Test Bank for Business Analytics 2nd Jaggia

Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

CORRECT ANSWERS ARE LOCATED IN THE 2ND HALF OF THIS DOC.

TRUE/FALSE - Write 'T' if the statement is true and 'F' if the statement is false.

1) The process of retrieving, cleaning, integrating, transforming, and enriching data to support analysis is called data wrangling.
   - ◎ true
   - ◎ false

2) A foreign key (FK) is the only unique identifier in a table structure.
   - ◎ true
   - ◎ false

3) In R, the following represents how to receive results from column 3, row 2 > myData[3,2].
   - ◎ true
   - ◎ false

4) In R, to sort data in descending order, we use a negative parameter in the order function.
   - ◎ true
   - ◎ false

5) Simple mean imputation is the best route for replacing large quantities of missing variables within a data set without distorting the relationship among variables.
   - ◎ true
   - ◎ false

6) To view only a portion of the data that is of interest, subsetting is used.
   - ◎ true
   - ◎ false

7) Converting data from one structure to another is called data transformation.
   - ◎ true
   - ◎ false

8) Subsetting is a technique used to convert numerical values into categorical variables.
   - ◎ true
   - ◎ false

9) A dummy variable takes on a value of 1 or 0 to describe two categories of a variable.
   - ◎ true
   - ◎ false

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

TBEXAM.COM

10) Megan took a phone survey where each question posed had an answer range of unsatisfied to completely satisfied describing her purchase experience. Because the categories are in equal increments, the category can be recoded into a number transforming the category into what is called a category score.
- ◎ true
- ◎ false

11) The strategy of removing observations with missing data is called omission.
- ◎ true
- ◎ false

12) Changing an individual's date of birth to age, combining height and weight to create body mass index, calculating percentages, or converting values to natural logarithms are examples of transforming categorical data.
- ◎ true
- ◎ false

MULTIPLE CHOICE - Choose the one alternative that best completes the statement or answers the question.

13) Which of the following is *NOT* a process of the data management system?
- A) acquire
- B) distribute
- C) store
- D) summarize

TBEXAM.COM

14) Which term represents data items, events, or things stored in a database file?
- A) instance
- B) entity
- C) settings
- D) quantitative

15) Mary in the accounting department has been assigned a specific vehicle as her company car to perform audits. This represents which type of relationship?
- A) 1 : 1
- B) 1 : M
- C) M : N
- D) M : M

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

16) Select, From, and Where keywords are statements used in _____.
A) DBMS
B) XML
C) SQL
D) JAVA

17) The primary purpose of a(n) _____ is to support decision-making and provide a composite view of the organization.
A) data warehouse
B) data mart
C) entity
D) attribute

18) A non-relational database structure that can support the storage of a wide range of data, including structured, semi-structured, and unstructured is called _____.
A) SQL
B) Free Range
C) NoSQL
D) Recreational

TBEXAM.COM

19) Mary has been tasked with reviewing a large data file. She wants to begin by first inspecting the number of values in each cell, both numeric and non-numeric, for any blank entries. The plan is to first find the blank or missing values for first review. Using Excel, what function(s) should she use to complete this task?
A) COUNT
B) COUNTA
C) COUNTIF
D) Both COUNT and COUNTA

20) Molly wants to view observations with missing values in Inventory. However, her data set is quite large. What functions should she use to complete her task in R?
A) > is.na (myData.Inventory)
B) > is.na (myData$Inventory)
C) > which (is.na(myData$Inventory))
D) > which (is.na(myData.Inventory))

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

21) In the presence of outliers in a data set, extremely small or large values, it is preferred to use the _____ instead of the _____ to impute missing variables.
A) median; mean
B) mean; median
C) subset; total
D) average; range

22) In a data set with 25 variables, if 13% of the values, randomly spread across observations, are missing (blank), what is the probable percent of complete and usable observations?
A) 87%
B) 13%
C) 3.08%
D) 4.67%

23) In a data set with 20 variables, if 8% of the values, randomly spread across observations, are missing (blank), what is the probable percent of complete and usable observations?
A) 92%
B) 8%
C) 18.87%
D) 15.29%

TBEXAM.COM

24) Using the simple mean imputation strategy, what value would be placed in the missing observation in x_1?

x_1	x_2
76	22
82	
89	32
	41
86	28

A) No value because excluded
B) 83
C) 81
D) 67

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

25) Using the simple mean imputation strategy, what value would be placed in the missing observation in x_1?

x_1	x_2
76	22
82	
91	32
	41
88	28

A) No value because excluded
B) 84
C) 83
D) 90

26) Using the omission strategy, what value would be placed in the missing observation in x_1?

x_1	x_2
76	22
82	
85	32
	41
82	28

TBEXAM.COM

A) No value because excluded
B) 81
C) 79
D) 65

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Edition Jaggia

Business Analytics Edition 2 by Jaggia

27) Using the omission strategy, what value would be placed in the missing observation in x_1?

x_1	x_2
76	22
82	
91	32
	41
88	28

A) No value because excluded
B) 84
C) 83
D) 90

28) When performing an analysis, one technique is called RFM. Which of the following is not reflective of RFM?
A) recency
B) frequency
C) monetary
D) relevancy

TBEXAM.COM

29) Mark wants to have a better understanding of his client base at the credit union. To do so, he is running a report to show loan amount approval with corresponding credit scores. He realized the data set is quite large and wants to create categories by grouping. To do this, he needs to do all the following ***except***
A) identify the value he wants to transform into smaller groups or bins.
B) remove 20% of the data to create a training set.
C) ensure the data sets are not overlapping.
D) identify how he wants the observations to be labeled in the bin.

30) In Microsoft Excel, Aimee is trying to create a new column called RFM. This column is merging multiple values into one cell. The function to accomplish this is called?
A) TRANSFORM
B) CONCATENATE
C) VARIABLE
D) VLOOKUP

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

31) The function that provides a natural logarithm in Excel is?
- A) INT function
- B) LN function
- C) YEARFRAC function
- D) VLOOKUP function

32) In R, Mary wants to understand the number of days between rain events in Chicago, IL. What function is used to find the number of rain events between today and January 1, 2026?
- A) difftime
- B) as.numeric
- C) diffdate
- D) floor

33) Using R, what is the formula that will allow for the weekday function to display the day of the week for November 15, 2020?
- A) >weekdays< (as.Date(“2020-11-15”)
- B) > format(as.Date(“2020-11-15”), “%d”)
- C) > weekdays(as.Date(“2020-11-15”))
- D) > Sys.Date(“2020-11-15”)

TBEXAM.COM

34) Four observations were binned into one group. In this group, the values are: 40, 45, 38, and 17. What is the average of the group?
- A) 37
- B) 36
- C) 34
- D) 35

35) Four observations were binned into one group. In this group, the values are: 40, 45, 38, and 33. What is the average of the group?
- A) 41
- B) 40
- C) 38
- D) 39

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

36) The following table contains 2 variables with 2 observations. A new variable was created named Sum. This is the sum of the values x_1 and x_2 for each observation. What is the average value of Sum if the chart is completed?

x_1	x_2	Sum
76	28	
60	32	

A) 98
B) 45
C) 104
D) 92

37) The following table contains 2 variables with 2 observations. A new variable was created named Sum. This is the sum of the values x_1 and x_2 for each observation. What is the average value of Sum if the chart is completed?

x_1	x_2	Sum
76	22	
82	32	

A) 106
B) 53
C) 98
D) 114

TBEXAM.COM

38) When too many variables are categorized in an analysis, several potential issues may occur. Which of the following is not one of the issues that may occur?
A) model performance suffers
B) rarely occurring categories may not be captured accurately
C) difficulty in differentiating among observations
D) an increase in the number of categories as the data set becomes larger

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

39) Henry wants to analyze income, but the sheer number of categories in the data's current form will make a clear analysis less meaningful. In Excel with Analytic Solver, how will Henry determine the frequency of each category to transform his data?

A) Income variable is selected and Analytic Solver produces frequency levels for each Income category from most to least frequent.
B) Inspect the frequency of Income category: >table(myData$Income).
C) Income variable is selected and Analytic Solver produces a new category for non-use variables.
D) Apply a limit to the number of categories from the drop-down to a reasonable number.

40) Using R, what function is used to evaluate the categories in the variable to identify the dummy variables?

A) referral
B) if
C) ifelse
D) view

41) In the following table, there are four observations with three variables. Which category is the best fit to be transferred into dummy variables?

Marital Status	Age	Income
Single	24	$45,000
Married	26	$33,000
Single	33	$53,000
Married	28	$59,000

TBEXAM.COM

A) age
B) marital status
C) income
D) none are a good fit for a dummy variable

42) Ann is analyzing a data set that contains two variables, Job Title and 401K. 401K contains the name of the three companies that carry the retirement accounts. It is mandatory to have an account, thus no observation is blank. If 401K was transformed to dummy variables, how many should be created?

A) 2
B) 3
C) 4
D) 1

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

43) Transform the marital status into dummy variables where Single = 1 and Married = 0. How many would have the category score of 0?

Marital Status	Age	Income
Married	24	$45,000
Married	26	$33,000
Married	33	$53,000
Married	28	$59,000
Single	36	$62,000
Married	29	$48,000

A) 1
B) 6
C) 5
D) 0

44) Transform the marital status into dummy variables where Single = 1 and Married = 0. How many would have the category score of 0?

Marital Status	Age	Income
Single	24	$45,000
Married	26	$33,000
Single	33	$53,000
Married	28	$59,000
Married	36	$62,000
Married	29	$48,000

TBEXAM.COM

A) 2
B) 6
C) 4
D) 0

45) Michael is examining a data set and trying to determine which category he can transform into a dummy variable. Of the four variables, Employee Number, Pay Rate, Hire Date, and Sex, which is the best fit to use a dummy variable?

A) employee number
B) pay rate
C) hire date
D) sex

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

46) Marcus wants to include the month of the year in the analysis as categories. How many dummy variables will be needed?
A) 12
B) 11
C) 6
D) 1

47) Kara is reviewing categories where a series of numbers represent the type of loan. She would prefer the actual name of the loan be retained when running her analysis. Using Microsoft Excel, what function will allow Kara to retain the category name instead of recording them in numbers?
A) log function
B) view function
C) IF function
D) head function

48) Using the following table view, Mark wants to create a relationship between the two tables. What will he need to add to establish a relationship?



A) primary key
B) foreign key
C) instances
D) entities

49) Which of the following Excel functions will Ibrahim use to determine how many employees make more than $20 per hour?
A) COUNT
B) COUNTA
C) COUNTIF
D) COUNTIFS

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

50) Which of the following Excel equations will identify the number of married individuals under the age of 30?

Marital Status	Age	Income
Single	24	$45,000
Married	26	$33,000
Single	33	$53,000
Married	28	$59,000
Married	36	$62,000
Married	29	$48,000

A) =COUNT(A2:A7, "=Married", B2:B7,"<30")
B) =COUNTA(A2:A7, "=Married", B2:B7,"<30")
C) =COUNTIF(A2:A7, "=Married", B2:B7,"<30")
D) =COUNTIFS(A2:A7, "=Married", B2:B7,"<30")

51) Using the imputation strategy for categorical values, what value would be placed in the missing observation in *Fav_Color*?

Age	*Fav_Color*
22	Purple
	Blue
32	Purple
41	Red
28	

TBEXAM.COM

A) No value because excluded
B) Blue
C) Purple
D) Red

52) What data preparation technique is Maeve using when she extracts a payroll data set into two separate files, one for hourly employees and one for salary employees?

A) Separating
B) Subsetting
C) Typesetting
D) Wrangling

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

53) Which of the following is NOT an example of categorical data transformation?

A) Category binning
B) Category reduction
C) Category scores
D) Dummy variables

54) The variable x_1 contains three categories ranging from "Poor" to "Good." Convert the category names into category scores into x_2 (i.e., 1 = "Poor", 2 = "OK", and 3 = "Good"). How many observations have a category score of 1?

x_1	x_2
OK	
Good	
Good	
Poor	
OK	
Good	

TBEXAM.COM

A) 0
B) 1
C) 2
D) 3

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

Answer Key

Test name: Chapter 02

1) TRUE
The definition of data wrangling is the process of retrieving, cleansing, integrating, transforming, and enriching data to support subsequent data analysis.

2) FALSE
A primary key (PK) is the main unique identifier in a table structure, whereas a foreign key is a primary key from a related entity.

3) FALSE
> myData[3,2] represents how to view the data in row 3, column 2.

4) FALSE
To place in descending order, we use the *decreasing* parameter in the order function.

5) FALSE
If the number of missing variables is relatively small, then the simple mean process fills in the observations without biasing the results. However, in large quantities, simple mean will distort the data leading to biased results.

TBEXAM.COM

6) TRUE
If only a portion of the data is of interest, such as sales by region, subsetting is the method to use to view only the needed information.

7) TRUE
Data transformation is used to convert data from one format or structure to another.

8) FALSE
9) TRUE
Dummy variables are used often to describe categories. For example, 1 = one result, such as male, and 0 equals female.

10) TRUE
Because the survey response is one of four options, then the categorical variable can be transformed into a category score for running an analysis.

11) TRUE

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

The omission strategy, also called complete-case analysis, recommends that observations with missing values be excluded from the analysis and is appropriate when only a small number of values are missing from the data.

12) FALSE
Each of these examples converts data into numerical data values as opposed to transforming them into categories.

13) D
A data management process is to acquire, organize, store, manipulate, and distribute data. However, summarize is not an option in data management.

14) B
An entity is a generalized category, representing people, places, or things to be stored in a database file.

15) A
In this situation, it is one person, Mary, assigned to a specific vehicle. So 1 : 1 relationship is the best fit for the scenario.

16) C
TBEXAM.COM
Structured Query Language (SQL) is driven by the statements Select, From, and Where, to specify tables, attributes, and criteria to retrieve.

17) A
A data warehouse or enterprise data warehouse is the central repository for data in an organization. The historical and comprehensive view allows management to make strategic decisions for the business.

18) C
A NoSQL offers the flexibility, performance, and scalability to handle high volumes of data. This next phase database will become common to handle the new world of growing data analysis.

19) D
Because the data is both numerical and non-numerical, then both the COUNT and the COUNTA function need to be used to find the blank or missing values quickly.

20) C

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

When a small data set is in use, then is.na function is fine, but when the set is larger, then which and is.na is used to quickly identify the missing values.

21) A
With outliers, the preference is median over mean for the missing values. The reasoning is the swing could impact the variable amounts. Thus, both Analytic Solver and R both have easy imputations to compute this function.

22) C
$(1 - 0.13)^{25} = 0.0308$ or 3.08%.

23) C
$(1 - 0.08)^{20} = 0.1887$ or 18.87%.

24) B
A simple mean is replacing the blank or missing variable with the mean or average of the present variables = (76 + 82 + 89 + 86) ÷ 4 = 83.25 or 83.

25) B
A simple mean is replacing the blank or missing variable with the mean or average of the present variables = (76 + 82 + 91 + 88) ÷ 4 = 84.25 or 84.

TBEXAM.COM

26) A
In the omission strategy, the missing values are excluded from the observation.

27) A
In the omission strategy, the missing values are excluded from the observation.

28) D
RFM is the acronym for recency, frequency, and monetary.

29) B
Binning is taking the entire data set, identifying the value to be binned into smaller groups, ensuring no data overlapping, and labeling the bin accordingly.

30) B
The CONCATENATE function allows for multiple cells to be merged into one cell.

31) B
In Excel, LN function provides a natural logarithm transformation.

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

TBEXAM.COM

32) A
The difftime function is used to determine the number of days between dates.

33) C
The weekdays is the function used to present the result of the day of the week. This example > weekdays(as.Date("2020-11-15")) comes back with the result of "Sunday".

34) D
40 + 45 + 38 + 17 = 140 ÷ 4 = 35.

35) D
40 + 45 + 38 + 33 = 156 ÷ 4 = 39.

36) A
First you need to sum the variables x_1 and x_2 for each row (104 and 92, respectively). (104 + 92) ÷ 2 = 98 is the average of Sum.

37) A
First you need to sum the variables x_1 and x_2 for each row (98 and 114, respectively). (98 + 114) ÷ 2 = 106 is the average of Sum.

TBEXAM.COM

38) D
If the results of a smaller set are applied to a larger data set, then errors may be created. The categories will not increase in numbers as the set becomes larger, just more data will reside under the same amount of categories.

39) A
Analytic Solver will produce results indicating the frequency levels from most to least frequent category for income.

40) C
The ifelse function evaluates the category and determines the assignment of the 1 or 0. For example, if the category is sex, then 1 for male, 0 for female.

41) B
Marital status can be transformed into 1 = married and 0 = single.

42) A

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

Business Analytics Edition 2 by Jaggia

The dummy variables would cover the three possible options for the company being used for the 401K funds. Given *k* categories of a variable, the general rule is to create *k* − 1 dummy variables, using the last category as reference. For 401k we only need to define two dummy variables (*k* − 1 = 3 − 1 = 2). Creating a third dummy would create data redundancy.

43) C
The category score for Marital is 0, thus there are 5 with that status.

44) C
The category score for Marital is 0, thus there are 4 with that status.

45) D
Sex would be the best solution because the options are minimal. Example: 1 = Female, 0 = Male

46) B
If a given *k* categories = 12, then *k* − 1, or 12 − 1 = 11 dummy variables.

47) C
An IF function allows for statements to be crafted to transform numbers into category names.

TBEXAM.COM

48) B
A foreign key is the primary key from another entity used to create a relationship between the tables.

49) C
COUNTIF is used to count the number of cells in a row or column that meet certain criteria such as “more than $20 per hour”. COUNTIFS is used when looking to count the number of cells that meet more than one criteria. COUNT and COUNTA are used to determine the number of cells in a row or column that contain numerical and non-numerical values, respectively.

50) D
COUNTIFS is used when looking to count the number of cells that meet more than one criteria whereas COUNTIF is used to count the number of cells in a row or column that meet just one criterion such as “=married”. COUNT and COUNTA are used to determine the number of cells in a row or column that contain any values, numerical and non-numerical, respectively.

51) C
In the case of categorical variables, the most frequent category is often used as the imputed value.

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Jaggia

TBEXAM.COM

Business Analytics Edition 2 by Jaggia

52) B
The process of extracting portions of a data set that are relevant to the analysis is called subsetting.

53) A
Three common approaches for transforming categorical data are category reduction, dummy variables, and category scores. Binning is the process of converting numerical values into categorical variables by grouping into bins, whereas dummy variables and category scores convert categorical data into numerical values and category reduction simplifies the number of categories.

54) B
When converted, x_2 should be 2, 3, 3, 1, 2, 3, respectively. Since "Poor" only occurs once, the category score 1 appears in only one observation.

TBEXAM.COM

TBEXAM.COM

TBEXAM.COM - THE WORLD'S LARGEST TEST BANK AND SOLUTIONS DATABASE WITH UNBEATABLE RATES
Access Complete Test Bank for Business Analytics 2nd Edition Jaggia