Automated Data Analysis Report

Table of Contents

1. Clustering Results

Best Parameters: {'epsilon': 4.180812541640952, 'min_samples': 4, 'silhouette': 0.5462857069724237},
Best Silhouette Score: 0.546

2. ANOVA Results

Results for capital-gain: F-value = 136807.398, P-value = 0.000
Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05
== group1 group2 meandiff p-adj lower upper reject ---------- -1 2 13.3387 0.0 13.1543 13.5231 True -1 3 1.0125 0.0 0.8339 1.191 True -1 4 -0.0 1.0 -0.1782 0.1782 False -1 5 -0.0 1.0 -0.1788 0.1788 False -1 6 0.0 1.0 -0.345 0.345 False 2 3 -12.3262 0.0 -12.3752 -12.2773 True 2 4 -13.3387 0.0 -13.3863 -13.2911 True 2 5 -13.3387 0.0 -13.3886 -13.2888 True 2 6 -13.3387 0.0 -13.6379 -13.0395 True 3 4 -1.0125 0.0 -1.025 -1.0 True 3 5 -1.0125 0.0 -1.0321 -0.9929 True 3 6 -1.0125 0.0 -1.3082 -0.7168 True 4 5 0.0 1.0 -0.016 0.016 False 4 6 0.0 1.0 -0.2955 0.2955 False 5 6 0.0 1.0 -0.2958 0.2958 False ----------

Results for capital-loss: F-value = 175887.433, P-value = 0.000
Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05
== group1 group2 meandiff p-adj lower upper reject ---------- -1 2 -6.3998 0.0 -6.5632 -6.2365 True -1 3 -6.3998 0.0 -6.5581 -6.2416 True -1 4 -6.3998 0.0 -6.5578 -6.2419 True -1 5 -1.8181 0.0 -1.9765 -1.6596 True -1 6 3.0551 0.0 2.7494 3.3609 True 2 3 0.0 1.0 -0.0434 0.0434 False 2 4 -0.0 1.0 -0.0422 0.0422 False 2 5 4.5818 0.0 4.5375 4.626 True 2 6 9.455 0.0 9.1898 9.7201 True 3 4 -0.0 1.0 -0.0111 0.0111 False 3 5 4.5818 0.0 4.5644 4.5992 True 3 6 9.455 0.0 9.1929 9.717 True 4 5 4.5818 0.0 4.5676 4.5959 True 4 6 9.455 0.0 9.1931 9.7168 True 5 6 4.8732 0.0 4.611 5.1354 True ----------

Results for positive_capital_gain: F-value = inf, P-value = 0.000
Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05
== group1 group2 meandiff p-adj lower upper reject ---------- -1 2 3.5999 0.0 3.5999 3.5999 True -1 3 3.5999 0.0 3.5999 3.5999 True -1 4 -0.0 0.0 -0.0 -0.0 True -1 5 -0.0 0.9999 -0.0 0.0 False -1 6 -0.0 1.0 -0.0 0.0 False 2 3 0.0 0.0 0.0 0.0 True 2 4 -3.5999 0.0 -3.5999 -3.5999 True 2 5 -3.5999 0.0 -3.5999 -3.5999 True 2 6 -3.5999 0.0 -3.5999 -3.5999 True 3 4 -3.5999 0.0 -3.5999 -3.5999 True 3 5 -3.5999 0.0 -3.5999 -3.5999 True 3 6 -3.5999 0.0 -3.5999 -3.5999 True 4 5 0.0 0.0 0.0 0.0 True 4 6 0.0 0.0127 0.0 0.0 True 5 6 0.0 1.0 -0.0 0.0 False ----------

Results for positive_capital_loss: F-value = inf, P-value = 0.000
Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05
== group1 group2 meandiff p-adj lower upper reject ---------- -1 2 -4.6989 0.0 -4.6989 -4.6989 True -1 3 -4.6989 0.0 -4.6989 -4.6989 True -1 4 -4.6989 0.0 -4.6989 -4.6989 True -1 5 -0.0 0.0 -0.0 -0.0 True -1 6 -0.0 1.0 -0.0 0.0 False 2 3 -0.0 0.0001 -0.0 -0.0 True 2 4 -0.0 0.0 -0.0 -0.0 True 2 5 4.6989 0.0 4.6989 4.6989 True 2 6 4.6989 0.0 4.6989 4.6989 True 3 4 -0.0 0.0 -0.0 -0.0 True 3 5 4.6989 0.0 4.6989 4.6989 True 3 6 4.6989 0.0 4.6989 4.6989 True 4 5 4.6989 0.0 4.6989 4.6989 True 4 6 4.6989 0.0 4.6989 4.6989 True 5 6 0.0 0.0 0.0 0.0 True ----------

Results for age_education_interaction: F-value = 264.158, P-value = 0.000
Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05
== group1 group2 meandiff p-adj

lower upper reject ---------- -1 2 0.0978 0.9996 -0.773 0.9686 False -1 3 -0.621 0.2879 -1.4644 0.2223 False -1 4 -1.2044 0.0006 -2.046 -0.3628 True -1 5 -0.723 0.1428 -1.5677 0.1216 False -1 6 -1.7138 0.0326 -3.3433 -0.0843 True 2 3 -0.7188 0.0 -0.95 -0.4876 True 2 4 -1.3021 0.0 -1.527 -1.0773 True 2 5 -0.8208 0.0 -1.0567 -0.5849 True 2 6 -1.8115 0.0035 -3.2248 -0.3983 True 3 4 -0.5833 0.0 -0.6424 -0.5243 True 3 5 -0.102 0.0212 -0.1946 -0.0093 True 3 6 -1.0927 0.2239 -2.4893 0.3038 False 4 5 0.4814 0.0 0.406 0.5567 True 4 6 -0.5094 0.9044 -1.9049 0.8861 False 5 6 -0.9908 0.3304 -2.3881 0.4066 False ----------

3. Cluster Variability

	antecedent support	consequent support	support	confidence	\
count	107.000000	107.000000	107.000000	107.000000	
mean	0.083740	0.243636	0.048353	0.620479	
std	0.045382	0.086054	0.023328	0.156351	
min	0.029992	0.125719	0.027116	0.420732	
25%	0.045193	0.182005	0.030403	0.467670	
50%	0.067379	0.202958	0.036565	0.577586	
75%	0.112572	0.256368	0.057313	0.780303	
max	0.202958	0.410846	0.110107	0.917808	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	107.000000	107.000000	107.000000	107.000000	107.000000	
mean	2.742092	0.028951	2.434190	0.654361	2.906542	
std	0.921510	0.015991	1.158408	0.125715	0.292443	
min	1.794985	0.013956	1.327059	0.488225	2.000000	
25%	2.034365	0.017920	1.459066	0.545493	3.000000	
50%	2.285053	0.023929	1.895184	0.600917	3.000000	
75%	3.255464	0.029946	3.161177	0.781146	3.000000	
max	5.463524	0.074919	7.168036	0.872680	3.000000	

		coverage
	count	107.000000
	mean	0.083740
	std	0.045382
	min	0.029992
	25%	0.045193
	50%	0.067379
	75%	0.112572
3	max	0.202958

	antecedent support	consequent support	support	confidence	\
count	84.000000	84.000000	84.000000	84.000000	
mean	0.104050	0.262216	0.058999	0.614567	
std	0.069485	0.080317	0.035814	0.187857	
min	0.029878	0.145305	0.025305	0.395207	
25%	0.051265	0.201912	0.031340	0.437138	
50%	0.076101	0.223768	0.043524	0.552176	
75%	0.139138	0.344592	0.068107	0.752674	
max	0.344592	0.437263	0.142643	0.991416	

		lift	leverage	conviction	zhangs_metric	total_items	coverage
	count	84.000000	84.000000	84.000000	84.000000	84.000000	84.000000
	mean	2.386512	0.033274	5.248958	0.635994	2.904762	0.104050
	std	0.455398	0.020977	12.523368	0.099000	0.295307	0.069485
	min	1.747838	0.012210	1.293575	0.462519	2.000000	0.029878
	25%	2.039033	0.018263	1.447832	0.567646	3.000000	0.051265
	50%	2.343404	0.024154	1.673277	0.619573	3.000000	0.076101
	75%	2.623155	0.041694	3.012710	0.709479	3.000000	0.139138
4	max	4.146542	0.087095	65.558819	0.931602	3.000000	0.344592

	antecedent support	consequent support	support	confidence	\
count	139.000000	139.000000	139.000000	139.000000	
mean	0.081087	0.257893	0.051181	0.687153	
std	0.051891	0.081795	0.027665	0.171889	
min	0.027663	0.112725	0.026279	0.428571	
25%	0.043568	0.200207	0.032503	0.537769	
50%	0.060858	0.219917	0.038036	0.643443	
75%	0.102351	0.298755	0.062241	0.848364	
max	0.219917	0.396957	0.124481	1.000000	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	139.000000	139.000000	139.000000	139.000000	139.000000	
mean	2.816366	0.031620	inf	0.673401	2.928058	
std	0.878689	0.018009	NaN	0.106102	0.259327	
min	1.952271	0.013625	1.397649	0.522678	2.000000	
25%	2.163704	0.019085	1.677896	0.585482	3.000000	
50%	2.466681	0.025496	2.155145	0.634110	3.000000	
75%	2.933749	0.037959	4.521399	0.795888	3.000000	
max	5.885226	0.080376	inf	0.867896	3.000000	

		coverage
	count	139.000000
	mean	0.081087
	std	0.051891
	min	0.027663
	25%	0.043568
	50%	0.060858
	75%	0.102351
5	max	0.219917

	antecedent support	consequent support	support	confidence	\
count	24.000000	24.000000	24.000000	24.000000	
mean	0.048925	0.138172	0.037634	0.785230	
std	0.010406	0.055561	0.005268	0.099019	
min	0.032258	0.077419	0.032258	0.666667	
25%	0.043548	0.108065	0.032258	0.714286	
50%	0.051613	0.129032	0.038710	0.750000	
75%	0.053226	0.148387	0.038710	0.833333	
max	0.070968	0.296774	0.051613	1.000000	

		lift	leverage	conviction	zhangs_metric	total_items	coverage
	count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
	mean	6.266899	0.031079	inf	0.868886	2.875000	0.048925
	std	1.880070	0.004561	NaN	0.056527	0.337832	0.010406
	min	3.369565	0.022685	2.748387	0.726667	2.000000	0.032258
	25%	5.054348	0.028377	3.203226	0.845805	3.000000	0.043548
	50%	5.794449	0.031384	3.470968	0.873016	3.000000	0.051613
	75%	7.799679	0.033933	5.293548	0.906081	3.000000	0.053226
2	max	10.219780	0.042456	inf	0.944820	3.000000	0.070968

	antecedent support	consequent support	support	confidence	\
count	3.260000e+02	3.260000e+02	3.260000e+02	326.0	
mean	9.090909e-02	9.090909e-02	9.090909e-02	1.0	
std	1.389912e-17	1.389912e-17	1.389912e-17	0.0	
min	9.090909e-02	9.090909e-02	9.090909e-02	1.0	
25%	9.090909e-02	9.090909e-02	9.090909e-02	1.0	
50%	9.090909e-02	9.090909e-02	9.090909e-02	1.0	
75%	9.090909e-02	9.090909e-02	9.090909e-02	1.0	
max	9.090909e-02	9.090909e-02	9.090909e-02	1.0	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	326.0	3.260000e+02	326.0	326.0	326.000000	
mean	11.0	8.264463e-02	inf	1.0	2.981595	
std	0.0	1.389912e-17	NaN	0.0	0.134617	
min	11.0	8.264463e-02	inf	1.0	2.000000	
25%	11.0	8.264463e-02	NaN	1.0	3.000000	
50%	11.0	8.264463e-02	NaN	1.0	3.000000	
75%	11.0	8.264463e-02	NaN	1.0	3.000000	
max	11.0	8.264463e-02	inf	1.0	3.000000	

		coverage
	count	3.260000e+02
	mean	9.090909e-02
	std	1.389912e-17
	min	9.090909e-02
	25%	9.090909e-02
	50%	9.090909e-02
	75%	9.090909e-02
-1	max	9.090909e-02

	antecedent support	consequent support	support	confidence	lift	\
count	212.00	212.00	212.00	212.0	212.0	
mean	0.25	0.25	0.25	1.0	4.0	
std	0.00	0.00	0.00	0.0	0.0	
min	0.25	0.25	0.25	1.0	4.0	
25%	0.25	0.25	0.25	1.0	4.0	
50%	0.25	0.25	0.25	1.0	4.0	
75%	0.25	0.25	0.25	1.0	4.0	
max	0.25	0.25	0.25	1.0	4.0	

	leverage	conviction	zhangs_metric	total_items	coverage
count	212.0000	212.0	212.0	212.000000	212.00
mean	0.1875	inf	1.0	2.933962	0.25
std	0.0000	NaN	0.0	0.248936	0.00
min	0.1875	inf	1.0	2.000000	0.25
25%	0.1875	NaN	1.0	3.000000	0.25
50%	0.1875	NaN	1.0	3.000000	0.25
75%	0.1875	NaN	1.0	3.000000	0.25
6 max	0.1875	inf	1.0	3.000000	0.25

4. Rule Metrics Comparison

mean	std	min	25%	50%	75%
04794584438373	0.15635136856818543	0.42073170731707316	0.467670011148272	0.5775862068965517	0.78030303030
45667263102756	0.1878566850199141	0.395206527281999	0.43713793237028264	0.5521755153060461	0.75267441860
71530498216469	0.17188939956318752	0.42857142857142855	0.5377686472819216	0.6434426229508197	0.84836427939
52302789802789	0.09901864756375088	0.6666666666666666	0.7142857142857143	0.75	0.83333333333
1.0	0.0	1.0	1.0	1.0	1.0
1.0	0.0	1.0	1.0	1.0	1.0
70232405796826	0.1766041616799795	0.39602248162559445	0.4603789693920703	0.556979297558966	0.74713024282

5. Top Unique Rules per Cluster

Cluster 3:

Rule: frozenset({'education_Masters'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]', 'sex_Male'}) (Support: 0.060, Confidence: 0.590, Lift: 1.812)
Rule: frozenset({'workclass_Self-emp-inc', 'marital-status_Married-civ-spouse'}) -> frozenset({'occupation_aggregated_Exec-managerial'}) (Support: 0.028, Confidence: 0.507, Lift: 2.417)
Rule: frozenset({'workclass_Self-emp-inc', 'native_country_aggregated_United-States'}) -> frozenset({'occupation_aggregated_Exec-managerial'}) (Support: 0.029, Confidence: 0.483, Lift: 2.299)
Rule: frozenset({'workclass_Self-emp-inc', 'sex_Male'}) -> frozenset({'occupation_aggregated_Exec-managerial'}) (Support: 0.029, Confidence: 0.480, Lift: 2.285)
Rule: frozenset({'workclass_Self-emp-inc', 'race_White'}) -> frozenset({'occupation_aggregated_Exec-managerial'}) (Support: 0.030, Confidence: 0.477, Lift: 2.273)

Cluster 4:

Rule: frozenset({'relationship_Wife'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.043, Confidence: 0.991, Lift: 2.267)
Rule: frozenset({'relationship_Wife', 'race_White'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.036, Confidence: 0.991, Lift: 2.266)
Rule: frozenset({'relationship_Wife', 'workclass_Private'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.030, Confidence: 0.990, Lift: 2.265)
Rule: frozenset({'relationship_Wife', 'native_country_aggregated_United-States'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.037, Confidence: 0.990, Lift: 2.264)
Rule: frozenset({'hours_per_week_binned_21-30', 'relationship_Own-child'}) -> frozenset({'marital-status_Never-married'}) (Support: 0.029, Confidence: 0.956, Lift: 2.775)

Cluster 5:

Rule: frozenset({'education_Doctorate', 'native_country_aggregated_United-States'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.028, Confidence: 1.000, Lift: 2.519)
Rule: frozenset({'education_Doctorate', 'race_White'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.030, Confidence: 1.000, Lift: 2.519)
Rule: frozenset({'Cluster_(4.0, 6.0]', 'education_Doctorate'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.033, Confidence: 0.979, Lift: 2.467)
Rule: frozenset({'education_Doctorate'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.033, Confidence: 0.979, Lift: 2.467)
Rule: frozenset({'education_Doctorate'}) -> frozenset({'Cluster_(4.0, 6.0]', 'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.033, Confidence: 0.979, Lift: 2.467)

Cluster 2:

Rule: frozenset({'education_HS-grad', 'age_education_interaction_(494.0, 1350.0]'}) -> frozenset({'hours_per_week_binned_41-50'}) (Support: 0.032, Confidence: 1.000, Lift: 3.370)
Rule: frozenset({'occupation_aggregated_Prof-specialty', 'race_Asian-Pac-Islander'}) -> frozenset({'education_Prof-school'}) (Support: 0.032, Confidence: 1.000, Lift: 3.370)
Rule: frozenset({'marital-status_Married-civ-spouse', 'race_Asian-Pac-Islander'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.039, Confidence: 0.857, Lift: 10.220)
Rule: frozenset({'workclass_Private', 'race_Asian-Pac-Islander'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.032, Confidence: 0.833, Lift: 9.936)
Rule: frozenset({'occupation_aggregated_Craft-repair', 'marital-status_Married-civ-spouse'}) -> frozenset({'education_HS-grad'}) (Support: 0.032, Confidence: 0.833, Lift: 5.616)

Cluster -1:

Rule: frozenset({'relationship_Unmarried', 'marital-status_Never-married'}) -> frozenset({'education_Some-college'}) (Support: 0.091, Confidence: 1.000, Lift: 11.000)
Rule: frozenset({'age_education_interaction_(494.0, 1350.0]', 'marital-status_Never-married'}) -> frozenset({'hours_per_week_binned_21-30'}) (Support: 0.091, Confidence: 1.000, Lift: 11.000)
Rule: frozenset({'hours_per_week_binned_21-30'}) -> frozenset({'relationship_Other-relative', 'native_country_aggregated_United-States'}) (Support: 0.091, Confidence: 1.000, Lift: 11.000)

Rule: frozenset({'relationship_Other-relative', 'native_country_aggregated_United-States'}) -> frozenset({'hours_per_week_binned_21-30'}) (Support: 0.091, Confidence: 1.000, Lift: 11.000)
Rule: frozenset({'hours_per_week_binned_21-30'}) -> frozenset({'relationship_Other-relative', 'occupation_aggregated_Prof-specialty'}) (Support: 0.091, Confidence: 1.000, Lift: 11.000)

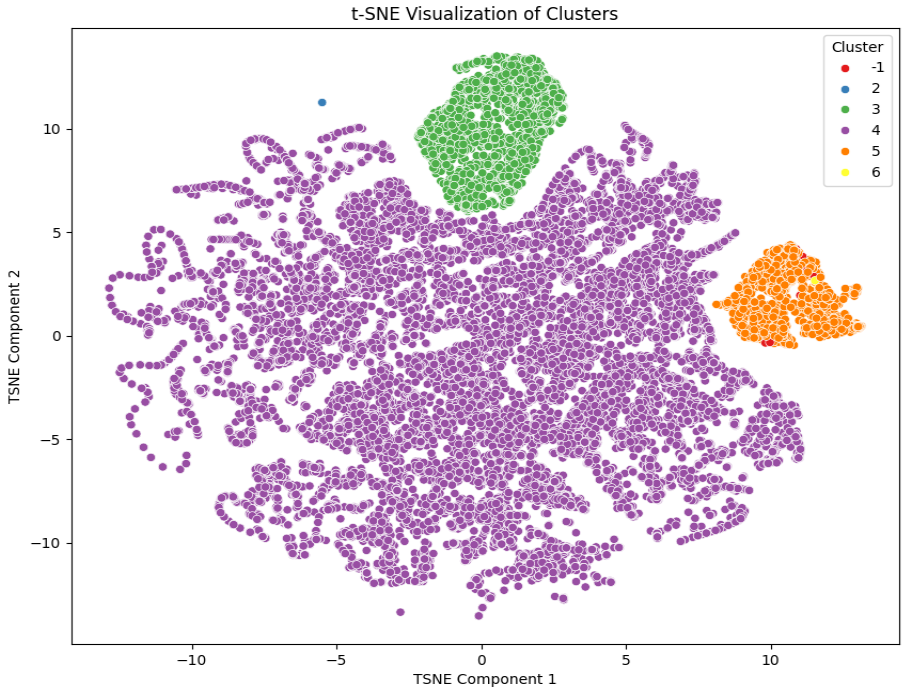
Cluster 6:

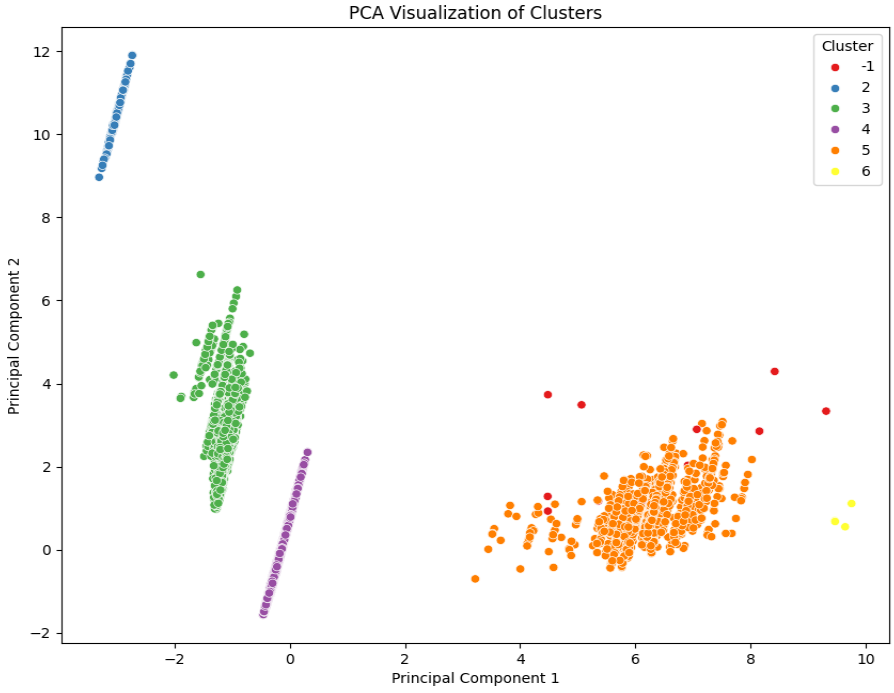
Rule: frozenset({'education_HS-grad'}) -> frozenset({'age_education_interaction_(261.0, 369.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'age_education_interaction_(369.0, 494.0]', 'relationship_Own-child'}) -> frozenset({'education_Some-college'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'age_education_interaction_(369.0, 494.0]'}) -> frozenset({'education_Some-college', 'relationship_Own-child'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'relationship_Own-child'}) -> frozenset({'education_Some-college', 'age_education_interaction_(369.0, 494.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'education_Some-college', 'marital-status_Separated'}) -> frozenset({'age_education_interaction_(369.0, 494.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)

6. Top 10 Common Rules Sorted by Absolute Coverage Difference

Rule: frozenset({'relationship_Own-child', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.292)
Rule: frozenset({'relationship_Own-child', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.275)
Rule: frozenset({'relationship_Own-child', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.271)
Rule: frozenset({'relationship_Not-in-family', 'native_country_aggregated_United-States', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.265)
Rule: frozenset({'marital-status_Never-married', 'relationship_Own-child', 'native_country_aggregated_United-States'}) (Abs Coverage Difference: 0.260)
Rule: frozenset({'marital-status_Never-married', 'relationship_Own-child', 'native_country_aggregated_United-States'}) (Abs Coverage Difference: 0.256)
Rule: frozenset({'relationship_Own-child', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.254)
Rule: frozenset({'relationship_Not-in-family', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.252)
Rule: frozenset({'relationship_Own-child', 'native_country_aggregated_United-States', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.242)
Rule: frozenset({'marital-status_Never-married', 'relationship_Own-child', 'native_country_aggregated_United-States'}) (Abs Coverage Difference: 0.241)

7. Cluster Visualizations


t-SNE Visualization of Clusters
Cluster
-1
2
3
4
5
6
TSNE Component 2
TSNE Component 1
10
5
0
−5
−10
−10
−5
0
5
10


PCA Visualization of Clusters
Cluster
-1
2
3
4
5
6
Principal Component 2
Principal Component 1
12
10
8
6
4
2
0
−2
−2
0
2
4
6
8
10