

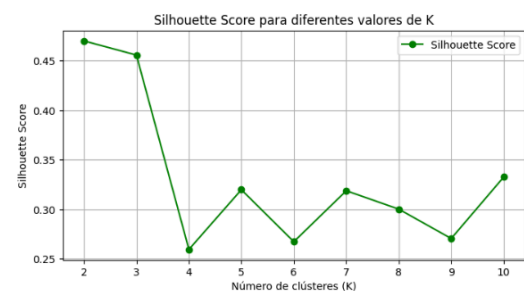
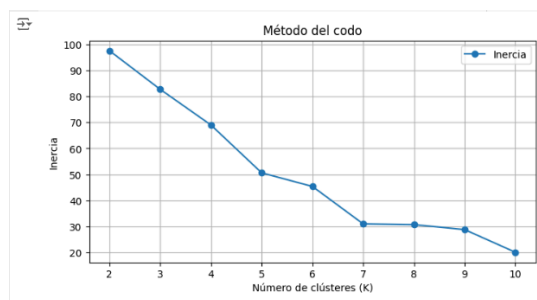
1. **Histogramas:** Fueron creados para estudiar la distribución de las variables cuantitativas, como la satisfacción general con la universidad.
2. **Nubes de palabras:** Se emplearon para analizar las respuestas en texto y resaltar las palabras más repetidas.

### 3. Modelo de agrupación: K-means

- **Estandarización**

Primero, las variables numéricas fueron normalizadas con StandardScaler para asegurar la consistencia.

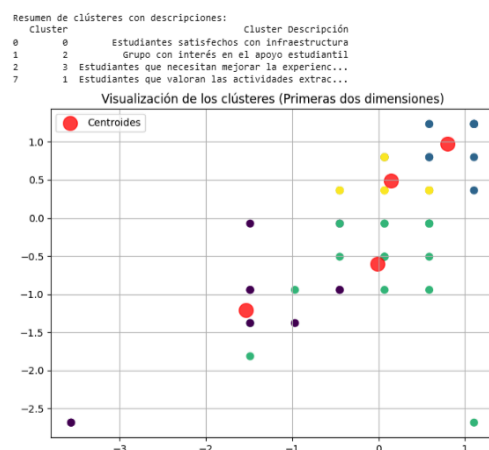
1. **Método del codo:** Se analizó la suma de las distancias desde los puntos a sus centroides para varios valores de K, valor de K que mostró una estabilización en la reducción de la inercia fue elegido como el número ideal de clústeres.
2. **Silhouette score:** Se calculó para evaluar qué tan bien se agrupan los elementos dentro de los clústeres y su distanciamiento entre sí.



- **Clústeres principales**

Se identificaron cuatro clústeres principales, el primero compuesto por estudiantes interesados en actividades extracurriculares, el segundo por aquellos que buscan mejorar su experiencia académica, el tercero enfocado en el apoyo estudiantil, y el cuarto por estudiantes satisfechos con la infraestructura universitaria. Estos grupos ofrecieron información valiosa para segmentar a los estudiantes y comprender mejor sus necesidades.

- **Visualización**



### 4. Modelo de clasificación: Naive Bayes

Los datos se separaron en dos conjuntos, el 80% se usó para entrenar el modelo y optimizar sus parámetros, y el 20% restante se reservó para probar el rendimiento del modelo y evitar el sobreajuste.

Se calcularon métricas esenciales como, la precisión, que mide el porcentaje de predicciones acertadas, el recall, que evalúa la habilidad del modelo para identificar correctamente cada clase, y F1-score, que combina precisión y recall, siendo especialmente útil en situaciones con clases desbalanceadas.

Procesando: ¿Con qué frecuencia participas en actividades extracurriculares organizadas por la universidad?

Reporte de clasificación - '¿Con qué frecuencia participas en actividades extracurriculares organizadas por la universidad?':

	precision	recall	f1-score	support
A veces	0.60	1.00	0.75	6
Casi nunca	1.00	0.33	0.50	3
Nunca	0.00	0.00	0.00	2
Siempre	1.00	0.00	0.00	1
accuracy			0.58	12
macro avg	0.65	0.33	0.31	12
weighted avg	0.63	0.58	0.50	12

Procesando: En promedio, ¿cuántas horas dedicas a actividades relacionadas con la universidad (clases, tareas, estudio, etc.) por semana?

Reporte de clasificación - 'En promedio, ¿cuántas horas dedicas a actividades relacionadas con la universidad (clases, tareas, estudio, etc.) por semana?':

	precision	recall	f1-score	support
10 - 20 horas	0.00	0.00	0.00	1
20 - 30 horas	0.50	0.50	0.50	6
Más de 30 horas	0.50	0.20	0.29	5
micro avg	0.36	0.33	0.35	12
macro avg	0.33	0.23	0.26	12
weighted avg	0.46	0.33	0.37	12

Procesando: ¿Qué aspecto de la vida universitaria consideras más gratificante y por qué?

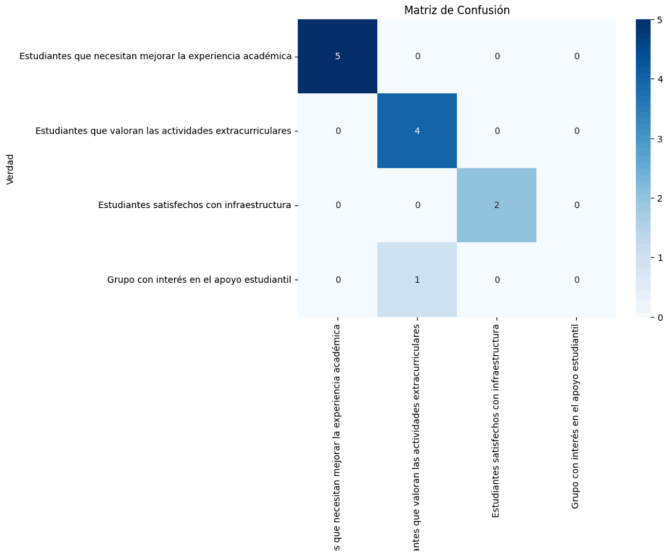
No hay suficiente después de filtro.

Procesando: Si pudieras mejorar un aspecto de tu experiencia universitaria, ¿cuál sería y cómo lo cambiarías?

No hay suficiente después de filtro.

Procesando: ¿Qué tan cómodo te sientes al expresar tus ideas en el ambiente universitario y por qué?

No hay suficiente después de filtro.



## Conclusión

El proyecto permitió descubrir patrones relevantes en los datos y explorar métodos avanzados de análisis y modelado. La aplicación de K-means resultó eficaz para clasificar a los estudiantes, identificando grupos que brindan una mejor comprensión de sus características y necesidades, como el interés en actividades extracurriculares y la satisfacción con la infraestructura. Este análisis resalta la importancia de utilizar modelos de machine learning para interpretar datos complejos, tanto numéricos como textuales, y obtener información valiosa para apoyar la toma de decisiones.