

# Automated Data Analysis Report

## Clustering Results

Best Parameters: {'epsilon': 3.8040859170159296, 'min\_samples': 4, 'silhouette': 0.4880201859408698}, Best Silhouette Score: 0.488

Train Silhouette Score: 0.488, Test Silhouette Score: 0.498

## ANOVA Results

Results for age: F-value = 38.494, 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 0 -0.9337 0.548 -2.7463 0.879 False -1 1
-0.6836 0.7676 -2.4977 1.1306 False -1 2 -0.3153 0.971 -2.1433 1.5128 False 0 1 0.2501 0.0 0.1737
0.3265 True 0 2 0.6184 0.0 0.3808 0.856 True 1 2 0.3683 0.0008 0.1199 0.6167 True
-----
```

Results for education-num: F-value = 106.421, 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 0 0.6124 0.8196 -1.1928 2.4177 False -1 1
0.9881 0.4961 -0.8186 2.7948 False -1 2 1.786 0.0568 -0.0346 3.6065 False 0 1 0.3756 0.0 0.2996
0.4517 True 0 2 1.1735 0.0 0.9369 1.4102 True 1 2 0.7979 0.0 0.5505 1.0453 True
-----
```

Results for capital-gain: F-value = 57861.408, 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 0 -2.7388 0.0 -3.3789 -2.0987 True -1 1
-2.8253 0.0 -3.4659 -2.1847 True -1 2 10.853 0.0 10.2075 11.4986 True 0 1 -0.0865 0.0 -0.1134
-0.0595 True 0 2 13.5919 0.0 13.508 13.6758 True 1 2 13.6783 0.0 13.5906 13.7661 True
-----
```

Results for capital-loss: F-value = 201971.779, 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 0 -4.5334 0.0 -4.8923 -4.1745 True -1 1
0.0431 0.9899 -0.3161 0.4022 False -1 2 -4.5334 0.0 -4.8953 -4.1715 True 0 1 4.5765 0.0 4.5613
4.5916 True 0 2 -0.0 1.0 -0.047 0.047 False 1 2 -4.5765 0.0 -4.6256 -4.5273 True
-----
```

Results for hours-per-week: F-value = 43.020, 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 0 1.5607 0.1198 -0.2515 3.3728 False -1 1
1.7868 0.0553 -0.0269 3.6004 False -1 2 2.3236 0.006 0.496 4.1511 True 0 1 0.2261 0.0 0.1498
0.3024 True 0 2 0.7629 0.0 0.5254 1.0004 True 1 2 0.5368 0.0 0.2885 0.7852 True
-----
```

Results for positive\_capital\_gain: F-value = 496.378, 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 0 -1.513 0.1223 -3.2773 0.2513 False -1 1
-1.8123 0.0417 -3.578 -0.0466 True -1 2 1.8123 0.044 0.0331 3.5916 True 0 1 -0.2993 0.0 -0.3737
-0.225 True 0 2 3.3253 0.0 3.0941 3.5566 True 1 2 3.6246 0.0 3.3828 3.8664 True
=====
```

Results for positive\_capital\_loss: F-value = 18487433.213, 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 0 -2.3351 0.0 -2.3734 -2.2969 True -1 1
2.3351 0.0 2.2968 2.3734 True -1 2 -2.3351 0.0 -2.3737 -2.2965 True 0 1 4.6702 0.0 4.6686 4.6718
True 0 2 -0.0 1.0 -0.005 0.005 False 1 2 -4.6702 0.0 -4.6755 -4.665 True
=====
```

Results for age\_education\_interaction: F-value = 140.020, 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 0 -1.2698 0.2682 -3.0714 0.5319 False -1
1 -0.8288 0.6389 -2.6319 0.9742 False -1 2 0.032 1.0 -1.7848 1.8489 False 0 1 0.4409 0.0 0.365
0.5168 True 0 2 1.3018 0.0 1.0657 1.5379 True 1 2 0.8608 0.0 0.6139 1.1078 True
=====
```

## Cluster Variability

```
{0: antecedent support consequent support support confidence \ count 182.000000 182.000000
182.000000 182.000000 mean 0.080317 0.181258 0.065087 0.842276 std 0.062253 0.066294
0.050362 0.186897 min 0.022148 0.034810 0.022148 0.449806 25% 0.043523 0.129324 0.033254
0.672234 50% 0.050605 0.222981 0.044682 0.965483 75% 0.083237 0.239548 0.076455 1.000000
max 0.243411 0.253884 0.222981 1.000000 lift leverage conviction zhangs_metric total_items \ count
182.000000 182.000000 182.000000 182.000000 182.000000 mean 5.525815 0.050729 inf 0.860124
2.945055 std 3.019219 0.038527 NaN 0.082389 0.228502 min 3.452059 0.016757 1.622436 0.741274
2.000000 25% 4.108270 0.026034 2.638277 0.792644 3.000000 50% 4.484697 0.034724 NaN
0.824137 3.000000 75% 4.484697 0.059005 NaN 0.949133 3.000000 max 13.268287 0.173260 inf
1.000000 3.000000 coverage count 182.000000 mean 0.080317 std 0.062253 min 0.022148 25%
0.043523 50% 0.050605 75% 0.083237 max 0.243411 , 1: antecedent support consequent support
support confidence \ count 184.000000 184.000000 184.000000 184.000000 mean 0.073916 0.166867
0.061412 0.861959 std 0.052685 0.040011 0.042978 0.162819 min 0.021115 0.065878 0.021115
0.471698 25% 0.036318 0.143370 0.031250 0.766758 50% 0.044764 0.181588 0.042652 0.911688
75% 0.102196 0.183277 0.082981 1.000000 max 0.183277 0.254223 0.181588 1.000000 lift leverage
conviction zhangs_metric total_items \ count 184.000000 184.000000 184.000000 184.000000
184.000000 mean 5.336811 0.049652 inf 0.874055 2.951087 std 1.048280 0.035256 NaN 0.066767
0.216275 min 3.865518 0.016540 1.699415 0.771143 2.000000 25% 4.694368 0.025555 3.557975
0.836589 3.000000 50% 5.456221 0.034557 9.027344 0.856007 3.000000 75% 5.506977 0.063956
NaN 0.910253 3.000000 max 10.385965 0.148614 inf 1.000000 3.000000 coverage count 184.000000
mean 0.073916 std 0.052685 min 0.021115 25% 0.036318 50% 0.044764 75% 0.102196 max
0.183277 , 2: antecedent support consequent support support confidence \ count 65.000000 65.000000
65.000000 65.000000 mean 0.066535 0.117817 0.057331 0.863557 std 0.031855 0.055339 0.029520
0.138327 min 0.042735 0.068376 0.042735 0.700000 25% 0.059829 0.068376 0.042735 0.714286
```

50% 0.059829 0.102564 0.042735 0.857143 75% 0.068376 0.136752 0.068376 1.000000 max  
0.247863 0.247863 0.196581 1.000000 lift leverage conviction zhangs\_metric total\_items coverage  
count 65.000000 65.000000 65.000000 65.000000 65.000000 65.000000 mean 8.794609 0.048647 inf  
0.921956 2.907692 0.066535 std 4.000200 0.021749 NaN 0.057876 0.291712 0.031855 min 4.034483  
0.032143 2.820513 0.785714 2.000000 0.042735 25% 5.571429 0.037475 3.200855 0.872727  
3.000000 0.059829 50% 7.800000 0.039813 NaN 0.936170 3.000000 0.059829 75% 14.625000  
0.051428 NaN 0.973214 3.000000 0.068376 max 14.625000 0.147856 inf 1.000000 3.000000  
0.247863 , -1: antecedent support consequent support support confidence lift \ count 886.0 886.0 886.0  
886.0 886.0 mean 0.5 0.5 0.5 1.0 2.0 std 0.0 0.0 0.0 0.0 0.0 min 0.5 0.5 0.5 1.0 2.0 25% 0.5 0.5 0.5 1.0  
2.0 50% 0.5 0.5 0.5 1.0 2.0 75% 0.5 0.5 0.5 1.0 2.0 max 0.5 0.5 0.5 1.0 2.0 leverage conviction  
zhangs\_metric total\_items coverage count 886.00 886.0 886.0 886.000000 886.0 mean 0.25 inf 1.0  
2.902935 0.5 std 0.00 NaN 0.0 0.296214 0.0 min 0.25 inf 1.0 2.000000 0.5 25% 0.25 NaN 1.0  
3.000000 0.5 50% 0.25 NaN 1.0 3.000000 0.5 75% 0.25 NaN 1.0 3.000000 0.5 max 0.25 inf 1.0  
3.000000 0.5 }

## Rule Metrics Comparison

0 1 2 -1 Whole Dataset count 182.000000 184.000000 65.000000 886.0 63.000000 mean 0.842276  
0.861959 0.863557 1.0 0.886183 std 0.186897 0.162819 0.138327 0.0 0.087088 min 0.449806  
0.471698 0.700000 1.0 0.711079 25% 0.672234 0.766758 0.714286 1.0 0.823729 50% 0.965483  
0.911688 0.857143 1.0 0.892037 75% 1.000000 1.000000 1.000000 1.0 0.954250 max 1.000000  
1.000000 1.000000 1.0 1.000000

## Top Unique Rules per Cluster

### Cluster 0:

Rule: frozenset({'education-num\_(13.0, 16.0]'}) -> frozenset({'sex\_Male', 'education\_Masters'})  
(Support: 0.035, Confidence: 0.450, Lift: 12.922)  
Rule: frozenset({'education\_Masters'}) -> frozenset({'occupation\_aggregated\_Prof-specialty',  
'education-num\_(13.0, 16.0]'}) (Support: 0.026, Confidence: 0.511, Lift: 10.793)  
Rule: frozenset({'occupation\_aggregated\_Prof-specialty', 'education-num\_(13.0, 16.0]'}) ->  
frozenset({'education\_Masters'}) (Support: 0.026, Confidence: 0.546, Lift: 10.793)  
Rule: frozenset({'occupation\_aggregated\_Prof-specialty', 'education-num\_(10.0, 13.0]'}) ->  
frozenset({'education\_Bachelors'}) (Support: 0.048, Confidence: 0.827, Lift: 4.974)  
Rule: frozenset({'education-num\_(13.0, 16.0]', 'marital-status\_Married-civ-spouse'}) ->  
frozenset({'occupation\_aggregated\_Prof-specialty'}) (Support: 0.027, Confidence: 0.588, Lift: 4.546)

### Cluster 1:

Rule: frozenset({'occupation\_aggregated\_Exec-managerial', 'education-num\_(13.0, 16.0]'}) ->  
frozenset({'education\_Masters'}) (Support: 0.039, Confidence: 0.807, Lift: 7.897)  
Rule: frozenset({'education\_Prof-school'}) -> frozenset({'occupation\_aggregated\_Prof-specialty',  
'education-num\_(13.0, 16.0]'}) (Support: 0.038, Confidence: 0.849, Lift: 7.674)  
Rule: frozenset({'education\_Prof-school'}) -> frozenset({'education-num\_(13.0, 16.0]',  
'hours-per-week\_(45.0, 99.0]'}) (Support: 0.028, Confidence: 0.623, Lift: 7.523)  
Rule: frozenset({'education-num\_(13.0, 16.0]', 'workclass\_Local-gov'}) ->  
frozenset({'education\_Masters'}) (Support: 0.022, Confidence: 0.765, Lift: 7.483)  
Rule: frozenset({'education\_Prof-school'}) -> frozenset({'occupation\_aggregated\_Prof-specialty',  
'hours-per-week\_(45.0, 99.0]'}) (Support: 0.025, Confidence: 0.566, Lift: 7.447)

### **Cluster 2:**

Rule: frozenset({'native\_country\_aggregated\_Other', 'workclass\_Private'}) -> frozenset({'race\_Asian-Pac-Islander'}) (Support: 0.043, Confidence: 0.714, Lift: 10.446)  
Rule: frozenset({'race\_Asian-Pac-Islander', 'workclass\_Private'}) -> frozenset({'native\_country\_aggregated\_Other'}) (Support: 0.043, Confidence: 1.000, Lift: 9.750)  
Rule: frozenset({'occupation\_aggregated\_Craft-repair'}) -> frozenset({'race\_White', 'age\_education\_interaction\_(261.0, 369.0]'}) (Support: 0.043, Confidence: 0.714, Lift: 9.286)  
Rule: frozenset({'occupation\_aggregated\_Craft-repair'}) -> frozenset({'age\_education\_interaction\_(261.0, 369.0]', 'native\_country\_aggregated\_United-States'}) (Support: 0.043, Confidence: 0.714, Lift: 9.286)  
Rule: frozenset({'race\_Asian-Pac-Islander', 'marital-status\_Married-civ-spouse'}) -> frozenset({'native\_country\_aggregated\_Other'}) (Support: 0.051, Confidence: 0.857, Lift: 8.357)

### **Cluster -1:**

Rule: frozenset({'sex\_Male'}) -> frozenset({'workclass\_Private'}) (Support: 0.500, Confidence: 1.000, Lift: 2.000)  
Rule: frozenset({'workclass\_Private'}) -> frozenset({'sex\_Male'}) (Support: 0.500, Confidence: 1.000, Lift: 2.000)  
Rule: frozenset({'race\_White', 'sex\_Male'}) -> frozenset({'workclass\_Private'}) (Support: 0.500, Confidence: 1.000, Lift: 2.000)  
Rule: frozenset({'race\_White', 'workclass\_Private'}) -> frozenset({'sex\_Male'}) (Support: 0.500, Confidence: 1.000, Lift: 2.000)  
Rule: frozenset({'sex\_Male'}) -> frozenset({'race\_White', 'workclass\_Private'}) (Support: 0.500, Confidence: 1.000, Lift: 2.000)

## **Cluster Visualizations**

