

Aplicación de la Metodología CRISP-DM para la determinación del Impacto de las Emociones en la interacción que tienen los jóvenes con las Redes Sociales

Corporación Universitaria Remington
Facultad de Ingenierías
Ingeniería de Sistemas

Brandon Alexander Maestre Saldarriaga¹
Juan Felipe Serna Sanchez²
Ivonne Castaño Osorio³
Jhon Edison Amortegui⁴
Opción de Trabajo de grado: Seminario
2024

¹Estudiante de noveno semestre de Ingeniería de Sistemas. Uniremington Pereira email: brandon.maestre.7709@miremington.edu.co

²Estudiante de noveno semestre de ingeniería de sistemas. Uniremington Pereira email: juan.serna.0028@miremington.edu.co

³Asesor Temático Uniremington Pereira. Email: ivonne.castano@uniremington.edu.co

⁴Asesor Metodológico. Uniremington Pereira. Email: john.amortegui@uniremington.edu.co

Dedicatoria

Queremos dedicarle el siguiente trabajo a nuestras familias y a las futuras generaciones las cuales vean este documento en un futuro como aporte a su conocimiento o como guía para su educación y que esperamos que en él encuentren todo lo necesario para desarrollar de mejor manera su proceso académico y/o para inspirarse y crear nuevas cosas, proyectos o ideas en un futuro lejano o no.

También queremos extender la dedicatoria a nuestros maestros los cuales nos instruyeron e inspiraron a lograr esto que estamos culminando a día de hoy ya que en el presente documento se ve reflejado el tiempo y conocimiento transmitido a nosotros sus estudiantes.

Agradecimientos

Queremos agradecer principalmente a Dios y a la Universidad por tan excelente oportunidad de crecimiento tanto profesional como personal, la cual nos ha ayudado a mejorar día a día nuestros conocimientos en nuestra carrera, en nuestro trabajo y en nuestra vida personal.

También queremos agradecer a cada una de las personas que estuvieron durante nuestro proceso académico como lo son nuestras familias, profesores y compañeros los cuales siempre estuvieron ahí para nosotros cuando más los necesitamos y nos brindaron un apoyo incondicional y esencial el cual nos ayudó durante todo este extenso pero gratificante proceso educativo.

Tabla de contenidos

Resumen	6
Palabras clave	6
Pregunta orientadora de la búsqueda	7
Metodología de búsqueda de la información	8
Sustentación teórica de la pregunta	15
Conclusiones	20
Referencias	21

Tabla de Gráficos

Gráfico 1. Emociones vs Tiempo de uso	12
Gráfico 2. Promedio de mensajes enviados vs Emociones.....	16
Gráfico 3. Análisis de LinkedIn.....	17
Gráfico 4. Análisis de Instagram	17
Gráfico 5. Análisis de Twitter.....	18
Gráfico 6. Análisis de Publicaciones por día vs Plataforma.....	19
Gráfico 7. Análisis de Tiempo de uso vs Plataforma.....	19

Resumen

El presente trabajo tiene como objetivo utilizar la ciencia de datos para dar una respuesta, a través de un análisis profundo y con la ayuda de diferentes herramientas, a la pregunta problematizadora que se desarrollará a lo largo del proyecto. Se utilizará un dataset que contiene datos de encuestas y registros de redes sociales, incluyendo información sobre edad, género, estado de ánimo y uso promedio (en minutos) de las plataformas. Además, el análisis incluirá la creación de gráficos que representen la relación entre estas variables y las emociones de los jóvenes. Este documento permitirá analizar la importancia de los datos como herramienta para comprender mejor qué emociones son las que mayor influencia tienen en las interacciones de los jóvenes en las redes sociales.

Palabras clave: Ciencia de Datos, Big Data, Análisis de Datos, Redes Sociales, Machine Learning, Emociones.

Pregunta orientadora de la búsqueda

El uso de internet y de las redes sociales ha crecido exponencialmente en los últimos años. Esto ha llevado a un aumento significativo en el número de usuarios, especialmente entre los jóvenes, que pasan gran parte de su día en estas plataformas. Este incremento ha suscitado un gran interés en entender por qué motivo son tan frecuentes las interacciones de los jóvenes y si algún estado de ánimo específico los atrae a buscar las redes sociales.

De acuerdo con el artículo: *Impacto del aumento del uso de Internet y las redes sociales en la salud mental de jóvenes y adolescentes*: “Algunas publicaciones especializadas el uso de Internet provocaría fundamentalmente emociones positivas en los adolescentes. El 96,9% reconoce sentir alegría o risa en la red; un 81,6%, tranquilidad o relajación. El 78,9%, placer o diversión y el 71,6%, apoyo o comprensión” (Observatorio Nacional de Tecnología y Sociedad et al., 2024, p. 11). Por lo tanto, se busca encontrar por medio del uso de análisis de datos posibles acercamientos a la siguiente pregunta **¿Qué emociones son las que mayor influencia tienen en las interacciones de los jóvenes en las redes sociales?**

Metodología de búsqueda de la información

Este proyecto será desarrollado haciendo uso de la metodología Crisp-DM la cual es ampliamente utilizada en tareas de minería de datos y análisis predictivo, su documentación fue extraída del sitio web de IBM (*Conceptos Básicos De Ayuda De CRISP-DM*). Comúnmente, esta metodología, al ser iterativa, permite mejorar los resultados a través de la repetición de ciertas fases, de ser necesario.

La metodología CRISP-DM está dividida en diez fases. La primera fase es el entendimiento del negocio, en la cual se buscan comprender y establecer los objetivos del proyecto, ya sea desde la perspectiva de un negocio o de un problema a resolver.

En la segunda fase, se encuentra el enfoque analítico, en el cual se define cómo se orientarán los datos para determinar qué tipo de análisis se emplea en el proyecto. La tercera fase son los requisitos de datos, donde se recopilan los datos necesarios para llevar a cabo el análisis, identificando su fuente y tipo, y verificando su calidad.

La cuarta fase es la recopilación de datos, en la cual se determina dónde se obtendrán los datos, ya sea de un sitio web como Kaggle, una base de datos propia o compartida. En la quinta fase se encuentra la comprensión de datos, en la cual se realiza un análisis detallado de los datos obtenidos para identificar sus atributos principales, permitiendo saber si las variables cumplen con los requisitos necesarios para el desarrollo del proyecto.

La sexta fase es la preparación de datos, donde se lleva a cabo la limpieza y preparación de los datos, garantizando un análisis exitoso. En la séptima fase está el

modelado, donde se utilizan diversas herramientas y técnicas para encontrar tendencias y patrones de comportamiento.

La octava fase consiste en la evaluación, donde se miden los modelos para determinar su efectividad y precisión en las tendencias. La novena fase es la de despliegue, donde los resultados se presentan de una manera más gráfica y comprensible para facilitar su interpretación.

La décima fase es la de revisión, donde se realiza una revisión final de todo lo realizado en el proyecto, permitiendo encontrar posibles mejoras a futuro y nuevos campos de investigación.

A continuación, se ahondará en dichas fases de la metodología, enfocándose en el estudio y análisis a realizar con los propios datos del proyecto. Se explicará de manera detallada lo realizado en cada paso, aportando también contexto para entender mejor el proceso en cuestión.

- **Entendimiento del Negocio:** en este proyecto es de suma importancia hacer un análisis detallado de los datos, dado que con ello se pueden identificar patrones, tendencias y correlaciones que de otra manera podrían pasar desapercibidos. “Esto brinda la capacidad de anticipar y prever eventos futuros con mayor precisión, lo que puede ser invaluable en una variedad de contextos” (Cruz, 2024, p. 10). Esto se puede observar al analizar cómo los estados de ánimo influyen en la interacción de los jóvenes en redes sociales; al identificar esta correlación, es posible encontrar una tendencia de comportamiento acerca de su uso mediante los resultados obtenidos con los análisis.

El uso de redes sociales ha venido creciendo en los últimos años. Common Sense Media (2019) informa que los jóvenes pasan un promedio de 7 horas diarias usando redes sociales. Los principales motivos pueden ser el fácil acceso a través de sus celulares, la búsqueda de aceptación social y el entretenimiento instantáneo que se genera en estas plataformas. En el informe se puede observar el rango de edad y cuáles son las redes que se utilizan:

Jóvenes (15-30 años): Los jóvenes son el grupo de edad más activo de los usuarios de redes sociales. Son hiperconectados, usan múltiples plataformas regularmente y son activos en comunidades en línea. Utilizan las redes sociales para mantener relaciones con sus amigos y familiares, compartir experiencias, recibir información, entretenimiento y descubrir nuevas formas de expresión e identidad. (Quiroz et al., 2023, p. 9)

Esto permite observar que las redes sociales más usadas por los jóvenes las cuales son, y en este rango de edad es donde mayor tiempo se pasa, por otra parte, ya el rango más adultos, se ve que se utiliza LinkedIn la cual:

Es la red profesional más grande del mundo en Internet. Esta web conecta a empresas y empleados, en ella puedes encontrar el trabajo o la pasantía adecuada para ti. Además, puedes fortalecer tus relaciones profesionales y aprender las habilidades que necesitas para tener éxito en tu carrera. (Jelpit, 2023)

Por otro lado, es interesante analizar el estudio realizado por Ellison (2016) el cual permite concluir como el estado de ánimo de los jóvenes puede influir en su interacción o tiempo de uso de las redes sociales, esto es debido a que muchos a sentirse solos o con

depresión tienden a usar estas plataformas como un medio de escape o de entretenimiento a su realidad. Lo cual termina en un ciclo de dependencia excesiva de las redes.

El artículo titulado: Las redes sociales y la salud mental dice lo siguiente con respecto a cómo influye el uso de las redes sociales en la condición mental de las personas:

El uso intensivo de las redes sociales podría estar ocultando otros problemas subyacentes, como el estrés, la depresión o el aburrimiento. Si pasa más tiempo en las redes sociales cuando se siente deprimido, solo o aburrido, puede estar usándolas como una forma de distraerse de los sentimientos desagradables o para calmar su estado de ánimo. Aunque puede resultar difícil al principio, permitirse sentir puede abrirles las puertas a formas más sanas de controlar su estado de ánimo. (Smith & Robinson, 2024)

Esto puede comprobarse de manera efectiva si se visualiza la **Gráfico 1** del presente documento, en la que se representa el tiempo de uso en relación a las emociones, allí podemos evidenciar que cuanto más negativa sea la emoción, más tiempo se invierte en redes sociales.

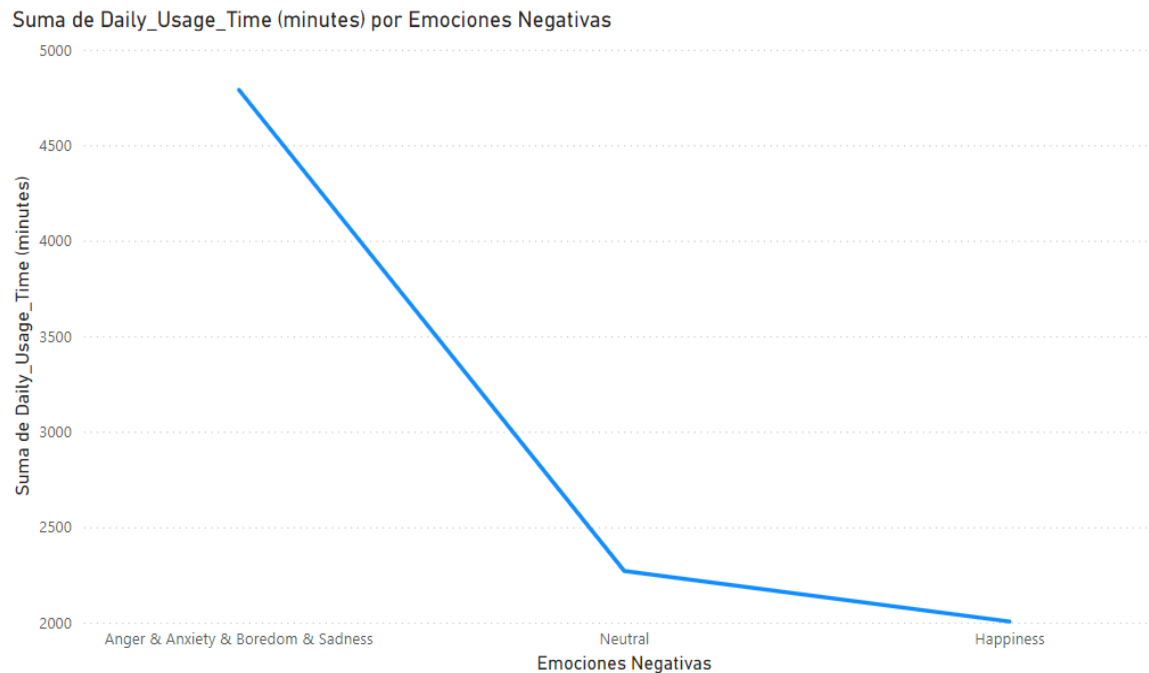


Gráfico 1. Emociones vs Tiempo de uso.

- Si se analiza el **Gráfico 1** se puede observar que cuánto más negativa sea la emoción que sienta una persona, más tiempo se invierte en redes sociales.
- Entre más felicidad sienten las personas, menos tiempo pasan en redes sociales.

Por otro lado, se puede observar como “A su vez, estos sentimientos afectan negativamente su estado de ánimo y empeoran los síntomas de depresión, ansiedad y estrés.” (Franco, 2021).

Este artículo permite ver de una manera más detallada lo mencionado, donde el psicólogo hace referencia a que cuando se usan las redes sociales de manera excesiva bajo los estados de ánimo negativos, estos tienden a empeorar debido a la creación de dependencia a las plataformas.

- **Enfoque Analítico:** Se determina el entendimiento del negocio y se puede definir el enfoque analítico de manera descriptiva/correlacional el cual se usará para resolver el problema planteado, haciendo uso de técnicas y herramientas de análisis y modelos los cuales permiten examinar los datos en detalle.
- **Requisitos de Datos:** En esta fase se recopilaron los datos requeridos para poder hacer el análisis correspondiente, lo cual incluye la identificación de la fuente y tipos de datos, así como también la calidad y la cantidad de los mismos.
- **Recopilación de Datos:** Se investigó un dataset extraído del sitio web “Kaggle” (BULUT, 2024) en el cual se tiene la información de la interacción de los jóvenes en redes sociales. Se analizó a detalle para determinar si era útil en el proyecto y si cumplió con los requisitos necesarios para ser utilizado.
- **Comprensión de Datos:** En esta fase se hace un análisis detallado de los diferentes datos recolectados para así identificar cuáles son sus atributos principales y verificar la calidad de los mismos, lo cual permitió encontrar que se cumple con las variables necesarias para el desarrollo del proyecto, encontrando la edad, género, uso promedio de redes sociales, tipo de interacción en cada plataforma y los estados de ánimo. Datos los cuales permitirán encontrar patrones de comportamiento y en cuales nos podemos centrar.
- **Preparación de Datos:** después de la debida limpieza a Al dataset en el cual se removió información en blanco y poco relevante para nuestro estudio, además de incluir la tabla de datos en Power BI para su mejor visualización, dichos datos se

encuentran en una base de datos la cual consta de 9 columnas, aproximadamente 16000 registros útiles distribuidos entre 7 de esas 9 columnas y gracias a dichos

Edad	Género	Plataforma	Tiempo de Uso Diario (minutos)".	Publicaciones Por Día	Me_Gusta Recibidos Por Día	Comentarios Recibidos Por Día	Mensajes Enviados Por Día	Emoción Dominante
27	Femenino	Snapchat	120	4	40	18	22	Neutral
21	No-binario	Snapchat	60	1	18	7	12	Neutral
28	No-binario	Snapchat	115	3	38	18	27	Ansiedad
27	Masculino	Telegram	105	3	48	20	28	Ansiedad
21	No-binario	Facebook	55	3	17	7	12	Neutral
25	Femenino	Instagram	160	6	85	26	30	Felicidad
25	Femenino	Instagram	160	6	85	26	30	Felicidad
21	No-binario	Facebook	55	3	17	7	12	Neutral
24	No-binario	Snapchat	85	3	33	20	18	Felicidad
33	Masculino	LinkedIn	45	1	10	5	12	Aburrimiento
32	Femenino	Instagram	140	5	70	22	33	Felicidad
23	No-binario	Facebook	105	2	25	12	18	Neutral
32	Femenino	Instagram	140	6	75	28	30	Felicidad
25	Femenino	Instagram	160	6	85	26	30	Felicidad
35	Masculino	Twitter	70	1	13	8	10	Aburrimiento

registros se pudo desarrollar la anterior tabla:

Tabla 1. Estadísticas

Sustentación teórica de la pregunta

El aumento de jóvenes usando las redes sociales cada vez va creciendo, y por ende también los que se vuelven adictos o crean dependencia de ellas, esto está creando problemas de su salud mental por diferentes causas pero con grandes consecuencias “varios estudios han encontrado una fuerte relación entre el uso excesivo de las redes sociales y un mayor riesgo de depresión, ansiedad, soledad, autolesión e incluso pensamientos suicidas.” (Smith & Robinson, 2024), esto se podrá comprobar más en detalle haciendo uso de gráficas las cuales mostrarán dichos patrones o comportamientos con los datos que se han obtenido para el estudio.

El dataset seleccionado aportó la información necesaria para llegar a las siguientes conclusiones:

- Las personas que sienten emociones negativas tienden a ser las que más tiempo dedican a las redes sociales y también las que más interacciones tienen.
- Alrededor del 15% de las personas las cuales se encuentran en la tabla de datos hacen uso de la red social Instagram estando en un rango de edad de entre los 27 a 32 años de edad.
- La cantidad de mensajes enviados en redes sociales varía en función de la emoción que se tenga al momento de entrar a la plataforma, es por esto que se

puede ver en el **Gráfico 2** que por lo general las personas felices o que sientan ansiedad tienden a ser las que más mensajes envían desde estas plataformas.

Suma de Mensajes_Enviados_Por_Día por Emoción_Dominante

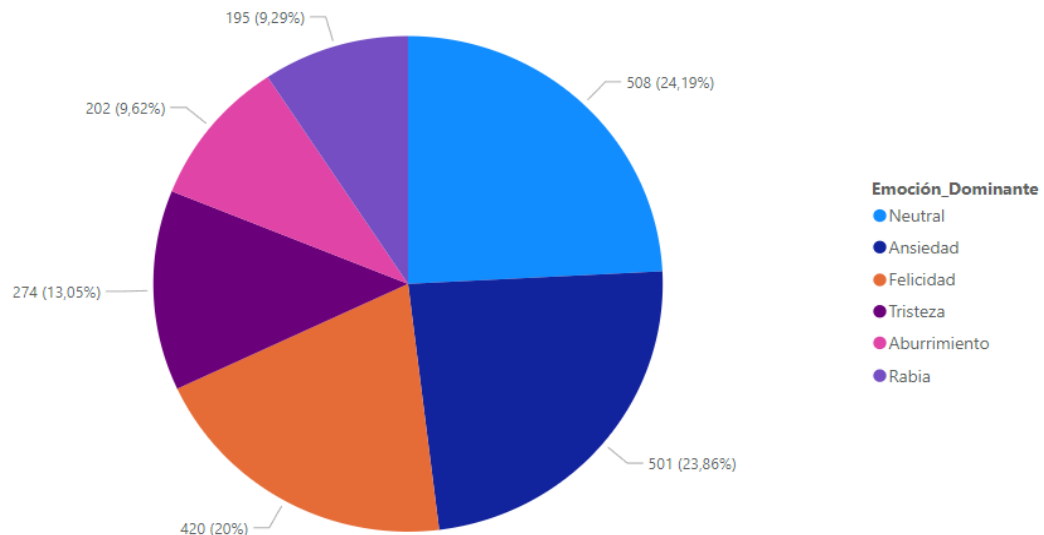


Gráfico 2. Promedio de mensajes enviados vs Emociones.

- De acuerdo a el **Gráfico 3** se puede evidenciar que las personas que hacen uso de la red social LinkedIn son personas que en su mayoría sienten aburrimiento, se teoriza que puede ser debido a que esta plataforma se usa como medio para encontrar un nuevo empleo por lo que podría ser una explicación a este curioso patrón

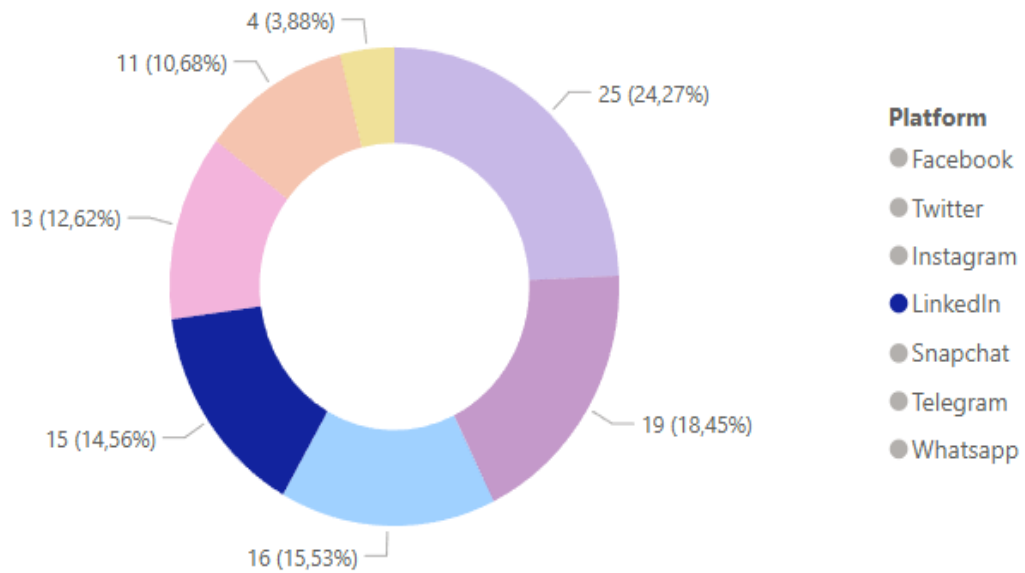


Gráfico 3. Análisis de LinkedIn.

- De acuerdo con el **Gráfico 4** se puede identificar que por lo general en su mayoría las personas que sienten felicidad tienden a usar más la red social Instagram.

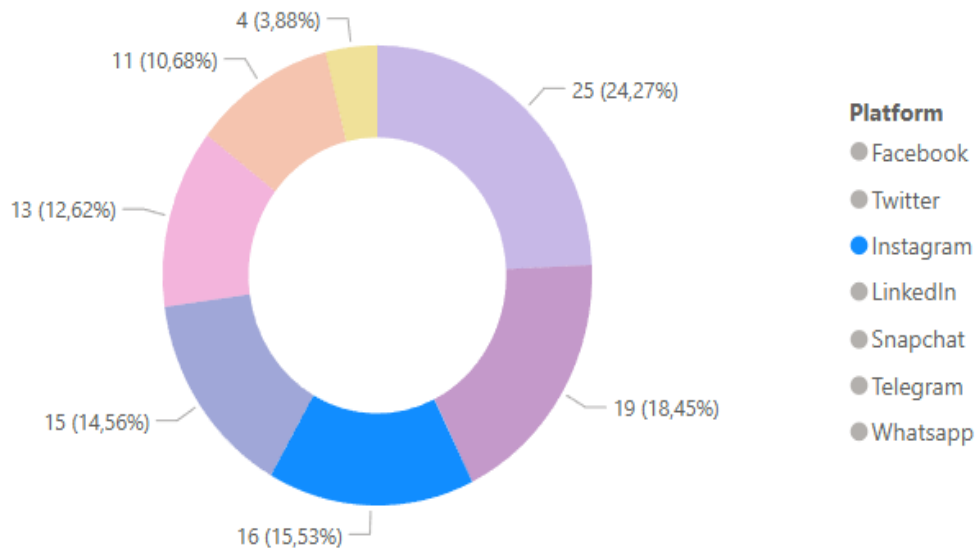


Gráfico 4. Análisis de Instagram.

- Se puede identificar que por lo general Twitter tiene una mayor cantidad de personas que sentían emociones negativas como rabia, tristeza, ansiedad y aburrimiento en comparación con las otras redes sociales. Para referencias véase el **Gráfico 5**.

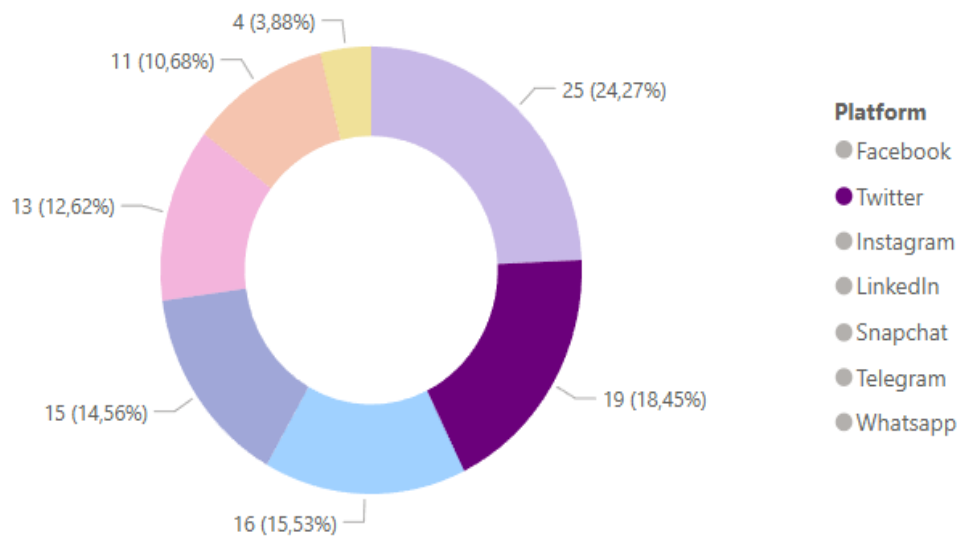


Gráfico 5. Análisis de Twitter.

- Se puede concluir que la red social más utilizada es Instagram al comparar la cantidad de posts publicados al día distribuidos por plataforma y en dicha gráfica se puede evidenciar de manera clara la superioridad de esta red social por sobre las demás, siendo esto información increíblemente útil para aquellas empresas o personas que pretendan hacer publicidad. Véase en el **Gráfico 6** y **Gráfico 7** para una mejor visualización.

Suma de Posts_Per_Day por Platform

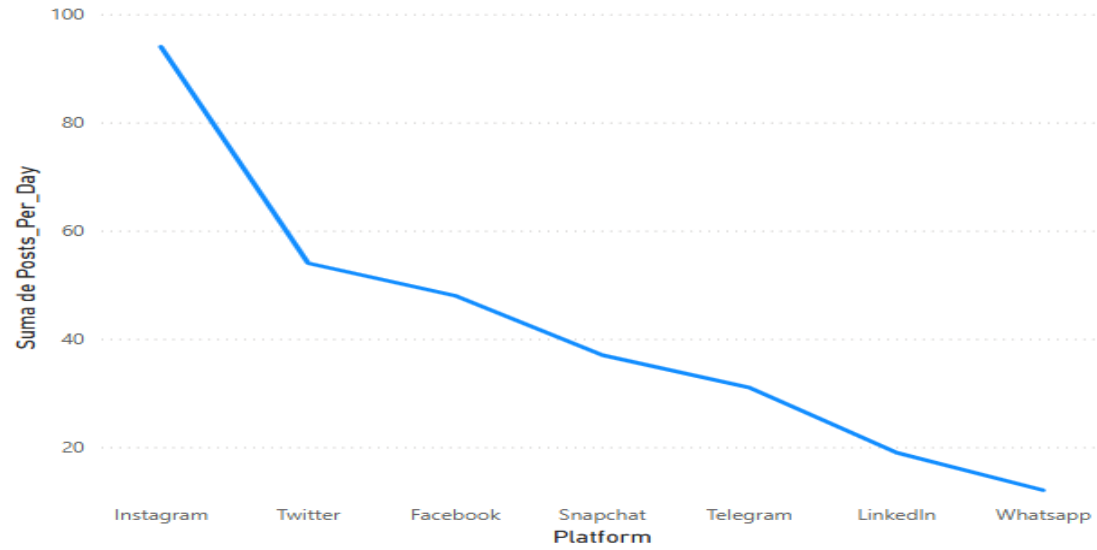


Gráfico 6. Análisis de Publicaciones por día vs Plataforma

Suma de Daily_Usage_Time (minutes) por Platform

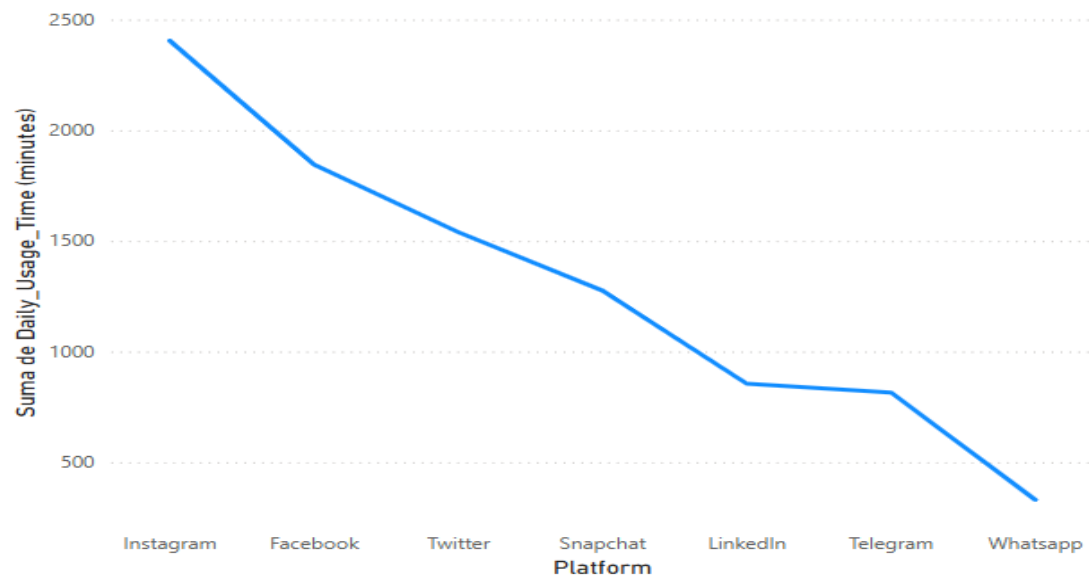


Gráfico 7. Análisis de Tiempo de uso vs Plataforma

Conclusiones

Gracias al Seminario de Big Data se pudo comprender de manera clara y eficaz como hacer un correcto análisis de los datos que se tienen, para extraer información interesante o valiosa para las compañías y personas, es por esto que podemos resaltar la labor que cumple el análisis de datos en negocios y estudios como el realizado en este caso para poder entender mejor alguna problemática, error o situación a mejorar, ya que estas metodologías nos permiten encontrar el trasfondo de los datos, algo así como la historia de los mismos para a partir de dicha historia, realizar un análisis exhaustivo y aportar valiosa información a todo tipo de proyectos o empresas que los requieran.

El análisis de datos es una herramienta poderosa con múltiples beneficios para todo aquel que sepa cómo desenvolverse en este tipo de ejercicios, lo único negativo que podríamos encontrarle de momento es que es necesario tener un conocimiento fundamental del uso de aplicaciones de visualización tipo Power BI, así como de otras herramientas como lo puede ser Excel ya que son esenciales para el desarrollo e implementación de este tipo de ejercicios.

Lista de referencias

BULUT, E. (2024). *Social Media Usage and Emotional Well-Being* [Explore the connection between social media usage patterns and emotional.].

<https://www.kaggle.com/>. Retrieved 6 15, 2024, from

<https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being/data>

Common Sense Media. (2019). The Common Sense census. : *Media use by tweens and teens*.

<https://www.commonsensemedia.org/sites/default/files/research/report/2019-census-8-to-18-full-report-updated.pdf>

Conceptos básicos de ayuda de CRISP-DM. (n.d.). IBM. Retrieved 6 15, 2024, from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Cruz Santos, E. F. (2024). *ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES DE PUNTOS CALIENTES EN EL*

AÑO 2021 Y 2022 UTILIZANDO ESTRATEGIAS DE MACHINE

LEARNING. <https://repositorio.uniremington.edu.co/>. Retrieved 6 15, 2024, from

<https://repositorio.uniremington.edu.co/server/api/core/bitstreams/6c1fdd27-37b8-4eeb-bffe-41d56bb98368/content>

Ellison, J. B. (2016). Sharing the small moments: Ephemeral social interaction on Snapchat. *Information, Communication & Society*, 19(19), 956-977.

10.1080/1369118X.2015.1084349

Franco Carranza, J. A. (2021, July 13). *Redes sociales y salud mental | Psicólogo Jorge Franco Carranza*. Psicólogo Jorge Alberto Franco Carranza.

Retrieved June 15, 2024, from

<https://psicologojorgefranco.com/2021/07/13/redes-sociales-y-salud-mental/>

Jelpit. (2023, May 30). *Qué es LinkedIn y para qué sirve esta red profesional*.

Jelpit. Retrieved June 15, 2024, from

<https://www.jelpit.com/blog/empresas/emprendimiento/que-es-linkedin-y-para-que-sirve-esta-red-profesional/>

Observatorio Nacional de Tecnología y Sociedad, Red.es., & Secretaría de Estado de Digitalización e Inteligencia Artificial. (2024, 03). *Impacto del*

aumento del uso de Internet y las redes sociales en la salud mental de jóvenes y adolescentes. Retrieved 06 12, 2024, from

<https://fundadeps.org/wp-content/uploads/2024/03/ONTSI-redes-sociales-salud-mental-jovenes-adolescentes.pdf>

Quiroz, C. D., Gómez, M. D., Cárdenas, H. C., & Mira, J. F. (2023). *Estadísticas sobre el uso de redes sociales en Colombia del 2021 al 2023*.

<https://repositorio.uniremington.edu.co>. Retrieved 6 15, 2024, from

<https://repositorio.uniremington.edu.co/server/api/core/bitstreams/f542005>

[9-81a3-4bff-beed-d81519bab4ea/content](https://repositorio.uniremington.edu.co/server/api/core/bitstreams/f542005-9-81a3-4bff-beed-d81519bab4ea/content)

Smith, M., & Robinson, L. (2024, 2 21). *Las redes sociales y la salud mental*.

Retrieved 6 15, 2024, from [https://www.helpguide.org/es/problemas-de-](https://www.helpguide.org/es/problemas-de-la-adolescencia/las-redes-sociales-y-la-salud-mental)

[la-adolescencia/las-redes-sociales-y-la-salud-mental](https://www.helpguide.org/es/problemas-de-la-adolescencia/las-redes-sociales-y-la-salud-mental)