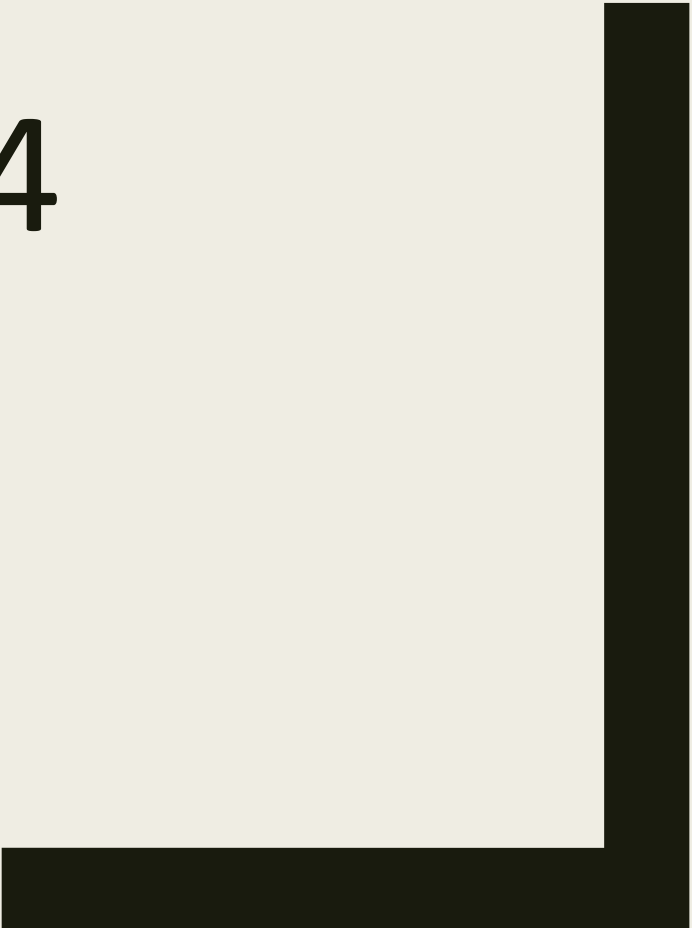




WORKSHOP 1

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





PART I: TEAM CONTRACT

As Group 5, we have prepared a separate document as a Group Contract and submitted it on Blackboard

PART II: SOFTWARE DOWNLOAD

PART III: VIDEO TUTORIAL





PART IV: RETENTION DATASET ANSWERS



Page 3 - 32

Question 1: How many observations (rows) exist in the Retention table?

Answer 1: 2626 Rows

Question 2: How many variables (columns) exist in the Retention table?

Answer 2: 21 Columns

.. Property	Value
ID	retention
Name	RETENTION
Variables	
Decisions	
Role	Raw
Notes	
Library	BAN210
Table	RETENTION
Sample Data Set	
Size Type	
Sample Size	
Type	DATA
No. Obs	2626
No. Cols	21
No. Bytes	332800
Segment	
Created By	u59397413
Create Date	20/1/22 9:33 PM
Modified By	u59397413
Modify Date	20/1/22 9:33 PM
Scope	Local

Page 3 - 32

Question 3: When the “Basic” option is selected as the Metadata Advisor Options of the Data Source Wizard, how are the initial roles and levels of the variables determined? How about the “Advanced” option?

Answer 3: If we choose the Basic option here, the initial role and level of the variables will be determined by the variable type and format values. If we choose the Advanced option here, initial roles and levels are based on the variable type, format values, and the number of distinct values contained in the variable

Page 3 - 33

Question 4: Based on the default settings are variables with missing values rejected? Explain.

Answer 4: Yes, the variables with missing values are rejected provided if the percentage of missing values is greater than 50%

Question 5: Based on the default settings, which numeric variables are assigned as nominal variables?

Answer 5: Based on the default settings, any numeric variable with fewer than 20 unique values is assigned as nominal variable

Page 3 - 33

Question 6: Is there a maximum number of levels for character variables in the default settings? Explain.

Answer 6: Any character variable with more than 20 levels is rejected.

Page 3 - 34

Question 7: How many input variables are selected? How many target variables? How many of the input variables are nominal? Are there any rejected variables? (Compare with your initial answer on the number of variables in this table.)

Answer 7: 20 input variable and 1 input variable is selected

There is only 1 input variable in Nominal and there are no rejected variables. If we compare these numbers with Answer 2 we have 21 variables where 20 are input and 1 target.

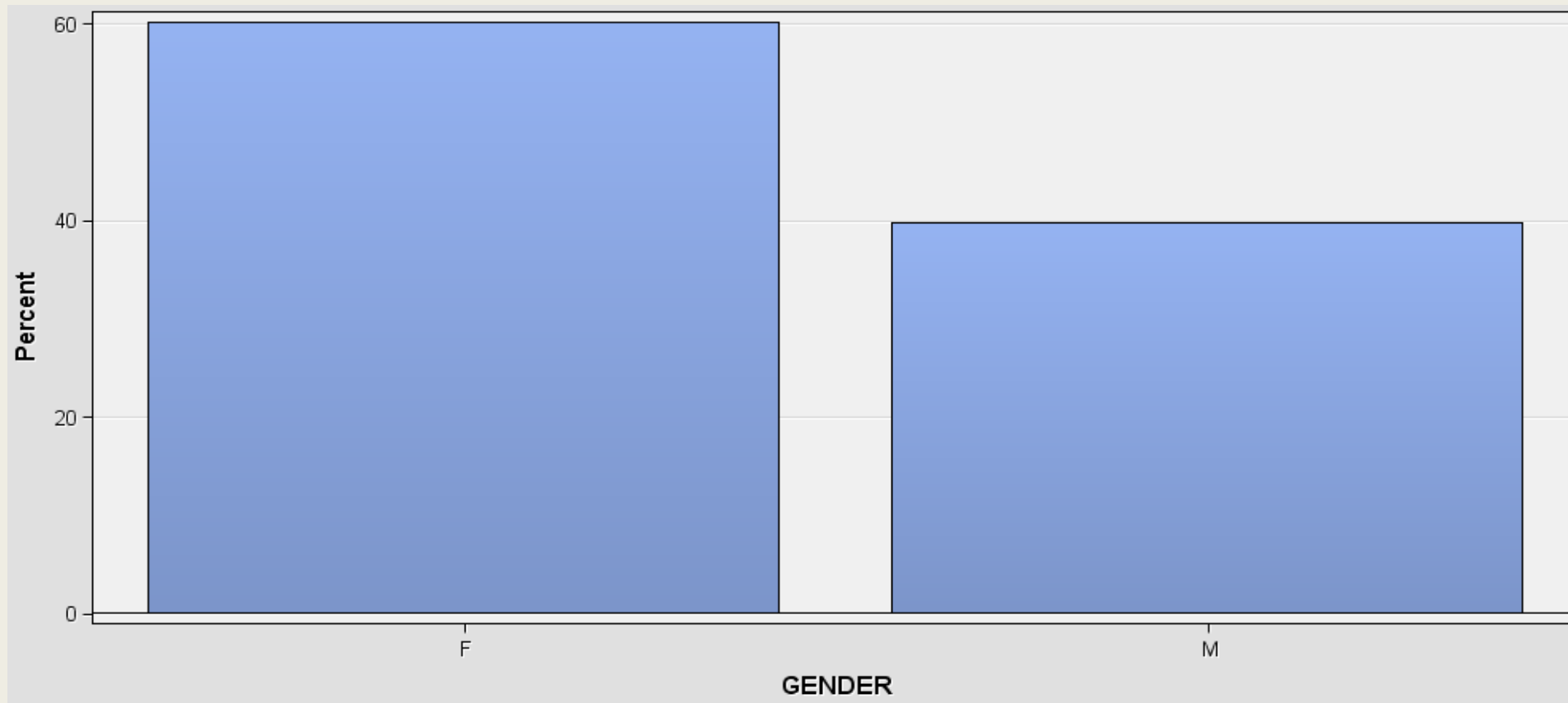
Question 8: What does the value of the max fetch size in sampling for exploration of data depend on? What is a typical value?

Answer 8: The max fetch size in sampling for exploration of data depends on the record length and the number of observations in the dataset. Max fetch size is usually close to 20,000 observations.

Page 3 - 36

Question 9: Create a BAR plot of the GENDER variable. What percentage of the students are female?

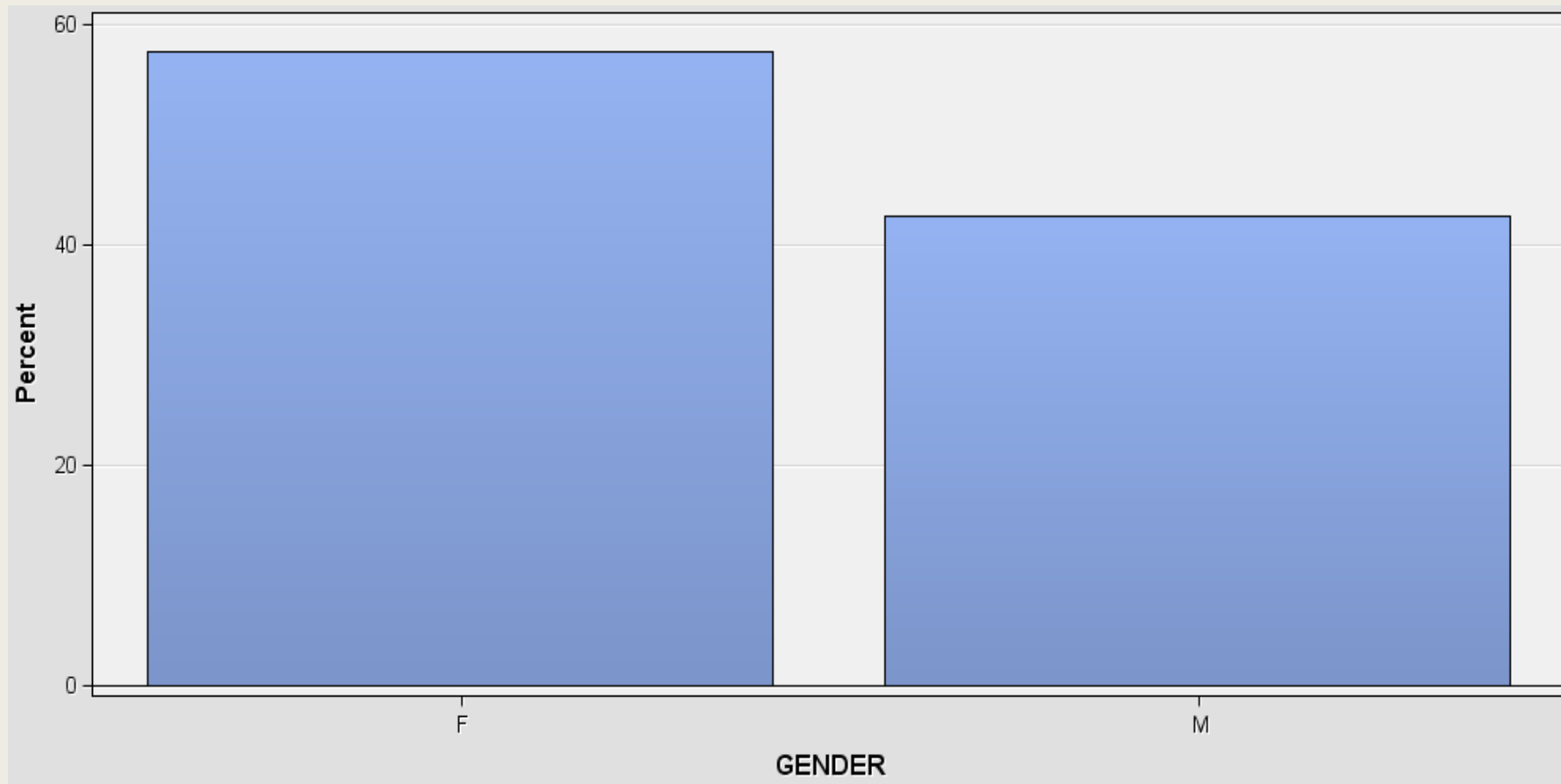
Answer 9: 60.24% of the students are female.



Page 3 - 36

Question 10: Highlight the students with Target=1 on the above GENDER bar plot. Select View > Copy and then paste the plot here.

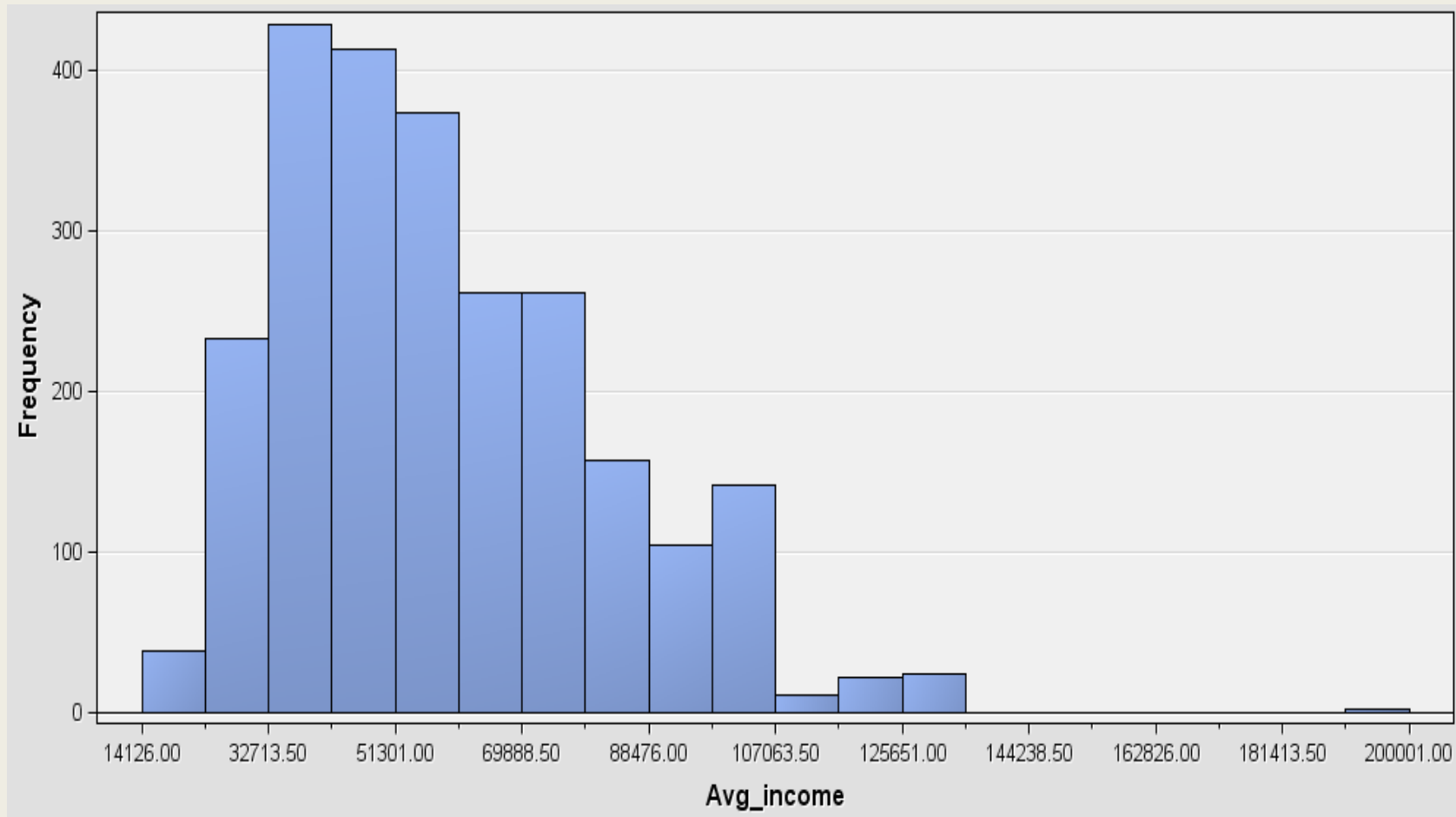
Answer 10: 57.49 percentage of the students are female where Target=1



Page 3 - 37

Question 11: Create a HISTOGRAM plot of the Avg_income variable with 20 bins. Paste here.

Answer 11:



Question 12: Drag and drop the Retention data source on to the diagram. Click on the Retention data source on the left panel, and make sure the creation properties are visible (your username and date) on the properties panel (see following figure). Take a screenshot (PtrScr on your keyboard) and paste here.

Answer 12:

The screenshot shows the SAS Enterprise Miner - Workshop1 interface. The left panel displays the 'Data Sources' tree with 'RETENTION' selected. The bottom panel shows the properties of the selected data source.

Property	Value
Variables	
Decisions	
Role	Raw
Notes	
Library	BAN210
Table	RETENTION
Sample Data Set	
Size Type	
Sample Size	
Type	DATA
No. Obs	2626
No. Cols	21
No. Bytes	332800
Segment	
Created By	u59397413
Create Date	20/1/22 9:33 PM
Modified By	u59397413
Modify Date	20/1/22 9:33 PM
Scope	Local

Name
Data Source name.

u59397413 as u59397413 Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)

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