Automated Data Analysis Report

Table of Contents

1. Clustering Results

Best Parameters: {'epsilon': 3.0945475991811424, 'min_samples': 4, 'silhouette': 0.4046043531253859}, Best Silhouette Score: 0.405

Train Silhouette Score: 0.323, Test Silhouette Score: 0.454

2. ANOVA Results

Results for wife_religion: F-value = 9842.113, 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 -- 0 3 -2.6512 0.0 -2.7162 -2.5861 True 0 4 0.1808 0.0 0.1254 0.2361 True 0 5 0.1808 0.0 0.1187 0.2428 True 3 4 2.8319 0.0 2.7884 2.8755 True 3 5 2.8319 0.0 2.7802 2.8836 True 4 5 -0.0 1.0 -0.0388 0.0388 False
----

Results for wife_working: F-value = 1213.322, 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 --- 0 3 -0.1743 0.0326 -0.3385 -0.01 True 0 4 0.5854 0.0 0.4455 0.7252 True 0 5 -1.7074 0.0 -1.864 -1.5508 True 3 4 0.7597 0.0 0.6497 0.8696 True 3 5 -1.5331 0.0 -1.6637 -1.4025 True 4 5 -2.2928 0.0 -2.3909 -2.1947 True
----

Results for media_exposure: 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 --- 0 3 -3.6903 0.0 -3.6903 -3.6903 True 0 4 -3.6903 0.0 -3.6903 -3.6903 True 0 5 -3.6903 0.0 -3.6903 -3.6903 True 3 4 -0.0 0.0 -0.0 -0.0 True 3 5 -0.0 0.1531 -0.0 0.0 False 4 5 0.0 0.0 0.0 0.0 True ---

Results for age_children_interaction: F-value = 10.132, 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 --- 0 3 -0.5884 0.0 -0.9168 -0.2599 True 0 4 -0.4105 0.001 -0.6901 -0.1309 True 0 5 -0.6211 0.0 -0.9342 -0.308 True 3 4 0.1779 0.1594 -0.0419 0.3977 False 3 5 -0.0327 0.9884 -0.2938 0.2284 False 4 5 -0.2106 0.0297 -0.4067 -0.0145 True
----

Results for edu_interaction: F-value = 74.302, 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 --- 0 3 1.7659 0.0 1.4609 2.0708 True 0 4 1.1435 0.0 0.8839 1.4031 True 0 5 1.2048 0.0 0.9141 1.4956 True 3 4 -0.6224 0.0 -0.8265 -0.4183 True 3 5 -0.561 0.0 -0.8035 -0.3186 True 4 5 0.0613 0.8221 -0.1208 0.2434 False
----

3. Cluster Variability

	antecedent support	consequent support	support	confidence	lift \
count	2.000000	2.000000	2.000000	2.000000	2.000000
mean	0.047872	0.186170	0.031915	0.675000	3.617157
std	0.007522	0.007522	0.000000	0.106066	0.423571
min	0.042553	0.180851	0.031915	0.600000	3.317647
25%	0.045213	0.183511	0.031915	0.637500	3.467402
50%	0.047872	0.186170	0.031915	0.675000	3.617157
75%	0.050532	0.188830	0.031915	0.712500	3.766912
max	0.053191	0.191489	0.031915	0.750000	3.916667

	leverage	conviction	zhangs_metric	total_items	coverage
count	2.000000	2.000000	2.000000	2.0	2.000000
mean	0.023031	2.640957	0.757803	3.0	0.047872
std	0.001040	0.838749	0.028249	0.0	0.007522
min	0.022295	2.047872	0.737828	3.0	0.042553
25%	0.022663	2.344415	0.747815	3.0	0.045213
50%	0.023031	2.640957	0.757803	3.0	0.047872
75%	0.023399	2.937500	0.767790	3.0	0.050532
max	0.023766	3.234043	0.777778	3.0	0.053191

0

	antecedent support	consequent support	support	confidence	lift \
count	8.000000	8.000000	8.000000	8.000000	8.000000
mean	0.169899	0.380388	0.100754	0.616861	1.613644
std	0.136250	0.050189	0.075136	0.123897	0.167796
min	0.071839	0.329023	0.045977	0.484848	1.442113
25%	0.093750	0.336207	0.054598	0.505827	1.498432
50%	0.102730	0.369971	0.066092	0.602724	1.516726
75%	0.178879	0.433908	0.112428	0.688590	1.792935
max	0.433908	0.433908	0.221264	0.803030	1.850692

	leverage	conviction	zhangs_metric	total_items	coverage
count	8.000000	8.000000	8.000000	8.00000	8.000000
mean	0.036827	1.780288	0.460455	2.75000	0.169899
std	0.024916	0.604799	0.091754	0.46291	0.136250
min	0.014095	1.288540	0.338690	2.00000	0.071839
25%	0.018161	1.347060	0.371956	2.75000	0.093750
50%	0.029252	1.579203	0.491880	3.00000	0.102730
75%	0.045098	1.895010	0.509173	3.00000	0.178879
max	0.075381	2.874005	0.601820	3.00000	0.433908

4

```
       antecedent support  consequent support   support  confidence      lift  \
count            8.000000            8.000000  8.000000    8.000000  8.000000
mean             0.136824            0.351914  0.084459    0.629029  1.814057
std              0.033875            0.083109  0.021094    0.120203  0.261525
min              0.090090            0.265766  0.063063    0.451613  1.622932
25%              0.102477            0.272523  0.063063    0.595328  1.681482
50%              0.150901            0.367117  0.081081    0.648649  1.694118
75%              0.166667            0.382883  0.108108    0.661486  1.830704
max              0.166667            0.509009  0.108108    0.826087  2.394453

       leverage  conviction  zhangs_metric  total_items  coverage
count  8.000000    8.000000       8.000000     8.000000  8.000000
mean   0.036505    1.873423       0.509244     2.875000  0.136824
std    0.007271    0.475674       0.066394     0.353553  0.033875
min    0.024694    1.322470       0.428194     2.000000  0.090090
25%    0.032490    1.665103       0.482531     3.000000  0.102477
50%    0.035103    1.756410       0.491667     3.000000  0.150901
75%    0.044294    2.054899       0.528212     3.000000  0.166667
5      max    0.044294    2.823198       0.646429     3.000000  0.166667
       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
3      max         NaN         NaN            NaN          NaN
```

4. Rule Metrics Comparison

mean	std	min	25%	50%	75%
0.675	0.10606601717798214	0.6	0.6375	0.675	0.7125
.168607555983384	0.12389674540338928	0.48484848484848486	0.5058269545024512	0.6027236652236653	0.68858974358
.290289580349188	0.12020348625264278	0.45161290322580644	0.5953282828282828	0.6486486486486487	0.66148648648
nan	nan	nan	nan	nan	nan
.258966153085457	0.07743962136650274	0.6116504854368933	0.6736051193557664	0.745942217172537	0.78633874025

5. Top Unique Rules per Cluster

Cluster 0:

Rule: frozenset({'age_children_interaction_(44.0, 90.0]', 'edu_interaction_(6.0, 12.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.032, Confidence: 0.750, Lift: 3.917)
Rule: frozenset({'age_children_interaction_(44.0, 90.0]', 'standard_of_living_index_4'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.032, Confidence: 0.600, Lift: 3.318)

Cluster 4:

Rule: frozenset({'age_children_interaction_(168.0, 768.0]', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.056, Confidence: 0.780, Lift: 1.798)
Rule: frozenset({'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.221, Confidence: 0.658, Lift: 1.517)
Rule: frozenset({'age_children_interaction_(44.0, 90.0]', 'standard_of_living_index_3'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.055, Confidence: 0.603, Lift: 1.494)
Rule: frozenset({'standard_of_living_index_4', 'age_children_interaction_(90.0, 168.0]'}) -> frozenset({'edu_interaction_(12.0, 16.0]'}) (Support: 0.076, Confidence: 0.602, Lift: 1.791)
Rule: frozenset({'standard_of_living_index_4'}) -> frozenset({'edu_interaction_(12.0, 16.0]'}) (Support: 0.221, Confidence: 0.510, Lift: 1.517)

Cluster 5:

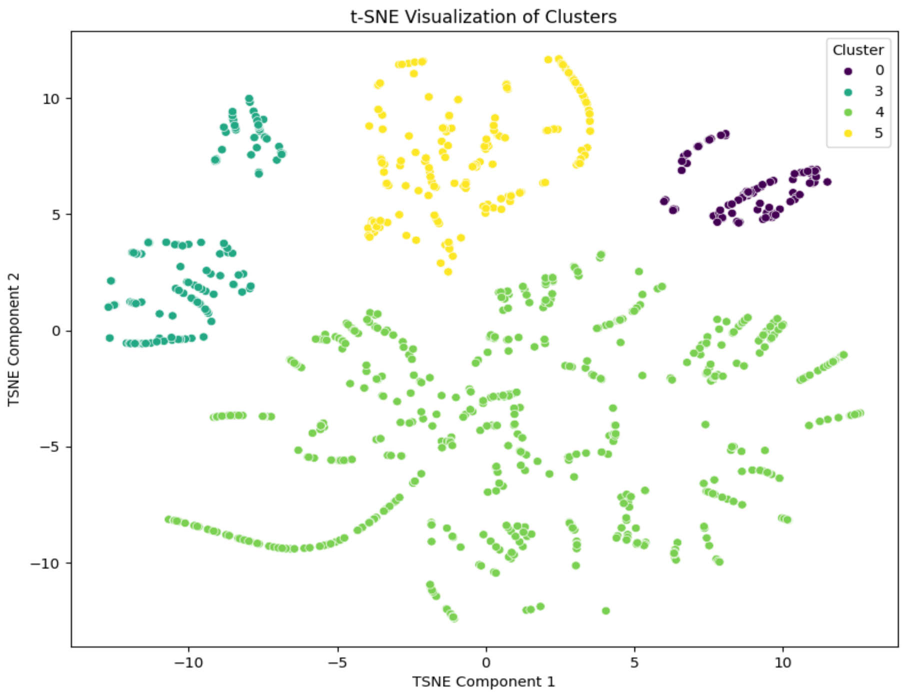
Rule: frozenset({'age_children_interaction_(44.0, 90.0]', 'husband_occupation_2'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.063, Confidence: 0.700, Lift: 1.992)
Rule: frozenset({'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3'}) (Support: 0.108, Confidence: 0.649, Lift: 1.694)
Rule: frozenset({'Cluster_(4.0, 5.0]', 'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3'}) (Support: 0.108, Confidence: 0.649, Lift: 1.694)
Rule: frozenset({'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3', 'Cluster_(4.0, 5.0]'}) (Support: 0.108, Confidence: 0.649, Lift: 1.694)
Rule: frozenset({'age_children_interaction_(44.0, 90.0]', 'edu_interaction_(6.0, 12.0]'}) -> frozenset({'husband_occupation_2'}) (Support: 0.063, Confidence: 0.636, Lift: 2.394)

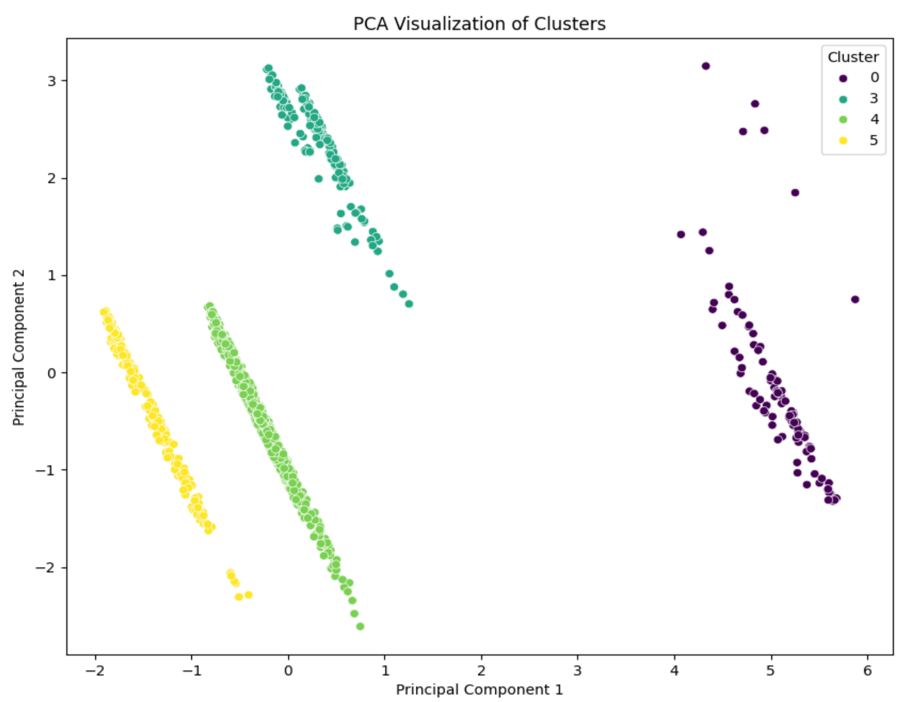
Cluster 3:

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_(90.0, 168.0]'}) (Abs Coverage Difference: 0.023)
Rule: frozenset({'standard_of_living_index_4', 'age_children_interaction_(90.0, 168.0]', 'edu_interaction_(12.0, 16.0]'}) (Abs Coverage Difference: 0.009)

7. Cluster Visualizations


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


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