CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

FRIDAY: 25 April 2025. Morning Paper.   **Time allowed: 3 hours.**

Answer ALL questions in SECTION I and SECTION II and any one (1) question in SECTION III. SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. SECTION II has three (3) practical questions each carrying twenty (20) marks. SECTION III has two (2) practical questions each carrying twenty (20) marks. Answer any one (1) question out of the two (2) questions in section III based on your specialisation.

Under SECTION II and SECTION III, you are required to create Ms Excel Worksheets and input your workings and solutions. You may use the Excel template within the question.

SECTION I (20 MARKS)

Answer ALL questions in this section.

QUESTION ONE
Which one of the following Ms Excel features allows you to select or highlight all cells that have formulas?
A.   Find it
B.   Replace ALL
C.   Go To Find
D.   Go To Special   (1 mark)

ANSWER: D

QUESTION TWO
Assuming cell A1 in Ms Excel is displaying the number "12000.7789". Which one of the following Ms Excel formulas should be used to round this number to the closest integer?
A.   =MROUND(A1,100)
B.   =MROUND(A1,10)
C.   =ROUND(A1,0)
D.   =ROUND(A1,1)   (1 mark)

ANSWER: C

QUESTION THREE
Which one of the following stages of the CRISP framework involves evaluating the results of the analysis?
A.   Interpret
B.   Insert
C.   Investigate
D.   Iterate   (1 mark)

ANSWER: A

QUESTION FOUR

Which one of the following types of data visualisation tools is **BEST** suited for showing the composition of a whole?

A.   Scatter plot
B.   Line graph
C.   Bar chart
D.   Pie chart   (1 mark)

ANSWER: D

QUESTION FIVE

Which one of the following is the first stage in the data lifecycle?

A.   Modeling data requirements
B.   Recording data
C.   Identifying data sources
D.   Removing data   (1 mark)

ANSWER: C

QUESTION SIX

In Ms Excel, which one of the following statements **BEST** describes the purpose of the VLOOKUP function?

A.   It multiplies values in a specified range
B.   It searches for a value in the leftmost column of a table and returns a value in the same row from a specified column
C.   It allows you to enter data into multiple cells at once
D.   It automatically generates charts based on selected data   (1 mark)

ANSWER: B

QUESTION SEVEN

In data analytics, which one of the following data models is typically used to organise data in a structure of tables with rows and columns, where each table has a unique key and relationships are established between tables?

A.   Dimensional Model
B.   Entity-Relationship (ER) Model
C.   Document Model
D.   Key-Value Model   (1 mark)

ANSWER: B

QUESTION EIGHT

Which one of the following **BEST** describes the difference between a "logical data model" and a "conceptual data model"?

A.   A logical data model defines the structure of data without specifying how it is stored while a conceptual data model defines the specific tables and columns
B.   A logical data model is created to outline the business processes whereas a conceptual data model outlines the technical database structure
C.   A conceptual data model is focused on high-level business concepts, while a logical data model is more detailed and defines specific attributes, relationships and data types
D.   A conceptual data model includes specific database platforms and storage details whereas a logical data model only includes high-level entities   (1 mark)

ANSWER: C

QUESTION NINE

One of the processes in the cross-industry standard process for data mining is data preparation. This includes the following tasks:

1.   Handling missing values
2.   Transforming variables
3.   Selecting features
4.   Combining datasets

Which one of the following choices presents the task in their **CORRECT** order?
A.   1, 2, 3, 4
B.   1, 4, 3, 2
C.   1, 3, 2, 4
D.   1, 2, 4, 3   (1 mark)

ANSWER: A

QUESTION TEN
Which one of the following graphs is the **CORRECT** category for distribution category?
A.   Scatter Plot
B.   Heat map
C.   Nested Pie Charts
D.   Box Plot   (1 mark)

ANSWER: D

QUESTION ELEVEN
The following statement relates to diagnostic analytics and its use in accounting:
Diagnostic analytics helps answer "Why did this happen?" by examining past performance and identifying the underlying reasons for certain outcomes. For example, diagnostic analytics can assist a business decide whether to approve or deny loans based on insights into the likelihood of default.

Which one of the following statements is **CORRECT**?
A.   The definition of diagnostic analytics is the only correct part
B.   The example of diagnostic analytics is the only correct part
C.   The entire statement is correct
D.   The entire statement is false   (1 mark)

ANSWER: B

QUESTION TWELVE
In the field of accounting and finance, prescriptive analytics is primarily used to_____.
A.   identify patterns and trends in historical financial data
B.   recommend actions or strategies to achieve desired financial outcomes
C.   summarise and report past financial transactions for decision-making
D.   predict future financial performance based on historical data   (1 mark)

ANSWER: B

QUESTION THIRTEEN
Which one of the following is considered as a key quality of effective data visualisation in data analytics?
A.   Using as many colours and shapes as possible to enhance visual appeal
B.   Displaying data without any labels or legends to keep the focus on the chart itself
C.   Simplifying complex data into clear, easy-to-understand visual formats that convey the message
D.   Including every possible data point even if it overcomplicates the visual representation   (1 mark)

ANSWER: C

QUESTION FOURTEEN
Which one of the following is **NOT** a core principle of data protection under data protection laws?
A.   Data minimisation
B.   Transparency and accountability
C.   Data retention for unlimited periods
D.   Purpose limitation   (1 mark)

ANSWER: C

QUESTION FIFTEEN

Which one of the following is **NOT** part of the 5Vs of Big Data?

A.   Volume
B.   Variety
C.   Velocity
D.   Validation   (1 mark)

ANSWER: D

QUESTION SIXTEEN

In the context of data analytics, which one of the following statements **CORRECTLY** shows the distinction between "data protection" and "data security" in their approach to managing data?

A.   Data protection focuses on ensuring data privacy and compliance with regulations while data security focuses on preventing unauthorised access and attacks
B.   Data protection and data security are identical since both aim to prevent data loss and ensure privacy
C.   Data protection involves encryption techniques while data security involves access control only
D.   Data protection is concerned with physical security measures while data security involves backup strategies
                                                                                                                               (1 mark)

ANSWER: A

QUESTION SEVENTEEN

Which one of the following is the primary purpose of data mapping in data analytics?

A.   To visualise data trends over time
B.   To convert data from one format to another
C.   To organise data into structured tables
D.   To cleanse data by removing errors   (1 mark)

ANSWER: B

QUESTION EIGHTEEN

Which one of the following statements is **TRUE** regarding data portability and data erasure under data protection laws?

A.   Data portability allows individuals to request that their data be erased without restriction while data erasure allows them to request a copy of their data in a structured format
B.   Data portability allows individuals to obtain their data in a commonly used machine-readable format and transfer it to another service provider while data erasure allows individuals to request the deletion of their personal data under certain conditions
C.   Data portability requires companies to retain individuals' data indefinitely while data erasure permits companies to store data without limitations
D.   Data portability and data erasure are not required under data protection laws   (1 mark)

ANSWER: B

QUESTION NINETEEN

Which one of the following is a primary challenge related to the speed of data in data analytics?

A.   Ensuring data security during processing
B.   Handling large volumes of data in real-time
C.   Simplifying data storage architectures
D.   Achieving data consistency across different systems   (1 mark)

ANSWER: B

QUESTION TWENTY

The basic data structure that can be thought of in an SQL table or a spreadsheet data representation is known as _____.

A.   series
B.   list
C.   dataframe
D.   points   (1 mark)

ANSWER: C

CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

FRIDAY: 6 December 2024. Morning Paper.   **Time allowed: 3 hours.**

Answer ALL questions in SECTION I and SECTION II and any one (1) question in SECTION III. SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. SECTION II has three (3) practical questions each carrying twenty (20) marks. SECTION III has two (2) practical questions each carrying twenty (20) marks. Answer any one (1) question out of the two (2) questions in section III based on your specialisation.

Under SECTION II and SECTION III, you are required to create Ms Excel Worksheets and input your workings and solutions. You may use the Excel template within the question.

SECTION I (20 MARKS)

Answer ALL questions in this section.

1.   Which one of the following is **NOT** a phase in the data analytics based on a CRISP lifecycle?
        A.   Data preparation
        B.   Data exploration
        C.   Data manipulation
        D.   Data modelling   (2 marks)

2.   Which one of the following statistical techniques is used to find relationships between variables in data analytics?
        A.   Regression analysis
        B.   Principal component analysis
        C.   Cluster analysis
        D.   Association rule learning   (2 marks)

3.   Which one of the following terms refers to the process of cleaning and preparing data for analysis?
        A.   Data mining
        B.   Data wrangling
        C.   Data extraction
        D.   Data modelling   (2 marks)

4.   Which one of the following tools is commonly used for data summarisation in Ms Excel?
        A.   Pivot tables
        B.   VLOOKUP
        C.   Power query
        D.   Scatter plots   (2 marks)

5. Which one of the following analyses is used to assess a project's vulnerability to different variables?
   A.   Break-even analysis
   B.   Sensitivity analysis
   C.   Scenario analysis
   D.   Ratio analysis   (2 marks)

6. Which one of the following tools is commonly used for cloud-based data management in business analytics?
   A.   SQL
   B.   Oracle
   C.   AWS
   D.   Excel   (2 marks)

7. A data analyst makes the following comments about functions in Ms Excel:
   1.   To determine the variance of a list of 50 numbers we use =VAR.P(A1:A50)
   2.   To determine the standard deviation of 50 numbers we use = VAR.P(A1:A50)
   3.   To determine the variance of a list of 50 numbers we use = VAR.S(A1:A10)

   Which one of the following statements is **TRUE**?
   A.   Only Statement 1 is correct
   B.   Only Statement 2 is correct
   C.   Only Statement 3 is correct
   D.   All the statements are correct   (2 marks)

8. Which one of the following choices provides the **CORRECT** order of data models?
   A.   Physical, logical and conceptual
   B.   Physical, conceptual and logical
   C.   Conceptual, logical and physical
   D.   Conceptual, physical and logical   (2 marks)

9. One of the processes in the cross-industry standard process for data mining is data understanding. This involves the following tasks:
   1.   Collect data
   2.   Verify data quality
   3.   Explore data
   4.   Describe data

   Which one of the following choices present the task in the **CORRECT** order?
   A.   1, 2, 3, 4
   B.   1, 4, 3, 2
   C.   1, 3, 2, 4
   D.   1, 2, 4, 3   (2 marks)

10. Susan Keni makes the following statement about data modelling and data evaluation in the cross-industry standard process for data mining:

   "The modelling phase focuses on technical model assessment, the evaluation phase looks more broadly at which model best meets the business and what to do next."

Which one of the following choices is **CORRECT**?
A. Only the statement on modelling phase is correct
B. Only the statement on evaluation phase is correct
C. The statement is correct for both modelling and evaluation phases
D. The statement is not correct for both modelling and evaluation phases (2 marks)

11. Which one of the following 'V's of big data means the ability to forecast data?
A. Veracity
B. Vaticination
C. Viscosity
D. Volatility (2 marks)

12. What does "visualisation of data relationships" refer to in business data analytics?
A. Comparing different variables
B. Presenting data in tabular form
C. Creating a dashboard
D. Highlighting associations between data points (2 marks)

13. One of the challenges of data analytics is obtaining real-time data from multiple sources. Which one of the following data systems would be **MOST** effective in assisting to resolve this challenge?
A. A data system that enables report building
B. A comprehensive and centralised data system
C. A data system that enables data integration
D. A data system that collects, organises and automatically alerts on trends (2 marks)

14. Victoria Morosa has been asked to pair the following accounting and finance tasks with the type of data analytics technique:

Task	Technique
Analysing ratios	Descriptive
Computing ratios	Prescriptive
Preparing forecast financial statements	Diagnostic
Preparing a budget	Predictive

Which one of the following tasks is **CORRECTLY** paired with the data analytics technique?
A. Analysing ratios
B. Computing ratios
C. Preparing forecast financial statements
D. Preparing a budget (2 marks)

15. Which one of the following is **NOT** a type of diagnostic activity in accounting and finance?
A. Estimating customer defaults in a certain period
B. Estimating the probability of stockouts
C. Computing the tax payable for a period
D. Identifying the appropriate stock levels (2 marks)

16. The following statements are qualities of good data visualisation:
    (i)   The visualisation should accurately represent the data and its trends
    (ii)   The reader should know what action to take after viewing your visualisation
    (iii)   Your visualisation should be easy to understand
    (iv)   Your message shouldn't take long to resonate

    Which of the statements above describes the quality of empowering?
    A.   Statement (i)
    B.   Statement (ii)
    C.   Statement (iii)
    D.   Statement (iv)   (2 marks)

17. The following statement applies to one of the data protection principles:
    "Any processing of personal data should be explained to the individual concerned on how the data is being collected, used, consulted, and the extent to which the data will be processed". This is the principle of_____.
    A.   lawfulness
    B.   fairness
    C.   transparency
    D.   integrity   (2 marks)

18. The following statement highlights a principle of data protection:
    "Personal data should be processed in a manner that ensures appropriate security and confidentiality of the personal data, including protection against unauthorised or unlawful access to or use of personal data and the equipment used for the processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

    Which one of the following statements is **CORRECT**?
    A.   The statement only applies to data integrity
    B.   The statement only applies to data confidentiality
    C.   The statement applies to both integrity and confidentiality
    D.   The statement applies to neither integrity nor confidentiality   (2 marks)

19. Under the data protection laws, a data subject has the right to receive their data in a structured, commonly used, machine-readable format, to transmit this ported data to another data controller or processor or to request the transfer to another data controller or processor where possible.
    This is referred to as the right to data _____.
    A.   access
    B.   portability
    C.   data erasure
    D.   data rectification   (2 marks)

20. Data mapping is an essential aspect of managing the risks associated with data and it is recommended for data protection. Which one of the following is **NOT** a component of data mapping?
    A.   Data modelling
    B.   Data migration
    C.   Data integration
    D.   Data transformation   (2 marks)

CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

FRIDAY: 23 August 2024. Morning Paper.   **Time allowed: 3 hours.**

Answer ALL questions in SECTION I and SECTION II and any one (1) question in SECTION III. SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. SECTION II has three (3) practical questions each carrying twenty (20) marks. SECTION III has two (2) practical questions each carrying twenty (20) marks. Answer any one (1) question out of the two (2) questions in section III based on your specialisation.

Under SECTION II and SECTION III, you are required to create Ms Excel Worksheets and input your workings and solutions. You may use the Excel template within the question.

SECTION I (20 MARKS)

Answer ALL questions in this section.

QUESTION ONE
Which one of the following is the **BEST** reason why an organisation may avoid cloud computing as a means of data management?
A.   Data costs
B.   Data scalability
C.   Data integrity
D.   Data safety

QUESTION TWO
The following are the main examples of data visualisation:
(i)   Comparison
(ii)   Composition
(iii)   Relationship
(iv)   Distribution

Which one 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 THREE
The following activities apply to data visualisation:
(i)   The format
(ii)   The content
(iii)   The structure
(iv)   The purpose

Which one of the following is the correct order of activities?
A.   (iii), (iv), (ii) and (i)
B.   (iv), (iii), (i) and (ii)
C.   (iii), (ii), (iv) and (i)
D.   (iv), (ii), (iii) and (i)

QUESTION FOUR
A method of storing data within a system that facilitates collocation of data in various schemata and structural forms is known as _____.
A.   data visualisation
B.   data lake
C.   big data management
D.   deep analytics

QUESTION FIVE
The Kenya Data Protection Act requires that a Data Protection Impact Assessment (DPIA) be carried out where a data processing activity creates high risk to the rights and freedoms of a data subject. Which one of the following is an example of high-risk data activity?
A.   Customer profiling
B.   Customer scoring
C.   Customer inquiry
D.   Customer feedback

QUESTION SIX
The following statements are on data security and data privacy:

(i)   Data security is the process of protecting corporate data and preventing data loss through unauthorised access.
(ii)   Data privacy aims to ensure that data is not shared within and outside the organisation.

Which one of the following statements is **TRUE**?
A.   Only statement (i) is correct
B.   Only statement (ii) is correct
C.   Both statements are correct
D.   None of the statements is correct

QUESTION SEVEN
Even when intentions are good, the outcome of data analysis can cause inadvertent harm to individuals or groups of people. In data ethics, this concept is referred to as _____.
A.   diverse impact
B.   disparate impact
C.   dissimilar impact
D.   contrasting impact

QUESTION EIGHT
According to cross-industry standard process for data mining, data modelling involves:
A.   Obtaining data and information from different sources, processing and storing for future reference
B.   Fixing or removing incorrect, corrupted, incorrectly formatted data and information
C.   Collecting data and information about business requirements from stakeholders and end users
D.   Creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures

QUESTION NINE
Which one of the following is **NOT** an array function in Ms Excel?
A.   MMULT
B.   TRANSPOSE
C.   VLOOKUP
D.   FREQUENCY

QUESTION TEN

Miriam Wambui is the Chief Finance Officer of Zelta Company Limited. She is using data analytics in estimating future risks that the company is facing and also cash budgeting with scenario analysis.

Which one of the following choices is **CORRECT**?

By carrying out risk management and cash budgeting, the category is classified as _____.
A.   predictive analytics for risk management and cash budgeting
B.   predictive analytics for risk management and prescriptive analytics for cash budgeting
C.   predictive analytics for cash budgeting and prescriptive analytics for risk management
D.   prescriptive analytics for risk management and cash budgeting

QUESTION ELEVEN

Which one of the following is **NOT** a strategy the data analyst will use in handling missing data to enable data analysis?
A.   Drop observations that have missing values
B.   Estimate missing value using other observations
C.   Assign an integer that will enable the values be processed as null values
D.   Highlighting the observation with a tag such as not available

QUESTION TWELVE

Which one of the following keyboard shortcuts is used to display the "function Arguments" dialog box in Ms Excel?
A.   Ctrl + F2
B.   Ctrl + Shift + F3
C.   Ctrl + Shift + A
D.   Ctrl + Shift + #

QUESTION THIRTEEN

Which one of the following software is **NOT** used in data visualisation?
A.   Python
B.   Power BI
C.   Tableau
D.   Datawrapper

QUESTION FOURTEEN

Which one of the following is the **BEST** way to address the issue of transparency in data analytics?
A.   Use proprietary algorithms for analysis
B.   Only use numerical data for analysis
C.   Document the analysis process and results
D.   Use a single data source for analysis

QUESTION FIFTEEN

Which one of the following Ms Excel functions is used to count the number of cells in a range that meet a certain criteria?
A.   COUNTA
B.   COUNTIF
C.   COUNT
D.   COUNTBLANK

QUESTION SIXTEEN

Which one of the following software will a data analyst use to clean data?
A.   Data stage
B.   SQL
C.   Tableu
D.   Microstrategy

QUESTION SEVENTEEN

Which one of the following Ms Excel features would allow the creation of a drop-down list of values that users can select from?
A.   Trace precedents
B.   Data validation
C.   Data table
D.   Text to column

QUESTION EIGHTEEN

Which one of the following cloud computing platforms is known for its data management and analytics capabilities?

A. Azure
B. AWS
C. IBM Cloud
D. Google Cloud

QUESTION NINETEEN

In accounting, which type of analytics is used to detect anomalies or irregularities in financial transactions?

A. Predictive analytics
B. Descriptive analytics
C. Diagnostic analytics
D. Prescriptive analytics

QUESTION TWENTY

Which Ms Excel tool can Alex Mutua use to figure out the interest rate at which he would make monthly payments of Sh.50,000 for the Sh.2 million bank loan he took for 5 years?

A. NPV
B. Vlookup
C. Scenario manager
D. Goal seek

CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

FRIDAY: 26 April 2024. Morning Paper.   **Time allowed: 3 hours.**

Answer ALL questions in SECTION I and SECTION II. SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. SECTION II has three (3) practical questions each carrying twenty (20) marks. SECTION III has two (2) practical questions each carrying twenty (20) marks. Answer any one (1) question out of the two (2) questions in section III.

Under SECTION II and SECTION III, 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 (20 MARKS)

Answer ALL questions in this section.

QUESTION ONE
Emma Atieno would like to determine the number of times revenues have exceeded Sh.12 million over the past 15 years. If the revenues are listed vertically in column A (From Cell A2 to Cell A16) of Excel, which of the following formulas will provide the correct output?
A.   COUNTIF(A2:A11,"=12")
B.   COUNTIF(A2:A11,">12")
C.   COUNTIF(A2:J11, ">12")
D.   COUNTIF(">12", A2:A11)

ANSWER: B

QUESTION TWO
A data analyst would like to determine the total revenue generated by a business in the first quarter of the year provided in Excel. Which of the following syntax in Excel will provide the correct output?
A.   =SUM([sum_range] range, criteria)
B.   = SUMIF(range, criteria, [sum_range])
C.   =SUMIFS([sum_range] range, criteria)
D.   = SUMIF(Criteria, range, [sum_range)

ANSWER: B

QUESTION THREE
Which one of the following data analytical tools will require the **LEAST** programming activity by a data analyst?
A.   Tableau
B.   Python
C.   R
D.   Power BI

ANSWER: A

QUESTION FOUR
Which one of the following "Vs" of big data means the ability to generate more copies of data?
A.   Value
B.   Validity
C.   Veracity
D.   Vagueness

ANSWER: C

QUESTION FIVE
The following statements are made by two data analysts regarding the "Vs" of big data:
Jane: "Visibility: Often the only way customers interact with models"
John: "Visualisation: Data science provides images into complex big data problems"

Which one of the following statements is **TRUE**?
A.   Only Jane is correct
B.   Only John is correct
C.   Both Jane and John are correct
D.   Both Jane and John are not correct

ANSWER: D

QUESTION SIX
Which one of the following approaches to data collection is more **EFFICIENT** in data cleaning?
A.   Online administered questionnaire
B.   Email administered questionnaire
C.   Physical administered questionnaire
D.   Interview

ANSWER: A

QUESTION SEVEN
Which one of the following ethical issues in data analytics pertains to the potential misuse of personal data for profiling or targeting?
A.   Data security
B.   Data cleaning
C.   Data privacy
D.   Data aggregation

ANSWER: C

QUESTION EIGHT
Which one of the following data visualisation charts presents comparison over many periods of time?
A.   Scatter graph
B.   Bubble chart
C.   Circular area chart
D.   Line chart

ANSWER: C

QUESTION NINE
The following activities are involved in the data understanding stage in the cross-industry standard process for data mining, **EXCEPT**_____.
A.   Checking for data completeness
B.   Checking for errors in data
C.   Checking for missing values in data
D.   Identifying appropriate data modelling

ANSWER: D

QUESTION TEN

You are given the following statements about data mining:
1.   Descriptive data mining is a type of analysis that extracts data that may help determine an outcome.
2.   Prescriptive data mining is a type of analysis that informs users of data of a given outcome.

Which one 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 not correct

ANSWER: D

QUESTION ELEVEN

The following statements are made about the qualities of good data visualisation:
1.   The visualisation should accurately represent the data and its trends
2.   The reader should know what action to take after viewing your visualisation
3.   Your visualisation should be easy to understand
4.   Your message should not take long to resonate

Which one of the following statements describes the quality of empowering?
A.   Statement 1
B.   Statement 2
C.   Statement 3
D.   Statement 4

ANSWER: B

QUESTION TWELVE

A scatter graph will likely be classified as a type of _____ data visualisation.
A.   Distribution
B.   Relationship
C.   Comparison
D.   Composition

ANSWER: B

QUESTION THIRTEEN

Which one of the following statements is the **CORRECT** definition of "data transformation"?
A.   The process that removes data that does not belong in a dataset
B.   The process of checking the integrity, accuracy and structure of data
C.   The process of converting data from one format or structure to another
D.   The process of creating a visual representation of data elements

ANSWER: C

QUESTION FOURTEEN

The following are various approaches to data risk management:
1.   Data risk sharing
2.   Data risk avoidance
3.   Data risk transfer
4.   Data risk acceptance
5.   Data risk reduction

Which one of the following answers provides the order of the **MOST** to the **LEAST** effective data risk management strategy?
A.   1, 2, 3, 4, 5
B.   4, 3, 1, 2, 5
C.   2, 1, 3, 5 ,4
D.   2, 3, 1, 5, 4

ANSWER: D

QUESTION FIFTEEN

The following are examples of data protection principles:
1.   Only collect what is sufficient for your specified purpose
2.   Process data bearing in mind the safety of individuals
3.   Explain why you are collecting data

Which statement among the three is the principle of retention?
A.   Statement 1
B.   Statement 2
C.   Statement 3
D.   None of the statements

ANSWER: D

QUESTION SIXTEEN

Limo Limited obtains data from clients to provide business consulting services. Recently, a client provided data for analysis, but a dispute arose and the process to provide consultancy to the client did not proceed. The client has requested Limo Limited to delete all the data and confirm that this is done. Limo Limited should_____.
A.   Delete the data
B.   Keep the data as per the requirements of the data protection laws
C.   Provide the data to a legal custodian for safe custody
D.   Seek legal advice before deleting the data

ANSWER: A

QUESTION SEVENTEEN

When analysing corporate financial ratios in Excel, which function will return the highest or lowest value in a dataset?
A.   MAX() and MIN()
B.   LARGE() and SMALL()
C.   TOP() and BOTTOM()
D.   HIGH() and LOW()

ANSWER: B

QUESTION EIGHTEEN

Based on the principles in the Unified Ethical Frame for Big Data Analytics (Abrahams, 2015), the attribute of data analysis being transparent and inclusive is provided by which of the following?
A.   Beneficial
B.   Fairness
C.   Progressive
D.   Respectful

ANSWER: D

QUESTION NINETEEN

What is the Excel function used to calculate the covariance between two data sets stored in cells B1:B10 and C1:C10?
A.   =SD(B1:B10, C1:C10)
B.   =VAR(B1:B10, C1:C10)
C.   =COVAR(B1:B10, C1:C10)
D.   =CORREL(B1:B10, C1:C10)

ANSWER: C

QUESTION TWENTY

In the context of data analytics, what does skepticism primarily refer to?
A.   A bias in favour of accepting data findings at face value
B.   The need to question and critically evaluate data sources and results
C.   The over-reliance on machine learning algorithms
D.   A positive attitude toward data-driven decision-making

ANSWER: B

RATES OF TAX (for employment income including wife's employment, self-employment and professional income).

Year of income 2023.

Assume the following rates of tax applied throughout the year of income 2023:

Monthly taxable pay (Sh.)		Annual taxable pay (Sh.)		Rate of tax % in each Sh.
1 - 24,000		1 - 288,000		10%
24,001 - 32,333		288,001 - 388,000		25%
32,334 - 500,000		388,001 - 6,000,000		30%
500,001 - 800,000		6,000,001 - 9,600,000		32.5%
Excess over 800,000		Excess over 9,600,000		35%

Personal relief Sh.2,400 per month (Sh.28,800 per annum).

Investment allowance: Capital expenditure incurred on:	Rate of investment allowance	Residual value (25% per year on equal instalments)	Prescribed benefit rates of motor vehicles provided by employer (i) Saloons, Hatch Backs and Estates	Monthly rates (Sh.)	Annual rates (Sh.)
(a) Buildings:			Up to 1200 cc	3,600	43,200
• Hotel building	50% in the first year of use	25%	1201 - 1500 cc	4,200	50,400
• Building used for manufacture	50% in the first year of use	25%	1501 - 1750 cc	5,800	69,600
• Hospital buildings	50% in the first year of use	25%	1751 - 2000 cc	7,200	86,400
• Petroleum or gas storage facilities	50% in the first year of use	25%	2001 - 3000 cc	8,600	103,200
• Educational/hostels building	10% per year on straight line basis		Over - 3000 cc	14,400	172,800
• Commercial building	10% per year on straight line basis				
(b) Machinery:			(ii) Pick-ups, Panel Vans (unconverted)		
• Machinery used for manufacture	50% in the first year of use	25%	Up to - 1750 cc	3,600	43,200
• Hospital equipment	50% in the first year of use	25%	Over - 1750 cc	4,200	50,400
• Ships or aircraft	50% in the first year of use	25%			
• Motor vehicles and heavy earthmoving equipment	25% per year on straight line basis		(iii) Land Rovers/Cruisers	7,200	86,400
• Computer software, calculators, copiers and duplicating machines	25% per year on straight line basis				
• Furniture and fittings	10% per year on straight line basis				
• Telecommunication equipment	10% per year on straight line basis				
• Film equipment by a local producer	25% per year on straight line basis				
• Machinery used to undertake operations under prospecting rights and exploration under mining rights	50% in the first year of use	25%			
• Other machinery	10% per year on straight line basis				
(c) Purchase/acquisition of right to use fibre optic cable by telecommunication operation	10% per year on straight line basis				
(d) Farm works	50% in the first year of use	25%			

Commissioner's prescribed benefit rates: Services	Monthly rates (Sh.)	Annual rates (Sh.)
(i) Electricity (Communal or from a generator)	1,500	18,000
(ii) Water (Communal or from a borehole)	500	6,000
Agriculture employees: Reduced rates of benefits		
(i) Water	200	2,400
(ii) Electricity	900	10,800

………………..…………………………………………………………….

CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

FRIDAY: 8 December 2023. Morning Paper.   **Time allowed: 3 hours.**

SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. **SECTION II** has two (2) practical questions each carrying thirty (30) marks. **SECTION III** has three (3) practical questions each carrying twenty (20) marks. Answer ALL questions in **SECTION I** and **SECTION II**. Answer any one (1) question out of the three (3) questions in **section III**.

Under **SECTION II** and **SECTION III**, 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.

RATES OF TAX (Including wife's employment, self-employment and professional income rates of tax).

Year of income 2022.

Monthly taxable pay (Sh.)			Annual taxable pay (Sh.)			Rate of tax % in each Sh.
1	-	24,000	1	-	288,000	10%
24,001	-	32,333	288,001	-	388,000	25%
Excess over	-	32,333	Excess over	-	388,000	30%

Personal relief Sh.2,400 per month (Sh.28,800 per annum).

Investment allowance: Capital expenditure incurred on:	Rate of investment allowance	Residual value (25% per year on equal instalments)	Prescribed benefit rates of motor vehicles provided by employer (i) Saloons, Hatch Backs and Estates		
				Monthly rates (Sh.)	Annual rates (Sh.)
(a) Buildings:					
• Hotel building	50% in the first year of use	25%	Up to 1200 cc	3,600	43,200
• Building used for manufacture	50% in the first year of use	25%	1201 - 1500 cc	4,200	50,400
• Hospital buildings	50% in the first year of use	25%	1501 - 1750 cc	5,800	69,600
• Petroleum or gas storage facilities	50% in the first year of use	25%	1751 - 2000 cc	7,200	86,400
• Educational/hostels building	10% per year on straight line basis		2001 - 3000 cc	8,600	103,200
• Commercial building	10% per year on straight line basis		Over - 3000 cc	14,400	172,800
(b) Machinery:			(ii) Pick-ups, Panel Vans (unconverted)		
• Machinery used for manufacture	50% in the first year of use	25%	Up to - 1750 cc	3,600	43,200
• Hospital equipment	50% in the first year of use	25%	Over - 1750 cc	4,200	50,400
• Ships or aircraft	50% in the first year of use	25%			
• Motor vehicles and heavy earth moving equipment	25% per year on straight line basis				
• Computer software, calculators, copiers and duplicating machines	25% per year on straight line basis				
• Furniture and fittings	10% per year on straight line basis				
• Telecommunication equipment	10% per year on straight line basis				
• Film equipment by a local producer	25% per year on straight line basis				
• Machinery used to undertake operations under prospecting rights and exploration under mining rights	50% in the first year of use	25%			
• Other machinery	10% per year on straight line basis				
(c) Purchase/acquisition of right to use fibre optic cable by telecommunication operation	10% per year on straight line basis		(iii) Land Rovers/Cruisers	7,200	86,400
(d) Farm works	50% in the first year of use	25%			

Commissioner's prescribed benefit rates: Services	Monthly rates (Sh.)	Annual rates (Sh.)
(i) Electricity (Communal or from a generator)	1,500	18,000
(ii) Water (Communal or from a borehole)	500	6,000
Agriculture employees: Reduced rates of benefits		
(i) Water	200	2,400
(ii) Electricity	900	10,800

Answer ALL questions in this section.

QUESTION ONE
Which of the following formulas in Excel will provide a subtotal of variables provided in a list of vertically listed cells?
A.   =sum (A1:H20)
B.   =sum (A1:A20)
C.   =sum (A1;H20)
D.   =sum (A;A20)

ANSWER: B

QUESTION TWO
Which of the following shortcut keys will enable an analyst to view the formula in a cell?
A.   Ctrl + ;
B.   Ctrl +:
C.   Ctrl + '
D.   Ctrl +]

ANSWER: C

QUESTION THREE
Which of the following is the **CORRECT** one for the HLOOKUP function in excel?
A.   = HLOOKUP(lookup_value, [range_lookup], table_array, column_index_num)
B.   = HLOOKUP(lookup_value, table_array, row_index_num, [range_lookup])
C.   = HLOOKUP(lookup_value, column_index_num, [range_lookup], table_array)
D.   = HLOOKUP([range_lookup], lookup_value, table_array, column_index_num)

ANSWER: B

QUESTION FOUR
The following statements relate to the 'Vs' of big data:
1.   **Volume:** It refers to the size of big data.
2.   **Veracity:** It refers to the assurance of integrity or credibility of the data.
3.   **Value:** How useful the data is in decision making.

Which of the following statements is **CORRECT**?
A.   Only statement 1 is correct
B.   Only statement 2 is correct
C.   Only Statement 3 is correct
D.   All the statements are correct

ANSWER: A

QUESTION FIVE
Which of the following 'Vs' of data provides for data as being authentic, trustworthy and original?
A.   Veracity
B.   Value
C.   Variability
D.   Variety

ANSWER: A

QUESTION SIX
The following are activities that are included in data modelling:
1.   Physical
2.   Conceptual
3.   Logical
4.   Implementation

Which of the following represents the **CORRECT** steps in data modelling?
A.   1,2,3,4
B.   2,3,4,1
C.   2,3,1,4
D.   3,1,2,4

ANSWER: C

QUESTION SEVEN
Which of the following is **NOT** a data cleaning technique?
A.   Remove duplicates.
B.   Remove irrelevant data.
C.   Standardise capitalisation.
D.   Convert data from text to numeric

ANSWER: D

QUESTION EIGHT
Part of the Cross-Industry Standard Process for data mining requires business understanding. Which of the following is the **CORRECT** description of business understanding?
A.   Defining specific goals and requirements for data mining
B.   Obtain an overview of the available data and their quality
C.   Data preparation for further validation
D.   Data is reviewed in the context of the business

ANSWER: A

QUESTION NINE
Which of the following questions is **NOT** relevant in the understanding of data as per the Cross-Industry Standard Process for data mining?
A.   Is the data complete (does it cover all the cases required)?
B.   Does the data contain errors?
C.   Are there missing values in the data?
D.   Is the data current?

ANSWER: D

QUESTION TEN
Data analytics can be analysed at four levels:
1.   What happened?
2.   What could happen?
3.   What should happen?
4.   Why it happened?

Which of the following is the alternative description of Level 4?
A.   Predictive
B.   Descriptive
C.   Diagnostic
D.   Prescriptive

ANSWER: C

QUESTION ELEVEN
Mhasibu is using data analytics to make management decisions. Some decisions and types of analytics are as follows:
1.   Prepare variance analysis report (Descriptive Analytics)
2.   Prepare flexible budgets evaluating various scenarios (Prescriptive)

Which of the following statement(s) is (are) **TRUE**?
A.   Statement One only
B.   Statement Two only
C.   Both Statements are True
D.   None of the statements are True

ANSWER: D

QUESTION TWELVE
Which of the following software does **NOT** focus only on data visualisation tools?
A.   Google Charts
B.   Python
C.   Grafana
D.   Tableu

ANSWER: B

QUESTION THIRTEEN
A pie chart will likely be classified as what type of data visualisation?
A.   Relationship
B.   Distribution
C.   Composition
D.   Comparison

ANSWER: B

QUESTION FOURTEEN
Under data ethics, when algorithms are used to predict past decisions by human beings, this may lead to which unethical behaviour?
A.   Human bias
B.   Discrimination
C.   Intrusion to privacy
D.   Lack of transparency

ANSWER: A

QUESTION FIFTEEN
While data protection deals solely with personal data, data security is directed to protect all sorts of information, regardless of whether it's digital or analog, comes from the natural person or not.

Which of the above is **CORRECT**?
A.   The definition of data protection only
B.   The definition of data security only
C.   Both definitions
D.   None of the definitions

ANSWER: C

QUESTION SIXTEEN
Which one of the following concepts does **NOT** apply in data protection of digital platforms?
A.   Encryption of personal data.
B.   Anonymization of personal data.
C.   Pseudonymization of personal data.
D.   Biometric access

ANSWER: D

QUESTION SEVENTEEN
Which of the following will **NOT** lead to a challenge in data privacy in an organisation?
A.   Data Privacy regulations
B.   Poor data visibility
C.   Too much data
D.   Too many devices

ANSWER: A

QUESTION EIGHTEEN
Companies must do everything they can to use customer data ethically, preserve customers' privacy, keep them informed of how their data is being used, provide consumers with options to opt in or out, and walk the fine line between serving up relevant, targeted content and overstepping boundaries.
This definition will likely be?
A.   Data confidentiality
B.   Data Security
C.   Data Privacy
D.   Data protection

ANSWER: A

QUESTION NINETEEN
Which of the following is **NOT** a benefit of data analytics?
A.   Customer retention
B.   Strategic analysis
C.   Profitability
D.   Risk identification

ANSWER: C

QUESTION TWENTY
Despite the appreciation and adoption of big data and analytics by organisations, which of the following is a risk of poor use of big data and analytics?
A.   Cost optimisation
B.   Staff dependence
C.   Infrastructure upgrade
D.   Poor heuristics

ANSWER: D

……………..……………………………………………………………………..

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
.................…..….

CPA POST ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

THURSDAY: 24 August 2023. Morning Paper.   **Time allowed: 3 hours.**

Answer ALL questions in SECTION I and SECTION II. Answer any one (1) question out of the three (3) questions in section III. SECTION I has twenty (20) Multiple Choice Questions each carrying one (1) mark. SECTION II has two (2) practical questions each carrying thirty (30) marks. SECTION III has three (3) practical questions each carrying twenty (20) marks.

Under SECTION II and SECTION III, 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.

RATES OF TAX (Including wife's employment, self-employment and professional income rates of tax).

Year of income 2022.

Monthly taxable pay (Sh.)			Annual taxable pay (Sh.)			Rate of tax % in each Sh.
1	-	24,000	1	-	288,000	10%
24,001	-	32,333	288,001	-	388,000	25%
Excess over	-	32,333	Excess over	-	388,000	30%

Personal relief Sh.2,400 per month (Sh.28,800 per annum).

Investment allowance: Capital expenditure incurred on:	Rate of investment allowance	Residual value (25% per year on equal instalments)	Prescribed benefit rates of motor vehicles provided by employer (i) Saloons, Hatch Backs and Estates	Monthly rates (Sh.)	Annual rates (Sh.)
(a) Buildings:					
• Hotel building	50% in the first year of use	25%	Up to 1200 cc	3,600	43,200
• Building used for manufacture	50% in the first year of use	25%	1201 - 1500 cc	4,200	50,400
• Hospital buildings	50% in the first year of use	25%	1501 - 1750 cc	5,800	69,600
• Petroleum or gas storage facilities	50% in the first year of use	25%	1751 - 2000 cc	7,200	86,400
• Educational/hostels building	10% per year on straight line basis		2001 - 3000 cc	8,600	103,200
• Commercial building	10% per year on straight line basis		Over - 3000 cc	14,400	172,800
(b) Machinery:			(ii) Pick-ups, Panel Vans (unconverted)		
• Machinery used for manufacture	50% in the first year of use	25%	Up to - 1750 cc	3,600	43,200
• Hospital equipment	50% in the first year of use	25%	Over - 1750 cc	4,200	50,400
• Ships or aircraft	50% in the first year of use	25%			
• Motor vehicles and heavy earth moving equipment	25% per year on straight line basis				
• Computer software, calculators, copiers and duplicating machines	25% per year on straight line basis				
• Furniture and fittings	10% per year on straight line basis				
• Telecommunication equipment	10% per year on straight line basis				
• Film equipment by a local producer	25% per year on straight line basis				
• Machinery used to undertake operations under prospecting rights and exploration under mining rights	50% in the first year of use	25%			
• Other machinery	10% per year on straight line basis				
(c) Purchase/acquisition of right to use fibre optic cable by telecommunication operation	10% per year on straight line basis		(iii) Land Rovers/Cruisers	7,200	86,400
(d) Farm works	50% in the first year of use	25%			

Commissioner's prescribed benefit rates: Services	Monthly rates (Sh.)	Annual rates (Sh.)
(i) Electricity (Communal or from a generator)	1,500	18,000
(ii) Water (Communal or from a borehole)	500	6,000
Agriculture employees: Reduced rates of benefits		
(i) Water	200	2,400
(ii) Electricity	900	10,800

Answer ALL questions in this section.

QUESTION ONE

How fast data is processed and accessed is represented by which of the following V's of Big Data?

A.   Velocity
B.   Validity
C.   Virality
D.   Veracity

ANSWER: A

QUESTION TWO

Data modelling is a process that helps one to understand the structures, form, and relationships of information. It requires one to first think of the model, then identify the model details and then how the model will be implemented. Which of the following is the correct name of data model implementation?

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

ANSWER: C

QUESTION THREE

The following activities are included in the Cross-Industry Standard Process for data mining:

1.   Data Preparation
2.   Data Understanding
3.   Data Deployment
4.   Business Understanding
5.   Data Evaluation
6.   Data modelling

Which of the following is the proper sequence of these activities?

A.   2,1,3,5,6,4
B.   1,3,5,6,4,2
C.   1,6,5,2,3,4
D.   4,2,1,6,5,3

ANSWER: D

QUESTION FOUR

Which of the following is **NOT** an activity in the modeling stage in the Cross-Industry Standard Process for Data Mining?

A.   Build
B.   Select
C.   Design
D.   Evaluate

ANSWER: D

QUESTION FIVE

For data visualisation, a bar chart will likely be classified as a_____.

A.   Distribution
B.   Relationship
C.   Comparison
D.   Composition

ANSWER: C

QUESTION SIX

A data analyst would like to present data in the form of a distribution. He will likely use all the following **EXCEPT**?

A.   Line chart
B.   Histogram
C.   Scatter graph
D.   Pie Chart

ANSWER: D

QUESTION SEVEN

Data Ethics are the norms of behaviour that promote appropriate judgements and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimising risks to individuals and society, and maximising the public good. All the following promote data ethics **EXCEPT**?

A. Complying with the law
B. Promoting transparency
C. Respecting privacy and confidentiality
D. Fairness

ANSWER: A

QUESTION EIGHT

Which of the following is **NOT** a concept of data integrity?

A. Data availability
B. Data currency
C. Data accuracy
D. Data validity

ANSWER: A

QUESTION NINE

With the high volume of data available for businesses, collecting meaningful data is a big challenge. To enable handling large volumes of data, data collection tools are critical. Which of the following will likely be a data collection software?

A. Excel
B. Python
C. R
D. Zoho

ANSWER: D

QUESTION TEN

A data analyst makes the following comments about functions in excel:

1. To determine the variance of a list of 10 numbers we use =VAR.P(A1:A10)
2. To determine the standard deviation of 10 numbers we use = VAR.P(A1:A10)
3. To determine the variance of a list of 10 numbers we use = VAR.S(A1:A10)

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

A. Only statement 1 is correct
B. Only statement 2 is correct
C. Only statement 3 is correct
D. All the statements are correct

ANSWER: C

QUESTION ELEVEN

The following are the various phases of data models:

1. Physical
2. Logical
3. Conceptual

Which of the following represents the correct order of the models?

A. 1,2,3
B. 2,3,1
C. 3,1,2
D. 3,2,1

ANSWER: D

QUESTION TWELVE

Which of the following techniques is **LEAST** effective in obtaining data for easy storage and analysis?

A. Interviews
B. Self-administered email questionnaire
C. Online questionnaire such as on google forms
D. Recorded interviews

ANSWER: A

QUESTION THIRTEEN

According to cross-industry standard process for data mining, data preparation involves:

A.   Obtaining data and information from different sources, processing and storing for future reference.
B.   Fixing or removing incorrect, corrupted, incorrectly formatted data and information.
C.   Collecting data and information about business requirements from stakeholders and end users.
D.   Creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures.

ANSWER: B

QUESTION FOURTEEN

The following statements are associated with big data:

(i)   Big data is a collection of data that is huge in volume, yet growing exponentially with time.
(ii)   Big data that can be stored, accessed and processed in the form of fixed format is termed as structured data.
(iii) Big data that is heterogeneous and contains a combination of simple text files and images is semi structured.

Which of the following choices is correct?
A.   Statements (i) and (ii) are correct
B.   Statements (i) and (iii) are correct
C.   Statements (ii) and (iii) are correct
D.   All the statements are correct

ANSWER: A

QUESTION FIFTEEN

Which of the following is **NOT** the correct definition of one of the 'V's of Big Data?
A.   Velocity: The speed at which big data is collected and processed
B.   Value: The usefulness of the information obtained from analysing big data
C.   Veracity: Multifactor, unstructured and dynamic nature of big data
D.   Volume: The size of big data

ANSWER: C

QUESTION SIXTEEN

The following are the different categories of data analytics:
1.   Diagnostic
2.   Predictive
3.   Descriptive
4.   Prescriptive

Which of the following is the correct order of value?
A.   3,1,4,2
B.   3,4,2,1
C.   2,3,4,1
D.   1,3,2,4

ANSWER: A

QUESTION SEVENTEEN

David Mongare is the audit manager of Mkaguzi Auditors. He is carrying out analytical reviews at two stages: at planning and at substantive testing. Which type of analytics relates to the two stages?
A.   Predictive analytics for planning and diagnostic analytics for substantive testing
B.   Diagnostic analytics for planning and predictive analytics for substantive testing
C.   Descriptive analytics for planning and diagnostic analytics for substantive testing
D.   Predictive analytics for planning and descriptive analytics for substantive testing

ANSWER: D

QUESTION EIGHTEEN

Which of the following is **NOT** a strategy that the data analyst will use in handling missing data to enable data analysis:
A.   Drop observations that have missing values
B.   Estimate the missing value using other observations
C.   Assign an integer that will enable the values to be processed as null values
D.   Highlighting the observation with a tag

ANSWER: D

QUESTION NINETEEN

You are given the following categories of data visualisation:

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

Which of the following choices represents the order from simple to complex in terms of analysing data for visualisation?

A. 5,3,2,4,1
B. 4,3,2,1,5
C. 2,3,1,4,5
D. 1,5,4,3,2

ANSWER: A

QUESTION TWENTY

Which of the following data visualisation charts will likely present more than two variables over time effectively?

A. Scatter Graph
B. Bubble Chart
C. Column Chart
D. Line Chart

ANSWER: D

……………..………………………………………………………………..

CPA ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

WEDNESDAY: 14 December 2022. 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 - TOTAL 40 MARKS

Question One
Data models evolve from conceptual (that is; a quick, high-level view of the business requirement) to logical (where the entities involved are expanded and include more detail) and finally the physical data model, which can be implemented with a specific database provider (like Oracle and SQL Server).

Which of the following choices is correct?
A.   The entire statement is true
B.   Only the statement on conceptual data model is true
C.   Only the statement on logical data model is true
D.   Only the statement on physical data model is true   (2 marks)

Question Two
Which of the following would be more appropriate to replace the question mark in the following diagram?
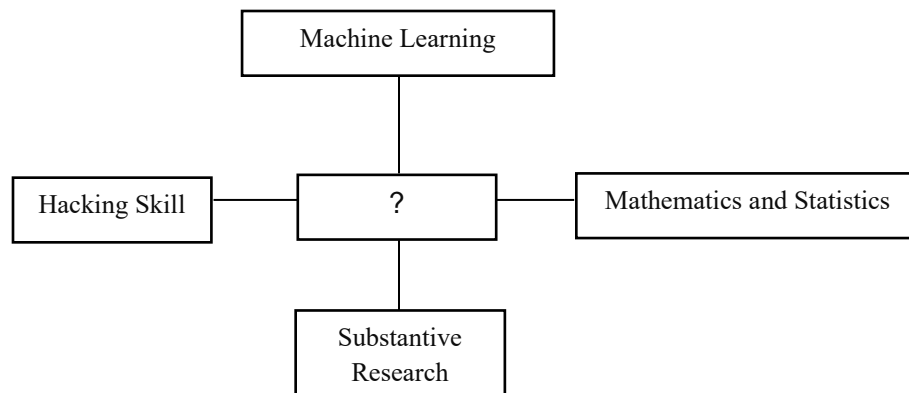
A.   Data analysis
B.   Data science
C.   Statistical inference
D.   Predictive modelling   (2 marks)

Question Three
According to cross-industry standard process for data mining, data modelling involves:
A.   Obtaining data and information from different sources, processing and storing for future reference
B.   Fixing or removing incorrect, corrupted, incorrectly formatted data and information
C.   Collecting data and information about business requirements from stakeholders and end users
D.   Creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures   (2 marks)

Question Four
Read the statements below and answer the question that follows:

(i)   Data mining relates to turning raw data into useful information.

(ii)   Data mining using built-in algorithms should guarantee a result.

Which of the following choices apply?
A.   The two statements are true
B.   Only the first statement is true
C.   Only the second statement is true
D.   None of the statements is true   (2 marks)

Question Five
Which of the following 'Vs' of data describes data as multifactor, unstructured and dynamic?
A.   Veracity
B.   Value
C.   Variability
D.   Variety
                                                                                                          (2 marks)

Question Six
Which of the following is an example of discrete data?
A.   Number of children
B.   Height of children
C.   Behaviour of children
D.   Test scores of children   (2 marks)

Question Seven
Ms Dare Mongare is the Chief Finance Officer of Modern Company Limited. She is using data analytics in estimating future risks that the company is facing and also cash budgeting, with scenario analysis.
By carrying out risk management and cash budgeting, she is applying:
A.   Predictive analytics for risk management and cash budgeting
B.   Predictive analytics for risk management and prescriptive analytics for cash budgeting
C.   Predictive analytics for cash budgeting and prescriptive analytics for risk Management
D.   Prescriptive analytics for risk management and cash budgeting   (2 marks)

Question Eight
Based on the principles in the Unified Ethical Frame for Big Data Analytics, which of the following applies to the principle of Fairness?
A.   Thinking through the potential impacts of our data use on all interested parties
B.   Sustainability of the data over time
C.   Transparency and inclusivity of the data
D.   Data benefiting both the business and customers   (2 marks)

Question Nine
Which of the following applications will likely NOT be used for cloud computing?
A.   Azure
B.   AWS
C.   SQL
D.   Alibaba Cl   (2 marks)

Question Ten

In data science, a relationship between two entities is called _____.

A.   Binary

B.   Quartenary

C.   Unary

D.   None of the above   (2 marks)

Question Eleven

Which of the following data visualisation tools will likely present a relationship of more than two variables effectively?

A.   Scatter Graph

B.   Bubble Chart

C.   Column Chart

D.   Line Chart   (2 marks)

Question Twelve

_____ graph displays information as a series of data points connected by straight line segments.

A.   Line

B.   Bar

C.   Scatter

D.   Histogram   (2 marks)

Question Thirteen

A data breach occurs when the organisation's data suffers a security incident resulting in a breach of confidentiality, availability or integrity. According to the applicable data protection law, in the case of a data breach, the organisation should:

A.   Notify the supervisory authority within 48 hours of the incident whether or not it poses a risk to the organisation and affected individuals

B.   Notify the supervisory authority within 48 hours of the incident, only if it poses a risk to the organisation and affected individuals

C.   Notify the supervisory authority within 72 hours of the incident whether or not it poses a risk to the organisation and affected individuals

D.   Notify the supervisory authority within 72 hours of the incident, only if it poses a risk to the organisation and individuals   (2 marks)

Question Fourteen

Data fishing is sometimes referred to as

A.   Data bagging

B.   Data dredging

C.   Data merging

D.   Data pooling   (2 marks)

Question Fifteen

One of the challenges of big data is the fact that there is so much data, so many techniques and models to analyse the data and several ways to interpret the findings and results. The data analyst should therefore be more sceptical in the following aspects **EXCEPT** on the:

A.   Sources of data

B.   Data collection

C.   Data analysis

D.   Assumptions   (2 marks)

Question Sixteen

Which of the following is NOT a principle of data protection as provided by the data protection laws in various countries?

A.   Process data lawfully

B.   Maximise data collection

C.   Ensure data quality

D.   Limit data processing   (2 marks)

Question Seventeen

Two challenges of big data and analytics is inaccessibility of data and low speed of access to data. Which one of the following actions by an organisation will address the two challenges?
A.   Providing access to all data to specified employees
B.   Providing access to selected data to all employees
C.   Restricting access to all data for specified employees
D.   A database management system   (2 marks)

Question Eighteen

Which of the following is an open source revision/version control system?
A.   Numpy
B.   Git
C.   Scipy
D.   Loft   (2 marks)

Question Nineteen

Alteryx is an example of a _____.
A.   Data management tool
B.   Data cleaning tool
C.   Data visualisation tool
D.   Data presentation tool   (2 marks)

Question Twenty

Which of the following formulas in Ms Excel will provide a subtotal of variables provided in a list of vertically listed cells?

A.   =sum (A1:H20)

B.   =sum (A1:A20)

C.   =sum (A1;H20)

D.   =sum (A;A20)

SECTION II – TOTAL 60 MARKS

Question Twenty one

You are provided with the following extracts of the statement of profit or loss for Sepetuka Limited:

Sepetuka Limited
Statement of profit or loss extract for the year ended 30 September:

Year	2019 Sh."000"	2020 Sh."000"	2021 Sh."000"	2022 Sh."000"
Sales	54,000	64,800	81,000	95,580
Cost of sales	(32,400)	(32,400)	(32,400)	(38,232)
Gross profit	21,600	32,400	48,600	57,348
Operating expenses	(10,800)	(10,125)	(21,094)	(14,934)
Operating profit	10,800	22,275	27,506	42,414
Depreciation	(600)	(800)	(750)	(900)
Profit before interest and tax	10,200	21,475	26,756	41,514
Finance costs	(5,000)	(7,000)	(9,000)	(8,000)
Profit before tax	5,200	14,475	17,756	33,514
Income tax expense	(1,560)	(4,343)	(5,327)	(10,054)
Profit after tax	3,640	10,132	12,429	23,460

Required:

(a)   Calculate and interpret the following ratios:

      (i)   Annual revenue growth rates for years 2020, 2021 and 2022.   (3 marks)

      (ii)   Three years cumulative average growth rate (CAGR) for year 2022.   (3 marks)

      (iii)   Effective tax rate for the period 2019 to 2022.   (2 marks)

(b)   Now assume the following for Sepetuka Limited:

1.   Revenue growth rates are forecast under three scenarios namely base case, optimistic case and pessimistic case. The base case growth rate for the first year forecast is the 2022 CAGR. This is expected to reduce by 2% annually until the last year of the forecast subject to a minimum of 15%. The optimistic case is 20% above the base case while the pessimistic case is 10% below the base case in all forecast periods.

2.   Gross profit margin for the first year of the forecast is the 3-year average for the period from year 2020 to year 2022. This is expected to reduce by 2% annually until the last year of the forecast subject to a minimum of 50%.

3.   Operating expense ratios are modelled as 3-year averages for the period from 2020 to 2022. These are assumed to remain constant over the forecast period.

4.   Depreciation to revenue ratio is assumed to remain constant as the 3-year average for the period 2020 to 2022.

5.   Finance costs are expected to reduce steadily as the loans are repaid. Use the reduction rate in year 2022 over the forecast period.

6.   Income tax expense is calculated as the historical effective tax rate.

Required:

Prepare five-year forecast statements of profit or loss for Sepetuka Limited from year 2023 to year 2027.

             (12 marks)
(Total: 20 marks)

Question Twenty two

Mrs Jane Wakwa is the Marketing Director of Vuma Limited, a company that makes and sells electronic devices.

The company is considering the launch of a new mobile phone model branded "Trex". The available data is not fully reliable though Jane still feels that she can make a recommendation on whether or not to launch "Trex".

Additional information:

1.   Trex is estimated to have a shelf life of five years commencing year 2023.

2.   Trex will require the purchase of a machine at a cost of Sh.100 million at the end of year 2022, after which the machine will be sold for Sh.20 million at the end of the fifth year.

3.   The selling price and cost structures of Trex (for the first year 2023) with expected inflation factors are as follows:

	Sh. (Per unit)	Inflation rate (%) - from year 2024 onwards
Selling price	5,000	2%
Material costs	2,000	4%
Direct labour costs	1,000	5%
Incremental fixed cost (excludes depreciation)	500	10%

4.   The company is eligible for capital allowances (depreciation for tax purposes) at the rate of 25% on reducing balance.

5.   At the end of the project when the machine is sold, any gain or loss on disposal will be considered for tax.

6.   The tax rate on income and capital allowances is at the rate of 30% per annum. Assume that the tax for a given period is paid in the same year.

7.   The project will require an initial investment in working capital of Sh.20 million which will be increasing by Sh.5 million at the end of each year to cater for general inflation. The whole amount together with the periodic increase will, however, revert at the end of the project.

8.   Experience has shown that demand for new products is not exactly known in year one but tends to be stable thereafter. Jane has come up with the following estimates of demand for year 2023.

Probability	Expected sales (Units)
30%	40,000
40%	30,000
30%	10,000

Jane expects an initial increase in demand in year 2024 of 25% then a decline of 50% in year 2025. This level will remain the same till the end of the project.

9. Vuma Limited has a real weighted average cost of capital (WACC) of 8% and general inflation is expected to be at 4%.

Due to the risk of the project, Jane feels that the relevant nominal WACC should be increased by 3%.

Required:
Compute the following:

(a) The weighted average cost of capital to be used to evaluate the project. (2 marks)

(b) The relevant cash flows over the project period. (15 marks)

(c) The net present value (NPV) of the project. Advise on the viability of the project. (3 marks)
(Total: 20 marks)

Question Twenty three
You are evaluating a four-year project with an initial investment of Sh.10,000,000 on 1 January 2023 and the following cash flow characteristics:

Date	Cash flow (Sh.)
1 January 2023	(10,000,000)
30 June 2023	2,750,000
31 July 2024	4,250,000
30 September 2025	3,250,000
31 December 2026	2,750,000

The discount rate is given as 8%.

Required:
(a) (i) Calculate the NPV and XNPV of the project and hence determine whether the project is viable. (6 marks)

(ii) Highlighting the cause of the difference between NPV and XNPV, explain which one you would use in your analysis and why. (4 marks)

(b) (i) Calculate the IRR and XIRR of the project and hence determine whether the project is viable. (6 marks)

(ii) Highlighting the cause of the difference between IRR and XIRR, explain which one you would use in your analysis and why. (4 marks)
(Total: 20 marks)

Question Twenty four

Bamuda Limited presented the following financial statements for the years ended 30 June 2021 and 30 June 2022:

Statement of profit or loss for the year ended 30 June 2022:

	Sh."million"	Sh."million"
Revenue		473
Cost of sales		(229)
Gross profit		244
Gain on financial assets at fair value		5
Investment income		6
		255
Other expenses		
Administration expenses	48	
Distribution costs	76	
Finance costs	17	(141)
Profit before tax		114
Income tax expense		(47)
Profit after tax for the year		67

Statement of financial position as at 30 June:

	Sh."million" 2022		Sh."million" 2021	
Assets:				
Non-current assets:				
Property, plant and equipment		327		264
Intangible assets		40		50
Financial assets through other comprehensive income (OCI)		22		10
		389		324
Current assets:				
Inventory	123		176	
Trade receivables	95		87	
Financial assets at fair value	65		30	
Cash and cash equivalents	29	312	0	293
Total assets		701		617
Equity and liabilities:				
Equity:				
Ordinary share capital (Sh.10 each)		230		150
Share premium		30		0
Revaluation reserve		36		67
Finacial assets through OCI - equity reserve		2		0
Retained profits		121		91
		419		308
Non-current liabilities:				
10% convertible loan stock		100		150
Current liabilities:				
Bank overdraft			22	
Trade payables	156		100	
Interest payable	7		3	
Income tax payable	19	182	34	159
Total liabilities		282		309
Total equity and liabilities		701		617

Additional information:

1. Property, plant and equipment held by Bamuda Limited are items of plant and equipment and freehold premises. During the year ended 30 June 2022, items of plant and equipment which originally cost Sh.40 million were disposed of resulting in a loss of Sh. 6 million charged in administrative expenses. These items had a net book value of Sh.28 million as at the date of disposal.
2. Depreciation charge for the year ended 30 June 2022 was Sh.43 million.

3.   Sh.50 million of convertible loan stock was converted to Sh.50 million ordinary share capital at par during the year ended 30 June 2022.

Required:

Prepare the statement of cash flows for Bamuda Limited for the year ended 30 June 2022 as per IAS 7 "Statement of Cash Flows".

(Total 20 marks)

Question Twenty five

Farmgate Company Limited produces and sells shovels.

The company provides the following data for quantities of shovels produced and the related production cost for the calendar year 2021:

Month (year 2021)	Quantity produced (units)	Production cost Sh.
January	150,000	18,000,000
February	120,000	14,000,000
March	200,000	23,000,000
April	170,000	19,000,000
May	120,000	16,000,000
June	250,000	30,000,000
July	220,000	27,000,000
August	90,000	11,000,000
September	180,000	24,000,000
October	300,000	32,000,000
November	280,000	29,000,000
December	350,000	36,000,000

Required:

(a)   Determine the variable cost per unit and the total variable costs using regression analysis.   (6 marks)

(b)   Estimate the total production cost if the following number of units are produced:

   (i)   100,000 units

   (ii)   150,000 units

   (iii)   200,000 units

   Include the lower and upper boundaries at 95% confidence intervals.

(12 marks)

(c)   Prepare a chart plotting the predicted Y versus the actual Y from the regression function determined in (a) above. Interpret your answer.   (2 marks)

(Total 20 marks)

....

CPA ADVANCED LEVEL

BUSINESS DATA ANALYTICS (PRACTICAL PAPER)

PILOT PAPER

December 2022.   Time allowed: 3 hours.

Answer ALL questions. This paper has TWO sections. SECTION I has twenty (20) Multiple Choice Questions of two (2) marks each. SECTION II has three (3) practical questions of twenty (20) marks each.

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.

SECTION I – TOTAL 40 MARKS

Question One
Which of the following choices provides the correct definition of a data model?
A.   A visual representation of data objects
B.   An organisation of data objects
C.   A conceptual representation of data objects
D.   A graphical representation of data objects   (2 marks)

Question Two
The Cross Industry Standard Process for Data Mining (CRISP) requires a data analyst to understand the business as the first step. Which of the following components does **NOT** form part of this activity?
A.   Specific goals and requirements for data mining
B.   An overview of the available data and their quality
C.   Inventory and resources available
D.   Cost benefit analysis   (2 marks)

Question Three
Which of the following is the most common goal of statistical modelling?
A.   Inference
B.   Descriptive analysis
C.   Sub-setting
D.   Risk analysis   (2 marks)

Question Four
A collection of interesting and useful patterns in database is called _____.
A.   Data Set
B.   Information
C.   Knowledge
D.   Analytics   (2 marks)

Question Five
Which of the following is included in a data cleaning report?
A.   Data structure
B.   Data content
C.   Data cleaning activities
D.   Data models   (2 marks)

Question Six
Which of the following short-key functions will an analyst use in excel to paste contents into a cell?
A.   Control + F
B.   Control + V
C.   Control + C
D.   Control + U   (2 marks)

Question Seven
The following statements relate to the formula for computing the standard deviation in Ms Excel:
(i)   STDEV.P = Computes the standard deviation of an entire sample
(ii)   STDEV.S = Computes the standard deviation of an entire sample
(iii)   STDEVA = Computes the standard deviation of a selected sample

Which of the following choices is correct?
A.   Statements (i) and (ii) are correct
B.   Statements (i) and (iii) are correct
C.   Statements (ii) and (iii) are correct
D.   All the statements are correct   (2 marks)

Question Eight
Which one is not an array function in Ms Excel?
A.   MMULT
B.   TRANSPOSE
C.   VLOOKUP
D.   None of the above   (2 marks)

Question Nine
_____ is the basic data structure that can be thought of in an SQL table or a spreadsheet data representation.
A.   Series
B.   List
C.   Dataframe
D.   Points   (2 marks)

Question Ten
Which of the following components does **NOT** fall within the realm of "Velocity" in big data?
A.   Processes
B.   Streams
C.   Real-Time
D.   Terabytes   (2 marks)

Question Eleven
Which of the following 'Vs' of data provides for data as being authentic, trustworthy and original?
A.   Veracity
B.   Value
C.   Variability
D.   Variety   (2 marks)

Question Twelve
The following statements relate to the definition of the types of data analytics:
(i)   Descriptive analytics provide historical data and information
(ii)   Prescriptive analytics provide data and information for decision making
(iii)   Predictive analytics provide data and information to explain deviation from expected

Which of the following is the correct answer?
A.   Only statement (i) is correct
B.   Only statement (ii) is correct
C.   Only statement (iii) is correct
D.   Only statements (i) and (ii) are correct   (2 marks)

Question Thirteen
Which one ranks highest among the data analytics tools?
A.   Prescriptive analytics
B.   Predictive analytics
C.   Descriptive analytics
D.   Summative analytics   (2 marks)

Question Fourteen
Which of the following is **NOT** a data cleaning strategy?
A.   Removing incorrect data
B.   Removing duplicate data
C.   Fixing data with structural errors
D.   Filling in for missing data   (2 marks)

Question Fifteen
Which of the following software will a data analyst use to clean data?
A.   Data Stage
B.   SQL
C.   Tableu
D.   Microstrategy   (2 marks)

Question Sixteen
The following activities apply to data visualisation:
1.   The format
2.   The content
3.   The structure
4.   The purpose

Which of the following is the correct order of the activities?
A.   3,4,2 and 1
B.   4,3,1 and 2
C.   3,2,4, and 1
D.   4, 2,3 and 1   (2 marks)

Question Seventeen
Which of the following graph will likely present better a relationship between variables in data visualisation?
A.   Exploded Pie Chart
B.   Stacked Bar Chart
C.   Progress Line graph
D.   Radar Chart   (2 marks)

Question Eighteen
Which of the following correctly defines composition as a data visualisation technique?
A.   The method of graphical representation of data and information
B.   The method that shows the differences or similarities between values, with and without axes
C.   The method of presenting a part-to-whole relationship of data and information
D.   The method that shows relationships and connections between the data and information   (2 marks)

Question Nineteen
Which of the following people in an organisation will **LEAST** likely be involved in implementing the data protection laws in a country?
A.   Chief Risk Officer
B.   Chief Legal Officer
C.   Chief Information Officer
D.   Chief Executive Officer   (2 marks)

Question Twenty

Under the data protection laws, organisations will have to take steps to demonstrate they know the data they hold, where it is stored, and who it is shared with. This is referred to as:

A. Data Portability
B. Data Inventory
C. Data Traceability
D. Data Management (2 marks)

SECTION II – TOTAL 60 MARKS

Question Twenty One

Amani Pharmaceuticals Limited is a mid-sized firm exploring prospects of future growth. The management of the company has prepared the statement of profit or loss for the year ended 30 June 2022 with forecasts of expected growth rates in the various line items for the next three years.

This information is presented below:

Amani Pharmaceuticals Limited
Statement of profit or loss for the year ended 30 June 2022:

		Worst	Base	Best
	Sh."000"	%	%	%
Revenue	4,500.00	2	5	10
Cost of goods sold	(3,200.00)	2	5	10
Gross Profit	1,300.00			
Other incomes - Investment income	15.00	2	3	4
Total Income	1,315.00			
Administration expenses	(210.00)	5	5	5
Distribution expenses	(210.00)	2	5	10
Earnings before interest, tax, depreciation and amortisation	895.00			
Depreciation and amortisation	(150.00)	5	10	10
Earnings before interest and tax	745.00			
Interest expense	(120.00)	10	10	10
Earnings before tax	625.00			
Income tax expense	(160.00)			
Profit after tax (for the period)	465.00			
Dividends paid	(280.00)	10.00	10.00	10.00
Retained Profit for the year	185.00			
Retained profit b/f	550.00			
Retained Profit c/f	735.00			
Total equity	2,200.00			
Total assets	4,500.00	5%	10%	10%

(Growth Rates: Scenario spans the Worst / Base / Best columns)

Additional information:
1. Assume a corporation tax rate of 30%.
2. The growth in equity is only attributable to changes in retained profits.
    Hint: In applying the growth rates, ignore changes in opening and closing balances.
3. Total equity will grow only as a result of retained profits.

Required:

(a) Prepare a spreadsheet showing the forecast statement of profit or loss for the next three years under each of the three scenarios. (10 marks)

(b) Compute the following metrics for evaluating performance under each scenario:

    (i) Net profit margin. (2 marks)

    (ii) Return on assets. (2 marks)

    (iii) Return on equity. (2 marks)

(c) Present a summarised dash board with suitable graphs to highlight the revenues and net profit under each scenario.
(4 marks)
(Total: 20 marks)

Question Twenty Two

You are provided with the following extracts of the statement of profit or loss and statement of financial position for Darubini Limited for the years ended 30 June 2020, 2021 and 2022:

Darubini Limited
Statement of profit or loss extract for the year ended 30 June (figures in Sh."000"):

	2020	2021	2022
Sales	54,000	64,800	81,000
Cost of sales	(32,400)	(32,400)	(32,400)
Gross profit	**21,600**	**32,400**	**48,600**
Operating expenses	(10,800)	(10,125)	(21,094)
Operating profit	**10,800**	**22,275**	**27,506**
Depreciation	(600)	(800)	(750)
Profit before interest and tax	**10,200**	**21,475**	**26,756**
Finance costs	(5,000)	(7,000)	(9,000)
Profit before tax	**5,200**	**14,475**	**17,756**
Income tax expense	(1,560)	(4,343)	(5,327)
Net profit	**3,640**	**10,133**	**12,429**

Statement of financial position:
Assets:
Non-current assets:

	2020	2021	2022
Property, plant and equipment	25,800	45,200	46,700
Current assets:			
Inventory	15,000	21,750	27,488
Receivables	4,500	8,750	15,460
Cash and cash equivalents	0	0	3,823
Total current assets	**19,500**	**30,500**	**46,771**
Total assets	**45,300**	**75,700**	**93,471**
Equity and Liabilities:			
Long-term liabilities:			
Long term borrowings	33,333	46,667	60,000
Current liabilities:			
Accounts payable	3,750	6,658	7,989
Bank overdraft	5,297	9,322	0
Sub-total	**9,047**	**15,980**	**7,989**
Equity:			
Paid up capital	10,000	10,000	10,000
Retained earnings	(7,080)	3,053	15,482
Total equity	**2,920**	**13,053**	**25,482**
Total equity and liabilities	**45,300**	**75,700**	**93,471**

Required:

(a)   Prepare the statement of cash flows for the company for the two years ended 30 June 2021 and 30 June 2022.

(8 marks)

(b)   Calculate and comment on the following key ratios:

    (i)   Revenue growth rates for the years ended 30 June 2021 and 30 June 2022.   (2 marks)

    (ii)   Two years' cumulative average revenue growth rate for the year ended 30 June 2022.   (2 marks)

    (iii)   Return on equity (ROE) for each of the three years. Discuss your results.   (2 marks)

(c)   Using the two-years cumulative average revenue growth rate calculated in (b) (ii) above, prepare a one year forecast of the statement of profit or loss for the year ending 30 June 2023. Assume that the vertical ratios for year ended 30 June 2022 apply for the 2023 forecast except that depreciation and finance costs remain constant. Tax is assumed to be at 30%.   (6 marks)

(Total: 20 marks)

Question Twenty Three

(a)   You are a financial management consultant in an investment firm. A potential client requests you to analyse an investment with the following features:

- The cost of the investment is Sh.10,000,000.

- The investment pays out a sum of Sh."X" at the end of the first year; this payout grows at the rate of 10 percent per year for 11 years.

Required:

(i)   If your discount rate is 15%, calculate the minimum value of "X" that should attract the investor to make the investment.   (3 marks)

(ii)   The respective "NPVs" of the investment if the discount rates were to range between 10% and 15%.
                                                                                                                                          (3 marks)

(iii)   Assume that you have three project scenarios as outlined below:

Scenario	Discount rate	Initial cost Sh.	Starting cash flow Sh.	Growth in initial cash flow
Scenario 1	15%	10,000,000	1,500,000	10%
Scenario 2	10%	15,000,000	2,000,000	8%
Scenario 3	13%	20,000,000	4,000,000	7%

Prepare detailed outcomes of the project under each of the above scenarios.   (8 marks)

(b)   Assume that an investor takes an investment loan of Sh.150,000,000. The loan is for 48 months at an annual interest rate of 15 percent (which the bank translates to a monthly rate of 15%/12   = 1.25%). The payments are to be made at the end of each of the next 48 months.

Required:

(i)   Calculate the monthly payment on the loan.   (2 marks)

(ii)   Assume that you have the following options to negotiate with the bank:

Option	Annual interest rate (%)	Term
Option 1	15	48 months
Option 2	18	60 months
Option 3	13	36 months

Use a two-way data table to determine which option would be the most favourable to the investor.
                                                                                                                                          (4 marks)
                                                                                                                          (Total: 20 marks)
                     ……………………………………………………………………………