



**Tecnológico
de Monterrey**

Tecnológico de Monterrey

Campus Puebla

Actividad 3 (Segmentación)

Mauricio Grau Gutierrez Rubio

A01734914

22 de agosto de 2025

En este trabajo se aplicó la técnica de K-Means Clustering sobre el dataset de Airbnb de la ciudad de Tokyo.

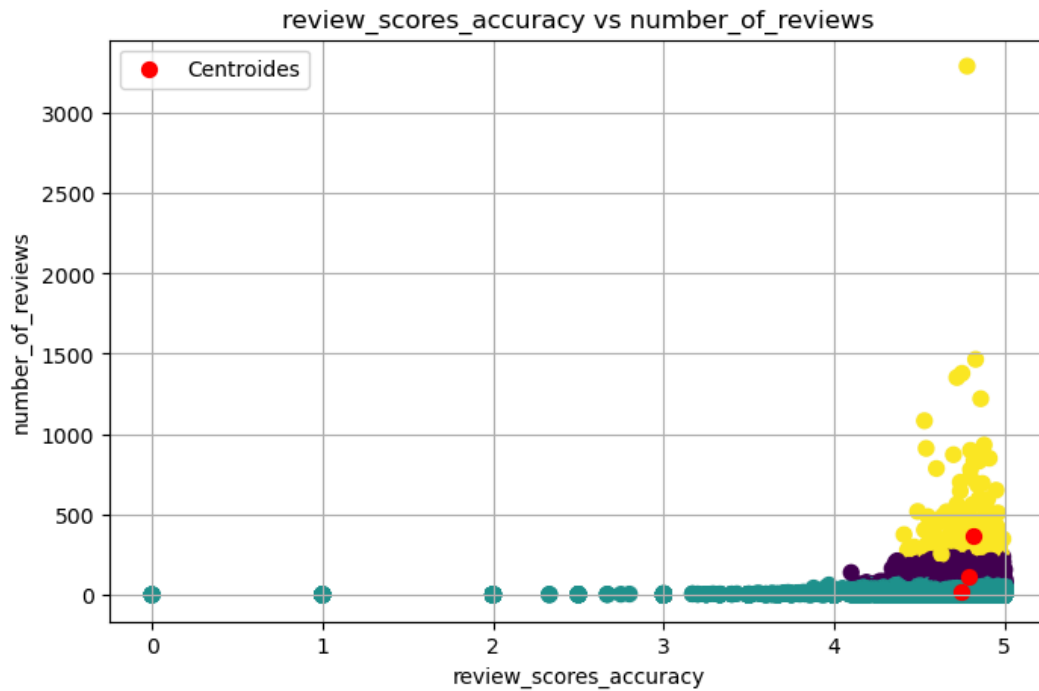
Durante el proceso se presentaron algunas complicaciones técnicas. La principal fue en el escenario (b), ya que la variable price estaba expresada en yenes con símbolos y separadores, lo que causó errores al entrenar el modelo. Fue necesario limpiar y transformar los precios a valores numéricos. También hubo que trabajar con valores faltantes (NaN), que se resolvieron aplicando estrategias de rellenado (forward fill).

(a) K= 3, Variables = “review_scores_accuracy” y “number_of_reviews

En este primer escenario se identificaron tres clusters bien diferenciados:

- Cluster 1: Alojamientos con muy pocas reseñas y calificaciones bajas o nulas, lo que refleja falta de interacción o anuncios nuevos.
- Cluster 2: Alojamientos con una cantidad moderada de reseñas, con calificaciones regulares en precisión. Representan la mayor parte de los anuncios.
- Cluster 3: Un grupo reducido de alojamientos con muchas reseñas y calificaciones muy altas, que corresponden a hosts consolidados y con alta demanda

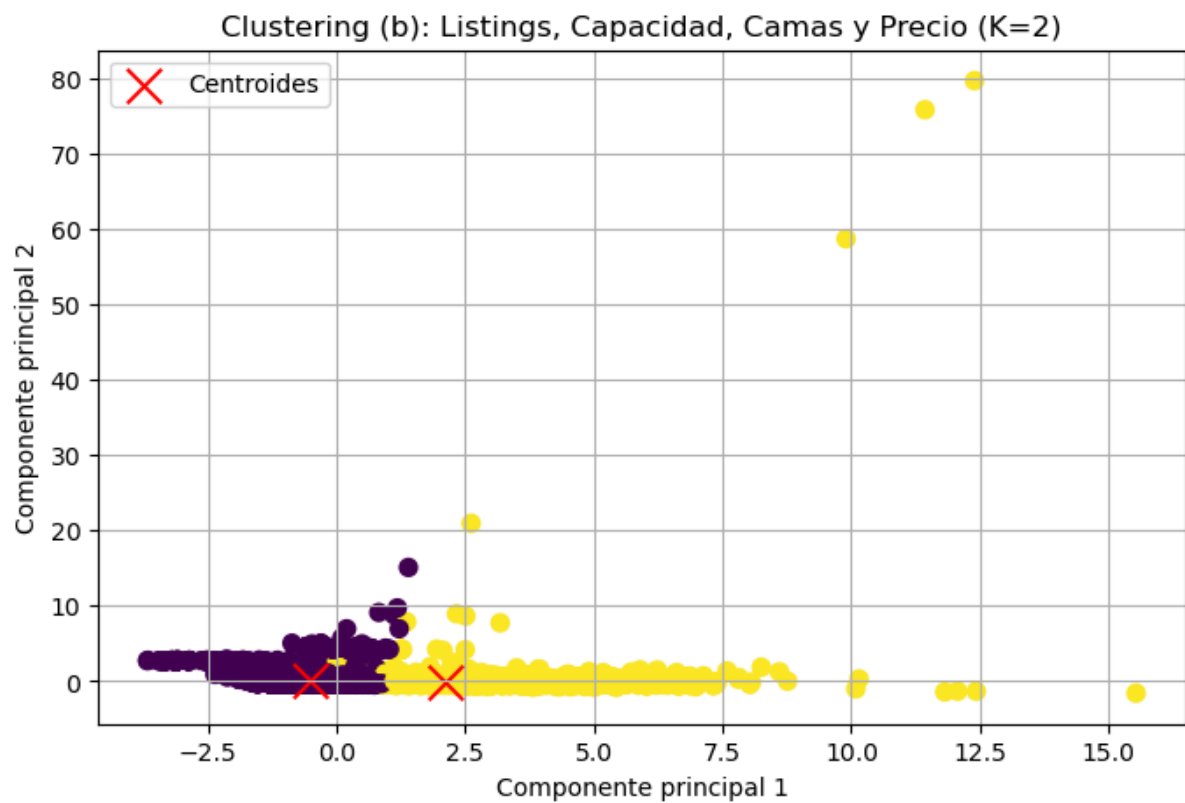
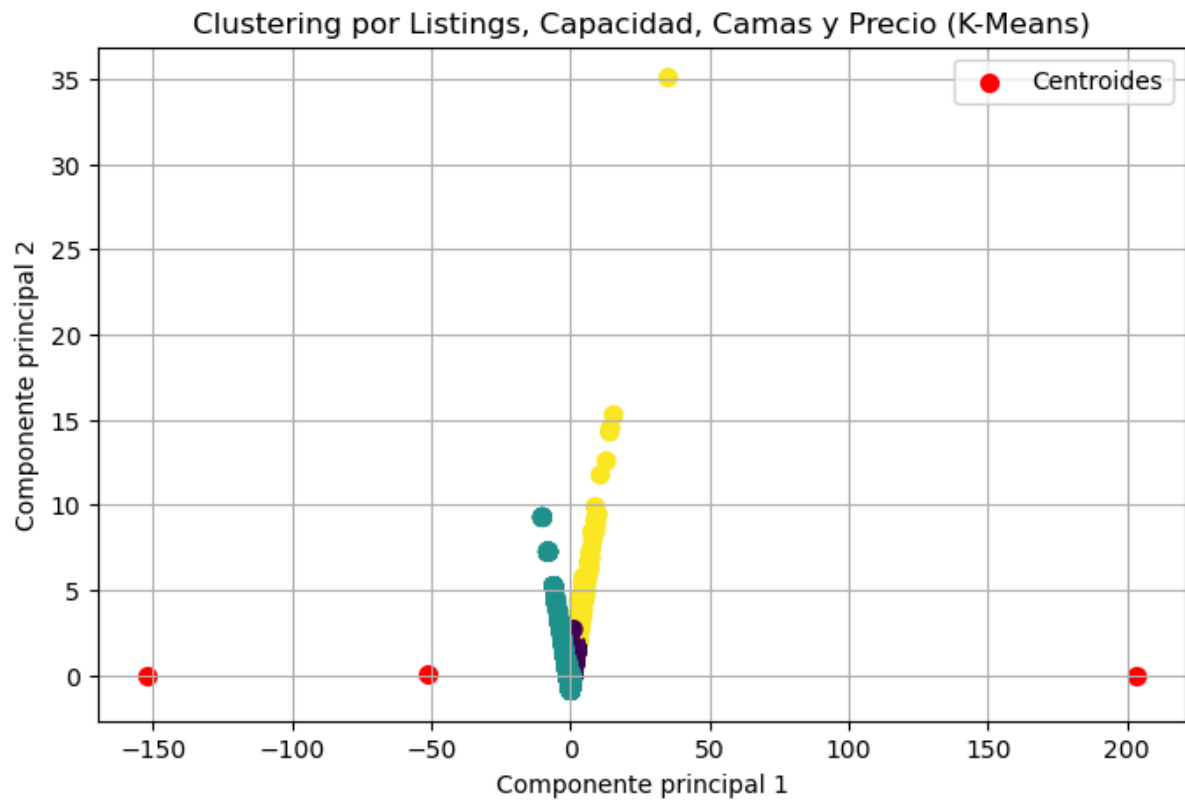
Este escenario permitió observar cómo se distribuye la popularidad de los anuncios: la mayoría tiene pocas reseñas, mientras que solo un pequeño grupo concentra una gran actividad.



(b) K= 2, Variables = “host_listings_count”, “accommodates”, “beds” y “price”

Este escenario presentó las mayores complicaciones debido al formato de la variable price (yenes con símbolos y comas). Tras la limpieza, el modelo generó dos clusters:

- Cluster 1: Alojamientos pequeños y de bajo costo, con menos camas y menor capacidad.
- Cluster 2: Alojamientos más grandes y costosos, con mayor número de camas y mayor capacidad de huéspedes.



(c) K= 5, Variables = “price”, “review_scores_value”,

“review_scores_location” y “price”

Se formaron cinco clusters que muestran la combinación de precio y calidad percibida:

- Alojamientos de bajo costo con buenas valoraciones.
- Alojamiento de precio medio con reseñas aceptables.
- Propiedades de alto costo con buenas calificaciones en ubicación y valor.
- Clústeres menores que representan combinaciones atípicas (ejemplo: precios altos pero con valoraciones regulares).

Este escenario evidencia que el mercado de Tokyo no se define solo por el precio, sino también por la percepción de valor y localización.

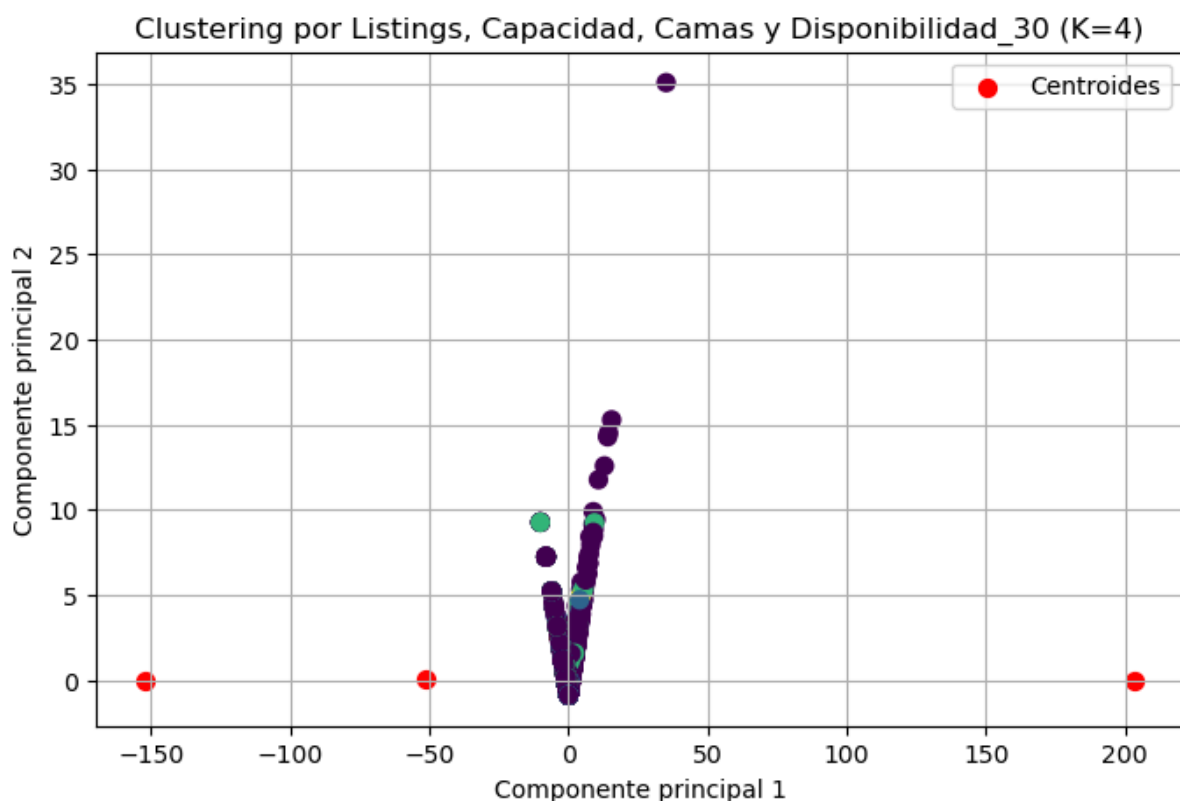


(d) K= 4, Variables = “host_listings_count”, “accommodates”, “beds” y “availability_30”

Los cuatro clusters obtenidos reflejaron distintos perfiles de anfitriones y propiedades:

- Cluster 1: Anfitriones con pocos anuncios y baja disponibilidad mensual.
- Cluster 2: Alojamientos grandes (más camas y accomodates) con disponibilidad variable.
- Cluster 3: Anfitriones con múltiples listings y alta disponibilidad (más profesionalizados).
- Cluster 4: Alojamientos medianos tanto en tamaño como en disponibilidad.

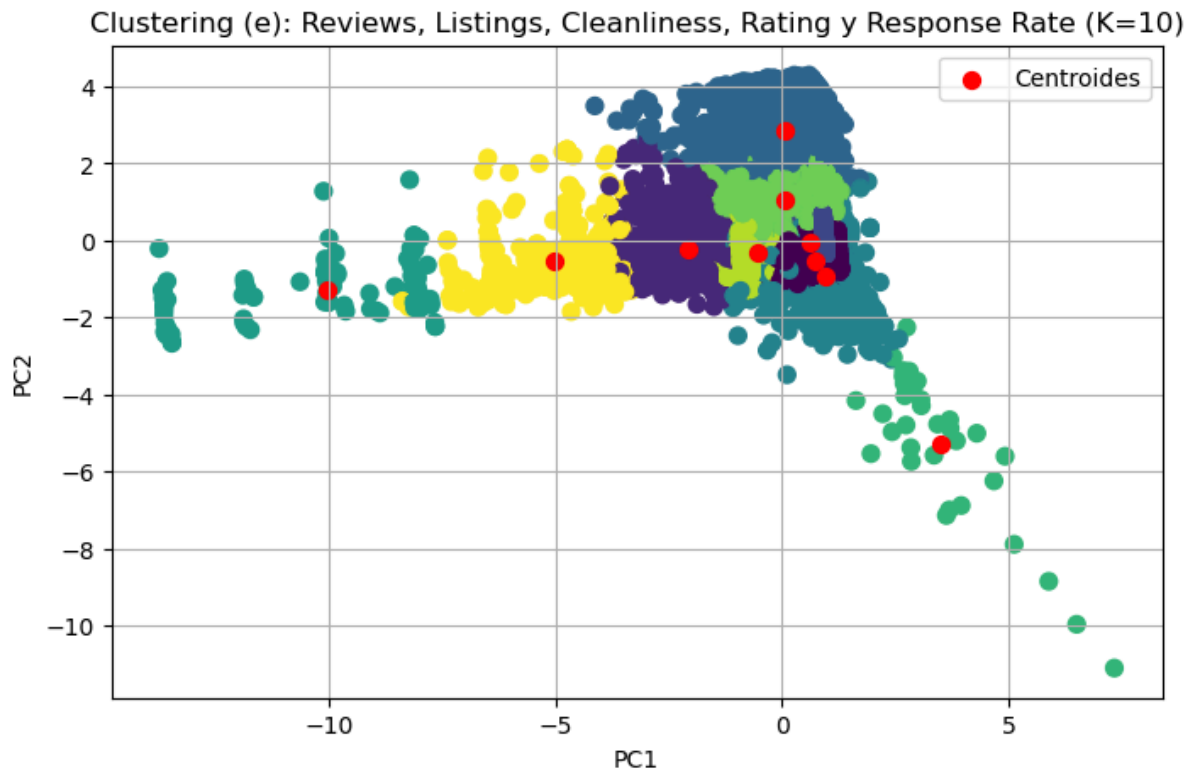
Este escenario permitió distinguir entre anfitriones ocasionales y profesionales.



**(e)K= 10, Variables = “reviews_per_month”,
“calculated_host_listings_count”, “review_scores_cleanliness”,
“review_scores_rating” y “host_response_rate”**

- Fue el escenario más detallado, con diez clusters que combinan frecuencia de reseñas, limpieza, calificaciones y tasa de respuesta:
- Clusters de alta actividad y buenas calificaciones, reflejando anfitriones muy comprometidos.
- Clusters con poca actividad y calificaciones bajas, que muestran menor fiabilidad.
- Diferencias claras según el nivel de limpieza percibida.
- Clusters diferenciados por la rapidez y frecuencia de respuesta de los anfitriones.

Este escenario muestra de forma clara la relación entre la calidad del servicio y el comportamiento del anfitrión.



Conclusión

- El análisis de clustering permitió identificar distintos perfiles de alojamientos en Tokyo:
- En el escenario (a) se observaron diferencias claras en la popularidad de los anuncios.
- El escenario (b) mostró la división entre alojamientos económicos y de lujo, aunque presentó complicaciones técnicas con la variable price.
- El escenario (c) resaltó la diversidad de combinaciones entre precio y valoraciones.
- En el escenario (d) se diferenciaron los hosts ocasionales y profesionales.

Finalmente, el escenario (e) ofreció la segmentación más detallada, reflejando la calidad del servicio y atención de los anfitriones.