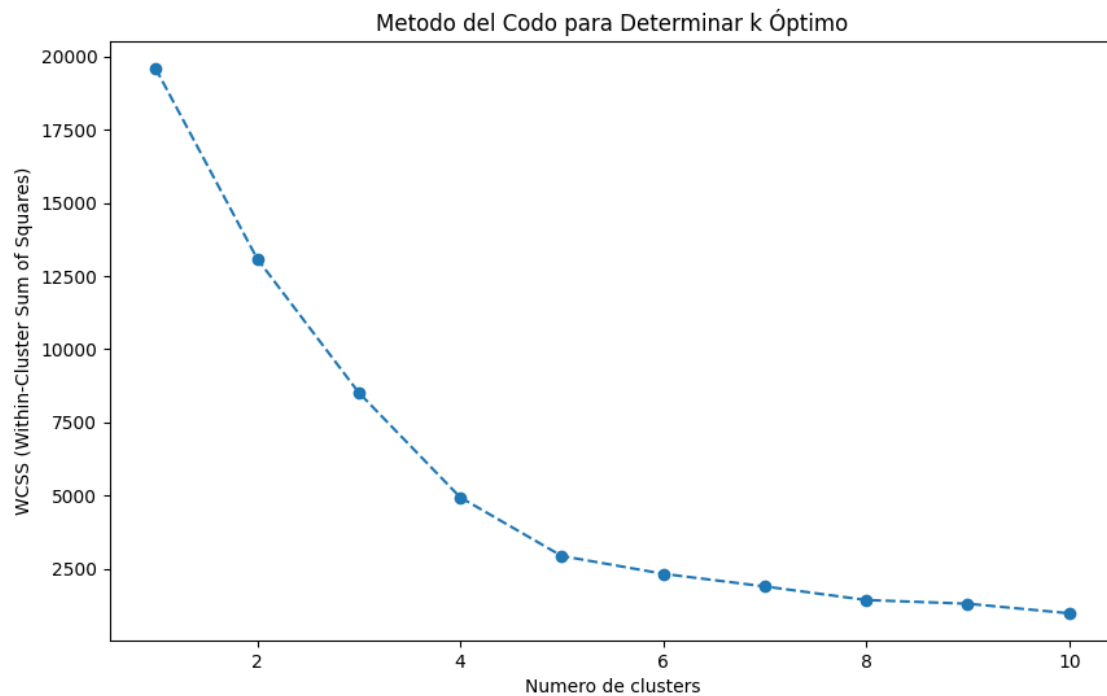
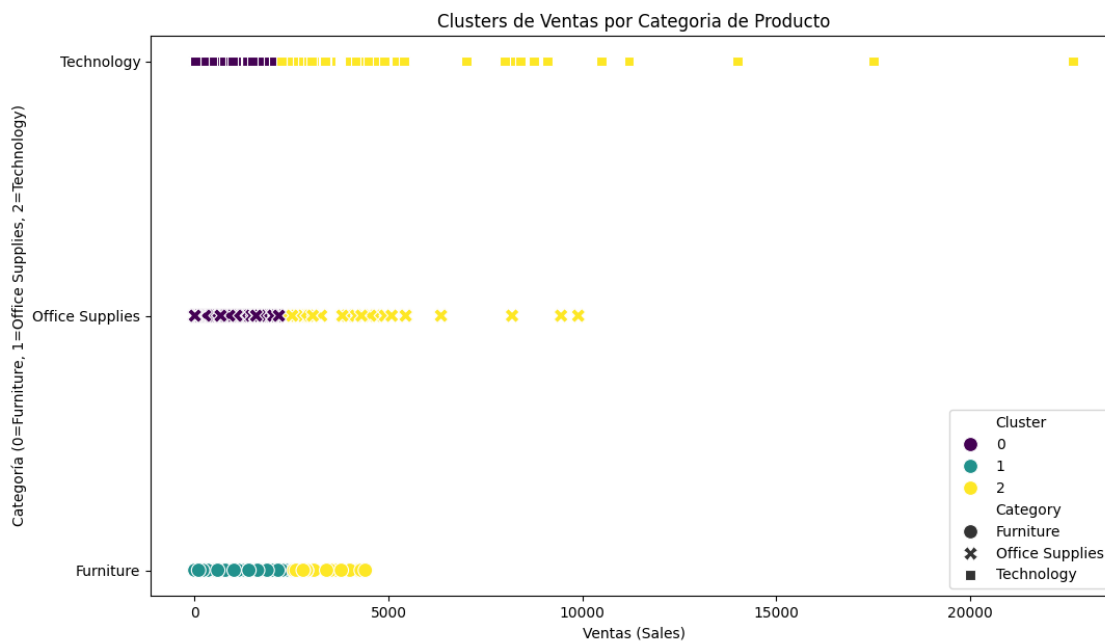


Data clustering

En esta practica, use los parametros 'Sales' y 'Category' asignándole un valor numérico a cada categoria para el modelo, estandaricé los datos usando StandardScaler para asegurar que todas las variables tuvieran el mismo peso, y determiné el número óptimo de clusters (k=3) mediante el método del codo, graficando la suma de cuadrados intra-cluster (WCSS) para diferentes valores de k. Aplicando K-Means con k=3.



Generé los clusters y los visualicé en un scatter plot que muestra las ventas contra las categorías numéricas, diferenciando cada cluster por color y cada categoría por marcadores.



Analicé la distribución de los clusters y las estadísticas por grupo, identificando patrones como transacciones de bajo valor en todas las categorías (Cluster 0), ventas altas en Technology y Furniture (Cluster 1), y transacciones medias predominantes en Office Supplies (Cluster 2).

Distribucion de clusters:

Cluster	
0	7633
1	2056
2	111

Name: count, dtype: int64

Distribucion de categorias por cluster:

Category	Furniture	Office Supplies	Technology
Cluster			
0	0	5882	1751
1	2056	0	0
2	22	27	62

Analisis estadistico por cluster:

				Sales		Category_Num			
				mean	median	min	max	count	mean
Cluster									
0				147.627459	39.992	0.444	2177.584	7633	1.229399
1				319.734459	180.784	1.892	2453.430	2056	0.000000
2				4300.201297	3083.430	2239.936	22638.480	111	1.360360

El análisis sugiere que las ventas en el Cluster 0 tiene las ventas promedio más bajas (147.63) y un rango amplio (0.444 a 2,177.584), lo que sugiere una variedad de productos con valores moderados mientras que Cluster 1 muestra ventas promedio más altas que el Cluster 0 (319.73) y un rango similar (1.892 a 2,453.430), indicando productos de muebles con un valor más alto y Cluster 2 destaca por ventas promedio muy altas (4,300.20) y un rango extremo (2,239.936 a 22,638.480), lo que sugiere que este cluster agrupa productos de lujo o de alto valor, posiblemente tecnología premium.