



Proyecto Integrador  
Dra. Grettel Barceló Alonso  
Dr. Luis Eduardo Falcón Morales

## Reporte de Métricas: Visualización de Espacios de Embeddings Usando UMAP

Junio 02 2024

**Equipo 37**

A00759664 - Joel Orlando Hernández Ramos

A01793486 - Juan Carlos Alvarado Carricarte

A00260430 - Juan Carlos Romo Cárdenas

# Introducción

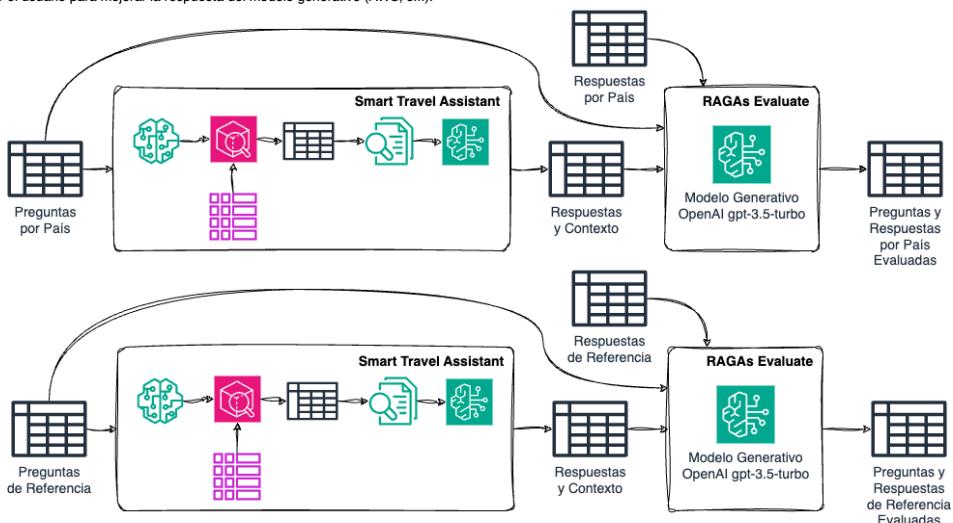
En la era digital actual, la personalización y eficiencia en la planificación de viajes se ha convertido en una prioridad para los consumidores globales. Respondiendo a esta demanda, nuestro proyecto ha desarrollado un asistente de inteligencia artificial avanzado, integrado en una innovadora plataforma de traveltech, diseñada para transformar la experiencia de selección de hoteles y destinos. Este sistema inteligente, apoyado por la robusta "knowledge base" de Amazon Bedrock, incorpora información detallada sobre más de 2 millones de hoteles.

El objetivo principal del proyecto ha sido hacer que la planificación de viajes sea más intuitiva y basada en datos confiables, mejorando significativamente la satisfacción del cliente y optimizando los tiempos de búsqueda y selección de alojamientos y destinos. Durante las fases iniciales, nuestro equipo se centró en el pre-procesamiento de datos y el desarrollo de una base de conocimientos sólida, estableciendo un modelo baseline que sirvió como punto de partida para futuras iteraciones y mejoras.

Con el propósito de alcanzar una mayor eficiencia, nuestro equipo desarrolló tres modelos alternativos al modelo baseline. Estos modelos fueron evaluados cualitativamente para determinar su eficacia y precisión en la generación de recomendaciones. El modelo que destacó por sus capacidades avanzadas, empleando técnicas de generación de texto asistida por recuperación de información (como se ilustra en la Figura 1 del sistema RAG versión 2.00 con LangChain), fue seleccionado para desarrollo y evaluación adicionales. Este modelo no solo mejoró la calidad de las recomendaciones, sino que también proporcionó una base más sólida para las interacciones intuitivas y personalizadas con los usuarios.

## Retrieval Augmented Generation (RAG) - KB V2.00 con LangChain

La técnica RAG hace uso de una base de conocimiento para extender o aumentar la pregunta, petición o sugerencia proveída por el usuario para mejorar la respuesta del modelo generativo (AWS, s.f.).



**Figura 1.** Esquema M étodológico del Modelo Final

En esta fase final del proyecto, hemos aplicado métricas cuantitativas para evaluar exhaustivamente el rendimiento de nuestro modelo seleccionado. Esta entrega detalla los resultados obtenidos mediante la visualización de embeddings en el espacio con UMAP y otras métricas esenciales, comparándolos con nuestro modelo baseline para demostrar las mejoras

logradas y la eficacia de nuestra solución. Este informe presenta los hallazgos de estas evaluaciones, subrayando cómo nuestro asistente inteligente no solo cumple con los objetivos propuestos sino que establece un nuevo estándar en la personalización y eficiencia de la planificación de viajes.

## Métricas

### Visualización de Espacios de Embeddings Usando UMAP

UMAP es una técnica avanzada de reducción de dimensionalidad que nos permite visualizar la estructura de los datos de alta dimensión en un espacio de baja dimensión. Esta herramienta es especialmente útil para interpretar cómo el modelo seleccionado organiza y comprende los vastos datos relacionados con más de 2 millones de hoteles.

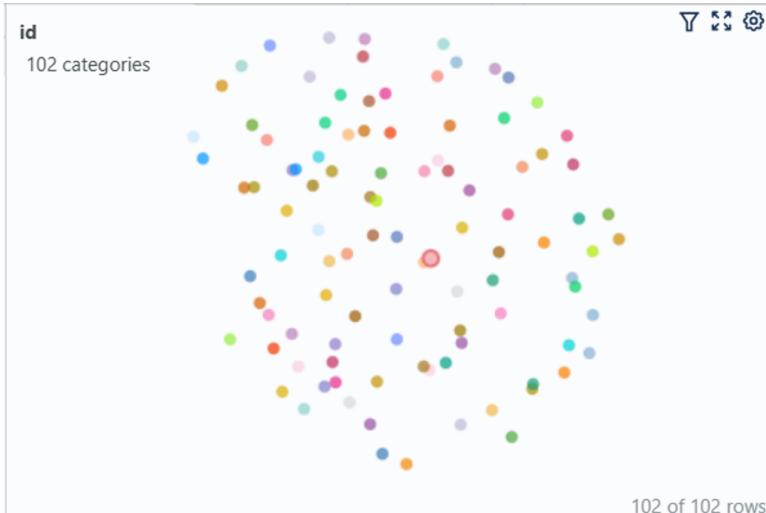
En el contexto de nuestro proyecto, UMAP nos ayuda a validar la eficacia del modelo en agrupar hoteles y destinos de manera significativa, reflejando similitudes y diferencias en sus características, lo cual es fundamental para personalizar recomendaciones. La capacidad de visualizar estos embeddings proporciona una ventana clara hacia el comportamiento del modelo, mostrando cómo interpreta y procesa las entradas para generar recomendaciones. Esta visualización no solo es una herramienta de diagnóstico para identificar áreas de mejora sino que también actúa como un método para comparar la performance del modelo seleccionado con el modelo baseline.

En este sentido, a continuación se muestran una serie de preguntas. Cada una de ellas con un análisis de los resultados obtenidos por UMAP.

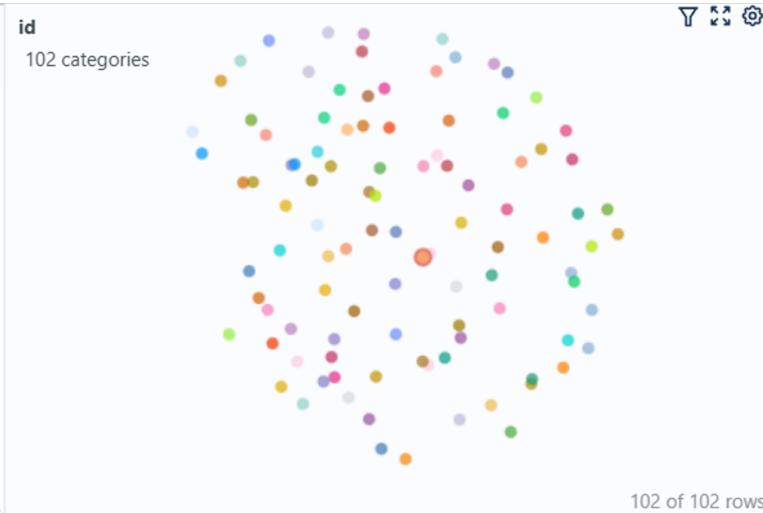
## 1. Can you suggest accommodation near Pilanesberg National Park in South Africa?

### Baseline

Pregunta



Respuesta



## Travel Assistant KB V2.00 con LangChain

Pregunta	Respuesta
<p>id 102 categories</p> 	<p>id 102 categories</p> 

102 of 102 rows

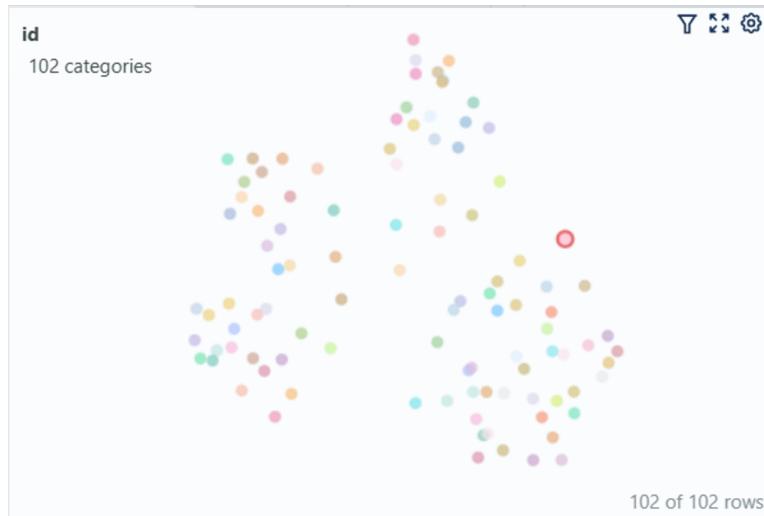
102 of 102 rows

2. Can you suggest accommodation near Pilanesberg National Park in the northwest province in South Africa?

**Baseline**



Pregunta



Respuesta



3. What attractions are close to Black Rhino Game Lodge, northwest province in South Africa?

## Baseline



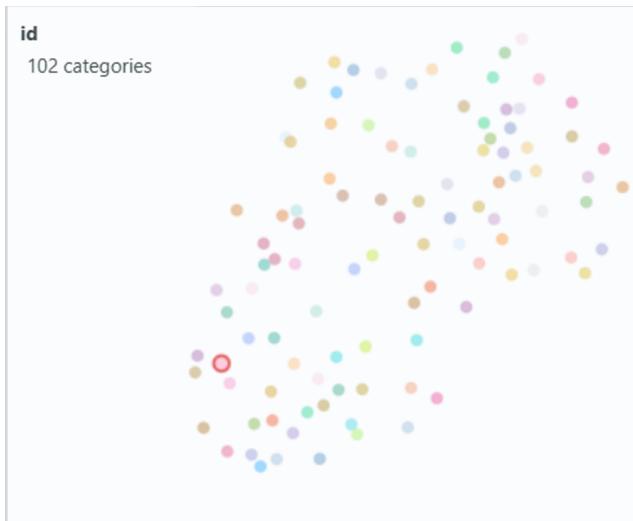
## Travel Assistant KB V2.00 con LangChain



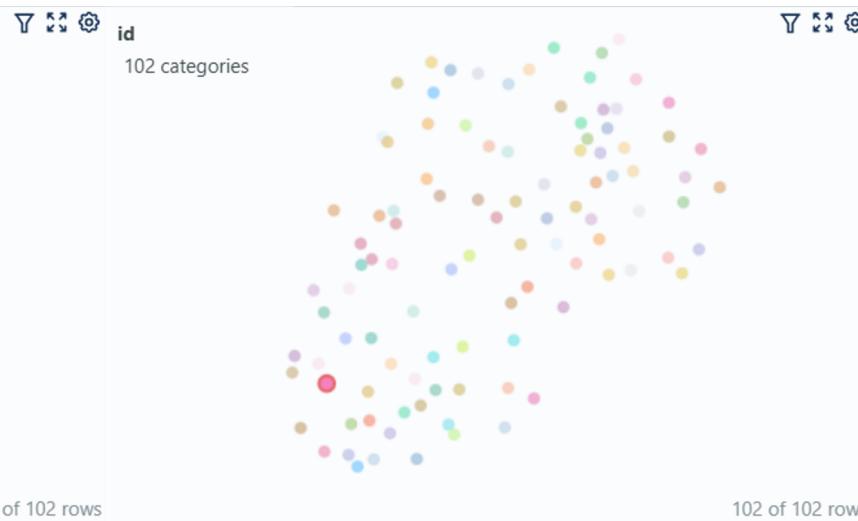
4. Can you suggest accommodation in Cape Town, Western Cape province in South Africa?

## Baseline

Pregunta



Respuesta



Travel Assistant KB V2.00 con LangChain

Pregunta



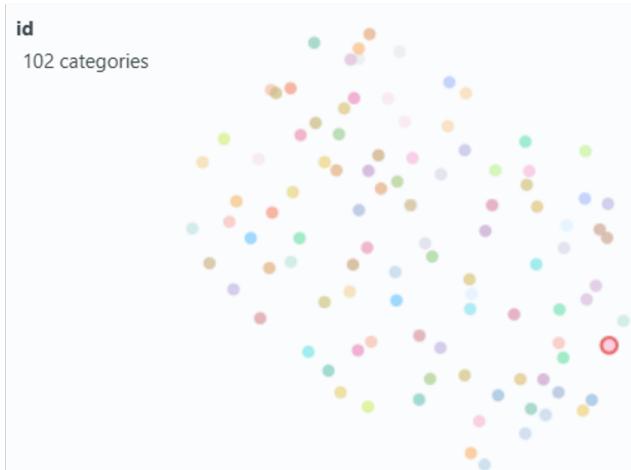
Respuesta



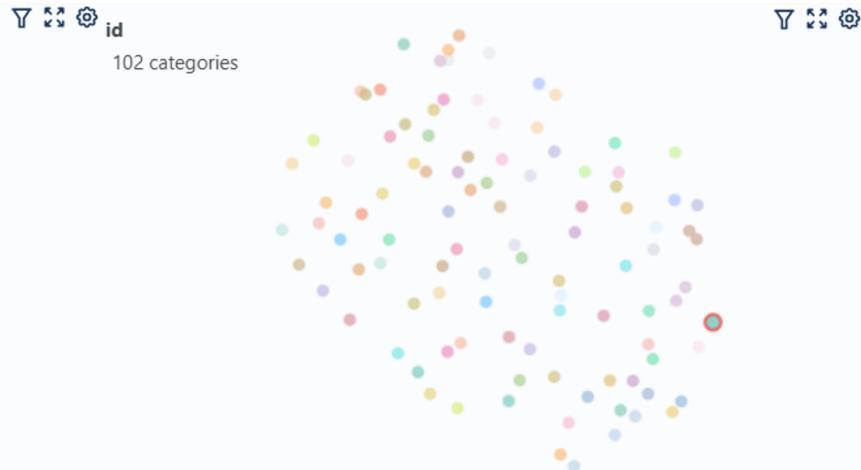
5. Can you suggest accommodation in cape town, western cape province in south africa, minimum rating of 4?

**Baseline**

Pregunta



Respuesta



102 of 102 rows

102 of 102 rows

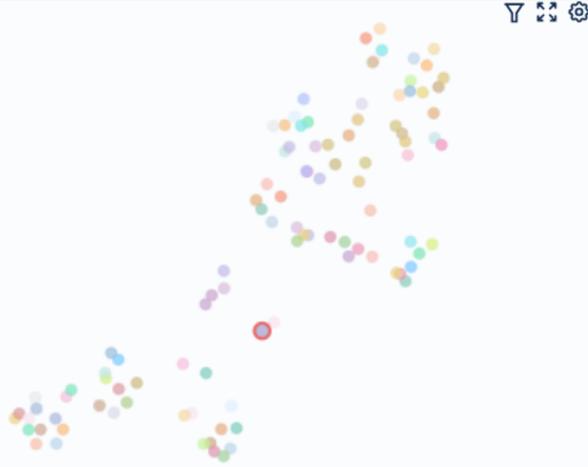
**Travel Assistant KB V2.00 con LangChain**

Pregunta



102 of 102 rows

Respuesta

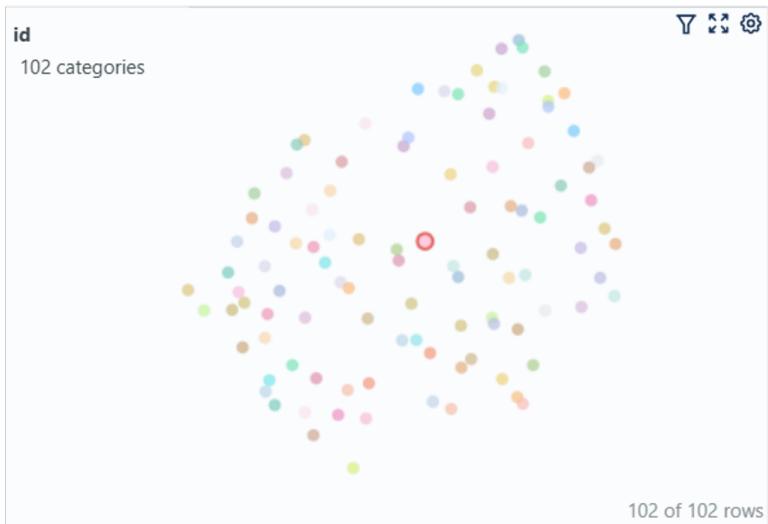


102 of 102 rows

6. Can you suggest accommodation in Hoedspruit, Limpopo province in South Africa?

## Baseline

Pregunta

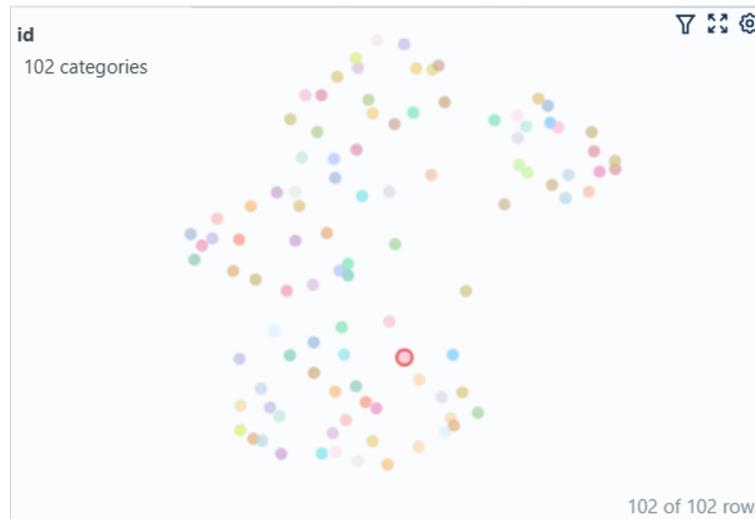


Respuesta

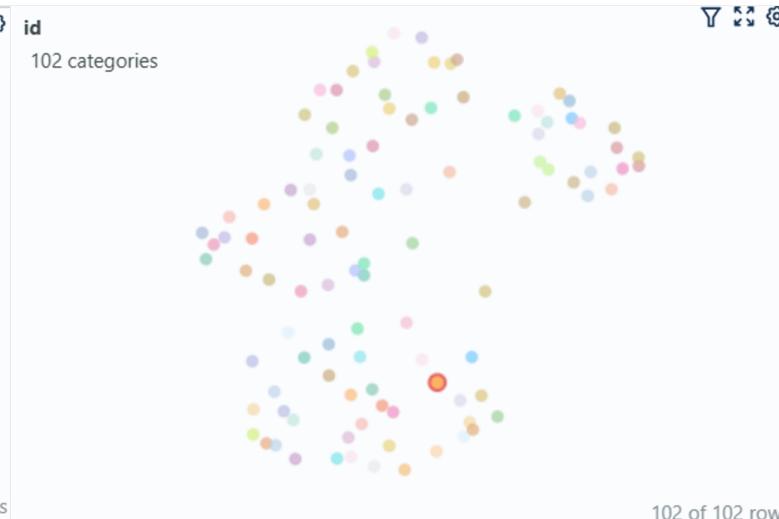


Travel Assistant KB V2.00 con LangChain

Pregunta



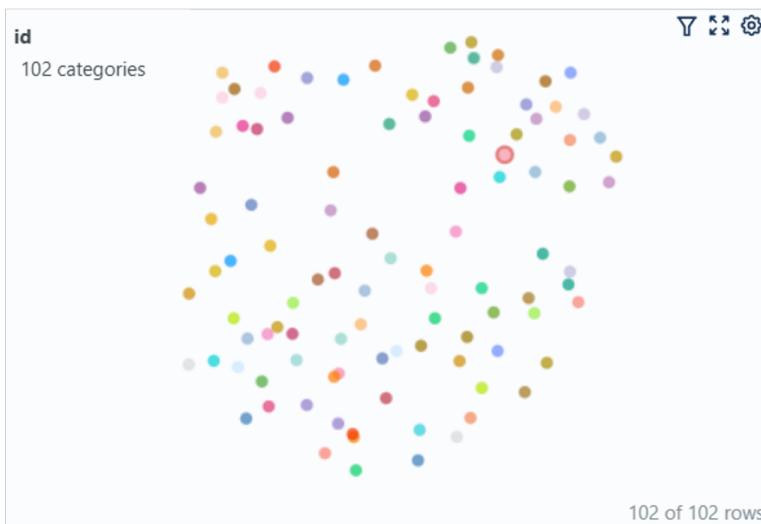
Respuesta



7. Can you suggest 5 places to stay in Monterrey, Nuevo Leon in Mexico?

## Baseline

Pregunta



Respuesta



Travel Assistant KB V2.00 con LangChain

Pregunta



id  
102 categories

Respuesta



id  
102 categories

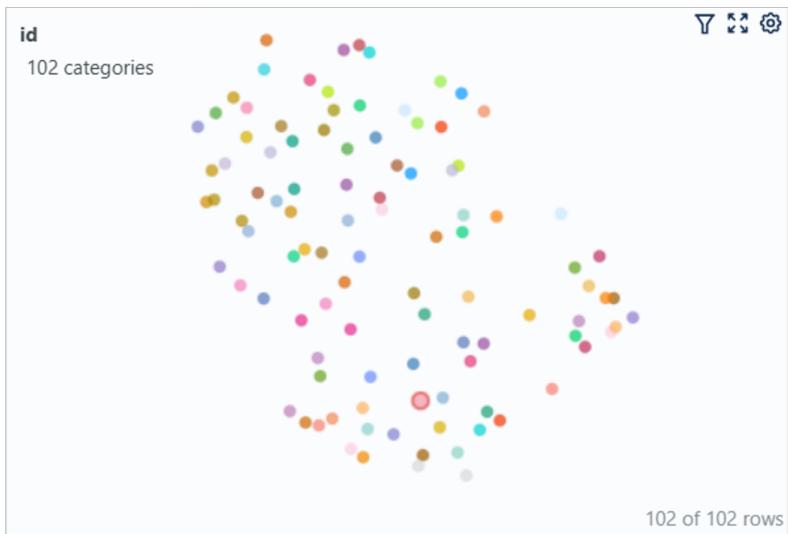
102 of 102 rows

102 of 102 rows

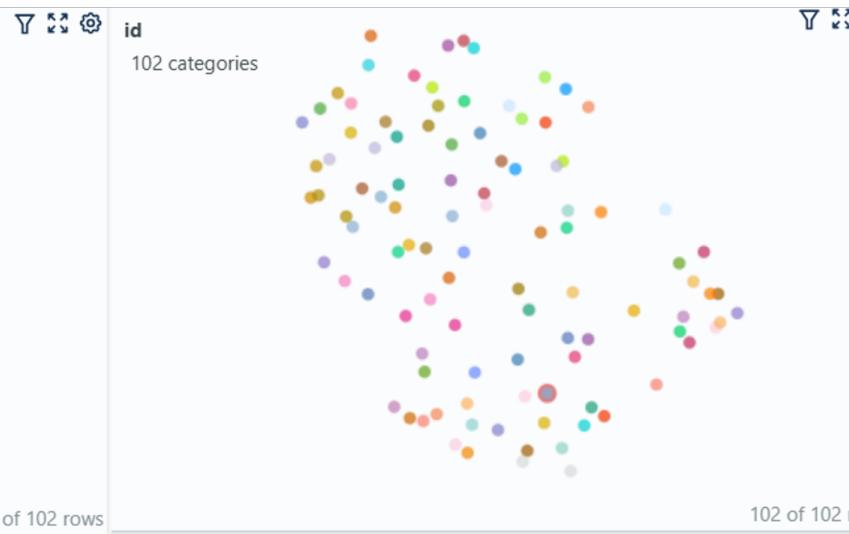
8. Can you suggest 5 places to stay in Monterrey, Nuevo Leon in Mexico, near Tec de Monterrey?

## Baseline

Pregunta

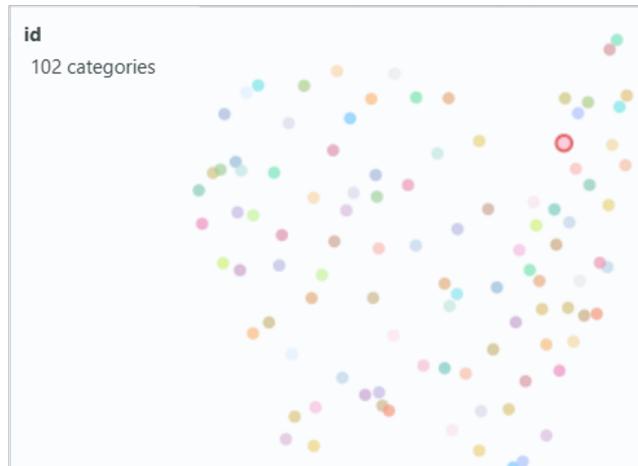


Respuesta

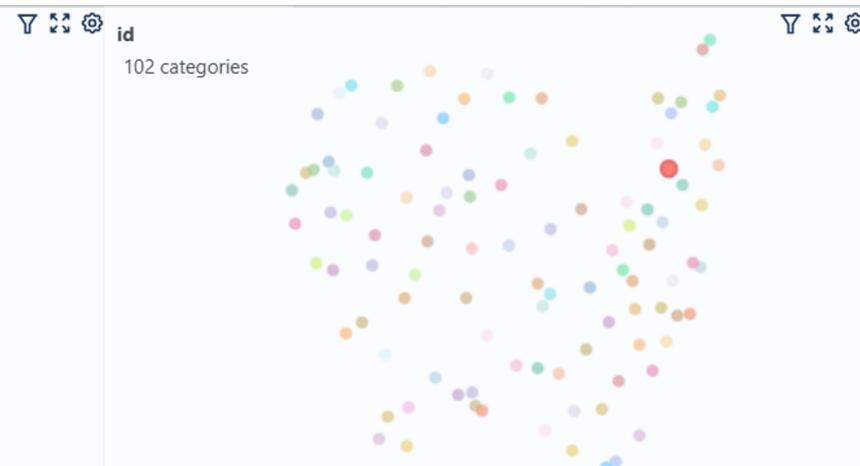


Travel Assistant KB V2.00 con LangChain

Pregunta



Respuesta

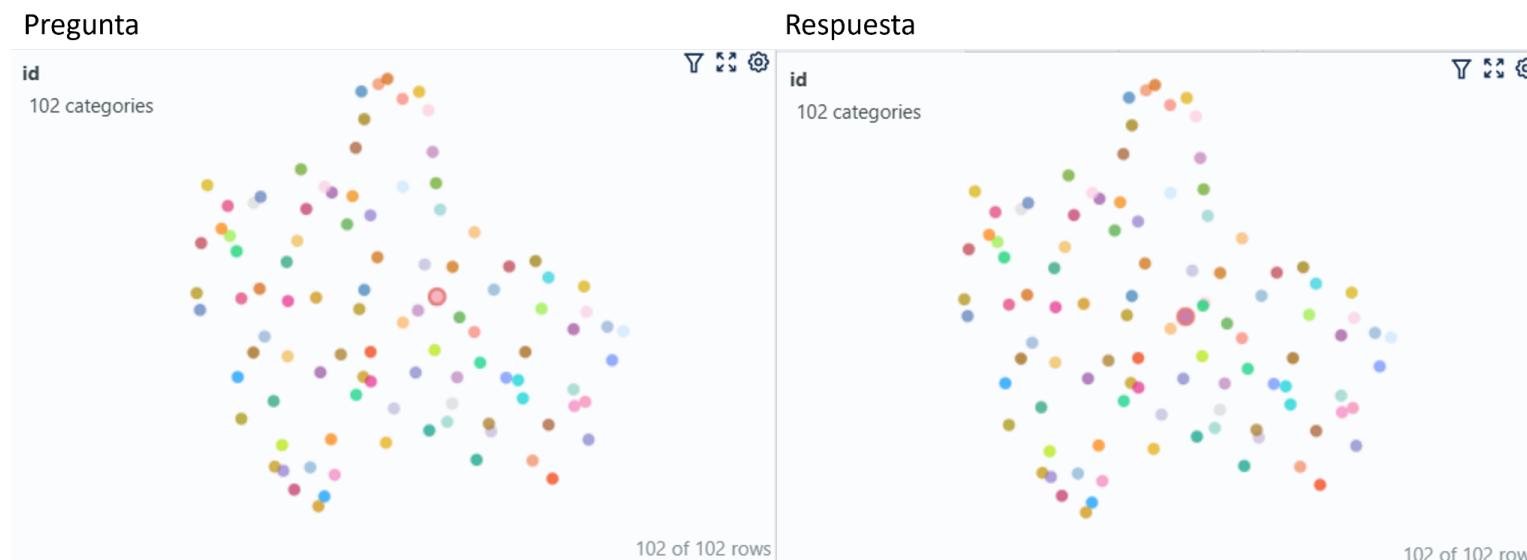


102 of 102 rows

102 of 102 rows

9. I want to stay with my wife for a romantic weekend in Mexico City. We want to have a nice massage within the hotel.

#### Baseline

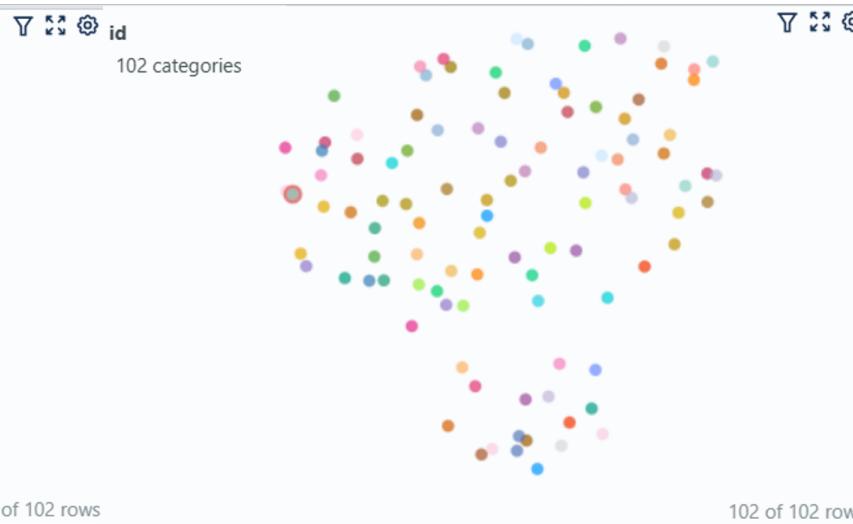


## Travel Assistant KB V2.00 con LangChain

Pregunta

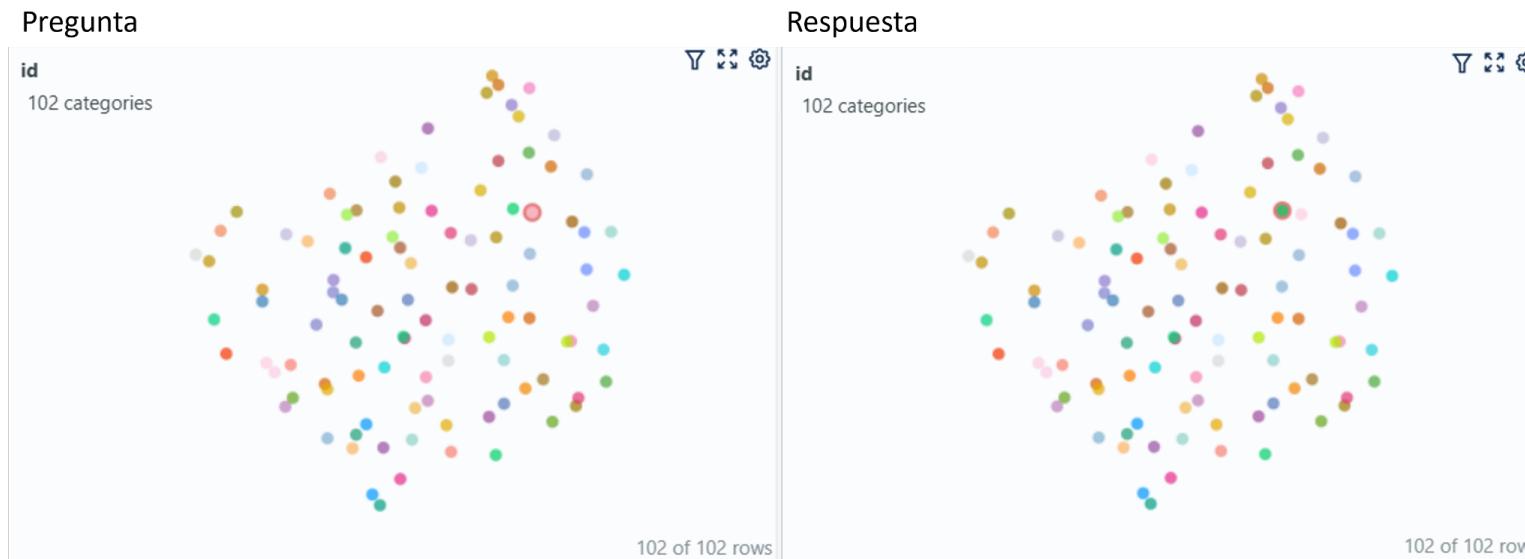


Respuesta



10. I want to have a golf weekend with my friends in Cancun. What hotels do you recommend that have their golf course?

**Baseline**



## Travel Assistant KB V2.00 con LangChain

Pregunta



Respuesta



## 11. Can you recommend hotels near Ueno Park in Tokyo?

### Baseline

Pregunta



Respuesta

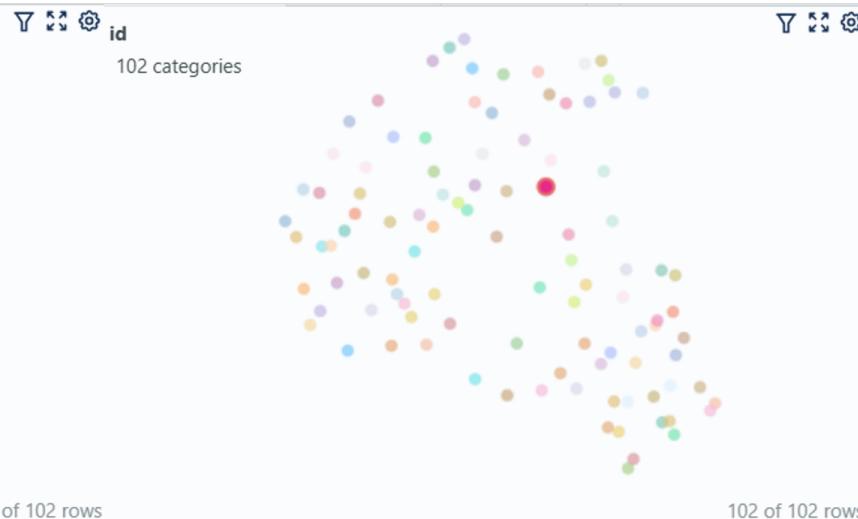


## Travel Assistant KB V2.00 con LangChain

Pregunta



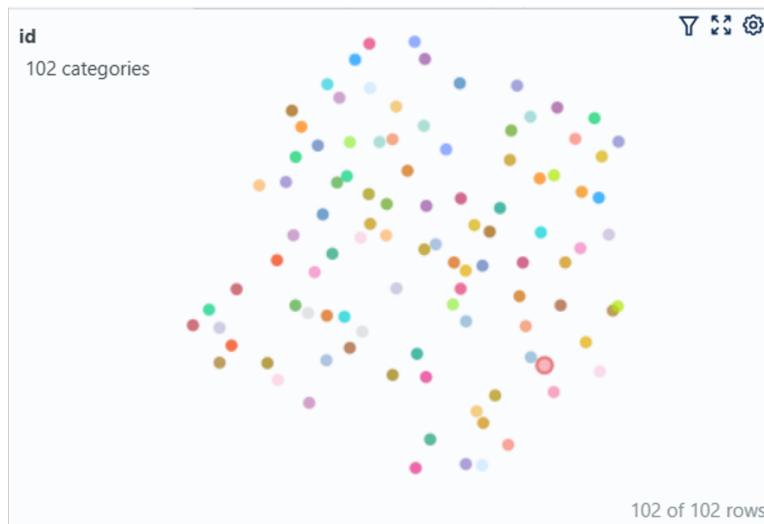
Respuesta



12. Can you recommend hotels in Hakone Japan? Ideally, I am looking for traditional Japanese ryokans.

**Baseline**

Pregunta



Respuesta

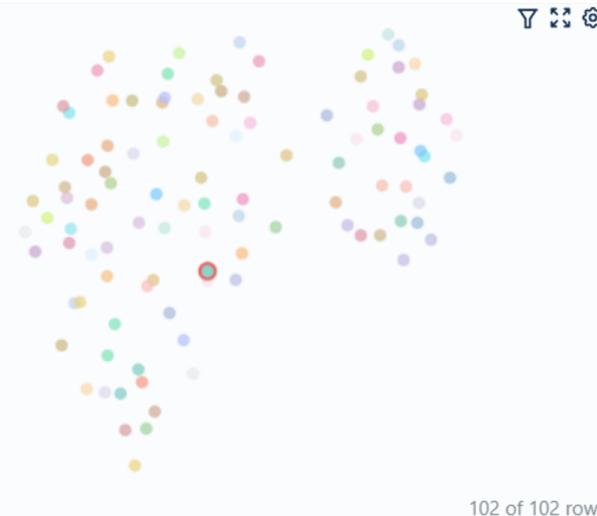


## Travel Assistant KB V2.00 con LangChain

Pregunta



Respuesta



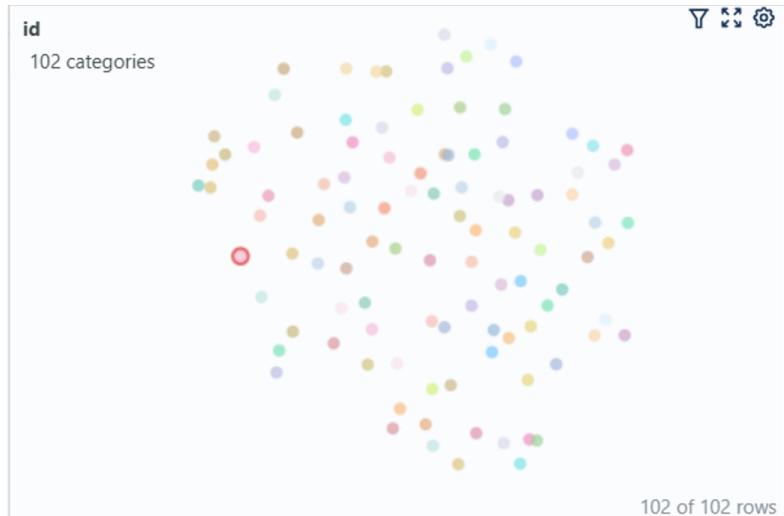
13. Can you recommend hotels in Chihuahua that are located near Periferico de la Juventud?

**Baseline**

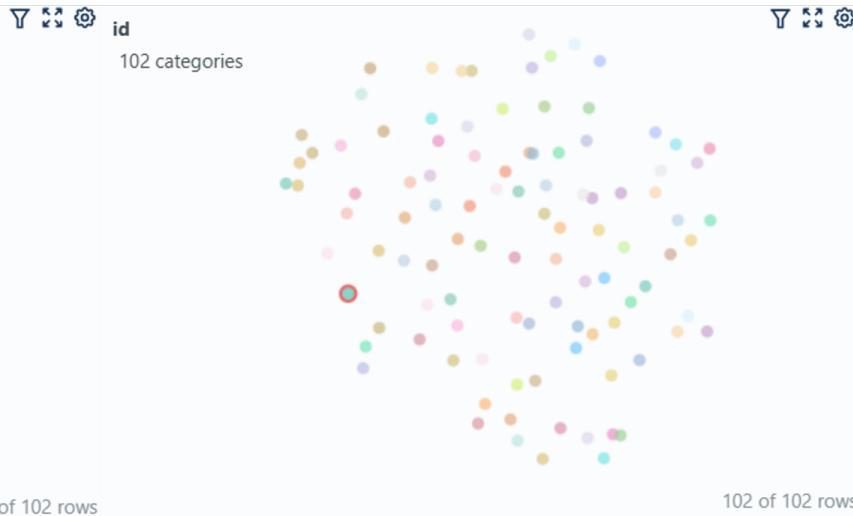


## Travel Assistant KB V2.00 con LangChain

Pregunta



Respuesta



## Observaciones:

La imagen muestra dos pares de visualizaciones de UMAP, un par por cada versión del modelo, donde cada punto representa un embedding de datos relacionado con hoteles o destinos. Es importante destacar que los gráficos presentados muestran que el modelo KB V2.00 tiende a formar clusters más definidos y separados reflejando una estructuración más clara del espacio de los datos. Esto contrasta con el modelo baseline, donde los embeddings parecen estar más dispersos y menos diferenciados, lo que podría interpretarse como una menor capacidad para diferenciar entre las consultas y los contextos.

Por ejemplo, las respuestas a la pregunta sobre alojamiento cerca del Pilanesberg National Park en Sudáfrica mostraron que el modelo KB V2.00 podía identificar y recomendar opciones que el modelo baseline no lograba distinguir adecuadamente. De igual manera, en esta imagen podemos darnos cuenta fácilmente de una marcada distribución en el espacio en forma de clusters. Esto sugiere que el modelo está efectivamente diferenciando varias categorías de datos basándose en similitudes en los embeddings de los hoteles. Cada cluster podría representar hoteles con características similares como ubicación geográfica, rango de precios, o tipo de alojamiento.

En la mayoría de los casos, en el modelo KB V2.00, es muy importante destacar que la pregunta y la respuesta, ambos puntos enfatizados por un círculo rojo en sus respectivas imágenes, se encuentran en el mismo cluster con una cercanía muy grande. Esto es significativo porque nos demuestra que la respuesta es altamente relevante para la pregunta. En otras palabras, el modelo ha identificado correctamente las características de la pregunta y ha encontrado una respuesta que coincide bien con estas características dentro de su espacio de representación.