Automated Data Analysis Report

Table of Contents

1. Clustering Results 2. ANOVA Results 3. Cluster Variability 4. Rule Metrics Comparison 5. Top Unique Rules per Cluster 6. Top 10 Common Rules 7. Cluster Visualizations

1. Clustering Results

Best Parameters: {'epsilon': 2.393369097964607, 'min_samples': 6, 'silhouette': 0.33287232534725236}, Best Silhouette Score: 0.333

2. ANOVA Results

Results for wife_religion: F-value = 13092.125, 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 1 1.6032 0.0 1.4889 1.7174 True -1 3 1.6032 0.0 1.4953 1.7111 True -1 5 -1.2024 0.0 -1.3145 -1.0903 True -1 6 -1.2024 0.0 -1.3201 -1.0847 True -1 7 1.6032 0.0 1.4933 1.713 True -1 8 1.6032 0.0 1.3945 1.8119 True 1 3 -0.0 1.0 -0.0421 0.0421 False 1 5 -2.8055 0.0 -2.8575 -2.7536 True 1 6 -2.8055 0.0 -2.8687 -2.7424 True 1 7 -0.0 1.0 -0.0468 0.0468 False 1 8 0.0 1.0 -0.1835 0.1835 False 3 5 -2.8055 0.0 -2.8415 -2.7696 True 3 6 -2.8055 0.0 -2.8563 -2.7548 True 3 7 0.0 1.0 -0.0281 0.0281 False 3 8 0.0 1.0 -0.1797 0.1797 False 5 6 -0.0 1.0 -0.0592 0.0592 False 5 7 2.8055 0.0 2.7642 2.8469 True 5 8 2.8055 0.0 2.6233 2.9877 True 6 7 2.8055 0.0 2.7508 2.8603 True 6 8 2.8055 0.0 2.6198 2.9912 True 7 8 0.0 1.0 -0.1808 0.1808 False ----------

Results for wife_working: F-value = 2977.025, 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 1 0.9581 0.0 0.7256 1.1905 True -1 3 1.4836 0.0 1.264 1.7032 True -1 5 1.4836 0.0 1.2555 1.7117 True -1 6 -0.8242 0.0 -1.0637 -0.5847 True -1 7 -0.8242 0.0 -1.0477 -0.6007 True -1 8 -0.8242 0.0 -1.2489 -0.3996 True 1 3 0.5255 0.0 0.4399 0.6112 True 1 5 0.5255 0.0 0.4199 0.6312 True 1 6 -1.7823 0.0 -1.9107 -1.6539 True 1 7 -1.7823 0.0 -1.8776 -1.687 True 1 8 -1.7823 0.0 -2.1557 -1.4089 True 3 5 -0.0 1.0 -0.0731 0.0731 False 3 6 -2.3078 0.0 -2.4111 -2.2045 True 3 7 -2.3078 0.0 -2.365 -2.2507 True 3 8 -2.3078 0.0 -2.6734 -1.9423 True 5 6 -2.3078 0.0 -2.4282 -2.1874 True 5 7 -2.3078 0.0 -2.392 -2.2236 True 5 8 -2.3078 0.0 -2.6786 -1.9371 True 6 7 -0.0 1.0 -0.1115 0.1115 False 6 8 -0.0 1.0 -0.3779 0.3779 False 7 8 0.0 1.0 -0.3679 0.3679 False ----------

Results for media_exposure: F-value = 6948.627, 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 1 1.6372 0.0 1.4817 1.7928 True -1 3 -2.183 0.0 -2.3299 -2.036 True -1 5 -2.183 0.0 -2.3356 -2.0303 True -1 6 -2.183 0.0 -2.3432 -2.0227 True -1 7 -2.183 0.0 -2.3325 -2.0334 True -1 8 -2.183 0.0 -2.4671 -1.8988 True 1 3 -3.8202 0.0 -3.8775 -3.7628 True 1 5 -3.8202 0.0 -3.8909 -3.7495 True 1 6 -3.8202 0.0 -3.9061 -3.7342 True 1 7 -3.8202 0.0 -3.884 -3.7564 True 1 8 -3.8202 0.0 -4.0701 -3.5703 True 3 5 -0.0 1.0 -0.0489 0.0489 False 3 6 -0.0 1.0 -0.0691 0.0691 False 3 7 -0.0 1.0 -0.0382 0.0382 False 3 8 -0.0 1.0 -0.2446 0.2446 False 5 6 0.0 1.0 -0.0806 0.0806 False 5 7 -0.0 1.0 -0.0563 0.0563 False 5 8 0.0 1.0 -0.2481 0.2481 False 6 7 -0.0 1.0 -0.0746 0.0746 False 6 8 0.0 1.0 -0.2529 0.2529 False 7 8 0.0 1.0 -0.2462 0.2462 False ----------

Results for age_children_interaction: F-value = 12.511, 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 1 -0.9782 0.0084 -1.8013 -0.155 True -1 3 -1.4548 0.0 -2.2323 -0.6772 True -1 5 -1.5288 0.0 -2.3366 -0.721 True -1 6 -1.6896 0.0 -2.5378 -0.8413 True -1 7 -1.6483 0.0 -2.4399 -0.8567 True -1 8 -2.4452 0.0 -3.949 -0.9413 True 1 3 -0.4766

0.0001 -0.7799 -0.1733 True 1 5 -0.5506 0.0003 -0.9247 -0.1765 True 1 6 -0.7114 0.0001 -1.1662 -0.2566 True 1 7 -0.6701 0.0 -1.0077 -0.3326 True 1 8 -1.467 0.0186 -2.7894 -0.1445 True 3 5 -0.074 0.9802 -0.3328 0.1848 False 3 6 -0.2348 0.484 -0.6007 0.1311 False 3 7 -0.1935 0.0715 -0.3959 0.0088 False 3 8 -0.9904 0.2651 -2.285 0.3042 False 5 6 -0.1608 0.9242 -0.5872 0.2656 False 5 7 -0.1195 0.9005 -0.4177 0.1786 False 5 8 -0.9164 0.3769 -2.2293 0.3966 False 6 7 0.0413 0.9999 -0.3534 0.436 False 6 8 -0.7556 0.6383 -2.0938 0.5826 False 7 8 -0.7969 0.5444 -2.0999 0.5061 False
---

Results for edu_interaction: F-value = 47.758, 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 1 -1.0374 0.0015 -1.8093 -0.2655 True -1 3 0.168 0.9937 -0.5611 0.8972 False -1 5 0.7759 0.0406 0.0184 1.5334 True -1 6 0.859 0.0246 0.0636 1.6544 True -1 7 0.2755 0.9295 -0.4668 1.0178 False -1 8 -0.9111 0.4755 -2.3213 0.4991 False 1 3 1.2054 0.0 0.921 1.4899 True 1 5 1.8133 0.0 1.4625 2.1641 True 1 6 1.8964 0.0 1.4699 2.3228 True 1 7 1.3129 0.0 0.9964 1.6294 True 1 8 0.1263 0.9999 -1.1138 1.3664 False 3 5 0.6079 0.0 0.3652 0.8505 True 3 6 0.6909 0.0 0.3478 1.034 True 3 7 0.1075 0.635 -0.0823 0.2972 False 3 8 -1.0791 0.1193 -2.2931 0.1348 False 5 6 0.0831 0.9964 -0.3168 0.4829 False 5 7 -0.5004 0.0 -0.78 -0.2208 True 5 8 -1.687 0.0011 -2.9182 -0.4558 True 6 7 -0.5835 0.0001 -0.9536 -0.2133 True 6 8 -1.7701 0.0007 -3.0249 -0.5152 True 7 8 -1.1866 0.0635 -2.4085 0.0353 False
---

3. Cluster Variability

		antecedent support	consequent support	support	confidence	lift	\
	count	8.000000	8.000000	8.000000	8.000000	8.000000	
	mean	0.166524	0.387272	0.096747	0.607547	1.566041	
	std	0.135460	0.047679	0.072409	0.105193	0.162729	
	min	0.075342	0.328767	0.047945	0.456311	1.387945	
	25%	0.086758	0.333333	0.053653	0.559852	1.465085	
	50%	0.104452	0.409247	0.061644	0.594118	1.496000	
	75%	0.171518	0.428082	0.104737	0.662126	1.715358	
	max	0.428082	0.428082	0.213470	0.769231	1.796923	

		leverage	conviction	zhangs_metric	total_items	coverage
	count	8.000000	8.000000	8.000000	8.00000	8.000000
	mean	0.033805	1.653912	0.435987	2.75000	0.166524
	std	0.023634	0.423086	0.096721	0.46291	0.135460
	min	0.013738	1.234589	0.311414	2.00000	0.075342
	25%	0.016914	1.404793	0.349562	2.75000	0.086758
	50%	0.026111	1.521366	0.465875	3.00000	0.104452
	75%	0.040476	1.738544	0.492161	3.00000	0.171518
3	max	0.070776	2.478311	0.579718	3.00000	0.428082

	antecedent support	consequent support	support	confidence	lift	\
count	7.000000	7.000000	7.000000	7.000000	7.000000	
mean	0.129919	0.383827	0.083019	0.652545	1.724631	
std	0.037538	0.062187	0.018742	0.092472	0.284717	
min	0.086792	0.283019	0.056604	0.590909	1.535205	
25%	0.094340	0.367925	0.071698	0.590909	1.535205	
50%	0.135849	0.384906	0.086792	0.638889	1.567644	
75%	0.166038	0.396226	0.098113	0.652174	1.797406	
max	0.166038	0.490566	0.098113	0.851852	2.304348	

	leverage	conviction	zhangs_metric	total_items	coverage
count	7.000000	7.000000	7.000000	7.000000	7.000000
mean	0.032719	1.931053	0.467281	2.857143	0.129919
std	0.003384	0.698417	0.075703	0.377964	0.037538
min	0.026145	1.503564	0.418030	2.000000	0.086792
25%	0.031734	1.503564	0.418030	3.000000	0.094340
50%	0.034204	1.640639	0.419024	3.000000	0.135849
75%	0.034204	1.963679	0.489009	3.000000	0.166038
7 max	0.036810	3.438679	0.619835	3.000000	0.166038

	antecedent support	consequent support	support	confidence	lift	\
count	0.0	0.0	0.0	0.0	0.0	
mean	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	NaN	

	leverage	conviction	zhangs_metric	total_items
count	0.0	0.0	0.0	0.0
mean	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN
1 max	NaN	NaN	NaN	NaN

		antecedent support	consequent support	support	confidence	\
	count	20.000000	20.000000	20.000000	20.000000	
	mean	0.100000	0.328571	0.092857	0.966667	
	std	0.048611	0.102073	0.033583	0.102598	
	min	0.071429	0.142857	0.071429	0.666667	
	25%	0.071429	0.214286	0.071429	1.000000	
	50%	0.071429	0.321429	0.071429	1.000000	
	75%	0.142857	0.428571	0.142857	1.000000	
	max	0.214286	0.428571	0.142857	1.000000	

		lift	leverage	conviction	zhangs_metric	total_items	coverage
	count	20.000000	20.000000	20.000000	20.000000	20.000000	20.000000
	mean	3.231667	0.061480	inf	0.736684	2.950000	0.100000
	std	1.046994	0.024690	NaN	0.120521	0.223607	0.048611
	min	2.333333	0.040816	2.142857	0.615385	2.000000	0.071429
	25%	2.333333	0.040816	NaN	0.615385	3.000000	0.071429
	50%	2.566667	0.053571	NaN	0.709790	3.000000	0.071429
	75%	4.666667	0.081633	NaN	0.846154	3.000000	0.142857
-1	max	4.666667	0.112245	inf	1.000000	3.000000	0.214286

		antecedent support	consequent support	support	confidence	lift	\
	count	0.0	0.0	0.0	0.0	0.0	
	mean	NaN	NaN	NaN	NaN	NaN	
	std	NaN	NaN	NaN	NaN	NaN	
	min	NaN	NaN	NaN	NaN	NaN	
	25%	NaN	NaN	NaN	NaN	NaN	
	50%	NaN	NaN	NaN	NaN	NaN	
	75%	NaN	NaN	NaN	NaN	NaN	
	max	NaN	NaN	NaN	NaN	NaN	

		leverage	conviction	zhangs_metric	total_items
	count	0.0	0.0	0.0	0.0
	mean	NaN	NaN	NaN	NaN
	std	NaN	NaN	NaN	NaN
	min	NaN	NaN	NaN	NaN
	25%	NaN	NaN	NaN	NaN
	50%	NaN	NaN	NaN	NaN
	75%	NaN	NaN	NaN	NaN
5	max	NaN	NaN	NaN	NaN

```
              antecedent support  consequent support  support  confidence  lift  \
count                        0.0                 0.0      0.0         0.0   0.0
mean                         NaN                 NaN      NaN         NaN   NaN
std                          NaN                 NaN      NaN         NaN   NaN
min                          NaN                 NaN      NaN         NaN   NaN
25%                          NaN                 NaN      NaN         NaN   NaN
50%                          NaN                 NaN      NaN         NaN   NaN
75%                          NaN                 NaN      NaN         NaN   NaN
max                          NaN                 NaN      NaN         NaN   NaN

              leverage  conviction  zhangs_metric  total_items
count              0.0         0.0            0.0          0.0
mean               NaN         NaN            NaN          NaN
std                NaN         NaN            NaN          NaN
min                NaN         NaN            NaN          NaN
25%                NaN         NaN            NaN          NaN
50%                NaN         NaN            NaN          NaN
75%                NaN         NaN            NaN          NaN
max                NaN         NaN            NaN          NaN
```

6

```
              antecedent support  consequent support  support  confidence  lift  \
count                        5.0                 5.0      5.0         5.0  5.00
mean                         0.2                 0.8      0.2         1.0  1.25
std                          0.0                 0.0      0.0         0.0  0.00
min                          0.2                 0.8      0.2         1.0  1.25
25%                          0.2                 0.8      0.2         1.0  1.25
50%                          0.2                 0.8      0.2         1.0  1.25
75%                          0.2                 0.8      0.2         1.0  1.25
max                          0.2                 0.8      0.2         1.0  1.25

              leverage  conviction  zhangs_metric  total_items  coverage
count             5.00         5.0           5.00     5.000000       5.0
mean              0.04         inf           0.25     2.800000       0.2
std               0.00         NaN           0.00     0.447214       0.0
min               0.04         inf           0.25     2.000000       0.2
25%               0.04         NaN           0.25     3.000000       0.2
50%               0.04         NaN           0.25     3.000000       0.2
75%               0.04         NaN           0.25     3.000000       0.2
max               0.04         inf           0.25     3.000000       0.2
```

8

4. Rule Metrics Comparison

mean	std	min	25%	50%	75%
075467511729772	0.10519266426148435	0.4563106796116505	0.5598518518518518	0.5941176470588235	0.662126400996
525451199364243	0.09247246603279749	0.5909090909090909	0.5909090909090909	0.6388888888888888	0.652173913043
nan	nan	nan	nan	nan	nan
666666666666666	0.10259783520851541	0.6666666666666666	1.0	1.0	1.0
nan	nan	nan	nan	nan	nan

nan	nan	nan	nan	nan	nan
1.0	0.0	1.0	1.0	1.0	1.0
794927922235962	0.10930876288179006	0.4896551724137931	0.6075268817204301	0.696078431372549	0.762376237623

5. Top Unique Rules per Cluster

Cluster 3:

Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.055, Confidence: 0.727, Lift: 1.699)
Rule: frozenset({'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.213, Confidence: 0.640, Lift: 1.496)
Rule: frozenset({'age_children_interaction_(42.0, 87.0]', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.048, Confidence: 0.600, Lift: 1.402)
Rule: frozenset({'standard_of_living_index_3', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.054, Confidence: 0.580, Lift: 1.486)
Rule: frozenset({'standard_of_living_index_4'}) -> frozenset({'edu_interaction_(12.0, 16.0]'}) (Support: 0.213, Confidence: 0.499, Lift: 1.496)

Cluster 7:

Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'husband_occupation_2'}) (Support: 0.057, Confidence: 0.652, Lift: 2.304)
Rule: frozenset({'husband_occupation_2', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.057, Confidence: 0.652, Lift: 1.858)
Rule: frozenset({'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3'}) (Support: 0.098, Confidence: 0.591, Lift: 1.535)
Rule: frozenset({'Cluster_(5.0, 8.0]', 'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3'}) (Support: 0.098, Confidence: 0.591, Lift: 1.535)
Rule: frozenset({'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3', 'Cluster_(5.0, 8.0]'}) (Support: 0.098, Confidence: 0.591, Lift: 1.535)

Cluster 1:

Cluster -1:

Rule: frozenset({'husband_occupation_4', 'age_children_interaction_(87.0, 164.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.071, Confidence: 1.000, Lift: 3.500)
Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'standard_of_living_index_4'}) -> frozenset({'husband_occupation_3'}) (Support: 0.071, Confidence: 1.000, Lift: 2.333)
Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) -> frozenset({'husband_occupation_3'}) (Support: 0.071, Confidence: 1.000, Lift: 2.333)
Rule: frozenset({'husband_occupation_3', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_3'}) (Support: 0.071, Confidence: 1.000, Lift: 4.667)
Rule: frozenset({'standard_of_living_index_3', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'husband_occupation_3'}) (Support: 0.071, Confidence: 1.000, Lift: 2.333)

Cluster 5:

Cluster 6:

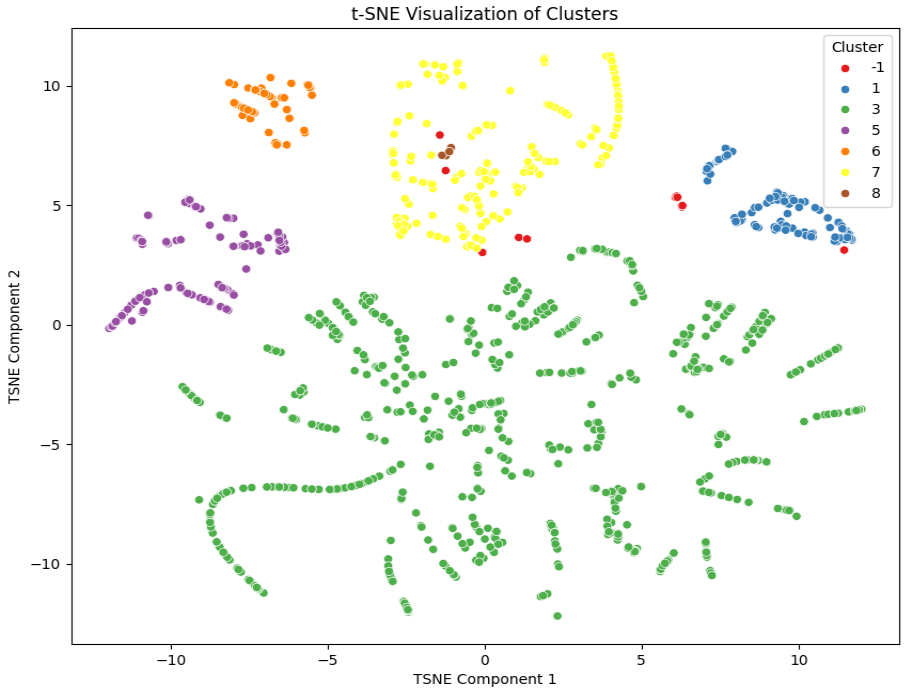
Cluster 8:

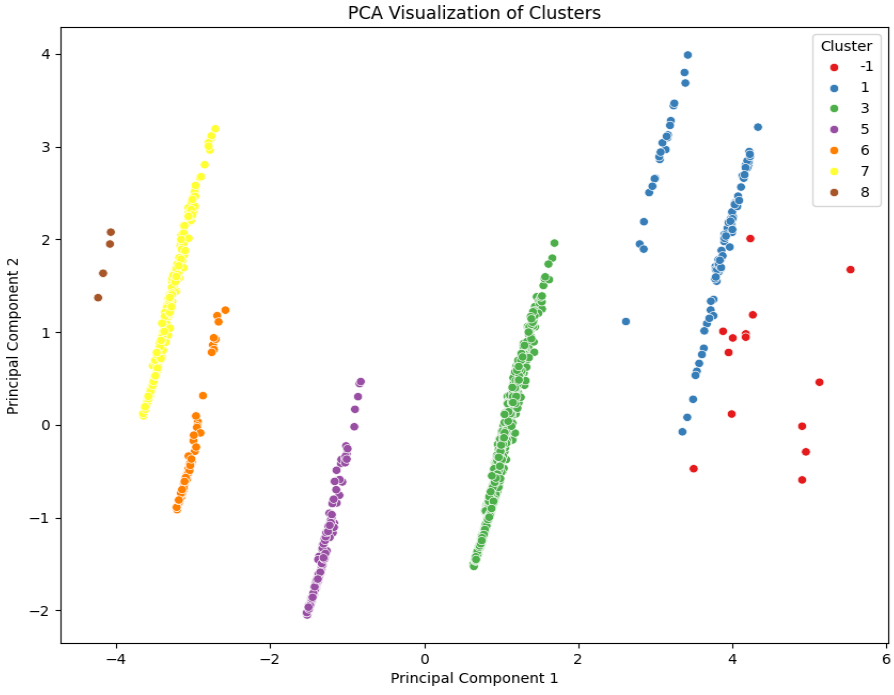
Rule: frozenset({'edu_interaction_(6.0, 12.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.200, Confidence: 1.000, Lift: 1.250)
Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'husband_occupation_4'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.200, Confidence: 1.000, Lift: 1.250)
Rule: frozenset({'edu_interaction_(6.0, 12.0]'}) -> frozenset({'husband_occupation_4', 'standard_of_living_index_4'}) (Support: 0.200, Confidence: 1.000, Lift: 1.250)
Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'Cluster_(5.0, 8.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.200, Confidence: 1.000, Lift: 1.250)
Rule: frozenset({'edu_interaction_(6.0, 12.0]'}) -> frozenset({'Cluster_(5.0, 8.0]', 'standard_of_living_index_4'}) (Support: 0.200, Confidence: 1.000, Lift: 1.250)

6. Top 10 Common Rules Sorted by Absolute Coverage Difference

Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.051)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.047)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.038)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.029)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.023)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.019)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.018)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.015)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.013)
Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.010)

7. Cluster Visualizations


t-SNE Visualization of Clusters
Cluster
-1
1
3
5
6
7
8
TSNE Component 2
TSNE Component 1
10
5
0
−5
−10
−10
−5
0
5
10


PCA Visualization of Clusters
Cluster
-1
1
3
5
6
7
8
Principal Component 2
Principal Component 1
4
3
2
1
0
−1
−2
−4
−2
0
2
4
6