WORKSHOP 2

By Group 5-
Anand Mohan Thakur
Josh Shaji
Poonam Bhaliyan
Poornima Singh
Prateek Ramjanam Singh

PART II: ASSOCIATION RULES

Answer 2, Page 5-76

Question: Save the output window as *WS2-pg5-76.lst* and submit with this workshop
Answer: Submitted the output file as *WS2-pg5-76.lst* on Black Board

Question: Based on the Statistics Plot, which two rules have the highest support percentage? What are the confidence percentages for these rules? Although the rules are based on the same item sets**,** why are the confidence values different?
Answer: Rules with the highest support percentage and their confidence percentages:
Rule 1: SVG-> CKNG, Confidence Percentage: 87.56, Support Percentage: 54.17
Rule 2: CKNG-> SVG, Confidence Percentage: 63.15, Support Percentage: 54.17

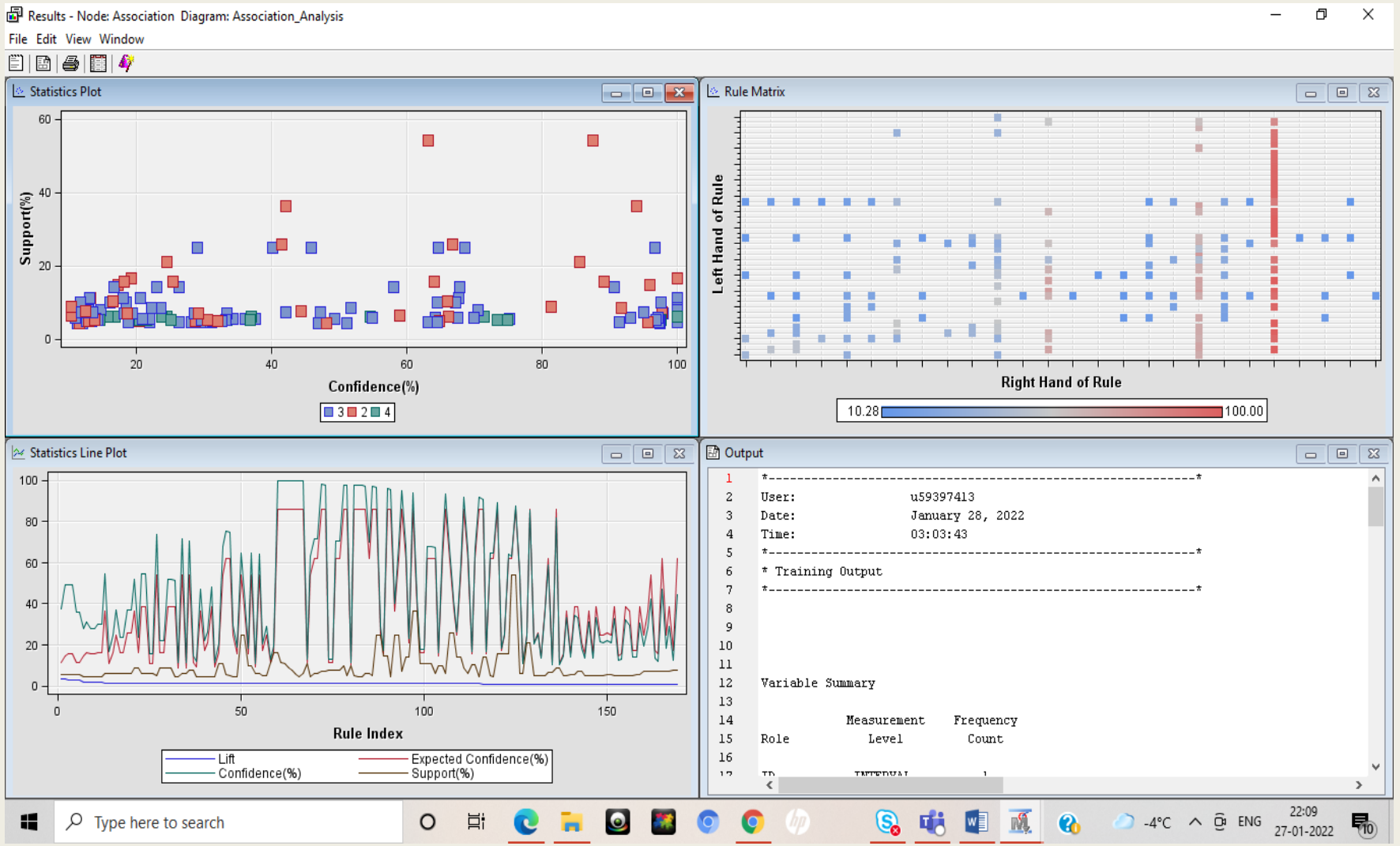
Rule (X→ Y) talks about the possibility of an occurrence of a consequent(Y) provided the antecedent is occurring(X).
In Rule 1:Conf(SVG-> CKNG)= Supp(SVG U CKNG)/Supp(SVG)
Here, we measure how often each item of CKNG appears in transactions that contains items in SVG

In Rule 2:
Conf(CKNG-> SVG)= Supp(CKNG U SVG)/Supp(CKNG)
And in this rule we measure how often each item of SVG appears in transactions that contains items in CKNG

The Confidence of vice versa is not necessarily has to be equal because the antecedent and consequent are swapped and hence the confidence in the occurrence of consequent provided with the occurrence of antecedent.

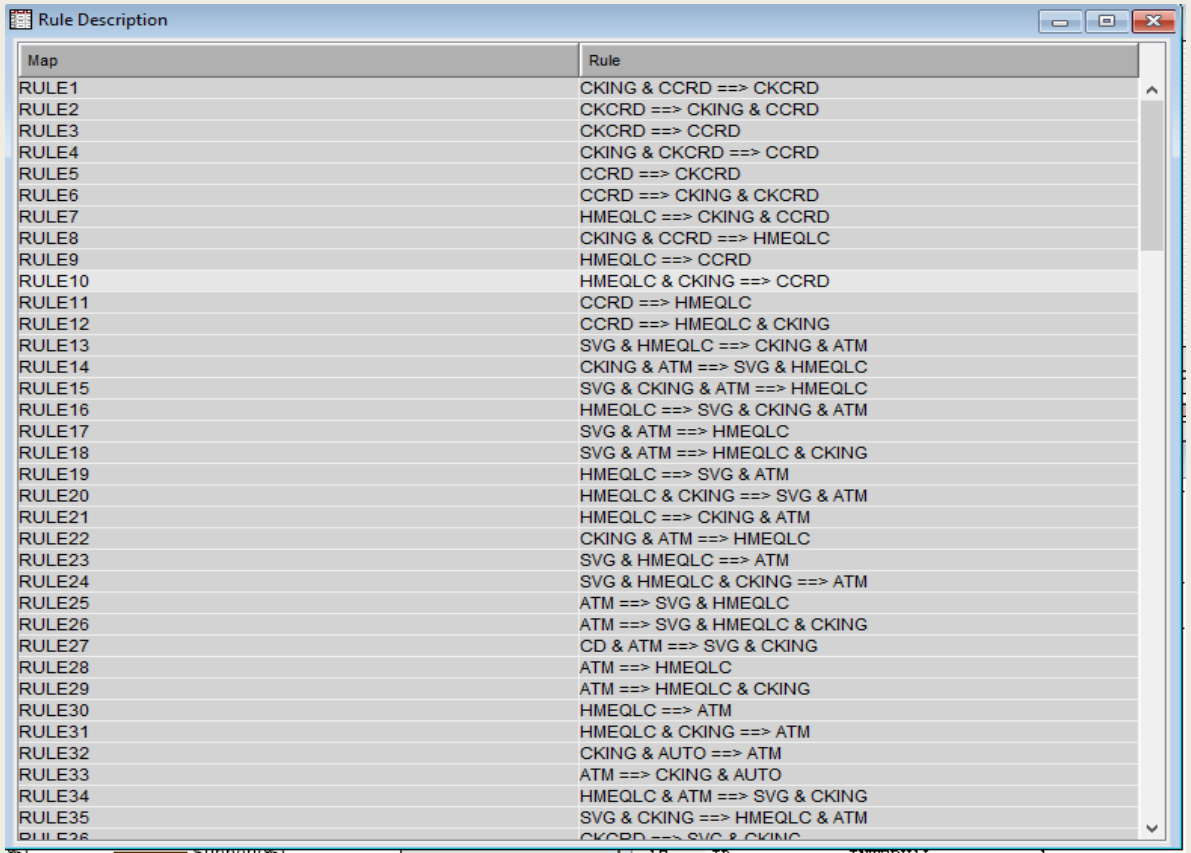
Answer 2, Page 5-73

Question: What is the maximum lift? Which rule has this lift value? What does this rule say?
Answer: The maximum lift value is 3.3 for Rule 2 (CKNG-> SVG).
Lift is the confidence divided by the expected confidence. It is the measure of strength of the association.
Here, Lift=3.3, meaning that a list having CKNG is 3.3 times likely to have SVG in it.



Answer 2, Page 5-78, 79

Question: At step 19 of page 5-78, what does Rule 10 say?
Answer: Rule 10- HMEQLC & CKING ==> CCRD. Rule 10 defines about the occurrence of a customer enquiring about the credit card whenever they are enquiring about the CKING(Checking account) and HMEQLC(Home Equity line of credit) altogether.
Question: At step 20 of page 5-79, by looking at the rule matrix,
Part A: What is the most common right hand of rules?
Answer: The most common right hand of rules is CKNG
Part B: What is the left hand of rules shown on the second row from the top?
Answer : The most common right hand of rules is IRA


Rule Description

Map	Rule
RULE1	CKING & CCRD ==> CKCRD
RULE2	CKCRD ==> CKING & CCRD
RULE3	CKCRD ==> CCRD
RULE4	CKING & CKCRD ==> CCRD
RULE5	CCRD ==> CKCRD
RULE6	CCRD ==> CKING & CKCRD
RULE7	HMEQLC ==> CKING & CCRD
RULE8	CKING & CCRD ==> HMEQLC
RULE9	HMEQLC ==> CCRD
RULE10	HMEQLC & CKING ==> CCRD
RULE11	CCRD ==> HMEQLC
RULE12	CCRD ==> HMEQLC & CKING
RULE13	SVG & HMEQLC ==> CKING & ATM
RULE14	CKING & ATM ==> SVG & HMEQLC
RULE15	SVG & CKING & ATM ==> HMEQLC
RULE16	HMEQLC ==> SVG & CKING & ATM
RULE17	SVG & ATM ==> HMEQLC
RULE18	SVG & ATM ==> HMEQLC & CKING
RULE19	HMEQLC ==> SVG & ATM
RULE20	HMEQLC & CKING ==> SVG & ATM
RULE21	HMEQLC ==> CKING & ATM
RULE22	CKING & ATM ==> HMEQLC
RULE23	SVG & HMEQLC ==> ATM
RULE24	SVG & HMEQLC & CKING ==> ATM
RULE25	ATM ==> SVG & HMEQLC
RULE26	ATM ==> SVG & HMEQLC & CKING
RULE27	CD & ATM ==> SVG & CKING
RULE28	ATM ==> HMEQLC
RULE29	ATM ==> HMEQLC & CKING
RULE30	HMEQLC ==> ATM
RULE31	HMEQLC & CKING ==> ATM
RULE32	CKING & AUTO ==> ATM
RULE33	ATM ==> CKING & AUTO
RULE34	HMEQLC & ATM ==> SVG & CKING
RULE35	SVG & CKING ==> HMEQLC & ATM
RULE36	CKCRD ==> SVG & CKING

Answer 2, Page 5-79

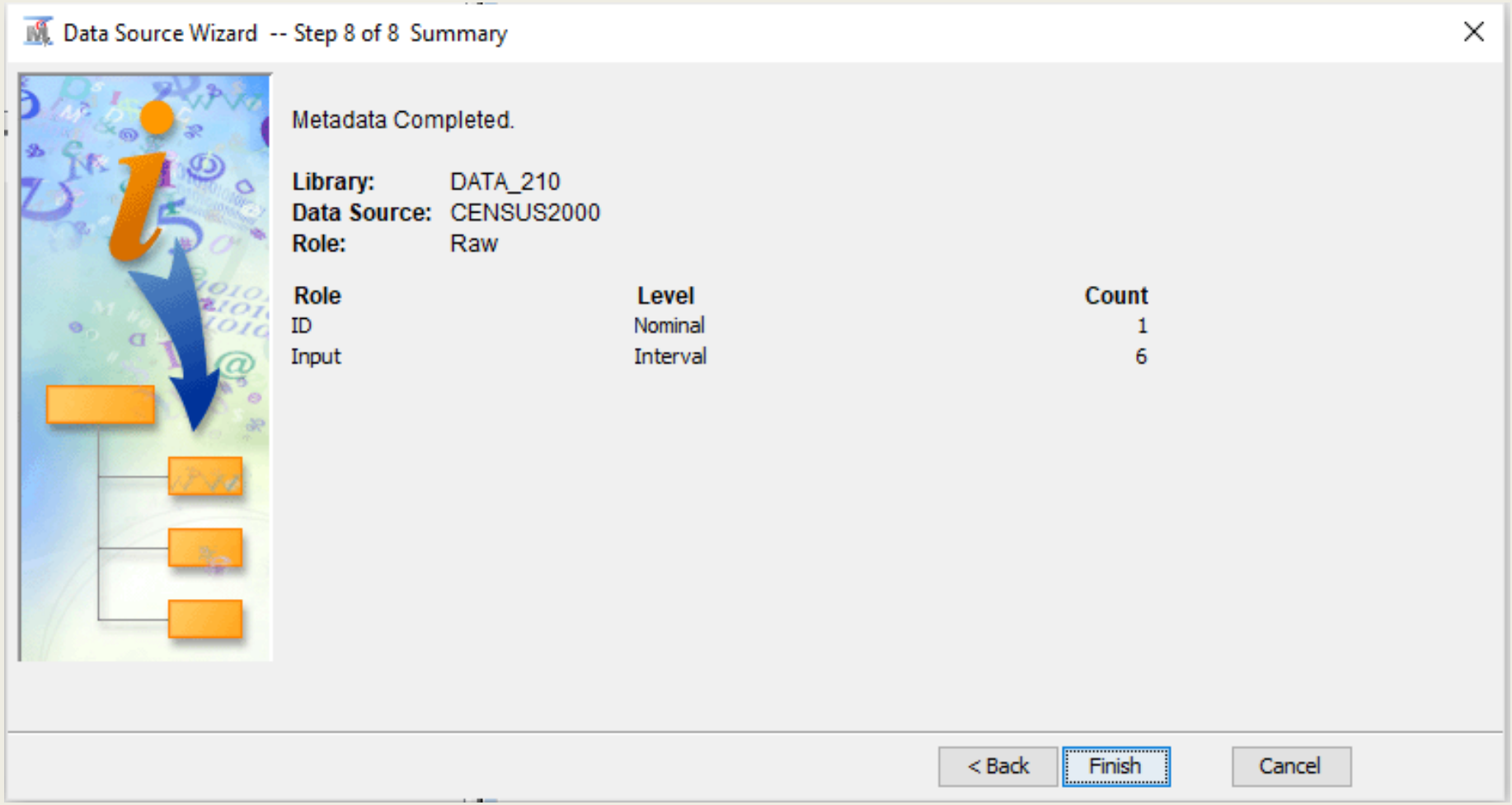
1. **Question:** List five possible actionable decisions from the analysis (Page 5-79, idea exchange).
   Answer:
   1. A saving and a checking account services to be provided together to the customer.
   2. A check card should be issued whenever a customer avails credit card and checking account service.
   3. A customer with a home equity line of credit can be issued checking account and a check card if they already don’t have a checking account with the bank.
   4. If a customer is already having a ATM debit card, then there is no need to issue them a checking account and a automobile installment loan. Chances are since they already hold an account, they might already have these services.
   5. A credit card is to be issued whenever a check/debit card is issued by the bank to the new customer.

PART III: CLUSTERING

Answer 3, Page 5-28

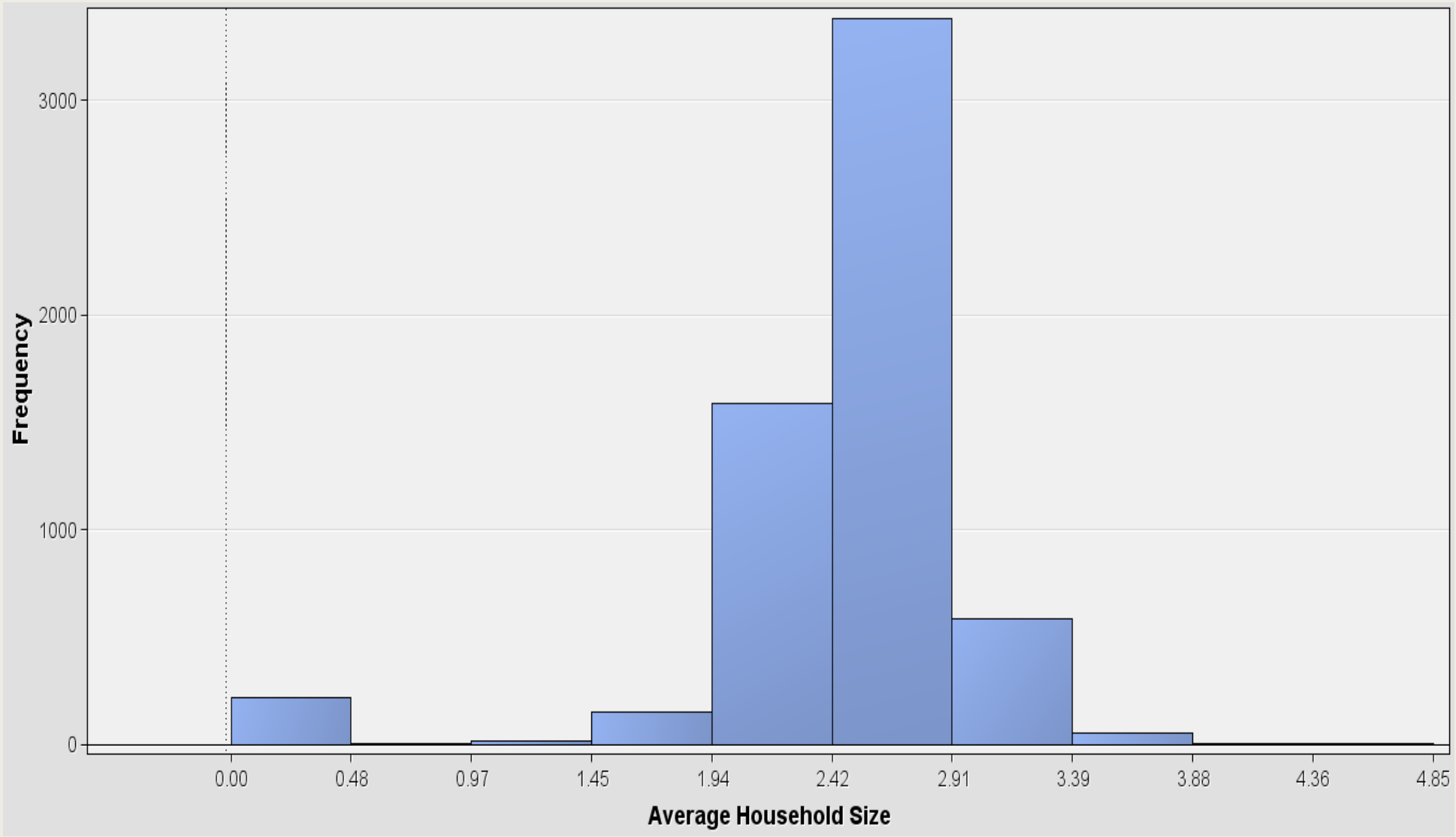
Question: At step 7 of page 5-28, replace the following image with a screenshot of the Data Source Wizard, showing the summary information:

Answer:



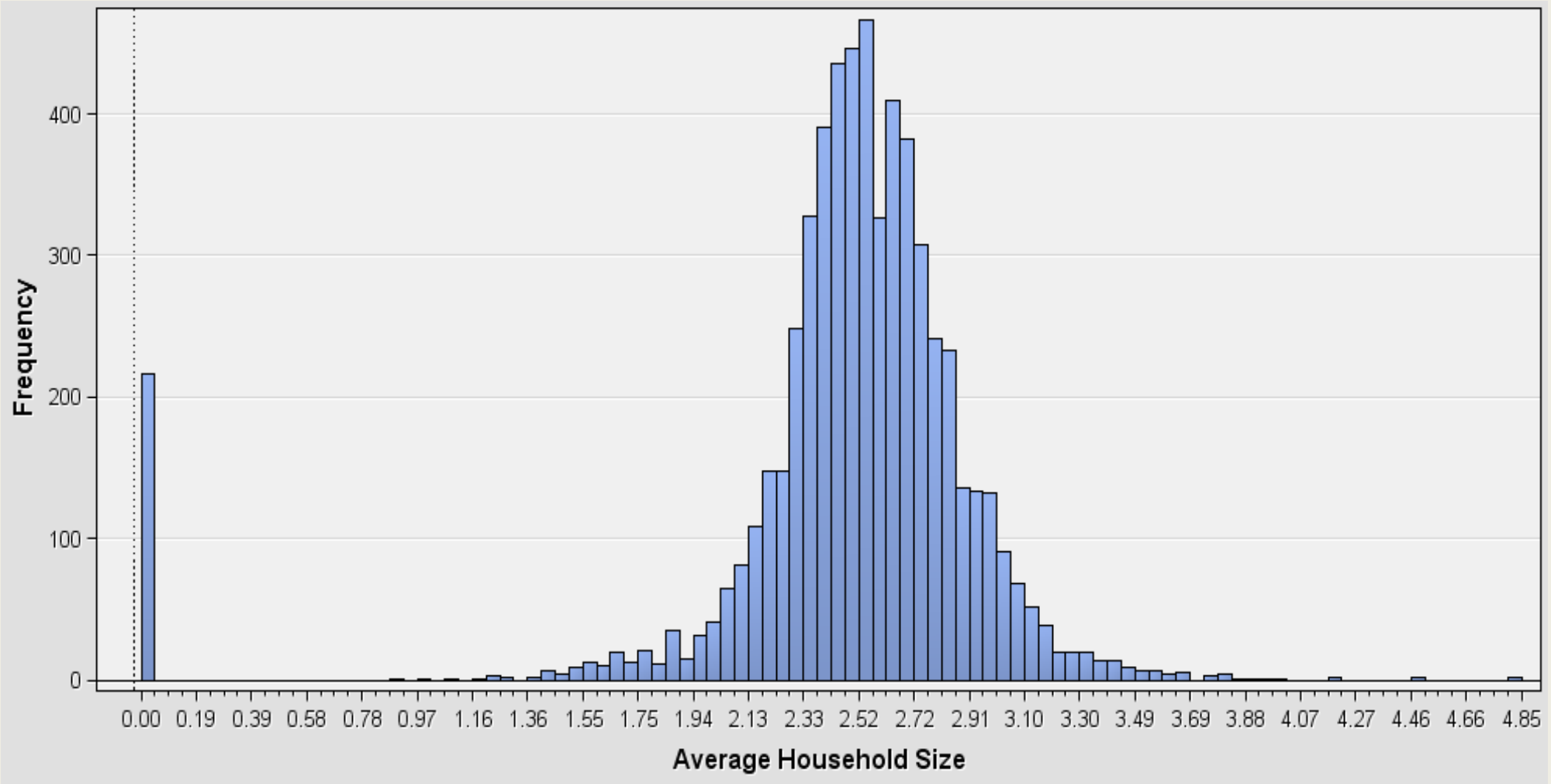
Answer 3, Page 5-29

Question: At step 4 of page 5-29, paste an image of your histogram here.
Answer:



Answer 3, Page 5-31

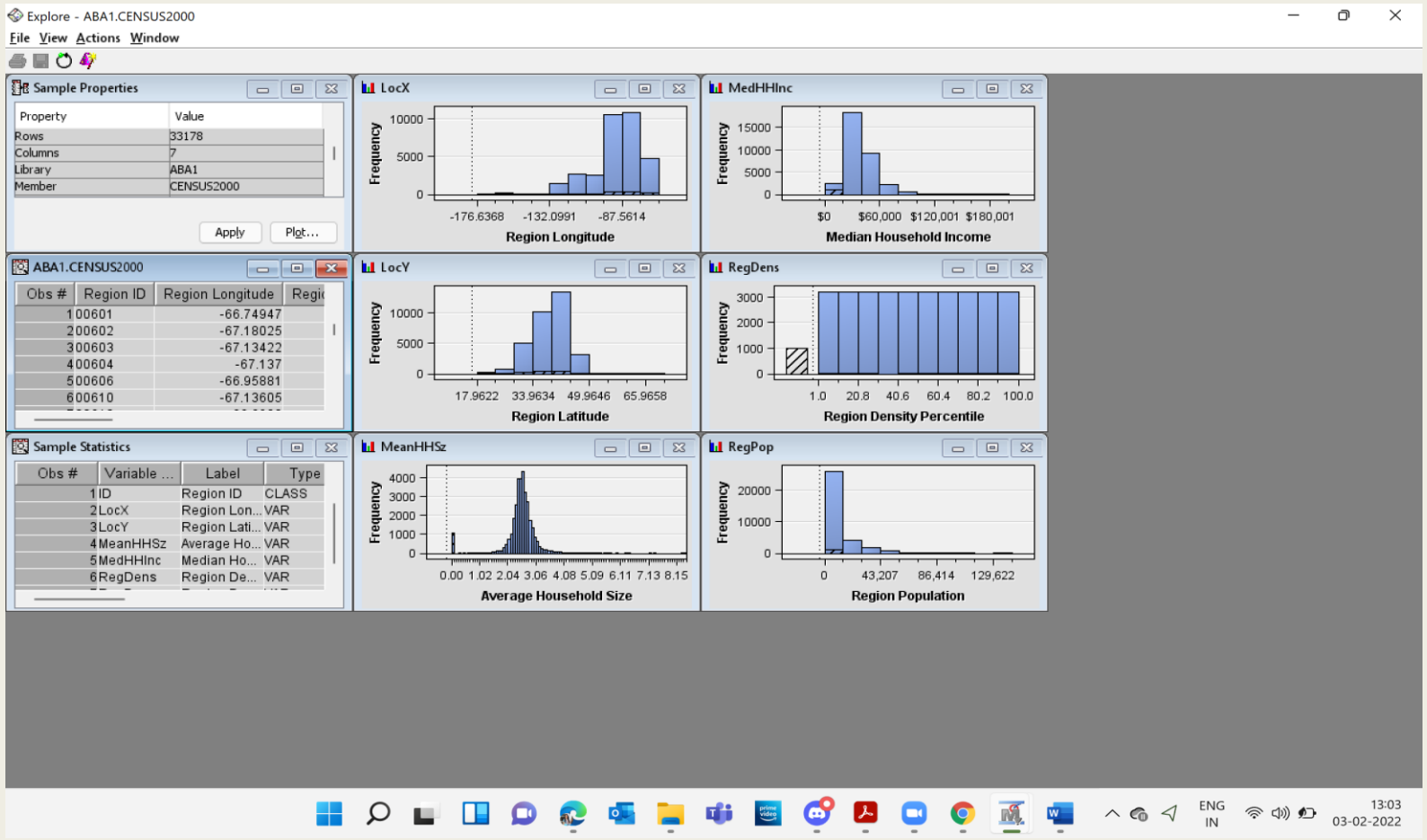
Question: At step 6 of page 5-31, paste an image of your histogram here.
Answer:



Answer 3, Page 5-32

Question: Paste an image supporting each of the observations made on page 5-32:
*The zero average household size seems to be evenly distributed across the longitude and latitude variables.
*It seems concentrated on low incomes and populations,
And also makes up the majority of the missing observations in the distribution of region density

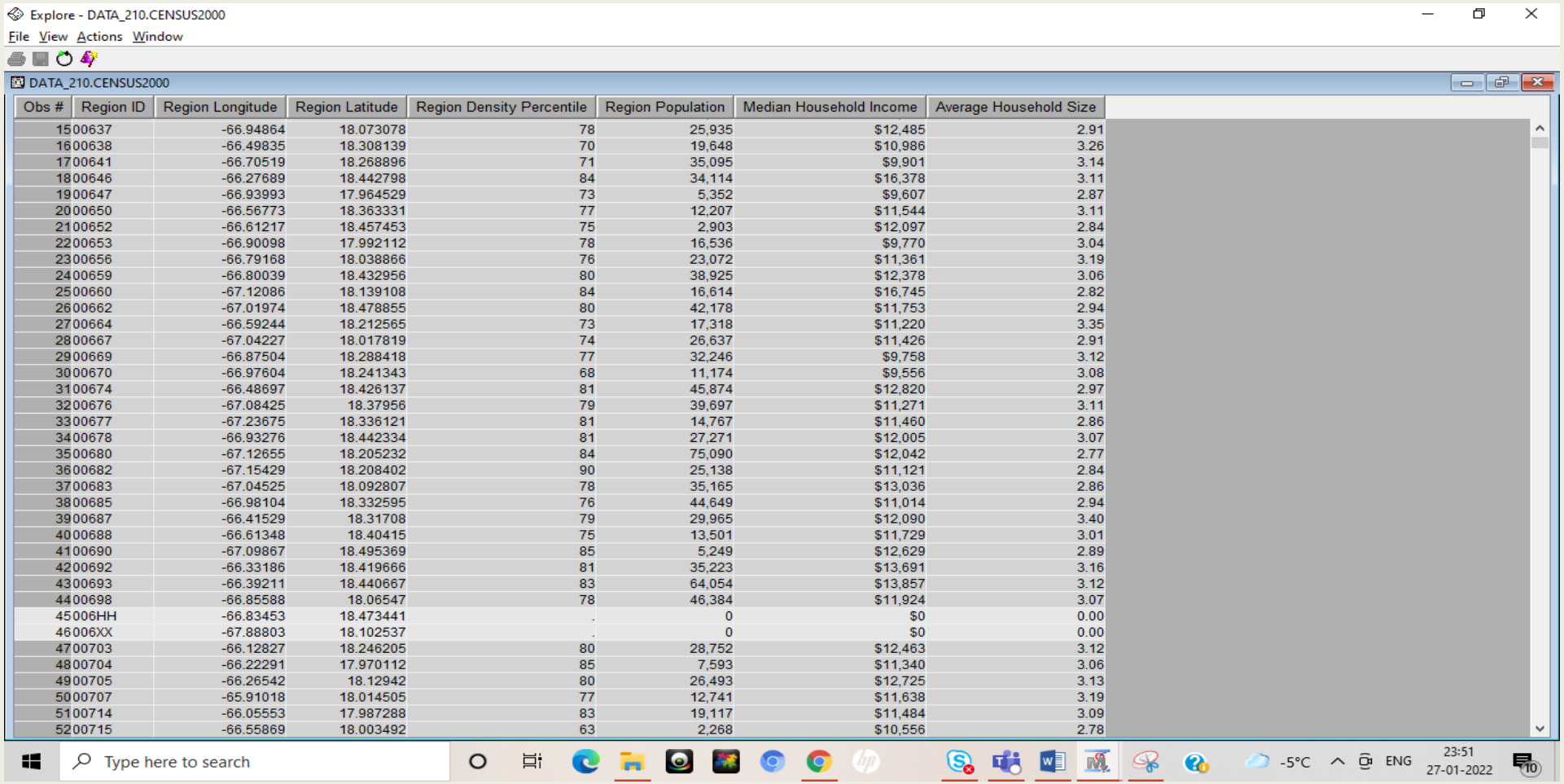
Answer:



Continued Answer 3

Answer 3, Page 5-32

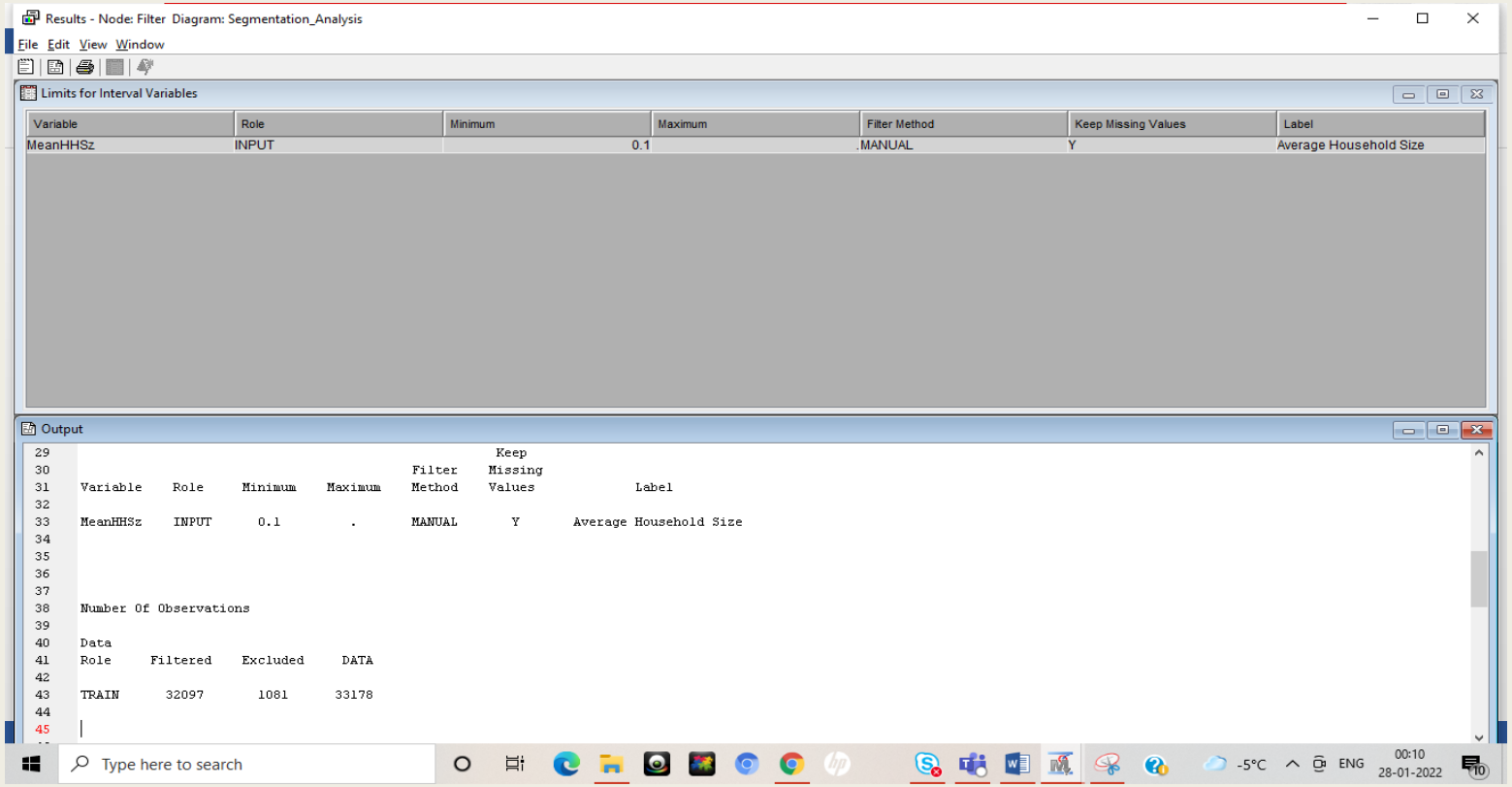
Question: Paste an image supporting each of the observations made on page 5-32:
*The zero average household size seems to be evenly distributed across the longitude and latitude variables.
*It seems concentrated on low incomes and populations,
And also makes up the majority of the missing observations in the distribution of region density
Answer:



Obs #	Region ID	Region Longitude	Region Latitude	Region Density Percentile	Region Population	Median Household Income	Average Household Size
1	500637	-66.94864	18.073078	78	25,935	$12,485	2.91
1	600638	-66.49835	18.308139	70	19,648	$10,986	3.26
1	700641	-66.70519	18.268896	71	35,095	$9,901	3.14
1	800646	-66.27689	18.442798	84	34,114	$16,378	3.11
1	900647	-66.93993	17.964529	73	5,352	$9,607	2.87
2	000650	-66.56773	18.363331	77	12,207	$11,544	3.11
2	100652	-66.61217	18.457453	75	2,903	$12,097	2.84
2	200653	-66.90098	17.992112	78	16,536	$9,770	3.04
2	300656	-66.79168	18.038866	76	23,072	$11,361	3.19
2	400659	-66.80039	18.432956	80	38,925	$12,378	3.06
2	500660	-67.12086	18.139108	84	16,614	$16,745	2.82
2	600662	-67.01974	18.478855	80	42,178	$11,753	2.94
2	700664	-66.59244	18.212565	73	17,318	$11,220	3.35
2	800667	-67.04227	18.017819	74	26,637	$11,426	2.91
2	900669	-66.87504	18.288418	77	32,246	$9,758	3.12
3	000670	-66.97604	18.241343	68	11,174	$9,556	3.08
3	100674	-66.48697	18.426137	81	45,874	$12,820	2.97
3	200676	-67.08425	18.37956	79	39,697	$11,271	3.11
3	300677	-67.23675	18.336121	81	14,767	$11,460	2.86
3	400678	-66.93276	18.442334	81	27,271	$12,005	3.07
3	500680	-67.12655	18.205232	84	75,090	$12,042	2.77
3	600682	-67.15429	18.208402	90	25,138	$11,121	2.84
3	700683	-67.04525	18.092807	78	35,165	$13,036	2.86
3	800685	-66.98104	18.332595	76	44,649	$11,014	2.94
3	900687	-66.41529	18.31708	79	29,965	$12,090	3.40
4	000688	-66.61348	18.40415	75	13,501	$11,729	3.01
4	100690	-67.09867	18.495369	85	5,249	$12,629	2.89
4	200692	-66.33186	18.419666	81	35,223	$13,691	3.16
4	300693	-66.39211	18.440667	83	64,054	$13,857	3.12
4	400698	-66.85588	18.06547	78	46,384	$11,924	3.07
45	006HH	-66.83453	18.473441	.	0	$0	0.00
46	006XX	-67.88803	18.102537	.	0	$0	0.00
4	700703	-66.12827	18.246205	80	28,752	$12,463	3.12
4	800704	-66.22291	17.970112	85	7,593	$11,340	3.06
4	900705	-66.26542	18.12942	80	26,493	$12,725	3.13
5	000707	-65.91018	18.014505	77	12,741	$11,638	3.19
5	100714	-66.05553	17.987288	83	19,117	$11,484	3.09
5	200715	-66.55869	18.003492	63	2,268	$10,556	2.78

Answer 3, Page 5-38, 40, 41

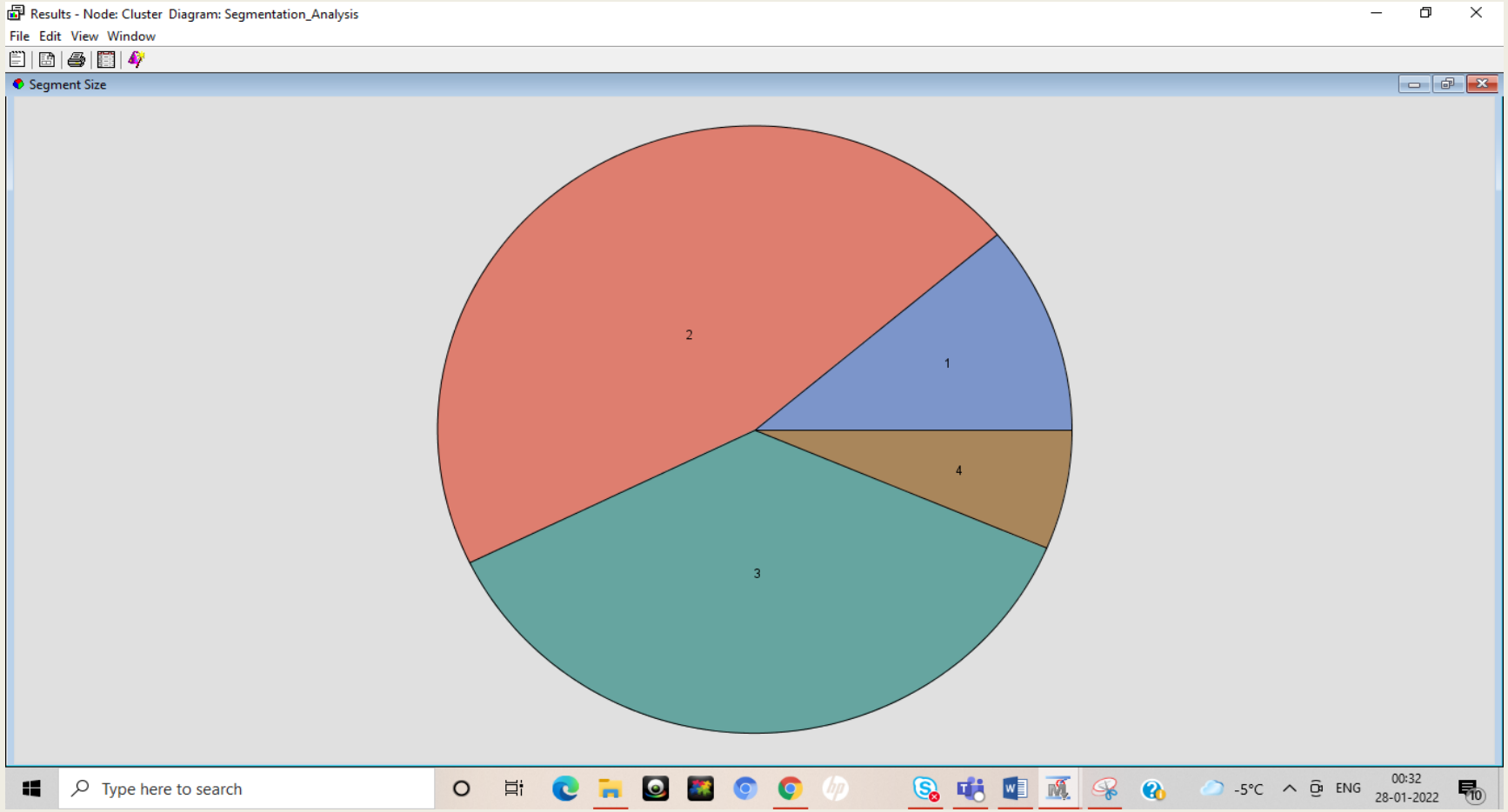
Question: At step 10 of page 5-38, save the output window as *WS2-pg5-38.lst* and submit with this workshop
Answer: Submitted the output file as *WS2-pg5-38.lst* on Black Board



Question: What standardization method is used (page 5-40 to 5-41)? What does it do and why?
Answer: Standardized distance method has been used to calculate distances.
This method subtracts the mean and divides by the standard deviation of the input values.
Standardization is required to create clusters of the selected input variables as the variables are on different measurement scales

Answer 3, Page 5-42, 45

Question: At step 1 of page 5-42, paste the segment size diagram here.
Answer:

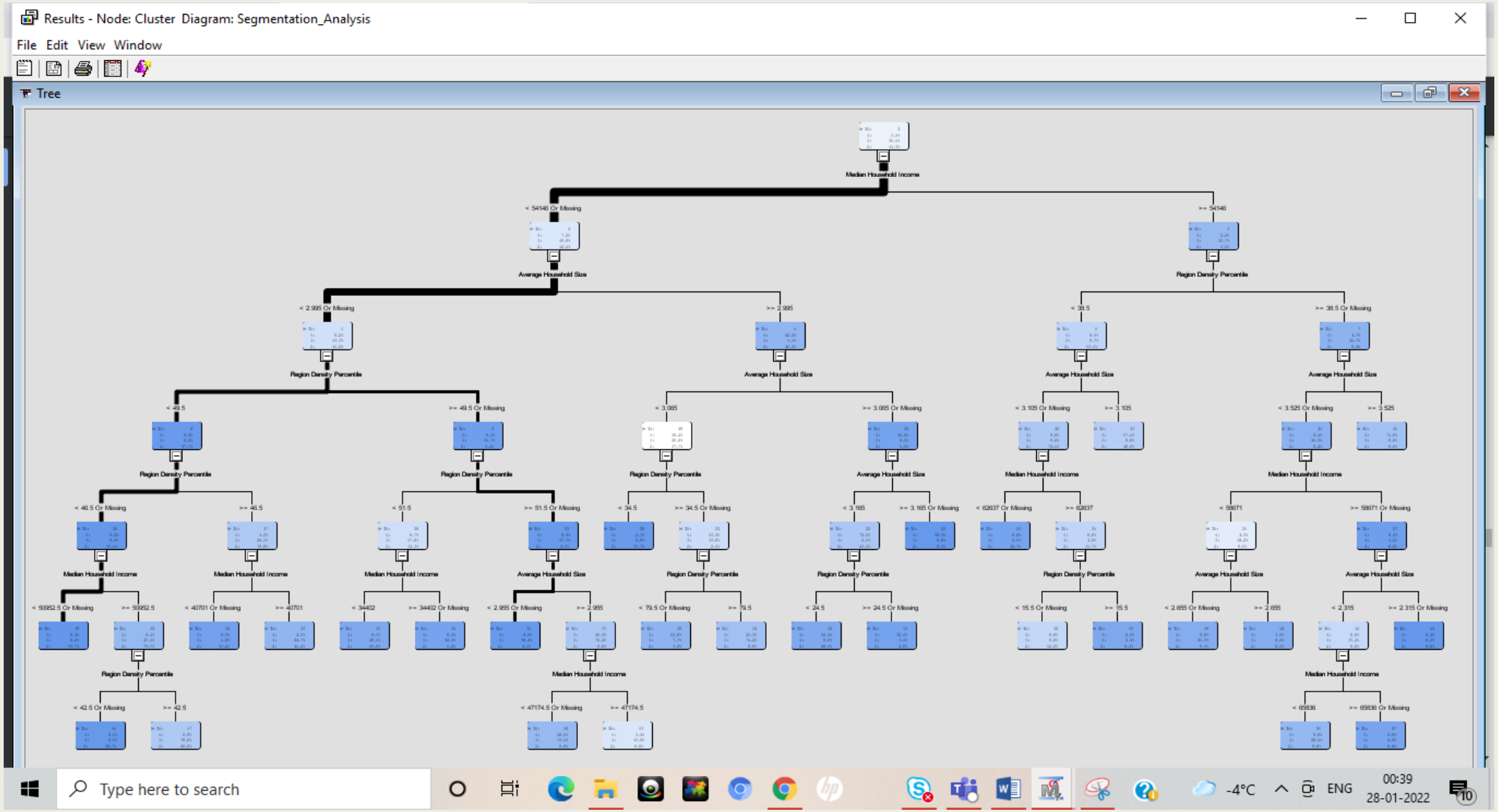


Question: At step 2 of page 5-45, which segment has the highest median household income? And what is the average household size in this segment?
Answer: Segment ID 1 has the highest median household income and the average household size is 2.76

Answer 3, Page 5-46

Question: Paste the Tree on page 5-46 here.

Answer:



GROUP WORK DECLARATION

We, **Group 5 (Anand Mohan Thankur, Josh Shaji, Poonam Bhaliyan, Prateek Ramjanam Singh, and Poornima Singh)** declare that the attached assignment is our own work in accordance with the Seneca Academic Policy. We have not copied any part of this assignment, manually or electronically, from any other source including web sites, unless specified as references. We have not distributed our work to other students.

	Name	Task(s)
1	Anand Mohan Thakur (149200206)	Consolidated the Workshop together on MS Teams
2	Josh Shaji (133557215)	Consolidated the Workshop together on MS Teams
3	Poonam Bhaliyan (121114219)	Consolidated the Workshop together on MS Teams
4	Prateek Ramjanam Singh (124483215)	Consolidated the Workshop together on MS Teams
5	Poornima Singh (125638213)	Consolidated the Workshop together on MS Teams

THANK YOU