

Caso de análisis:

La concesionaria AutoFlex, donde trabajamos como parte de una célula Data Science, está preocupada por la aceptación de la nueva versión de su marca estrella. Se han recibido muchos comentarios en redes sociales al respecto, y claramente se quiere optimizar el costo de procesamiento de esas opiniones. Para ello, desea interpretar esos comentarios utilizando el preprocesamiento de texto e implementar un modelo para resumir las opiniones de manera automática. El equipo de trabajo propone realizar un análisis de sentimientos que clasifique cada texto escrito, ya que la gerencia comercial tiene un conjunto de comentarios clasificados como "bueno", "malo" o "info" (necesita más información). Debemos realizar, primero, una descarga de la base, y terminar de sustentar metodológicamente el proyecto.

✓ Consignas

1. En primer lugar, debemos sustentar metodológicamente:

a. ¿Qué tipo de aplicación es un análisis de sentimientos? Por favor, en un párrafo de no más 6 líneas, explique en qué consiste, de manera que la gerencia de analítica lo comprenda fácilmente.

Respuesta: El análisis de sentimientos es una aplicación de procesamiento del lenguaje natural que identifica y clasifica las emociones expresadas en textos como positivos, negativos o neutrales. Utiliza algoritmos de aprendizaje automático y técnicas lingüísticas para evaluar opiniones de clientes en redes sociales, encuestas y reseñas, proporcionando a la gerencia de analítica información valiosa sobre las percepciones de los usuarios para apoyar la toma de decisiones.

b. ¿Qué tipo de procesamiento es necesario realizar primero? Exponga en un dibujo los pasos que va a realizar sobre los comentarios con el fin de convertirlos en data estructurada.

Respuesta: detallo los pasos:

1. Recolección de Datos
2. Preprocesamiento de Texto
3. Limpieza de Datos
4. Tokenización
5. Normalización
6. Eliminación de Palabras Irrelevantes
7. Etiquetado de Partes del Discurso
8. Extracción de Características
9. Análisis de Sentimientos y Clasificación

```

import matplotlib.pyplot as plt
import matplotlib.patches as patches

# Create a figure and axis
fig, ax = plt.subplots(figsize=(10, 8))

# Define the steps
steps = [
    "Recolección de Datos",
    "Preprocesamiento de Texto",
    " - Limpieza de Datos",
    " - Tokenización",
    "Normalización",
    "Eliminación de Palabras Irrelevantes",
    "Etiquetado de Partes del Discurso",
    "Extracción de Características",
    "Análisis de Sentimientos y Clasificación",
    "Conversión a Datos Estructurados"
]

# Define y positions for each step
y_positions = list(range(len(steps) * 2, 0, -2))

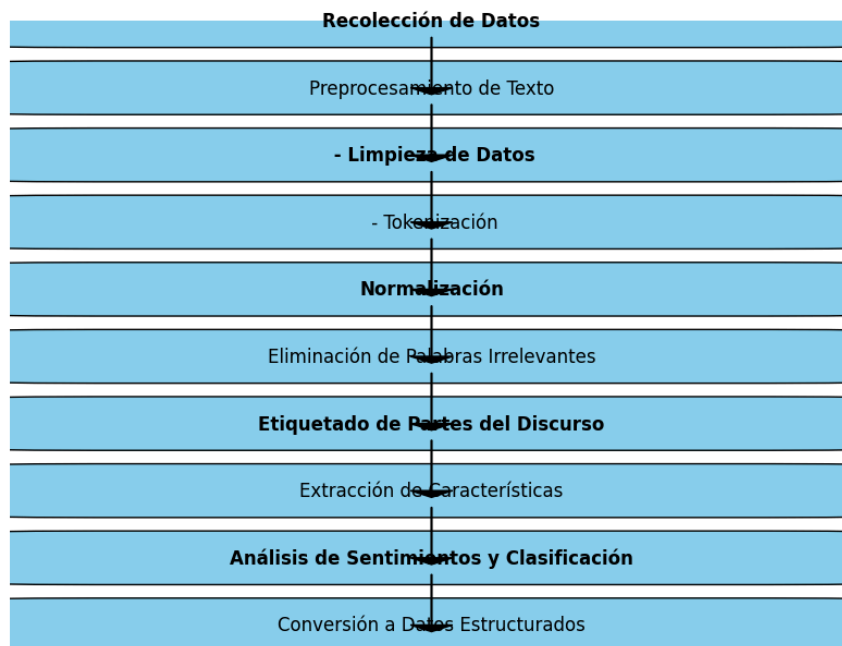
# Add rectangles for each step
for i, (step, y) in enumerate(zip(steps, y_positions)):
    ax.add_patch(patches.FancyBboxPatch((0.1, y-0.5), 0.8, 1, boxstyle="round,pad=0.3", edgecolor='black', facecolor='skybl
    ax.text(0.5, y, step, ha='center', va='center', fontsize=12, weight='bold' if i%2==0 else 'normal'))

# Add arrows
for i in range(len(steps)-1):
    ax.arrow(0.5, y_positions[i] - 0.5, 0, -1.5, head_width=0.05, head_length=0.2, fc='black', ec='black')

# Set limits and hide axes
ax.set_xlim(0, 1)
ax.set_ylim(0, len(steps) * 2)
ax.axis('off')

# Show the plot
plt.show()

```



✓ instalar paquetes

```
!pip install category_encoders
```

[Show hidden output](#)

Cargar Variables

```
import pandas as pd
import numpy as np
import category_encoders as ce
```

2. La gerencia de marketing le ha entregado una tabla con diferentes comentarios de los clientes, la cual se llama "comentarios.csv".
Primeramente, realice una lectura de los datos. Para ello, use el método `read_csv` en vez de `read_table`; utilice como separador la coma (.).
Indique cuántos registros tiene la tabla y cuántas columnas; visualice los 20 primeros registros.

```
df = pd.read_csv("/content/comentarios.csv", sep=",")
df.info()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 600 entries, 0 to 599
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   orden        600 non-null    int64
1   comentario    600 non-null    object
2   tipo         600 non-null    object
dtypes: int64(1), object(2)
memory usage: 14.2+ KB
```

✓ Analisis de la carga de datos

se obtuvo un dataframe con 600 entradas y 3 columnas, con porcentajes de tipo de comentarios clasificados como bueno, malo, info

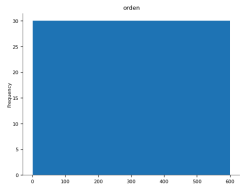
```
display(df)
```

	orden	comentario	tipo	
	0	1	No les creo nada ,	malo
	1	2	Está bonita la camioneta	bueno
	2	3	Quiero más información	info
	3	4	:(igual el crédito ha de ser muuuucho	malo
	4	5	Quisiera, más información, pero nadie me respo...	info

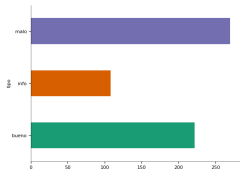
	595	596	piezas de respuestos escasas y demoran mucho e...	malo
	596	597	No sube muy bien las pendientes	malo
	597	598	es la mejor SUV que he tenido	bueno
	598	599	el crédito directo resulta muy caro	malo
	599	600	la recomiendo a ojos cerrados	bueno

600 rows x 3 columns

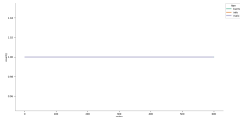
Distributions



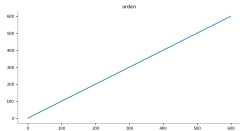
Categorical distributions



Time series



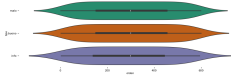
Values



Faceted distributions

<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed i









Next steps:

[Generate code with df](#)

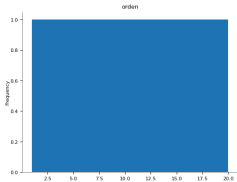
☒ [View recommended plots](#)

df.head(20)



	orden	comentario	tipo
			
			
0	1	No les creo nada ,	malo
1	2	Está bonita la camioneta	bueno
2	3	Quiero más información	info
3	4	:(igual el crédito ha de ser muuuucho	malo
4	5	Quisiera, más información, pero nadie me respo...	info
5	6	cuota chica, pero digan la última cuota, de se...	malo
6	7	Yo tengo uno claro la versión anterior „me ha...	bueno
7	8	No lo compres, son fabricados en méxico	malo
8	9	   , amo esta marca	bueno
9	10	más IVA? Ni que la weá la uses como camioneta	malo
10	11	Cuotas de \$174.900 + pie del 49,2% (7 millones...	malo
11	12	publicidad engañosa	malo
12	13	es mi tesoro, yo tengo uno, y es lo mejor	bueno
13	14	Dan poca información, necesito mayor info	info
14	15	info por fa	info
15	16	Los respuestos son demasiados caros	malo
16	17	excelente!	bueno
17	18	Es el mejor auto que he tenido	bueno
18	19	se daña a la primera, y los respuestos por las...	malo
19	20	más detalle del valor	info

Distributions



Categorical distributions

