Automated Data Analysis Report

Table of Contents

1. Clustering Results

Best Parameters: {'epsilon': 4.935886088617702, 'min_samples': 4, 'silhouette': 0.740598372599426},
Best Silhouette Score: 0.741

Train Silhouette Score: 0.376, Test Silhouette Score: 0.676

2. ANOVA Results

Results for capital-gain: F-value = 102723.863, 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 13.6783 0.0 13.506 13.8506 True -1 2 0.0 1.0 -0.1629 0.1629 False -1 4 1.0492 0.0 0.8858 1.2127 True -1 5 -0.0 1.0 -0.1638 0.1638 False -1 6 0.0 1.0 -0.3454 0.3454 False 1 2 -13.6783 0.0 -13.7348 -13.6218 True 1 4 -12.6291 0.0 -12.6871 -12.5711 True 1 5 -13.6783 0.0 -13.7374 -13.6192 True 1 6 -13.6783 0.0 -13.9881 -13.3685 True 2 4 1.0492 0.0 1.0347 1.0637 True 2 5 -0.0 1.0 -0.0183 0.0183 False 2 6 -0.0 1.0 -0.3047 0.3047 False 4 5 -1.0492 0.0 -1.0718 -1.0266 True 4 6 -1.0492 0.0 -1.3542 -0.7443 True 5 6 0.0 1.0 -0.3052 0.3052 False ----------

Results for capital-loss: F-value = 140639.753, 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 -5.5055 0.0 -5.6536 -5.3573 True -1 2 -5.5055 0.0 -5.6455 -5.3654 True -1 4 -5.5055 0.0 -5.646 -5.3649 True -1 5 -0.953 0.0 -1.0939 -0.8121 True -1 6 3.9355 0.0 3.6385 4.2326 True 1 2 -0.0 1.0 -0.0486 0.0486 False 1 4 -0.0 1.0 -0.0499 0.0499 False 1 5 4.5525 0.0 4.5017 4.6033 True 1 6 9.441 0.0 9.1746 9.7074 True 2 4 -0.0 1.0 -0.0125 0.0125 False 2 5 4.5525 0.0 4.5367 4.5682 True 2 6 9.441 0.0 9.179 9.703 True 4 5 4.5525 0.0 4.533 4.5719 True 4 6 9.441 0.0 9.1787 9.7033 True 5 6 4.8885 0.0 4.6261 5.151 True
----

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 1 3.6246 0.0 3.6246 3.6246 True -1 2 -0.0 0.0 -0.0 -0.0 True -1 4 3.6246 0.0 3.6246 3.6246 True -1 5 0.0 0.9968 -0.0 0.0 False -1 6 0.0 1.0 -0.0 0.0 False 1 2 -3.6246 0.0 -3.6246 -3.6246 True 1 4 0.0 0.0 0.0 0.0 True 1 5 -3.6246 0.0 -3.6246 -3.6246 True 1 6 -3.6246 0.0 -3.6246 -3.6246 True 2 4 3.6246 0.0 3.6246 3.6246 True 2 5 0.0 0.0 0.0 0.0 True 2 6 0.0 0.0 0.0 0.0 True 4 5 -3.6246 0.0 -3.6246 -3.6246 True 4 6 -3.6246 0.0 -3.6246 -3.6246 True 5 6 -0.0 0.9998 -0.0 0.0 False ----------

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 1 -4.6702 0.0 -4.6702 -4.6702 True -1 2 -4.6702 0.0 -4.6702 -4.6702 True -1 4 -4.6702 0.0 -4.6702 -4.6702 True -1 5 0.0 0.6037 -0.0 0.0 False -1 6 -0.0 1.0 -0.0 0.0 False 1 2 0.0 0.0 0.0 0.0 True 1 4 -0.0 0.0 -0.0 -0.0 True 1 5 4.6702 0.0 4.6702 4.6702 True 1 6 4.6702 0.0 4.6702 4.6702 True 2 4 -0.0 0.0 -0.0 -0.0 True 2 5 4.6702 0.0 4.6702 4.6702 True 2 6 4.6702 0.0 4.6702 4.6702 True 4 5 4.6702 0.0 4.6702 4.6702 True 4 6 4.6702 0.0 4.6702 4.6702 True 5 6 -0.0 0.894 -0.0 0.0 False ----------

Results for age_education_interaction: F-value = 216.320, 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 0.2784 0.9162 -0.5105 1.0673 False -1 2 -1.0722 0.0006 -1.818 -0.3264 True -1 4 -0.4826 0.4411 -1.2308 0.2657 False -1 5 -0.5824 0.2316 -1.3324 0.1676 False -1 6 -1.5797 0.0505 -3.1612 0.0019 False 1 2 -1.3506 0.0 -1.6092 -1.092 True 1 4 -0.761 0.0 -1.0266 -0.4954 True 1 5 -0.8608 0.0 -1.1314 -0.5903 True 1 6 -1.8581 0.0026 -3.2766 -0.4397 True 2 4 0.5896 0.0 0.5232 0.656 True 2 5 0.4897 0.0 0.4059 0.5736 True 2 6 -0.5075 0.9056 -1.9025 0.8874 False 4 5 -0.0998 0.0661 -0.2033 0.0037 False 4 6 -1.0971 0.2197 -2.4933 0.2992 False 5 6 -0.9973 0.3228 -2.3945 0.3999 False --

3. Cluster Variability

		antecedent support	consequent support	support	confidence	\
	count	83.000000	83.000000	83.000000	83.000000	
	mean	0.104125	0.263196	0.059184	0.617148	
	std	0.069965	0.080833	0.036020	0.187646	
	min	0.029569	0.145504	0.025451	0.399310	
	25%	0.051160	0.212501	0.031744	0.440633	
	50%	0.075933	0.223496	0.044119	0.565044	
	75%	0.139164	0.345373	0.068635	0.757404	
	max	0.345373	0.438523	0.143165	0.991189	

		lift	leverage	conviction	zhangs_metric	total_items	coverage
	count	83.000000	83.000000	83.000000	83.000000	83.000000	83.000000
	mean	2.388806	0.033406	5.282772	0.636931	2.903614	0.104125
	std	0.450555	0.021110	12.445754	0.098381	0.296913	0.069965
	min	1.755118	0.011927	1.292688	0.467623	2.000000	0.029569
	25%	2.045526	0.018618	1.444429	0.570716	3.000000	0.051160
	50%	2.311114	0.023478	1.702638	0.613188	3.000000	0.075933
	75%	2.621471	0.041407	2.998351	0.711009	3.000000	0.139164
2	max	4.112995	0.087515	63.727590	0.933793	3.000000	0.345373

	antecedent support	consequent support	support	confidence	\
count	147.000000	147.000000	147.000000	147.000000	
mean	0.075314	0.256720	0.048884	0.703390	
std	0.050836	0.082657	0.028067	0.170253	
min	0.025707	0.073693	0.025707	0.436620	
25%	0.040274	0.200514	0.030848	0.571429	
50%	0.059126	0.215938	0.036847	0.679245	
75%	0.093402	0.302485	0.058269	0.852755	
max	0.215938	0.399314	0.125107	1.000000	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	147.000000	147.000000	147.000000	147.000000	147.000000	
mean	2.909279	0.030786	inf	0.682048	2.918367	
std	0.925584	0.018445	NaN	0.103968	0.274740	
min	2.009260	0.013895	1.405398	0.534357	2.000000	
25%	2.215313	0.018680	1.747030	0.593385	3.000000	
50%	2.504292	0.022603	2.640592	0.622558	3.000000	
75%	3.506888	0.035780	4.641011	0.790896	3.000000	
max	7.681000	0.081808	inf	0.911191	3.000000	

	coverage
count	147.000000
mean	0.075314
std	0.050836
min	0.025707
25%	0.040274
50%	0.059126
75%	0.093402
max	0.215938

	antecedent support	consequent support	support	confidence	\
count	113.000000	113.000000	113.000000	113.000000	
mean	0.085664	0.248119	0.051897	0.652948	
std	0.048383	0.093110	0.024983	0.149541	
min	0.030649	0.131429	0.027532	0.437838	
25%	0.044156	0.180260	0.032727	0.495536	
50%	0.066494	0.206234	0.040000	0.637755	
75%	0.112208	0.301299	0.064935	0.795918	
max	0.194805	0.417662	0.108052	0.932203	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	113.000000	113.000000	113.000000	113.000000	113.000000	
mean	2.874718	0.031921	2.667643	0.674165	2.929204	
std	0.969473	0.017330	1.293131	0.129059	0.257627	
min	1.853723	0.014725	1.365784	0.487838	2.000000	
25%	2.021835	0.019518	1.552536	0.552853	3.000000	
50%	2.362268	0.025944	2.215947	0.641679	3.000000	
75%	3.422222	0.037441	3.472403	0.793776	3.000000	
max	5.350195	0.074657	8.589481	0.879032	3.000000	

		coverage
	count	113.000000
	mean	0.085664
	std	0.048383
	min	0.030649
	25%	0.044156
	50%	0.066494
	75%	0.112208
4	max	0.194805

	antecedent support	consequent support	support	confidence	\
count	22.000000	22.000000	22.000000	22.000000	
mean	0.058275	0.112277	0.043901	0.760823	
std	0.006262	0.047937	0.003002	0.089156	
min	0.042735	0.068376	0.042735	0.714286	
25%	0.059829	0.085470	0.042735	0.714286	
50%	0.059829	0.102564	0.042735	0.714286	
75%	0.059829	0.128205	0.042735	0.750000	
max	0.068376	0.307692	0.051282	1.000000	

	lift	leverage	conviction	zhangs_metric	total_items	coverage
count	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000
mean	7.344746	0.037515	inf	0.907050	2.863636	0.058275
std	1.756132	0.003530	NaN	0.050948	0.351250	0.006262
min	3.250000	0.029586	3.021368	0.723214	2.000000	0.042735
25%	5.785714	0.035320	3.118590	0.877518	3.000000	0.059829
50%	7.644886	0.037548	3.200855	0.925465	3.000000	0.059829
75%	8.357143	0.038133	3.615385	0.936364	3.000000	0.059829
1 max	10.446429	0.045146	inf	0.961818	3.000000	0.068376

	antecedent support	consequent support	support	confidence	\
count	2.340000e+02	2.340000e+02	2.340000e+02	234.0	
mean	7.142857e-02	7.142857e-02	7.142857e-02	1.0	
std	2.781507e-17	2.781507e-17	2.781507e-17	0.0	
min	7.142857e-02	7.142857e-02	7.142857e-02	1.0	
25%	7.142857e-02	7.142857e-02	7.142857e-02	1.0	
50%	7.142857e-02	7.142857e-02	7.142857e-02	1.0	
75%	7.142857e-02	7.142857e-02	7.142857e-02	1.0	
max	7.142857e-02	7.142857e-02	7.142857e-02	1.0	

	lift	leverage	conviction	zhangs_metric	total_items	\
count	234.0	2.340000e+02	234.0	2.340000e+02	234.000000	
mean	14.0	6.632653e-02	inf	1.000000e+00	2.982906	
std	0.0	1.390754e-17	NaN	1.112603e-16	0.129900	
min	14.0	6.632653e-02	inf	1.000000e+00	2.000000	
25%	14.0	6.632653e-02	NaN	1.000000e+00	3.000000	
50%	14.0	6.632653e-02	NaN	1.000000e+00	3.000000	
75%	14.0	6.632653e-02	NaN	1.000000e+00	3.000000	
max	14.0	6.632653e-02	inf	1.000000e+00	3.000000	

	coverage
count	2.340000e+02
mean	7.142857e-02
std	2.781507e-17
min	7.142857e-02
25%	7.142857e-02
50%	7.142857e-02
75%	7.142857e-02
-1 max	7.142857e-02

	antecedent support	consequent support	support	confidence	lift	\
count	212.00	212.00	212.00	212.0	212.0	
mean	0.25	0.25	0.25	1.0	4.0	
std	0.00	0.00	0.00	0.0	0.0	
min	0.25	0.25	0.25	1.0	4.0	
25%	0.25	0.25	0.25	1.0	4.0	
50%	0.25	0.25	0.25	1.0	4.0	
75%	0.25	0.25	0.25	1.0	4.0	
max	0.25	0.25	0.25	1.0	4.0	

	leverage	conviction	zhangs_metric	total_items	coverage
count	212.0000	212.0	212.0	212.000000	212.00
mean	0.1875	inf	1.0	2.933962	0.25
std	0.0000	NaN	0.0	0.248936	0.00
min	0.1875	inf	1.0	2.000000	0.25
25%	0.1875	NaN	1.0	3.000000	0.25
50%	0.1875	NaN	1.0	3.000000	0.25
75%	0.1875	NaN	1.0	3.000000	0.25
6 max	0.1875	inf	1.0	3.000000	0.25

4. Rule Metrics Comparison

mean	std	min	25%	50%	75%
171477424500899	0.18764582612303182	0.3993103448275862	0.44063303817731947	0.5650438946528332	0.7574041811
033898210509437	0.1702528482568039	0.4366197183098592	0.5714285714285715	0.6792452830188679	0.8527551942
52947757603941	0.14954099490694897	0.4378378378378378	0.4955357142857143	0.6377551020408162	0.7959183673
60822510822511	0.089155921603344	0.7142857142857143	0.7142857142857143	0.7142857142857143	0.7499999999
1.0	0.0	1.0	1.0	1.0	1.0
1.0	0.0	1.0	1.0	1.0	1.0
258966153085457	0.07743962136650274	0.6116504854368933	0.6736051193557664	0.745942217172537	0.7863387402

5. Top Unique Rules per Cluster

Cluster 2:

Rule: frozenset({'relationship_Wife'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.042, Confidence: 0.991, Lift: 2.260)
Rule: frozenset({'race_White', 'relationship_Wife'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.035, Confidence: 0.991, Lift: 2.259)
Rule: frozenset({'workclass_Private', 'relationship_Wife'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.029, Confidence: 0.991, Lift: 2.259)
Rule: frozenset({'native_country_aggregated_United-States', 'relationship_Wife'}) -> frozenset({'marital-status_Married-civ-spouse'}) (Support: 0.036, Confidence: 0.990, Lift: 2.257)
Rule: frozenset({'relationship_Own-child', 'hours_per_week_binned_21-30'}) -> frozenset({'marital-status_Never-married'}) (Support: 0.029, Confidence: 0.957, Lift: 2.770)

Cluster 5:

Rule: frozenset({'education_Doctorate', 'marital-status_Married-civ-spouse'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.031, Confidence: 1.000, Lift: 2.504)
Rule: frozenset({'education_Doctorate', 'sex_Male'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.027, Confidence: 1.000, Lift: 2.504)
Rule: frozenset({'occupation_aggregated_Prof-specialty', 'education_Doctorate'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.026, Confidence: 1.000, Lift: 2.504)
Rule: frozenset({'education_Doctorate', 'native_country_aggregated_United-States'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.030, Confidence: 1.000, Lift: 2.504)
Rule: frozenset({'education_Doctorate', 'race_White'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.033, Confidence: 1.000, Lift: 2.504)

Cluster 4:

Rule: frozenset({'workclass_Private', 'education_Masters'}) -> frozenset({'age_education_interaction_(494.0, 1350.0]'}) (Support: 0.041, Confidence: 0.788, Lift: 1.886)

Rule: frozenset({'hours_per_week_binned_51+', 'relationship_Not-in-family'}) -> frozenset({'marital-status_Never-married'}) (Support: 0.033, Confidence: 0.583, Lift: 3.403)
Rule: frozenset({'workclass_Self-emp-inc', 'native_country_aggregated_United-States'}) -> frozenset({'hours_per_week_binned_51+'}) (Support: 0.035, Confidence: 0.563, Lift: 1.869)
Rule: frozenset({'workclass_Self-emp-inc', 'marital-status_Married-civ-spouse'}) -> frozenset({'hours_per_week_binned_51+'}) (Support: 0.033, Confidence: 0.562, Lift: 1.867)
Rule: frozenset({'occupation_aggregated_Craft-repair', 'hours_per_week_binned_41-50'}) -> frozenset({'education_HS-grad'}) (Support: 0.043, Confidence: 0.497, Lift: 1.906)

Cluster 1:

Rule: frozenset({'occupation_aggregated_Prof-specialty', 'race_Asian-Pac-Islander'}) -> frozenset({'education_Prof-school'}) (Support: 0.043, Confidence: 1.000, Lift: 3.250)
Rule: frozenset({'workclass_Private', 'race_Asian-Pac-Islander'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.043, Confidence: 1.000, Lift: 9.750)
Rule: frozenset({'marital-status_Married-civ-spouse', 'race_Asian-Pac-Islander'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.051, Confidence: 0.857, Lift: 8.357)
Rule: frozenset({'race_Asian-Pac-Islander', 'sex_Male'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.043, Confidence: 0.833, Lift: 8.125)
Rule: frozenset({'race_Asian-Pac-Islander'}) -> frozenset({'native_country_aggregated_Other'}) (Support: 0.051, Confidence: 0.750, Lift: 7.312)

Cluster -1:

Rule: frozenset({'occupation_aggregated_Other', 'marital-status_Married-civ-spouse'}) -> frozenset({'education_5th-6th'}) (Support: 0.071, Confidence: 1.000, Lift: 14.000)
Rule: frozenset({'occupation_aggregated_Sales', 'age_education_interaction_(494.0, 1350.0]'}) -> frozenset({'marital-status_Separated'}) (Support: 0.071, Confidence: 1.000, Lift: 14.000)
Rule: frozenset({'occupation_aggregated_Sales', 'race_Black'}) -> frozenset({'marital-status_Separated'}) (Support: 0.071, Confidence: 1.000, Lift: 14.000)
Rule: frozenset({'marital-status_Separated'}) -> frozenset({'occupation_aggregated_Sales', 'race_Black'}) (Support: 0.071, Confidence: 1.000, Lift: 14.000)
Rule: frozenset({'race_Black', 'education_Some-college'}) -> frozenset({'marital-status_Separated'}) (Support: 0.071, Confidence: 1.000, Lift: 14.000)

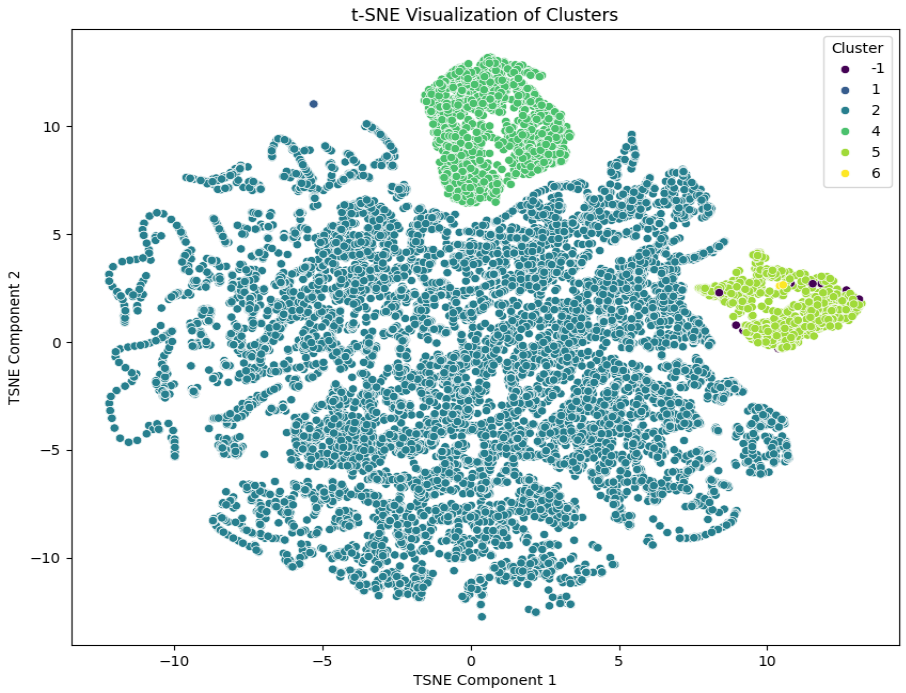
Cluster 6:

Rule: frozenset({'education_HS-grad'}) -> frozenset({'age_education_interaction_(261.0, 369.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'native_country_aggregated_United-States', 'education_Some-college'}) -> frozenset({'age_education_interaction_(369.0, 494.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'education_Some-college'}) -> frozenset({'marital-status_Separated', 'age_education_interaction_(369.0, 494.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'Cluster_(2.0, 6.0]', 'education_Some-college'}) -> frozenset({'age_education_interaction_(369.0, 494.0]'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)
Rule: frozenset({'Cluster_(2.0, 6.0]', 'age_education_interaction_(369.0, 494.0]'}) -> frozenset({'education_Some-college'}) (Support: 0.250, Confidence: 1.000, Lift: 4.000)

6. Top 10 Common Rules Sorted by Absolute Coverage Difference

Rule: frozenset({'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.294)
Rule: frozenset({'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.273)
Rule: frozenset({'native_country_aggregated_United-States', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.261)
Rule: frozenset({'native_country_aggregated_United-States', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.257)
Rule: frozenset({'workclass_Private', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.242)
Rule: frozenset({'race_White', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.241)
Rule: frozenset({'native_country_aggregated_United-States', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.237)
Rule: frozenset({'workclass_Private', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.236)
Rule: frozenset({'native_country_aggregated_United-States', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.236)
Rule: frozenset({'race_White', 'marital-status_Never-married', 'relationship_Own-child'}) (Abs Coverage Difference: 0.233)

7. Cluster Visualizations




PCA Visualization of Clusters
Cluster
-1
1
2
4
5
6
Principal Component 2
Principal Component 1
12
10
8
6
4
2
0
−2
−4
0
2
4
6
8
10