

hey,

Reto Datathon 2024

Cesar Romeo Vega Rodriguez

Rodrigo Arroyo Oropeza

Victoria Alejandra Velazquez Mendoza

Lucero Montserrat Salazar Loera

Statistics Stars



Presentación

hey, banco



@heybanco

PARTE 1:

Problemática

PARTE 2:

Conozcamos a nuestros
Clientes

PARTE 3:

Beneficios y Sugerencias

PARTE 4:

Conclusiones

PROBLEMÁTICA

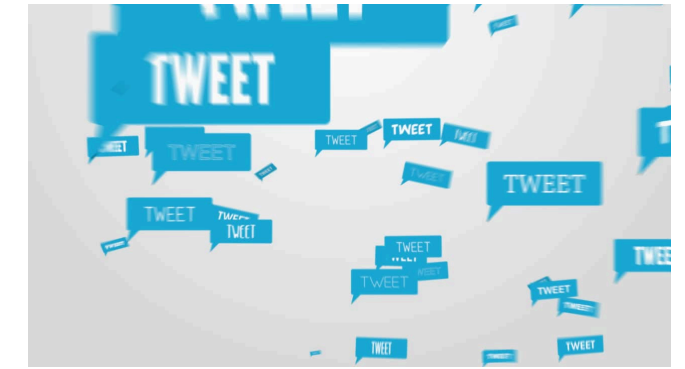
**¿Alguna vez quisiste
conocer lo que tus clientes
mencionan en redes
sociales?**

Estadísticas

1

HAY 18 MILLONES DE USUARIOS

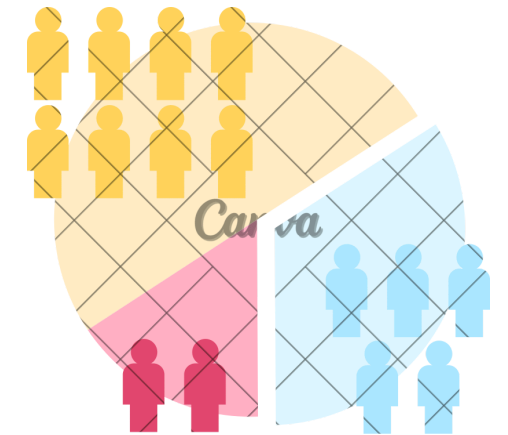
activos en México a finales de Enero 2024 (DataReportal, 2024)



2

EL 61.2% DE LOS USUARIOS MAYORES A 18 AÑOS

son hombres, mientras que solo el 38.8% son mujeres (DataReportal, 2024)



3

EL 56% DE LOS USUARIOS MEXICANOS DE REDES SOCIALES

usan X como su plataforma favorita (Statista, 2024)

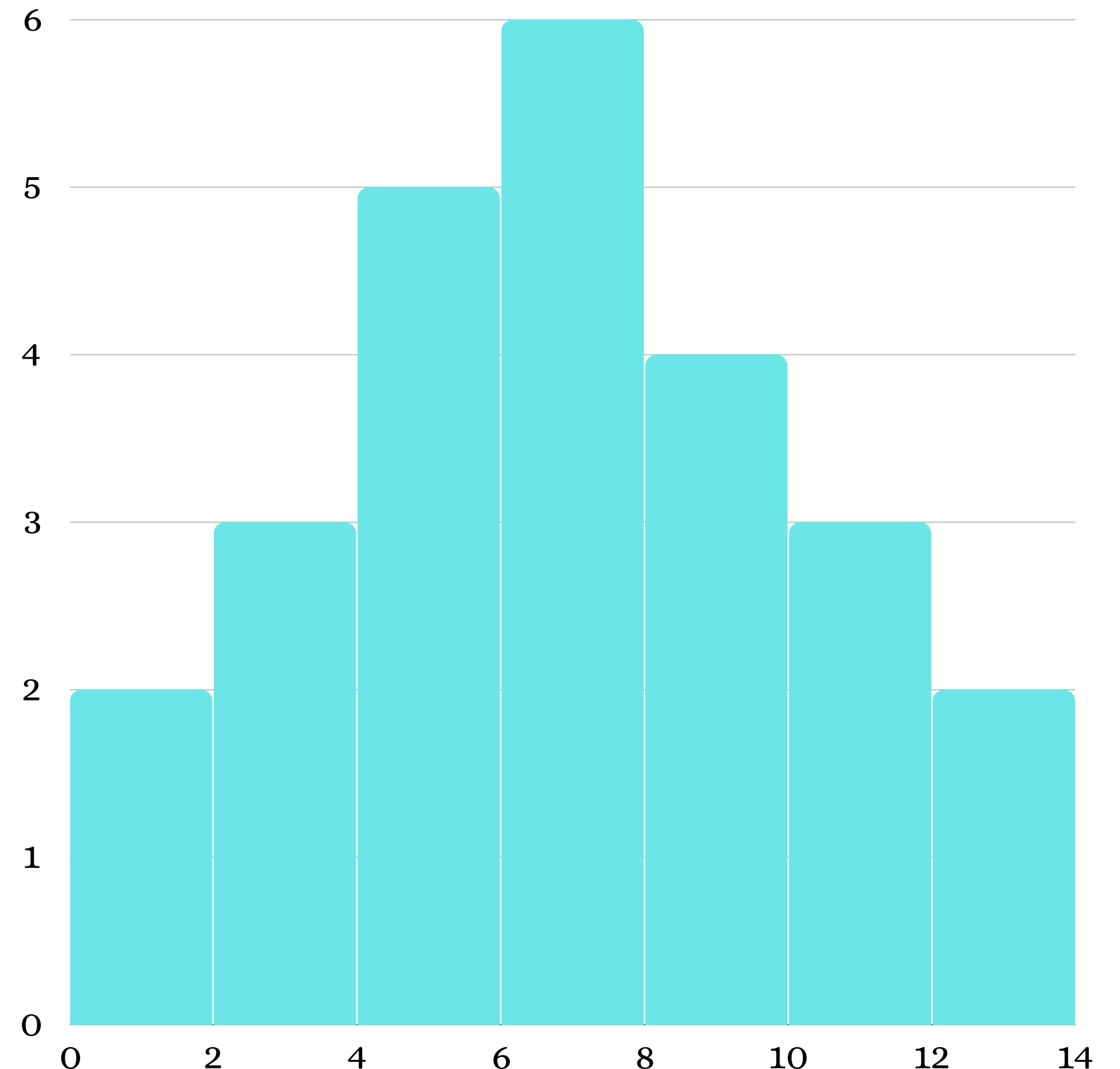
4

EL 87% DE LOS CLIENTES COMPARTEN LAS BUENAS EXPERIENCIAS CON OTROS,

Mientras que el 95% comparte las experiencias negativas. (Fuente: Zendesk)

¿CÓMO LLEGAREMOS AHÍ?

- Limpieza de Texto (Stopwords)
- *LEMANTIZACION
 - Análisis de sentimiento
 - Definición de “Tópicos”





LIBRERIAS USADAS

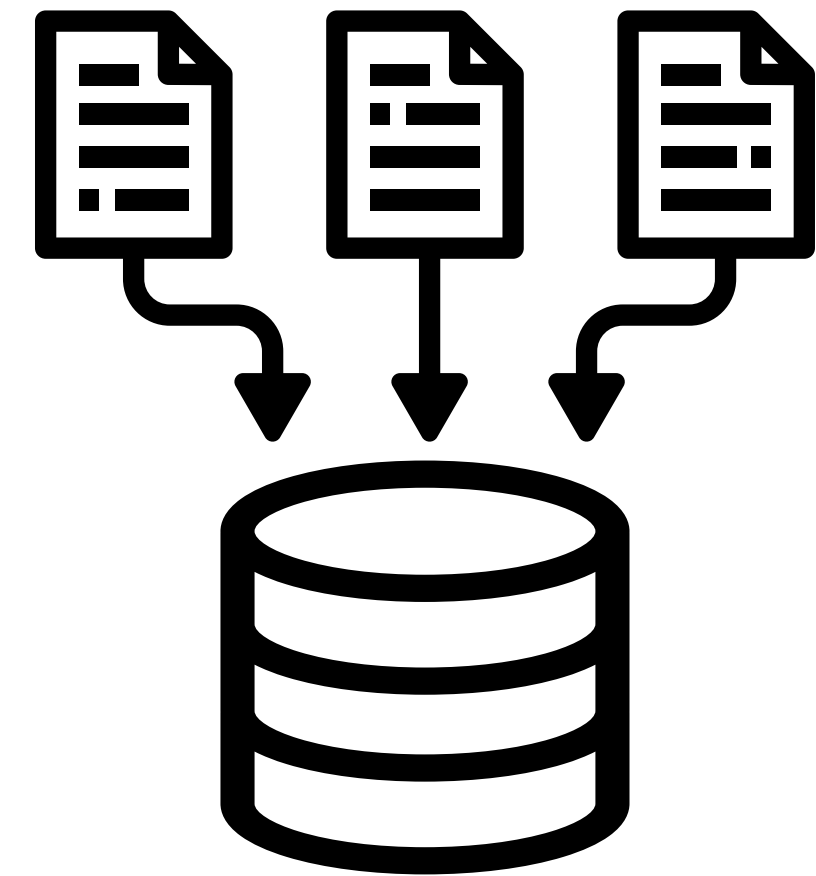
Nuestro script de **Python** utiliza diversas bibliotecas para análisis de datos y procesamiento de lenguaje natural (NLP), las cuales son:

- **NumPy:** Manipulación de datos numéricos.
- **Pandas:** Análisis de datos tabulares.
- **Matplotlib y Seaborn:** Visualización de datos.
- **NLTK y TextBlob:** Procesamiento de texto.
- **VADER Sentiment Analyzer:** Análisis de sentimientos.
- **Googletrans:** Traducción de texto.
- **spaCy:** Procesamiento avanzado de lenguaje natural.
- **tm:** Preprocesamiento, análisis y manipulación de datos textuales

TABLA FINAL

La tabla inicial estaba compuesta por fecha, tiempo y tweet. Aquí hay una explicación de las nuevas columnas:

- **year:** El año en el que se publicó el tweet.
- **month:** El mes en el que se publicó el tweet.
- **texto_tokenizado:** El contenido del tweet tokenizado en una lista de palabras para las categorías.
- **Razon/Categoria:** La razón o categoría asociada al tweet.
- **Porcentaje:** El porcentaje de palabras claves asociado al tweet.
- **tweet_english:** El contenido del tweet traducido al inglés.
- **neg, neu, pos:** Los puntajes de negatividad, neutralidad y positividad del tweet, respectivamente.
- **compound:** El puntaje compuesto del sentimiento del tweet.
- **sentiment:** El sentimiento asociado al tweet (positivo, negativo o neutral).
- **tweet_limpio:** El contenido del tweet limpio, despues de eliminar las “stopwords” y palabras sin importancia.



NUESTRA HERRAMIENTA...



Power BI

hey,



**CONOZCAMOS LO QUE
DICEN LOS CLIENTES**

Beneficios

- Validación de tweets a través del tiempo
- Satisfacción de clientes basada en tópicos únicos
- Identificación de Palabras Clave

Sugerencias

- Mejora del Tablero y análisis con variables:
“Usuario”, “Sub-categorías” y “Localización”
- Mejora en clasificación de tópicos con más
palabras clave

¡Muchas
gracias!

