mediados de semana, con un promedio de 2.91 (donde 0 es lunes y 6 es domingo).
 - o Hora: La mayoría de los videos se publican alrededor de las 6:30 p.m.
- Tasa de Engagement: La tasa de engagement promedio es del 7.47%. Esto indica que, en promedio, el 7.47% de los espectadores interactúan con el video a través de likes o comentarios. La tasa de engagement más alta observada es del 20.47%, mientras que la más baja es del 0.77%.
- Distribución de Datos: Las desviaciones estándar para likes, vistas y comentarios son relativamente altas, lo que indica una variabilidad significativa en la interacción de los videos. Esto sugiere que, aunque Vegetta tiene un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto.

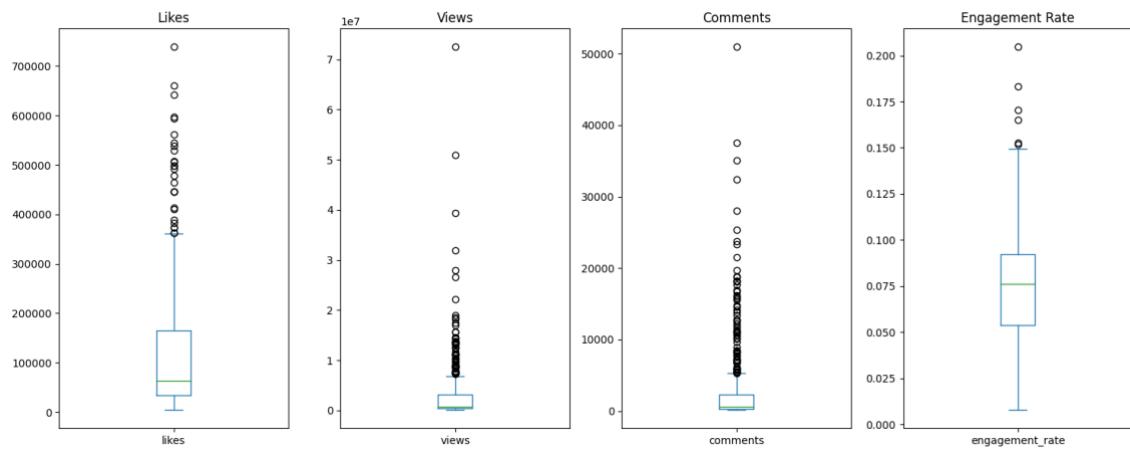


Figura 151 Boxplot Vegetta

Basándonos en los boxplots, se observa que hay una notable cantidad de valores atípicos en el extremo superior para likes, vistas y comentarios, lo que sugiere que algunos videos de Vegetta han tenido un rendimiento excepcionalmente alto en comparación con la mayoría. En cuanto a la tasa de engagement, la distribución parece más uniforme y centrada, sin muchos outliers, lo que indica una consistencia en la proporción de interacciones en relación con las vistas. La forma equilibrada de la caja en la tasa de engagement sugiere que, independientemente de la cantidad de vistas, la interacción de la audiencia con los videos tiende a ser constante.

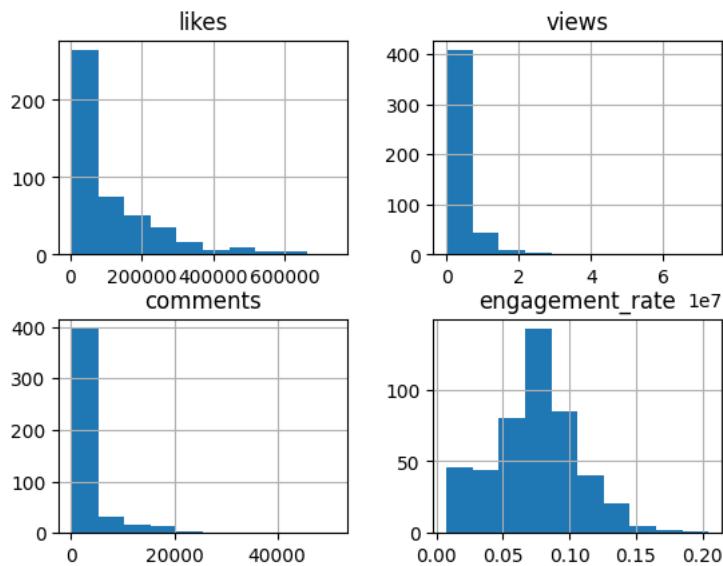


Figura 152 Histograma Vegetta

A partir de los histogramas, se observa que los likes, vistas y comentarios presentan una distribución sesgada hacia la izquierda, lo que indica que la mayoría de los videos de Vegetta tienen un número relativamente bajo de likes, vistas y comentarios, pero hay algunos videos que se destacan con cifras significativamente más altas. Por otro lado, la tasa de engagement muestra una distribución normal, lo que sugiere que la mayoría de los videos tienen una tasa de interacción promedio, con muy pocos videos que se desvían extremadamente hacia el alto o bajo engagement. Esto refleja una consistencia en cómo la audiencia interactúa con el contenido de Vegetta, independientemente de la cantidad de vistas que tenga el video.

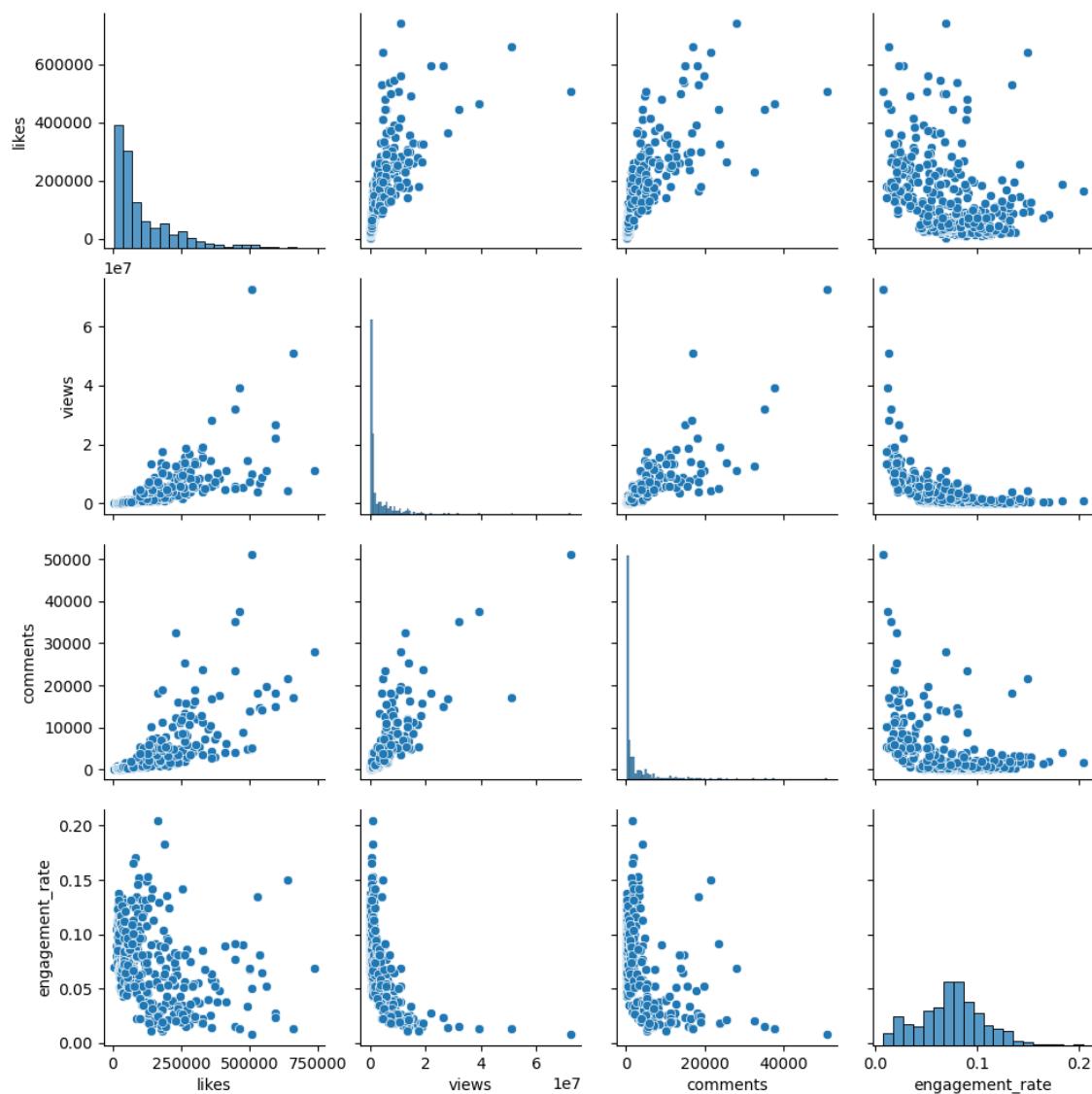


Figura 153 Scatterplot Vegetta

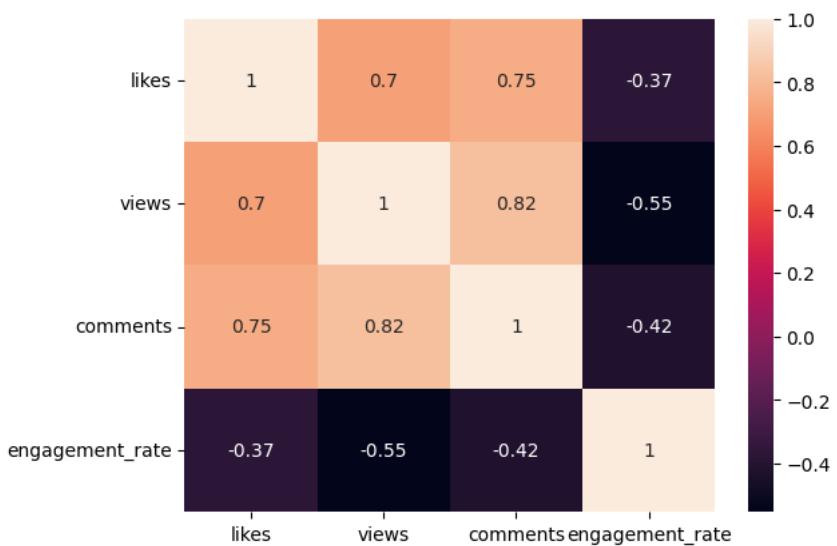


Figura 154 Matriz correlación Vegetta

En cuanto a la correlación, los likes están fuertemente correlacionados con las vistas y los comentarios, lo que indica que cuando un video recibe más likes, también tiende a tener más vistas y comentarios. Sin embargo, la tasa de engagement tiene una correlación negativa con los likes, vistas y comentarios. Esto sugiere que, aunque un video pueda tener muchas vistas, likes y comentarios, no necesariamente significa que tenga una alta tasa de interacción en proporción a las vistas. Es decir, no todos los espectadores que ven un video interactúan con él a través de likes o comentarios.

Serie Temporal

- Análisis Mensual: se observan fluctuaciones notables en la cantidad de vistas. Destacan picos significativos en 2018 y, posteriormente, en 2015, lo que indica que ha tenido momentos de gran popularidad en esos años. Es notable que Vegetta ha sido consistente en su producción de contenido, publicando videos desde hace muchos años hasta la fecha actual.

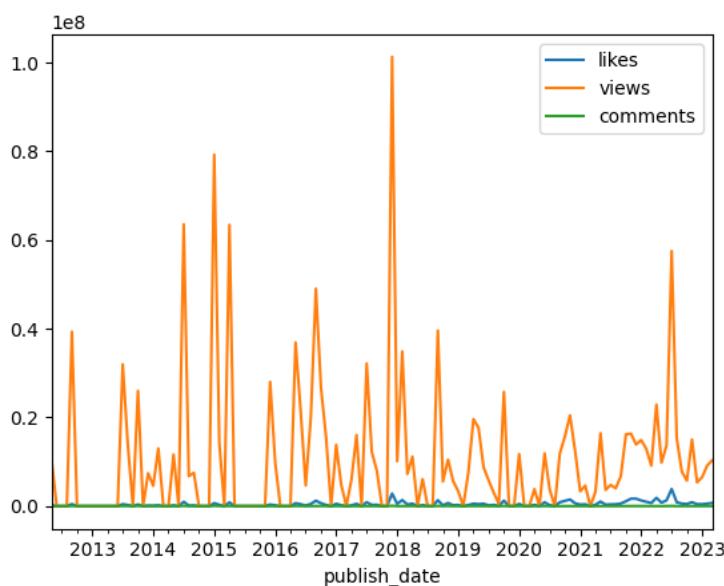


Figura 155 Serie Temporal mes Vegetta

- Análisis Semanal: Vegetta mantiene una frecuencia constante de publicaciones, distribuyendo sus videos a lo largo de todos los días de la semana. Aunque hay ligeras variaciones en las vistas según el día, no se identifica un patrón específico que indique un día preferido por su audiencia.

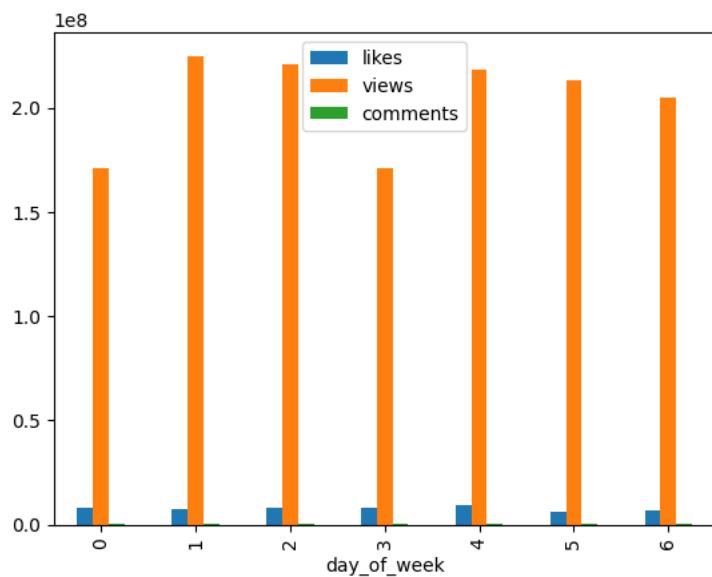


Figura 156 Serie Temporal semana Vegetta

- Análisis Horario: Vegetta muestra una clara preferencia por las tardes y primeras horas de la noche, especialmente entre las 17:00 y las 22:00 horas. Las 21:00 horas parecen ser su momento predilecto para lanzar nuevos contenidos.

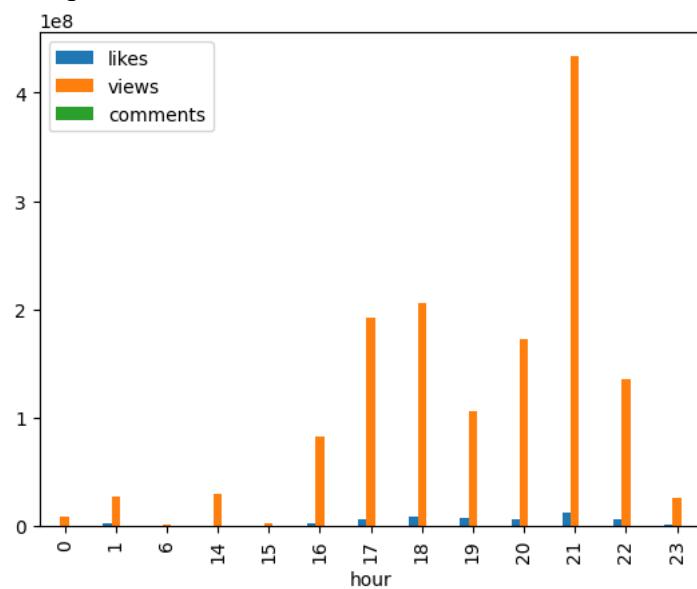


Figura 157 Serie Temporal hora Vegetta

El video más popular de Vegetta es "PLANETA VEGETTA: UNA GRAN AVENTURA NOS ESPERA". Este título sugiere que el video podría ser el comienzo o una parte significativa de una serie o aventura dentro del canal. La popularidad de este video podría deberse a varios factores, como un contenido emocionante, una promoción efectiva o simplemente el interés generado por el título. Sería interesante analizar el contenido del video y las reacciones de los espectadores para comprender completamente por qué este video en particular ha resonado tanto con la audiencia.

Minería de Textos



Figura 158 Wordclooud título Vegetta

En la nube de palabras relacionada con los títulos de los videos de Vegetta, se destaca que "karmaland" es la palabra más frecuente, lo que indica que es un tema o serie recurrente en su canal. Otros términos como "pubg", "victoria" y "resumen" sugieren que Vegetta juega y comenta sobre videojuegos, posiblemente obteniendo victorias en juegos como PUBG. La presencia de palabras como "aventura", "pirata" y "apocalipsisminecraft" sugiere que también se sumerge en mundos temáticos y aventuras específicas en juegos como Minecraft. Además, las palabras "vegetta", "arkadio" y "willy" podrían estar relacionadas con su identidad en línea, colaboradores o personajes recurrentes en sus videos.



Figura 159 Wordcloud descripción Vegetta

En la descripción de los videos de Vegetta, la palabra "vegetta" es la más frecuente, lo que indica que a menudo se refiere a sí mismo o a su marca en sus descripciones. La palabra "canal" es también muy común, lo que sugiere que a menudo habla sobre su canal o hace referencias a él. Las palabras "esperar", "amigo" y "abrazo" sugieren un tono amigable y cercano con su audiencia. Además, hay una fuerte presencia de términos relacionados con las redes sociales como "facebook", "twitch", "twitter" e "instagram", lo que indica que promociona activamente sus otras plataformas sociales en las descripciones de sus videos. La mención de "samuel" podría referirse a una persona específica, tal vez un colaborador o alguien relevante para su contenido. En general, Vegetta parece tener un enfoque muy interactivo y promocional en las descripciones de sus videos, buscando conectar con su audiencia y dirigirlos a sus otras plataformas.



Figura 160 Wordcloud transcripción Vegeta

En las transcripciones de los videos de Vegetta, la palabra "valer" es la más frecuente, lo que indica que es una expresión que utiliza con regularidad. Las palabras "ver", "decir" y "cosa" son también comunes, lo que sugiere que a menudo comenta sobre lo que está viendo o haciendo en sus videos. "Chico" y "venir" sugieren que a menudo se dirige o hace referencia a otras personas en sus videos. La presencia de "música" indica que a menudo usa música en sus videos. Las palabras "querer", "dejar", "pasar" y "casa" sugieren que a menudo habla sobre sus deseos, acciones y posiblemente sobre su vida personal. "Gustar" y "mirar" reflejan su interacción con el contenido que presenta. La palabra "eh" es una interjección que podría usar para llamar la atención o enfatizar algo. En general, estas palabras reflejan el estilo comunicativo y los temas recurrentes en los videos de Vegetta.

Common Phrases:

- valer valer valer: 1303
- mira mira mira: 738
- pa pa pa: 453
- ay ay ay: 444
- ta ta ta: 277
- ver ver ver: 241
- bla bla bla: 237
- chica niño niñas: 198
- uy uy uy: 195
- niño niñas mujer: 194
- niñas mujer hombr: 194
- ah valer valer: 179
- chico chica niño: 178
- vale valer valer: 174
- toma toma toma: 162
- mujer hombr viceversar: 159
- correr correr correr: 143
- god of war: 142
- olé olé olé: 138
- venir venir venir: 134

Figura 161 Frases comunes Vegetta

Las frases comunes en las transcripciones de Vegetta muestran una tendencia hacia la repetición de palabras para énfasis. "Valer valer valer" es la frase más repetida, lo que indica que es una expresión que utiliza con mucha frecuencia para enfatizar o reaccionar a algo. Frases como "mira mira mira", "pa pa pa", "ay ay ay" y "ta ta ta" sugieren que a menudo repite palabras para expresar sorpresa, emoción o énfasis. La presencia de frases como "chica niño niñas", "niño niñas mujer" y "niñas mujer hombr" sugiere que a menudo se refiere a diferentes grupos de personas o géneros en sus videos. La frase "god of war" indica que ha hablado sobre este popular videojuego en varias ocasiones. En general, estas frases comunes reflejan el estilo comunicativo enérgico y enfático de Vegetta, así como los temas y reacciones recurrentes en sus videos.

```
Topics:
Topic: 0, Words: 0.028*"valer" + 0.011*"ver" + 0.010*"cosa" + 0.010*"decir" + 0.007*"chico" + 0.007*"querer" + 0.007*"venir" + 0.006*"dejar" + 0.006*"música"
Topic: 1, Words: 0.020*"valer" + 0.014*"ver" + 0.013*"eh" + 0.010*"chico" + 0.007*"tío" + 0.007*"casa" + 0.006*"decir" + 0.006*"gente" + 0.005*"mirar"
Topic: 2, Words: 0.022*"valer" + 0.012*"ver" + 0.009*"decir" + 0.008*"cosa" + 0.007*"juego" + 0.006*"venir" + 0.006*"música" + 0.006*"gustar" + 0.006*"querer" + 0.006*"chico"
Topic: 3, Words: 0.014*"valer" + 0.009*"ver" + 0.009*"música" + 0.008*"coche" + 0.007*"gustar" + 0.007*"juego" + 0.007*"venir" + 0.007*"chico" + 0.007*"decir" + 0.007*"pasar"
Topic: 4, Words: 0.029*"valer" + 0.010*"ver" + 0.009*"dejar" + 0.009*"decir" + 0.009*"cosa" + 0.007*"tener" + 0.007*"querer" + 0.006*"quedar" + 0.006*"poder" + 0.006*"mirar"
```

Figura 162 Topic Modeling Vegetta

Basándonos en los tópicos, podemos interpretar lo siguiente sobre el contenido de Vegetta:

- Tópico 0: Se centra en acciones y observaciones generales. La palabra "valer" es prominente, lo que podría ser una afirmación o una forma de dar énfasis. También se mencionan términos como "ver", "cosa", "casa" y "decir", lo que sugiere que Vegetta a menudo comenta sobre lo que está viendo o experimentando, posiblemente en su hogar o en un entorno familiar.
- Tópico 1: Este tópico parece centrarse en las reacciones y comentarios de Vegetta. La presencia de palabras como "eh", "tío" y "gente" sugiere que Vegetta a menudo se dirige o responde a su audiencia o a otros personajes de manera informal y amigable.
- Tópico 2: Aquí, el enfoque parece estar en el entretenimiento y el juego. Con palabras como "juego" y "música", es probable que Vegetta comente sobre los videojuegos que está jugando o sobre la música de fondo que está usando. También hay una mezcla de emociones y opiniones con palabras como "gustar" y "querer".
- Tópico 3: Este tópico tiene un enfoque en actividades y pasatiempos. La mención de "coche" y "juego" sugiere que Vegetta podría estar jugando videojuegos relacionados con carreras o comentando sobre vehículos. La palabra "pasar" también indica una narración de eventos o acciones.
- Tópico 4: Se centra en las opiniones y decisiones de Vegetta. Las palabras "dejar", "tener", "quedar" y "poder" sugieren que a menudo habla sobre sus elecciones, lo que decide hacer o no hacer, y sus habilidades o capacidades en ciertas situaciones.

En general, estos tópicos ofrecen una visión de los temas y estilos de comunicación predominantes en el contenido de Vegetta, mostrando una combinación de comentarios sobre juegos, interacciones personales y reflexiones generales.

```
Unique Terms:
eh: 37.181714482229125
juego: 23.325616918923792
coche: 21.985042393242356
jugar: 17.21432736738637
willy: 16.96065076094824
xxxx: 16.266694543671512
zona: 15.725133149038491
ángel: 14.71071548740683
muerto: 14.247759822945088
directo: 14.060646691610813
```

Figura 163 Términos únicos Vegetta

Los términos únicos más destacados en el contenido de Vegetta son "eh", que es una interjección comúnmente utilizada; "juego" y "jugar", que indican un enfoque en videojuegos; "coche", que podría referirse a juegos de carreras o comentarios sobre vehículos; "willy", que podría ser una mención a otro creador o personaje; "xxxx", que es un marcador de lugar o censura; "zona", "ángel", "muerto" y "directo", que sugieren temas variados, desde lugares específicos en juegos hasta transmisiones en vivo o "streamings".

Network

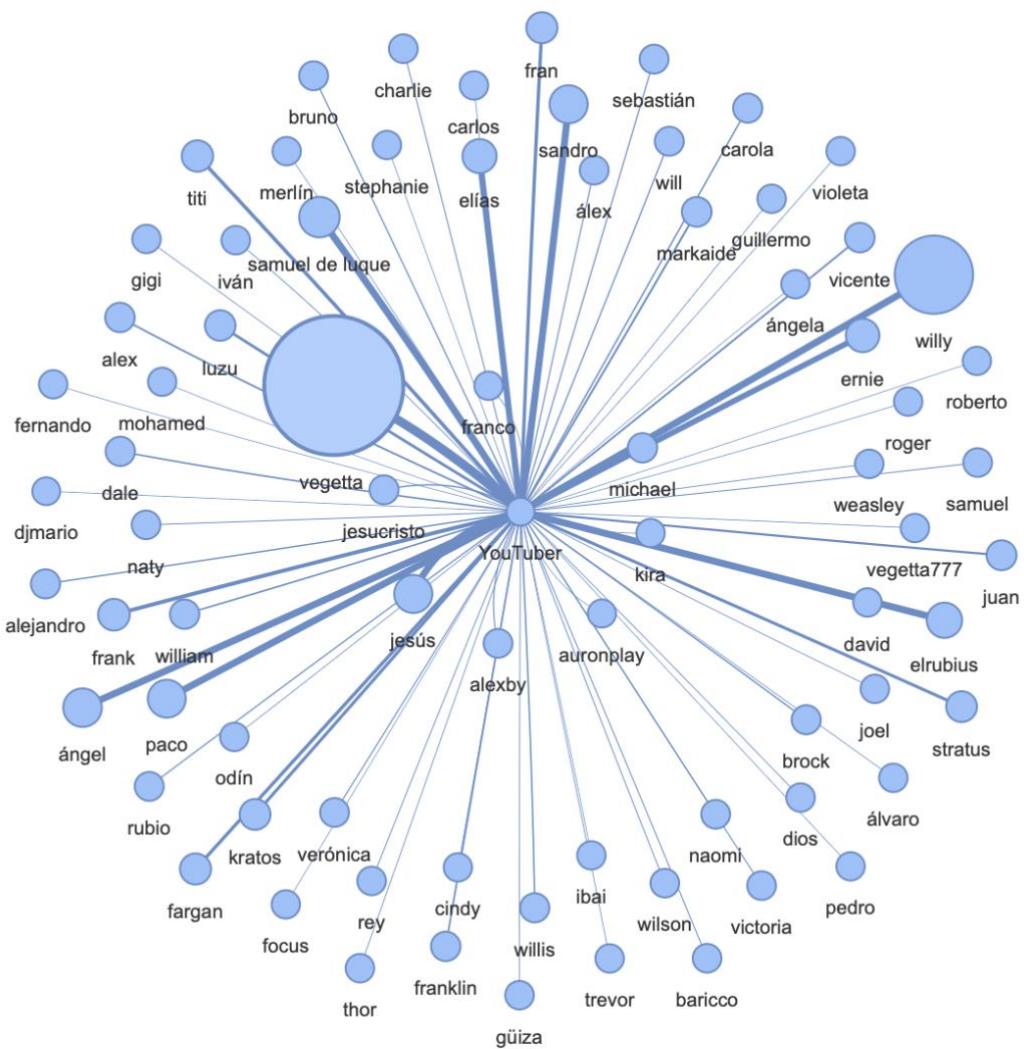


Figura 164 Network Vegetta

Basándonos en la red de Vegetta, se puede interpretar que "Vegetta" es el YouTuber central o principal, con la mayor cantidad de conexiones o menciones, sumando un total de 3961. Le sigue "Willy" con 1799 menciones, lo que sugiere que tiene una relación o colaboración cercana con Vegetta. Otros YouTubers como "Samuel de Luque", "Ángel", "Jesús", "Sandro" y "Paco" también tienen un número significativo de menciones, lo que indica que podrían ser colaboradores frecuentes o tener una relación estrecha con Vegetta.

Además, hay otros YouTubers conocidos como "elrubius" y "luzu" en la lista, lo que sugiere que Vegetta podría haber tenido interacciones o colaboraciones con ellos en algún momento.

En general, esta red muestra las conexiones y relaciones de Vegetta con otros YouTubers, ya sea a través de colaboraciones, menciones o cualquier otro tipo de interacción. Es evidente que Vegetta tiene una amplia red de conexiones en la comunidad de YouTube, lo que podría ser un indicador de su influencia y popularidad en la plataforma.

Conclusión de los resultados

Tras analizar los resultados de análisis de los influencers de YouTube mencionados, se pueden extraer las siguientes conclusiones:

- Características y Patrones de Lenguaje: Los influencers tienden a utilizar un lenguaje coloquial y cercano, lo que facilita la conexión con su audiencia, especialmente con el público joven. Palabras y frases repetidas, como "valer", "ver", "eh", entre otras, indican una tendencia a usar jerga o expresiones populares que resuenan con su audiencia. Además, la presencia de música de fondo y el uso de ciertos términos, como "música", sugieren que incorporan elementos multimedia para enriquecer su contenido y hacerlo más atractivo.
 - Interacción y Engagement: La tasa de interacción y los comentarios reflejan que estos influencers tienen una audiencia activa y comprometida. Las fluctuaciones en las vistas y likes, así como la presencia de outliers, sugieren que, aunque tienen un promedio generalmente alto de interacción, hay videos que son excepcionalmente populares y otros que no lo son tanto. Esto puede deberse a la naturaleza del contenido, la relevancia del tema o la calidad de la producción.
 - Red de Conexiones: Algunos influencers, como Vegetta, tienen una amplia red de conexiones con otros influencers o personas de figura, lo que indica colaboraciones y una posible estrategia de cross-promotion. Estas colaboraciones pueden aumentar la visibilidad y el alcance de su contenido.
 - Valores para la Audiencia: Los influencers parecen ofrecer entretenimiento, humor y, en algunos casos, educación o insights sobre temas específicos. Su capacidad para conectar con la audiencia y generar engagement sugiere que ofrecen un valor significativo, ya sea en forma de entretenimiento, información o una combinación de ambos.

Network de la lista predefinida de influencers

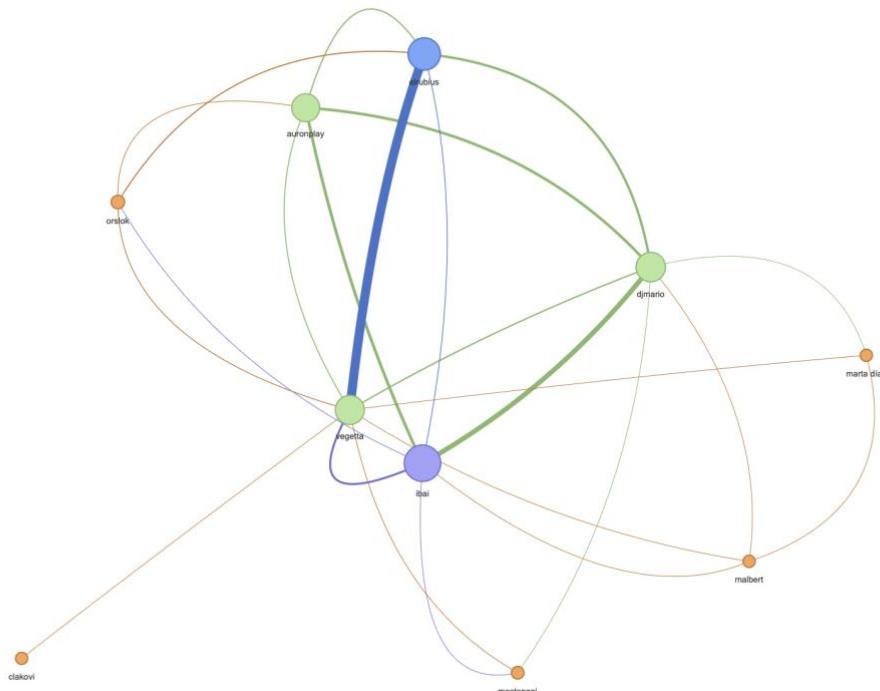


Figura 165 Network all influencers

En el análisis de la red de interacciones entre los influencers seleccionados, se observa una estructura compleja que revela patrones de colaboración y nexos interpersonales en el ámbito digital, particularmente en el mundo del gaming y entretenimiento.

Ibai, Auronplay, DjMario, ElRubius y Vegetta emergen como figuras centrales en esta comunidad, estableciendo conexiones robustas entre sí. Estas interacciones pueden ser indicativas de

colaboraciones en transmisiones en vivo, participaciones en torneos de videojuegos o apariciones en vlogs y otros formatos de entretenimiento. La densidad de sus conexiones sugiere que comparten una audiencia similar y que sus colaboraciones son estratégicas para mantener y ampliar su base de seguidores.

Por otro lado, influencers como Orslok, Mostopapi y Marta Díaz, aunque con conexiones menos intensas, representan a una generación emergente de creadores de contenido. Su presencia en la red indica que están en proceso de consolidación y que buscan establecer alianzas con influencers ya establecidos para ampliar su alcance. Específicamente, Marta Díaz, con un enfoque más diversificado que abarca temas de moda, belleza y estilo de vida, sugiere que la red no está limitada únicamente al ámbito del gaming, evidenciando un esfuerzo por diversificar contenido y llegar a audiencias más amplias.

Las conexiones entre estos influencers pueden ser el resultado de colaboraciones estratégicas, donde se busca combinar audiencias y aprovechar la popularidad de cada uno para generar contenido de mayor impacto, especialmente en eventos especiales, lanzamientos de productos o campañas promocionales.

En conclusión, la red ofrece una visión panorámica de las interacciones entre influencers, subrayando la naturaleza interconectada del mundo de los influencers y la importancia de las colaboraciones para ampliar el alcance y la influencia en la plataforma. Estas conexiones y colaboraciones reflejan una estrategia dinámica y adaptativa en un terreno digital en constante evolución.

Consideraciones éticas y sociales

En el desarrollo de este estudio, se han recolectado y analizado datos procedentes exclusivamente de la plataforma YouTube. Es crucial subrayar que la información utilizada está disponible al público en general, garantizando que no se ha accedido a datos personales o confidenciales en ningún momento del proceso. La motivación detrás de este proyecto es puramente académica y educativa, sin ninguna intención de obtener beneficios comerciales o de otra índole.

Además, en cuanto al análisis textual, se ha abordado con un enfoque riguroso y objetivo. Hemos sido meticulosos en evitar interpretaciones que puedan ser influenciadas por opiniones personales o prejuicios, asegurando que no se comprometa la reputación o imagen de los influencers mencionados. Es nuestra responsabilidad y compromiso garantizar que la investigación se realice con los más altos estándares éticos y profesionales.

Es importante también reconocer que vivimos en una era donde la información y la privacidad son esenciales. Por ello, hemos tomado todas las precauciones necesarias para asegurar que los datos se manejen con el debido respeto y cuidado, considerando siempre las implicaciones éticas y sociales que conlleva el análisis de datos en plataformas públicas.

Conclusión

Hoy en día, vivimos en una era de digitalización sin precedentes, donde las redes sociales nos conectan más que nunca. Ante la aparición de nuevas plataformas y el crecimiento exponencial de usuarios, surge la pregunta: ¿Qué valor nos aportan los contenidos generados por los influencers?

En este estudio, nos adentramos en el ecosistema de las redes sociales, identificando a sus dos protagonistas principales: la audiencia y el creador de contenido o influencer. Exploramos el panorama de las redes sociales tanto a nivel global como en España, analizando la interacción mediática de los adolescentes, quienes constituyen una gran proporción de los usuarios, y la relación parasocial que se establece entre ellos y los influencers.

Basándonos en la investigación de la Dra. Sue Aran-Ramspott, nos enfocamos en los influencers y aplicamos técnicas de Inteligencia Artificial, específicamente Procesamiento de Lenguaje Natural (NLP), para analizar las características y patrones lingüísticos de los influencers más populares entre los jóvenes en España. Gracias a la IA, procesamos grandes volúmenes de datos de manera eficiente, permitiendo que la tecnología realice el trabajo pesado y nos brinde información valiosa sobre la marca personal de estos influencers y el valor que transmiten. Además, mediante el reconocimiento de entidades nombradas (NER), generamos redes que muestran las interacciones entre estos influencers y otros actores.

Resulta fascinante observar que los contenidos más populares son de entretenimiento. Destacan temas como reacciones, colaboraciones, streaming de videojuegos y vlogs diarios. Las colaboraciones entre influencers, especialmente aquellos dentro del mismo nicho, fortalecen sus lazos. Es notable el uso de lenguaje coloquial, incluso malsónante, en algunos influencers como Auronplay. Sin embargo, elementos como el humor, la música adecuada, la cercanía con la audiencia y la construcción de una marca personal sólida son esenciales para el éxito.

En cuanto a los valores transmitidos, si bien es cierto que el uso repetido de insultos puede influir negativamente en la comunicación de los más jóvenes, también se destacan valores positivos y educativos impartidos por estos influencers. A pesar de todo, la relación entre el influencer y su audiencia se mantiene sólida, con el influencer refiriéndose a sus seguidores como su "familia" o fans.

Por lo último, es esencial comprender hasta qué punto el respaldo de técnicas avanzadas de NLP facilita el análisis de las relaciones entre los influencers, los valores que transmiten y su audiencia virtual, especialmente en investigaciones de comunicación de esta índole. La ventaja es significativa. Mediante el uso de estas técnicas, no es necesario acceder individualmente a cada plataforma para conocer a cada influencer, explorar sus canales y examinar los contenidos que publican. Estas herramientas nos asisten en la extracción y organización de información valiosa de extensas bases de datos, permitiéndonos interpretarla de manera eficiente. En contraste con los métodos tradicionales de investigación, que requieren días de trabajo humano para analizar una muestra con datos limitados, el NLP nos ofrece una economía considerable de tiempo y esfuerzo. Gracias a estructuras reutilizables, es posible procesar grandes volúmenes de datos, en nuestro caso miles de registros, los cuales son limitados únicamente por la fuente de datos disponible y la capacidad del procesador. Esta limitación, además, puede mitigarse alquilando servidores en la nube, lo que permite procesar datos en tiempo real, actualizar resultados constantemente y trabajar sin restricciones.

Coste Económico

En relación al aspecto financiero de este proyecto, se abordarán los gastos principales incurridos durante el desarrollo del análisis. Los costes más significativos se asociaron al salario del especialista encargado del estudio y al uso de un servicio de traducción.

Para comenzar, es esencial entender el panorama salarial de un Data Scientist en España. El sueldo medio anual para este perfil es de 40.500 €. Si consideramos una jornada laboral estándar de 40

horas semanales a lo largo de 52 semanas, esto nos da un total de 2080 horas al año. Esto se traduce en una tarifa horaria de 19,47€.

Desglosando las horas de trabajo:

- Recopilación de datos: 20 horas
- EDA (Análisis Exploratorio de Datos): 11 horas
- Preprocesamiento (Limpieza de datos): 40 horas
- Minería de Textos: 22 horas
- Creación de redes: 33 horas
- Interpretación de resultados: 25 horas
- Redacción de la memoria: 40 horas

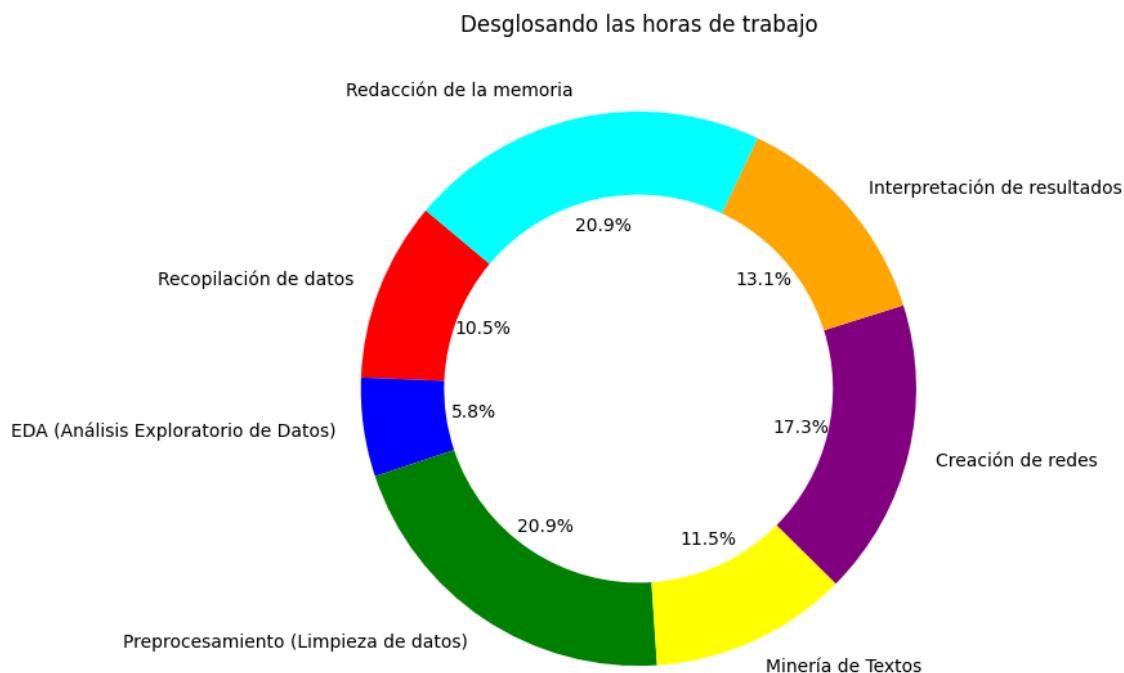


Figura 166 Horas de trabajo

Por otro lado, en las etapas iniciales del proyecto, durante el proceso de limpieza de datos, se contempló la idea de traducir todo el texto no español al español. Esta decisión resultó en un gasto de 55 euros (€20.00 por cada 1,000,000 de caracteres, más 5€ de suscripción). Aunque solo se tradujo el contenido de un hablante inglés, y a pesar de que el resto de los influencers son hispanohablantes que ocasionalmente usan términos en inglés, la traducción fue clara y comprensible. Sin embargo, al llegar al contenido de Emma Chamberlain, se optó por no continuar con la traducción.

En resumen, el coste total del proyecto ascendió a 3.773,77 €.

Limitaciones y Líneas de Futuro

Es esencial reconocer las limitaciones de este estudio. A pesar de que la implementación de técnicas de Inteligencia Artificial ha demostrado ser una herramienta valiosa y eficaz, existen ciertas deficiencias en el proceso de preprocesamiento, especialmente en la tokenización. Es probable que se obtengan resultados más precisos si se utiliza un modelo pre-entrenado de lenguaje más avanzado o específico para el ámbito de las redes sociales.

En la visualización de la nube de palabras, se detectaron duplicados, lo que podría deberse a variaciones en los formatos de texto. Una solución podría ser estandarizar todos los textos a un formato único, como UTF-8.

Las redes que muestran las interacciones entre influencers podrían enriquecerse incorporando datos demográficos extraídos de YouTube.

Este estudio se centró principalmente en los influencers. De cara al futuro, sería beneficioso recopilar y analizar datos de la audiencia, como los comentarios. Esto permitiría comprender las reacciones y valores que la audiencia percibe y valora. Además, se podría analizar la actividad de los usuarios en función del momento en que publican comentarios, proporcionando a los influencers información valiosa sobre cuándo es más propicio interactuar para maximizar el engagement. Adicionalmente, se podría profundizar en el análisis de los valores transmitidos por los influencers, utilizando un corpus etiquetado diseñado específicamente para predecir los valores que estos comunicadores generan en su audiencia.

Al final, como se ha mencionado en la conclusión, las técnicas avanzadas de NLP marcan un hito transformador en la investigación comunicativa, particularmente al estudiar la relación entre influencers y su audiencia en medios digitales. Estas herramientas no solo agilizan y perfeccionan la recolección y el análisis de datos, sino que también trascienden las barreras tradicionales relacionadas con el tiempo, el esfuerzo y el alcance. No obstante, es crucial reconocer que los resultados pueden variar con la incorporación de nuevos datos en el futuro. Por ello, se sugiere integrar y maximizar el uso de infraestructuras en la nube, sincronizando datos en tiempo real. Esta adaptación permitirá captar perspectivas más detalladas y contemporáneas, redefiniendo la manera en que abordamos y comprendemos el mundo digital y sus actores principales.

Bibliografía

- [1] L. Hudders, S. De Jans, y M. De Veirman, «The commercialization of social media stars: a literature review and conceptual framework on the strategic use of social media influencers», *Int J Advert*, vol. 40, n.º 3, 2021, doi: 10.1080/02650487.2020.1836925.
- [2] F. M. Solís y D. M. Critikián, «Social media and Generation Z», *Doxa Comunicacion*, vol. 2023, n.º 36, 2023, doi: 10.31921/doxacom.n36a1707.
- [3] «Estadísticas uso de redes sociales en 2023 (informe España y mundo)». <https://unavidaonline.com/estadisticas-redes-sociales/> (accedido 26 de agosto de 2023).
- [4] R. Conde y B. Casais, «Micro, macro and mega-influencers on instagram: The power of persuasion via the parasocial relationship», *J Bus Res*, vol. 158, 2023, doi: 10.1016/j.jbusres.2023.113708.
- [5] Gómez & Marín, «Impacto que generan las redes sociales en la conducta del adolescente y en sus relaciones interpersonales en Iberoamérica los últimos 10 años», *Tesis*, 2017.
- [6] S. Aran-Ramspott, Á. Moro Inchaurteta, y L. Botella García del Cid, «¿De qué me sirven los YouTubers e Instagramers? Preferencias de los adolescentes en España», *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes*, vol. 20, n.º 2, 2022, doi: 10.7195/ri14.v20i2.1875.
- [7] «(25) Inteligencia artificial y marketing: La inteligencia artificial sigue avanzando y se espera que tenga un impacto significativo en el marketing digital | LinkedIn». <https://www.linkedin.com/pulse/inteligencia-artificial-y-marketing-la-sigue-se-que-campos-de-laire/?originalSubdomain=es> (accedido 26 de agosto de 2023).
- [8] F. M. Solís y D. M. Critikián, «Social media and Generation Z», *Doxa Comunicacion*, vol. 2023, n.º 36, 2023, doi: 10.31921/doxacom.n36a1707.
- [9] «The 5 Characteristics of Social Media Websites That Go Viral | Entrepreneur». <https://www.entrepreneur.com/startng-a-business/the-5-characteristics-of-social-media-websites-that-go-viral/299903> (accedido 26 de agosto de 2023).
- [10] Instituto Nacional de Estadística, «Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación (TIC) en los Hogares», 2022. Accedido: 26 de agosto de 2023. [En línea]. Disponible en: https://www.ine.es/prensa/tich_2022.pdf
- [11] D. Gudiño, M. J. Fernández-Sánchez, M. T. Becerra-Traver, y S. Sánchez, «Social Media and the Pandemic: Consumption Habits of the Spanish Population before and during the COVID-19 Lockdown», *Sustainability (Switzerland)*, vol. 14, n.º 9, may 2022, doi: 10.3390/SU14095490.
- [12] W. Tafesse y B. P. Wood, «Followers' engagement with instagram influencers: The role of influencers' content and engagement strategy», *Journal of Retailing and Consumer Services*, vol. 58, 2021, doi: 10.1016/j.jretconser.2020.102303.
- [13] T. Hartmann y C. Goldhoorn, «Horton and Wohl revisited: Exploring viewers' experience of parasocial interaction», *Journal of Communication*, vol. 61, n.º 6, pp. 1104-1121, dic. 2011, doi: 10.1111/J.1460-2466.2011.01595.X.
- [14] B. Gómez Nieto, «El influencer: herramienta clave en el contexto digital de la publicidad engañosa», *methaodos revista de ciencias sociales*, vol. 6, n.º 1, 2018, doi: 10.17502/m.rcs.v6i1.212.
- [15] D. Shang y M. Ghriga, «Examining the impacts of key influencers on community development», *Computers, Materials and Continua*, vol. 60, n.º 3, 2019, doi: 10.32604/cmc.2019.08217.

- [16] J. Durau, S. Diehl, y R. Terlutter, «Motivate me to exercise with you: The effects of social media fitness influencers on users' intentions to engage in physical activity and the role of user gender», *Digit Health*, vol. 8, 2022, doi: 10.1177/20552076221102769.
- [17] «La tendencia influencer marketing | Generation Spain». <https://spain.generation.org/blog/la-tendencia-influencer-marketing/> (accedido 26 de agosto de 2023).
- [18] «What is Natural Language Processing? | IBM». <https://www.ibm.com/topics/natural-language-processing> (accedido 26 de agosto de 2023).
- [19] «Advanced Information Networking and Applications: Proceedings of the 36th ... - Google Books». https://books.google.es/books?dq=integration+of+nlp+in+youtube+content+analysis&hl=en&id=jiFnEAAQBAJ&lpg=PA194&ots=szK-qriuRK&pg=PA194&sa=X&sig=ACfU3U087eHIEemid0UCFlaT4sgNJlpfxA&source=bl&redir_esc=y#v=onepage&q=integration%20of%20nlp%20in%20youtube%20content%20analysis&f=false (accedido 26 de agosto de 2023).
- [20] D. Rao y B. McMahan, *Natural Language Processing with PyTorch*, vol. 53, n.º 9. 2019.
- [21] Y. Vasiliev, «Natural Language Processing with Python and spaCy: A Practical Introduction - Yuli Vasiliev - Google Libros», 2020, Accedido: 26 de agosto de 2023. [En línea]. Disponible en: https://books.google.com.pe/books?id=Au-_DwAAQBAJ&printsec=frontcover&dq=natural+language+processing&hl=es&sa=X&ved=0ahUKEwjCysfs-rHpAhVzH7kGHYyJBwiQ6AEliQEwCQ#v=onepage&q=natural+language+processing&f=false
- [22] «tf-idf basics of information retrieval | by Uniqtech | Data Science Bootcamp | Medium». <https://medium.com/data-science-bootcamp/tf-idf-basics-of-information-retrieval-48de122b2a4c> (accedido 26 de agosto de 2023).
- [23] «Understanding Word N-grams and N-gram Probability in Natural Language Processing | by Sunny Srinidhi | Towards Data Science». <https://towardsdatascience.com/understanding-word-n-grams-and-n-gram-probability-in-natural-language-processing-9d9eef0fa058> (accedido 26 de agosto de 2023).
- [24] «Stemming? Lemmatization? What?. Taking a high-level dive into what... | by Hunter Heidenreich | Towards Data Science». <https://towardsdatascience.com/stemming-lemmatization-what-ba782b7c0bd8> (accedido 26 de agosto de 2023).
- [25] «Different Techniques in Topic Modeling: LDA, Mallet LDA, STM & HDP | by Ramya Balakrishnan | Medium». <https://balakrishnanr-3390.medium.com/in-natural-language-processing-topic-modeling-is-a-type-of-statistical-modeling-that-is-used-to-8640636f3641> (accedido 26 de agosto de 2023).
- [26] «What Is Named Entity Recognition (NER)? | Definition from TechTarget». <https://www.techtarget.com/whatis/definition/named-entity-recognition-NER> (accedido 26 de agosto de 2023).
- [27] S. P. Borgatti y T. J. Grosser, «Structural Equivalence: Meaning and Measures», *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, pp. 621-625, mar. 2015, doi: 10.1016/B978-0-08-097086-8.43120-X.
- [28] J. Brender, «Descriptions of Methods and Techniques», *Handbook of Evaluation Methods for Health Informatics*, pp. 73-225, 2006, doi: 10.1016/B978-012370464-1.50007-1.
- [29] «A Guide to Social Network Analysis and its Use Cases | LatentView». <https://www.latentview.com/blog/a-guide-to-social-network-analysis-and-its-use-cases/> (accedido 26 de agosto de 2023).

- [30] U. Brandes, «Social Network Algorithms and Software», *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, pp. 454-460, mar. 2015, doi: 10.1016/B978-0-08-097086-8.43121-1.
- [31] «API Reference | YouTube Data API | Google for Developers». <https://developers.google.com/youtube/v3/docs?hl=en> (accedido 27 de agosto de 2023).

ANEXO

Fuente de datos y código