



Tecnológico de Monterrey

Actividad 3: Segmentación

Miranda Eugenia Colorado Arróniz A01737027

22 de agosto de 2025

Analítica de datos y herramientas de inteligencia artificial I (Gpo 101)

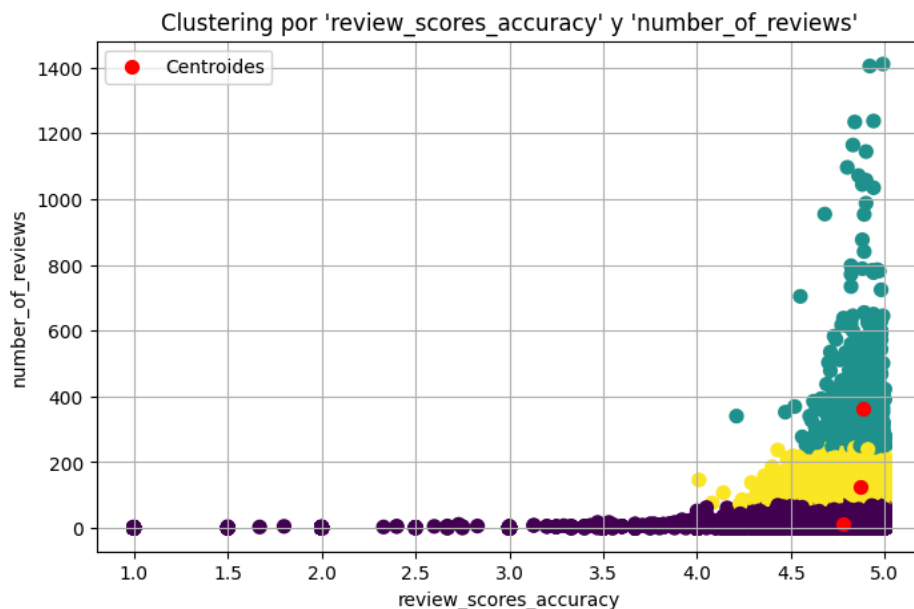
Dr. Alfredo García Suárez

Inciso a.

Obtener los siguientes clusters aplicando el algoritmo de Kmeans:

K= 3, Variables = “review_scores_accuracy” y “number_of_reviews”

Hice la segmentación con tres clusters, usando las variables “review_scores_accuracy” y “number_of_reviews”. Primero preparé el dataset, eliminando valores nulos y seleccionando las dos columnas mencionadas. Después apliqué el algoritmo de K-means, obtuve las etiquetas de los clusters y los centroides, que representan el promedio de cada grupo. Finalmente, grafiqué los resultados en una gráfica donde cada punto está coloreado según su cluster y los centroides aparecen en rojo. Los centroides muestran cómo se agrupan los datos y la gráfica evidencia una buena separación entre los clusters, lo que permite identificar fácilmente las características principales de cada grupo.



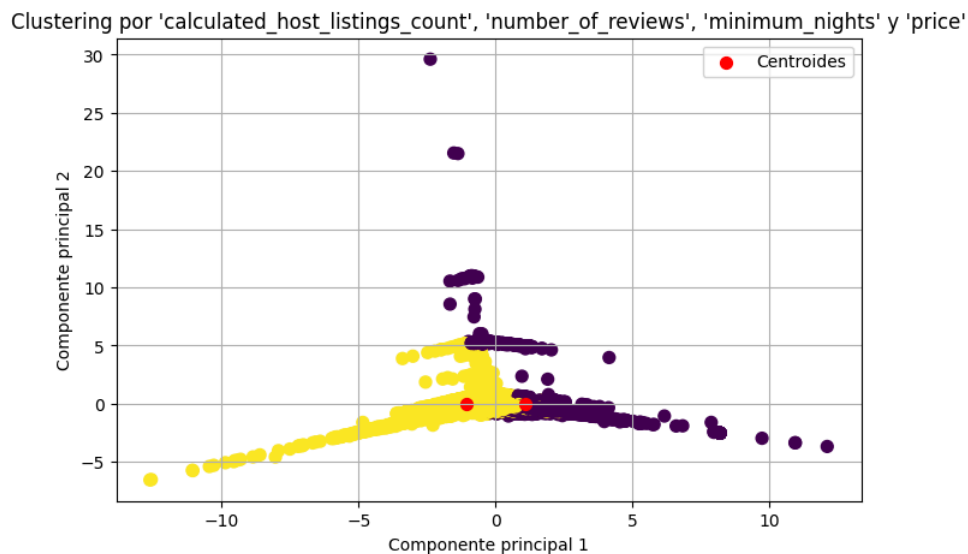
Inciso b.

Obtener los siguientes clusters aplicando el algoritmo de Kmeans:

K= 2, Variables = "calculated_host_listings_count", "number_of_reviews", "minimum_nights" y "price"

Para este inciso usé dos clusters y las variables “calculated_host_listings_count”, “number_of_reviews”, “minimum_nights” y “price”, asegurándome primero de limpiar y preparar el dataset, eliminando valores nulos y transformando la columna de precios a formato numérico. Después escalé los datos para que todas las variables tuvieran el mismo peso y apliqué el modelo de K-means, obteniendo así las etiquetas de los clusters y los centroides que representan el promedio de cada grupo en las cuatro variables seleccionadas.

En la gráfica hay una clara separación entre los dos grupos, lo que indica que el modelo logró segmentar adecuadamente los datos y permite identificar las características principales de cada cluster.

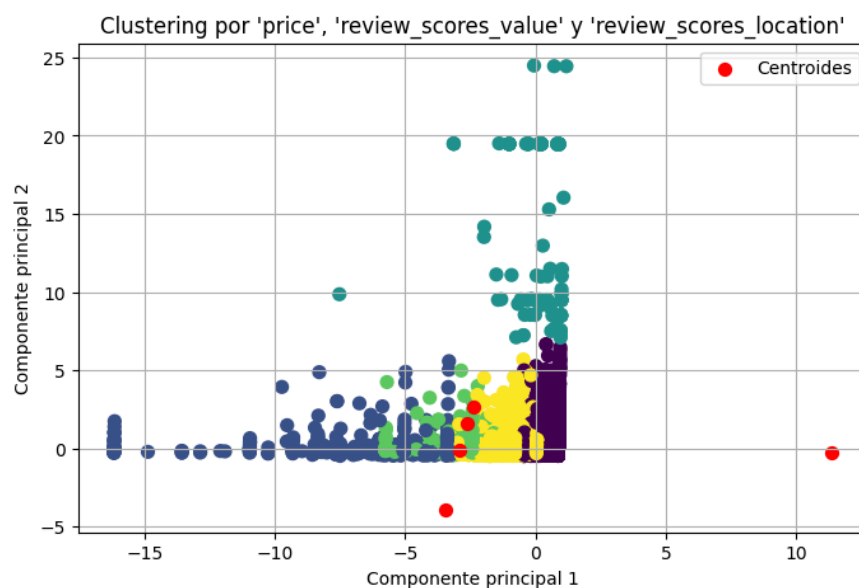


Inciso c.

Obtener los siguientes clusters aplicando el algoritmo de Kmeans:

K= 5, Variables = “price”, “review_scores_value” y “review_scores_location”

Esta vez usé cinco clusters y las variables “price”, “review_scores_value” y “review_scores_location”. Limpié el dataset, sustituyendo los valores nulos y transformando la columna de precios a formato numérico, normalicé los datos y apliqué el modelo de K-means, obteniendo las etiquetas de los clusters y los centroides. En la gráfica los datos se agrupan en cinco segmentos bien definidos, lo que indica que el modelo logró separar los datos según el precio y las calificaciones de valor y ubicación.



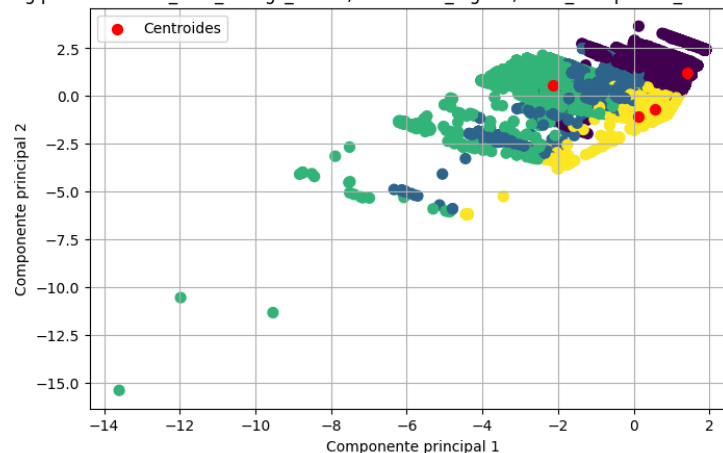
Inciso d.

Obtener los siguientes clusters aplicando el algoritmo de Kmeans:

K= 4, Variables = “calculated_host_listings_count”, “minimum_nights”, “host_acceptance_rate” y “availability_30”

En este inciso usé cuatro clusters y las variables “calculated_host_listings_count”, “minimum_nights”, “host_acceptance_rate” y “availability_30”. Analizando los centroides y la gráfica, se ve que los datos se agruparon en cuatro segmentos bien definidos y permite identificar las diferencias principales entre los grupos.

Clustering por 'calculated_host_listings_count', 'minimum_nights', 'host_acceptance_rate' y 'availability_30'



Inciso e.

Obtener los siguientes clusters aplicando el algoritmo de Kmeans:

K= 10, Variables = “number_of_reviews”, “calculated_host_listings_count”, “review_scores_cleanliness”, “review_scores_rating” y “host_acceptance_rate”

Para este último ejercicio usé diez clusters y las variables “number_of_reviews”, “calculated_host_listings_count”, “review_scores_cleanliness”, “review_scores_rating” y “host_acceptance_rate”. La gráfica se ve un poco distinta a las demás debido a que usé 10 centroides, lo que hizo que los datos se vieran muy pequeños en la gráfica.

Clustering por 'number_of_reviews', 'calculated_host_listings_count', 'review_scores_cleanliness', 'review_scores_rating' y 'host_acceptance_rate'

