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 = 20296.886, P-value = 0.000 Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05 ======== group1 group2 meandiff p-adi 1.8704 0.0 1.7712 1.9696 True -1 4 1.8704 0.0 1.7697 1.971 True -1 5 -0.9352 0.0 -1.0377 -0.8327 True -1 6 -0.9352 0.0 -1.0422 -0.8282 True 1 3 -0.0 1.0 -0.0359 0.0359 False 1 4 -0.0 1.0 -0.0398 0.0398 False 1 5 -2.8055 0.0 -2.8498 -2.7613 True 1 6 -2.8055 0.0 -2.8593 -2.7518 True 3 4 0.0 1.0 -0.0237 0.0237 False 3 5 -2.8055 0.0 -2.8361 -2.7749 True 3 6 -2.8055 0.0 -2.8488 -2.7623 True 4 5 -2.8055 0.0 -2.8406 -2.7704 True 4 6 -2.8055 0.0 -2.8521 -2.759 True 5 6 -0.0 1.0 -0.0504 0.0504 False -----Results for wife_working: F-value = 3630.540, P-value = 0.000 Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05 ======= group1 group2 meandiff p-adi 1.3462 0.0 1.119 1.5735 True -1 4 -0.9616 0.0 -1.1922 -0.731 True -1 5 1.3462 0.0 1.1114 1.5811 True -1 6 -0.9616 0.0 -1.2067 -0.7165 True 1 3 0.5255 0.0 0.4434 0.6077 True 1 4 -1.7823 0.0 -1.8734 -1.6912 True 1 5 0.5255 0.0 0.4242 0.6269 True 1 6 -1.7823 0.0 -1.9055 -1.6591 True 3 4 -2.3078 0.0 -2.3621 -2.2536 True 3 5 -0.0 1.0 -0.0701 0.0701 False 3 6 -2.3078 0.0 -2.4069 -2.2087 True 4 5 2.3078 0.0 2.2274 2.3882 True 4 6 0.0 1.0 -0.1066 0.1066 False 5 6 -2.3078 0.0 -2.4233 -2.1923 True Results for media_exposure: F-value = 10811.880, P-value = 0.000 Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05 ======== group1 group2 meandiff p-adi -2.5468 0.0 -2.6819 -2.4117 True -1 4 -2.5468 0.0 -2.6839 -2.4097 True -1 5 -2.5468 0.0 -2.6864 -2.4072 True -1 6 -2.5468 0.0 -2.6925 -2.4011 True 1 3 -3.8202 0.0 -3.869 -3.7713 True 1 4 -3.8202 0.0 -3.8743 -3.766 True 1 5 -3.8202 0.0 -3.8804 -3.7599 True 1 6 -3.8202 0.0 -3.8934 -3.7469 True 3 4 -0.0 1.0 -0.0323 0.0323 False 3 5 -0.0 1.0 -0.0417 0.0417 False 3 6 -0.0 1.0 -0.0589 0.0589 False 4 5 0.0 1.0 -0.0478 0.0478 False 4 6 0.0 1.0 -0.0634 0.0634 False 5 6 0.0 1.0 -0.0687 0.0687 False Results for age_children_interaction: F-value = 15.977, P-value = 0.000 Tukey-HSD Test Results: Multiple Comparison of Means - Tukey HSD, FWER=0.05 ========= group1 group2 meandiff p-adi 3 -1.7529 0.0 -2.5622 -0.9436 True -1 4 -1.9621 0.0 -2.7835 -1.1407 True -1 5 -1.8269 0.0 -2.6633 -0.9905 True -1 6 -1.9877 0.0 -2.8605 -1.1148 True 1 3 -0.4766 0.0001 -0.7692 -0.184 True 1 4 -0.6858 0.0 -1.0103 -0.3614 True 1 5 -0.5506 0.0002 -0.9115 -0.1897 True 1 6 -0.7114 0.0001 -1.1501 -0.2727 True 3 4 -0.2092 0.025 -0.4025 -0.016 True 3 5 -0.074 0.9587 -0.3237 0.1756 False 3 6 -0.2348 0.4035 -0.5878 0.1181 False 4 5 0.1352 0.7583 -0.1511 0.4215 False 4 6 -0.0256 1.0 -0.4054 0.3542 False 5 6 -0.1608 0.8751 -0.5721 0.2505 False ------

Results for edu_interaction: F-value = 55.231, P-value = 0.000

3. Cluster Variability

		ent support consequent support support confidence \							
	count	42.000000 42.000000 42.000000							
	mean	0.168162							
	std	0.104944							
	min	0.062785							
	25%	0.090468							
	50%	0.130708							
	75%	0.226313							
	max	0.428082							
lift leverage conviction zhangs_metric total_items coverage									
	count 42.000000								
	mean 1.29158								
	std 0.167568								
	min 1.134799								
	25% 1.175932								
	50% 1.208708								
	75% 1.378380								
3	max 1.796923	3 0.070776 2.478311 0.579718 3.000000 0.428082							
	antece	dent support consequent support support confidence \							
	count	33.000000 33.000000 33.000000							
	mean	0.212121 0.359626 0.113302 0.536347							
	std	0.139517							
	min	0.073529							
	25%	0.099265							
	50%	0.136029							
	75%	0.349265							
	max	0.496324							
		everage conviction zhangs_metric total_items coverage							
		33.000000 33.000000 33.000000 33.000000							
	mean 1.50556								
	std 0.225781								
	min 1.283951								
	25% 1.374316								
	50% 1.392888								
	75% 1.574074								
4	max 2.365217	7 0.074989 3.399816 0.630522 3.000000 0.496324							

```
antecedent support consequent support support confidence \
                                     16.000000 16.000000 16.000000
          count
                     16.000000
           mean
                       0.230198
                                      0.257426 0.094678
                                                          0.430119
            std
                      0.101197
                                     0.100256 0.049170
                                                         0.166763
                                     0.148515 0.039604
            min
                       0.069307
                                                          0.266667
            25%
                       0.165842
                                      0.193069 0.056931
                                                          0.298529
           50%
                       0.198020
                                      0.198020 0.089109
                                                          0.320856
            75%
                       0.336634
                                      0.341584 0.111386
                                                          0.591176
                       0.435644
                                      0.435644 0.198020
                                                          0.714286
            max
               lift leverage conviction zhangs_metric total_items coverage
    count 16.000000 16.000000 16.000000
                                           16.000000
                                                       16.000000 16.000000
                                                       2.500000 0.230198
     mean
            1.687226 0.037607
                                1.383560
                                            0.519623
      std
           0.265425 0.020920 0.315130
                                            0.129621
                                                      0.516398 0.101197
      min
           1.377273 0.014214
                                1.102723
                                            0.341564
                                                       2.000000 0.069307
     25%
            1.507574 0.019974
                                1.143299
                                            0.422688
                                                       2.000000 0.165842
     50%
            1.644969 0.034751
                                1.246463
                                            0.496250
                                                       2.500000 0.198020
     75%
                                                       3.000000 0.336634
            1.796449 0.047936
                                1.510166
                                            0.615792
1
     max
            2.142424 0.078032
                                2.122772
                                            0.769335
                                                       3.000000 0.435644
              antecedent support consequent support support confidence \
                     18.000000
                                     18.000000 18.000000 18.000000
          count
           mean
                       0.219157
                                      0.336015 0.108046
                                                          0.508400
                                     0.037537 0.028875
                      0.076342
            std
                                                         0.058202
                       0.096552
                                     0.275862 0.062069
                                                          0.408163
            min
            25%
                       0.175862
                                      0.337931 0.089655
                                                          0.500000
                                      0.337931 0.113793
           50%
                       0.213793
                                                          0.516129
                       0.275862
                                      0.370690 0.137931
                                                          0.542869
           75%
                                      0.379310 0.137931
            max
                       0.337931
                                                          0.642857
               lift leverage conviction zhangs_metric total_items coverage
    count 18.000000 18.000000 18.000000
                                           18.000000
                                                       18.000000 18.000000
            1.516875 0.036047
                                                       2.777778 0.219157
     mean
                                1.363338
                                            0.434456
      std
           0.120650 0.008902
                               0.135964
                                            0.057102
                                                      0.427793 0.076342
                                                       2.000000 0.096552
           1.387560 0.019263
                                1.223543
                                            0.321429
      min
     25%
            1.457398 0.032580
                                1.313793
                                            0.392157
                                                       3.000000 0.175862
     50%
            1.479592 0.038098
                                                       3.000000 0.213793
                                1.368276
                                            0.439145
     75%
                                1.374384
                                            0.479092
                                                       3.000000 0.275862
            1.527321 0.044709
5
            1.902332 0.044709
                                1.853793
                                            0.525021
                                                       3.000000 0.337931
     max
```

anteced	lent support conse	quent support suppo	ort confidence \						
count	18.000000	18.000000 18.000000	18.000000						
mean	0.197347	0.644279 0.140962							
std	0.140811	0.171269 0.076828	0.207872						
min	0.044776	0.373134 0.044776	0.518519						
25%	0.067164	0.432836 0.067164	0.579091						
50%	0.119403	0.776119 0.119403	1.000000						
75%	0.361940	0.776119 0.208955							
max	0.402985	0.776119 0.283582	1.000000						
lift leverage conviction zhangs_metric total_items coverage									
count 18.000000			8.000000 18.000000						
mean 1.3099	73 0.033526	inf 0.305400 2.8	388889 0.197347						
std 0.055343	0.019359 N	aN 0.085126 0.3	323381 0.140811						
min 1.224038	0.010025 1.289	009 0.234375 2	.000000 0.044776						
25% 1.288462	0.015037 1.407	7 564 0.247986 3	3.000000 0.067164						
50% 1.288462	2 0.026732 N	laN 0.257578 3.	000000 0.119403						
75% 1.293793	3 0.047449 N	laN 0.362245 3.	000000 0.361940						
6 max 1.47021	9 0.066830	inf 0.476190 3.0	00000 0.402985						
anteced	lent support conse	auent eunnort eunn	ort confidence \						
count	29.000000	29.000000 29.0000							
mean	0.097701	0.442529 0.09770							
std	0.032035	0.099860 0.03203							
min	0.083333	0.250000 0.08333							
25%	0.083333	0.416667 0.08333							
50%	0.083333	0.500000 0.08333							
75%	0.083333	0.500000 0.08333							
max	0.166667	0.500000 0.16666							
	01.0000.	0.00000							
lift leverage conviction zhangs_metric total_items coverage									
count 29.00000		29.0 29.000000 -	29.0 29.000000						
mean 2.427	586 0.054837	inf 0.619122	3.0 0.097701						
std 0.77593	1 0.022298	NaN 0.116632	0.0 0.032035						
	000 0.041667	inf 0.545455	3.0 0.083333						
	00 0.041667	NaN 0.545455	3.0 0.083333						
50% 2.0000		NaN 0.545455	3.0 0.083333						
75% 2.4000		NaN 0.636364	3.0 0.083333						
-1 max 4.000	000 0.125000	inf 0.900000	3.0 0.166667						

4. Rule Metrics Comparison

mean	std	min	25%	50%	75%
91100884440497	0.09427321097816364	0.32085561497326204	0.44472389585981975	0.49671319417765825	0.5104427736
3467925967239	0.11166707123207359	0.37894736842105264	0.46666666666666	0.548148148148148	0.5909090909
1194057168322	0.1667626043350457	0.2666666666666666	0.2985294117647059	0.32085561497326204	0.5911764705
3995887126361	0.05820159786792579	0.40816326530612246	0.5	0.5161290322580645	0.5428692699
6045641601197	0.2078719998105836	0.5185185185185185	0.5790909090909091	1.0	1.0

1.0 0.0 1.0 1.0 1.0 1.0

9802563448316 0.1380484715492421 0.3770491803278688 0.6289384502656165 0.7142857142857142 0.7824620041

5. Top Unique Rules per Cluster

Cluster 3:

Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.048, Confidence: 0.600, Lift: 1.402) Rule: frozenset({'age_children_interaction_(42.0, 87.0]', 'standard_of_living_index_3'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.054, Confidence: 0.580, Lift: 1.486) Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'age_children_interaction_(87.0, 164.0]'}) -> frozenset({'husband_occupation_3'}) (Support: 0.045, Confidence: 0.520, Lift: 1.221) Rule: frozenset({'age_children_interaction_(87.0, 164.0]', 'standard_of_living_index_3'}) -> frozenset({'husband_occupation_3'}) (Support: 0.037, Confidence: 0.508, Lift: 1.193) Rule: frozenset({'age_children_interaction_(87.0, 164.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.116, Confidence: 0.505, Lift: 1.180)

Cluster 4:

Rule: frozenset({'Cluster_(3.0, 4.0]', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.272, Confidence: 0.685, Lift: 1.381) Rule: frozenset({'edu_interaction_(12.0, 16.0]'}) -> frozenset({'Cluster_(3.0, 4.0]', 'standard_of_living_index_4'}) (Support: 0.272, Confidence: 0.685, Lift: 1.381) Rule: frozenset({'husband_occupation_2', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.055, Confidence: 0.652, Lift: 1.867) Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'age_children_interaction_(42.0, 87.0]'}) -> frozenset({'husband_occupation_2'}) (Support: 0.055, Confidence: 0.652, Lift: 2.365) Rule: frozenset({'Cluster_(3.0, 4.0]', 'standard_of_living_index_2'}) -> frozenset({'husband_occupation_3'}) (Support: 0.096, Confidence: 0.591, Lift: 1.576)

Cluster 1:

Rule: frozenset({'husband_occupation_2', 'standard_of_living_index_3'}) -> frozenset({'age_children_interaction_(164.0, 768.0]'}) (Support: 0.050, Confidence: 0.714, Lift: 1.640) Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'standard_of_living_index_2'}) -> frozenset({'husband_occupation_2'}) (Support: 0.109, Confidence: 0.688, Lift: 2.042) Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'husband_occupation_2'}) -> frozenset({'standard_of_living_index_2'}) (Support: 0.109, Confidence: 0.611, Lift: 1.715) Rule: frozenset({'standard_of_living_index_3'}) -> frozenset({'age_children_interaction_(164.0, 768.0]'}) (Support: 0.119, Confidence: 0.600, Lift: 1.377) Rule: frozenset({'husband_occupation_2'}) -> frozenset({'standard_of_living_index_2'}) (Support: 0.198, Confidence: 0.588, Lift: 1.650)

Cluster 5:

Rule: frozenset({'age_children_interaction_(164.0, 768.0]'}) -> frozenset({'husband_occupation_2'}) (Support: 0.117, Confidence: 0.548, Lift: 1.446)

Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'Cluster_(4.0, 6.0]'}) -> frozenset({'husband_occupation_2'}) (Support: 0.117, Confidence: 0.548, Lift: 1.446) Rule: frozenset({'age_children_interaction_(164.0, 768.0]'}) -> frozenset({'Cluster_(4.0, 6.0]', 'husband_occupation_2'}) (Support: 0.117, Confidence: 0.548, Lift: 1.446) Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'standard_of_living_index_4'}) -> frozenset({'husband_occupation_2'}) (Support: 0.069, Confidence: 0.526, Lift: 1.388) Rule: frozenset({'Cluster_(4.0, 6.0]', 'standard_of_living_index_3'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.110, Confidence: 0.516, Lift: 1.527)

Cluster 6:

Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'Cluster_(4.0, 6.0]'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.119, Confidence: 1.000, Lift: 1.288) Rule: frozenset({'age_children_interaction_(164.0, 768.0]'}) -> frozenset({'Cluster_(4.0, 6.0]', 'standard_of_living_index_4'}) (Support: 0.119, Confidence: 1.000, Lift: 1.288) Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'husband_occupation_2'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.045, Confidence: 1.000, Lift: 1.288) Rule: frozenset({'age_children_interaction_(87.0, 164.0]', 'husband_occupation_2'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.209, Confidence: 1.000, Lift: 1.288) Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'husband_occupation_2'}) -> frozenset({'standard_of_living_index_4'}) (Support: 0.060, Confidence: 1.000, Lift: 1.288)

Cluster -1:

Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'husband_occupation_3'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.083, Confidence: 1.000, Lift: 2.000) Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'standard_of_living_index_3'}) -> frozenset({'edu_interaction_(6.0, 12.0]'}) (Support: 0.083, Confidence: 1.000, Lift: 2.000) Rule: frozenset({'husband_occupation_3', 'standard_of_living_index_3'}) -> frozenset({'edu_interaction_(12.0, 16.0]'}) (Support: 0.083, Confidence: 1.000, Lift: 4.000) Rule: frozenset({'age_children_interaction_(87.0, 164.0]', 'edu_interaction_(12.0, 16.0]'}) -> frozenset({'husband_occupation_3'}) (Support: 0.083, Confidence: 1.000, Lift: 2.000) Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'standard_of_living_index_3'}) -> frozenset({'age_children_interaction_(87.0, 164.0]'}) (Support: 0.083, Confidence: 1.000, Lift: 2.000)

6. Top 10 Common Rules Sorted by Absolute Coverage Difference

Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'husband_occupation_2', 'standard_of_living_index_4'}) (Abs Coverage Difference: 0.236)

Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'husband_occupation_2', 'standard_of_living_index_4'}) (Abs Coverage Difference: 0.225)

Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'standard_of_living_index_4', 'husband_occupation_2'}) (Abs Coverage Difference: 0.200)

Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'standard_of_living_index_3'}) (Abs Coverage Difference: 0.177)

Rule: frozenset({'edu_interaction_(12.0, 16.0]', 'standard_of_living_index_4'}) (Abs Coverage Difference: 0.163)

Rule: frozenset({'husband_occupation_3', 'edu_interaction_(6.0, 12.0]'}) (Abs Coverage Difference: 0.150)

Rule: frozenset({'age_children_interaction_(164.0, 768.0]', 'standard_of_living_index_4'}) (Abs

Coverage Difference: 0.140)

Rule: frozenset({'edu_interaction_(6.0, 12.0]', 'standard_of_living_index_3'}) (Abs Coverage Difference:

0.115)

Rule: frozenset({'husband_occupation_3', 'edu_interaction_(6.0, 12.0]'}) (Abs Coverage Difference:

0.115)

Rule: frozenset({'standard_of_living_index_4', 'edu_interaction_(12.0, 16.0]', 'age_children_interaction_(87.0, 164.0]'}) (Abs Coverage Difference: 0.114)

7. Cluster Visualizations



