

Segmentación de clientes



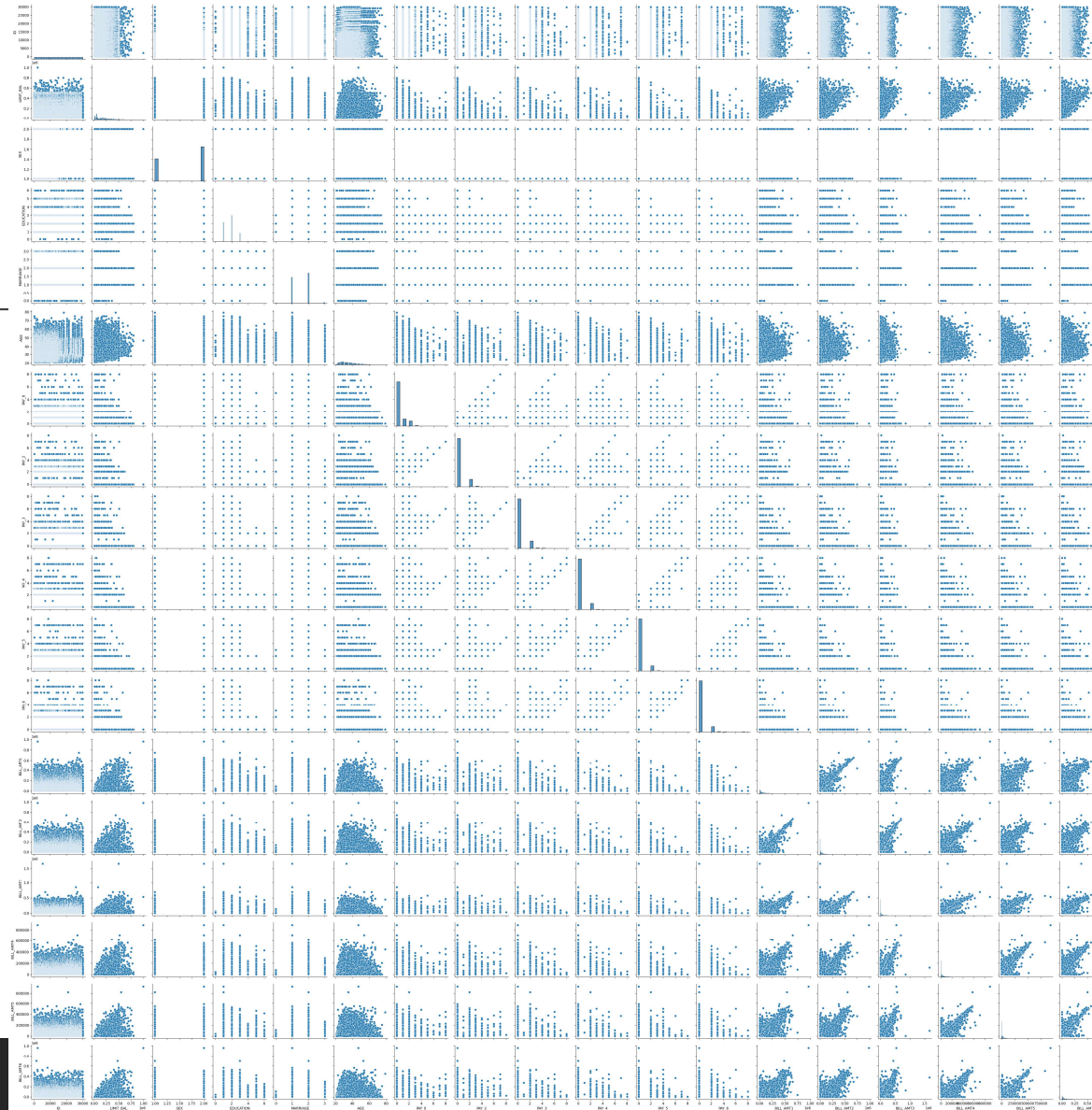
Conjunto de datos

El de ID se omite

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	BILL_AMT1
1	20000	2	2	1	24	3913
2	120000	2	2	2	26	2682
3	90000	2	2	2	34	29239
4	50000	2	2	1	37	46990
5	50000	1	2	1	57	8617
6	50000	1	1	2	37	64400
7	5.00E+05	1	1	2	29	367965
8	1.00E+05	2	2	2	23	11876

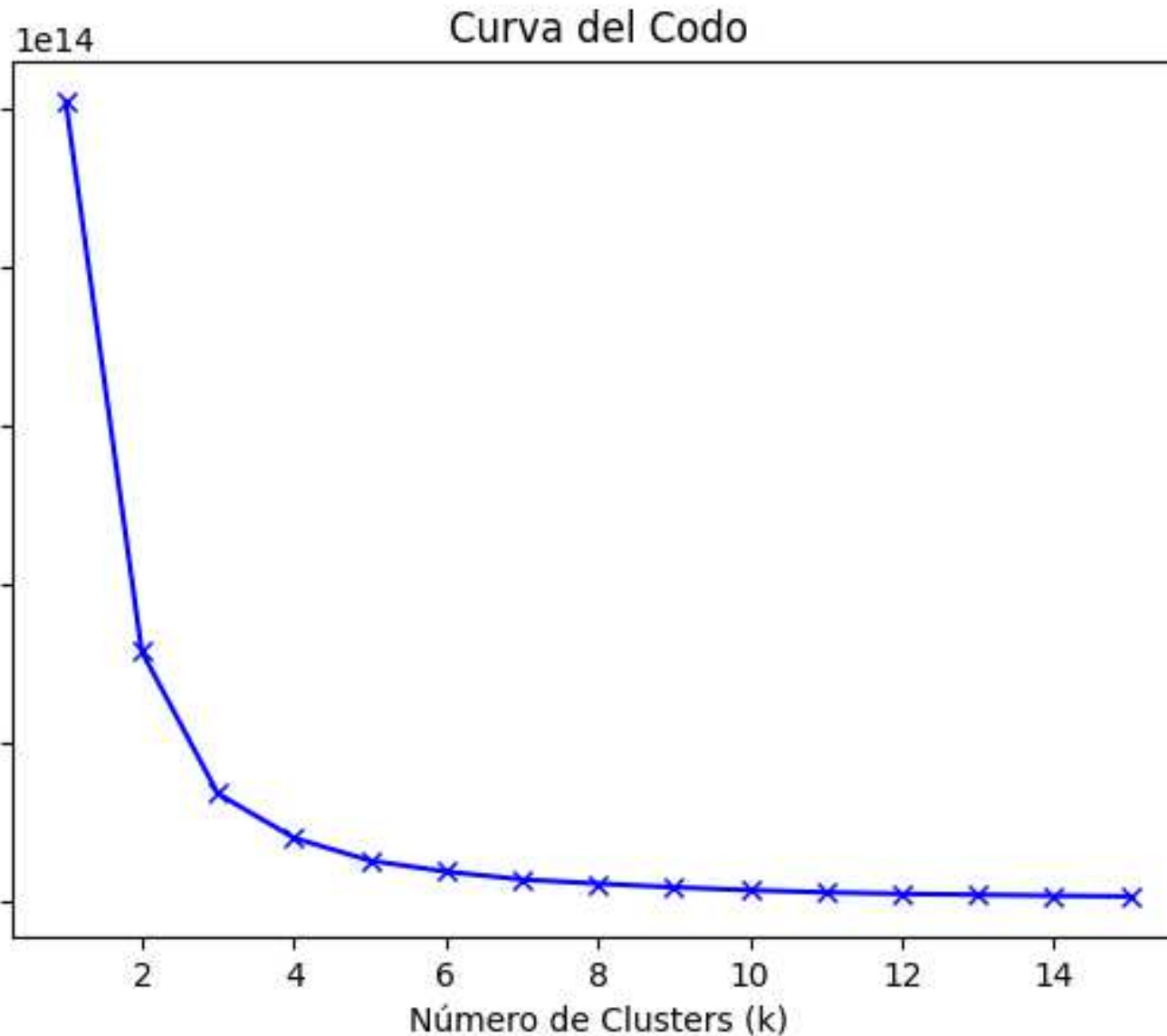
Preprocesamiento

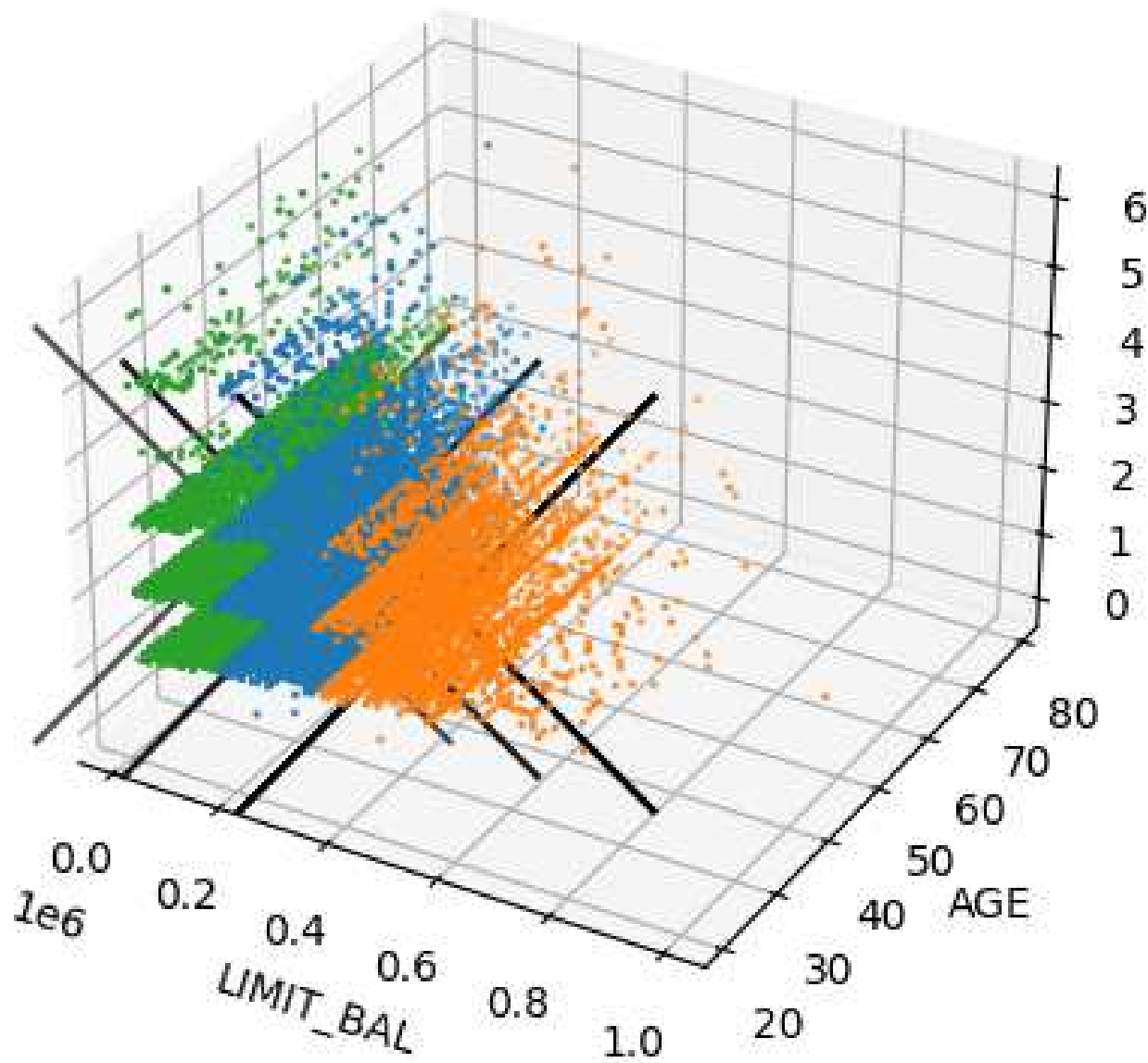
Al principio tenía
24 columnas,
pero muchas se
repetían entre
ellas



Selección de clusters

Mediante el método del codo, se ve claramente que el número de clusters indicado es 3





Visualizacion
de
resultados

```

az = plt.axes(projection='3d')
for i in range(3):
    az.scatter3D(X2[yhat2==i,0], X2[yhat2==i,4], X2[yhat2==i,2], label='Cluster '+str(i+1), zorder=1, s=1,)
    az.scatter3D(kmeans_2.cluster_centers[:,0], kmeans_2.cluster_centers[:,4], kmeans_2.cluster_centers[:,2], color='black', marker='x', label='Centroides', s=10000, zorder=50000)
ax.legend()
plt.xlabel('LIMIT_BAL')
plt.ylabel('AGE')
#ax.set_zlabel('EDUCATION')
plt.title('Clusters de los datos')
plt.show()

```

Clusters de los datos



Codigo ;)