

Student name: _____

1) Which of the following is *NOT* a process of the data management system?

- A) acquire
- B) distribute
- C) store
- D) summarize

2) Which term represents data items, events, or things stored in a database file?

- A) instance
- B) entity
- C) settings
- D) quantitative

3) 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

4) Select, From, and Where keywords are statements used in _____.

- A) DBMS
- B) XML
- C) SQL
- D) JAVA

5) 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

6) A non-relational database structure that can support the storage of a wide ranges of data, including structured, semi-structured, and unstructured is called _____.

- A) SQL
- B) Free Range

- C) NoSQL
- D) Recreational

7) 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

8) Molly wants to view observations with missing values in Inventory. However, her data set is quite large. What

- A) `> is.na (myData.Inventory)`
- B) `> is.na (myData$Inventory)`
- C) `> which (is.na(myData$Inventory))`

functions should she use to complete her task in R?

- D) `> which (is.na(myData.Inventory))`

9) In the presence of outliers in a data set, extremely small or large values, it is preferred to use the _____

- A) median; mean
- B) mean; median

instead of the _____ to impute missing variables.

- C) subset; total
- D) average; range

10) In a data set with 18 variables, if 11% of the values, randomly spread across observations, are missing (blank), what is the probable percent of complete and usable

observations?

- A) 89%
- B) 11%

- C) 12.27%
- D) 7.70%

11) 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%

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

x_1	x_2
72	22
78	
91	32
	41
88	28

- A) 18
- B) 82
- C) 80
- D) 66

13) 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) 17
- B) 84
- C) 83
- D) 90

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

x_1	x_2
76	22
82	
96	32
	41
93	28

- A) No value because excluded
- B) 87
- C) 85
- D) 69

15) 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

16) 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

17) 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

- 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.

create categories by grouping. To do this, he needs to do all the following *except*

- D) identify how he wants the observations to be labeled in the bin.

18) In Analytic Solver, 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

19) The function that provides a natural logarithm in Excel is?

- A) The INT function
- B) The LN function
- C) The YEARFRAC function

- D) The VLOOKUP function

20) 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

- A) `difftime`
- B) `as.numeric`

events between today and January 1, 2019?

- C) `diffdate`
- D) `floor`

21) 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")`

22) Four observations were binned into one group. In this group, the values are: 40, 45, 66, & 33. What is the average of the group?

- A) 48
- B) 47

- C) 45
- D) 46

23) Four observations were binned into one group. In this group, the values are: 40, 45, 38, & 33. What is the average of the group?

- A) 41
- B) 40

- C) 38
- D) 39

x_1	x_2	Sum
76	24	
82	58	

- A) 120
- B) 67
- C) 112
- D) 128

25) 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 was completed?

x_1	x_2	Sum
76	22	
82	32	

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

26) When too many variables are categorized in an analysis, several potential issues may occur. Which of the

- A) model performance suffers.
- B) rarely occurring categories may not be captured accurately.
- C) difficulty in differentiating among observations.

following is not one of the issues that may occur?

- D) an increase in the number of categories as the data set becomes larger.

27) 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.

28) 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

29) 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

A) age

B) marital status

C) income

D) none are a

good fit for a dummy variable.

30) 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

31) Transform the marital status into category scores where Single = 1 and Married = 0. How many would have the

category score of 0?

Marital	Age	Income	Married	36	\$62,00
Married	24	\$45,000			0
Single	26	\$33,000	Married	29	\$48,00
Single	33	\$53,000			0
Single	28	\$59,000			

- A) 1
- B) 6

- C) 3
- D) 0

32) Transform the marital status into category scores 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

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

33) 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,

- A) employee number
- B) pay rate

Hire Date, and Sex, which is the best fit to use a dummy variable?

- C) hire date
- D) sex

34) 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

35) 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 Analytic Solver, 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

36) 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

37) The process of retrieving, cleaning, integrating, transforming, and enriching data to support analysis is called data wrangling.

- ☐ true
- ☐ false

38) A foreign key (FK) is the only unique identifier in a table structure.

- ☐ true
- ☐ false

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

- ☐ true
- ☐ false

40) In R, to sort data in descending order, we use a negative parameter in the order function.

☐ true

☐ false

41) 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

42) To view only a portion of the data that is of interest, subsetting is used.

- ☐ true
- ☐ false

43) Converting data from one structure to another is called data transformation.

- ☐ true
- ☐ false

44) Subsetting is a technique used to convert numerical values into categorical variables.

- ☐ true
- ☐ false

45) A dummy variable takes on a value of 1 or 0 to describe two categories of a variable.

- ☐ true
- ☐ false

46) 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

- ☐ true
- ☐ false

recoded into a number transforming the category into what is called a category score.

Answer Key

Analytics

Test name: Chapter 02 Test Bank Business

1) D

2) B

3) A

4) C

5) A

6) C

7) D

8) C

9) A

10) C

11) C

12) B

13) B

14) A

15) A

16) D

17) B

18) B

19) B

- 20) A
- 21) C
- 22) D
- 23) D
- 24) A
- 25) A
- 26) D
- 27) A
- 28) C
- 29) B
- 30) A
- 31) C
- 32) C
- 33) D
- 34) B
- 35) C
- 36) B
- 37) TRUE
- 38) FALSE
- 39) FALSE
- 40) FALSE

41) FALSE

42) TRUE

43) TRUE

44) FALSE

45) TRUE

46) TRUE