



**CPA ADVANCED LEVEL**

**BUSINESS DATA ANALYTICS (PRACTICAL PAPER)**

**THURSDAY: 27 April 2023. Morning Paper.**

**Time allowed: 3 hours.**

Answer **ALL** questions in **SECTION I** and any **THREE (3)** questions in **SECTION II**. **SECTION I** has twenty (20) Multiple Choice Questions each carrying two (2) marks. **SECTION II** has five (5) practical questions each carrying twenty (20) marks.

Under **SECTION II**, you are required to create Ms Excel Worksheets with the name of the entity in each question and input your workings and solutions. You may use the Excel template within the question.

**SECTION I (40 MARKS)**

**Question One**

The acronym “CRISP” stands for:

- A. Cross-industry standard process for data mining
- B. Cross-industry statistical process for data mining
- C. Cross-industry standard practice for data mining
- D. Cross-industry statistical package for data mining

**Question Two**

You are given the following formulas for computing variance and standard deviation of a population and sample:

- 1. =VAR.S()
- 2. =VAR.P()
- 3. =STDEV.P()
- 4. =STDEV.S()

Which of the following statements in **CORRECT**?

- A. Statements 1 and 3 contain the formulas for variance and standard deviation of a population
- B. Statements 2 and 4 contain the formulas for variance and standard deviation of a sample
- C. Statements 1 and 4 contain the formulas for variance and standard deviation of a sample
- D. Statements 2 and 3 contain the formulas for variance and standard deviation of a sample

**Question Three**

Which of the following focuses on the discovery of previously unknown properties on the data?

- A. Data mining
- B. Big data
- C. Data wrangling
- D. Data archiving

**Question Four**

A data analyst would like to determine the number of times revenues have exceeded Sh.10 million over the past 10 years. If the revenues are listed vertically in column A (From Cell A2 to Cell A11) of Excel, which of the following formulas will provide the correct output? :

- A. COUNTIF(A2:A11,“=10”)
- B. COUNTIF(A2:A11,“>10”)
- C. COUNTIF(A2:J11, “>10”)
- D. COUNTIF(“>10”, A2:A11)

### Question Five

“Alteryx” is an example of a \_\_\_\_\_.

- A. Data management tool
- B. Data cleaning tool
- C. Data visualisation tool
- D. Data analysis tool

### Question Six

Which of the following approaches to data collection will require **SIGNIFICANT** data cleansing?

- A. Online administered questionnaire
- B. Email administered questionnaire
- C. Physically administered questionnaire
- D. All of the above

### Question Seven

Which of the following function can be used to find data with unique codes arranged in the top-most row of the dataset in MS Excel.

- A. MATCH
- B. HLOOKUP
- C. VLOOKUP
- D. SET UP

### Question Eight

Which of the following reasons will make an organisation **AVOID** cloud computing as a means for data management?

- A. Data Costs
- B. Data Scalability
- C. Data Integrity
- D. Data Safety

### Question Nine

Correlation analysis is an example of:

- A. Predictive analytics
- B. Prescriptive analytics
- C. Descriptive analytics
- D. Exploratory analytics

### Question Ten

The following are the main examples of data visualisation:

1. Comparison
2. Composition
3. Relationship
4. Distribution

Which of the following summarises the order of the examples from simple to complex?

- A. Distribution, comparison, composition, relationship
- B. Distribution, composition, comparison, relationship
- C. Comparison, composition, distribution, relationship
- D. Relationship, composition, distribution, comparison

### Question Eleven

One or more of the following activities is/are not from Phase 1 of Data Science Life Cycle.

- A. Learning the target domain
- B. Developing initial hypothesis
- C. Visualise initial hypothesis
- D. All of the above

**Question Twelve**

The following statements relate to the 'Vs' of big data:

1. Variability is the evolving nature of data sources
2. Variability is the different types of data structures

Which of the following statements is **CORRECT**?

- A. Only statement 1 is correct
- B. Only statement 2 is correct
- C. Both statements are correct
- D. Both statements are wrong

**Question Thirteen**

The following is considered by many to be the most important language for Data Science:

- A. Ruby
- B. R
- C. Java
- D. MS Excel 2010

**Question Fourteen**

Which of the following choices best represents the correct flow of data models:

- A. Conceptual, logical and physical
- B. Physical, logical and conceptual
- C. Logical, physical and conceptual
- D. None of the above

**Question Fifteen**

Choose the correct keyword for this definition: A graphical representation of a data set:

- A. Data Set
- B. Investigative Cycle
- C. Visualisation
- D. Data Plot

**Question Seven**

The \_\_\_\_\_ data model gives the data analyst the chance to gain an overview of the system to be analysed without being concerned with the details of how it will be analysed.

- A. Conceptual
- B. Logical
- C. Physical
- D. Rational

**Question Seventeen**

A bank collected data on visitors' viewing habits at the bank's website. Which technique can be best used to identify pages commonly viewed during the same visit to the website?

- A. Clustering
- B. Classification
- C. Association rules
- D. Panel analysis

**Question Eighteen**

The following statements apply to data mining:

1. Predictive data mining is a type of analysis that extracts data that may be helpful in determining an outcome.
2. Description data mining is a type of analysis that informs users of that data of a given outcome.

Which of the following is **CORRECT**?

- A. Only statement 1 is true
- B. Only statement 2 is true
- C. Both statements 1 and 2 are true
- D. Both statements are not true

**Question Nineteen**

Which of the following steps is performed by a data scientist after collecting the data?

- A. Data integration
- B. Data replication
- C. Data cleansing
- D. Data manipulation

**Question Twenty**

Which of the following best describes the work of a data architect?

- A. Utilise large sets of data to gather information that meets their company's needs
  - B. Work with businesses to determine the best usage of the information yielded from data
  - C. Develop data solutions that are optimised for performance and design applications
  - D. Evaluate data to reach at logical conclusions
- .....