SWENG 545: FINAL EXAM

OBJECTIVE:
For the past seven years, the Graphics, Visualization and, Users Department of the Georgia Institute of Technology in Atlanta has conducted an international survey of World Wide Web usage as a public service in order to provide information concerning the demographics and trends of Internet access. The objective of this assignment is to obtain a profile of the "typical" Internet user by applying clustering techniques. Such data mining task can be advantageous to e-commerce marketers so that they may tailor their advertisements to a particular people-set. These tasks may also assist to software engineers and system designers would be interested in understanding why a particular subset of the population is still uncomfortable using computers. In addition, the following tasks have been proposed, what are the typical groups of web users, explain differences and similarities among groups, and suggest methods of better targeting the most important customers. To complete this objective and additional tasks the General Demographics dataset from the GVU's Tenth WWW User Survey (October 1998), R Studio, and Excel will be used.

DESCRIPTION:
The General Demographics dataset is divided into eight sections, including General Demographics, Technology Demographics, Online Privacy and Security, Web and Internet Use, Software Filters and Content Rating (Vanderbilt), Everyday Life (Vanderbilt), Electronic Commerce, and Specialized Questionnaires. Each data section includes a unique user identifier. In preparation for the clustering task the data sections General Demographics and Web and Internet Use were combined. Initially the General Demographics data set contained 5,022 observations and 106 variables of both numeric and nominal value and the Web and Internet Use data set contained 3,291 observations and 126 variables of both numeric and nominal value as well. Once all NAs were removed from the two data sets the General Demographics data set contained 4,767 observations and 105 variables of both numeric and nominal value and the Web and Internet Use data set contained 3,291 observations and 125 variables of both numeric and nominal value as well. The two data sets were then merged into one data set titled "prep" containing 3,084 observations and 229 variables of both numeric and nominal value. The new data set "prep" contains only the users found in both of the original General Demographics and Web and Internet Use data sets since unique users will not help capture trends as smoothly during the clustering analysis.

PREPROCESSING:
The preprocessing tasks are necessity in ensuring less relevant or even error filled data is not used to produce business decisions. Without quality preprocessing several negative factors affecting a dataset such as noise, incomplete, inconsistent, or unmanageable size would never be addressed in the data mining process. Preprocessing ensures an adequate amount and proper format for a dataset is achieved before any substantial analysis is performed.

Task 1: Data Integration
The data integration task is a necessity in the data preprocessing process to ensure any redundancies or inconsistencies in the data are handled, resulting in an increase of the accuracy of the subsequent data mining process. The issues examined in this step were data integration, schema integration, value conflict detection, and redundancy detection. To address this task in the *prep* data set the column titled "who" contained all the individual user IDs. This is great for the identification of user observations during the latter sections of this exam, so, for the initial goal of cluster analysis the variable "who" will be a distraction. As a result, the variable "who" will be removed from the *prep* data set for now, but the user IDs will remain, reducing the total number of variables in the *prep* data set to 228.

Task 2: Data Cleaning

The data cleaning task is a necessity in the data preprocessing process to ensure bias isn't introduced, limit the generalization of findings, and to avoid the result of misleading conclusions. The identification and handling of missing data was already completed before the preprocessing stage. In addition, the following issues were examined as well, identification of outliers and the smoothing of out noisy data, correction of inconsistent data, and the resolve of redundancy caused by data integration. Variables (or factors) with only 1 level will not contain clusters of observations during the application of clustering techniques, therefore these 1 level variables act as noise within the data set. To address this issue variables with only 1 level were removed from the *prep* data set during this task.

Task 3: Data Transformation
The data transformation task is a necessity in the data preprocessing process to ensure the mapping of the data from its given format into the suitable format for visualization and model generation.

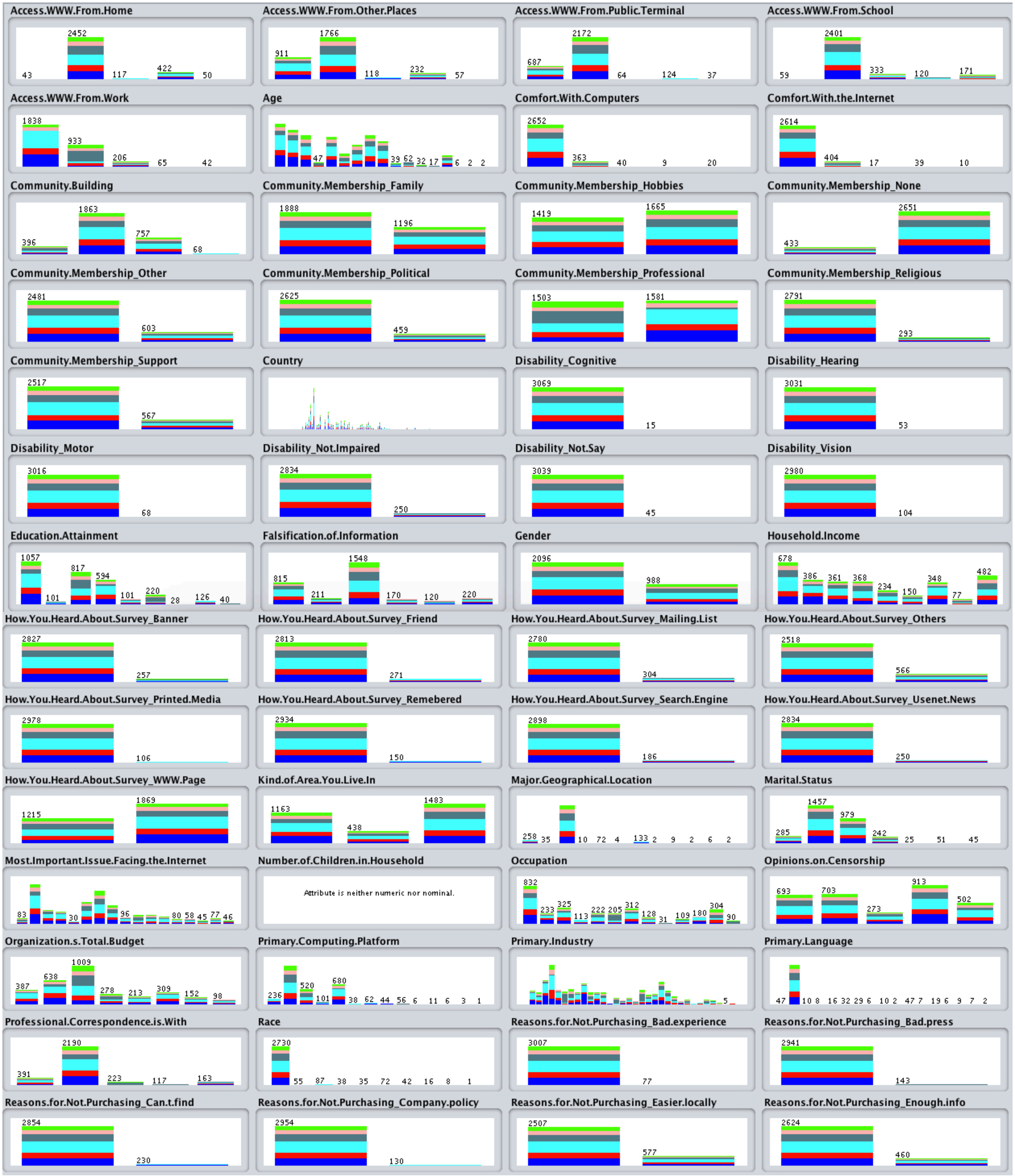
-Smoothing and Discretization-
In order to remove noise and convert continuous variables into nominal range variables binary levels of 0 and 1 were converted into N(no) and Y(yes).

Task 4: Data Reduction
The data reduction task is a necessity in the data preprocessing process to decrease the probability of obtaining an invalid model.

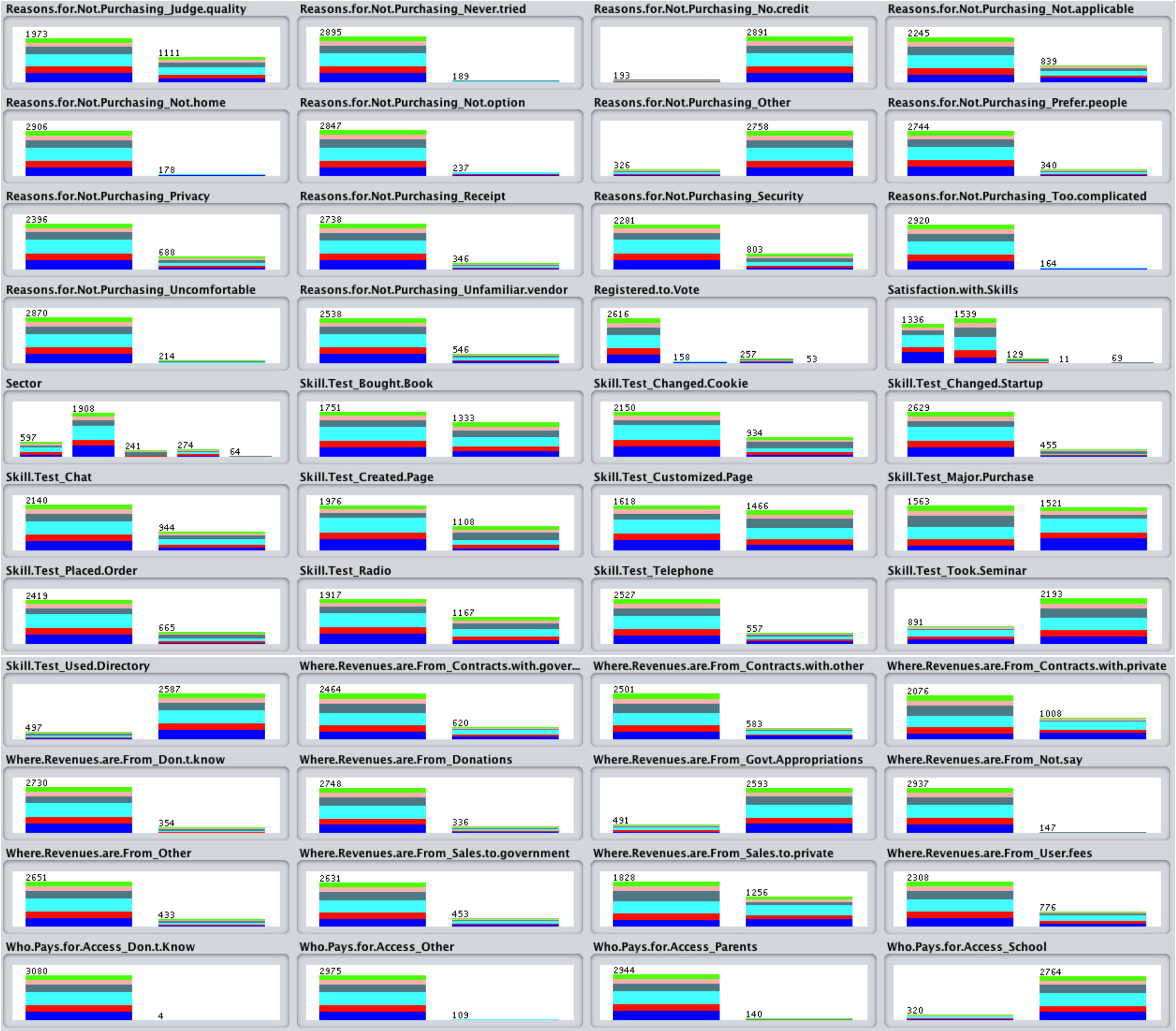
1) Merged two files into one = Prep data set now has 3084 observations and 229 variables
2) 1 level variable removal = New_Prep data set now has 3084 observations and 226 variables
3) User.Id as variable removal = Prep data set now has 3084 observations and 225 variables

Then saved file as csv format and continued pre-processing in Excel. In Excel I removed all apostrophes and converted the numeric instances for the variable Number.of.Children.in.Household to zero, one, two, three, and 4 or more, so that the instances would be recognized as a nominal value. Then uploaded the Prep data set into Weka for further analysis.

Weka Visualizations
The Prep data set has 3084 instances and 225 attributes all of nominal value.



Attribute is neither numeric nor nominal.

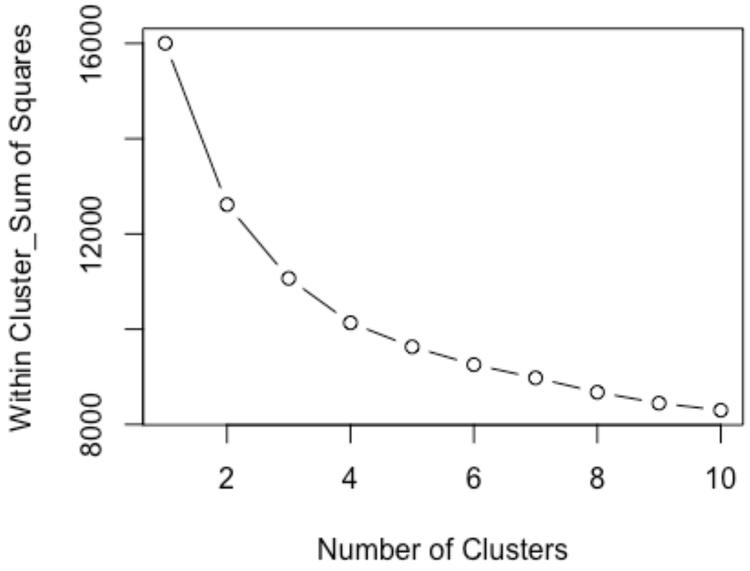


Section 4: *Clustering Technique: K-means Algorithm*

The K-means Algorithm is a partitional clustering approach that clusters into a predefined number of clusters. The K-means Algorithm associates each cluster with a center point called, centroid, which does no have to be a real point in the data set. For each point in the data set we calculate the distance to each centroid and finally the point is assigned to the cluster with the closest centroid.

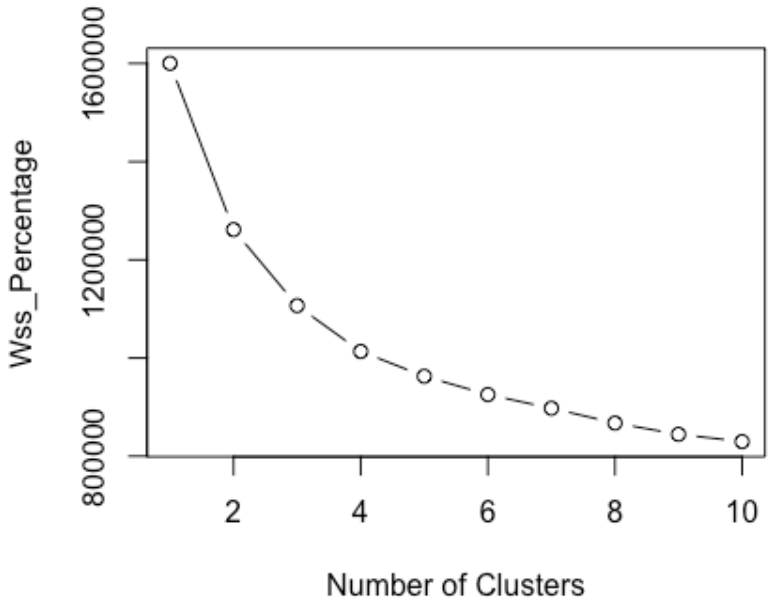
To perform the K-means Algorithm on the Prep dataset the SimpleKMeans operator was used under the Cluster tab in Weka Explorer. This technique was performed on the Prep dataset.

Gower's Distance / KMeans: Within Cluster_Sum of Squares



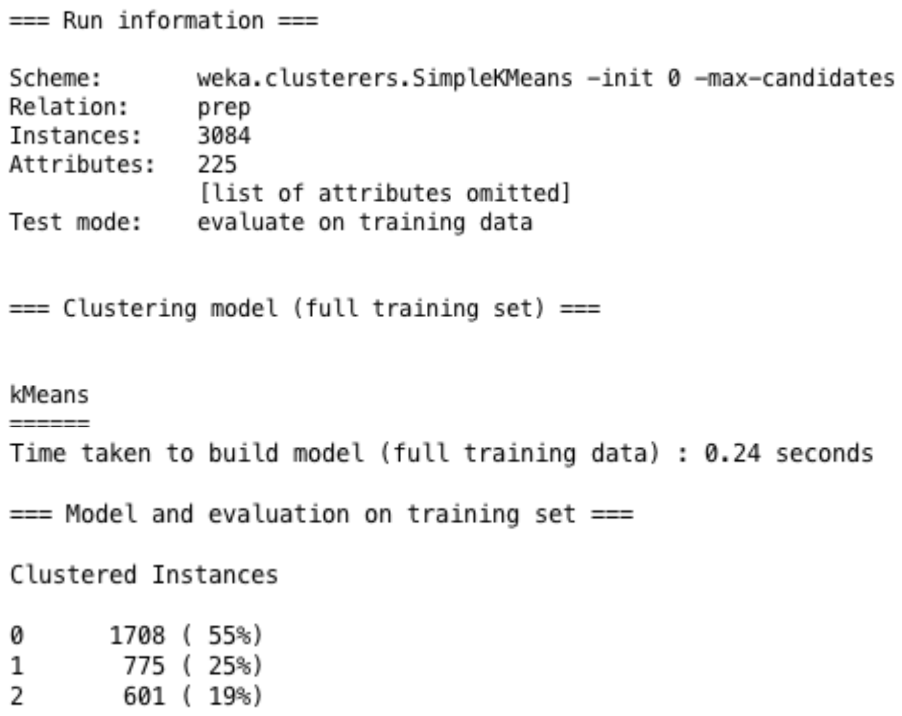
In this plot it is apparent the elbow occurs at k = 3.

KMeans: Wss Percentage Drop



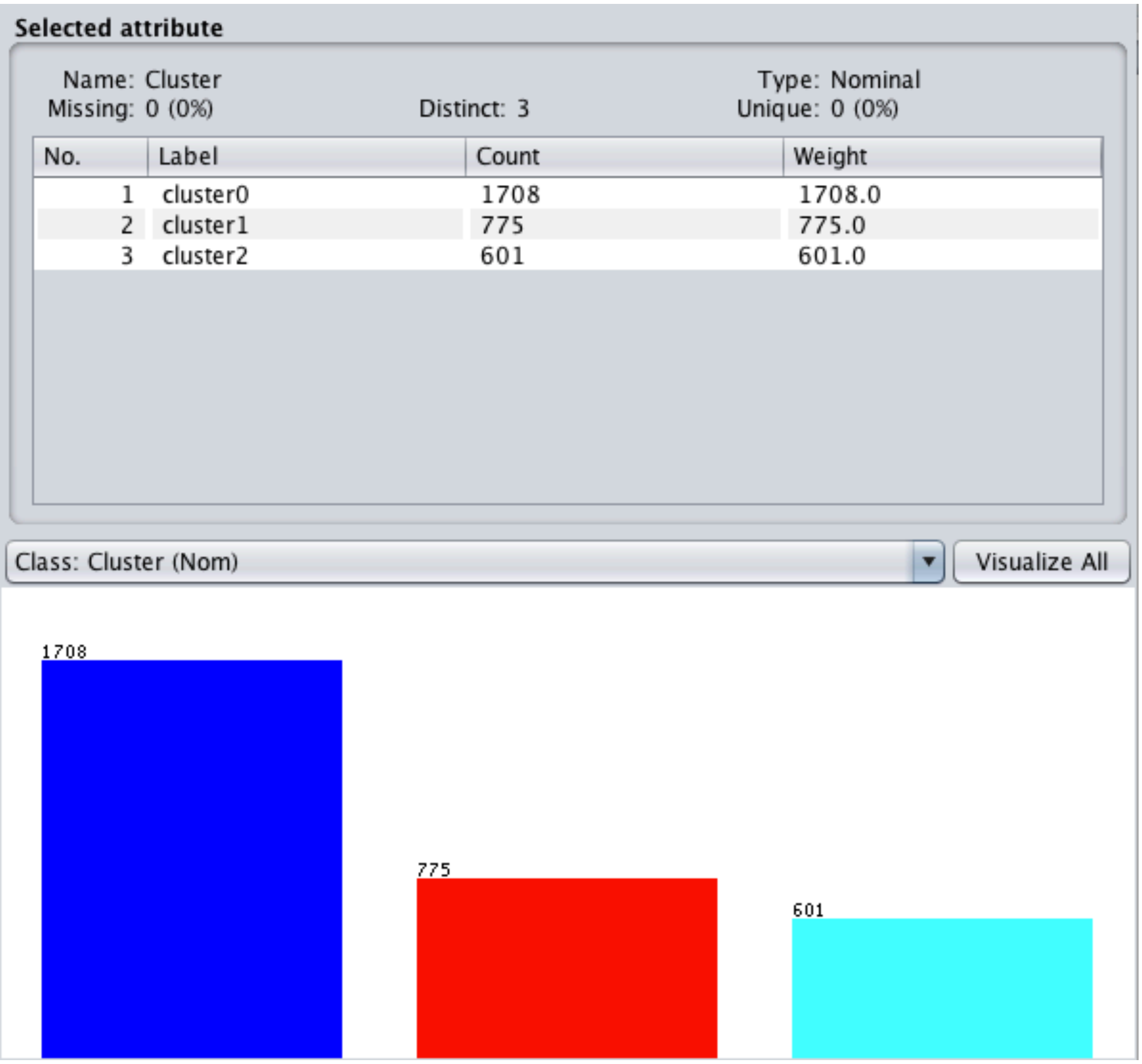
In the plot it appears the best number of clusters is also when k = 3 with minimum wss.

Simplified KMeans with Number of Folds Set to 3



```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -init 0 -max-candidates
Relation:     prep
Instances:    3084
Attributes:   225
              [list of attributes omitted]
Test mode:    evaluate on training data


=== Clustering model (full training set) ===


kMeans
======
Time taken to build model (full training data) : 0.24 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      1708 ( 55%)
1       775 ( 25%)
2       601 ( 19%)
```

Therefore, for this analysis 3 clusters were identified with a total of 3084 observations and 226 attributes. Cluster 0 contains 1708 instances making up 55% of the observations from the data set. Cluster 1 contains 775 instances making up 25% of the observations from the data set. Cluster 2 contains 601 instances making up 19% of the observations from the data set.



Selected attribute

Name: Cluster
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	cluster0	1708	1708.0
2	cluster1	775	775.0
3	cluster2	601	601.0

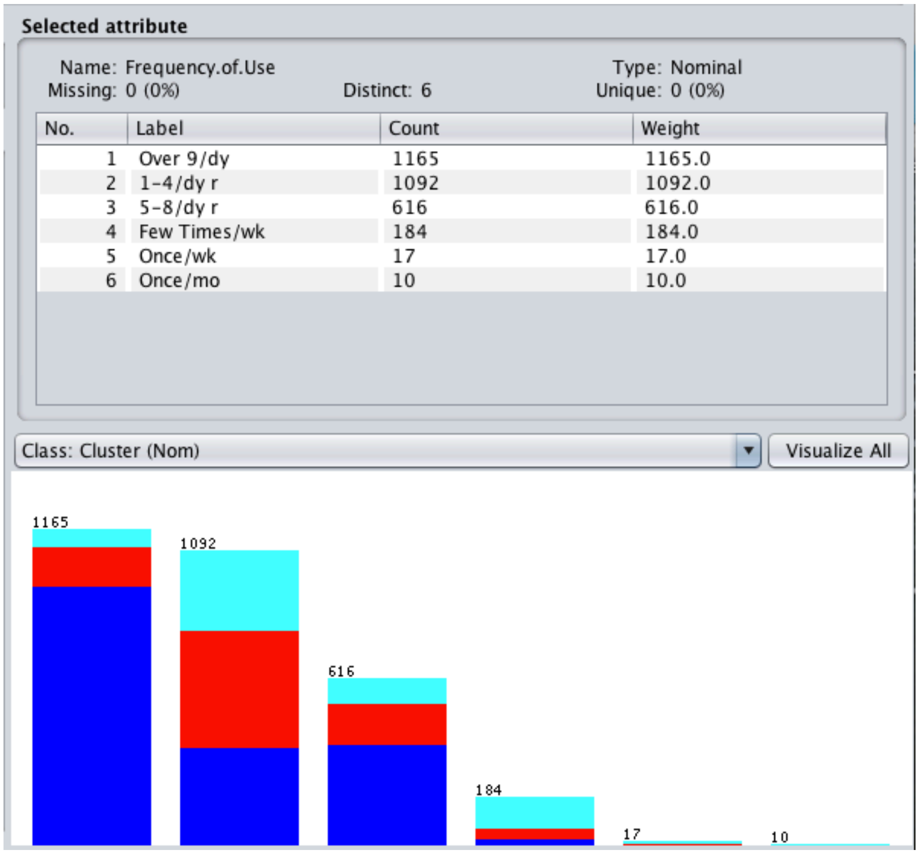
Class: Cluster (Nom)

Cluster 0 will be called Blue moving forward at 1708, Cluster 1 will be called Red moving forward at 775, and Cluster 2 will be called Teal moving forward at 601.

Section 5: *Evaluation - Profiling Clusters for Analysis*
From the 226 attributes the following 7 profiles were identified each containing subsets of information used to develop characteristics of each cluster.

- Profiling Clusters -
1) Web Access Profile

- Subgroups -
A) Frequency of Use



Selected attribute

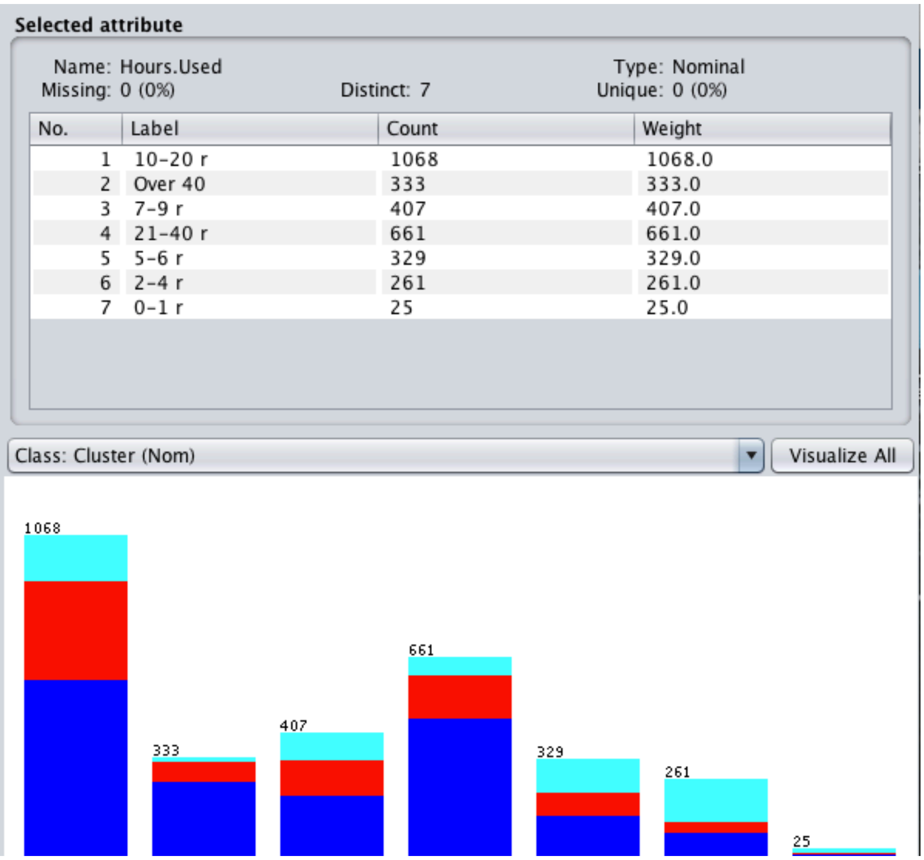
Name: Frequency.of.Use
Missing: 0 (0%)
Distinct: 6
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Over 9/dy	1165	1165.0
2	1-4/dy r	1092	1092.0
3	5-8/dy r	616	616.0
4	Few Times/wk	184	184.0
5	Once/wk	17	17.0
6	Once/mo	10	10.0

Class: Cluster (Nom)

Cluster Blue – High Use, Cluster Red – High Use, Cluster Teal – High Use

B) Hours Used



Selected attribute

Name: Hours.Used
Missing: 0 (0%)
Distinct: 7
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	10-20 r	1068	1068.0
2	Over 40	333	333.0
3	7-9 r	407	407.0
4	21-40 r	661	661.0
5	5-6 r	329	329.0
6	2-4 r	261	261.0
7	0-1 r	25	25.0

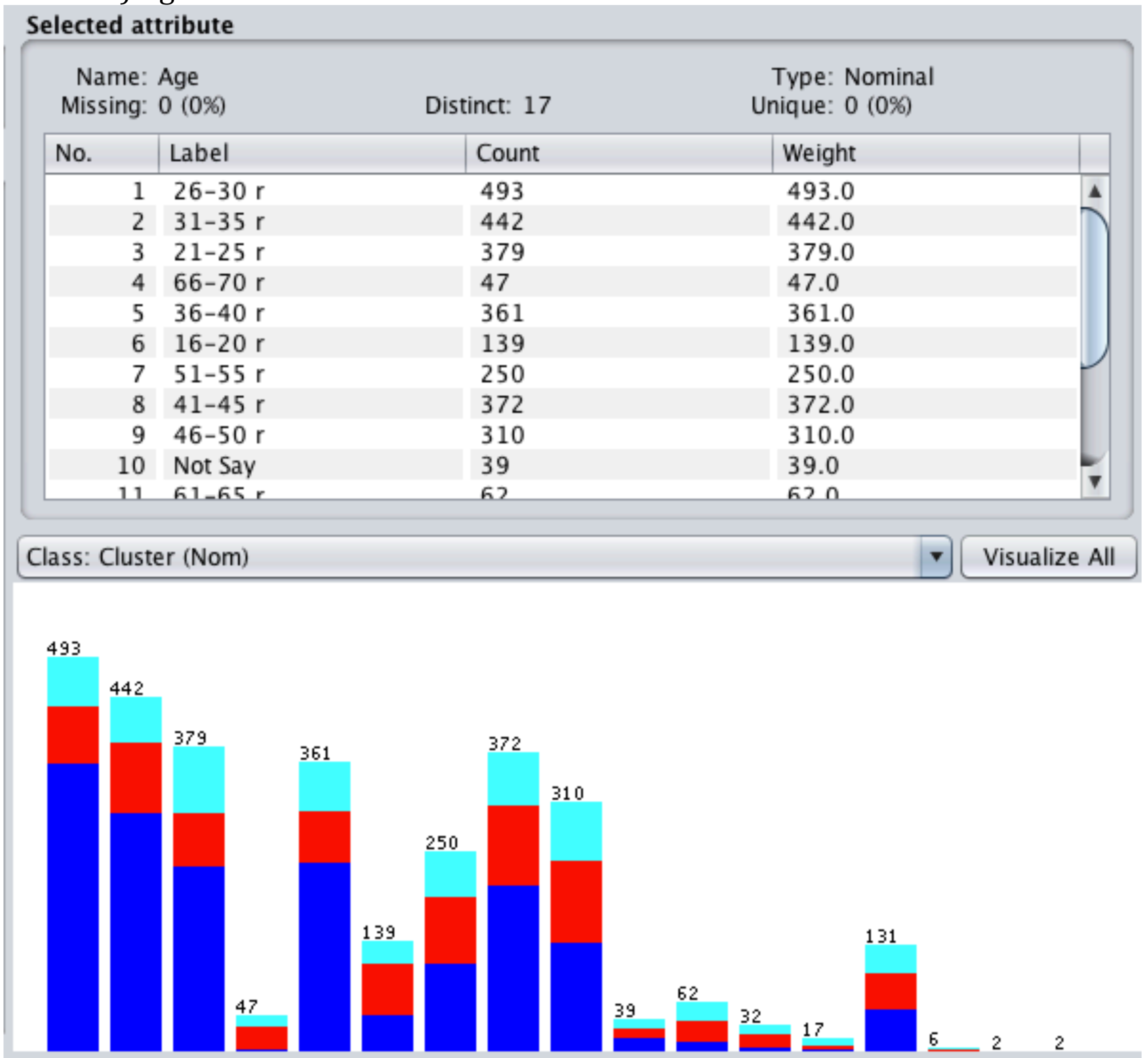
Class: Cluster (Nom)

Cluster Blue – High - Moderate Use, Cluster Red – Moderate, Cluster Teal – Low - Moderate

2) User Profile

- Subgroup –

A) Age



Selected attribute

Name: Age Type: Nominal

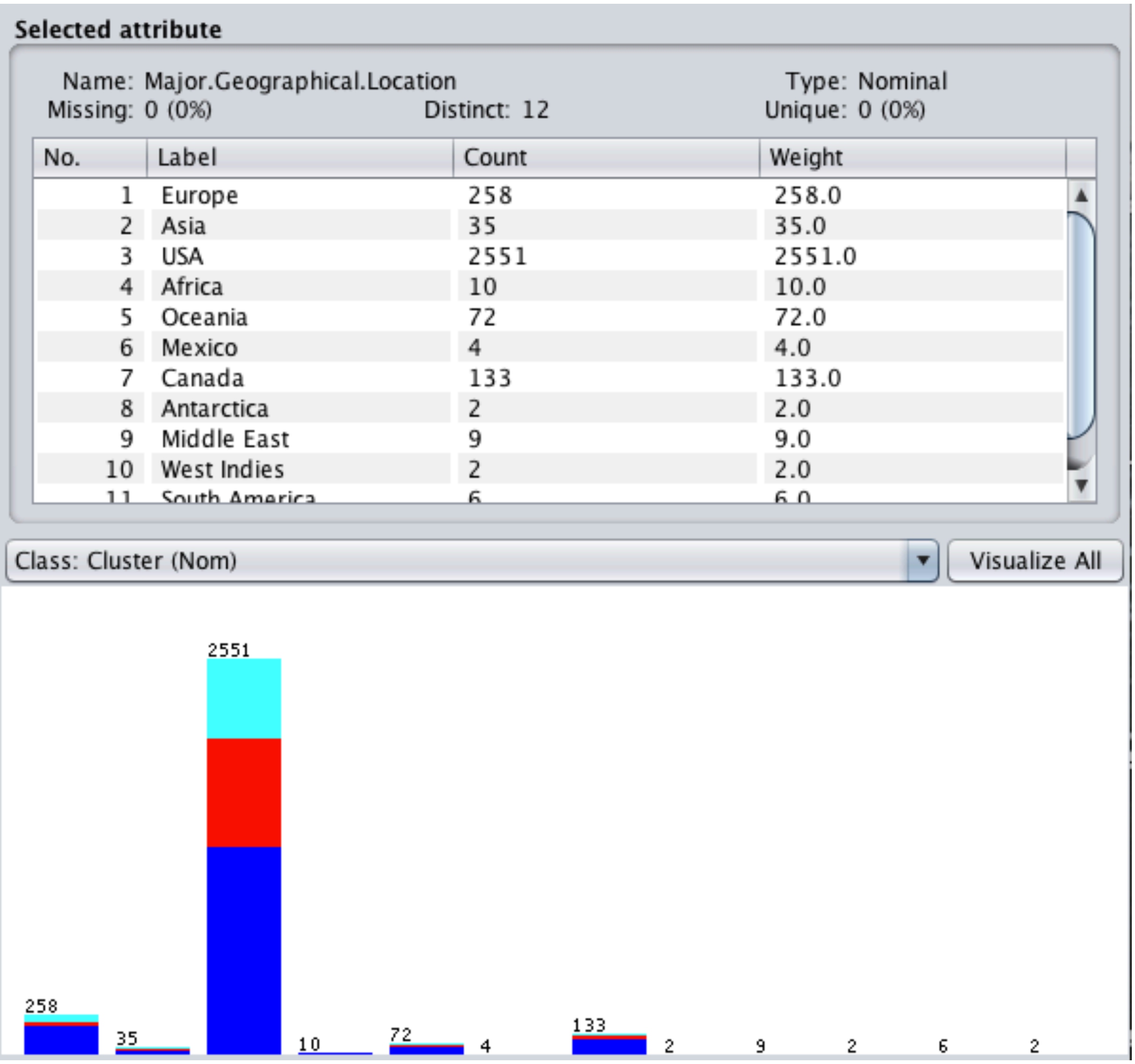
Missing: 0 (0%) Distinct: 17 Unique: 0 (0%)

No.	Label	Count	Weight
1	26-30 r	493	493.0
2	31-35 r	442	442.0
3	21-25 r	379	379.0
4	66-70 r	47	47.0
5	36-40 r	361	361.0
6	16-20 r	139	139.0
7	51-55 r	250	250.0
8	41-45 r	372	372.0
9	46-50 r	310	310.0
10	Not Say	39	39.0
11	61-65 r	62	62.0

Class: Cluster (Nom)

Cluster Blue – Ages 21 to 35 and 36 to 40, Cluster Red – Not Clear, Cluster Teal – Not Clear

B) Major Geographical Location



Selected attribute

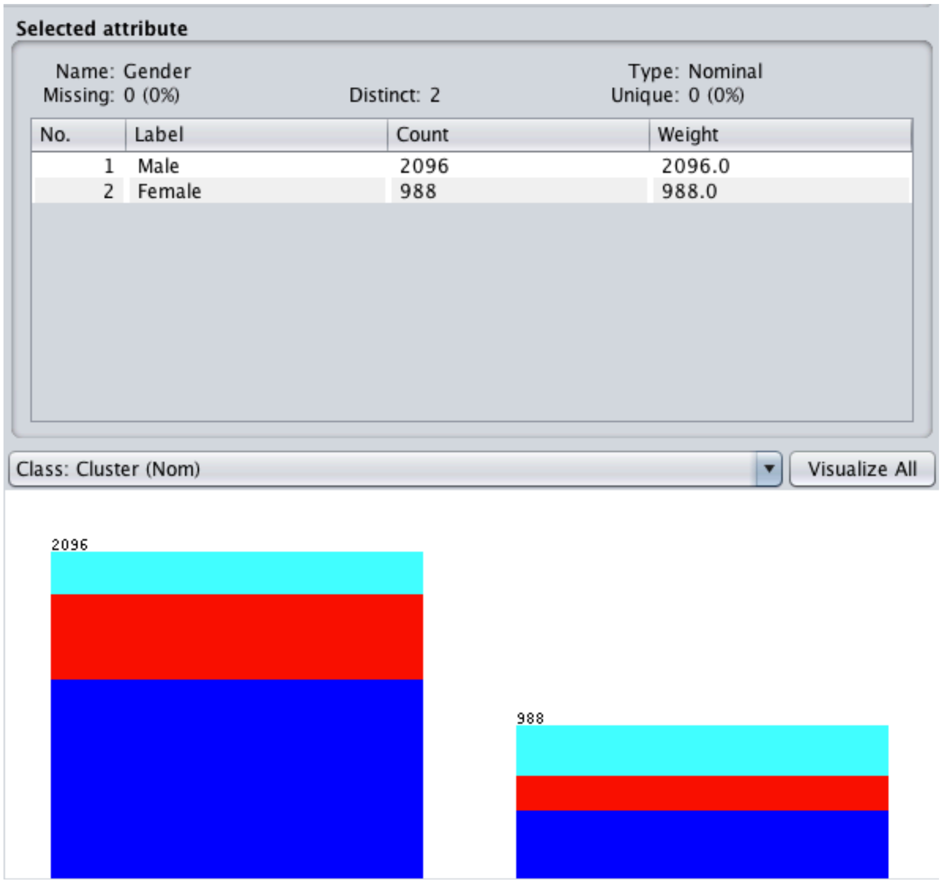
Name: Major.Geographical.Location
Type: Nominal
Missing: 0 (0%)
Distinct: 12
Unique: 0 (0%)

No.	Label	Count	Weight
1	Europe	258	258.0
2	Asia	35	35.0
3	USA	2551	2551.0
4	Africa	10	10.0
5	Oceania	72	72.0
6	Mexico	4	4.0
7	Canada	133	133.0
8	Antarctica	2	2.0
9	Middle East	9	9.0
10	West Indies	2	2.0
11	South America	6	6.0

Class: Cluster (Nom)

Cluster Blue – USA, Cluster Red – USA, Cluster Teal - USA

C) Gender



Selected attribute

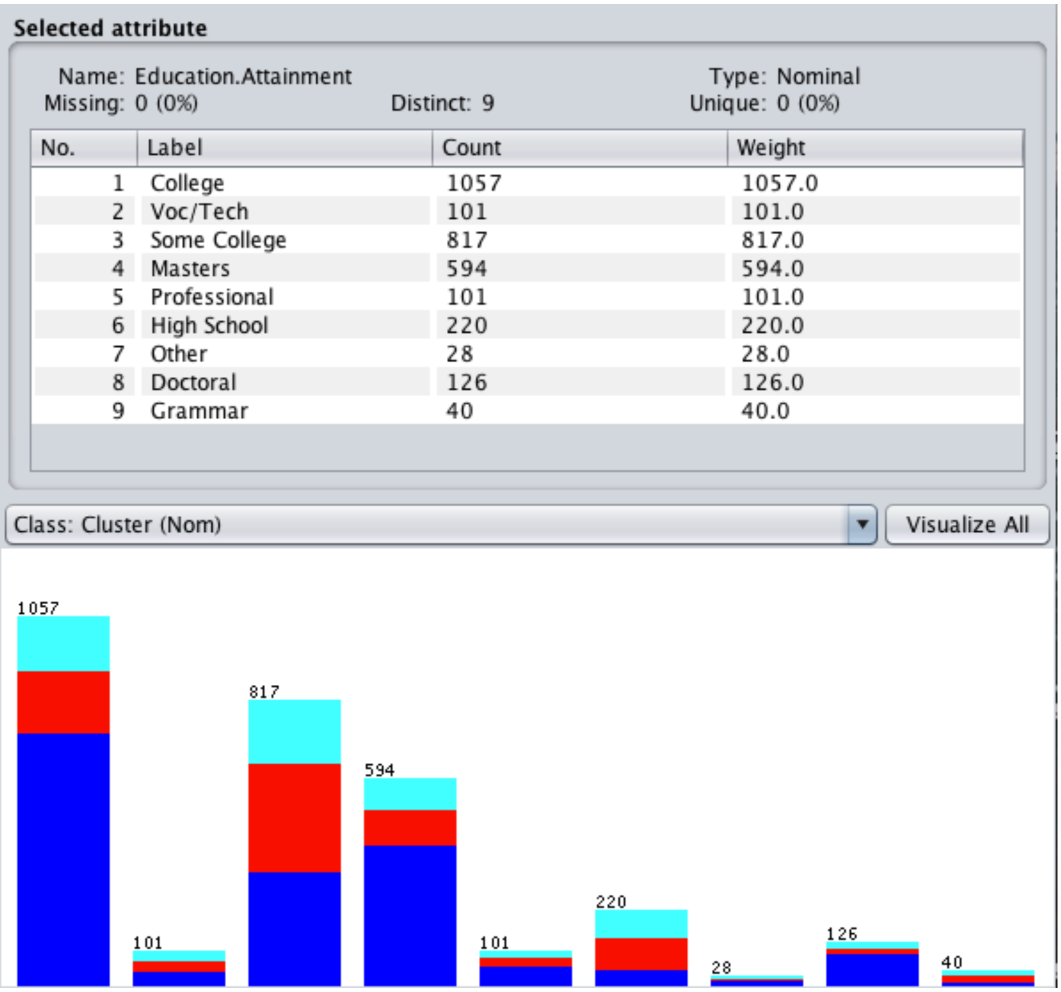
Name: Gender   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	Male	2096	2096.0
2	Female	988	988.0

Class: Cluster (Nom)

Cluster Blue – Mostly Male, Cluster Red – Mostly Male, Cluster Teal – Slightly More Female

D) Education



Selected attribute

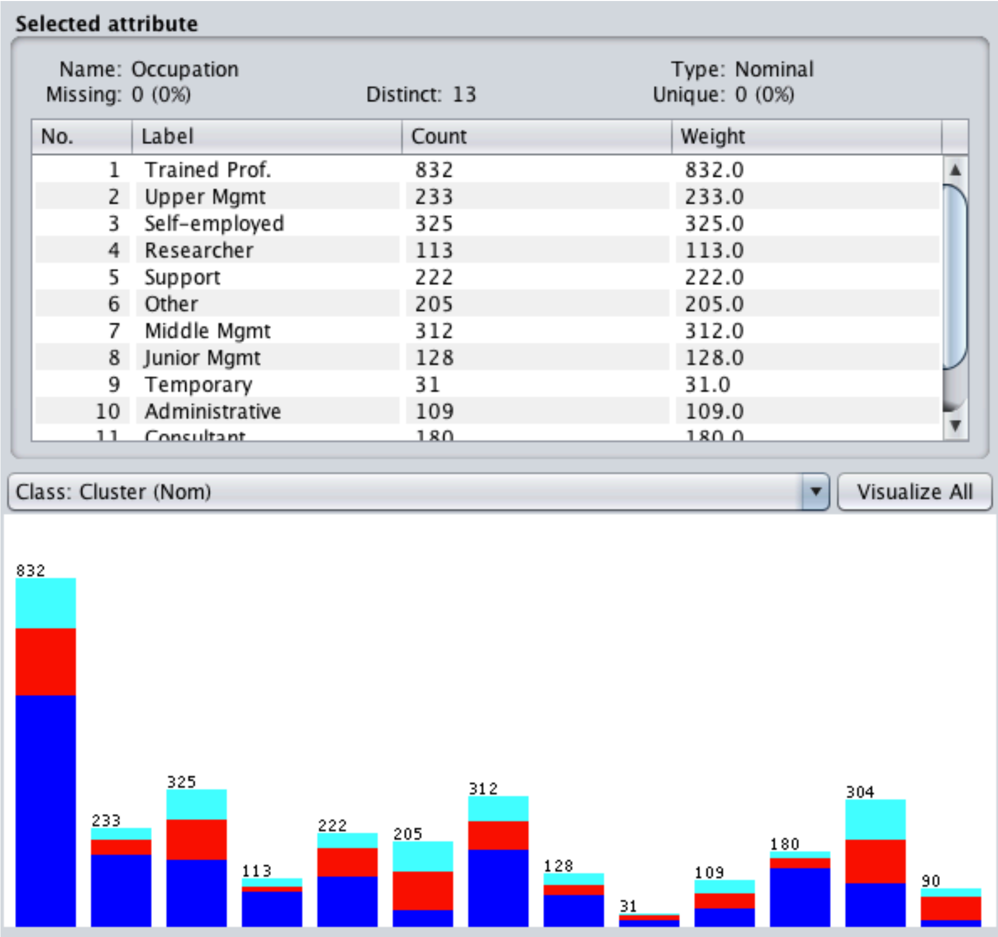
Name: Education.Attainment Type: Nominal
Missing: 0 (0%) Distinct: 9 Unique: 0 (0%)

No.	Label	Count	Weight
1	College	1057	1057.0
2	Voc/Tech	101	101.0
3	Some College	817	817.0
4	Masters	594	594.0
5	Professional	101	101.0
6	High School	220	220.0
7	Other	28	28.0
8	Doctoral	126	126.0
9	Grammar	40	40.0

Class: Cluster (Nom)

Cluster Blue – Some College – Masters, Cluster Red – Some College – College , Cluster Teal - Some College – College

E) Occupation



Selected attribute

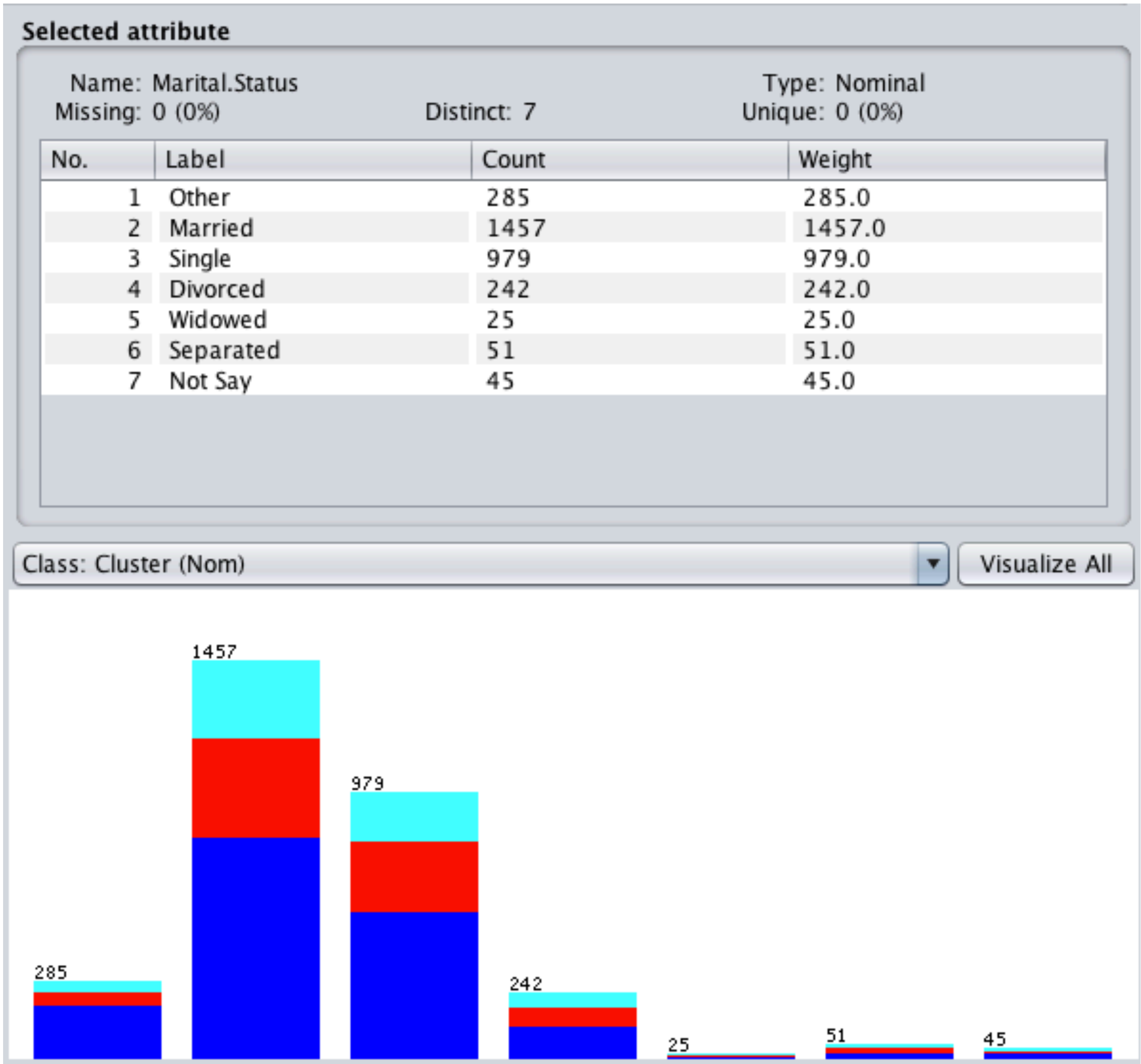
Name: Occupation
Missing: 0 (0%)
Distinct: 13
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Trained Prof.	832	832.0
2	Upper Mgmt	233	233.0
3	Self-employed	325	325.0
4	Researcher	113	113.0
5	Support	222	222.0
6	Other	205	205.0
7	Middle Mgmt	312	312.0
8	Junior Mgmt	128	128.0
9	Temporary	31	31.0
10	Administrative	109	109.0
11	Consultant	180	180.0

Class: Cluster (Nom)

Cluster Blue – Trained Professional, Cluster Red – Trained Professional, Cluster Teal – Not Clear

F) Martial Status



Selected attribute

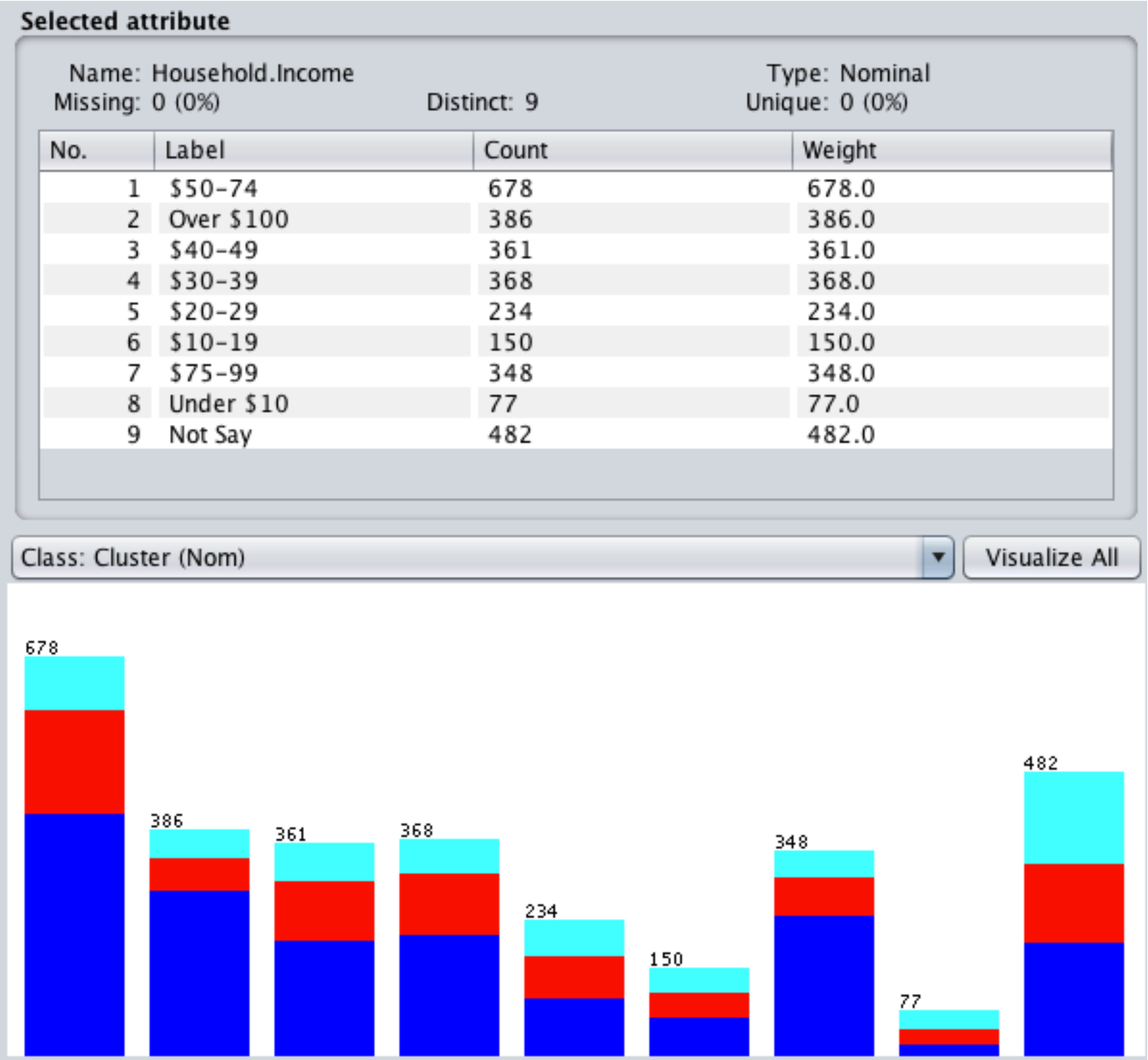
Name: Marital.Status
Missing: 0 (0%)
Distinct: 7
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Other	285	285.0
2	Married	1457	1457.0
3	Single	979	979.0
4	Divorced	242	242.0
5	Widowed	25	25.0
6	Separated	51	51.0
7	Not Say	45	45.0

Class: Cluster (Nom)

Cluster Blue – Married or Single, Cluster Red - Married or Single, Cluster Teal - Married or Single

G) Household Income



Selected attribute

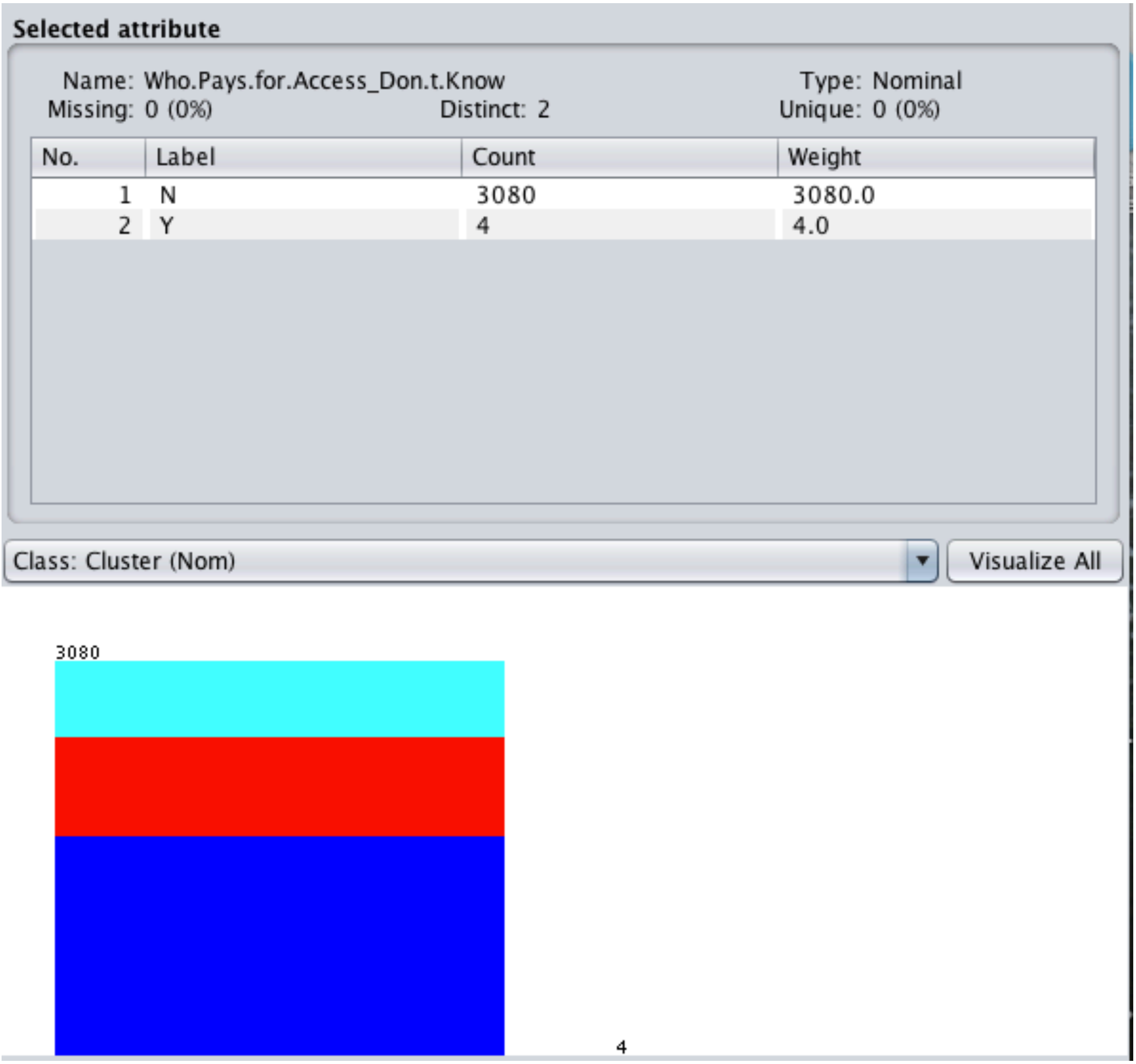
Name: Household.Income | Type: Nominal
Missing: 0 (0%) | Distinct: 9 | Unique: 0 (0%)

No.	Label	Count	Weight
1	$50-74	678	678.0
2	Over $100	386	386.0
3	$40-49	361	361.0
4	$30-39	368	368.0
5	$20-29	234	234.0
6	$10-19	150	150.0
7	$75-99	348	348.0
8	Under $10	77	77.0
9	Not Say	482	482.0

Class: Cluster (Nom)

Cluster Blue - $50-74 and Over $100, Cluster Red - $50-74 , Cluster Teal – Not Clear

H) Who Pays For Access



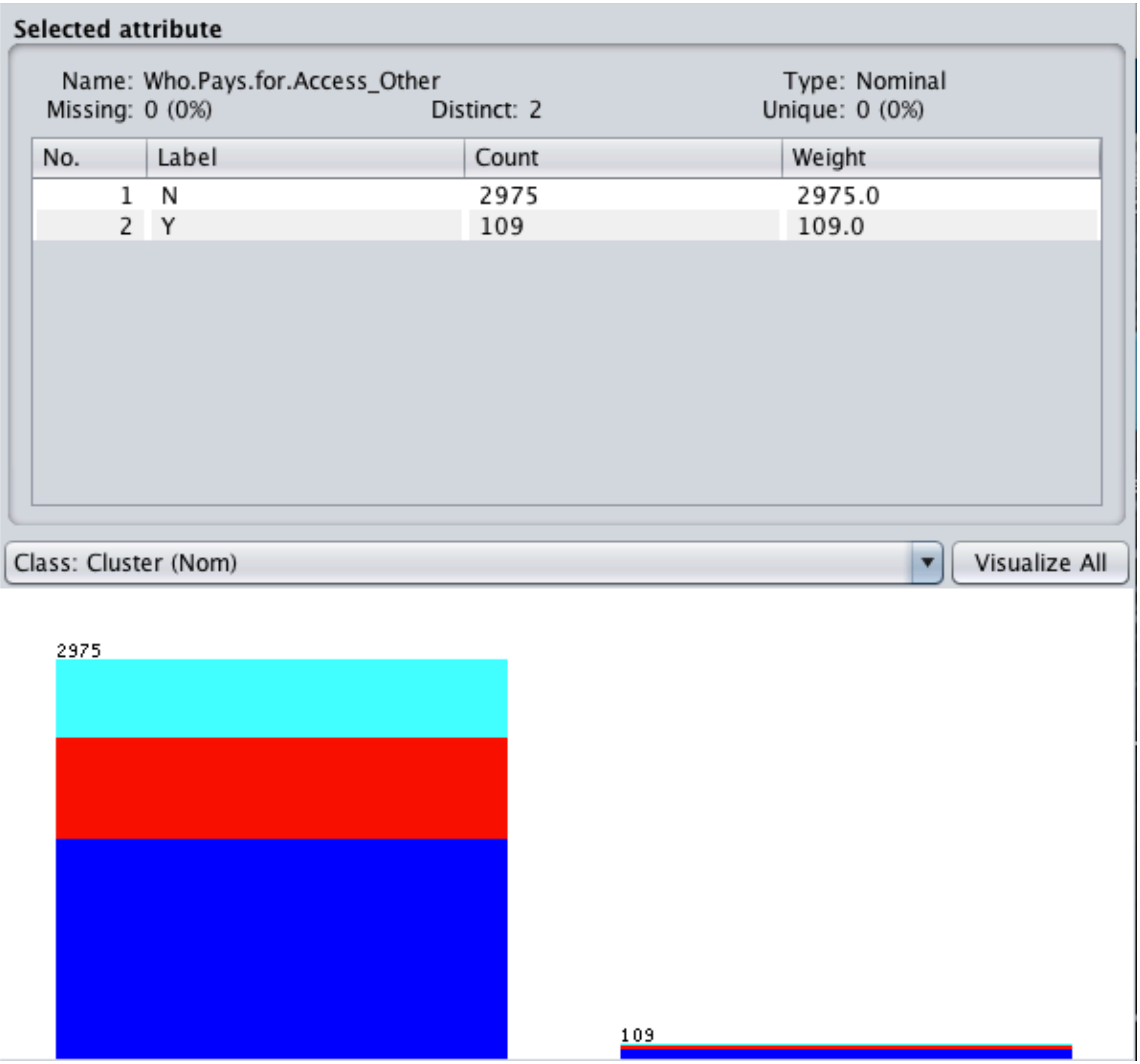
Selected attribute

Name: Who.Pays.for.Access_Don.t.Know   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	N	3080	3080.0
2	Y	4	4.0

Class: Cluster (Nom)

Not a reason for any Cluster



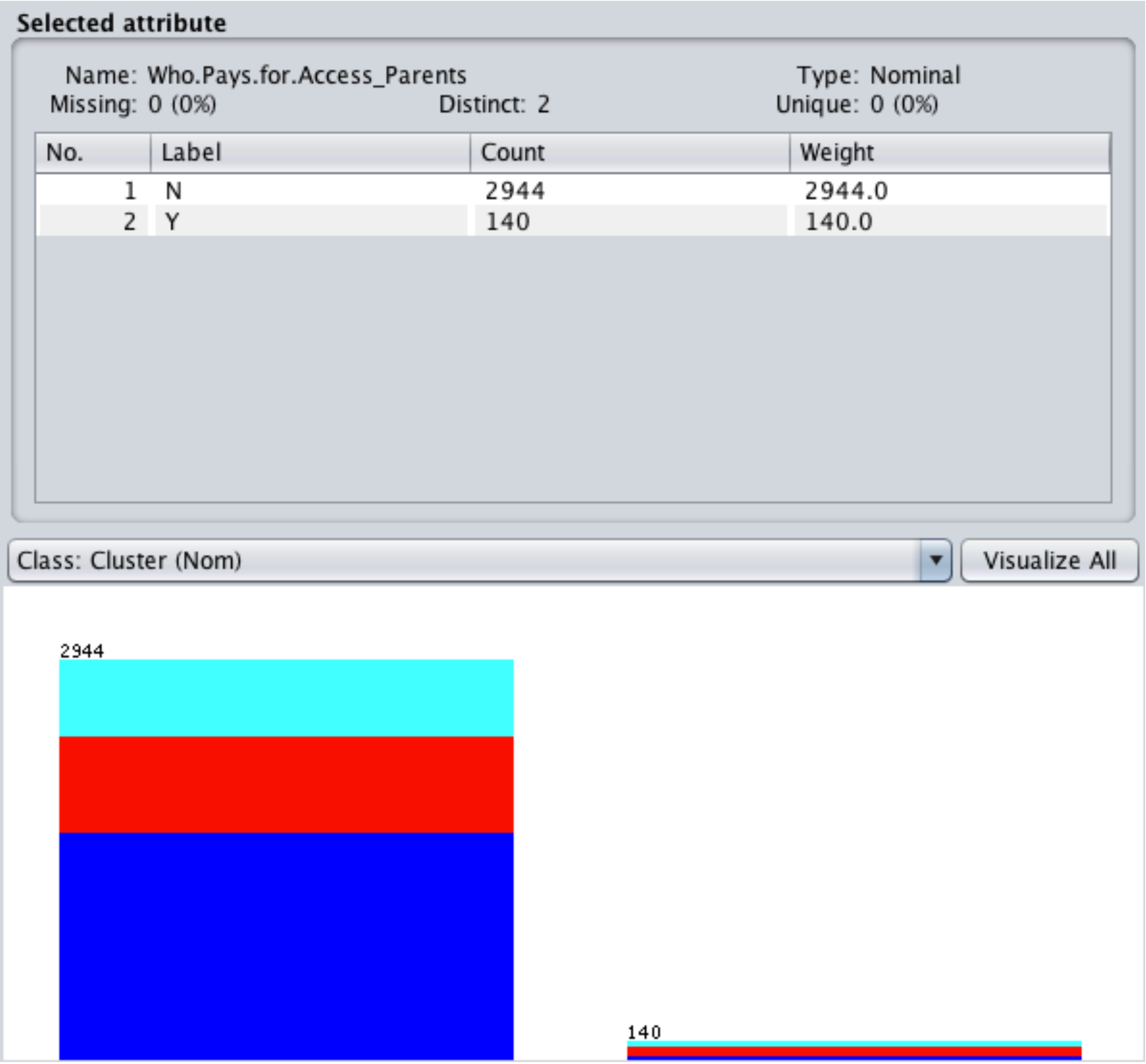
Selected attribute

Name: Who.Pays.for.Access_Other
Type: Nominal
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)

No.	Label	Count	Weight
1	N	2975	2975.0
2	Y	109	109.0

Class: Cluster (Nom)

Not a Reason For Any Cluster



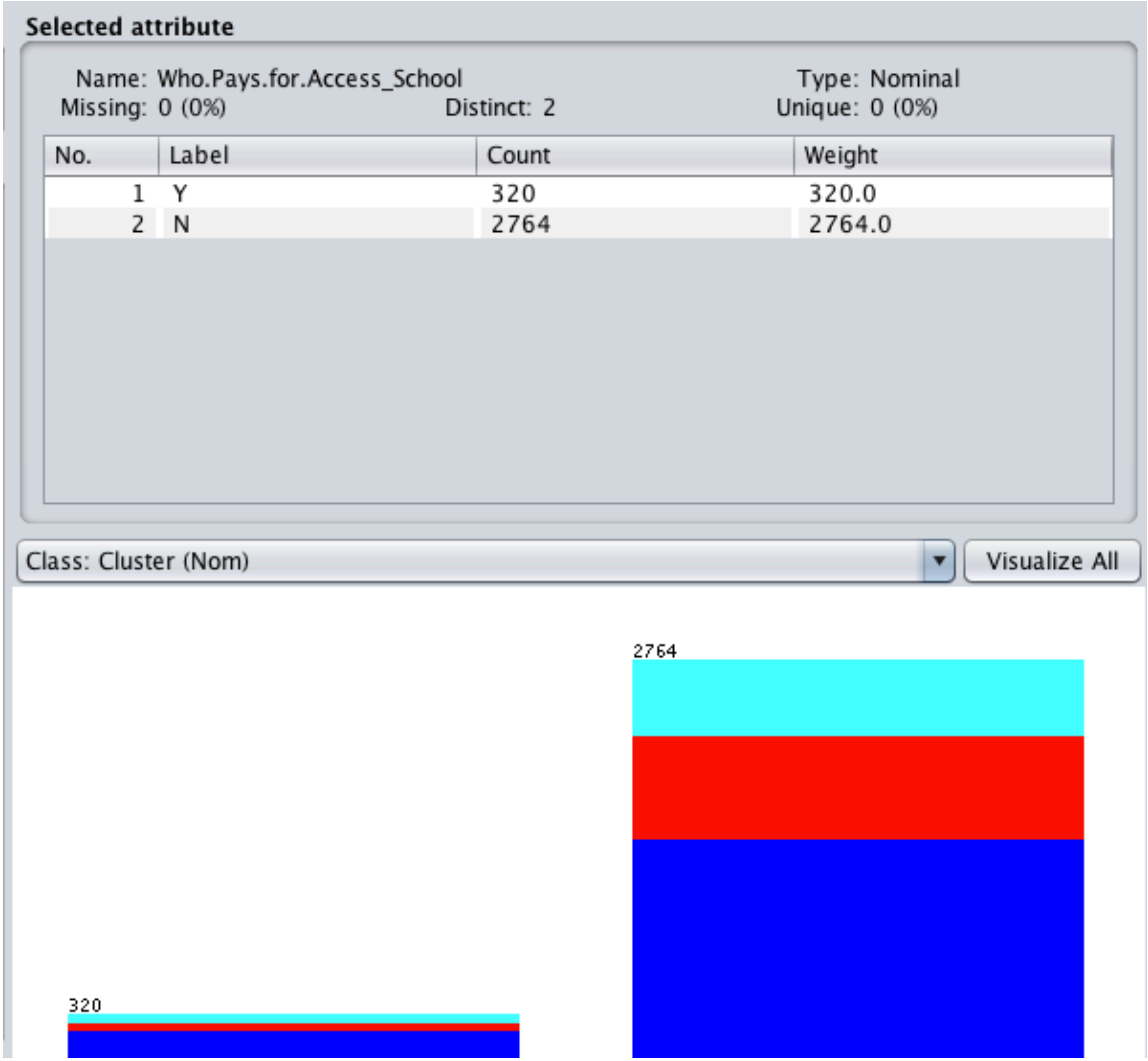
Selected attribute

Name: Who.Pays.for.Access_Parents
Type: Nominal
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)

No.	Label	Count	Weight
1	N	2944	2944.0
2	Y	140	140.0

Class: Cluster (Nom)

Not a Reason for Any Cluster



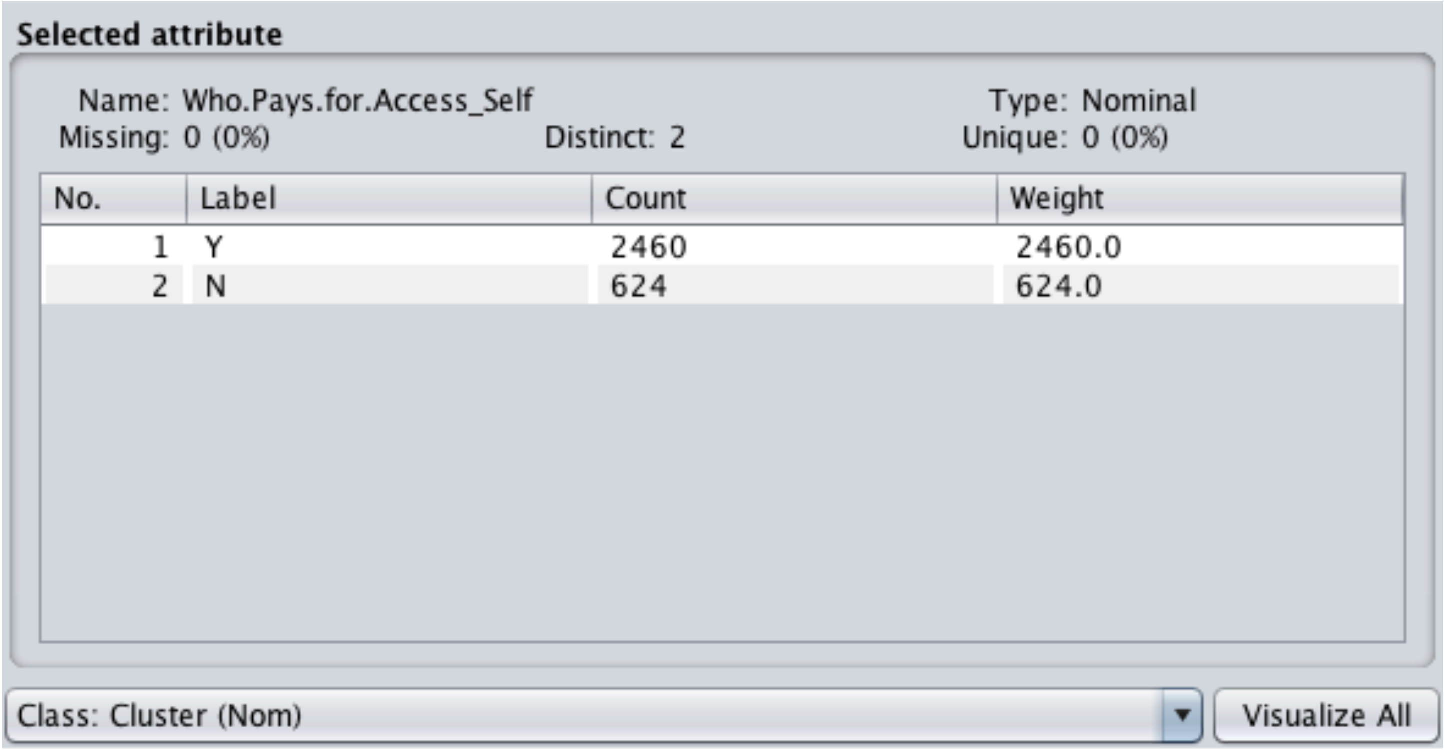
Selected attribute

Name: Who.Pays.for.Access_School   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	Y	320	320.0
2	N	2764	2764.0

Class: Cluster (Nom)

Not a Reason For Any Cluster

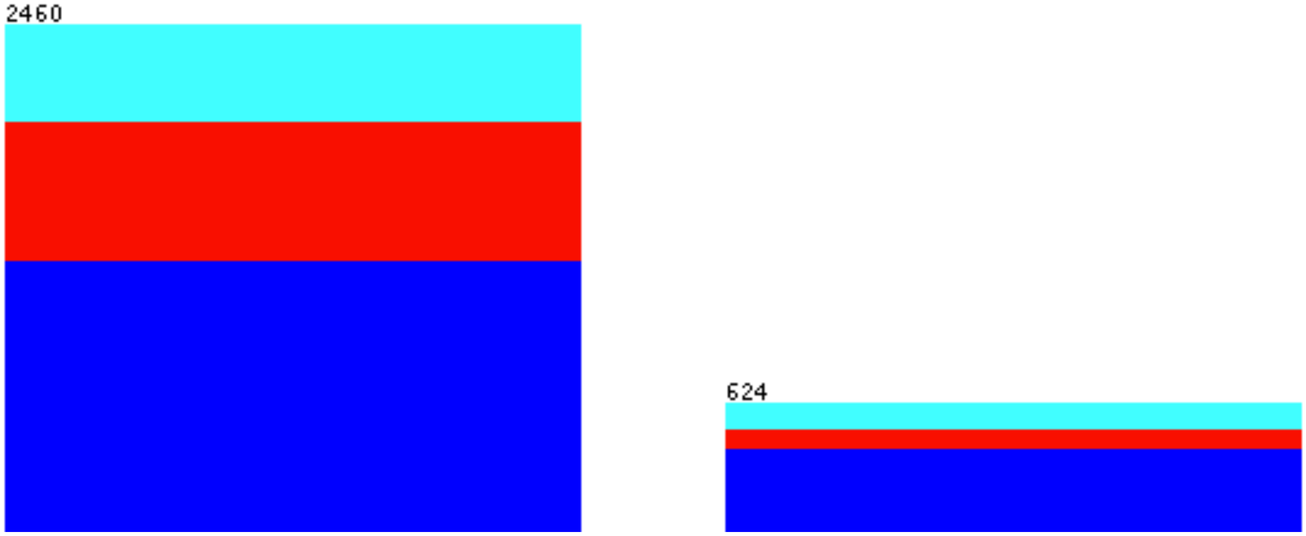


Selected attribute

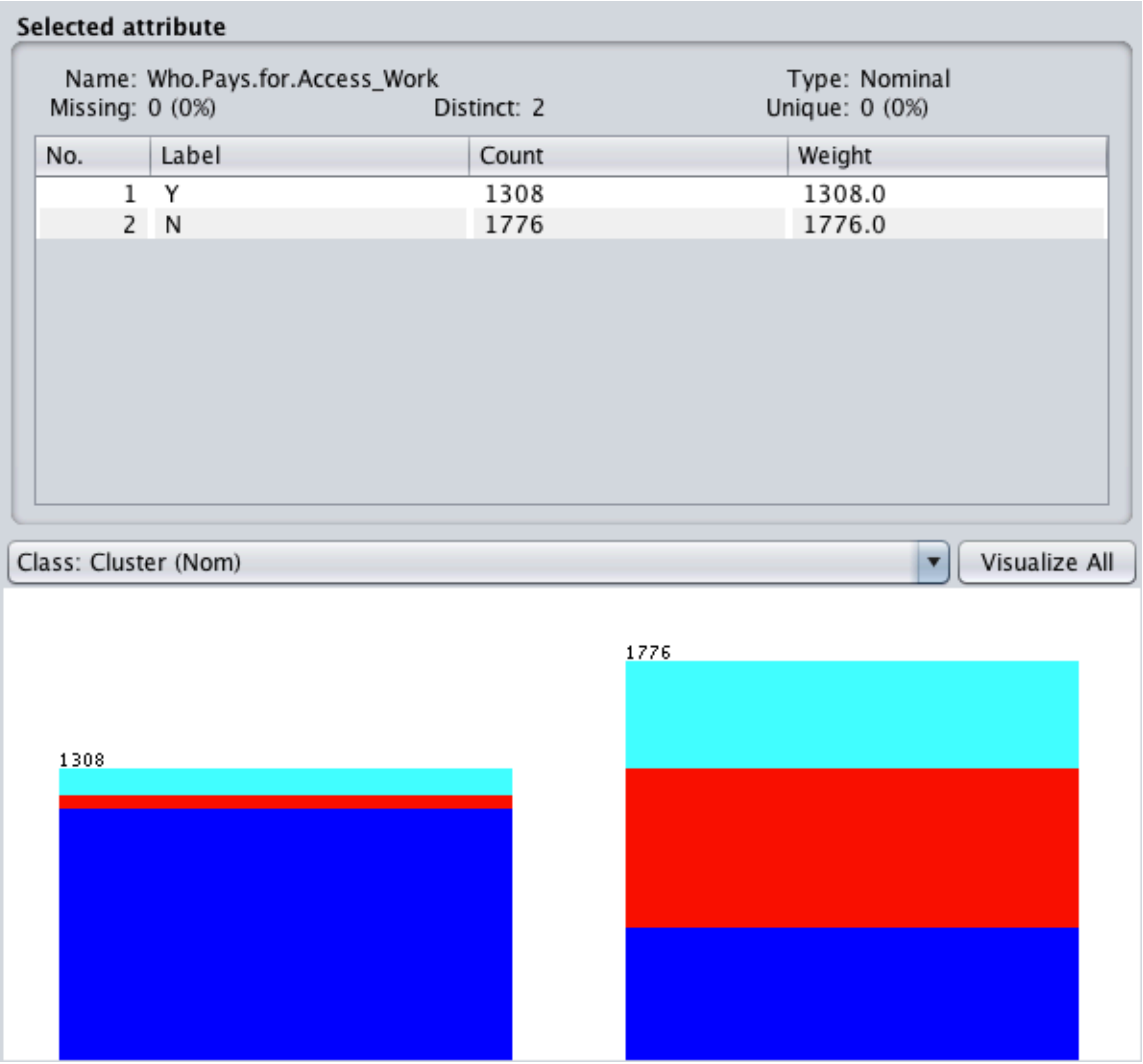
Name: Who.Pays.for.Access_Self   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	Y	2460	2460.0
2	N	624	624.0

Class: Cluster (Nom)



All Clusters say Self



Selected attribute

Name: Who.Pays.for.Access_Work Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

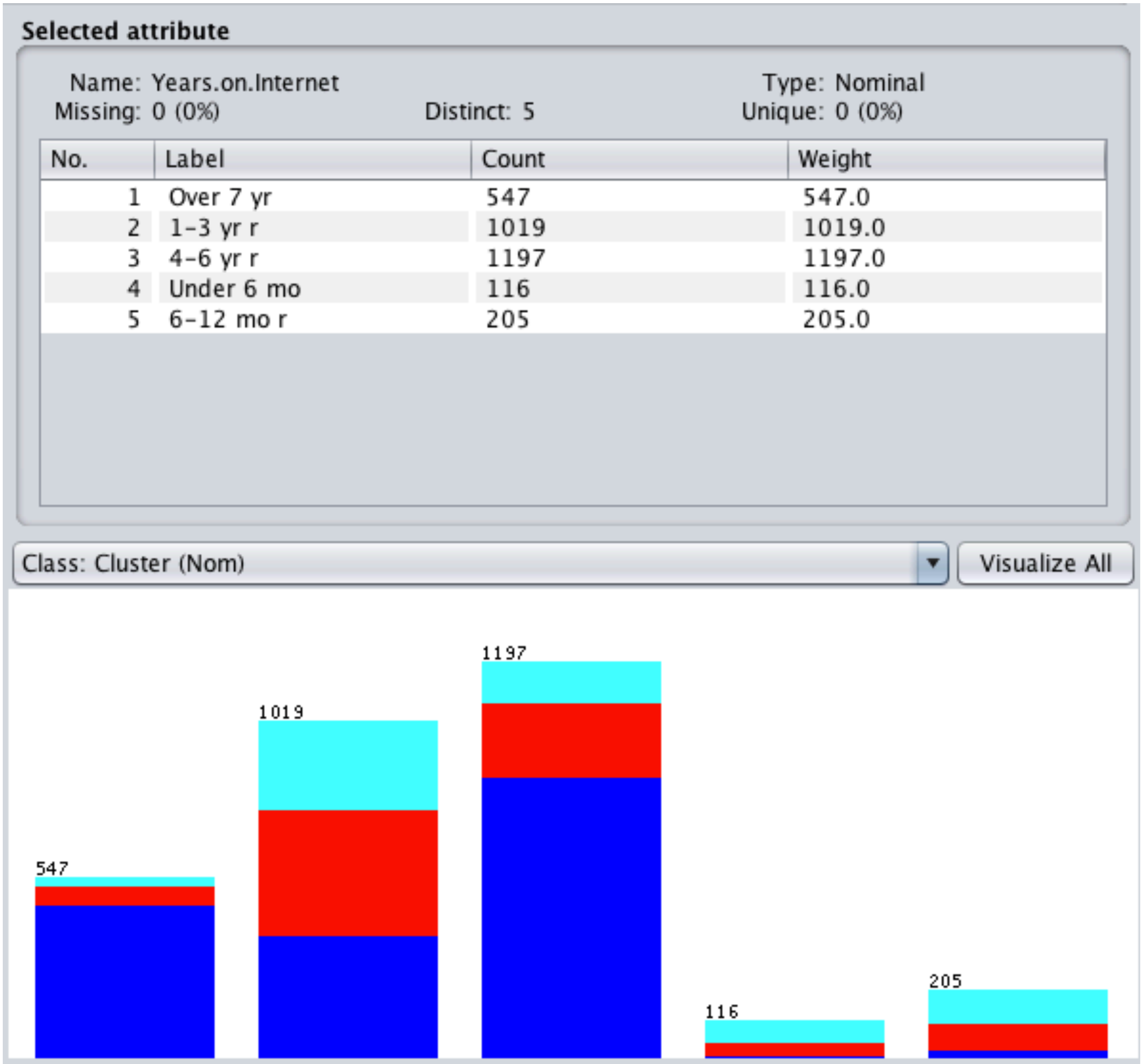
No.	Label	Count	Weight
1	Y	1308	1308.0
2	N	1776	1776.0

Class: Cluster (Nom)

Cluster Blue – Yes Work, Cluster Red – No Work Doesn’t Pay, Cluster Teal - No Work Doesn’t Pay

3) User Prior Experience

- subgroup –
A) Years on Internet



Selected attribute

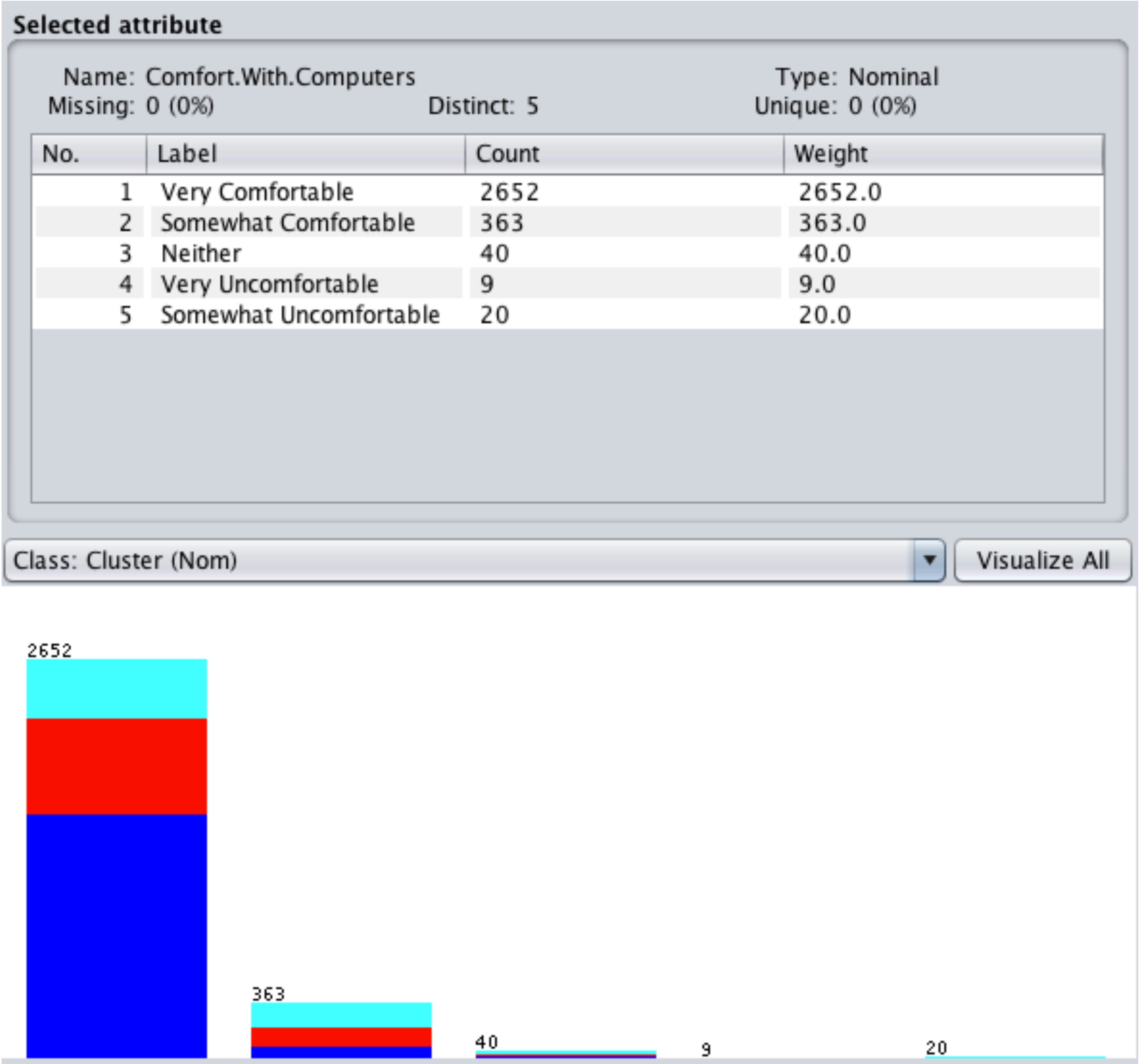
Name: Years.on.Internet Type: Nominal
Missing: 0 (0%) Distinct: 5 Unique: 0 (0%)

No.	Label	Count	Weight
1	Over 7 yr	547	547.0
2	1-3 yr r	1019	1019.0
3	4-6 yr r	1197	1197.0
4	Under 6 mo	116	116.0
5	6-12 mo r	205	205.0

Class: Cluster (Nom)

Cluster Blue - High to Moderate Time, Cluster Red – Moderate, Cluster Teal - Moderate

B)Comfort With Computers



Selected attribute

Name: Comfort.With.Computers	Type: Nominal
Missing: 0 (0%)	Distinct: 5	Unique: 0 (0%)

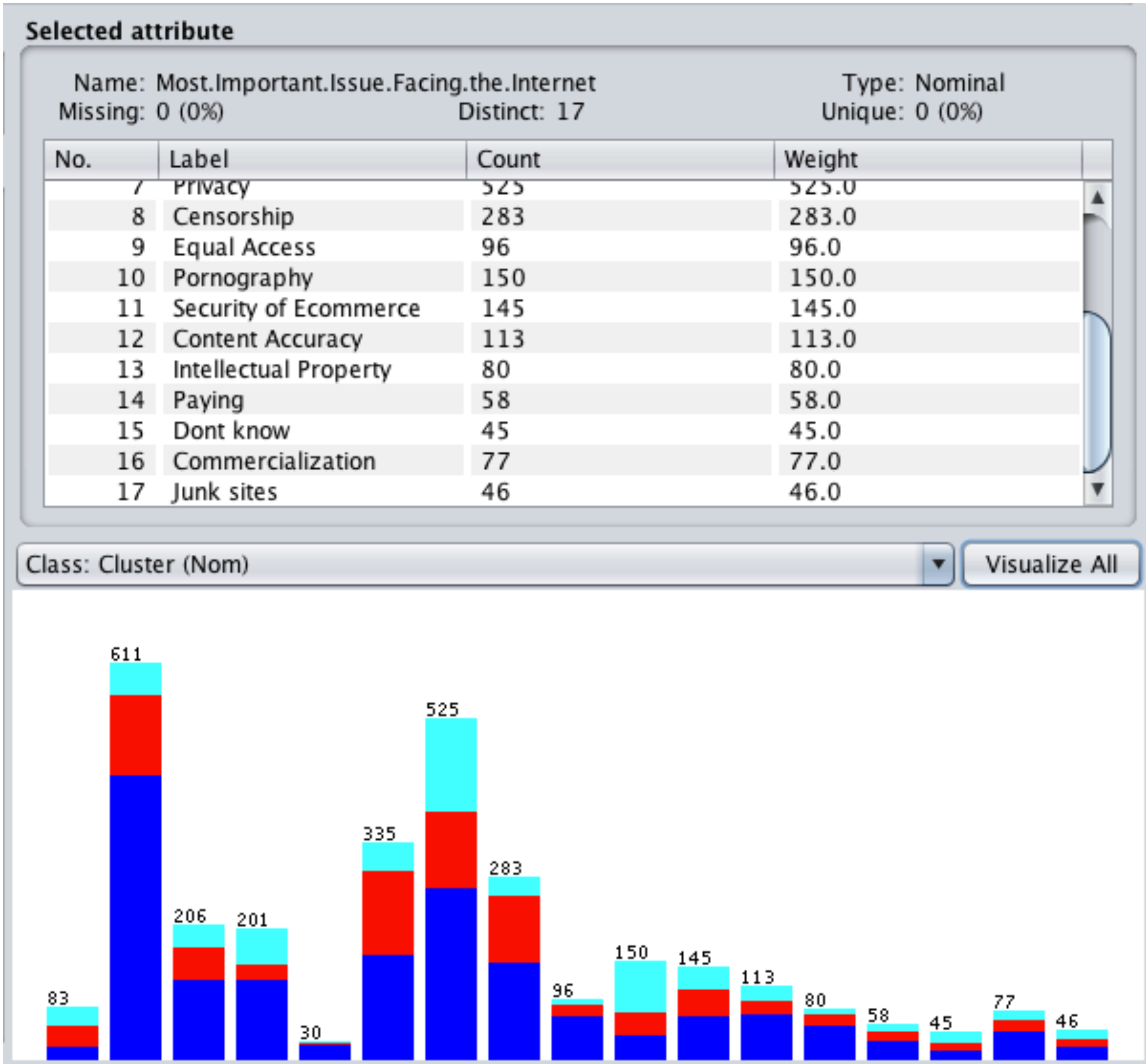
No.	Label	Count	Weight
1	Very Comfortable	2652	2652.0
2	Somewhat Comfortable	363	363.0
3	Neither	40	40.0
4	Very Uncomfortable	9	9.0
5	Somewhat Uncomfortable	20	20.0

Class: Cluster (Nom)

Cluster Blue – Very Comfortable, Cluster Red - Very Comfortable, Cluster Teal - Very Comfortable

4) User Opinions

A) Most Important Issues Facing The Internet



Selected attribute

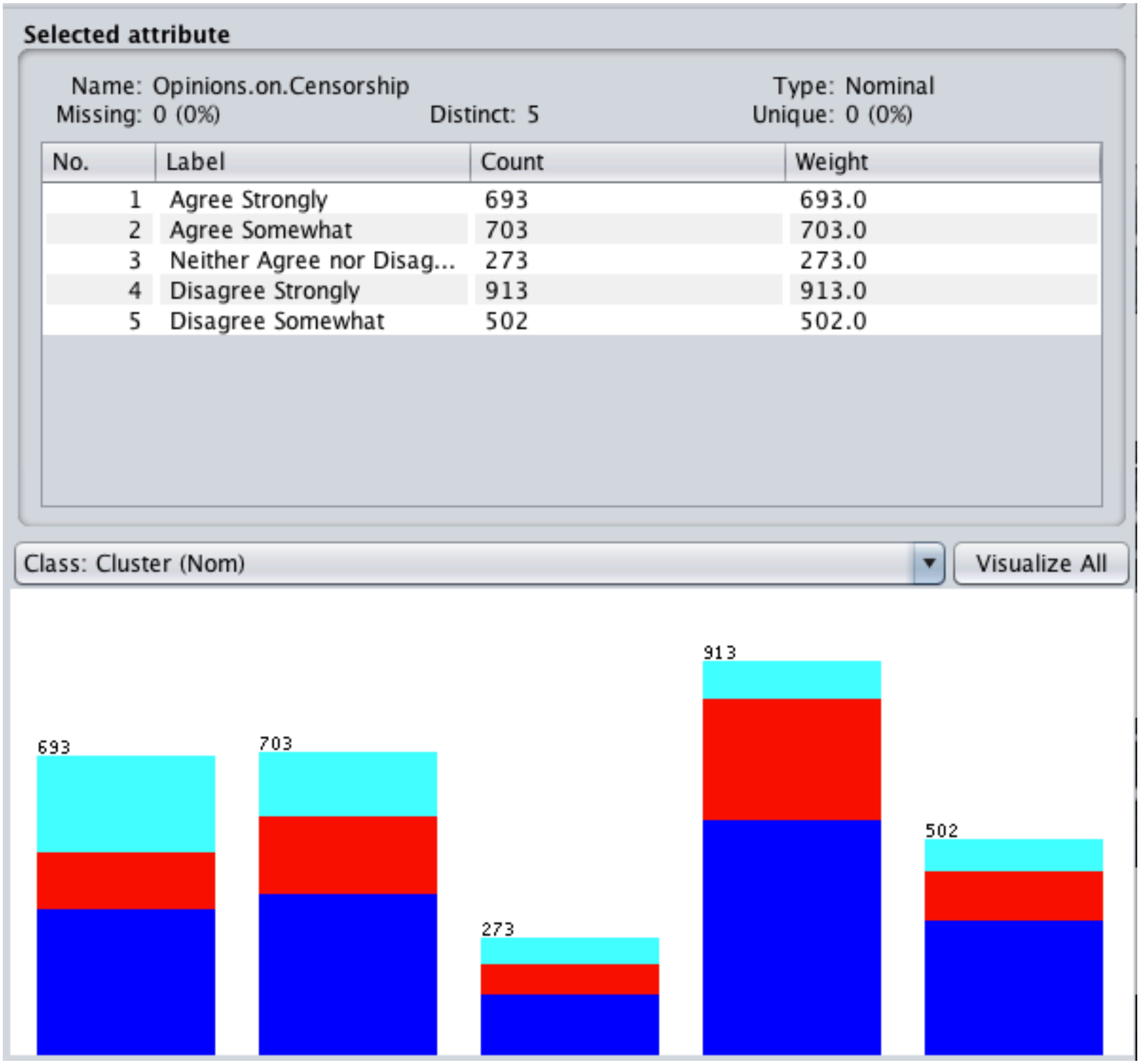
Name: Most.Important.Issue.Facing.the.Internet Type: Nominal
Missing: 0 (0%) Distinct: 17 Unique: 0 (0%)

No.	Label	Count	Weight
7	Privacy	525	525.0
8	Censorship	283	283.0
9	Equal Access	96	96.0
10	Pornography	150	150.0
11	Security of Ecommerce	145	145.0
12	Content Accuracy	113	113.0
13	Intellectual Property	80	80.0
14	Paying	58	58.0
15	Dont know	45	45.0
16	Commercialization	77	77.0
17	Junk sites	46	46.0

Class: Cluster (Nom)

Cluster Blue – Speed and Government Regulation, Cluster Red – Speed, Government Regulation, Privacy, Censorship, Cluster Teal – Privacy and Pornography

B) Opinions on Censorship



Selected attribute

Name: Opinions.on.Censorship
Type: Nominal
Missing: 0 (0%)
Distinct: 5
Unique: 0 (0%)

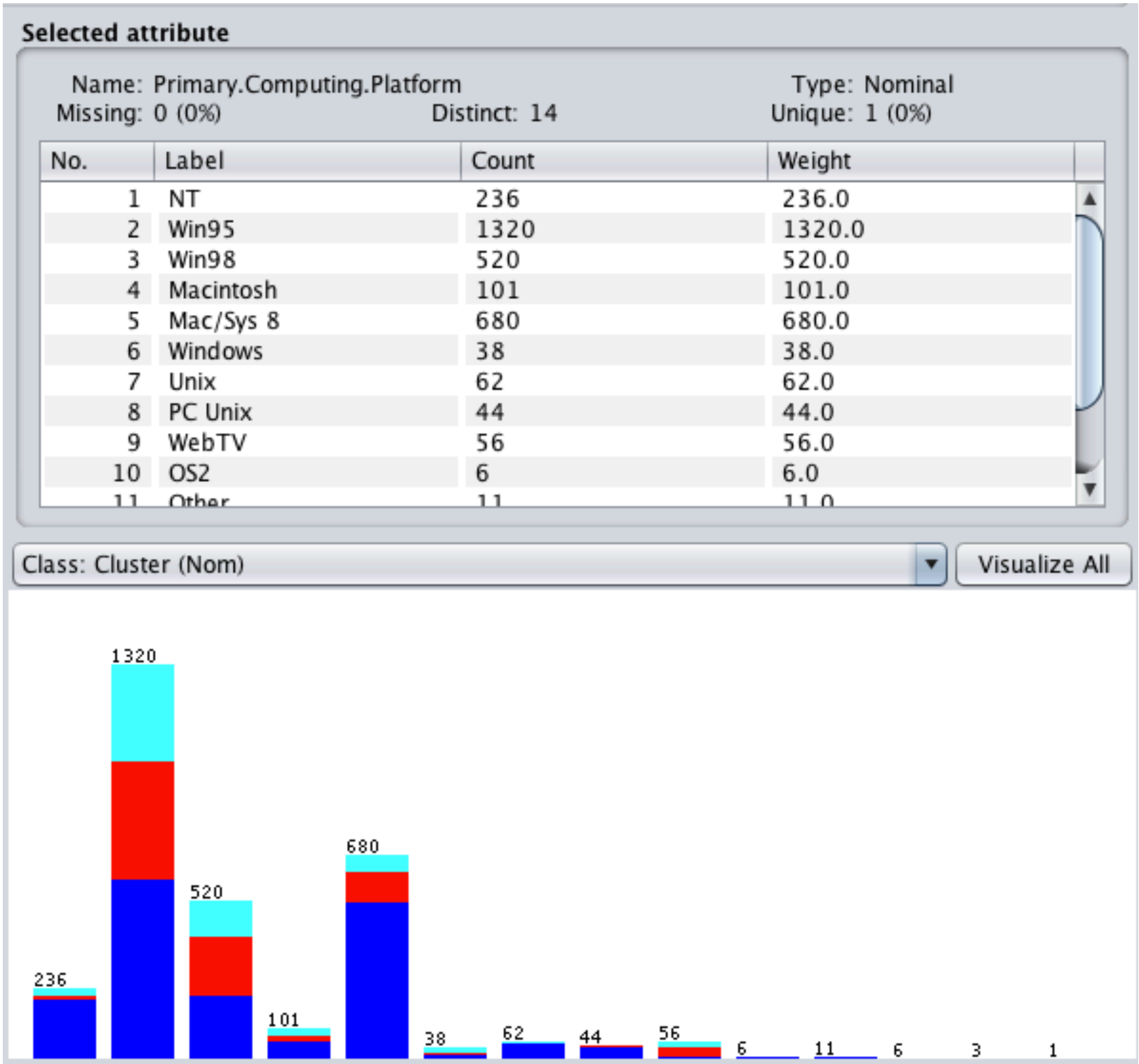
No.	Label	Count	Weight
1	Agree Strongly	693	693.0
2	Agree Somewhat	703	703.0
3	Neither Agree nor Disag...	273	273.0
4	Disagree Strongly	913	913.0
5	Disagree Somewhat	502	502.0

Class: Cluster (Nom)

Cluster Blue – Disagree, Cluster Red – Disagree, Cluster Teal – Agree

5) User Web Settings/Applications

A) Primary Computing Platform



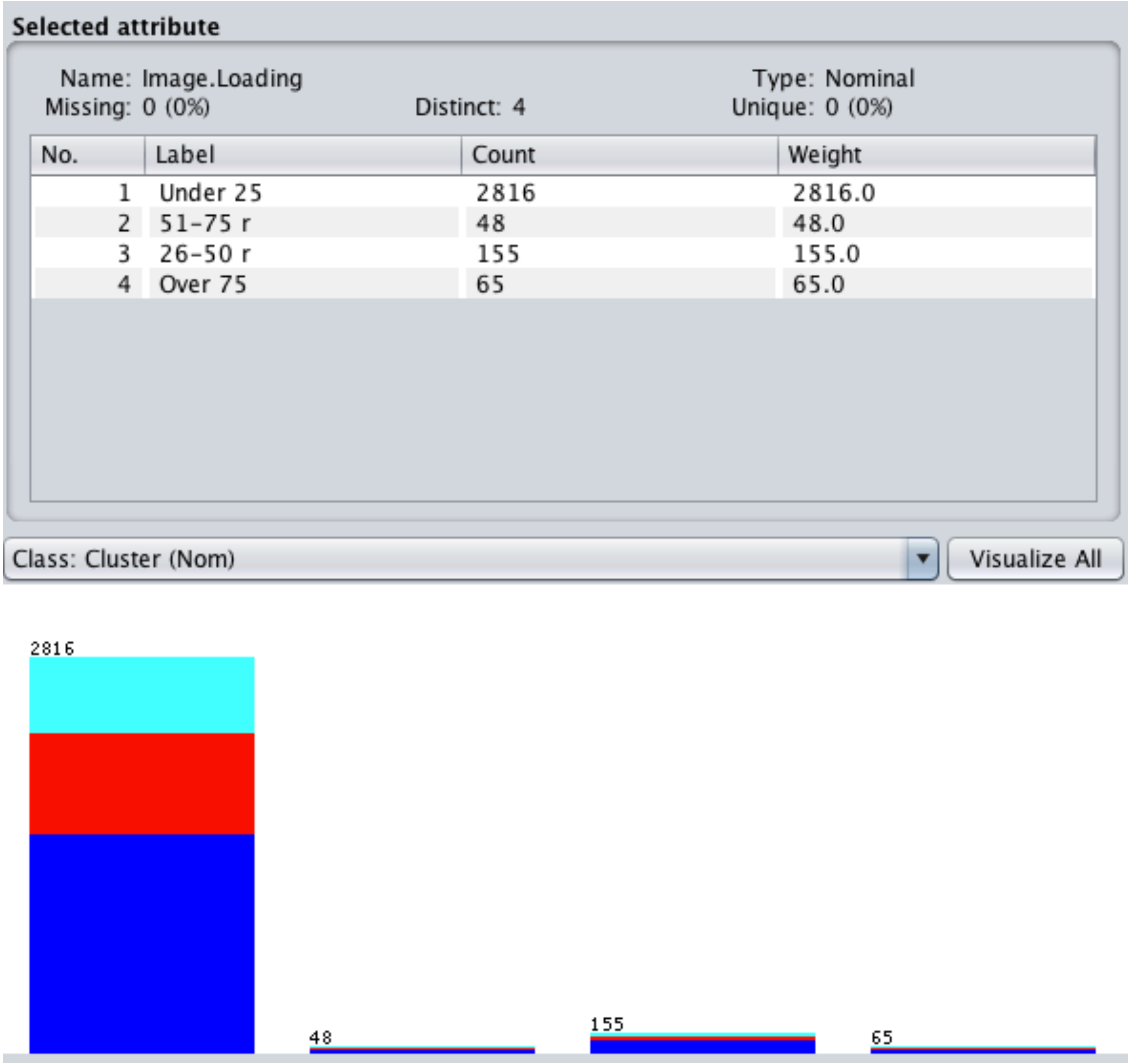
Selected attribute

Name: Primary.Computing.Platform Type: Nominal
Missing: 0 (0%) Distinct: 14 Unique: 1 (0%)

No.	Label	Count	Weight
1	NT	236	236.0
2	Win95	1320	1320.0
3	Win98	520	520.0
4	Macintosh	101	101.0
5	Mac/Sys 8	680	680.0
6	Windows	38	38.0
7	Unix	62	62.0
8	PC Unix	44	44.0
9	WebTV	56	56.0
10	OS2	6	6.0
11	Other	11	11.0

Cluster Blue – Win95 and Mac/Sys8, Cluster Red – Win95 and Win98, Cluster Teal – Win 95 and Win98

B) Image Loading



Selected attribute

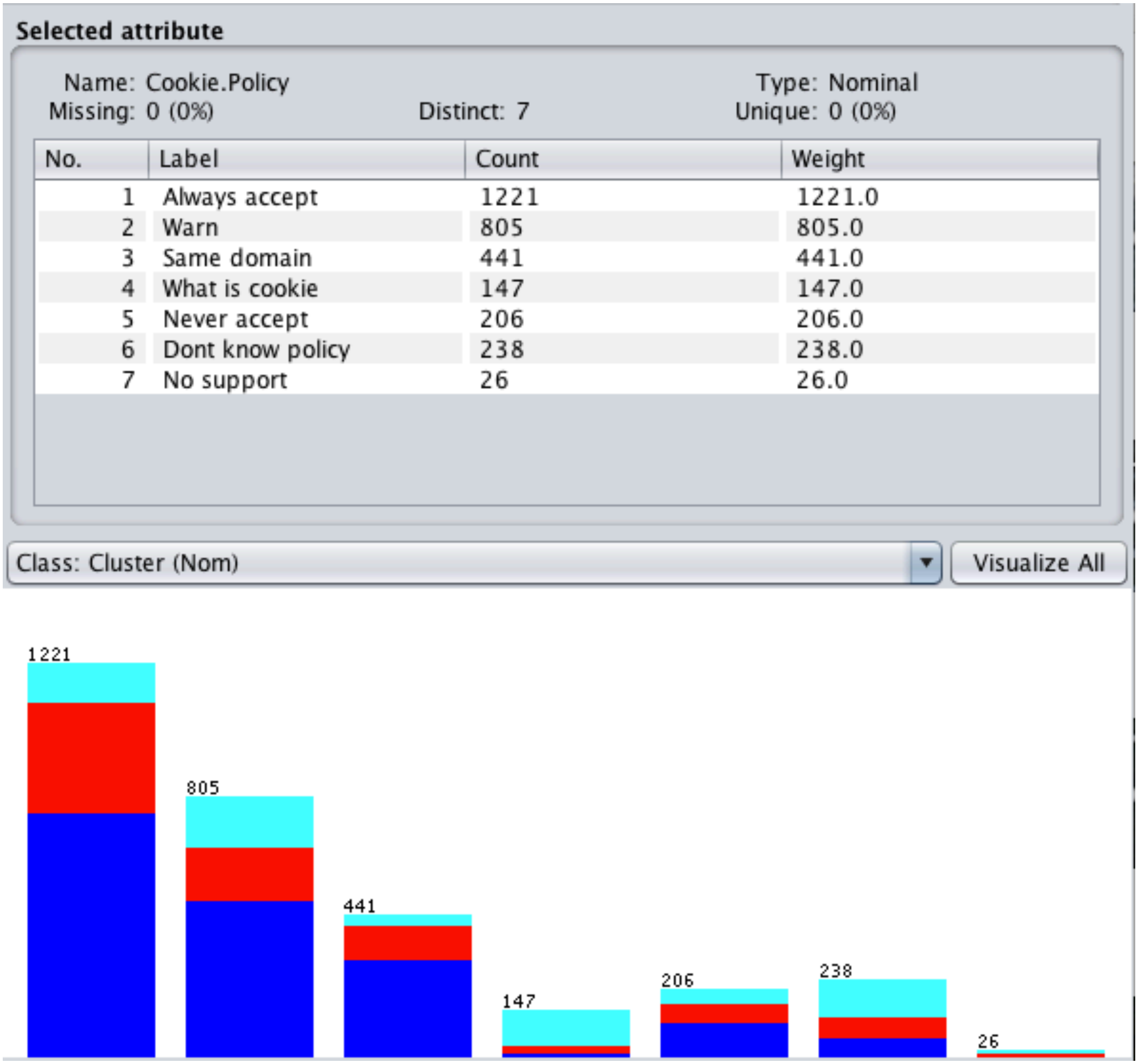
Name: Image.Loading   Type: Nominal
Missing: 0 (0%)   Distinct: 4   Unique: 0 (0%)

No.	Label	Count	Weight
1	Under 25	2816	2816.0
2	51-75 r	48	48.0
3	26-50 r	155	155.0
4	Over 75	65	65.0

Class: Cluster (Nom)

All clusters have low image loading frequency

C) Cookies Policy



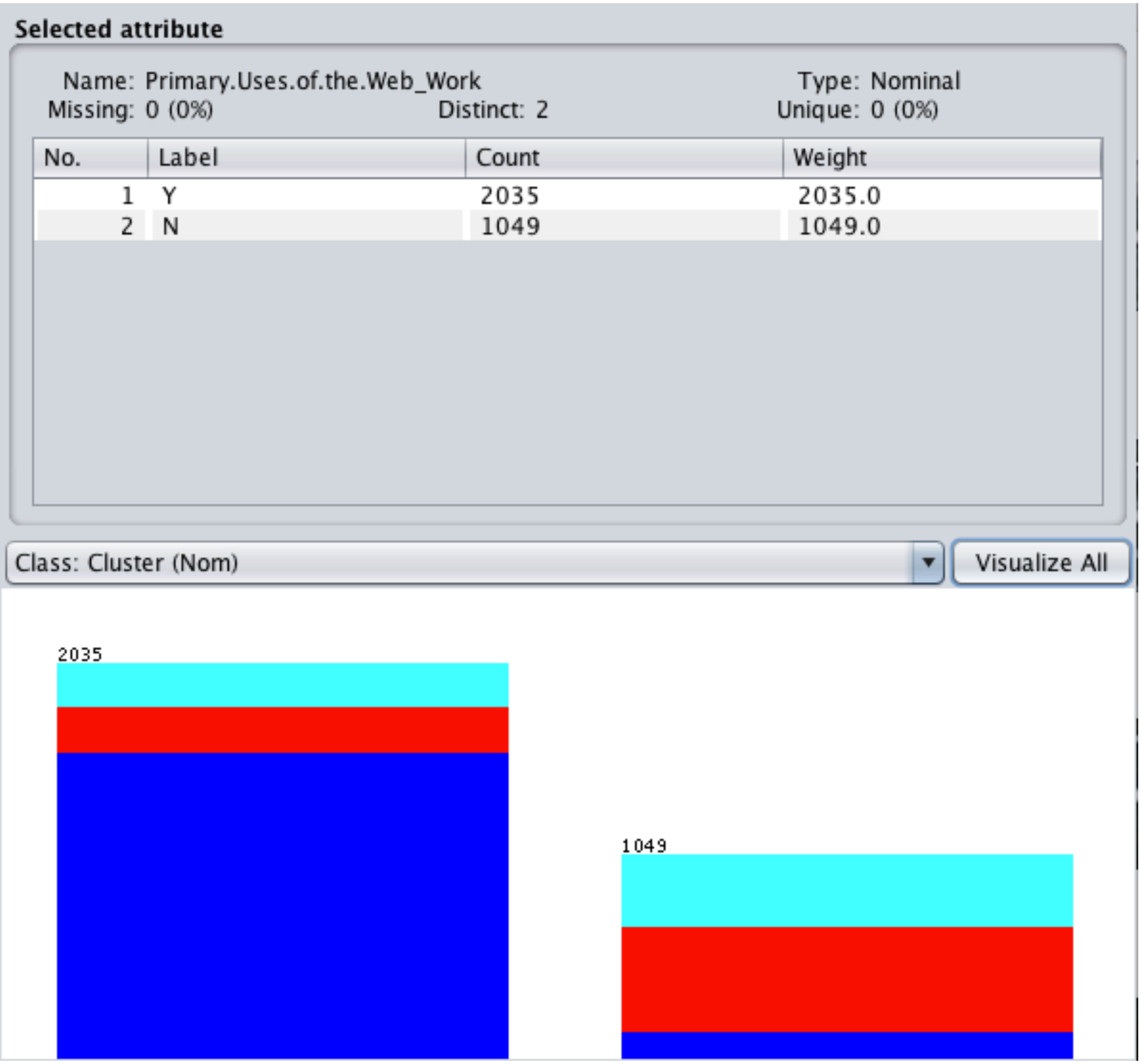
Name: Cookie.Policy
Type: Nominal
Missing: 0 (0%)
Distinct: 7
Unique: 0 (0%)

No.	Label	Count	Weight
1	Always accept	1221	1221.0
2	Warn	805	805.0
3	Same domain	441	441.0
4	What is cookie	147	147.0
5	Never accept	206	206.0
6	Dont know policy	238	238.0
7	No support	26	26.0

Cluster Blue- Mostly Always Accept Cookies, Cluster Red- Mostly Always Accept Cookies, Cluster Teal- No clear trend

6**) Primary Use of The Web**

A) Work



Selected attribute

Name: Primary.Uses.of.the.Web_Work

Type: Nominal

Missing: 0 (0%)

Distinct: 2

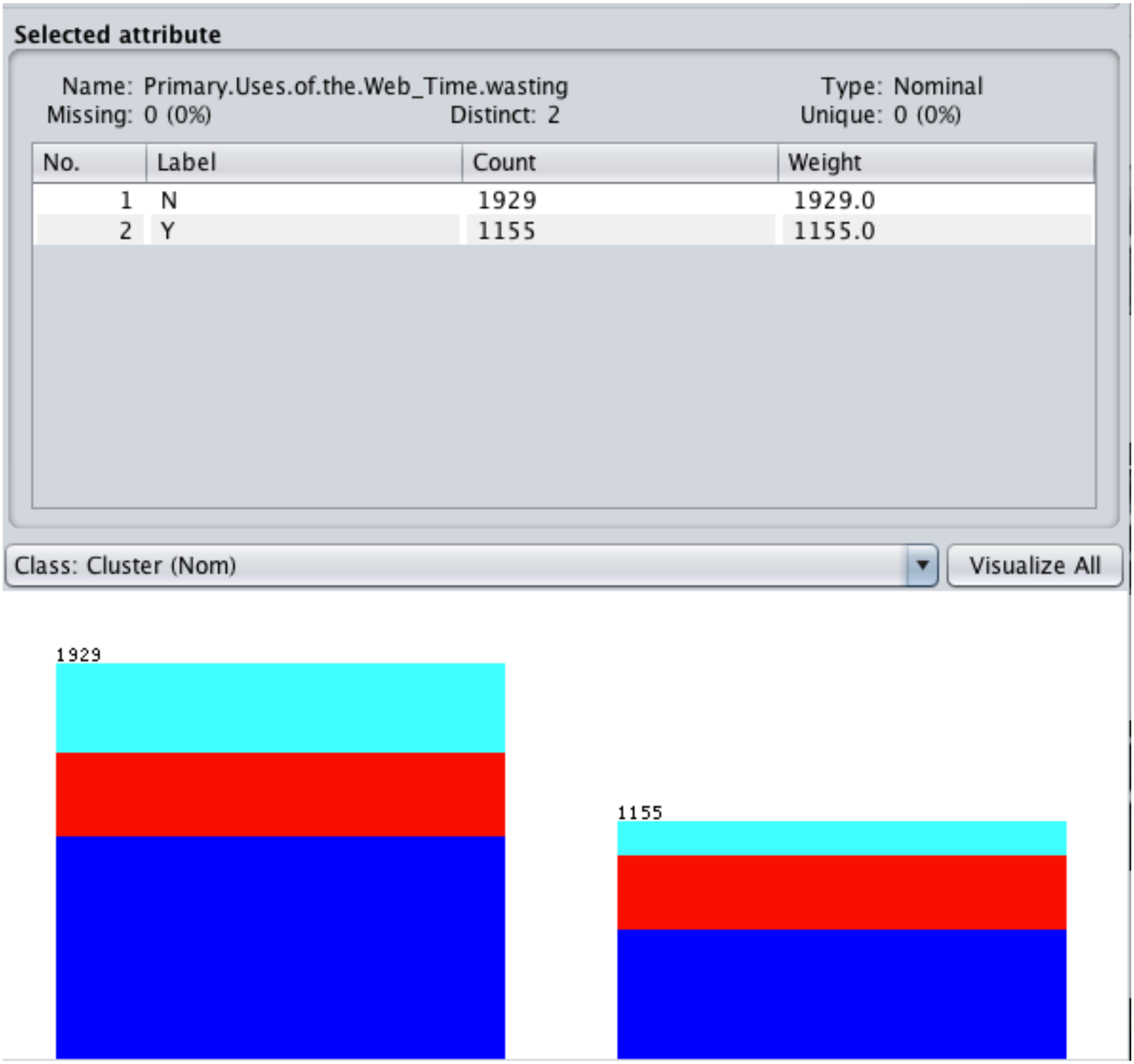
Unique: 0 (0%)

No.	Label	Count	Weight
1	Y	2035	2035.0
2	N	1049	1049.0

Class: Cluster (Nom)

Cluster Blue- Yes. The web is primarily used for work, Cluster Red- No. The web is not primarily used for work, Cluster Teal- No. The web is not primarily used for work.

B) Time Wasting



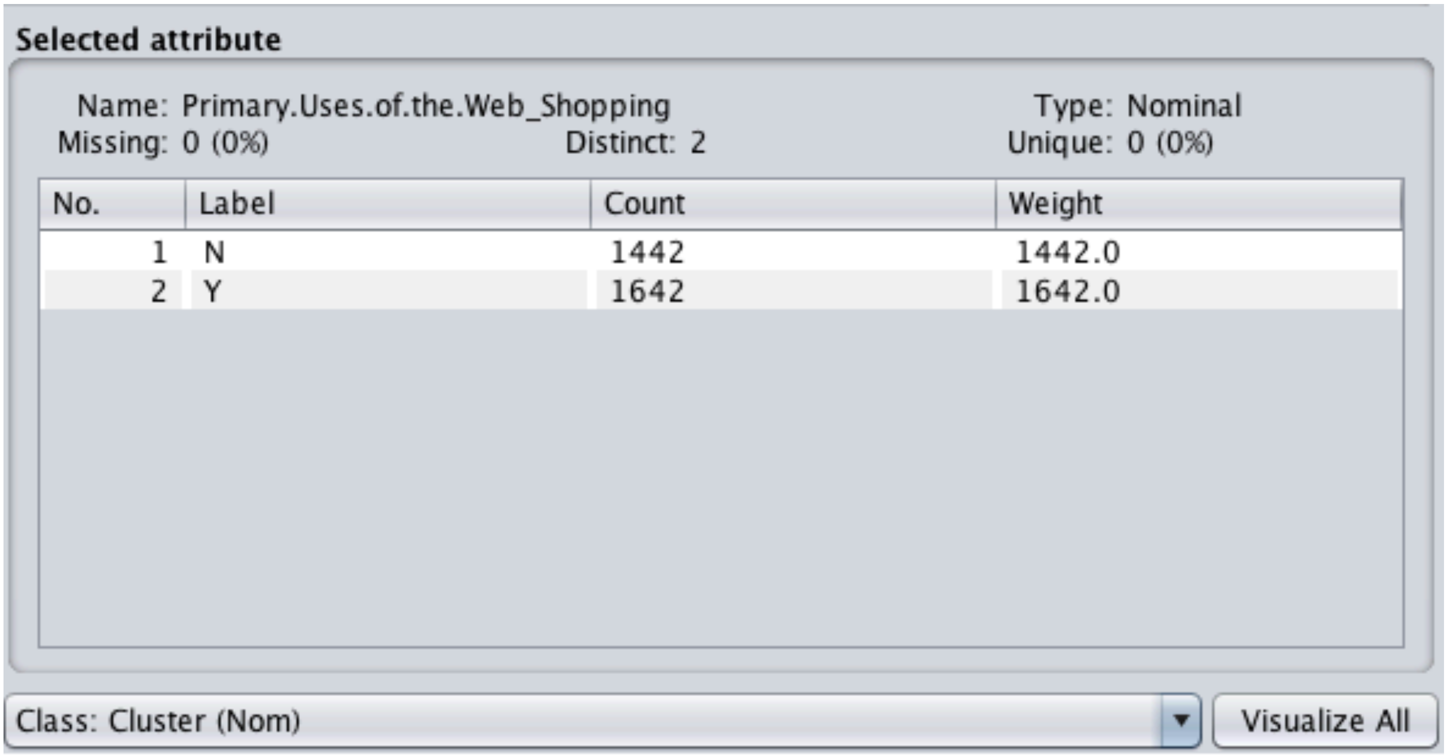
Selected attribute

Name: Primary.Uses.of.the.Web_Time.wasting
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	N	1929	1929.0
2	Y	1155	1155.0

Cluster Blue- Mostly No, the web is not primarily used for to waste time, Cluster Red- No clear trend, Cluster Teal- No, the web is not primarily used for to waste time

C) Shopping



Selected attribute

Name: Primary.Uses.of.the.Web_Shopping

Type: Nominal

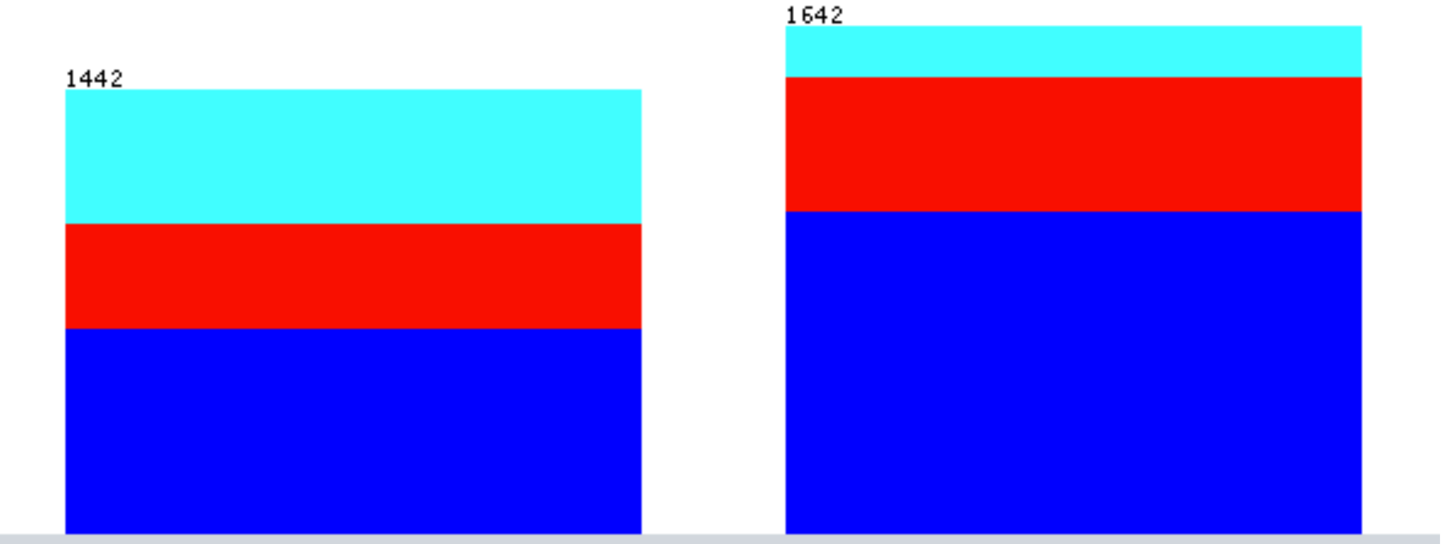
Missing: 0 (0%)

Distinct: 2

Unique: 0 (0%)

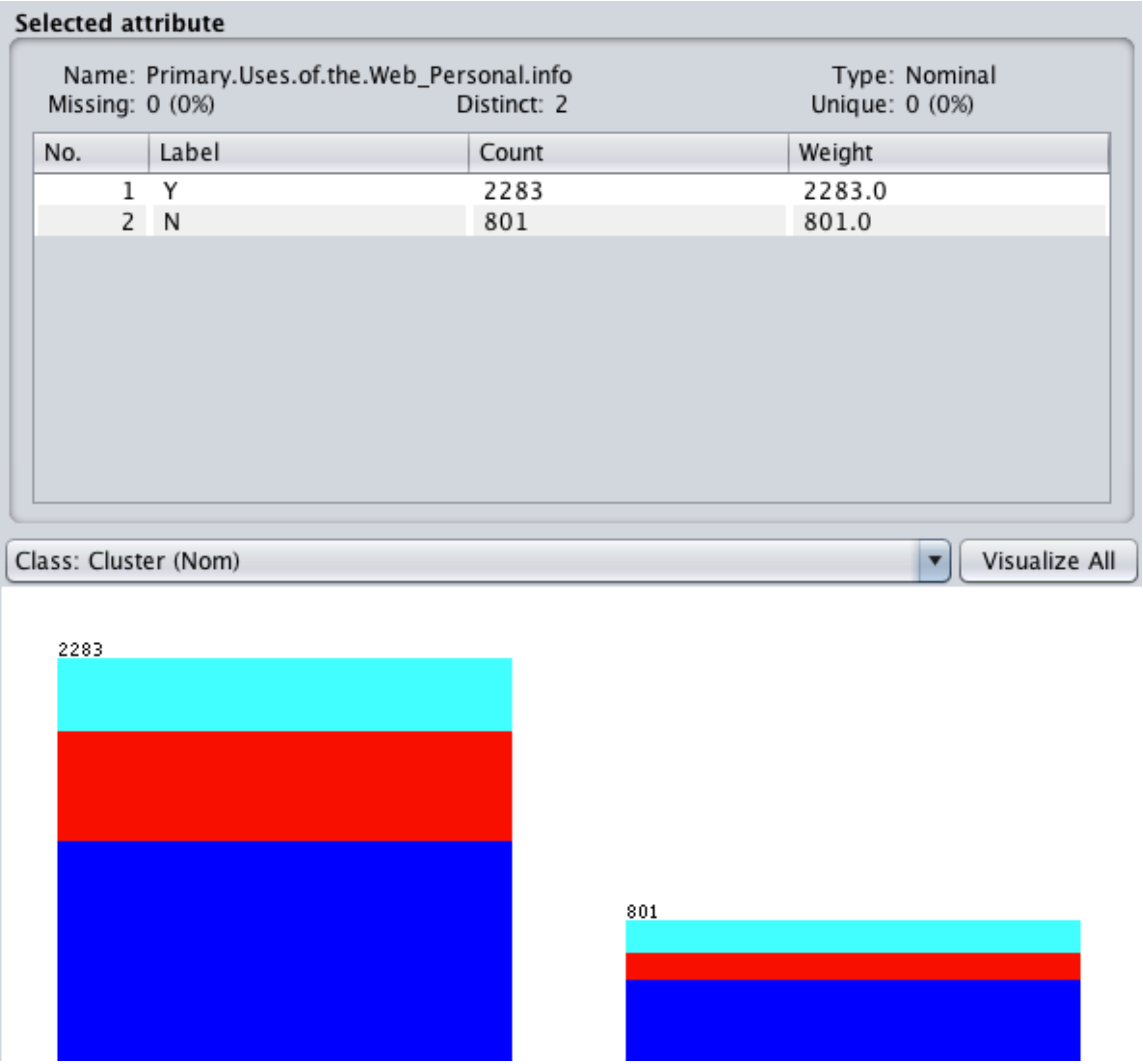
No.	Label	Count	Weight
1	N	1442	1442.0
2	Y	1642	1642.0

Class: Cluster (Nom)



Cluster Blue- Yes, the web is primarily used for shopping, Cluster Red- No clear trend, Cluster Teal- No, the web is not primarily used for shopping

D) Personal Information



Selected attribute

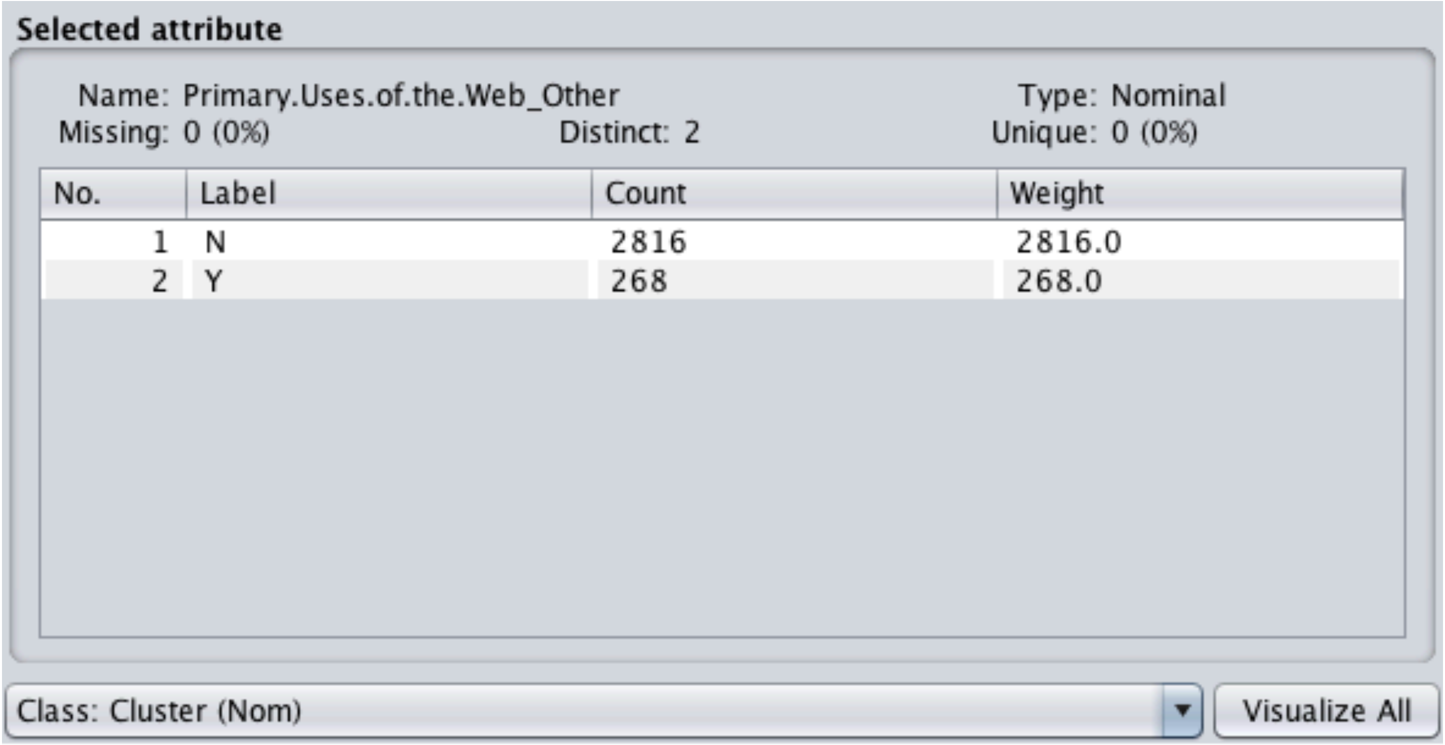
Name: Primary.Uses.of.the.Web_Personal.info Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count	Weight
1	Y	2283	2283.0
2	N	801	801.0

Class: Cluster (Nom)

All clusters primarily use the web for personal information.

E) Other

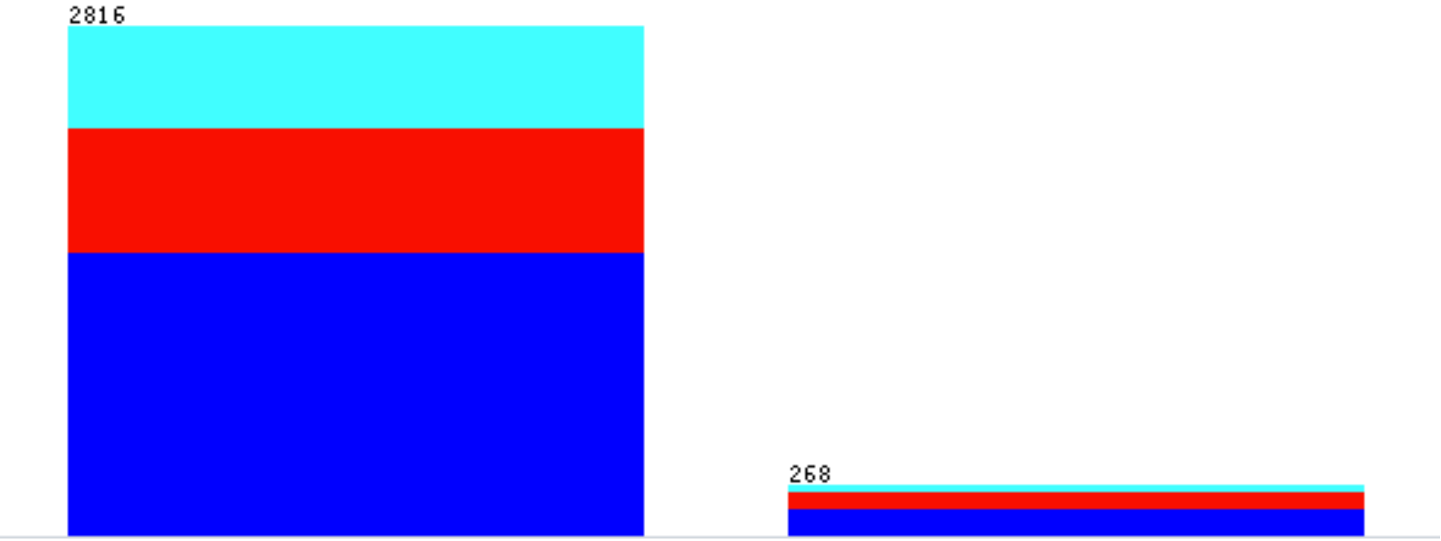


Selected attribute

Name: Primary.Uses.of.the.Web_Other
Type: Nominal
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)

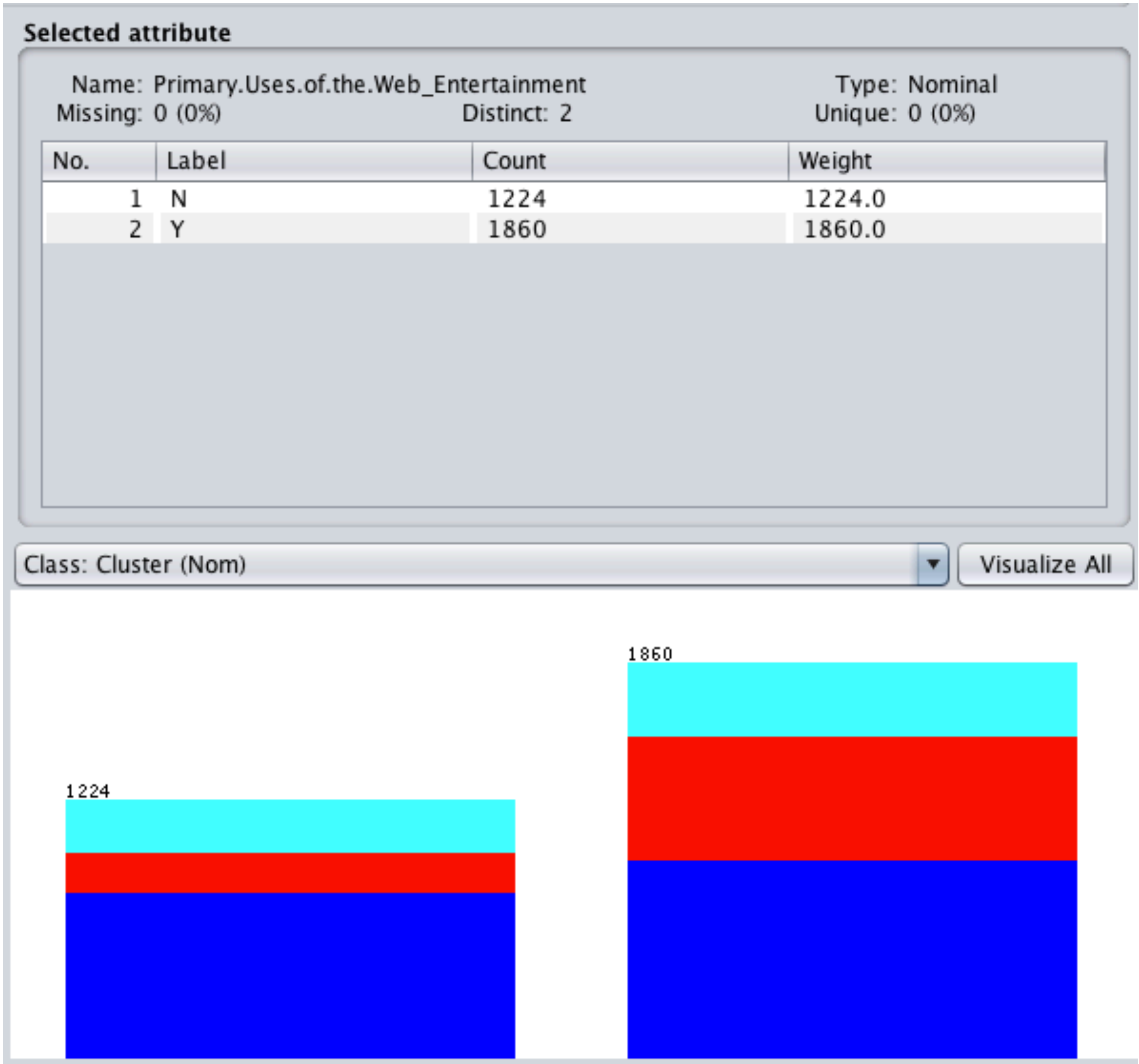
No.	Label	Count	Weight
1	N	2816	2816.0
2	Y	268	268.0

Class: Cluster (Nom)



All clusters primarily use the web for reasons other than those specified in this survey.

F) Entertainment



Selected attribute

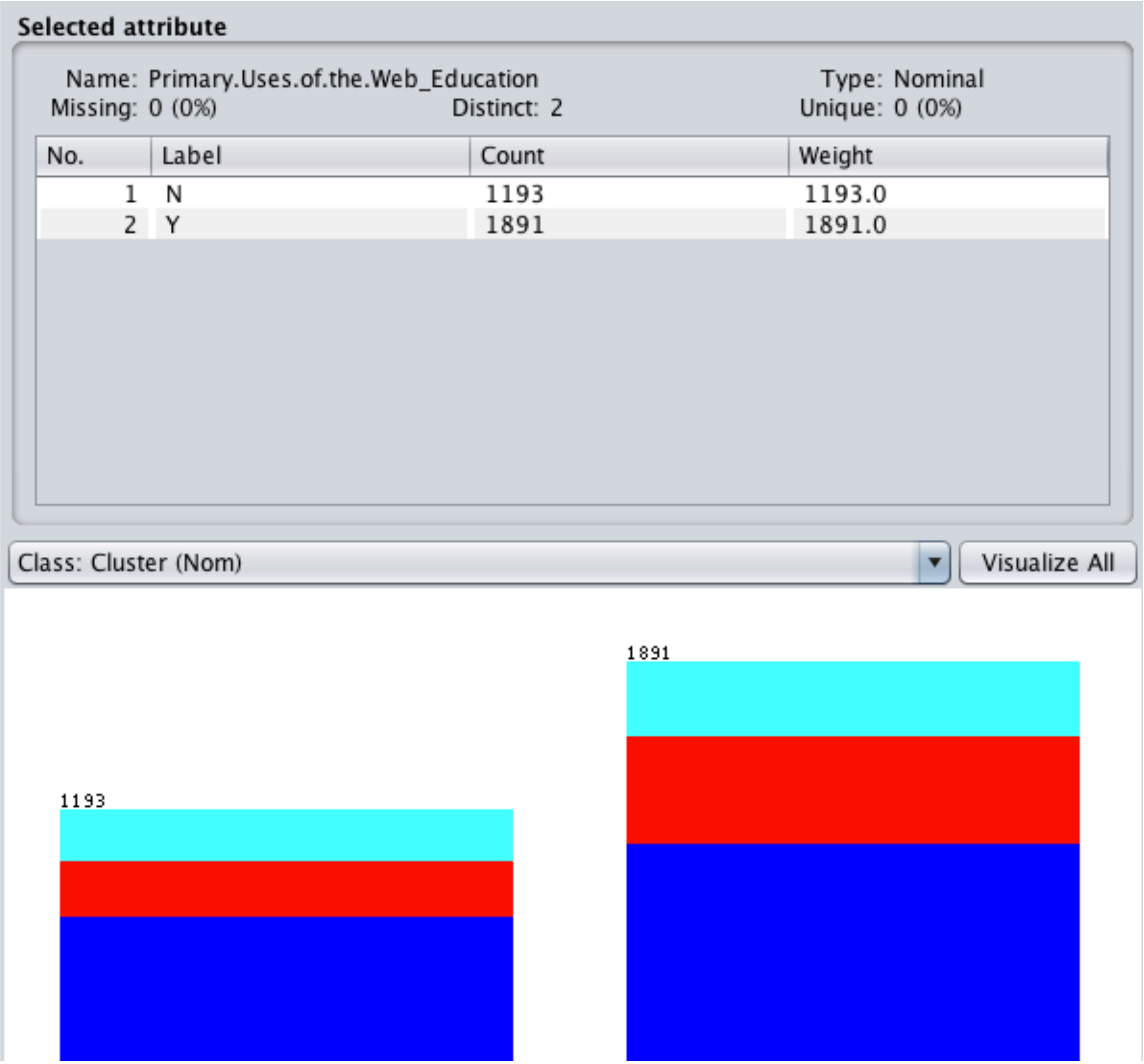
Name: Primary.Uses.of.the.Web_Entertainment   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	N	1224	1224.0
2	Y	1860	1860.0

Class: Cluster (Nom)

Cluster Blue- No clear trend, Cluster Red- Mostly Yes, the web is primarily used for entertainment. Cluster Teal- No clear trend

G) Education



Selected attribute

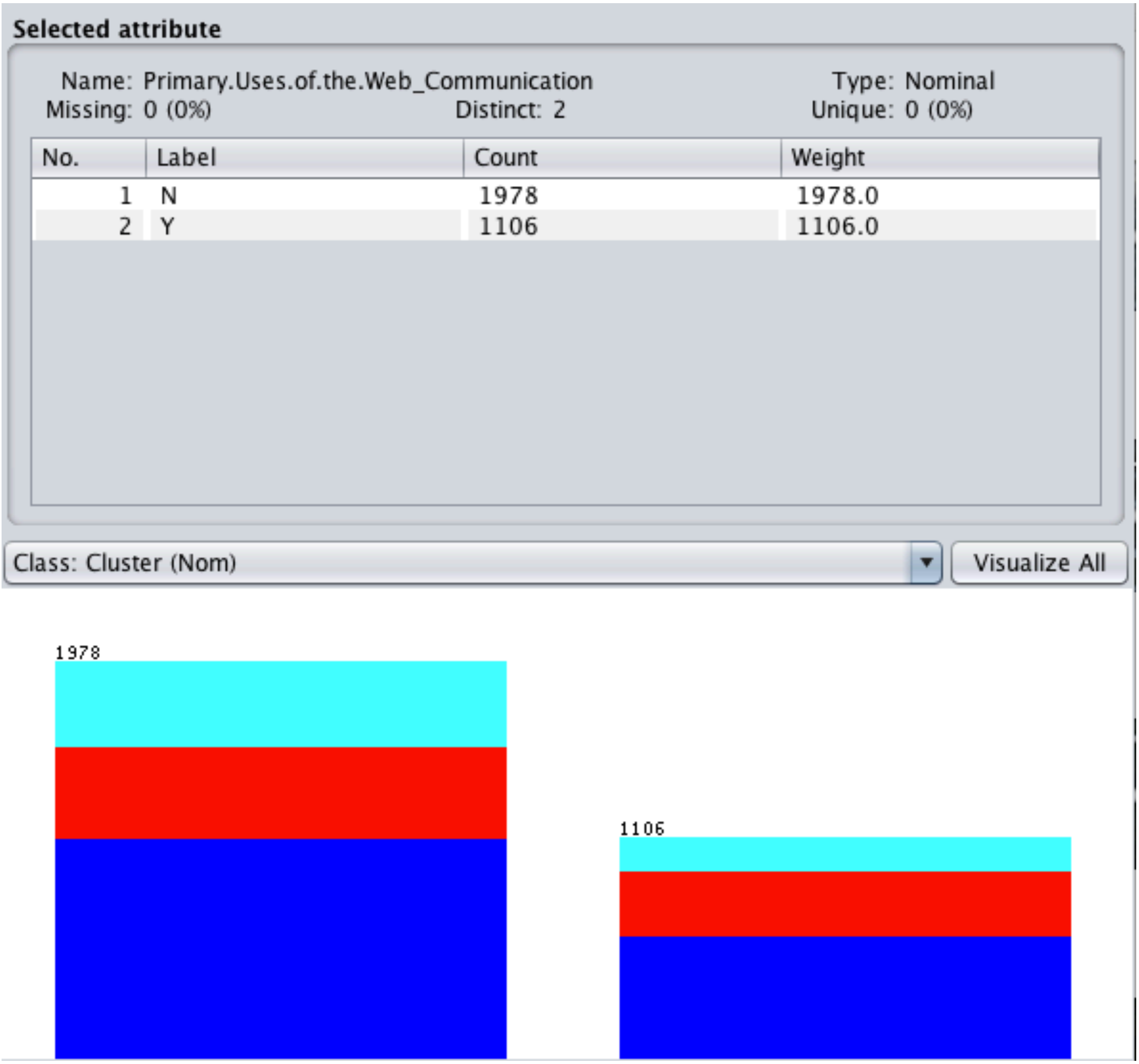
Name: Primary.Uses.of.the.Web_Education
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	N	1193	1193.0
2	Y	1891	1891.0

Class: Cluster (Nom)

Cluster Blue- Yes, the web is primarily used for education, Cluster Red- Yes, the web is primarily used for education, Cluster Teal- No clear trend

H) Communication



Selected attribute

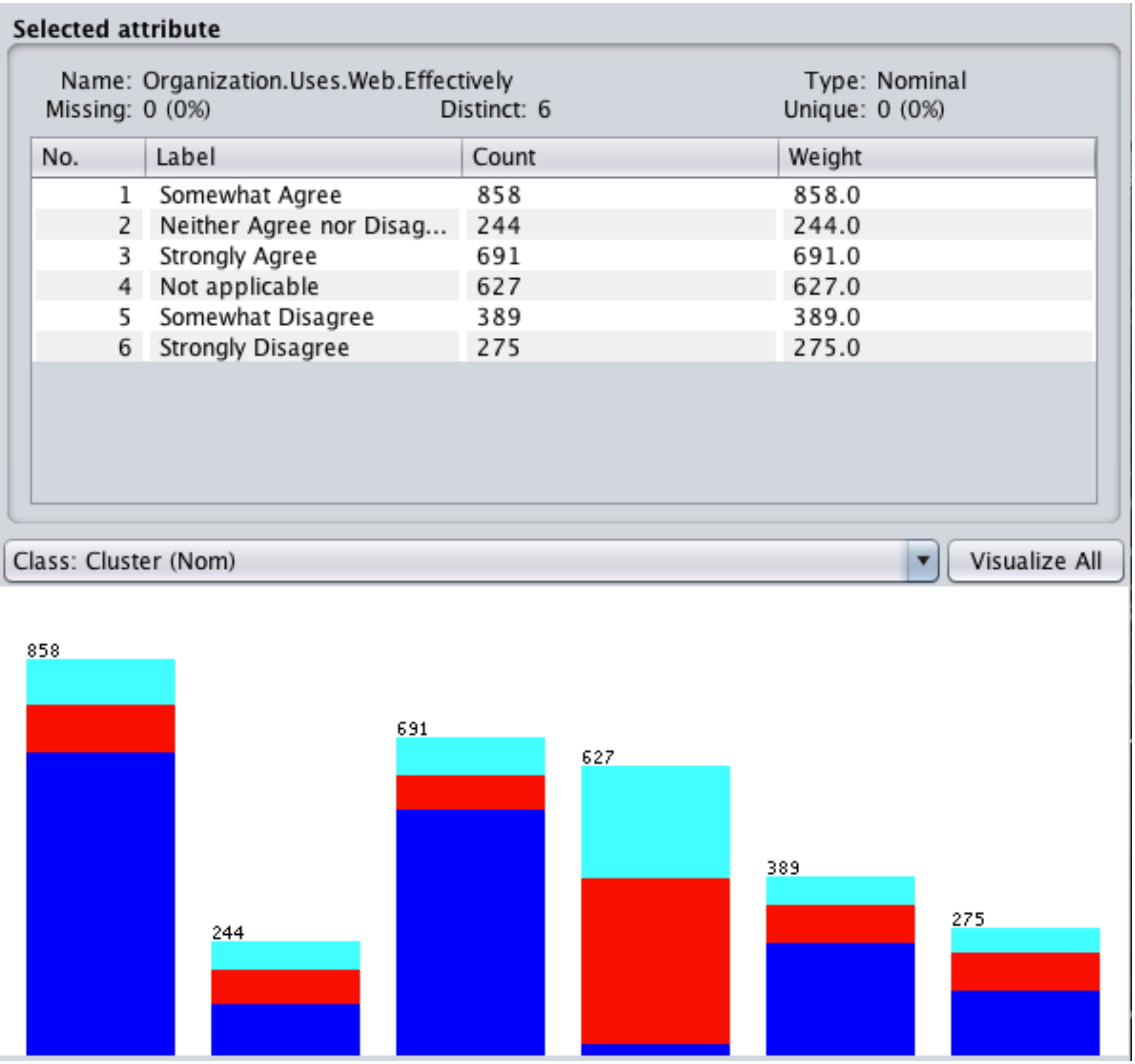
Name: Primary.Uses.of.the.Web_Communication   Type: Nominal
Missing: 0 (0%)   Distinct: 2   Unique: 0 (0%)

No.	Label	Count	Weight
1	N	1978	1978.0
2	Y	1106	1106.0

Class: Cluster (Nom)

Cluster Blue- Yes, the web is primarily used for communication, Cluster Red- No clear trend, Cluster Teal- Yes, the web is primarily used for communication

7) Organization Use Profile



Selected attribute

Name: Organization.Uses.Web.Effectively
Type: Nominal
Missing: 0 (0%)
Distinct: 6
Unique: 0 (0%)

No.	Label	Count	Weight
1	Somewhat Agree	858	858.0
2	Neither Agree nor Disag...	244	244.0
3	Strongly Agree	691	691.0
4	Not applicable	627	627.0
5	Somewhat Disagree	389	389.0
6	Strongly Disagree	275	275.0

Class: Cluster (Nom)

Cluster Blue- Agrees that the organization uses the web effectively, Cluster Red- No clear trend, Cluster Teal- No clear trend

Section 6: *Summary – Answering Question 1 & 2*
The typical groups of web users are as follows:

Cluster Blue (Originally Cluster 0):

- High Frequency of Use of the Web
- Ages 21 to 35 and 36 to 40
- Lives in the USA and Europe
- Mostly Male
- Some College to Masters Degree
- Trained Professional as Occupation
- Most Likely Married or Possibly Single

- Household Income of $50-74k or Over $100k
- Pays for Their Own Access to the Web
- Has High to Moderate Experience With the Internet
- Very Comfortable Using the Internet
- Speed and Government Regulation Are Important to Them
- Disagree With Censorship
- Uses Computing Platform Win95 and Mac/Sys8
- Low Image Loading Frequency
- Mostly Always Accepts Cookies
- The Web is Primarily Used for Work, Shopping, Personal Information, Education, and Communication
- Feels their Organization Uses the Web Effectively

Cluster Red (Originally Cluster 1):

- High Frequency of Use of the Web
- Age 41 to 50
- Lives in the USA
- Mostly Male
- Some College to College Degree
- Trained Professional as Occupation
- Most Likely Married or Possibly Single
- Household Income of $50-74k
- Pays for Their Own Access or Work Pays for Their Access to the Web
- Has Moderate Experience With the Internet
- Very Comfortable Using the Internet
- Speed, Government Regulation, Privacy, Censorship Are Important to Them
- Disagree With Censorship
- Uses Computing Platform Win95 and Win98
- Low Image Loading Frequency
- Mostly Always Accepts Cookies
- The Web is Primarily Used for Personal Information, Entertainment, and Education

Cluster Teal (Originally Cluster 2):

- High Frequency of Use of the Web
- Age 21 to 25
- Lives in the USA
- Slightly More Female than Male
- Some College to College Degree
- Trained Professional as Occupation
- Most Likely Married or Possibly Single
- Household Income Not Available or $50-74k
- Pays for Their Own Access or Work Pays for Their Access to the Web
- Has Moderate Experience With the Internet
- Very Comfortable Using the Internet
- Privacy and Pornography Are Important to Them
- Agrees With Censorship
- Uses Computing Platform Win95 and Win98
- Low Image Loading Frequency
- Cookie Patterns Unclear

- The Web is Primarily Used for Personal Information and Communication

Suggestions for Targeting the Most Important Customers:

After identifying the characteristics of the three clusters it is apparent that Cluster Blue (originally Cluster 0) is the cluster containing the most important customers. Marketing and advertising strategies that emphasize work productivity and personal and work organization since these customers are highly educated, trained professionals. In addition, most customers in this group are married so family or children based products and marketing should be considered as well. Lastly, with high and frequent internet usage online marketing strategies and advertisements would also prove to be effective.