



**Data Queens**

# **Desafío Hey, Banco**

**Business Case**

**Datathon 2024**

05 de mayo de 2024

# ÍNDICE

<b>Resumen ejecutivo</b>	<b>3</b>
<b>Antecedentes</b>	<b>3</b>
<b>Alternativas</b>	<b>4</b>
<b>Solución</b>	<b>4</b>
<b>Análisis de factibilidad</b>	<b>5</b>
Factibilidad Técnica	5
Factibilidad de Datos	5
Factibilidad Operativa	5
Factibilidad Financiera	5
<b>Caso de éxito</b>	<b>6</b>
<b>Referencias</b>	<b>6</b>

## Resumen ejecutivo

El análisis de datos es esencial en la actualidad, especialmente en un mundo donde la cantidad de datos generados crece exponencialmente. Según una encuesta realizada por Rethink Data en 2020, el 68% de los datos generados por las empresas no son aprovechados, lo que subraya la necesidad de implementar soluciones efectivas para analizar y utilizar estos datos de manera significativa.

Una de las áreas clave donde se pueden aplicar estas soluciones es en el análisis de texto y sentimientos en redes sociales. Las redes sociales son una fuente invaluable de información, y comprender el sentimiento detrás de las publicaciones puede proporcionar una visión profunda de la opinión pública y las tendencias emergentes.

Para abordar esta necesidad, nuestra solución combina técnicas de procesamiento de lenguaje natural (NLP) y visualización de datos para analizar el sentimiento en textos, identificar palabras clave y emojis más utilizados, y presentar estos hallazgos de manera visual en un dashboard interactivo. Este proporciona una manera clara y comprensible de interpretar los datos y encontrar áreas de oportunidad para mejorar los procesos internos.

En términos de factibilidad, la implementación técnica de la solución es viable, aunque puede requerir recursos computacionales significativos, especialmente para el análisis de sentimientos con BERT. La disponibilidad y calidad de los datos son factores críticos a considerar, así como la experiencia del usuario al interactuar con el dashboard de visualización de datos.

Este enfoque ya ha demostrado ser exitoso en casos anteriores, como el de la marca Old Spice, que experimentó un aumento significativo en su presencia en línea y su tráfico web al comprender y dirigirse a su audiencia de manera efectiva.

En resumen, nuestra solución ofrece una forma efectiva de comprender el sentimiento y la semántica en textos, identificar temas clave y patrones de comportamiento, y presentar estos hallazgos de manera visual para una mejor comprensión y toma de decisiones.

## Antecedentes

De acuerdo a una encuesta realizada por Rethink Data (Gonzalez, 2020), donde se encuestaron a más de 1500 líderes empresariales globales, el 68% de los datos generados por las empresas no son aprovechados. Esta es una cifra importante ya que con el avance de las tecnologías, cada vez crece más la cantidad de datos generados en los aspectos diarios. Incluso, el mismo estudio predijo un aumento del 42.2% anual de crecimiento en la cantidad de datos empresariales. En la actualidad se generan aproximadamente 328.77 millones de terabytes diariamente (Duarte, 2023)

Esta información, de ser analizada, es capaz de proveer perspectivas sobre los procesos, usuarios y plantilla de trabajadores. En la era digital en la que vivimos, los datos representan cada vez más una ventaja competitiva para las empresas, ya que permiten identificar tendencias, anticipar cambios en el mercado, personalizar experiencias para los clientes y optimizar la toma de decisiones estratégicas. Sin embargo, el desafío radica en cómo las empresas pueden aprovechar al máximo esta gran cantidad de datos y convertirlos en insights accionables que impulsen el crecimiento y la innovación empresarial.

## Alternativas

Para la realización del análisis de sentimientos exploramos diferentes alternativas que involucran distintas librerías y herramientas para el análisis de los datos.

- **TextBlob:** Utiliza un enfoque basado en reglas y lexicones para analizar el sentimiento. Proporciona una polaridad del sentimiento y se puede utilizar directamente para análisis de sentimientos en varios idiomas, incluido el español.
- **BERT:** Utiliza un modelo pre entrenado de aprendizaje profundo para comprender el contexto y la semántica del lenguaje. Proporciona una clasificación de sentimiento más avanzada y precisa, pero puede requerir más recursos computacionales y conocimientos técnicos para implementar.
- **Pattern.es:** Proporciona análisis de sentimientos específicamente para texto en español. Es similar a TextBlob pero está optimizado para el idioma español. Sin embargo, puede tener limitaciones en la precisión del análisis de sentimientos en ciertos casos.
- **Emojis:** Introducimos el manejo de emojis en el análisis de sentimientos para capturar mejor el sentimiento expresado en los textos que contienen emojis. Las diferentes implementaciones incluyen buscar emojis específicos en el texto, agregarlos al final del texto o eliminarlos antes de realizar el análisis de sentimientos.

## Solución

Las redes sociales son una herramienta de expresión para la población, las cuales tienen un alto nivel de accesibilidad. Para utilizar las publicaciones relacionadas con Hey, Banco nuestra solución utiliza una combinación de técnicas de procesamiento de lenguaje natural (NLP) y visualización de datos para analizar el sentimiento en textos, identificar palabras clave y emojis más utilizados, y presentar estos hallazgos de manera visual en un dashboard. La solución consta de los siguientes pasos:

El análisis de sentimientos se logró mediante un modelo preentrenado de BERT (*Bidirectional Encoder Representations from Transformers*) para el análisis de los textos. BERT es conocido por su capacidad para comprender el contexto y la semántica del lenguaje, lo que lo hace ideal para tareas de análisis de sentimientos más avanzadas y precisas.

Aplicamos el algoritmo TF-IDF (Term Frequency-Inverse Document Frequency) para extraer las palabras y oraciones clave más relevantes de los textos. TF-IDF asigna un peso a cada palabra en función de su frecuencia en el texto y su frecuencia inversa en el conjunto de documentos, lo que nos permite identificar las palabras más importantes y distintivas en cada comentario.

Además, se realizó un recuento de los emojis más utilizados en los textos. Los emojis son una parte importante de la comunicación en línea y pueden proporcionar información adicional sobre el sentimiento expresado en un texto.

Por último, los hallazgos se unifican en un dashboard interactivo. El dashboard presenta visualmente estos datos de manera clara y comprensible, utilizando gráficos y tablas para resaltar patrones y tendencias en los comentarios. A partir de esta visualización, se pueden interpretar los resultados para encontrar áreas de oportunidad y mejorar los procesos internos.

Con esta solución, proporcionamos una manera efectiva de comprender el sentimiento y la semántica en textos, identificar temas clave y patrones de comportamiento, y presentar estos hallazgos de manera visual para una mejor comprensión y toma de decisiones.

## Análisis de factibilidad

### Factibilidad Técnica

**Análisis de Sentimientos con BERT:** El uso de BERT para el análisis de sentimientos requiere recursos computacionales significativos, especialmente si se trabaja con grandes volúmenes de texto. Sin embargo, con el aumento de la disponibilidad de hardware y la optimización de bibliotecas como transformers, es técnicamente factible implementar esta parte de la solución.

**Extracción de Palabras Clave con TF-IDF:** TF-IDF es un algoritmo simple y eficaz que se puede implementar fácilmente con bibliotecas de procesamiento de texto como scikit-learn. Su implementación técnica es relativamente sencilla y no requiere recursos computacionales significativos.

### Factibilidad de Datos

**Disponibilidad de Datos:** La factibilidad de la solución depende de la disponibilidad de datos adecuados para realizar el análisis de sentimientos, la extracción de palabras clave y la identificación de emojis. Es importante asegurarse de que haya suficientes datos etiquetados para entrenar y validar el modelo de análisis de sentimientos.

**Calidad de los Datos:** La calidad de los datos es crucial para el éxito de la solución. Es importante que los datos estén limpios, bien estructurados y representativos del dominio de aplicación para obtener resultados precisos y significativos.

### Factibilidad Operativa

**Implementación y Mantenimiento:** La implementación de la solución requerirá la colaboración de un equipo técnico con experiencia en procesamiento de lenguaje natural y visualización de datos. Una vez implementada, la solución requerirá mantenimiento regular para garantizar que los modelos y algoritmos estén actualizados y funcionando correctamente.

**Interacción con Usuarios:** Es importante considerar la experiencia del usuario al diseñar el dashboard de visualización de datos. Se deben realizar pruebas de usuario para garantizar que el dashboard sea intuitivo, fácil de usar y proporcione información relevante de manera clara y comprensible.

### Factibilidad Financiera

**Costos de Desarrollo y Mantenimiento:** La solución implicará costos asociados con el desarrollo inicial, la implementación y el mantenimiento continuo del sistema. Esto incluye los costos de hardware, software, recursos humanos y cualquier licencia de software necesaria como la de power bi pro.

Beneficios Potenciales: Se deben evaluar los beneficios potenciales de la solución en términos de mejora de la toma de decisiones, eficiencia operativa y satisfacción del cliente. Esto puede ayudar a justificar los costos asociados con la implementación y el mantenimiento de la solución.

## Caso de éxito

La marca reconocida por su línea de cuidado personal varonil, Old Spice, se vio desafiada por una pérdida significativa del 30 % en sus ventas durante el 2009. La empresa para combatir esta pérdida, realizó una encuesta en línea donde se reveló un hallazgo inesperado para la empresa: el 70 % de los productos de ducha masculinos eran comprados por mujeres para sus parejas masculinas. Este descubrimiento fue fundamental y marcó un cambio importante en la estrategia de la marca.

Con base en esto, Old Spice tomó una decisión estratégica: enfocar su campaña de marketing de 2010 en las mujeres. Reconociendo que las mujeres desempeñaban un papel clave en la compra de sus productos; la marca se propuso capturar su atención y fidelidad a través de sus mensajes publicitarios. Asimismo, logró protagonizar el anuncio clave de esta estrategia por el exjugador de la NFL Isaiah Mustafá. El comercial humorístico y orientado directamente hacia las mujeres, rápidamente se volvió viral en redes sociales, alcanzando más de 105 millones de visualizaciones en YouTube. Este éxito se tradujo en un aumento del 2300 % en seguidores en Twitter y un aumento del 300 % en el tráfico web hacia su sitio oficial para Old Spice, confirmando la efectividad de su estrategia centrada en las mujeres y destacando la importancia de comprender a fondo a la audiencia para el éxito en el marketing.

## Referencias

Moliner, J. (2023, 29 abril). *5 empresas con gran éxito en las redes sociales* - Joe Moliner. Joe Moliner. <https://www.joemoliner.com/blog/redes-sociales/>

Gonzalez, R. (2020, 03 noviembre). *El 68% de los datos que generan las empresas no se aprovechan*. Big Data Magazine.

Los datos y su importancia en la actualidad. (2024). Dataiq.mx. <https://blog.dataiq.mx/la-importancia-de-los-datos>

Duarte, F. (2023). *Amount of Data Created Daily*. Exploding Topics. <https://explodingtopics.com/blog/data-generated-per-day>