K. J. SOMAIYA INSTITUTE OF MANAGEMENT STUDIES & RESEARCH
Vidyavihar, Mumbai- 400077

Program: PGPRM –1st year Tri- III (Batch 2017-19)
Subject: Retail Finance and Analytics
(End Term Examination)

Maximum Marks: 50 **Date: 26/03/2018**
Duration: 3 Hours

Instructions

1. **All questions are compulsory**
2. Please write answers in points
3. State relevant examples wherever applicable

QUESTION 1 Case : (20 Marks)

Sonia is a program director for a major health insurance provider. Recently she has been reading in medical journals and other articles, and found a strong emphasis on the influence of weight, gender and cholesterol on the development of coronary heart disease.
The research she's read confirms time after time that there is a connection between these three variables, and while there is little that can be done about one's gender, there are certainly life choices that can be made to alter one's cholesterol and weight.
She begins brainstorming ideas for her company to offer weight and cholesterol management programs to individuals who receive health insurance through her employer. She is concerned with helping those who have suffered heart attacks. She wants to help them improve lifestyle choices, including management of weight and stress, in order to improve their chances of not suffering a second heart attack.

Sonia is wondering if, with the right training data, she can predict the chances of her company's policy holders suffering second heart attacks. The details of the variables are:

- Age: The age in years of the person, rounded to the nearest whole year.
- Marital_Status: The person's current marital status, indicated by a coded number: 0– Single, never married; 1–Married; 2–Divorced; 3–Widowed.
- Gender: The person's gender: 0 for female; 1 for male.
- Weight_Category: The person's weight categorized into one of three levels: 0 for normal weight range; 1 for overweight; and 2 for obese.
- Cholesterol: The person's cholesterol level, as recorded at the time of their treatment for their most recent heart attack
- Stress_Management: A binary attribute indicating whether or not the person has previously attended a stress management course: 0 for no; 1 for yes.
- Trait_Anxiety: A score on a scale of 0 to 100 measuring the level of each person's natural stress levels and abilities to cope with stress. A short time after each person in each of the two data sets had recovered from their first heart

attack, they were administered a standard test of natural anxiety. Their scores are tabulated and recorded in this attribute along five point increments. A score of 0 would indicate that the person never feels anxiety, pressure or stress in any situation, while a score of 100 would indicate that the person lives in a constant state of being overwhelmed and unable to deal with his or her circumstances.

- 2nd_Heart_Attack: This attribute exists only in the training data set. It will be our label, the prediction or target attribute. In the training data set, the attribute is set to 'yes' for individuals who have suffered second heart attacks, and 'no' for those who have not.

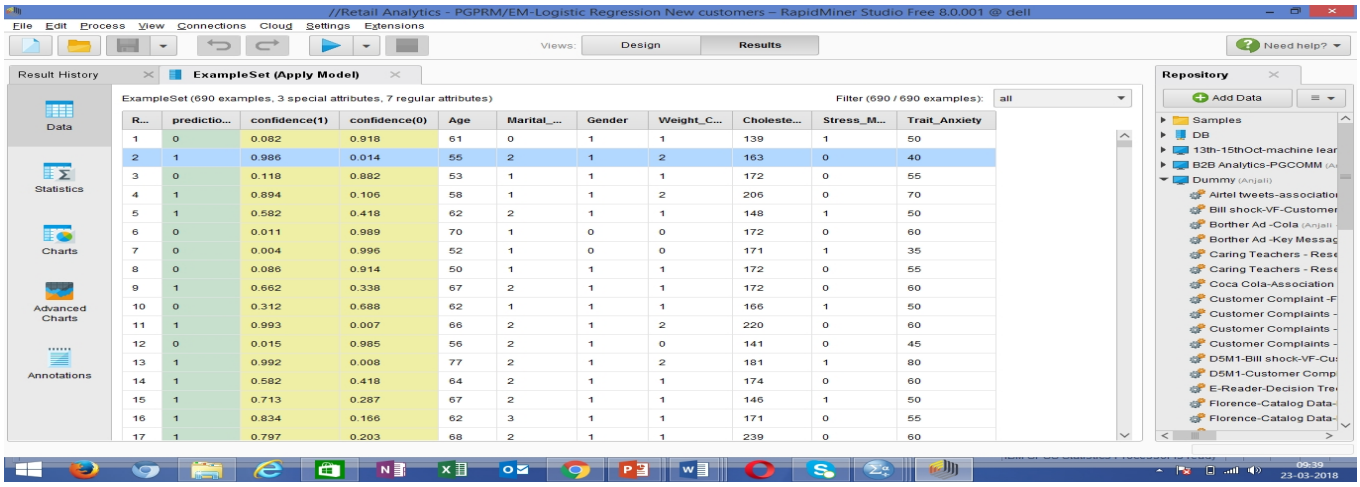
You were given two datasets to work on.

OUTPUTS FOR REFERENCE:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age	.119	.078	2.334	1	.127	1.127
	Marital_Status	1.278	.531	5.803	1	.016	3.591
	Gender	.215	.851	.064	1	.801	1.240
	Weight_Category	4.056	.976	17.255	1	.000	57.725
	Cholesterol	.009	.015	.345	1	.557	1.009
	Stress_Management	-.071	.949	.006	1	.941	.932
	Trait_Anxiety	-.054	.065	.670	1	.413	.948
	Constant	-12.419	4.690	7.011	1	.008	.000

a. Variable(s) entered on step 1: Age, Marital_Status, Gender, Weight_Category, Cholesterol, Stress_Management, Trait_Anxiety.



Questions

i) You were asked to construct hypothesis based on the above data? Which are the independent variables and which is the dependent variable?

ii) Which technique did you use to test your assumption? Linear regression or Logistic Regression? Why?

iii) Which are the important variables in predicting a second heart attack? Explain

iv) Why did you undertake modelling in RapidMiner? How did you validate the Model?

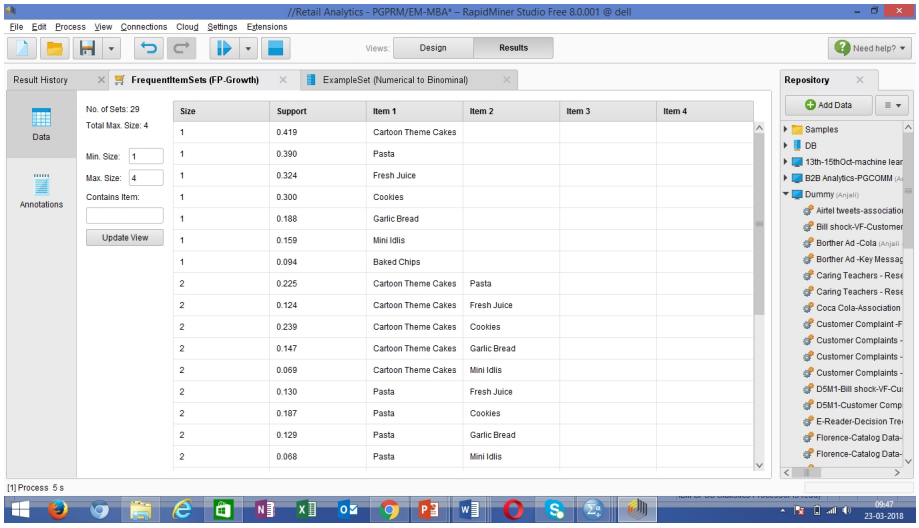
v) How does this Modelling Exercise Help Sonia in identifying and reaching out to potential Heart Attack candidates? Explain by referring to Respondent number 2 in the output given above.

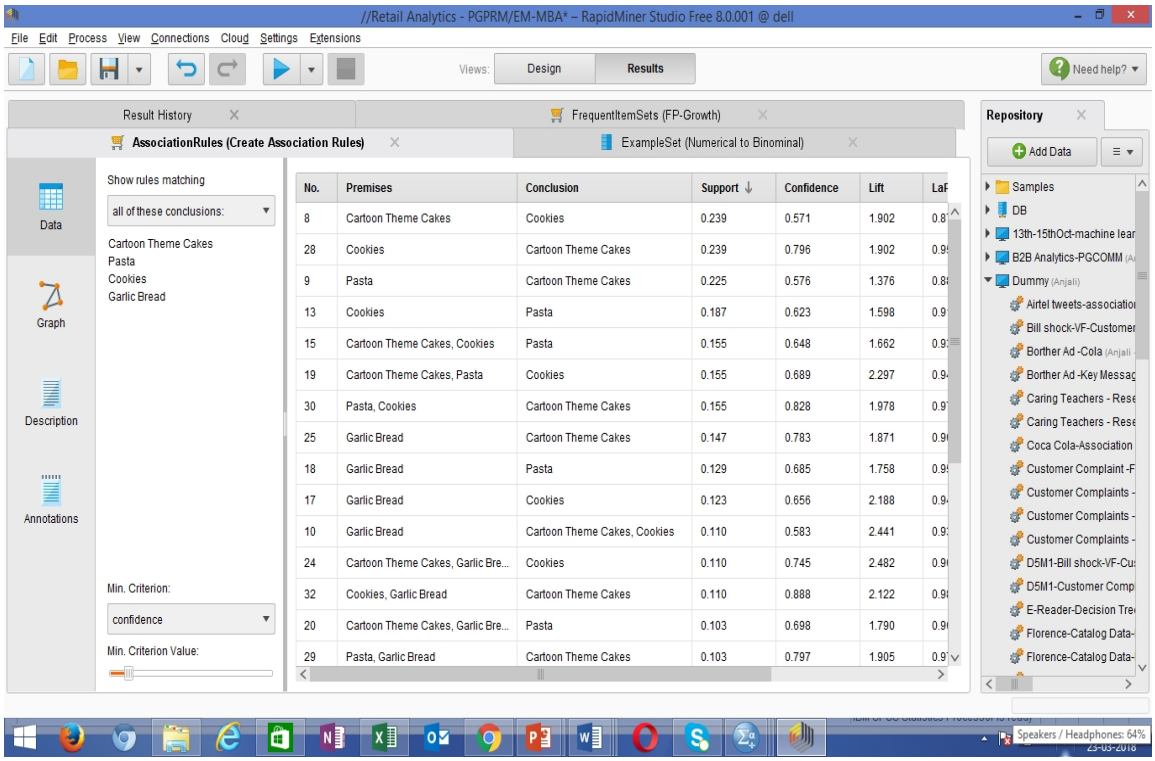
QUESTION 2 **(10 Marks)**

Two partners who specialise in offering birthday menus are quite popular. They make everything fresh and at home and offer a limited menu with emphasis on healthy snacks. The products are freshly made and without any artificial flavours or colours.

The partners have a sense of which products sell well by looking at the sales data however in order to improve sales they are looking at offering some combinations at discounted price.

OUTPUT FOR REFERENCE





Questions:

i) Which products sell well alone? Justify

ii) Which products need to be bundled. Which technique have you used to understand association? Explain the rationale for the combinations you are recommending.

QUESTION 3 **(10 Marks)**

Explain **ANY TWO** with suitable examples

A) Difference between Logistic Regression and Decision Tree

B) Difference between Analytics and Market Research

C) Advantages of Market Basket Analysis

QUESTION 4 EXPLAIN ANY ONE **(10 Marks)**

A) For the Paints Dealer category, what was the approach used to identify the problem existing in the Bangalore territory?

B) How did Linear Regression help in site selection of La Quinta?

********** End of Paper***********