

Análisis de datos con IA

Tarea 2: Introducción a Ciencia de Datos

Ana Valeria Deloya Andrade

Para esta tarea, fue seleccionado ChatGPT como la Inteligencia Artificial a la que se le pedirá analizar los datos de la fiesta de pizza.

A continuación se muestra la respuesta generada a cada una de las 3 preguntas que se le hicieron junto con una opinión mía al final sobre el análisis.

- **¿Puedes resumir este conjunto de datos de pizza?**

ChatGPT Mejora tu plan a Go! × Memoria llena ↑ Compartir Añadir personas ...

Sí 😊 Aquí tienes un resumen claro y general del conjunto de datos de pizza que compartiste:

📊 Descripción general del conjunto de datos

El conjunto de datos recoge respuestas de varias personas a dos preguntas principales:

1. ¿Cuántas rebanadas de pizza pueden comer?
2. ¿Cuál es su tipo de pizza favorito?

Cada registro contiene:

- Un número entero (de 0 a 12, aproximadamente) que indica cuántas rebanadas puede comer la persona.
- Una categoría de pizza favorita (por ejemplo: Pepperoni, Cheese, Mushroom, Veggie, Olives, Sausage o "I don't like pizza").

Insights from Data and AI L2 - ...

🍕 Preferencias de pizza

- Pepperoni es, por mucho, el tipo de pizza más frecuente en el conjunto de datos.
- Cheese aparece como la segunda opción más popular.
- Mushroom y Veggie/Veggi tienen una presencia moderada.
- Olives y Sausage aparecen con menor frecuencia.
- Un grupo notable de personas respondió "I don't like pizza", lo cual es interesante porque aun así reportan cuántas rebanadas comerían (a veces 0, pero no siempre).

📊 Número de rebanadas

- El rango de rebanadas va principalmente de **0 a 12**.
- Los valores más comunes se concentran entre **2 y 5 rebanadas**, lo que sugiere un consumo típico moderado.
- Hay casos extremos:
 - **0 rebanadas**, casi siempre asociados a "I don't like pizza".
 - **10-12 rebanadas**, que indican consumidores altos.
- No hay una relación estricta entre "no me gusta la pizza" y comer 0 rebanadas; algunas personas dicen no gustarles pero reportan más de 0.

🔍 Observaciones interesantes

- El conjunto de datos parece **repetir bloques de respuestas**, lo que sugiere que puede ser un dataset educativo o sintético.
- Hay **inconsistencias ortográficas** ("Veggie" vs "Veggi") que serían importantes de normalizar antes de hacer análisis estadísticos.
- Es un buen dataset para practicar:
 - Conteo de frecuencias
 - Promedios y distribuciones
 - Limpieza de datos categóricos
 - Visualizaciones (barras, histogramas)

• ¿Qué tendencias observas?

En el conjunto de datos se observan varias **tendencias claras**:

💡 Preferencias de sabor

- Pepperoni domina ampliamente como la pizza favorita. Aparece mucho más que cualquier otro tipo.
- Cheese es la segunda opción más popular, lo que sugiere preferencia por sabores clásicos.
- Veggie/Veggi y Mushroom tienen una presencia constante pero menor, asociadas a preferencias más específicas.
- Olives y Sausage son claramente minoritarias.
- Un grupo no despreciable dice "I don't like pizza", lo cual es una categoría relevante dentro del conjunto.

⌚ Consumo de rebanadas

- La mayoría de las personas reporta poder comer entre **2 y 5 rebanadas**, lo que indica un consumo típico moderado.
- Valores altos (**8 a 12 rebanadas**) existen, pero son menos frecuentes y representan consumidores intensivos.
- El valor **0 rebanadas** aparece casi exclusivamente con "I don't like pizza", aunque no de forma absoluta.

🔗 Relación entre gusto y consumo

- Quienes prefieren Pepperoni o Cheese tienden a reportar **más rebanadas** en promedio.
- Las personas que dicen **no gustarles la pizza** generalmente reportan pocas rebanadas, pero hay **inconsistencias** (algunos indican más de 0), lo que sugiere:
 - Interpretación subjetiva de la pregunta, o
 - Datos simulados con fines didácticos.

📊 Calidad y estructura de los datos

- Hay **repetición de patrones**, lo que apunta a que el dataset fue diseñado para práctica y no recolectado de forma completamente natural.
- Existen **variaciones ortográficas** (Veggie/Veggi) que afectarían análisis automáticos si no se limpian.
- El rango limitado de valores (0–12) facilita análisis estadísticos básicos.

🧠 Tendencia general

En conjunto, los datos reflejan que:

- **Las preferencias tradicionales (Pepperoni y Cheese) no solo son más comunes, sino que se asocian con mayor consumo.**
- **El rechazo a la pizza es minoritario**, pero suficientemente frecuente como para ser analizado como grupo separado.

• ¿Cuánta pizza deberíamos pedir basándonos en estos datos?

Basándonos en las **tendencias del conjunto de datos**, se puede hacer una recomendación práctica.

💡 ¿Cuánta pizza pedir?

🕒 Consumo típico observado

- **La mayoría de las personas come entre 2 y 5 rebanadas.**
- El valor más representativo (promedio aproximado) está alrededor de **3–4 rebanadas por persona**.
- Hay personas que comen mucho más, pero son minoría.

Recomendación general

 Pidan entre 3 y 4 rebanadas por persona como regla base.

Como una pizza estándar suele tener 8 rebanadas, esto se traduce en:

- 1 pizza para cada 2 personas (si quieren ir seguros).
- 1 pizza para cada 3 personas (si el grupo come moderadamente).

Ajustes según el grupo

- Si sabes que la mayoría ama la pizza (Pepperoni / Cheese) → inclínate a 4 rebanadas por persona.
- Si es un grupo mixto o hay personas que comen poco → 3 rebanadas por persona.
- Si hay personas que dicen "no me gusta la pizza" → considera un pequeño margen extra, porque el dataset muestra que aun así algunos comen.

Ejemplo rápido

- $10 \text{ personas} \times 3.5 \text{ rebanadas} \approx 35 \text{ rebanadas}$
- $35 \div 8 \approx 4.5 \text{ pizzas} \rightarrow \text{pide 5 pizzas}$

Conclusión

La decisión óptima, basada en los datos, es pedir un poco más de lo necesario, porque el consumo real suele concentrarse en valores medios y quedarse corto es más común que desperdiciar.

● Observaciones, comentarios y opinión de los resultados

En mi experiencia los chatbots en ocasiones tienden a inventar información, por lo que en este caso pensé que podría hacerlo pues le adjunté el pdf con los datos para poder hacerle las preguntas.

Para la primer pregunta enlistó correctamente las categorías o tipos de pizza (Pepperoni, Cheese, Mushroom, Veggie, Olives, Sausage, "I don't like pizza") y el rango de rebanadas (0 a 12). Además de agregar la observación de que una persona que haya opinado "I don't like pizza" puede comer más de 0 rebanadas, igual fue de lo primero que me llamó la atención al ver los datos. Me sorprendió que incluso notó la inconsistencia en la escritura del tipo favorito de pizza que en ocasiones aparece como "Veggie" y en otras como "Veggi".

De igual forma también me sorprendió que en su respuesta a la segunda pregunta incluyó el detalle de que hay personas que sí eligieron un tipo de pizza pero pusieron que comerían 0 rebanadas, siendo que no comer ninguna rebanada no es exclusivo de los "I don't like pizza". También menciona que hay una repetición de patrones, lo que quiere decir que se dio cuenta que los datos no fueron recolectados de forma natural sino que el dataset fue diseñado con fines prácticos. Desconozco si esto último sea verdad pero dado que son datos que se nos proporcionaron para realizar una tarea, pienso que sí lo es.

Finalmente para la tercer pregunta, antes de leer su respuesta pensé que daría una cantidad más elevada de pizzas a pedir considerando que hay personas que decían comer 6 o 7 rebanadas. Pues aunque la mayor cantidad de rebanadas es 12, no creo que haya alguien que se coma 12 rebanadas en una comida porque sería equivalente a comer 3 pizzas. Siendo así, que la Inteligencia Artificial sugiere que se limite el número de rebanadas por persona para pedir 5 pizzas en total.

Considero que hizo un buen análisis, no encontré que haya inventado o cambiado los datos que se le dieron a analizar, leyendo correctamente la información proporcionada. Me gustó que limitara el

número de rebanadas por persona aunque pienso que aún así debería comprarse 1 o 2 pizzas extra pues teniendo en cuenta que iría mucha gente a la fiesta puede haber un punto en que no sea posible seguir llevando un conteo correcto de cuántas rebanadas ha comido cada quien y puede existir el caso en que la gente que tome rebanadas de más intencionalmente.

Pienso que es bueno hacer uso de tecnologías como los chatbots de Inteligencia Artificial como herramientas cuando los datos que se tienen que analizar son muchos, siempre y cuando se verifique después que la respuesta generada por éstos sea precisa y correcta. Utilizando la información que proporciona como una ayuda y no como una verdad indiscutible. Me pareció interesante esta actividad.