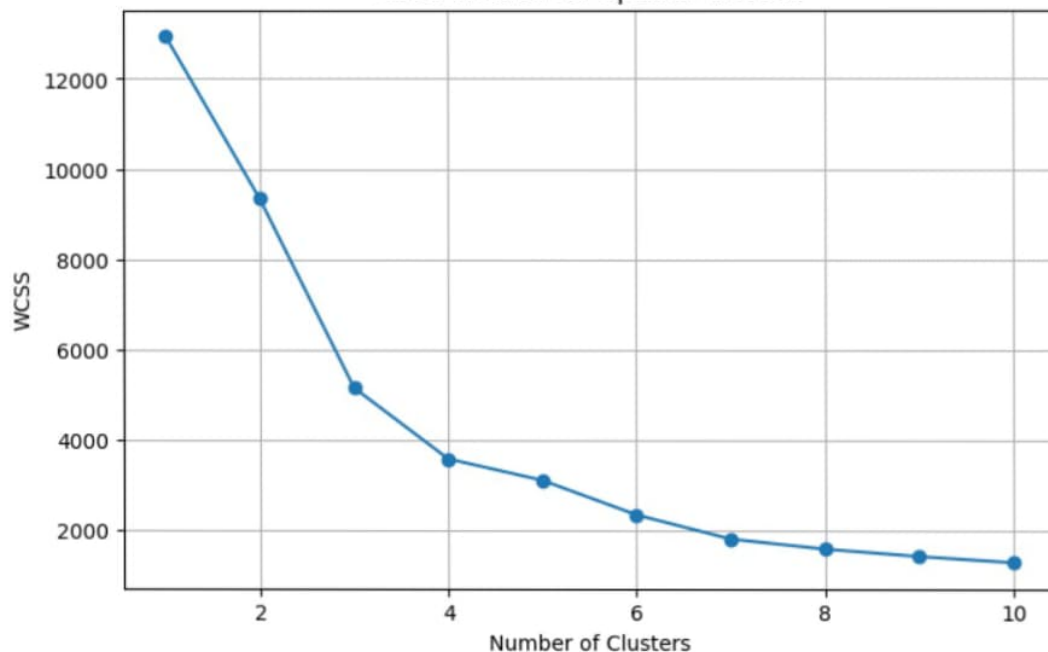


Choose files online_retail_II.xlsx
online_retail_II.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 45622278 bytes, last modified: 11/07/2025 - 100% done
Saving online_retail_II.xlsx to online_retail_II.xlsx
Columns: ['invoice', 'stockcode', 'description', 'quantity', 'invoicedate', 'price', 'customerid', 'country']
/tmp/ipython-input-1-1398629217.py:35: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

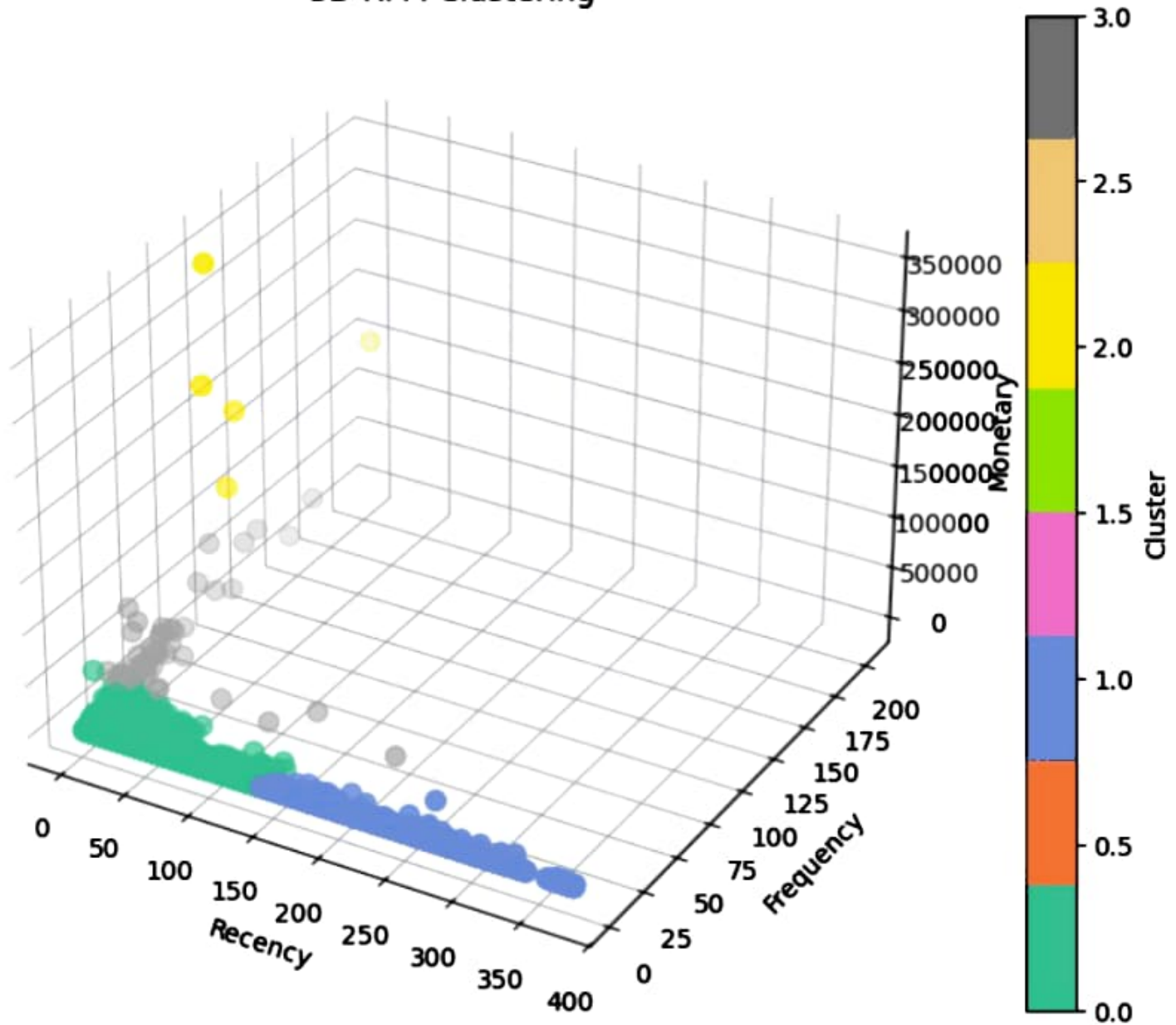
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['totalprice'] = df['quantity'] * df['price']
/tmp/ipython-input-1-1398629217.py:38: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df['invoicedate'] = pd.to_datetime(df['invoicedate'])

Elbow Method for Optimal Clusters



3D RFM Clustering



Sample of RFM Clusters:

	CustomerID	Recency	Frequency	Monetary	Cluster
0	12346.0	165	11	372.86	1
1	12347.0	3	2	1323.32	0
2	12348.0	74	1	222.16	0
3	12349.0	43	3	2671.14	0
4	12351.0	11	1	300.93	0

2D RFM Segmentation

