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 Duration: 3 Hours Date: 26/03/2018

(20

Instructions

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

<u>QUESTION 1</u> Case : 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, <u>she can predict the chances of her</u> <u>company's policy holders suffering second heart attacks</u>. 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									
		В	S.E.	Wald	df	Sig.	Exp(B)		
	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		
Chan 13	Weight_Category	4.056	.976	17.255	1	.000	57.725		
Step 1 ^a	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		

Variables in the Equation

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

	-	-	\subset			Views	Des	ign	Results				Need help?
esult History	\times	📕 Examp	leSet (Apply Mo	del) ×									Repository ×
	Exampl	🕒 Add Data 🛛 = 👻											
Data	R	predictio	confidence(1)	confidence(0)	Age	Marital	Gender	Weight_C	Choleste	Stress_M	Trait_Anxiety		Samples
Data	1	0	0.082	0.918	61	0	1	1	139	1	50	^	DB
	2	1	0.986	0.014	55	2	1	2	163	0	40		13th-15thOct-machine lea
Σ	3	0	0.118	0.882	53	1	1	1	172	0	55		 B2B Analytics-PGCOMM (A Dummy (Aniali)
Statistics	4	1	0.894	0.106	58	1	1	2	206	0	70		Airtel tweets-association
Charts	5	1	0.582	0.418	62	2	1	1	148	1	50		Bill shock-VF-Custome
	6	0	0.011	0.989	70	1	0	0	172	0	60		🗬 Borther Ad -Cola (Anja)
	7	0	0.004	0.996	52	1	0	0	171	1	35		Borther Ad -Key Messa Caring Teachers - Res
	8	0	0.086	0.914	50	1	1	1	172	0	55		Caring Teachers - Res
	9	1	0.662	0.338	67	2	1	1	172	0	60		Coca Cola-Association
Advanced	10	0	0.312	0.688	62	1	1	1	166	1	50		🛷 Customer Complaint -
Charts	11	1	0.993	0.007	66	2	1	2	220	0	60		Customer Complaints
	12	0	0.015	0.985	56	2	1	0	141	0	45		Customer Complaints
	12	1	0.992	0.008	77	2	1	2	181	1	80		P5M1-Bill shock-VF-Cu
Annotations	13	1	0.582	0.418	64	2	1	2	181	0	60		D5M1-Customer Com
													🧬 E-Reader-Decision Tre
	15	1	0.713	0.287	67	2	1	1	146	1	50		Florence-Catalog Data
	16	1	0.834	0.166	62	3	1	1	171	0	55		Florence-Catalog Data
	17	1	0.797	0.203	68	2	1	1	239	0	60	~	<

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.

				Views: Design	Results			Need help
esult History	🗙 🛒 Frequer	ntitemSets (FP-Growth)	🛛 🗶 📕 Exam	pleSet (Numerical to Binomina	al) ×			Repository ×
	No. of Sets: 29	Size	Support	Item 1	Item 2	Item 3	Item 4	🕒 Add Data 🛛
Data	Total Max. Size: 4	1	0.419	Cartoon Theme Cakes				^ ▶ 🔚 Samples
Dolla	Min. Size: 1	1	0.390	Pasta				DB
		1	0.324	Fresh Juice				I3th-15thOct-machine I
	Max. Size: 4							B2B Analytics-PGCOMI/ B2B Analytics-PGCOMI/
Annotations	Contains Item:	1	0.300	Cookies				Dummy (Anjali) P Airtel tweets-associa
		1	0.188	Garlic Bread				Bill shock-VF-Custor
	Update View	1	0.159	Mini Idlis				Borther Ad -Cola (An)
		1	0.094	Baked Chips				Porther Ad -Key Mes
			0.225		0			Caring Teachers - R
		2	0.225	Cartoon Theme Cakes	Pasta			Caring Teachers - R
		2	0.124	Cartoon Theme Cakes	Fresh Juice			🧬 Coca Cola-Associati
		2	0.239	Cartoon Theme Cakes	Cookies			🦿 Customer Complain
		2	0.147	Cartoon Theme Cakes	Garlic Bread			Customer Complain
		2	0.069	Cartoon Theme Cakes	Mini Idlis			Customer Complain
								Customer Complain
		2	0.130	Pasta	Fresh Juice			P D5M1-Bill Shock-VP-
		2	0.187	Pasta	Cookies			E-Reader-Decision
		2	0.129	Pasta	Garlic Bread			Plorence-Catalog Da
		2	0.068	Pasta	Mini Idlis			P Florence-Catalog Da
		-						V <

OUTPUT FOR REFERENCE

		•	Views:	Design Results					Need help?		
	Result History $ imes$			🛒 FrequentitemSets (FP-I	Growth) ×				Repository ×		
	🛒 AssociationRules (Create Ass	ociation	Rules) ×	🚦 ExampleSet	(Numerical to Bir	nominal)			🕂 Add Data 🛛 🚍		
	Show rules matching	No.	Premises	Conclusion	Support ↓	Confidence	Lift	LaF	Samples		
Data	all of these conclusions:	8	Cartoon Theme Cakes	Cookies	0.239	0.571	1.902	0.8 ^	DB		
	Cartoon Theme Cakes	Theme Cakes 28 Cookies	Cookies	Cartoon Theme Cakes	0.239	0.796	1.902	0.9!	 I3th-15thOct-machine le B2B Analytics-PGCOMM 		
Craph Description	Pasta Cookies Garlic Bread Min. Criterion:	9	Pasta	Cartoon Theme Cakes	0.225	0.576	1.376	0.8	 Dummy (Anjali) 		
		13	Cookies	Pasta	0.187	0.623	1.598	0.9	🧬 Airtel tweets-associati		
		15	Cartoon Theme Cakes, Cookies	Pasta	0.155	0.648	1.662	0.9:	Bill shock-VF-Cust		
		19	Cartoon Theme Cakes, Pasta	Cookies	0.155	0.689	2.297	0.9	🛷 Borther Ad -Cola (Anja)		
		30	Pasta, Cookies	Cartoon Theme Cakes	0.155	0.828	1.978	0.9	🖨 Caring Teachers - Re		
		25	Garlic Bread	Cartoon Theme Cakes	0.147	0.783	1.871	0.9	🧬 Caring Teachers - Re		
						0.685	1.758	_	Coca Cola-Association		
			Garlic Bread	Pasta	0.129		0.9	Customer Complaint			
		17	Garlic Bread	Cookies	0.123	0.656	2.188	0.9	Customer Complaints		
		10	Garlic Bread	Cartoon Theme Cakes, Cookies	0.110	0.583	2.441	0.9:	😴 Customer Complaints		
		24	Cartoon Theme Cakes, Garlic Bre	Cookies	0.110	0.745	2.482	0.9	P5M1-Bill shock-VF-C		
		32	Cookies, Garlic Bread	Cartoon Theme Cakes	0.110	0.888	2.122	0.9	💣 D5M1-Customer Com		
		20	Cartoon Theme Cakes, Garlic Bre	Pasta	0.103	0.698	1.790	0.9	E-Reader-Decision Tr Florence-Catalog Data		
	Min. Criterion Value:	29	Pasta, Garlic Bread	Cartoon Theme Cakes	0.103	0.797	1.905	0.9' 🗸	Florence-Catalog Data		
	-	<						>			

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 Marks)

(10

(10

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 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?

####