

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

2. ANOVA Results

Results for capital-gain: F-value = 228027.175, 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 13.2937 13.3837 True 1 3
1.0125 0.0 0.9949 1.0301 True 1 4 -0.0 1.0 -0.0143 0.0143 False 2 3 -12.3262 0.0 -12.3703 -12.2821
True 2 4 -13.3387 0.0 -13.3816 -13.2958 True 3 4 -1.0125 0.0 -1.0238 -1.0012 True
-----
```

Results for capital-loss: F-value = 259301.391, 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 -4.6088 0.0 -4.6511 -4.5665 True 1 3
-4.6088 0.0 -4.6254 -4.5922 True 1 4 -4.6088 0.0 -4.6223 -4.5954 True 2 3 0.0 1.0 -0.0415 0.0415
False 2 4 -0.0 1.0 -0.0403 0.0403 False 3 4 -0.0 1.0 -0.0106 0.0106 False
-----
```

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 2 3 0.0 0.0 0.0 0.0 True 2 4 -3.5999 0.0
-3.5999 -3.5999 True 3 4 -3.5999 0.0 -3.5999 -3.5999 True -----
```

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 2 3 -0.0 0.0 -0.0 -0.0 True 2 4
-0.0 0.0 -0.0 -0.0 True 3 4 -0.0 0.0 -0.0 -0.0 True -----
```

Results for age_education_interaction: F-value = 436.789, 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 0.8181 0.0 0.6055 1.0306 True 1 3
0.0993 0.0118 0.016 0.1825 True 1 4 -0.4841 0.0 -0.5517 -0.4165 True 2 3 -0.7188 0.0 -0.9272
-0.5104 True 2 4 -1.3021 0.0 -1.5048 -1.0995 True 3 4 -0.5833 0.0 -0.6366 -0.5301 True
-----
```

3. Cluster Variability

3	antecedent support	consequent support	support	confidence \
	count	107.000000	107.000000	107.000000
	mean	0.083740	0.243636	0.048353
	std	0.045382	0.086054	0.023328
	min	0.029992	0.125719	0.027116
	25%	0.045193	0.182005	0.030403
	50%	0.067379	0.202958	0.036565
	75%	0.112572	0.256368	0.057313
	max	0.202958	0.410846	0.110107
				0.917808
	lift	leverage	conviction	zhangs_metric
	count	107.000000	107.000000	107.000000
	mean	2.742092	0.028951	2.434190
	std	0.921510	0.015991	1.158408
	min	1.794985	0.013956	1.327059
	25%	2.034365	0.017920	1.459066
	50%	2.285053	0.023929	1.895184
	75%	3.255464	0.029946	3.161177
	max	5.463524	0.074919	7.168036
				0.872680
	total_items \			
	count	107.000000	107.000000	107.000000
	mean	2.742092	0.028951	2.434190
	std	0.921510	0.015991	1.158408
	min	1.794985	0.013956	1.327059
	25%	2.034365	0.017920	1.459066
	50%	2.285053	0.023929	1.895184
	75%	3.255464	0.029946	3.161177
	max	5.463524	0.074919	7.168036
				0.872680
	coverage			
	count	107.000000		
	mean	0.083740		
	std	0.045382		
	min	0.029992		
	25%	0.045193		
	50%	0.067379		
	75%	0.112572		
	max	0.202958		
4	antecedent support	consequent support	support	confidence \
	count	84.000000	84.000000	84.000000
	mean	0.104050	0.262216	0.058999
	std	0.069485	0.080317	0.035814
	min	0.029878	0.145305	0.025305
	25%	0.051265	0.201912	0.031340
	50%	0.076101	0.223768	0.043524
	75%	0.139138	0.344592	0.068107
	max	0.344592	0.437263	0.142643
				0.991416
	lift	leverage	conviction	zhangs_metric
	count	84.000000	84.000000	84.000000
	mean	2.386512	0.033274	5.248958
	std	0.455398	0.020977	12.523368
	min	1.747838	0.012210	1.293575
	25%	2.039033	0.018263	1.447832
	50%	2.343404	0.024154	1.673277
	75%	2.623155	0.041694	3.012710
	max	4.146542	0.087095	65.558819
				0.931602
	total_items			
	count	84.000000	84.000000	84.000000
	mean	2.386512	0.033274	5.248958
	std	0.455398	0.020977	12.523368
	min	1.747838	0.012210	1.293575
	25%	2.039033	0.018263	1.447832
	50%	2.343404	0.024154	1.673277
	75%	2.623155	0.041694	3.012710
	max	4.146542	0.087095	65.558819
				0.931602
	coverage			
	count	84.000000		
	mean	0.104050		
	std	0.069485		
	min	0.029878		
	25%	0.051265		
	50%	0.076101		
	75%	0.139138		
	max	0.344592		

	antecedent support consequent support support confidence \						
	count	121.000000	121.000000	121.000000	121.000000		
	mean	0.076982	0.254484	0.048076	0.681212		
	std	0.048871	0.081904	0.025380	0.172412		
	min	0.026694	0.112252	0.026010	0.417323		
	25%	0.043121	0.199863	0.031485	0.531429		
	50%	0.060233	0.220397	0.034908	0.647887		
	75%	0.092402	0.296372	0.056810	0.825397		
	max	0.220397	0.396988	0.123203	1.000000		
	lift leverage conviction zhangs_metric total_items \						
	count	121.000000	121.000000	121.000000	121.000000	121.000000	
	mean	2.826871	0.029753	inf	0.672363	2.900826	
	std	0.874105	0.016434	NaN	0.102405	0.300138	
	min	1.973658	0.013525	1.370859	0.524092	2.000000	
	25%	2.186272	0.018517	1.665150	0.586575	3.000000	
	50%	2.469574	0.024001	2.187918	0.634925	3.000000	
	75%	3.028221	0.034722	4.186069	0.786848	3.000000	
	max	5.826753	0.079154	inf	0.865708	3.000000	
1	coverage						
	count	121.000000					
	mean	0.076982					
	std	0.048871					
	min	0.026694					
	25%	0.043121					
	50%	0.060233					
	75%	0.092402					
	max	0.220397					
	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		
2	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
	max	10.219780	0.042456	inf	0.944820	3.000000	0.070968

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
12121151123937	0.17241245748440304	0.4173228346456693	0.5314285714285715	0.6478873239436619	0.82539682539
52302789802789	0.09901864756375088	0.6666666666666666	0.7142857142857143	0.75	0.83333333333
49058149800485	0.17700274639572697	0.39316325501663807	0.46019592388244573	0.5493871665465033	0.74370860927

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({'occupation_aggregated_Other', 'age_education_interaction_(261.0, 369.0]']) -> frozenset({'education_HS-grad'}) (Support: 0.034, Confidence: 0.562, Lift: 2.191)
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({'hours_per_week_binned_41-50', 'age_education_interaction_(261.0, 369.0]']) -> frozenset({'education_HS-grad'}) (Support: 0.057, Confidence: 0.504, Lift: 1.964)
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)

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 1:

Rule: frozenset({'education_Doctorate', 'race_White'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]']) (Support: 0.031, Confidence: 1.000, Lift: 2.519)
Rule: frozenset({'education_Doctorate', 'native_country_aggregated_United-States'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]']) (Support: 0.029, Confidence: 1.000, Lift: 2.519)
Rule: frozenset({'education_Doctorate'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]']) (Support: 0.034, Confidence: 0.980, Lift: 2.470)
Rule: frozenset({'education_Doctorate', 'marital-status_Married-civ-spouse'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]']) (Support: 0.029, Confidence: 0.977, Lift: 2.470)

2.460)

Rule: frozenset({'education_Doctorate', 'sex_Male'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.026, Confidence: 0.974, Lift: 2.454)

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)

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_Not-in-family', 'race_White', 'marital-status_Never-married'}) (Abs Coverage Difference: 0.271)

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)

7. Cluster Visualizations

