

Question 1

Match the task and purpose to the appropriate Big Data analytics method.

analyzing customer subscription cancellations and correlating them with customer comments and ratings to determine the most common reasons why users cancel subscriptions

analyzing customer ordering history to find popular products within last 6 months

analyzing historical data of temperature to prepare latest weather forecast

reviewing sales data of top 5 competitors to determine the marketing strategy of own product for next quarter

predictive analytics

prescriptive analytics

diagnostic analysis

descriptive analytics

Place the options in the following order:

Analyzing customer subscription cancellations and correlating them with customer comments and ratings to determine the most common reasons why users cancel subscriptions	Diagnostic analysis
Analyzing customer ordering history to find popular products within last 6 months	Descriptive analytics
Analyzing historical data of temperature to prepare latest weather forecast	Predictive analytics
Reviewing sales data of top 5 competitors to determine the marketing strategy of own product for next quarter	Prescriptive analytics

Question 2

Match the type of data analysis with the question it can answer.

Predictive

Prescriptive

Diagnostic

Descriptive

Why did something happen?

What should be done next?

What happened?

What might happen in the future?

Question 3

Which statement is true concerning data analytics?

Prescriptive analytics uses historical data to provide regular reports on events that have already happened.



Predictive analytics uses simulation models and forecasting to suggest what could happen.

Descriptive analytics uses a feedback system to track the outcome of actions taken.

Descriptive analytics attempts to predict what will happen next.

Question 4

A data analyst uses Excel and Tableau to identify patterns and correlations in a dataset to draw conclusions. Which phase of the analytic process is the analyst currently working on?



analyzing the data

preparing the data

presenting the results

investigating the data

Question 5

Which type of variable is used for qualitative values such as gender or eye color?



nominal

ordinal

continuous

discrete

Question 6

What are three resources provided by the Kaggle web site? (Choose three.)



code



community competitions



publicly available datasets

data analysis integration with Excel

data analysis integration with Tableau

API for major social media networks

Question 7

Match the variable with the description.

interval



quantitative values that can specify if a value for a variable exists

discrete



quantitative value from a finite set of values

ordinal



qualitative values in which order matters

ratio



quantitative values from a given range

Question 8

SUM					
=COUNT(D2:D7)					
	B	C	D	E	
1					
2			10		
3			7		
4			5		
5			9		
6			12		
7			3		
8			=COUNT(D2:D7)		

Refer to the exhibit. What is the resulting output from the formula in cell D8?



6

7.6

12

46

Question 9

Which statement is true about the ELT and ETL process?



In the ELT process, the transform step occurs on the stored data as it is used.

In the ELT process, the load step occurs after the transform step.

In the ETL process, the load step occurs before the transform step.

In the ETL process, the extract step occurs after the load step.

Question 10

Which two tasks are performed as part of the transform step of the ETL data process? (Choose two.)



removing the blank records



standardizing formats such as date, time, and location

transferring the data into a database

collecting and recording the data

communicating results to decision-makers

Question 11

An analyst in an online order company is researching products that online customers spend the most time browsing for on the website but do not buy. The result of the analysis will enable the company to release quick discount notices on the website to encourage customers to buy those products. What relevant data is required to do the research?

brands of products

inventory of products

customer order history



streaming customer-viewing data

Question 12

What are two plain-text file types that are compatible with numerous applications and use a standard method of representing data records? (Choose two.)

DOC

PDF



JSON



XML

XLS

Question 13

Match the Microsoft Excel function with its description.

LEN	it is used to create a sum of values based on multiple criteria
COUNTA	it is used to return the number of characters in a cell
CONCATENATE	it is used to combine the data from one or more columns into a single column
SUMIF	it is used to return a value or the reference to a value from within a table or range
INDEX	it is used to convert text to all capital letters (UPPERCASE)
UPPER	it is used to return the number of blank or non-blank cells in a range

Question 14

A data analyst needs to organize sales data for analysis. Which Excel function can order the data by date sold, with the most recent sales listed first?

Sort & Filter
Find & Select
Conditional Formatting
Text to Columns

Question 15

Which option in Microsoft Excel represents a formula with an absolute reference?

=B\$=(C\$1\$-F\$1\$)



=(A1-\$B1)

=(B\$1-B\$3)

=(A\$1-\$B\$1))

Dollar symbols in a cell reference indicate to Microsoft Excel to treat the cell reference as absolute and to always refer to the value of the cell regardless of whether it is moved or where the formula is located. An absolute reference is designated in a formula by the addition of a dollar sign (\$) before the column and row. As an absolute reference was used to refer to Cell B1, the formula will automatically update to wherever the contents of this cell is moved.

Question 16

How can data analysts use the Conditional Formatting tool in Excel to aid in data analysis of bike sales records?



to show only sales from the United States

to list sales order numbers from lowest to highest

to highlight values in the revenue column that are above average

to separate the contents of the product description column into separate columns

By using conditional formatting, a data analyst can highlight cells that contain certain values or that meet certain criteria.

Question 17

A data analyst wants to compare the average life expectancy and GDP for forty countries. Which type of visual representation would best suit this task?

line graph

pie chart

bar chart



scatter plots

Question 18

A learner is analyzing a large volume of data in a Microsoft Excel spreadsheet and wishes to find duplicate data values. The data has been organized in a table where each row has different but related forms of data in each column. Which Microsoft Excel function can be used to do this?



VLOOKUP

IF

ISNA

SUM

Question 19

	A	B	C	D	E	F
1	Good	Evil				
2	A-Bomb	Abominatio	#N/A			
3	Abe Sapien	Abraxas				
4	Abin Sur	Absorbing Man				
5	Adam Monroe	Air-Walker				
6	Adam Strange	Ajax				
7	Agent 13	Alex Mercer				
8	Agent Bob	Alien				
9	Agent Zero	Amazo				
10	Alan Scott	Ammo				

Refer to the exhibit. A learner is analyzing a data spreadsheet in Microsoft Excel and notices a formula function "`=VLOOKUP(B1,A2:A10,1,FALSE)`". When clicking on cell C2, the value displayed in the field in cell C2 is "#N/A". Why is this value displayed?



No value is entered in cell C3.

No duplicate value for the value entered in cell B2 is found.

The formula is entered in the wrong cell.

The cell range is incorrect.

VLOOKUP can also be used to help with data cleaning by finding duplicates. With VLOOKUP you can compare two columns (or lists) and find duplicate values. The formula is written in cell C2 as `=VLOOKUP(B1,A2:A10,1,FALSE)`. This formula compares the value in B1 with the range A2:A10. If there is no duplicate then an error is displayed as "#N/A". If a duplicate was found it would display the name of the duplicate.

Question 20

What is a statistical analysis result that a descriptive statistical analysis will not provide?

calculation of how widely dispersed the data is

calculation of the smallest or largest value

visualization of particular data trends occurring



predictions made about other data sets that are not in the population

Question 21

What is referenced by the FROM statement in an SQL query?



the table containing the data

the column where the data is located

the database to be queried

the row where the data is located

Question 22

Drag the SQL Data Manipulation Language (DML) statements to the correct description.

ORDER BY



used to structure the resulting query in ascending (ASC) order by default

WHERE



returns the selection from the table name specified

FROM



is a conditional statement used to filter your returned results

SELECT



returns the fields specified

Question 23

Movie		
MovieID	Title	Release_date
33	Unforgiven	1992-08-07
35	The Simpsons Movie	2007-07-25
38	Eternal Sunshine of the Spotless Mind	2004-03-19
55	Amores perros	2000-06-16
58	Pirates of the Caibbean: Dead Man's Chest	2006-06-20

Refer to the exhibit. A portion of the Movie table from the Movies database is shown. A data analyst is writing a query that will return the title and release date of movies released after 2000. The analyst further needs the list ordered by the release date. Which query will return the required results?



```
SELECT Title, Release_date
FROM Movie
WHERE Release_date > '2000-12-31'
ORDER BY Release-date
```

```
SELECT Movie
FROM Title, Release_date
WHERE Release_date > '2000-12-31'
ORDER BY Release-date
```

```
SELECT Title, Release_date
FROM Release_date > '2000-12-31'
WHERE Movie
ORDER BY Release-date
```

```
SELECT Release_date > '2000-12-31'
FROM Title, Release_date
WHERE Movie
ORDER BY Release-date
```

User
user_id
content_id
user_name
user_email
user_dob

Question 24

Refer to the exhibit. Which option will correctly display the user names and email addresses of users in this table?

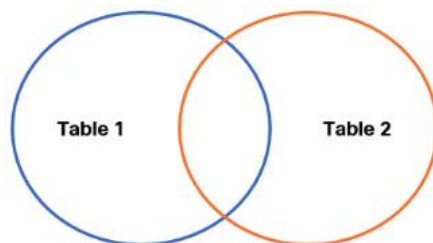


SELECT user_id, user_name, user_email
FROM User

SELECT User
FROM user_id, user_name, user_email

SELECT User
WHERE user_id, user_name, user_email

SELECT user_id, user_name, user_email
WHERE User



Question 25

Refer to the exhibit. Match the JOIN operation with the proper description.

LEFT JOIN



Returns all records from Table 1 and only the matched records from Table 2.

FULL JOIN



Returns all records from Table 2 and the matched records from Table 1.

RIGHT JOIN



Returns only records where the selected fields have matching values in Table 1 and Table 2.

INNER JOIN



Returns all records from Table 1 and Table 2.

Question 26

What is the purpose of the DML language commands in SQL?



to select, update, insert or delete records and values from a table

to create, modify or delete objects in the database schema

to manage updates to the database storage

to change rights and permissions to data in the database

Question 27

What does a field refer to in database schema terminology?

a unique record within the database

a collection data in rows and columns

a specific attribute of an individual record



a collection of attributes of an individual item

A field within the database schema refers to a specific attribute of an individual record. Fields are named by column headings. An individual movie title in the Title column would be a field in the movie table.

Movie table:
MovieID – primary key
Title – name of the movie
Date – date the movie was released
Cost – total cost of production

Review table:
ID – Review ID – primary key
Movie Title – name of the movie
Author Name – name of the reviewer
Comment – review text
Date – date review is posted
Score – number of stars awarded

Question 28

Refer to the exhibit. A data analyst writes a SQL query to extract information from multiple tables in the Movies database. To complete the ON command, what needs to be entered in place of the question mark in this query?

```
SELECT m.Title, m.Date, r.AuthorName, r.Score, r.Comment, r.Date
FROM Review AS r
JOIN Movie AS m
ON m.Title = ?;
```



r.MovieTitle;

r.ID;

r.Comment;

r.Score;

r.Date;

Question 29

Why would you use Tableau in your data analysis?

To perform mathematical calculations using formulas and built-in functions.

To communicate with and query databases.



To create visualizations.

To display data in a tabular format.

Question 30

Which two are benefits of organizing information from multiple datasets into a dashboard? (Choose two)



The information is easily secured.

The information is easily analyzed.

The information is easily visualized.



The information is easily tracked.

Question 31

What Tableau Public feature is used to consolidate multiple views and visualizations?



dashboards

worksheets

screens

objects

Question 32

You are preparing a presentation that needs a visualization showing the relative levels of coffee production in the coffee regions of the world. Which type of visualization in Tableau is well suited for displaying this type of data?

heat map



bubble map

area graph

bar chart

Question 33

What are two methods that ensure confidentiality? (Choose two.)

nonrepudiation



encryption

availability



authentication

integrity

authorization

Question 34

A data analyst conducting a study stops collecting more data once the evidence starts to support the hypothesis. What type of bias has the analyst introduced into the findings?

selection

information



confirmation

interpretation

Question 35

Which statement is true in respect of avoiding bias in data analytics?



avoid selecting only data and methods that support your assumptions

focus on the larger patterns and trends and include outliers in the overall analysis

use a small sample set that supports your hypothesis

don't allow the data to confirm your conclusions

Question 36

Which type of bias in data analysis can be caused by the influence of outliers?

selection



information

confirmation

interpretation

Question 37

Which three features should your data project portfolio contain to ensure that it is considered favorably by employers and recruiters? (Choose three.)



it is organized and easy to navigate



there are links to all your projects on the homepage



each project has a clear outcome or conclusion

it contains many multi-colored charts and graphics that show the results of the data analysis

there are links to software, so non-technical reviewers can replicate your analysis

it uses a layout and presentation that showcases your web design and markup skills

Question 38

What are the three functions of programming languages optimized for data analysis? (Choose three)



statistical analysis



data cleaning



data visualization

data communications across networks

virtualization of processing hardware

fast response time to user input

Question 39

Which data analytic tool can create interactive documents containing executable program code and markdown text?

Tableau

Excel

Kaggle



Jupyter Notebooks

Question 40

Which feature of Python reduces the coding requirements when data tasks such as exploratory data analysis and machine learning are required?



The availability of code libraries such as NumPy and Scikit-learn.

It is free and open source.

It is both an interpreted and object-oriented language.

Python code based on common English words.

It uses dynamic data typing.