Question 1 : option a

Question 2 : option a

Question 3 : option b

Question 4 : option d

Question 5 : option c

Question 6 : option b

Question 7 : option b

Question 8 : option a

Question 9 : option c

Question 10 answer : normal distribution divides data into normal data, mean , median and mode are all equal , the curve is symmetric at the center , they equally distribute like half is in right and half is in left , the total area approximately would be 1 .

Question 11 answer : mean , median ,and mode , is one of the most common methods of imputing values when dealing with missing data.

Also we can use fillna(), bfill(), ffill()

Question 12 answer : we have to distribute data into two groups and we have to changes in single variable these two versions to two similarly sized audiences and analyze which one performed better over a specific period of time.

Question 13 answer: it’s not good practice and it ignore the relationship between variable such as correlation, mean substitution cannot account for dependence structure among features. It’s used when the missing values are numerical and the distribution of the variable is approximately normal.

Question 14 answer: it’s a data analysis technique, that predict value for unknown data by using other related data, also relationship between one independent variable and one dependent variable.

Question 15 answer: two types of branches:

A : descriptive

B : inferential