




ARTIFICIAL INTELLIGENCE

# PROJECT

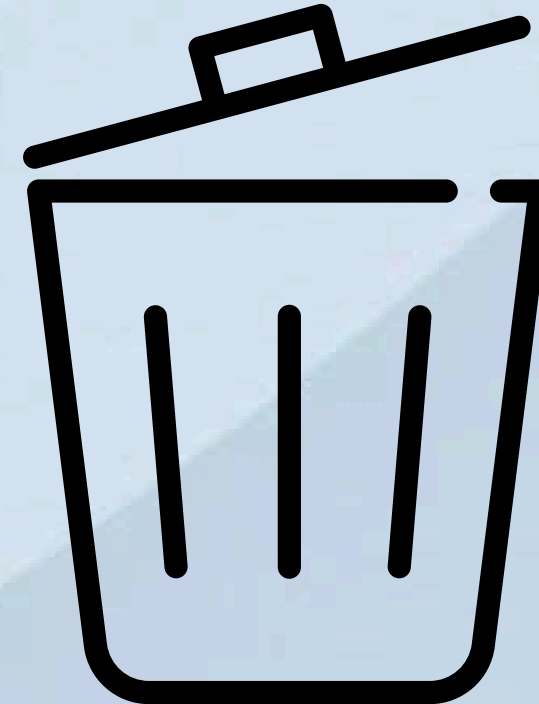
By Leonardo Berdejo, Bote Sánchez, Danny  
Cano Y., Camila Carrera, Steve Rosero



# PREPROCESAMIENTO

## Requerimientos:

- Muestra del 20% de datos del archivo
- Se considerará únicamente la lista corta de causas de muerte (lc1)
- Eliminaciones en la columna lc1:
  - Defunciones ocurridas fuera de Ecuador (prov\_fall sea Null)
  - Muertes violentas (64)
  - Muertes por COVID-19 (66 y 67)
  - Causas mal definidas (99)
  - Resto de causas (88)
  - Valores vacíos (valores con Null)



# PREPROCESAMIENTO

A. Se eliminarán todos los datos que no contengan información alguna en la columna lc1:

```
#Se crea nuevo dataset para almacenar los preprocesamientos  
df_1 = df.dropna(subset=['lc1'])  
df_1
```

B. Se eliminan los datos en la columna lc1 que se mencionaron en los requerimientos:

```
lc1_del = ['64 Agresiones (Homicidios)', 'COVID-19, virus identificado', 'COVID-19, virus no identificado', '99 Causas mal definidas', '88 Resto de causas']  
df_1 = df_1[~df_1['lc1'].isin(lc1_del)]  
df_1
```

La columna **prov\_fall** será útil para eliminar las filas de defunciones ocurridas fuera de Ecuador, se eliminarán las filas que tengan a prov\_fall vacía entendiéndose como que fallecieron en territorio externo a cualquier provincia (Ecuador)

```
df_1 = df_1.dropna(subset=['prov_fall']) #Eliminación de valores nulos  
df_1
```



# PREPROCESAMIENTO

Limpiamos el dataset de valores NaN

```
# Convertir todas las columnas categóricas a tipo categórico si no lo están
categorical_columns = df_1.select_dtypes(include=['category']).columns
df_1[categorical_columns] = df_1[categorical_columns].astype('category')

# Añadir la categoría 'Desconocido' a cada columna categórica si aún no está presente
for col in categorical_columns:
    if 'Desconocido' not in df_1[col].cat.categories:
        df_1[col] = df_1[col].cat.add_categories('Desconocido')

# Reemplazar todos los valores NaN en el DataFrame df_1 con la cadena 'Desconocido'
df_1_clean = df_1.fillna('Desconocido')
# Convertir todas las columnas a tipo de datos cadena
df_1_clean = df_1_clean.astype(str)
```

✓ 1m 0.1s

```
#Visualizamos el Dataset limpio
df_1_clean
```

✓ 5.0s

C. Se genera varios modelos con variables diferentes:

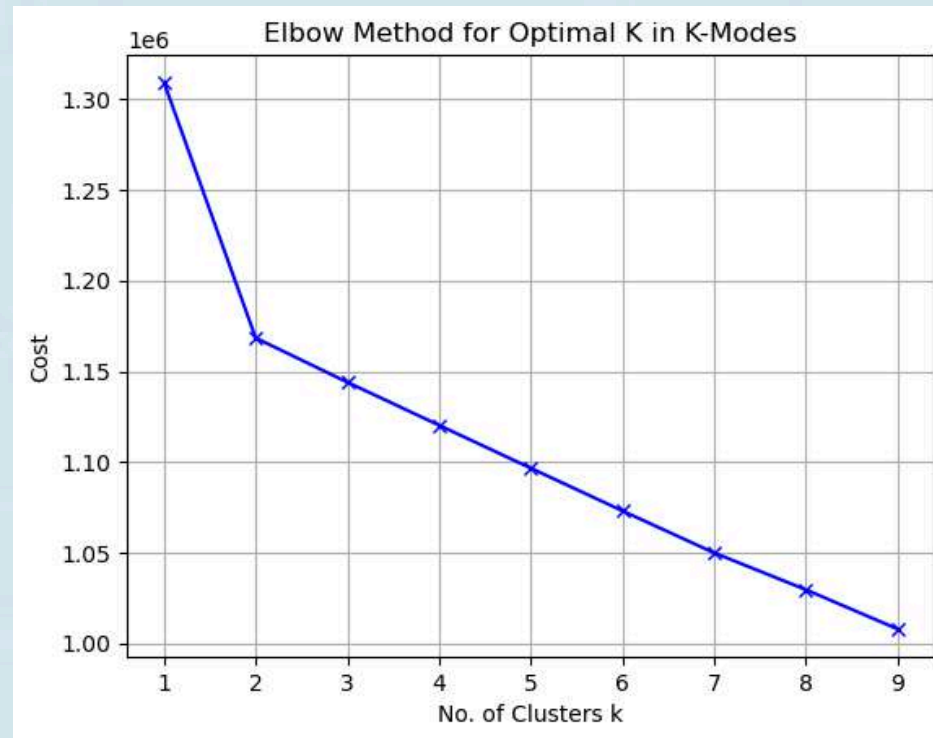
# MODELO FINAL 1

Variables consideradas en Modelo 1:

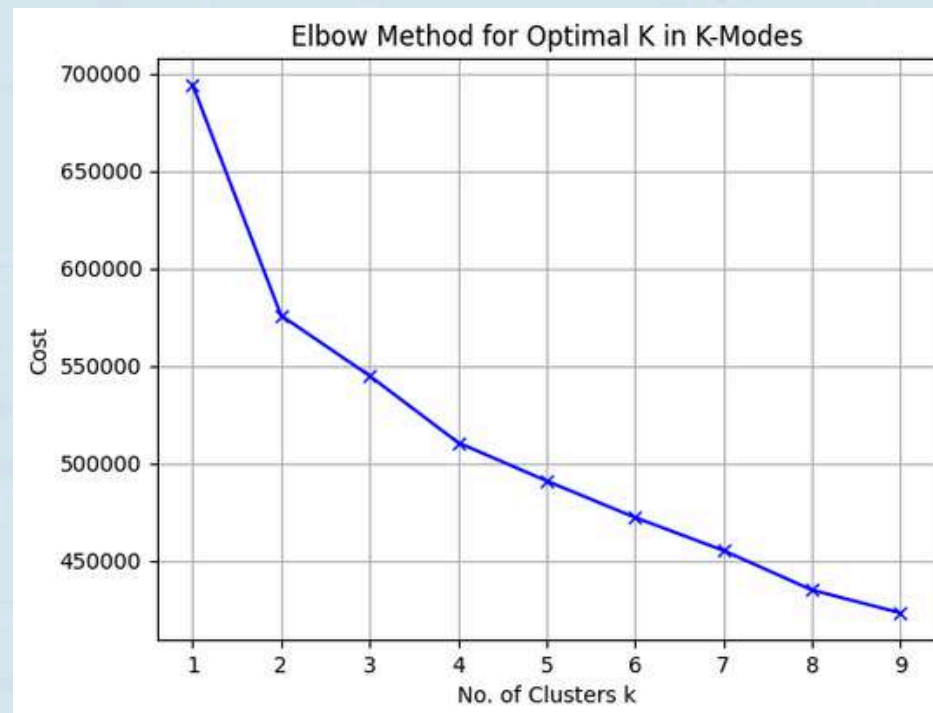
```
columns_to_keep = ['sexo', 'area_fall', 'lc1']  
df_selected = df[columns_to_keep]  
  
# Obtener el 20% de los datos aleatorios  
df_prueba3 = df_selected.sample(frac=0.2, random_state=40);
```

	sexo	area_fall	lc1
935151	Hombre	Urbana	47 Enfermedades crónicas de las vías respirato...
175592	Mujer	Urbana	9 Neoplasia maligna del estómago
861969	Mujer	Urbana	25 Neoplasias benignas in situ y de comportami...
811518	Mujer	Urbana	26 Diabetes Mellitus
1062163	Mujer	Urbana	57 Accidentes de transporte terrestre
...	...	...	...
1203391	Mujer	Rural	41 Insuficiencia cardíaca, complicaciones y en...
164562	Hombre	Urbana	51 Cirrosis y otras enfermedades del hígado
1082784	Mujer	Rural	56 Malformaciones congénitas, deformidades y a...
1051014	Hombre	Urbana	57 Accidentes de transporte terrestre
215231	Hombre	Urbana	39 Paro cardíaco
251022 rows × 3 columns			

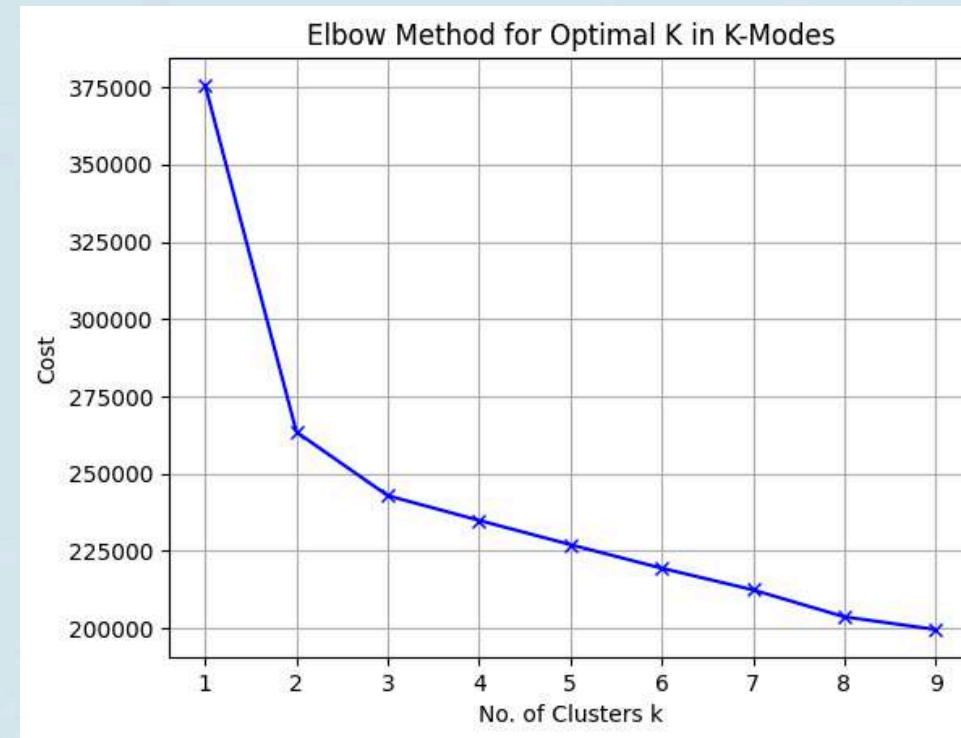
# MODELADO



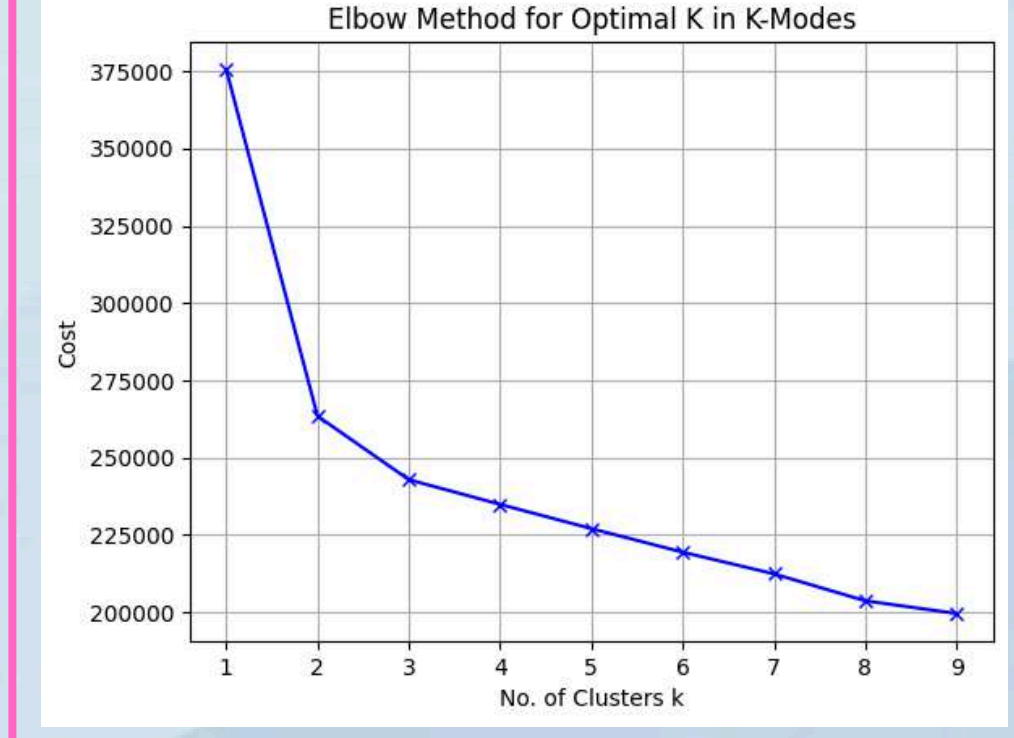
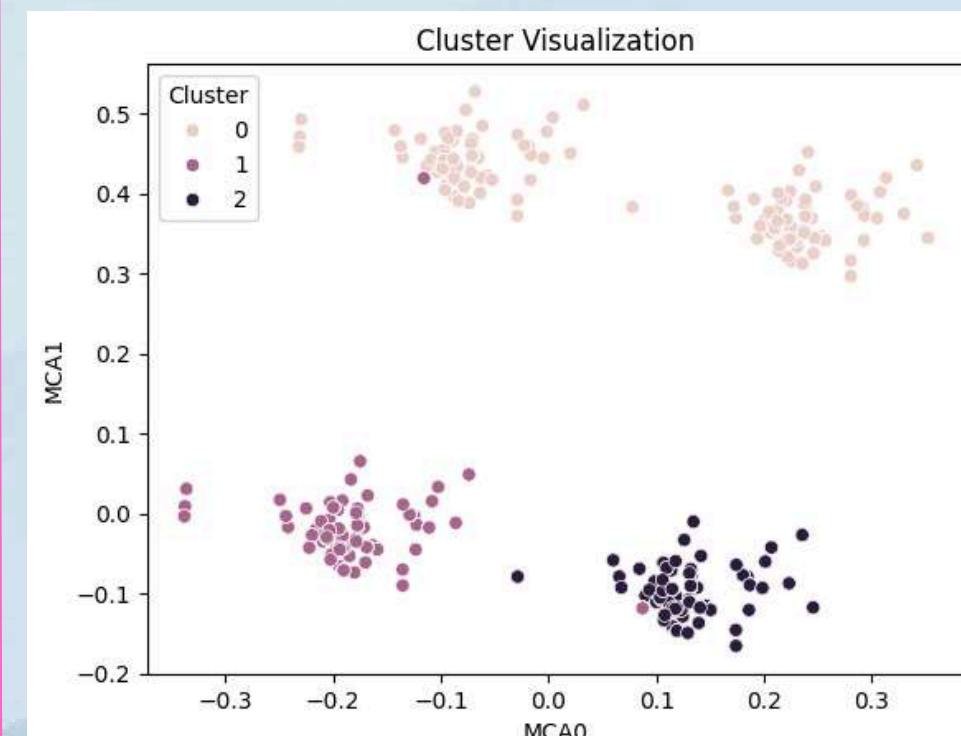
ACSS con K = 2: 0.068



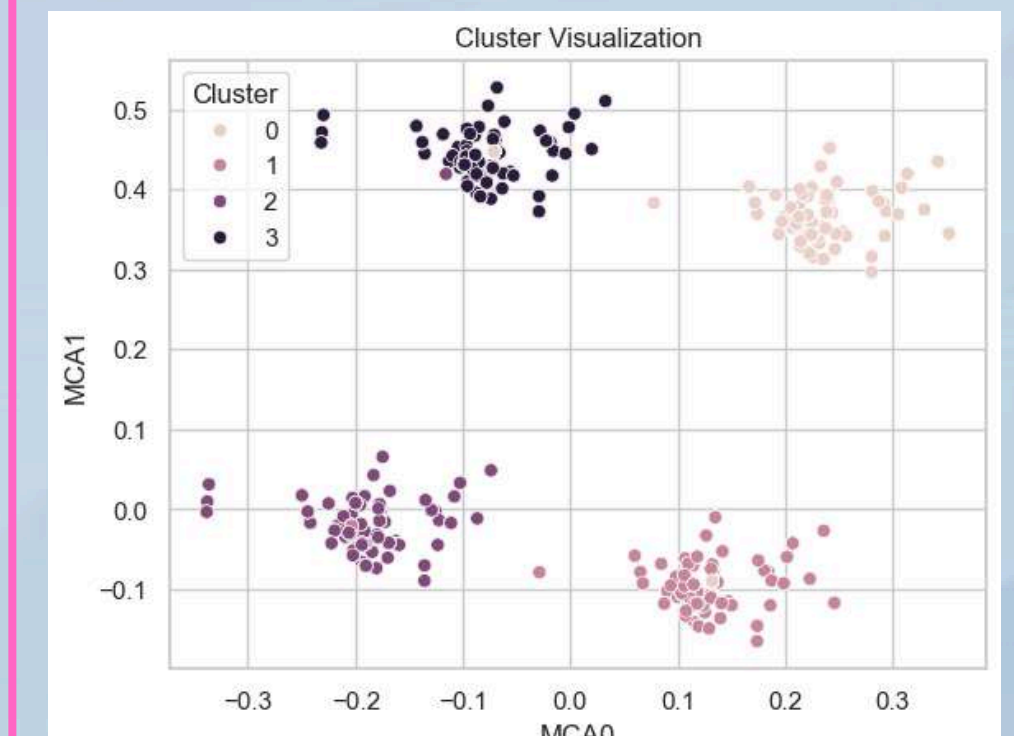
ACSS con K = 4: 0.13



ACSS con K = 3: 0.42

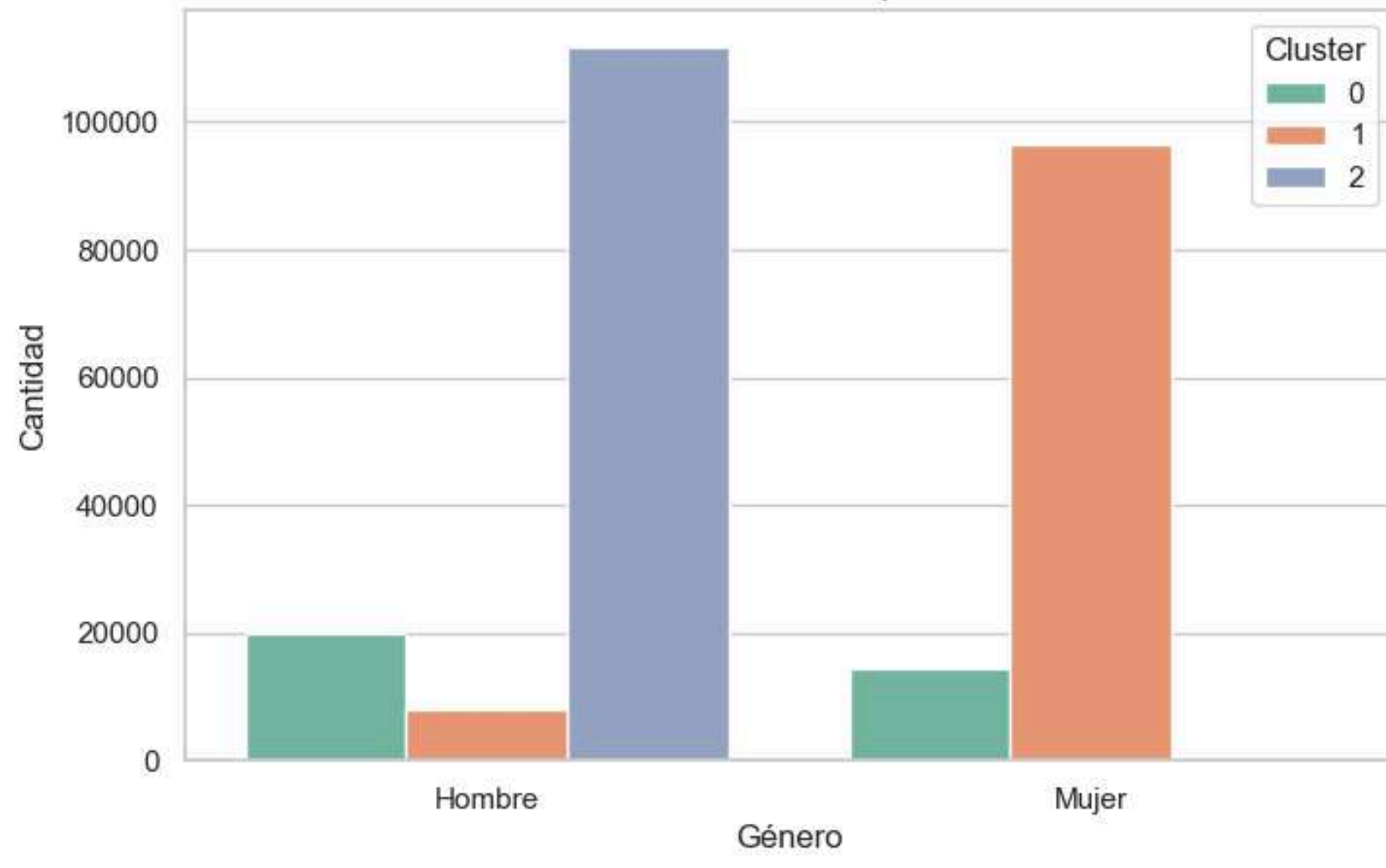


ACSS con K = 4: 0.377

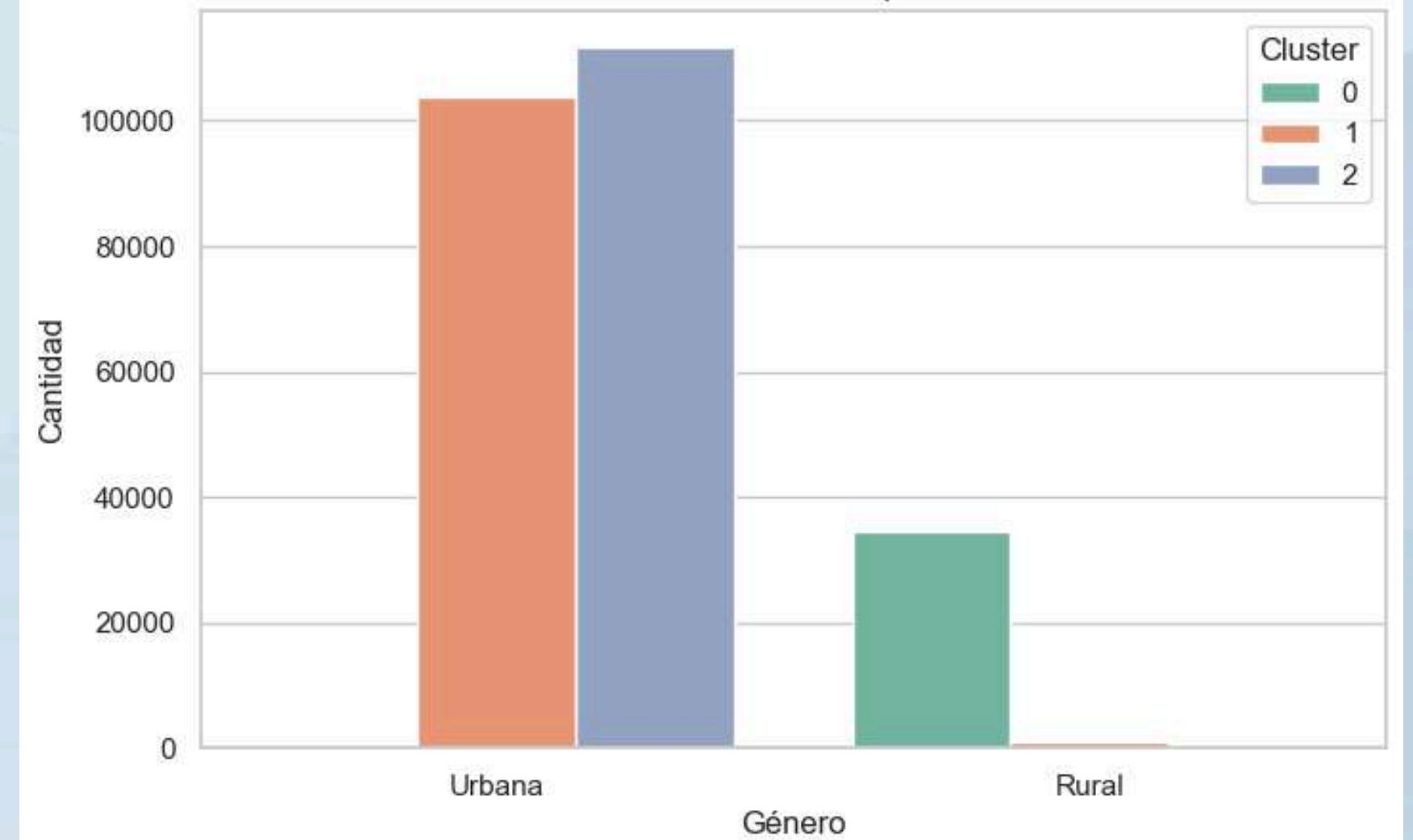


# RESULTADOS K:3

Distribución de Género por Cluster

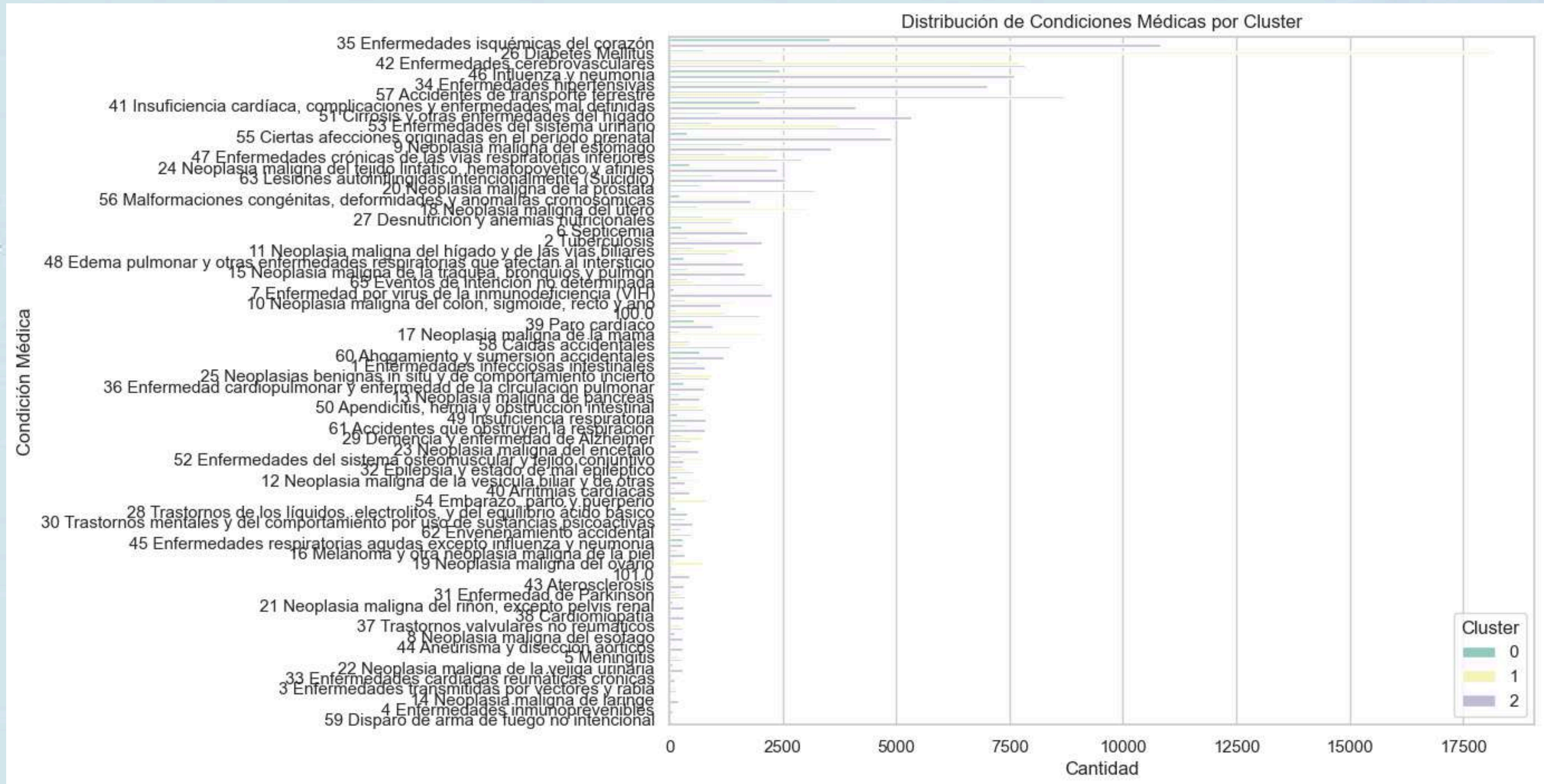


Distribución de Género por Cluster



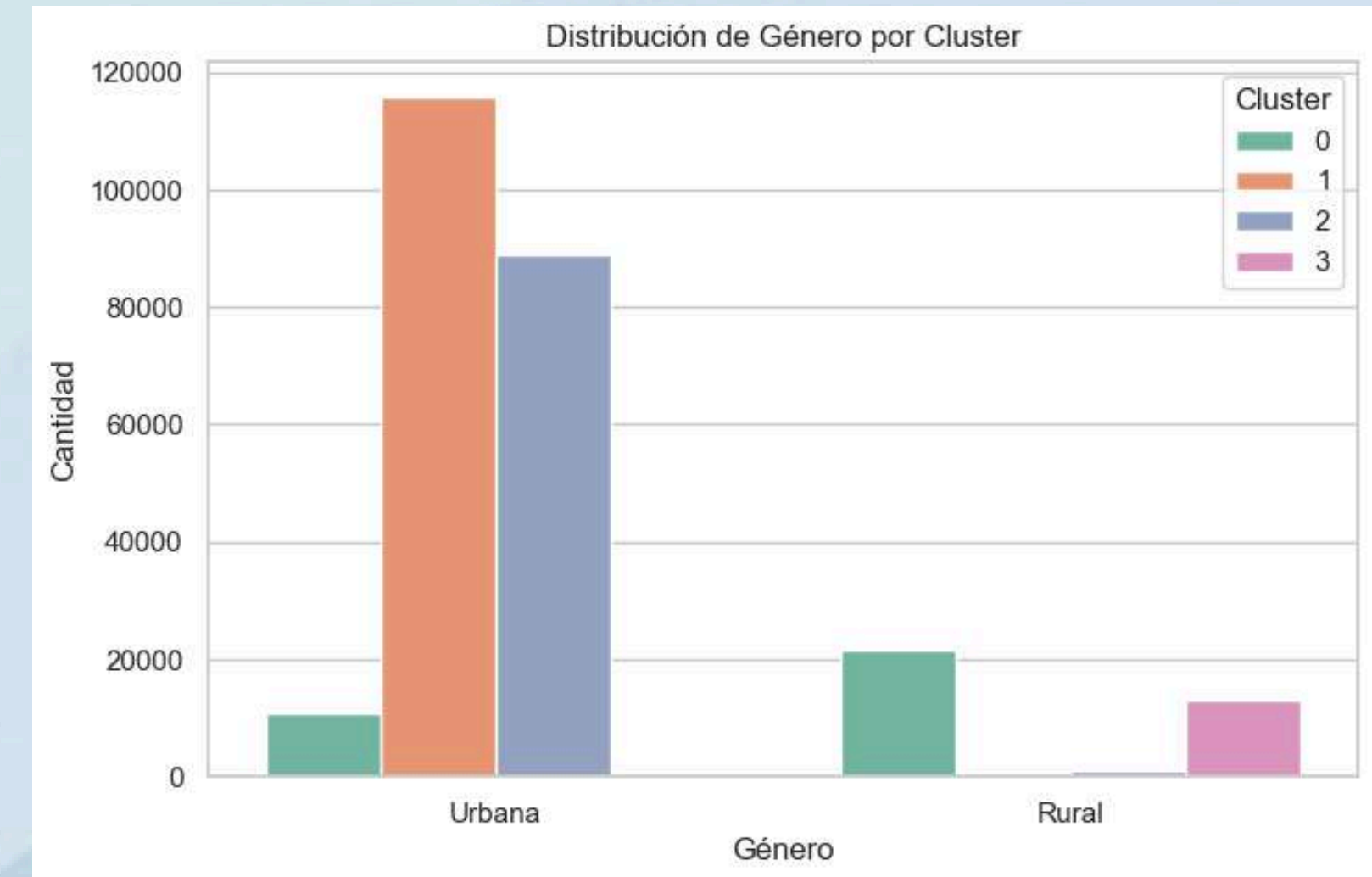
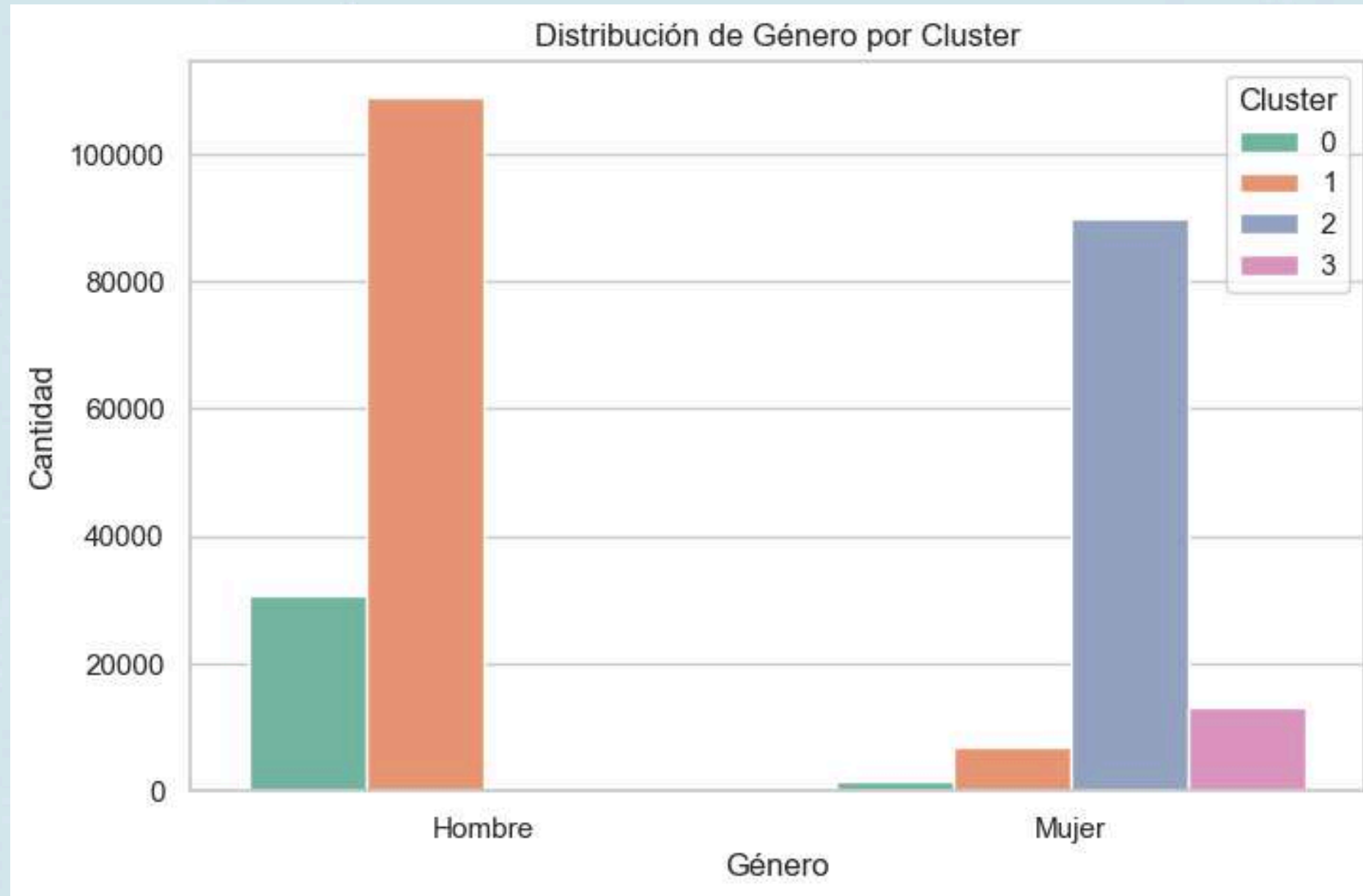


# RESULTADOSK:3

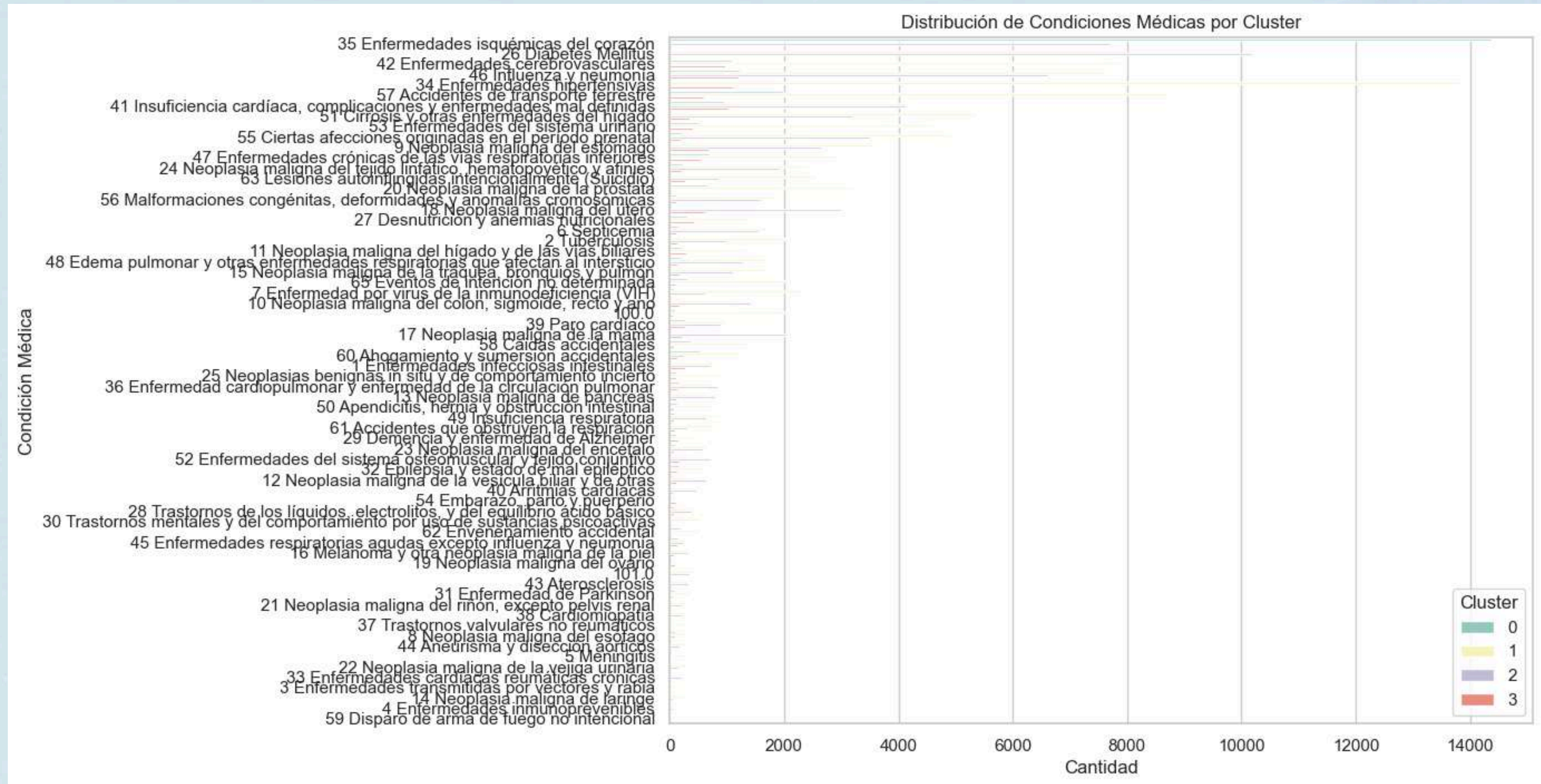




# RESULTADOS K:4



# RESULTADOS K: 4



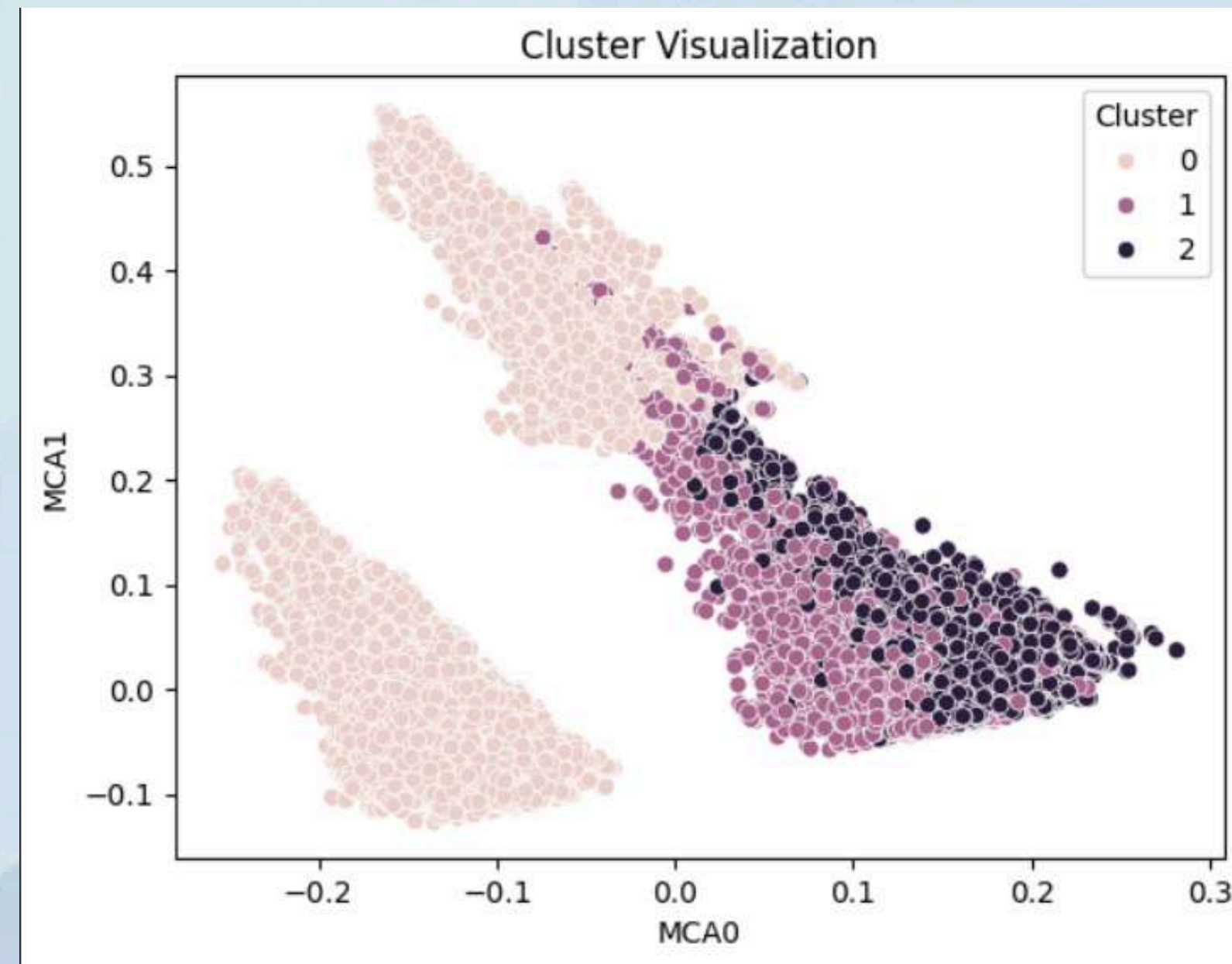
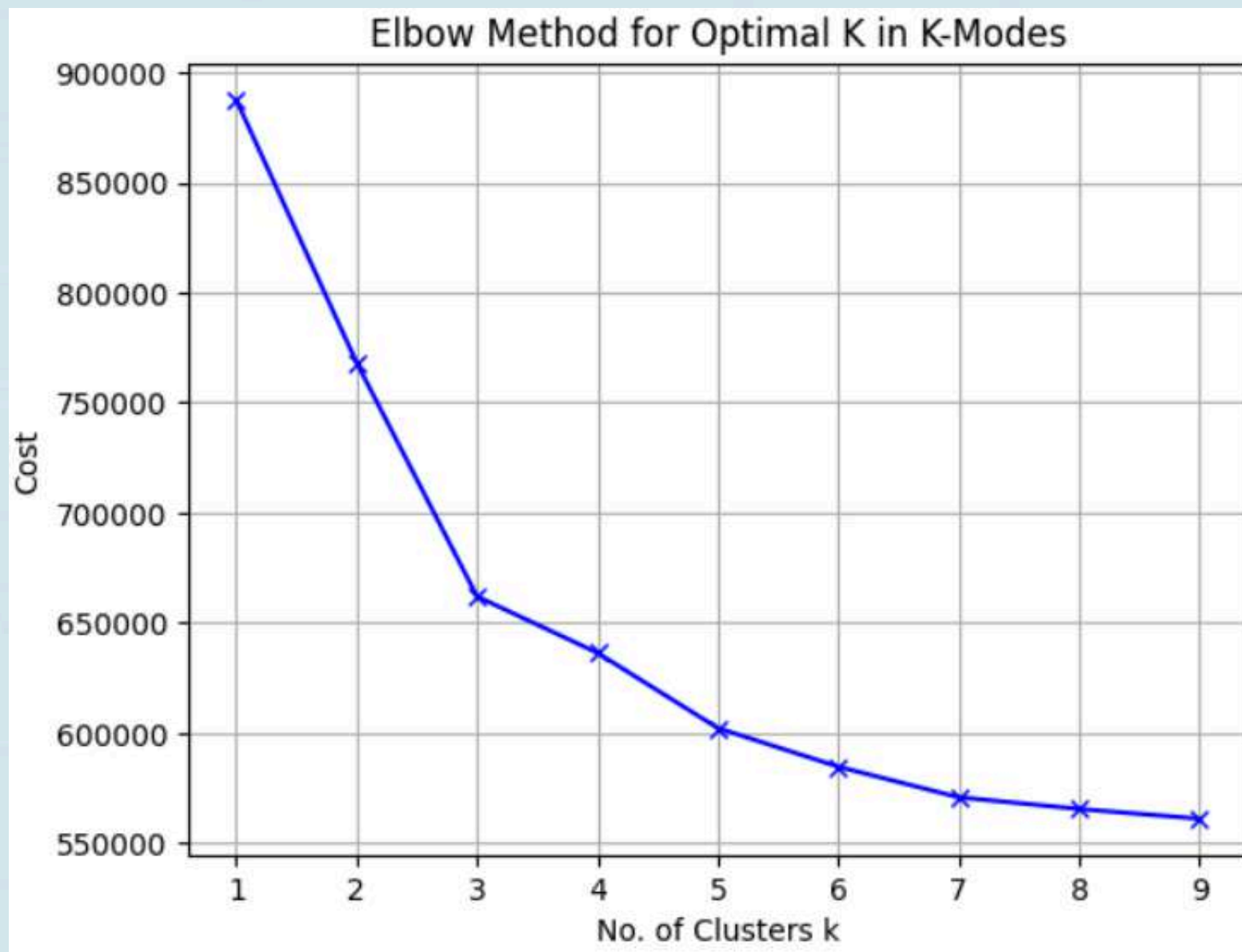
# MODELO FINAL 2

Variables consideradas en Modelo 2:

```
columns_to_keep = ['edad', 'lc1', 'lugar_ocur', 'area_fall', 'autopsia']  
df_selected = df[columns_to_keep]  
  
# Obtener el 20% de los datos aleatorios  
df_prueba8 = df_selected.sample(frac=0.2, random_state=40);
```



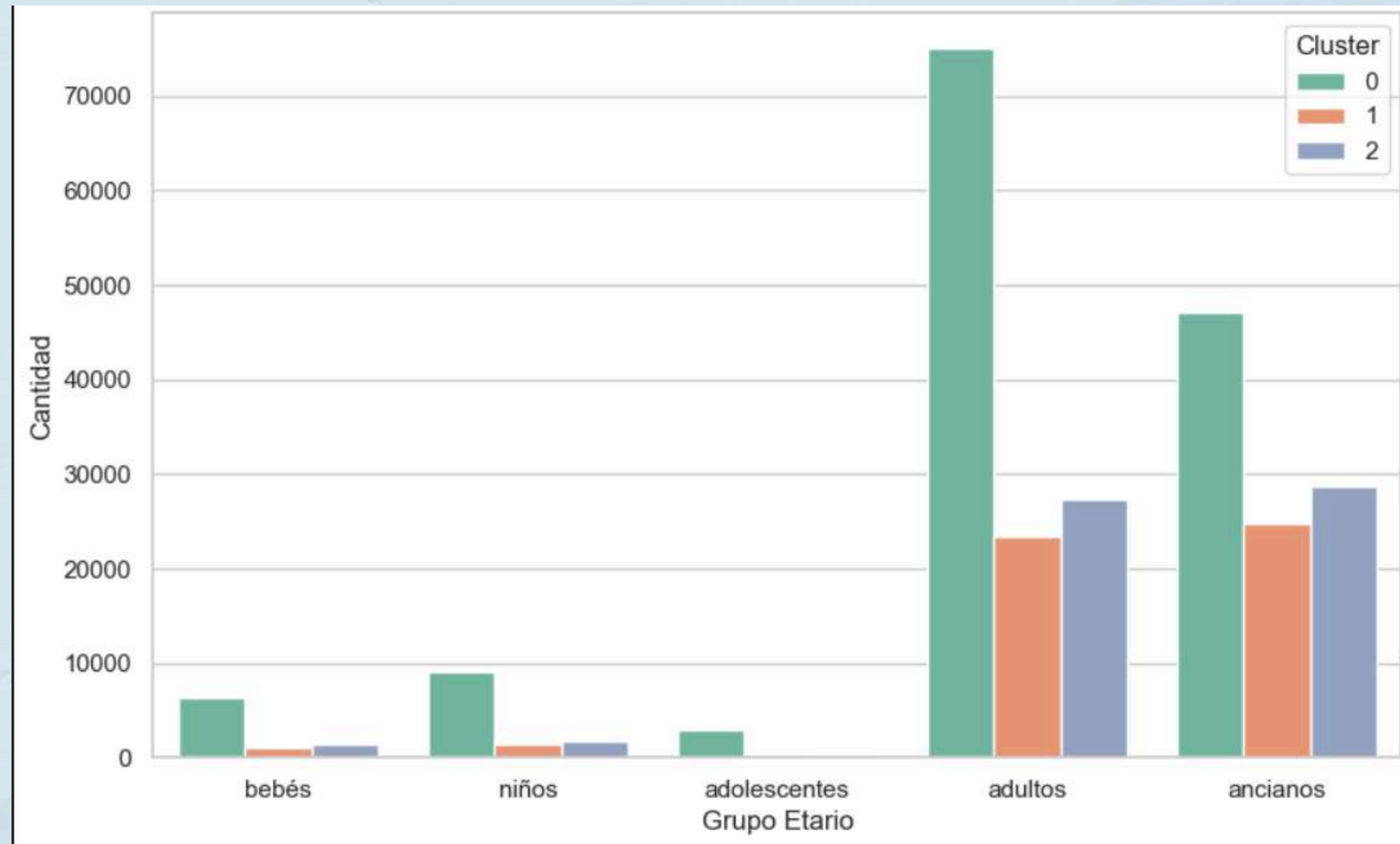
# MODELADO



ACSS con K = 3: 0.36

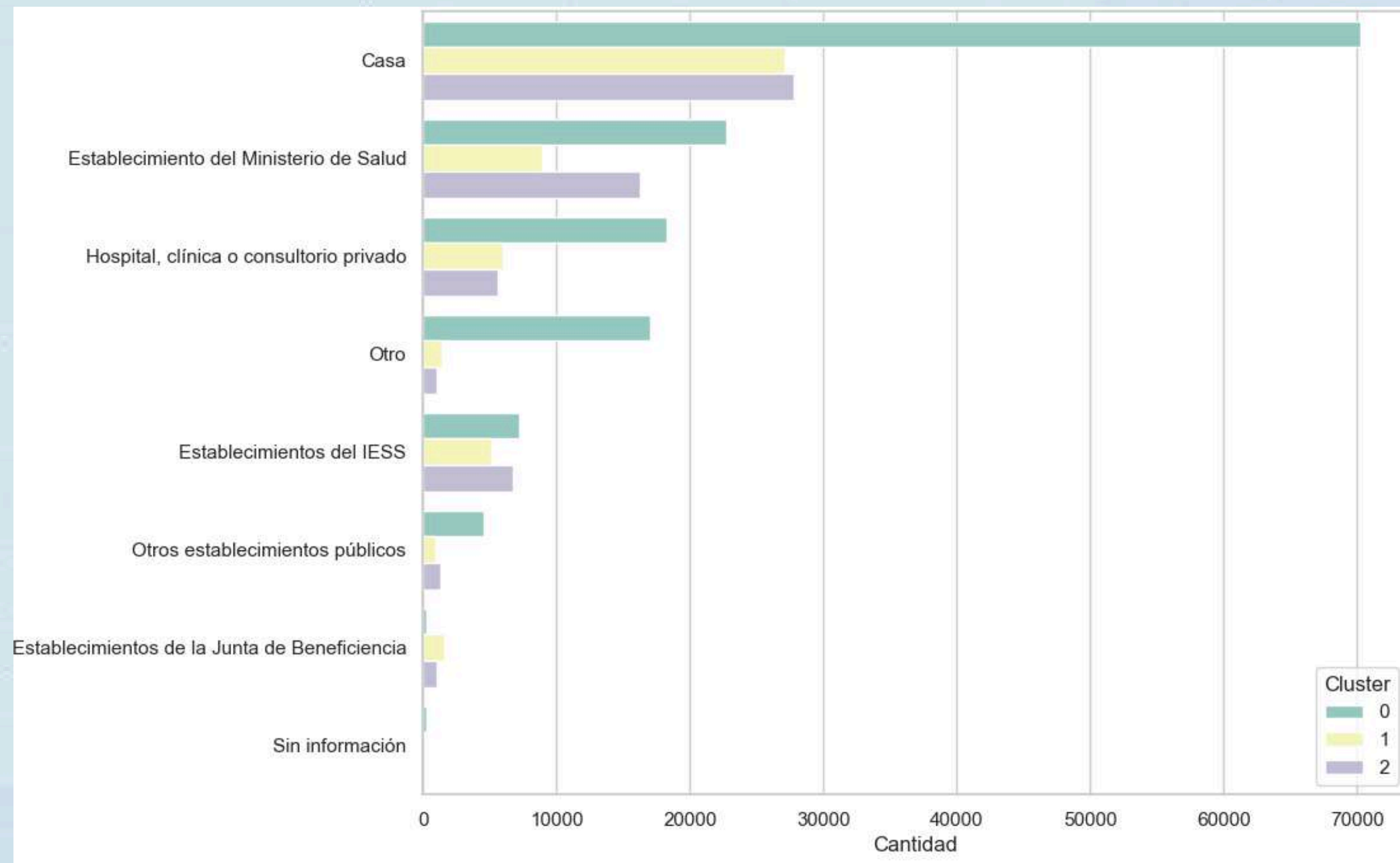
# RESULTADOS K=3

Distribución por grupos de edad:



# RESULTADOS K=3

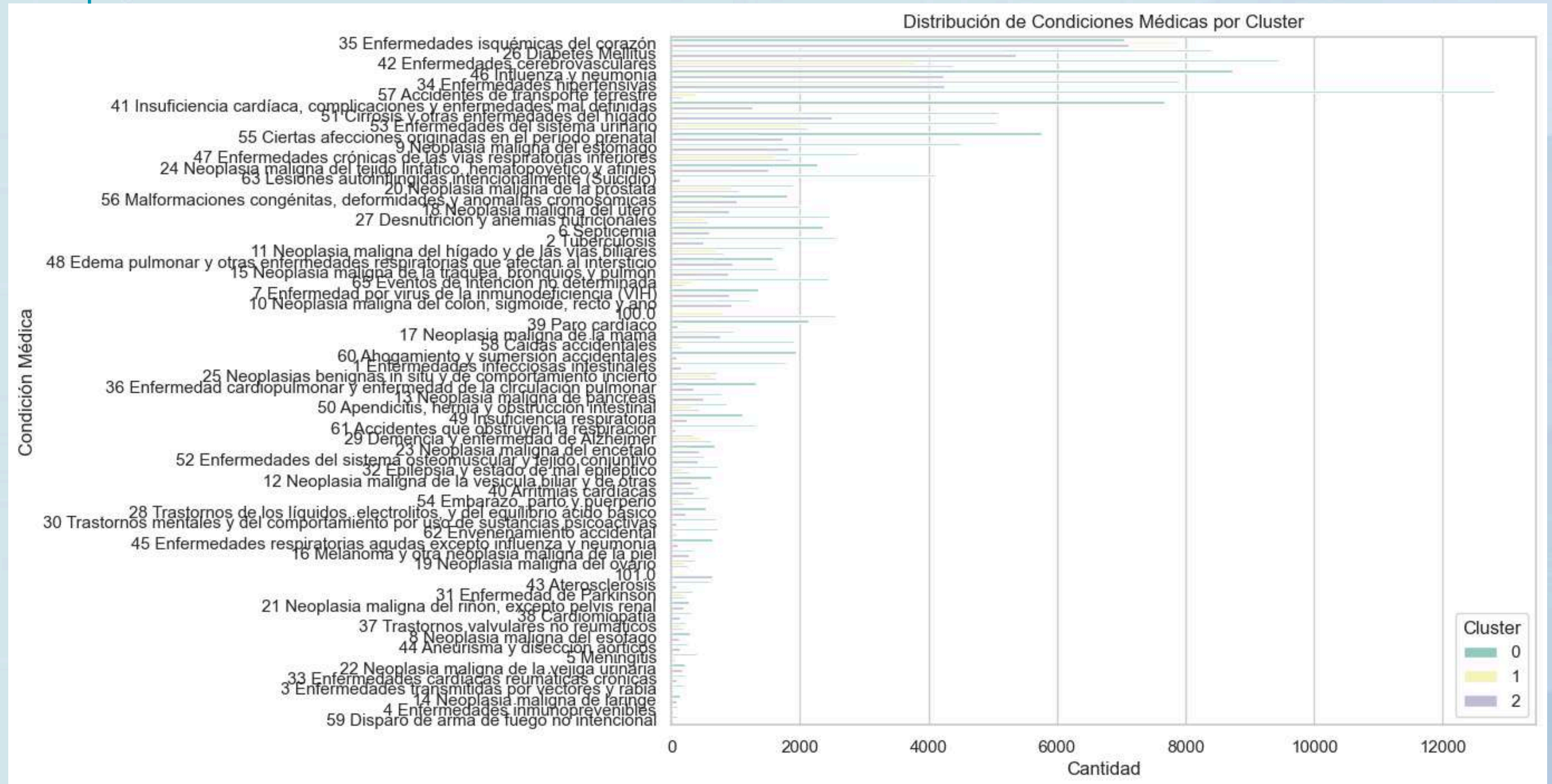
Distribución por lugar de fallecimiento:





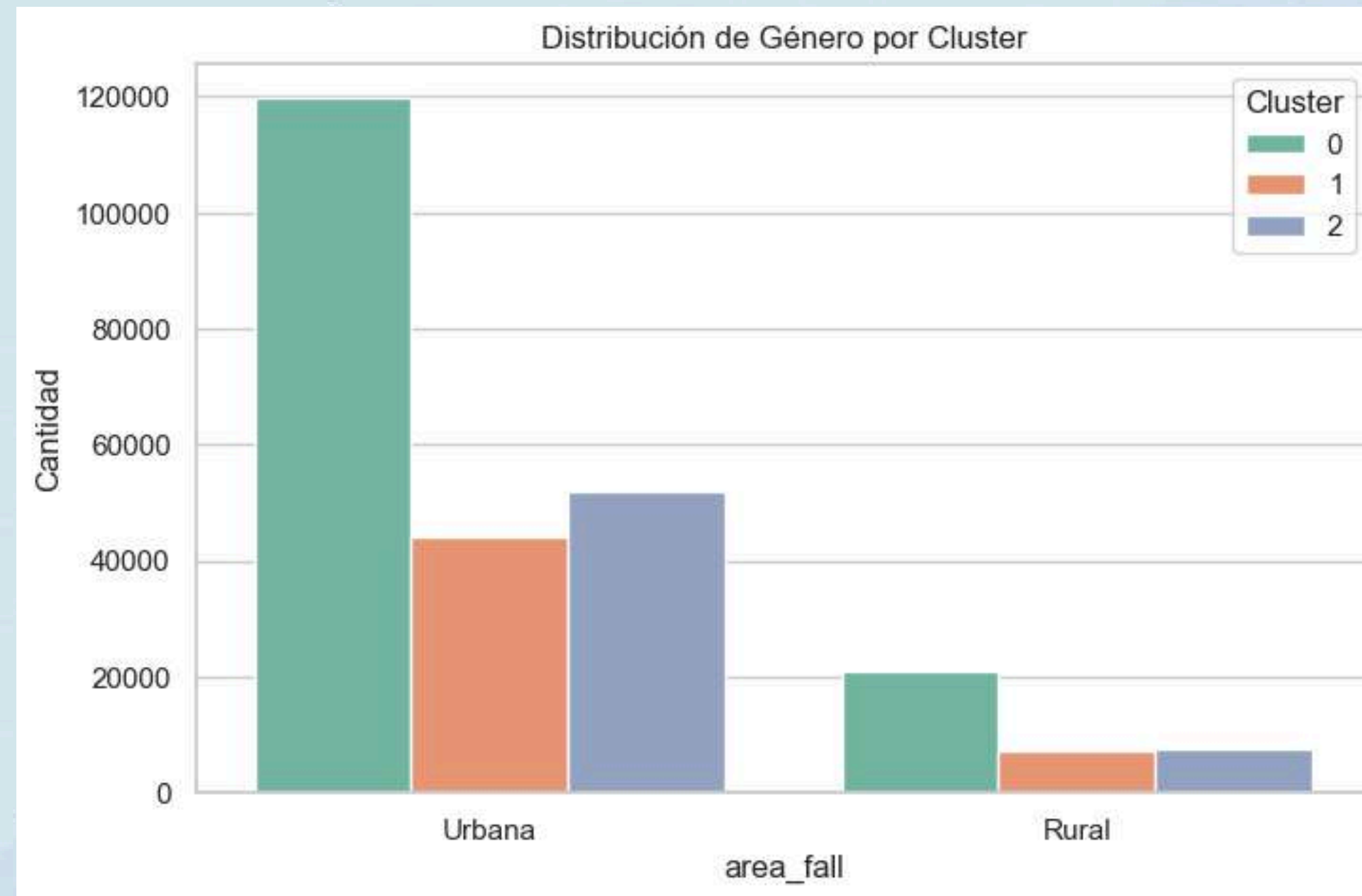
# RESULTADOS K=3

Distribución por Causa de fallecimiento :



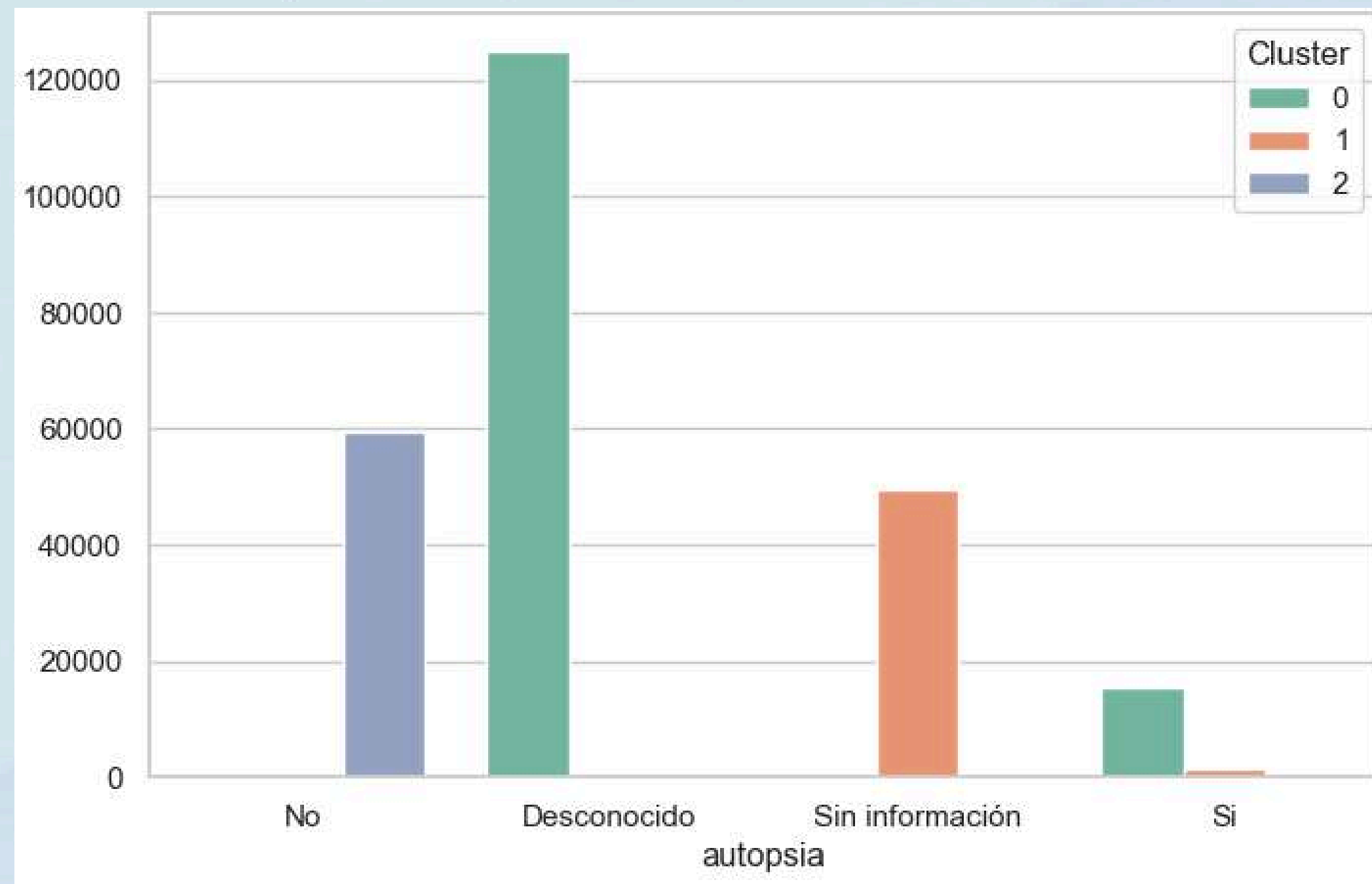
# RESULTADOS K=3

Distribución por área de fallecimiento :




# RESULTADOS K=3

Distribución por autopsia:





# CONCLUSIONES

The background of the slide features a series of overlapping, wavy, organic shapes in various shades of blue and green, creating a sense of depth and movement. The colors transition from a light, airy blue on the left to a more vibrant green on the right, with darker blue and green tones filling the lower and right portions of the frame.

# GRACIAS POR SU ATENCIÓN



Collaborative Online International Learning

Equipo 4

