

# Análisis de datos con IA - Mariana Jiménez Martínez

La IA que utilice fue **Chat GPT (la versión gratuita)**

Este es el **prompt** que utilice:

The screenshot shows a Chat GPT interface with a black border. At the top, it asks, "¿Con qué puedo ayudarte?" Below this is a message card with a red icon and the text "Insights from Data and AI L2 - Datos PDF". The main area contains the following text:  
Analiza los datos de este archivo PDF, y responde lo siguiente:  
¿Puedes resumir este conjunto de datos de pizza?  
¿Qué tendencias observas?  
¿Cuánta pizza deberíamos pedir basándonos en estos datos?  
At the bottom right of the message area are two small icons: a downward arrow and an upward arrow.

¿Cómo interpretó la IA los datos del pdf?

- Del texto que envió, dividió las respuestas en tres secciones/partes, y citó las fuentes que usó, que fue el PDF.

Respuesta a la pregunta: **¿Puedes resumir este conjunto de datos de pizza?**

## 1) Resumen del conjunto de datos

El archivo contiene pares de datos:

(Número de rebanadas que una persona puede comer, Tipo de pizza favorita).

El conjunto incluye decenas de observaciones, con valores de consumo entre 0 y 12 rebanadas y tipos de pizza que se repiten:

Cheese, Pepperoni, Mushroom, Veggie/Veggi, Olives, Sausage, I don't like pizza.

Hallazgos generales:

- Pepperoni aparece con mucha más frecuencia que cualquier otro tipo.
- Cheese es el segundo tipo más común.
- Hay varias menciones a "I don't like pizza" con consumo 0–8 rebanadas, aunque "0" es lo más frecuente.
- Se observan valores altos (10–12 rebanadas) asociados especialmente con Pepperoni, Cheese, Olives, y Veggi.

La respuesta de la IA fue que, el archivo se divide en dos datos principales, el número de pizza que una persona puede comer y el tipo de pizza favorita.

- Lo cuál es correcto, se divide en Number of Slices I Can Eat y Favorite Kind of Pizza.

También dice que el rango es de 12 rebanadas de pizza. Específicamente dice que los valores de consumo son entre 0 y 12 rebanadas.

- También es correcto el mayor número en Number of Slices I Can Eat es 12 y el menor es 0.

Los tipos de pizza que menciona que se repiten son Cheese, Pepperoni, Mushroom, Veggie/Veggi, Olives, Sausage, I don't like pizza.

- Esto es incorrecto, aunque unió la categoría Veggi con Veggie. Busqué en internet, y parece ser lo mismo, en el documento no dan contexto si es un error tipográfico o si es una distinta categoría. Si es la misma categoría y solo es un error tipográfico está en lo correcto la IA.

→ Hice una tabla en excel para la frecuencia en la que aparece cada tipo de pizza en Favorite Kind of Pizza, en cada página de la 1 a la 17, para comparar si es correcto lo que respondió la IA.

# de la página	# de veces que se repitió "Cheese"	# de veces que se repitió "Pepperoni"	# de veces que se repitió "Mushroom"	# de veces que se repitió "I don't like pizza"	# de veces que se repitió "Veggi"	# de veces que se repitió "Veggie"	# de veces que se repitió "Olives"	# de veces que se repitió "Sausage"	Total
1	8	11	4	5	1	1	1	0	31
2	7	11	4	4	2	1	1	1	31
3	7	11	3	5	2	1	1	1	31
4	7	11	4	4	2	1	1	1	31
5	7	10	4	5	2	1	1	1	31
6	8	12	3	3	2	1	1	1	31
7	7	13	5	4	2	0	0	0	31
8	8	11	4	5	1	1	1	0	31
9	7	11	4	4	2	1	1	1	31
10	7	11	3	5	2	1	1	1	31
11	6	11	4	5	2	1	1	1	31
12	8	10	4	4	2	1	1	1	31
13	7	13	3	3	2	1	1	1	31
14	7	12	4	4	2	1	1	0	31
15	6	11	4	5	2	1	1	1	31
16	8	10	4	5	1	1	1	1	31
17	4	7	2	2	2	1	1	1	20
TOTAL	119	186	63	72	31	16	16	13	516

Dice que el "Pepperoni aparece con mucha más frecuencia que cualquier otro tipo."

- Es correcto se repite 186 veces "Pepperoni" en la categoría Favorite Kind of Pizza.

Además dice que el "Cheese" es el segundo más frecuente.

- Esto es correcto se repite 119 veces siendo el segundo más frecuente.

En la parte de "I don't like pizza", respondió que hay un consumo de entre 0 a 8 rebanadas.

- Eso es incorrecto, el valor mínimo es 0 y el más alto es 12, se repitieron 2 veces "I don't like pizza" con un valor de 12. Pero está correcto en que se repite más es 0.
  - En esta parte no debería de haber un valor de más de 0 en I don't like pizza, ya que no es un tipo de pizza, por lo que hay un error en los datos, pero aún así se equivocó al decir que hay un valor de entre 0 a 8.

Respondió que se observan valores altos de entre 10 a 12, en Pepperoni, Cheese, Olives y Veggi.

- Está incorrecto, entre los valores de 10 a 12 que son los más altos, hay Olives, Pepperoni, Mushroom, Cheese, I don't like pizza (aunque no es un tipo de pizza) y Veggi. Omitió al menos una categoría.
  - Observación: Aquí puedo observar que puso individualmente la categoría Veggi sola en vez de combinarla con Veggie. Lo cual pienso que contradice los datos que me dió.

Respuesta a la pregunta: ¿Qué tendencias observas?

## 2) Tendencias observadas

### Preferencias

- Pepperoni domina claramente como tipo favorito.
- Cheese, Veggie/Veggi y Olives aparecen regularmente pero menos.
- Un porcentaje notable declara "I don't like pizza".

### Consumo

- Los valores más comunes de consumo parecen oscilar entre 2 y 5 rebanadas.
- Hay grupos que comen 10–12 rebanadas, pero son minoría.
- Los que dicen "I don't like pizza" generalmente declaran 0 rebanadas, aunque existen valores mayores (lo que sugiere datos ruidosos o interpretaciones distintas).

### Relación tipo–cantidad

- Las cantidades más altas (10–12) se asocian repetidamente con Pepperoni, Olives, Cheese y Veggi.
- Tipos como Mushroom y Sausage aparecen, pero en volúmenes pequeños.



Del análisis de tendencias, respondió que hay tres tendencias: las preferencias, el consumo y la relación tipo-cantidad.

Respondé que el pepperoni domina claramente como tipo favorito

- Es correcto anteriormente mencioné que se repitió 180 veces

La IA respondió que Cheese, Veggie/Veggi y Olives aparecen regularmente pero menos.

- Esto es incorrecto, anteriormente me dijo que el segundo más frecuente era Cheese. Lo cuál nos dice que inventó o alucinó estos datos. Aunque Veggie, Veggi y Olives aparecen menos frecuentemente. Los menos repetidos como favoritos son: Sausage, Olives, Veggie y Veggi.

Respondió que un porcentaje notable declara que no le gusta la pizza.

- Aquí respondió correctamente, aproximadamente al 14% no le gusta la pizza.
- Para estás respuestas hice una tabla en la cuál se muestra la frecuencia en la que se repite cada número por página, de la 1 a la 17.

PÁGINA	# que se repite													TOTAL
	# --- 0	# --- 1	# --- 2	# --- 3	# --- 4	# --- 5	# --- 6	# --- 7	# --- 8	# --- 9	# --- 10	# --- 11	# --- 12	
1	2	2	8	7	1	4	0	0	2	1	2	1	1	31
2	1	4	8	5	2	5	1	1	1	0	0	1	2	31
3	3	5	4	5	3	7	1	1	0	0	1	1	0	31
4	2	7	4	5	4	4	2	1	1	1	0	0	0	31
5	1	5	8	9	1	2	1	1	1	1	0	1	0	31
6	0	3	5	10	6	4	0	1	0	1	0	0	1	31
7	2	2	5	10	1	4	0	1	2	1	0	1	2	31
8	1	1	9	6	1	7	0	1	1	0	2	2	0	31
9	1	8	6	3	3	6	2	0	0	0	0	0	2	31
10	5	4	3	6	3	3	2	2	1	0	1	1	0	31
11	1	7	5	7	4	4	1	0	0	2	0	0	0	31
12	0	5	7	11	2	2	0	1	1	0	0	1	1	31
13	1	1	4	10	5	5	0	2	0	1	0	1	1	31
14	1	1	8	7	1	5	0	1	2	1	2	2	0	31
15	1	5	8	4	2	7	1	0	1	0	0	0	2	31
16	4	5	3	6	4	5	1	1	0	0	1	1	0	31
17	1	5	2	3	2	2	2	1	1	0	0	0	0	20
TOTAL:	27	70	97	114	45	76	14	15	14	10	9	13	12	516

En la sección consumo:

Dice que los valores más comunes son entre 2 y el 5.

- Está parcialmente correcto, los valores más repetidos son: el 3, 2, 5, y el 1. El 4 no entra entre estos.

La IA respondió que los valores 10 y 12, son minoría.

- Eso es correcto, son los valores que menos se repiten.

Responde que los que dicen “I don't like pizza” generalmente tienen valor de 0, pero que en algunos casos el valor supera 0 y eso es un dato incorrecto.

- Está correcto, la mayoría de “I don't like pizza” tienen un valor de 0. Y tiene razón en decir que está erróneo si supera 0.

En Relación Tipo-Cantidad la IA responde: “Las cantidades más altas (10-12) se asocian repetidamente con Pepperoni, Olives, Cheese, y Veggi”

- Anteriormente explique que le falta una categoría que es la de Mushroom (y “I don't like pizza” aunque no sea una categoría y sea dato erróneo)

También dice que “Tipos como Mushroom y Sausage aparecen, pero en volúmenes pequeños”

- Lo cuál es otra vez incorrecto, Mushroom está en 4to lugar y Sausage está en último lugar en relación del tipo de pizza con la cantidad que se repite. Los que menos se repiten son Sausage, Olives y Veggie.

Respuesta a la pregunta: ¿Cuánta pizza deberíamos pedir basándonos en estos datos?

### 3) ¿Cuánta pizza deberíamos pedir?

Para estimar, necesitamos un consumo promedio. Dado que el archivo contiene demasiadas líneas y no está estructurado, usaré una aproximación basada en patrones observados en el contenido:

#### Estimación razonada

- La mayoría de los valores de consumo están entre 2 y 5.
- Distribución visual aproximada del archivo:
  - ~20–25% de valores  $\geq 8$  rebanadas
  - ~60% de valores entre 2 y 5
  - El resto entre 0–1

Aproximación del promedio ponderado:

- (0–1 rebanada)  $\rightarrow \sim 15\%$
- (2–5 rebanadas, promedio  $\sim 3.5$ )  $\rightarrow \sim 60\%$
- (8–12 rebanadas, promedio  $\sim 10$ )  $\rightarrow \sim 25\%$

Cálculo rápido:

$$(0.15 \times 1) + (0.60 \times 3.5) + (0.25 \times 10)$$

$$\approx 0.15 + 2.1 + 2.5$$

$$\approx 4.75 \text{ rebanadas por persona}$$



Aquí la IA dice que tiene demasiadas líneas y que va a encontrar un patrón, lo cuál es un problema al hacer una estimación de la pizza que deberíamos pedir, dice que va a usar una “aproximación basada en patrones observados” en lugar de leer cada línea y hacer un cálculo exacto.

## Recomendación

≈ 5 rebanadas por persona.

Equivale a 2 personas por pizza de 10 rebanadas, o 1.5 personas por pizza estándar de 8 rebanadas.

Si el grupo es grande:

- 1 pizza por cada 1.5 personas es una regla simple y segura.

Si deseas, puedo:

- Calcular un promedio exacto extrayendo cada número del PDF y procesándolo.
- Construir una tabla limpia de todo el dataset.
- Generar gráficos de distribución.

Fuentes

**Respuesta de la IA:** Repite que el consumo está entre 2 y 5, distribuyendo aproximadamente:

Entre 20% y 25% de valores son mayores que 8 rebanadas

El 60% de valores entre 2 y 5

El resto entre 0 y 1. (entre 15% y 20%)

La IA hace la aproximación del promedio ponderado:

(0-1 rebanada) - 15%

(2-5 rebanadas, promedio 3.5) - 60%

(8-12 rebanadas, promedio 10) - 25%

Cálculo rápido:  $(0.15 \times 1) + (0.60 \times 3.5) + (0.25 \times 10) = 0.15 + 2.1 + 2.5 = 4.75$  rebanadas por persona.

Y responde redondeando que serían un promedio de 5 rebanadas por persona.

Por lo que el cálculo de la IA está incorrecto debido a que no leyó línea por línea, solo sacó un promedio, de lo que aproximo por los patrones.

Haciendo los cálculos manualmente, sería algo así:

$(0 \times 27) + (1 \times 70) + (2 \times 97) + (3 \times 114) + (4 \times 45) + (5 \times 76) + (6 \times 14) + (7 \times 15) + (8 \times 14) + (9 \times 10) + (10 \times 9) + (11 \times 13) + (12 \times 12) =$

$70 + 194 + 342 + 180 + 380 + 84 + 105 + 112 + 90 + 90 + 143 + 144 = \text{TOTAL ES } 1934.$

Necesitaría 1934 rebanadas de pizza. 241.75 pizzas normales de 8 rebanadas se redondea a 242

Entre 516 casos, aquí pienso que es el número de personas

$1934 / 516 = 3.74806202$

Es decir en promedio necesitaría 3.74 rebanadas de pizza por persona

**CONCLUSIÓN:** Llegué a una conclusión, la IA es una excelente herramienta, pero para analizar datos no es lo ideal (al menos específicamente IAs en general, tal vez sea diferente con una IA específicamente entrenada en ello), llega a no leer todos los datos que hay y solo hacer una estimación. Salta categorías y datos, llega a inventar datos en algunas ocasiones, y en general no lee todo el documento. Pienso que la IA, no llegará a reemplazar a los analistas de datos y su trabajo de depuración de datos, ya que es muy complejo el proceso. Pero la IA puede ayudar a proporcionar datos, que ayuden al proceso de analizar datos.