9

Select the wrong option with respect to K-Means clustering:

- [x] The number of clusters is an output of the algorithm.
- [ ] The number of clusters should be identified considering both the robustness and convergence of the algorithm.
- [ ] The number of cases in each cluster should be considered in defining the parameter k.
- [ ] The convergence of the algorithm can be assessed through the iteration history.

10

What is a dummy variable?

- [x] A variable which takes the value of 0 or 1 to indicate the absence or presence of a given category.
- [ ] It is the equivalent of the Discount Rate in the CLV calculation.
- [ ] It is a continuous variables ranging from 0 to 1.
- [ ] It is a categorical variable with more than 2 categories.

11

Take into consideration a Binomial Logistic Regression Model. What does the following formula represent?

$$\frac{\text{deviance(fitted logistic, saturated model)}}{\text{deviance(null model, saturated model)}}$$

- [ ] It represents the R squared of the model
- [ ] It represents the F squared of the model
- [x] None of the other
- [ ] It represents the support of the model

12

Which of the following analytical techniques represent clustering methods?

- ○ K-means, Varimax, Dendrogram
- ○ Partition methods, ANOVA, hierarchical methods
- ○ Varimax, Oblimin, Quartimax
- ● None of the above

13

"churn" is a dataset containing the information of 1,000,000 customers including a unique identifier and a column "Churn" which is 1 if the customer is a churner, 0 otherwise. What is the objective of the following piece of code in R?

```
sample <- sample(c(TRUE, FALSE), nrow(churn), replace=TRUE,
prob=c(0.7,0.3))
t1  <- churn[sample, ]
t2   <- churn[!sample, ]
```

- ○ To create two dummy variables t1 and t2.
- ● To split data from "churn" into training and test sets.
- ○ To create two datasets t1 and t2, with the same variables contained in "churn" and a new variable called "sample".
- ○ To create two datasets t1 and t2, with the same columns and rows contained in "churn", but with different values.

14

With respect to Prediction Accuracy assessment of a model, which of the following is false?

- [x] Precision and sensitivity are synonymes.
- [ ] Recall depends on the number of true positives and false negatives.
- [ ] Sensitivity depends on the number of true positives and false negatives.
- [ ] Precision depends on the number of true positives and false positives.

15

With respect to RFM Analysis, which of the following is true?

- [x] New customers typically skew the models influencing monetary values, since those tend to buy cheap products to test the company's products
- [ ] Recency, Frequency, and Monetary values are usually more than enough to fully predict the purchasing behaviours of all the customers.
- [ ] New customers do not typically influence frequency values.
- [ ] The number of transactions is a good metric of monetary value

16

What is the goal of churn analysis?

- [ ] Understand why customers churn
- [ ] Understand which customers are at risk of churning
- [ ] Investigate how to retain the customers at risk of churning
- [x] All of the above

17

Attribution Modelling:

- ◯ Helps quantifying the impact of marketing inputs on sales
- ◯ Is a top-down approach that aims at giving the credit for the sale to the right channel
- ● May use the so called "removal effect" to understand the effect of each touchpoint
- ◯ Uses aggregate data on marketing investments, contextual elements and conversions

18

In the process of developing a marketing mix model:

- ◯ The first step is to map the possible touchpoints and customer journeys
- ◯ It is important to take into account that the contribution of the digital touchpoints is usually over-estimated by these models
- ◯ It is not possible to use non-parametric models
- ● It is important to include also environmental, macroeconomic and contextual variables that may influence the response