CST8502 - Lab 1 Answers

Student Name: Alvin Litani Liauw
Student Number: 41118874

During demo time, you should be able to answer questions on the process that you did in this lab.

When you take screenshots, make sure that you are getting the shot of the full Weka.

Part 1 - Data Exploration in Weka

10. Fill this table:

Flower Type	Count
Iris-setosa	50
Iris- versicolor	50
Iris-virginica	50

11. Fill this table:

Attribute	Type	Missing	Distinct	Unique	Minimum	Maximum	Mean	StdDev
sepallength	Numeric	0	35	9	4.3	7.9	5.843	0.828
sepalwidth	Numeric	0	23	5	2	4.4	3.054	0.434
petallength	Numeric	0	43	10	1	6.9	3.759	1.764
petalwidth	Numeric	0	22	2	0.1	2.5	1.199	0.763

12. Explanation of Distinct: List of elements that appear after removing duplicates

    Explanation of Unique: List of elements that appear only once

    Selected numbers: 41,11,88,74,111,18,87,18,41,11

    #Distinct: 7

    #Unique: 4

    Distinct numbers: 41,11,88,74,111,18,87

    Unique numbers: 88,74,111,87

Part 2 - Data Preparation and Cleaning

3. Student number: 41118874

   Row number of the new instance: 75

   Salary: 18874

   Screenshot:

73	41010222	Kerrie	Middlehur	kmiddlehurstl@bigcartel.com	59845 Green Ridge Way	U.S.A.	1 USD	86000
74	41010223	Temp	Geratt	tgerattb@cyberchimps.com	262 Bonner Plaza	U.S.A.	1 USD	100000
75	41118874	Alvin	Liauw	alvin_liauw@test.com	411 Liauw Street	Canada	8 CAD	18874
76	41010224	Lenora	Beggan	lbeggan1i@howstuffworks.com	36266 Clarendon Alley	Germany	2 EUR	37028
77	41010225	Micheil	Mendez	mmendez2q@earthlink.net	6423 Sauthoff Alley	China	3 CHY	140000
78	41010226	Brinn	Barracks	bbarracks1q@examiner.com	9782 Dovetail Park	Germany	2 EUR	70000
79	41010227	Johan	Cherrie	jcherriep@twitter.com	616 Barby Pass	U.S.A.	1 USD	22000

7. Which are the four important attributes that are relevant to analyse this dataset?

   Country, Branch, Currency, Salary

8. For the nominal attributes of the above question, fill in the following table:

Attribute Name: Country		**Attribute Name: Branch**	
Label	**Count**	**Label**	**Count**
China	**39**	**1**	**39**
U.S.A.	**39**	**2**	**39**
Germany	**38**	**3**	**40**
Mexico	**38**	**4**	**36**
Canada	**1**	**6**	**1**
Japan	**1**	**8**	**1**

Attribute Name: Currency	
Label	**Count**
CHY	**41**
USD	**39**
EUR	**38**
MXD	**36**
INR	**1**
CAD	**1**

9. Sorted list of anomalies (Sort by ID column) with the reason:

Id	first_name	last_name	email	Address	Country	Branch	Currency	Salary	Reason
40010144	Maressa	Commucci	mcommucci3e@techcrunch.com	3 Prairieview Alley	Mexico	4	MXD	40	This person is being paid way less than others paid in MXD.
40010148	Burton	Dudden	bdudden39@japanpost.jp	050 Nova Court	Mexico	4	INR	45999	The only person being paid in INR.
40010155	Maximilian	Camies	mcamiesv@so-net.ne.jp	7201 Cambridge Park	U.S.A.	1	USD	4000	This person is being paid way less than others paid in USD.
40010158	Darcy	Addie	daddie1k@jalbum.net	836 Marquette Pass	Germany	2	EUR	60500999	This person is being paid way more than others in EUR.
41010219	Chelsie	Mulles	cmulles3x@amazon.co.uk	8 Talmadge Circle	Mexico	6	MXD	36000	The only person who is in branch 6.
41010237	Elsworth	Skells	eskells35@spotify.com	54486 Carberry Park	Mexico	4	CHY	150000	The only person who is in Mexico being paid CHY.
41110305	Misha	Tunna	mtunnaj@dailymotion.com	15 Kipling Drive	U.S.A.	1	USD	32000999	This person is being paid way more than others in USD.
41110334	Bel	Hodgin	bhodgin2g@msu.edu	60 Bellgrove Court	Japan	3	CHY	600000	The only person in Japan.

41110348	Nelson	McRinn	nmcrinn3p@economist.com	56053 Buell Terrace	Mexico	2	MXD	19999	The only person who in branch 2 who is being paid MXD and in Mexico.
41118874	Alvin	Liauw	alvin_liauw@test.com	411 Liauw Street	Canada	8	CAD	18874	The only person in Canada and also being paid CAD.

(Add more rows as required. You must find at least 8 anomalies)

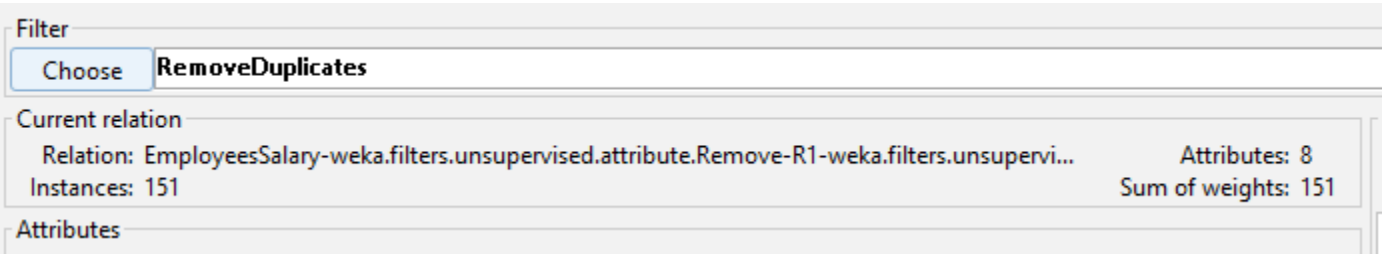
9. How many instances do you have now? _156_

13.

   a. How many instances do you have now? _151_

   b. How many duplicates (how many got removed): _5_

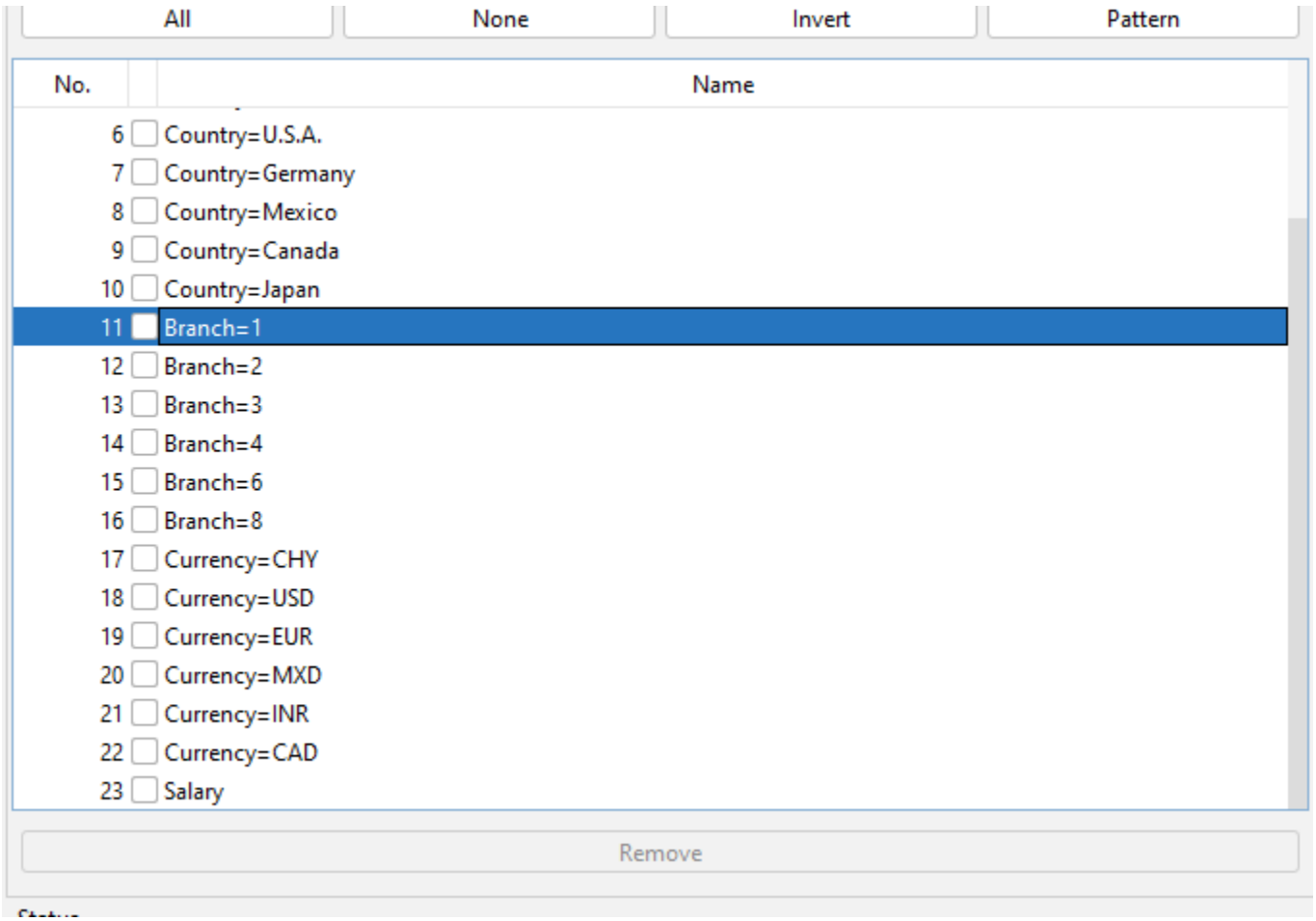
14. Take a screenshot and paste it here.



Nominal to Binary

16. How many nominal attributes do you have? 3

18. Take a screenshot and paste it in lab2 document.



21. Take a screenshot of the file while it is opened in Notepad++. Header should be visible.

In order to get the credit for this lab:

1. Show the EmployeesSalary file in Weka during demo.
2. Show EmployeesSalaryNoDupBinary.arff in Weka.
3. Show the answers in lab1_Answers.doc
4. Upload lab1_Answers.doc, EmployeesSalaryNoDuplicates.arff and EmployeesSalaryNoDupBinary.arff in Brightspace before the submission due date.