



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 2 December 2024. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two (2) sections. SECTION I has twenty (20) short response questions. Each question is allocated two (2) marks. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are indicated in the question.

Required resources:

- **A computer**
- **Python programming language**
- **Pycharm IDE**
- **Jupyter Notebook**

SECTION I (40 MARKS)

1. The python mode of programming where a group of commands are put into a file before executing the file sequentially is known as _____ . (2 marks)
2. Illustrate the output from the following Python code snippet as used in Python data visualisation:
for i in range (10):
if i % 2 == 0:
continue
print(i) (2 marks)
3. The principle of effective data visualisation defined as avoiding overcomplicating the visualisation with too many data points or chart types is referred as _____ . (2 marks)
4. Write a Python script to add a title “Simple line Plot” to a graph. (2 marks)
5. What is the largest machine learning library with extensive APIs for performing tensor computations? (2 marks)
6. The unnecessary or distracting elements in data visualisations that do not enhance understanding are called _____ . (2 marks)
7. Given the Python list: students = ["Kamau", "Otieno", "Juma", "Kamante", "Kodonyo", "Njok", "Bakari"], write a Python statement to insert the student "Ngetitch" between "Kamante" and "Kodonyo". (2 marks)
8. Given the Python dataset named “df_clients”, illustrate how you would display some statistical information from the dataset. (2 marks)
9. The SciPy library subpackage that includes the basic functions to solve eigenvalue problems, decompositions, matrix functions, special matrix functions, matrix equation solver functions and low level routines is referred to as _____ . (2 marks)
10. The python function used to exclude the missing values from a data frame is referred to as _____ . (2 marks)
11. A design pattern in Python that enables programmers to add new functionality to an existing object without altering its structure is referred to as _____ . (2 marks)

www.chopi.co.ke

12. Identify the type of visualisation in Figure 1 that is used in exploratory data analysis to illustrate statistical distributions in a dataset. (2 marks)

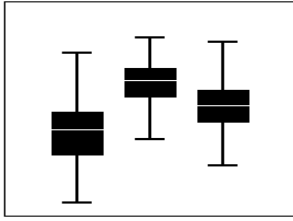
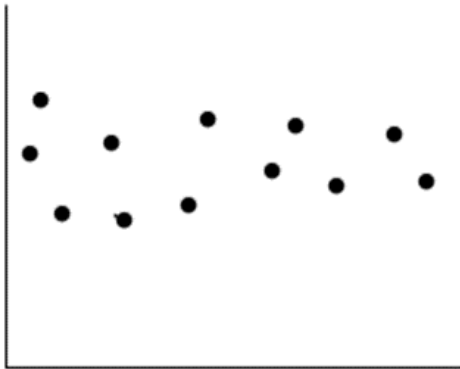


Figure 1

13. The plots used to determine whether a dataset or time series is random are known as _____. (2 marks)
14. Identify the correlation represented by the scatter plot below: (2 marks)



15. The process of extracting a portion of a list or string using a start and end index as used in Python programming is referred to as _____. (2 marks)
16. State the concept in Python programming which refers to a combination of the condition and action being repeated until the condition becomes false. (2 marks)
17. In Python data visualisation, a file that contains definitions and functions that can be imported into other programs is known as _____. (2 marks)
18. Indicate the mathematical operator in Python that is used for performing exponentiation operations. (2 marks)
19. Write the matplotlib statement that you could use to display the whole chart without displaying the axes as used in Python data visualisation. (2 marks)
20. State the output from the execution of the Python code: `pow(5, 2, mod=9)`. (2 marks)

SECTION II (60 MARKS)

21. Create a Word document called “Question 21” and use it to save and capture the screenshots for Question 21. Create a Python script called “student” to perform the tasks in questions (a) to (g).
- (a) Use the table shown below to create a CSV dataset called “stud” that contains the details on students’ performance in a class and save it on the most appropriate location on your local machine. (3 marks)

GRADE	NUMBER
A	25
B	40
C	30
D	15
E	5

- (b) Write a Python script to import the matplotlib and panda's libraries on the Python scripts "student" assigning the appropriate alias names. (3 marks)
- (c) Load the dataset from the CSV file into a DataFrame called "studgrades". (2 marks)
- (d) Display the contents of the dataset for verification purposes. (2 marks)
- (e) Visualise your results by using a blue bar chart labelling both axes with the appropriate captions. (3 marks)
- (f) Write a script to calculate and display the number of students in a class. (3 marks)
- (g) Write a script to calculate and display the percentage performance for each grade. (4 marks)

Save "Question 21" document and upload.

(Total: 20 marks)

22. Create a word processing document named "Question 22" and use the word processor document to save your answers to questions (a) and (b).

- (a) Given the dataset of A= (5,6,9,8,7,3,18,3,7,5,14,13,9,7) write the python codes that will perform the following tasks:
 - (i) Compute and print the mean and standard deviation. (5 marks)
 - (ii) Compute and print the t-test and p-values. (5 marks)
- (b) Given the data points U = (1,4,6,9) and V = (3,1,2,5), write the python code that will create the interpolation function, define the interpolation values and plot the results. (10 marks)

Capture screenshots to demonstrate how you have performed the above tasks.

Save "Question 22" document and upload.

(Total: 20 marks)

23. Create a Word document called "Question 23" and use it to save and capture the screenshots for Question 23.

Create a Python script called "employees" to perform the tasks in questions (a) to (e).

- (a) Study the table provided below showing the number of employees in each department for a non-governmental organisation and use it to answer questions (i) to (iii).

DEPARTMENTS	EMPLOYEES
HR	20
IT	50
Finance	15
Marketing	30
Sales	45

- (i) Using Python list, create a list of the departments and the number of employees from the data provided in the table and import the matplotlib library. (4 marks)
- (ii) Create the visualisation for the distribution of employees in the department using a labelled histogram chart adding titles and labels, grouping the data points into equal intervals of 5 bins with black edge color. (4 marks)
- (iii) Create a labelled box plot to visualise the employees' data. (2 marks)
- (b) Using the dataset provided in (a), define a Python dictionary to store the data in an object called "emprec". (3 marks)

- (c) Extract the keys and values of the data dictionary created in (b) and display the results on the console screen. (2 marks)
- (d) Write a Python script to create a labelled pie chart to visualise the data dictionary created in (b) above. (3 marks)
- (e) Write a Python script to visualise the data using a labelled scatter plot. (2 marks)

Save "Question 23" document and upload.

(Total: 20 marks)

.....

Chopi.co.ke



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 19 August 2024. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two (2) sections. SECTION I has twenty (20) short response questions. Each question is allocated two (2) marks. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are indicated in the question.

Required resources:

- **A computer**
- **Python programming language**

SECTION I (40 MARKS)

1. State the python data type that includes list, tuple and range. (2 marks)
2. The python variables which are used for configuration of data that varies between deployment stations and for sensitive data that should not be stored directly in the code are referred to as _____. (2 marks)
3. In python, a one-dimensional array that holds data of any type is known as _____. (2 marks)
4. State the syntax for a python method that can be used to clear all the data from the set data structure. (2 marks)
5. Which function in Python generates random numbers from a normal distribution? (2 marks)
6. The high-level, deep learning API that is written in python and is used to make the implementation of neural networks easy is known as _____. (2 marks)
7. The python environment variable which is used to ignore all import statements while calling the python interpreter is called _____. (2 marks)
8. State the python function that can be used to get keyboard data from the user in python. (2 marks)
9. A fast and space efficient multidimensional array that provides vectorised arithmetic operations and sophisticated broadcasting capabilities in NumPy is called _____. (2 marks)
10. What is the name of the powerful framework for obtaining, processing and storing web data using Python? (2 marks)
11. The high-level neural networks API written in python and capable of running on top of TensorFlow and is designed to be modular, fast and easy to use is known as _____. (2 marks)
12. State the Python library for interactive data visualisation that generates plots using HTML and JavaScript, designed for modern web browsers to offer high-performance interactivity. (2 marks)
13. Which Python library is used in scientific and technical computing to improve the performance of optimisation, integration, interpolation and linear algebra? (2 marks)

14. After installing Pandas, state the command you would use to import the library to the python script or notebook. (2 marks)
15. Which Pandas method is used to determine pairwise correlation of columns that exclude NA/null values? (2 marks)
16. State the name given to the Python data visualisation function for analysing data by grouping it, based on one or more columns and then performing operations such as aggregation on those groups. (2 marks)
17. Which tool transforms basic columnar data into a two-dimensional table, offering a comprehensive summary of the dataset? (2 marks)
18. Which fundamental concept in statistical hypothesis testing allows you to measure the degree of evidence against the null hypothesis based on actual data? (2 marks)
19. What is the name given to two-dimensional, size-mutable and heterogeneous tabular data structure in pandas? (2 marks)
20. The test that assesses if there is a statistically significant difference between the expected frequencies and the reserved frequencies in one or more categories of a contingency table is known as _____. (2 marks)

SECTION II (60 MARKS)

21. Create a word processing document named “Question 21” and use the word processor document to save your answers to questions (a) to (e) below.

- (a) Create the data set below and save it as “**stockdataset.csv**”. (4 marks)

Date	Opening	Maximum	Minimum	Closing
21/04/2021	120467	123200	122400	123987
23/04/2021	430676	423000	422765	435876
28/04/2021	232657	222451	221543	234900
30/04/2021	373834	376544	367089	378799
2/5/2021	431290	434167	439088	432153
9/5/2021	431290	431167	429088	432153
11/5/2021	482747	485667	484655	487707
13/5/2021	540203	542803	541345	543261
14/5/2021	592660	597660	596655	598815
15/5/2021	373834	375544	366787	367007
3/5/2021	121987	123400	122900	124567
4/5/2021	433876	423090	422995	436233
5/5/2021	232900	222651	221443	235675
7/5/2021	373834	375544	368089	376599

- (b) Write the python code that will import the data set into a data frame called “Stock data” and display the output after executing the code. (4 marks)
- (c) Write the python code which will ensure that the Date variable is a datetime variable and sort the data in ascending order by the date. (5 marks)
- (d) Write the python code which will create and display a line plot of date against closing and opening values. (4 marks)
- (e) Write the python code which will add the legends Start and End for the opening and closing stocks respectively. (3 marks)

Capture screenshots to demonstrate how you have performed the above tasks.

Save and upload “Question 21”.

(Total: 20 marks)

22. Create a word document called “Question 22” to save and capture the screenshots for Questions 22. Create a Python script called “Question 22” to perform the tasks in questions (a) to (d) below.

Use the table below and the questions that follows:

Programming language	Popularity
Java	22.2
Python	17.6
PHP	8.8
JavaScript	8
C#	7.7
C++	6.7

- (a) Import the Pandas and Matplotlib libraries and assign the aliases using the common convention. (2 marks)
- (b) Create a Python dictionary called “popu” using the table provided above. (3 marks)
- (c) Generate and display a bar chart using Matplotlib, including the grid lines, customise the minor grid and sets the title of the plot to “Popularity of Programming Languages”, the X-axis (Languages) and Y-axis (Popularity) respectively. (9 marks)
- (d) Using relevant Python libraries, write a program to generate 2000 random numbers between 1 and 500 and visualise the data using a well labeled horizontal bar chart and save the script as “horizontalB”. (6 marks)

Save and upload “Question 22”.

(Total: 20 marks)

23. Create a word document called “Question 23” to save and capture the screenshots for Questions 23. Create a Python script called “Question23” to perform the tasks in questions (a) to (b) below.

Use the table below to answer the questions that follow:

Name	Age	Allowance
Mary	20	20,000
Olive	18	18,000
Nick	21	17,000
Sam	19	8,000

- (a) (i) Create Python dictionary called “casual” and use it to create a DataFrame called “emp”. (4 marks)
- (ii) Write the Python code to display the second row and the row where the name Nick exists. (2 marks)
- (iii) Write the Python code to add a new column for gender for each record labelled as ‘M’ for male and ‘F’ for female. (2 marks)
- (iv) Write the Python code to rename the column for “Age” to “Allowance”. (2 marks)
- (b) Data visualisation relates to the use of visual tools like charts, graphs and maps to visually represent and understand trends, outliers and patterns in data.

Write a Python program to create bar plot from a DataFrame.

Sample Data Frame:

```
a b c d e
2 4,8,5,7,6
4 2,3,4,2,6
6 4,7,4,7,8
8 2,6,4,8,6
10 2,4,3,3,2
```

- (i) Import the Pandas dataframe, numpy and Matplotlib libraries and assign the aliases using the common convention. (3 marks)
- (ii) Create the array called 'a' and dataframe 'df' with the frame provided above. (3 marks)
- (iii) Plot a bar chart from the data above. Turn on the grid and minor ticks with line width set to 0.5 and color green and black. (4 marks)

Save and upload "Question 23".

(Total: 20 marks)

.....



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 22 April 2024. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two (2) sections. SECTION I has twenty (20) short response questions. Each question is allocated two (2) marks. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are shown at the end of the question.

Required resources

- **A computer**
- **Python program**

SECTION I (40 MARKS)

1. The python library that includes built-in functions for solving differential equations is referred to as: (2 marks)
2. The python function that defines the values of how data items are categorised is called: (2 marks)
3. The open-source Python library that extends the capabilities of Pandas to allow for easy manipulation and analysis of geospatial data is called: (2 marks)
4. State the Python inbuilt method used to generate N random numbers between 0 and 1 sampled from a uniform data set. (2 marks)
5. _____ is the measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. (2 marks)
6. The python package that creates a parse tree for parsed web pages based on specific criteria that can be used to extract, navigate, search and modify data from HTML is referred to as: (2 marks)
7. What is the measure of the difference between the highest and lowest values in a dataset? (2 marks)
8. The basic component of a plot in data visualisation that displays the horizontal and vertical lines that help readers gauge the values of data points is known as: (2 marks)
9. _____ in data visualisation involves presenting data in a narrative format to convey a specific message or insight effectively. (2 marks)
10. Write down the python commands to import the pandas library then load a dataset from the csv file known as "level3.csv" into a data frame object called "ddma". (2 marks)
11. The symbol used to represent data points in a scatter plot such as a dot, circle or a square is referred to as: (2 marks)
12. State the name of the python command that can be used to add a subplot to an existing 2-D plot in order to draw a 3-D plot. (2 marks)

13. The fundamental aspect of Python syntax used to define the structure of code blocks, such as loops, conditional statements and function definitions typically done using spaces or tabs is called _____. (2 marks)
14. _____ function is used in matplotlib to save a plot as an image such as PNG or JPEG format. (2 marks)
15. The lightweight data-interchange format that is human-readable and easy for both humans and machines and represents data in key-value pairs is known as: (2 marks)
16. _____ provides a comprehensive environment for writing, debugging and running Python code such as Jupyter Notebook, Spyder and IDLE. (2 marks)
17. _____ refers to using one value to describe multiple data points for example calculating average value of many student scores. (2 marks)
18. The python module that is used to read from or write to Excel files is referred to as: (2 marks)
19. State the name given to the Python data visualisation function for analysing data by categorising it based on one or more columns and then performing operations such as aggregation on those groups. (2 marks)
20. _____ is a python module with a highly-productive interface for solving machine learning (ML) problems with a focus on modern deep learning. (2 marks)

SECTION II (60 MARKS)

21. Create a word document called “Question 21” to save and capture the screenshots for Question 21. Create a Python script called “visual21” to perform the tasks in questions (a) to (b).

- (a) Use the table below to answer the questions that follow:

Year	Sales
2015	10000
2016	12000
2017	15000
2018	18000
2019	20000
2020	25000

- (i) Write the python code to import the Pandas and Matplotlib libraries and assign the aliases using the common convention. (2 marks)
 - (ii) Create a Python dictionary called “allData” using the table provided above. (3 marks)
 - (iii) Create a two-dimensional Pandas DataFrame called “dfSales” using the data dictionary created in (ii) above. (2 marks)
 - (iv) Write the python code to save the DataFrame “dfSales” to a file named “sales_data.csv”. (2 marks)
 - (v) Write the python code to generate and display a solid line plot with dots markers using Matplotlib, including the grid lines and sets the title of the plot to “Yearly Sales Trend”, the X-axis (Years) and Y-axis (Sales) respectively. (5 marks)
- (b) Using relevant Python libraries, write a program to generate 100 random numbers between 1 and 10, visualise the data using a well labelled boxplot and save the script as “boxplot”. (6 marks)

Capture a screenshot to demonstrate how you have performed the above task.

Save and upload “Question 21”.

(Total: 20 marks)

22. Create a word document called “Question 22” and use it to save and capture the screenshots for Questions 22.

Create a Python script called “visual22” to perform the tasks in questions (a) to (b).

(a) Use the table below to answer the questions that follow:

Name	Age	Salary
Dan	25	50000
Anita	30	60000
Susan	35	55000
Caleb	40	80000

- (i) Use the table shown above to create Python dictionary called “staff” and use it to create a DataFrame called “emp”. (4 marks)
- (ii) Write the Python code to display the second row and display the row where name is “Susan”. (2 marks)
- (iii) Write the Python code to add a new column for gender for each record labeled as ‘M’ for male and ‘F’ for female. (2 marks)
- (iv) Write the Python code to rename the column for “Age” to “Years”. (2 marks)

(b) Write python codes to answer the data visualisation questions in b (i) to b (iii).

- (i) Display a multiple plot lines for years against name and gender respectively using the relevant library. (4 marks)
- (ii) Use the Matplotlib library to create a new figure and a set of subplots within that figure. Set the title for the subplots to “Age Vs. Name”, the X-Axis as “Years” and Y-axis as “Name”. (3 marks)
- (iii) Create a line plot for “Years” against the “Names”, on a set of axes using the data from a DataFrame. The data points should be represented by circular markers in blue with a dashed line connecting them. (3 marks)

Capture a screenshot to demonstrate how you have performed the above task.

Save and upload “Question 22”.

(Total: 20 marks)

23. Create a word document called “Question 23” to save and capture the screenshots for Questions 23. Create a Python script called “visual23” to perform the tasks in questions (a) to (e).

- (a) Write a Python script to import the data visualisation libraries Seaborn and matplotlib, load the inbuilt dataset “tips” and display the data. (5 marks)
- (b) Using Seaborn, Create and display a scatter plot for “tips” against the “total bill” from the dataset loaded in question 23 (a). Using Python, Label the title, X and Y axes. (5 marks)
- (c) Write a Python script to display only rows where the total bill is greater than 30. (3 marks)
- (d) Write a Python script to calculate the tip percentage of the total bill. (4 marks)
- (e) Write a Python script to calculate the tip percentage of the total bill. (3 marks)

Capture a screenshot to demonstrate how you have performed the above task.

Save and upload “Question 23”.

(Total: 20 marks)

.....



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 4 December 2023. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two (2) sections. SECTION I has twenty (20) short response questions. Each question is allocated two (2) marks. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are shown at the end of the question.

Required resources

- **A computer**
- **Python program**

SECTION I (40 MARKS)

1. When using python, which command can be typed at the command prompt to install the core SciPy packages? (2 marks)
2. The python environment variable that contains a list of directories where Python searches for modules when there is need to import them in to the scripts is referred to as _____. (2 marks)
3. The Numpy array attribute that returns a tuple with each index having the number of corresponding elements is called _____. (2 marks)
4. The comprehensive software tool that provides features for writing, debugging, and running Python code with features like code completion, debugging tools and project management in Python data visualization is called _____. (2 marks)
5. The python function activity that gives the function a name, specifies the parameters that are to be included in the function and structures the blocks of code is called _____. (2 marks)
6. The python multi-line statements are usually denoted by _____. (2 marks)
7. A two-dimensional labeled data structure with columns of potentially different data types similar to a spreadsheet commonly used in Pandas structure for data manipulation, analysis and modeling is referred to as _____. (2 marks)
8. The Bokeh python data visualisation interface that provides high flexibility to application developers is referred to as _____. (2 marks)
9. Write the equation of a simple linear regression model used in Python data visualisation. (2 marks)
10. The python pandas method that returns the headers and a specified number of rows starting from the top is referred to as _____. (2 marks)

11. Write the python code that will import pyplot from Matplotlib. (2 marks)
12. Write the Python statement to create a histogram to visualise the distribution of data as used in Python data visualisation. (2 marks)
13. The python package that extends the datatypes used by pandas to allow spatial operations on geometric types is referred to as _____ . (2 marks)
14. Write the python function that can be used to find the mode of a dataframe. (2 marks)
15. You have been given a dataset containing information about customer purchases at a retail store. State the exploratory data analysis function that you would use in Python to understand the data types and check for missing values. (2 marks)
16. Write the python code that will load and read a CSV dataset in Python using pandas. (2 marks)
17. Which python function is used to find the skewness value? (2 marks)
18. Write the python code to generate a random 1x10 distribution for 2 occurrences. (2 marks)
19. Which is the python library that supports the statistical hypothesis test for independence between categorical variables? (2 mark)
20. The activity of processing the data in various formats for the purpose of analysing or getting them ready to be used with another set of data is referred to as: (2 marks)

SECTION II (60 MARKS)

21. Create a folder on the desktop called “PDV2023” to save your work. Create a word processing document called “Question 21” and use it to save and capture the screenshots for Questions 21. Create a Python script called “Quadratic” to perform the tasks in questions (a) to (f). (20 marks)

A quadratic equation is created using the data values 9, 27, 81, 243, 729, 2187. You are required to:

- (a) Import the matplotlib and numpy libraries and give them relevant alias names. (2 marks)
- (b) Use the data values to create a list called “lstquad” and one-dimensional array called “quad” using Numpy and display the array. (4 marks)
- (c) Generate a line plot graph to visualise the array, add a label to the x-axis and y-axis as “X-AXIS” and “Y-AXIS” respectively and set the title of the plot as “Quadratic equation plot”. Save your line plot as an image file called “plot.png”. (5 marks)
- (d) Customise the appearance of the plot in question 21 (c) to change the line style to dashed and red color? (2 marks)
- (e) Create a scatter plot with green circular markers ('o') and add a label to identify it as a scatter plot. (3 marks)
- (f) Create a bar plot with the red, green and blue colors and labels “A, B, C, D, E, F”. (4 marks)

Save “Question 21” and upload.

(Total: 20 marks)

22. Create a word processing document called “Question 22” to save and capture the screenshots for Questions 22. Save it in the “PDV2023” folder. Create a Python script called “3Dplot” to perform the tasks in questions (a) to (g). (10 marks)
 - (a) Create three list data structures with mylst1, mylst2 and mylst3 of size 5 and populate them with arbitrary integer values between 1 and 25. (4 marks)

- (b) Create and display a three dimension scatter plot by importing the relevant Python libraries. (5 marks)
- (c) Customise and display your scatter plot to have star markers of size of the data point to 500. (3 marks)
- (d) Modify the 3D scatter plot code to add relevant labels to the x, y, and z axes and provide a title for the plot. (2 marks)
- (e) Write a Python code to retrieve and display a slice of the first three elements from “mylst2”. (2 marks)
- (f) Write a Python code to get and display the last element of “mylst3” using negative indexing. (2 marks)
- (g) Write a Python code to extract and display a slice of the elements from index 1 to 4 from “mylst1”. (2 marks)

Save “Question 22” and upload.

(Total: 20 marks)

23. Create a word document called “Question 23” to save and capture the screenshots for Questions 23 and save it in the “PDV2023” folder. Create a Python script called “Timeseries” to perform the tasks in questions (a) to (g) below. Use the CSV dataset provided below to answer the questions that follow:

```
Date, Total Sales
2023-01-01,15000
2023-02-01,17500
2023-03-01,18500
2023-04-01,22000
2023-05-01,25000
2023-06-01,28000
2023-07-01,30000
2023-08-01,31000
2023-09-01,28500
2023-10-01,25000
2023-11-01,21500
2023-12-01,19000
```

- (a) Write a python code to import the pandas and matplotlib libraries using the relevant aliases. (2 marks)
- (b) Create a Python dataset as a dictionary and assign it to an object called “data”. (4 marks)
- (c) Create and display the pandas DataFrame named “sales”. (2 marks)
- (d) Write the Python code to convert the 'Date' column to a datetime object. (2 marks)
- (e) Plot the time series for the dataset to produce the output as shown in figure 1. (6 marks)

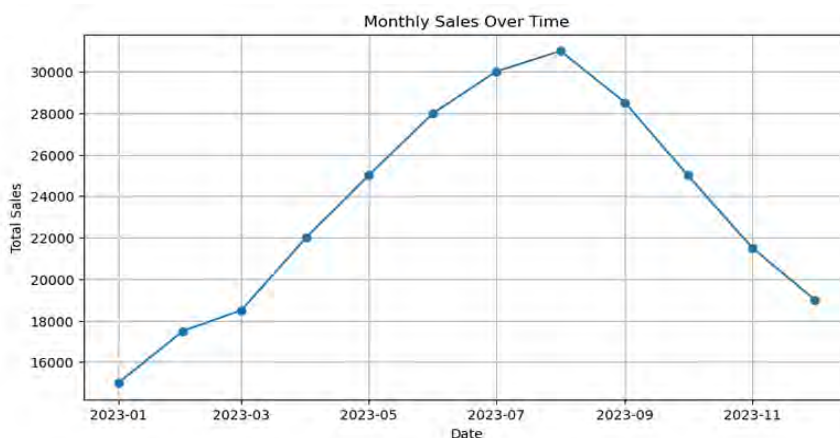


Figure 1

(f) Write the Python code to retrieve and display the total sales for 1st January 2023 from the “data” dictionary. (2 marks)

(g) Write the Python code to access and display the list of dates from the 'data' dictionary in reverse order. (2 marks)

Save “Question 23” and upload.

(Total: 20 marks)

.....

www.chopi.co.ke



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 21 August 2023. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two (2) sections. SECTION I has twenty (20) short response questions. Each question is allocated two (2) marks. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are shown at the end of the question.

Required resources

- **A computer**
- **Python program**

SECTION I (40 MARKS)

1. The name given to a set of horizontal and vertical lines that help with reading values from a plot in data visualisation chart is called _____ . (2 marks)
2. The linear regression module that helps to conduct statistical tests and estimate models is called _____ . (2 marks)
3. The open source python library which is used in mathematics, engineering, scientific and technical computing is referred to as _____ . (2 marks)
4. State the method used to create a scatter plot in python. (2 marks)
5. The type of distribution that is used to model the number of events that occur within a fixed interval of time or space, given the average rate of occurrence as used in statistical data analysis is called _____ . (2 marks)
6. Name two libraries that support Python Data Aggregation. (2 marks)
7. State the python function that will be used to label the title of a graph as “Car Graph”. (2 marks)
8. State the name given to the Python library which is used to create 2D graphs and plots using python scripts. (2 marks)
9. A file containing Python code that can be imported and used in other programs to allow the organisation of code into separate files for better modularity and reusability is called _____ . (2 marks)
10. The python package designed for geospatial data processing in order to produce maps and other geospatial data analyses is referred to as _____ . (2 marks)
11. The overall window that houses all the subplots and elements of visualisation as used in Python data visualisation is known as _____ . (2 marks)
12. State the function used by the Python Pandas library to support the detection of missing values. (2 marks)

13. State the function of the pandas library which is used to read the contents of a comma separated values (CSV) file into the python environment. (2 marks)
14. A single value or observation within a dataset represented by a dot or marker in visualisation is called _____ . (2 marks)
15. Name the python function that is used to read the entire html file. (2 marks)
16. The interactive visual displays that consolidate and present multiple data visualisations in a single view is referred to as _____ . (2 marks)
17. What is the name given to the open-source Python library that allows programmers to build powerful tools, dashboards and complex applications _____ . (2 marks)
18. Name the command used for installing Python Scikit-learn (Sklearn) libraries. (2 marks)
19. The python object that consists of contiguous one-dimensional segment of computer memory, combined with an indexing scheme that maps each item to a location in the memory block is called? (2 marks)
20. In python, which name best describes the data structure that stores only items that are of the same single data type? (2 marks)

SECTION II (60 MARKS)

QUESTION 21

Create a word document called “Question 21” and use it to save solutions to questions (a) to (g) below.

The table below contains an employee data set.

Employee	Age	Salary	Department
John	35	50,000	Sales
Emily	28	60,000	Marketing
Michael	42	70,000	Finance
Sophia	31	55,000	Sales
William	45	80,000	Finance
Olivia	33	65,000	Marketing

Required:

Using the Employees data set given above, create a Python script called “Employees” to perform the tasks in questions (a) to (g) below:

- (a) Import the matplotlib and numpy libraries and give them relevant alias names. (2 marks)
- (b) Use the dataset to create a data dictionary data structure called “employees”. (4 marks)
- (c) Extract the ages, salaries, departments from the data dictionary. (3 marks)
- (d) Use a relevant function to extract the number of department and store the results in an object called “department_counts”. (2 marks)

- (e) Create the data visualisation using a bar chart showing the number of departments from the employees dataset with a well labeled x-axis as “Department” y-axis as “Count” and title as “Employee Distribution by Department”. (4 marks)
- (f) Create a pie chart for age distribution with the title “Employee Age Distribution”. (2 marks)
- (g) Create a well labeled scatter plot for salary against age. (3 marks)

Save question 21 document and upload.

(Total: 20 marks)

QUESTION 22

Create a word document called “Question 22” and use it to save solutions to questions(a) to (g) below.

Use the data set provided below to answer the questions that follow:

A	B	C
1	2	3
2	4	6
3	6	9
4	8	12
5	10	15

Create a python script called “Boxplot” to perform the tasks in question (a) to (g).

- (a) Import the appropriate data visualisation library for creating a box plot. (2 marks)
- (b) Use the dataset above to create a Data Frame called “boxplot”. (4 marks)
- (c) Create a box plot of size 8 by 6. (3 marks)
- (d) Plot the box plot using the values from the Data Frame created in (b) above. (2 marks)
- (e) Sets the x-axis tick positions to be evenly spaced from 1 to the number of columns in the Data Frame. (3 marks)
- (f) Set the x-axis label and the y-axis as “Variables” and “Values” respectively. (3 marks)
- (g) Set the title of the plot as “Box Plot” and display the plot. (3 marks)

Save question 22 and upload.

(Total: 20 marks)

QUESTION 23

Create a word document called “Question 23” and use it to save solutions to questions (a) and (b) below.

- (a) Create a python script called “Floors” that will perform the tasks below:

The number of floors for the tallest buildings in Nairobi is as follows; 44,38,35,23,35,27,31. You are required to:

- (i) Create a list data structure called “floors” and display the output on the console screen. (3 marks)
- (ii) Display the last three elements of the list, accessing the list from right to the left. (3 marks)
- (iii) Use a relevant list function to display the size of the list. (3 marks)

- (iv) Display the elements of the list in ascending order using the relevant function. (3 marks)
- (v) Use a list function to add another building with 28 floors at the end of the list. (2 marks)
- (b) The LeadTech LTD decides to increase the salary for its programmers based on their performance. The programmers' rating ranges from lowest (1) to the highest (5). If the programmer's rating is greater than 4, the programmer gets an increment of 10%, else an increment of 5%.
Write a Python script called "salary" and assign a constant salary for an employee and the rating. The program should then calculate the programmer's increment and display the current, and new salary. (6 marks)

Save question 23 document and upload.

(Total: 20 marks)

.....

www.chopi.co.ke



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 24 April 2023. Morning Paper.

Time Allowed: 3 hours.

Answer ALL questions. This paper has two sections. SECTION I has twenty (20) short response questions of two (2) marks each. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are shown at the end of the question.

Required resources

- **A computer**
- **Python program**

SECTION I (40 MARKS)

1. Name the data type for the python command `x = ["John", "Alice", "Victor"]` (2 marks)
2. A 3-Dimension plot can be represented in a 2-Dimension graph using a python data visualisation technique called: (2 marks)
3. The python syntax that does not mix tabs with spaces and uses blank lines to separate top-level function and class definitions is referred to as: (2 marks)
4. State the command that can be used to load python into memory from the command prompt. (2 marks)
5. The python library that is extensively used for scientific and technical computations due to its efficient routines is referred to as: (2 marks)
6. _____ is a python environment variable that contains the path of an initialisation file containing python source code and is executed every time the interpreter is started. (2 marks)
7. Which type of data visualization is used to represent word frequency and to show the most commonly used words in a text dataset? (2 marks)
8. The process of transforming raw data into a format that can be used for analysis that involves cleaning, transforming, and aggregating the data to make it suitable for the analysis is called: (2 marks)
9. The block of code that performs a specific task in python, can be reused throughout a program and accepts arguments is referred to as: (2 marks)
10. The python data operation which involves the processing of data into different formats such as merging, grouping and concatenating is referred to as: (2 marks)
11. In python, the term that refers to a series of data points in which each data point is associated with a timestamp is referred to as: (2 marks)
12. The python term used to describe the use of python in statistical data analysis which is a special case of the binomial distribution where a single experiment is conducted so that the number of observations is 1 is referred to as: (2 marks)

13. Which python function is used to find the linear regression relationship of two variables? (2 marks)
14. Given the Python list, age=[34,23,13,61,22], what would be the output when the statement print(age[::-1]) (2 marks)
15. What name is given to the technique used to convert categorical variables into numerical data that can be used in the analysis? (2 marks)
16. Using python, illustrate how to create a 1D array in numpy with three integers numbers between 1 and 5. (2 marks)
17. Write the python code that will return the number of elements in the “boxes” array. (2 marks)
18. What kind of data visualisation chart is created using Seaborn sns.displot(data, bins=20, kde=False)? (2 marks)
19. State the two python functions that can be used to convert user input into a whole number. (2 marks)
20. State a commercial tool used for data visualisation in Python programming? (2 marks)

SECTION II (60 MARKS)

21. Create a word processing document named “Question 21” and use the word processor document to save your answers to questions (a) to (e).

- (a) Create a file named “Student” using the data below and save it in csv format. (4 marks)

A	B	C	D	E
Student number	Student name	AFR marks	AFM marks	AAA marks
KAD234	Anne Morgan	45	76	65
KAD345	Morris Ogechi	67	78	43
KAD453	Alice Reegan	54	47	56
KAD453	Alice Reegan	54	47	56
KXE567	Donald Benard		65	63
KHY678	Violet Jones	45	73	34
KUT779	Michael Nketia	56		49
KNM555	Kevin Ndiga	76	42	
KLT666	Joseph Borch	54	45	73
KRT778	Anita Ngunyi	45	76	65
KDE887	Steven Mike	67	58	81

- (b) Write the python code that will return a new data frame with no empty cells. (4 marks)
- (c) Write the python code that will remove all rows with NULL values. (4 marks)
- (d) Write the python code that will replace NULL values with the number 76. (4 marks)
- (e) Write the python code that will draw a line in a diagram from position (1, 4) to position (7, 10). (4 marks)

Upload Question 21

(Total: 20 marks)

22. Create a word processing document named “Question 22” and use the word processor document to save your answers to questions (a) to (g)

- (a) Write the code to define and display a 2-Dimensional arrays for the data provided in tables A and B using Numpy. (4 marks)

A

1	2
3	4

B

5	6
7	8

- (b) Write the code to calculate the product of the two matrices A and B created in 22 (a) and display the result. (2 marks)
- (c) Write the code get the inverse of the array B using a relevant python function and display the result. (2 marks)
- (d) Write the code get the transpose of the array B using a relevant python function and display the result. (2 marks)
- (e) Plot A and B as scatter plots with red and blue colors respectively using seaborn library. (4 marks)
- (f) Adding labels to the X and Y axis of the scatter plot using the relevant plot functions. (3 marks)
- (g) Add legend and show the plot for the scatter plot. (3 marks)

Upload Question 22

(Total: 20 marks)

23. Create a word processing document named “Question 23” and use the word processor document to save your answers to questions (a) to (f).

Write the Python program to do the following:

- (a) Import pandas and matplotlib libraries. (2 marks)
- (b) Use the table shown below to create a dataset called “employees” using python dictionary data structure. (4 marks)

Employee	Age	Salary	Job Group
Tim Kabi	32	50000	Group A
Raymond Tanui	28	55000	Group B
Seth Orlale	42	60000	Group A
Abigael Mugi	35	65000	Group C
Fabian Tom	40	70000	Group B

- (c) Use pandas to create a data frame called “employees”. (2 marks)
- (d) Create a well labeled bar plot with X and Y axes, legend and a title, to group average “Salary” by “Job Group”. (4 marks)
- (e) Create a well labeled histogram to show the distribution of ages. (4 marks)
- (f) Create a well labeled scatter plot for “Age” versus “Salary”. (4 marks)

Upload Question 23.

(Total: 20 marks)

.....



DIPLOMA IN DATA MANAGEMENT AND ANALYTICS (DDMA)

LEVEL III

PYTHON DATA VISUALISATION

MONDAY: 5 December 2022. Morning Paper.

Time Allowed: 3 hours.

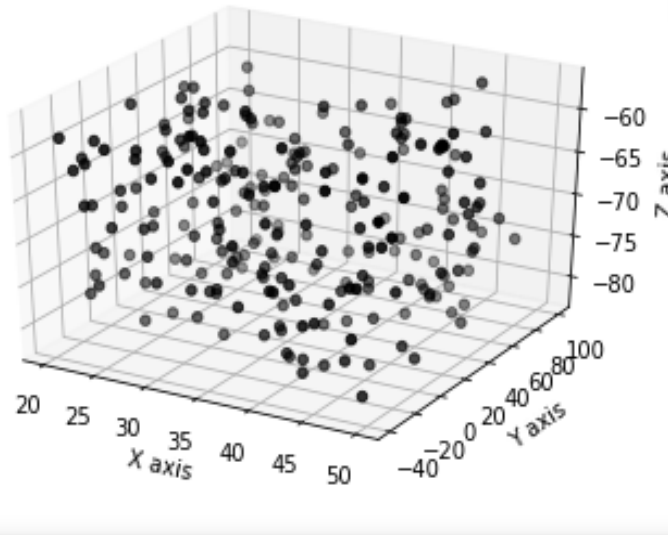
Answer ALL questions. This paper has two sections. SECTION I has twenty (20) short response questions of two (2) marks each. SECTION II has three (3) practical questions of sixty (60) marks. Marks allocated to each question are shown at the end of the question.

Required resources

- A computer
- Python program

SECTION I (40 MARKS)

1. A Scatterplot is a type of plot used to look at a predictive or correlational relationship between variables on a Cartesian coordinate. Write the name of the scatter plot shown in the figure below. (2 marks).



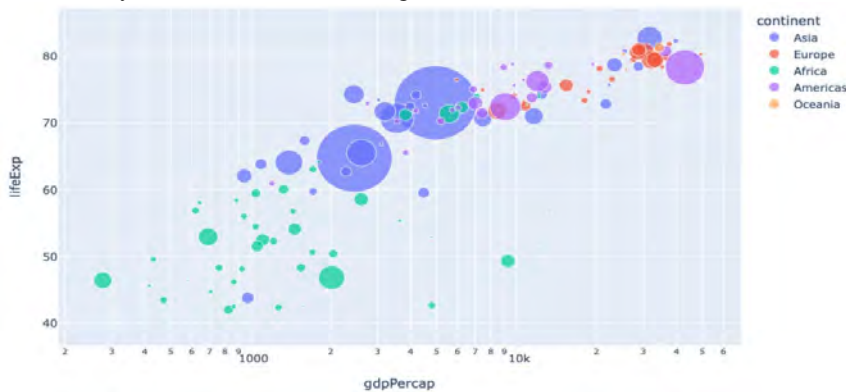
www.chopi.co.ke

2. Name the set of tools and techniques supporting the analysis of geospatial data through the use of interactive visualisation to allow data exploration and decision-making processes. (2 marks)
3. State the name given to the visualisation of a single feature of the dataset by use of charts such as histograms and pie charts. (2 marks)
4. Outline how you create an empty list called age in Python programming. (2 marks)
5. The main cause of type I and type II errors in statistical analysis is due to observations that lacks certain information, on the variables in data sets. This tends to compromise the quality of data visualisation. State the term used to describe the lack of such information as used in data pre-processing. (2 marks)

6. The term given to things that we use while coding, to organise multiple pieces of data in different ways such as lists, tuples, trees and graphs in Python programming is _____. (2 marks)
7. Variables whose value cannot be determined before they happen are termed as random variables. State the name given to the type of random variable that can take any value from a range of values. (2 marks)
8. A scientific discipline used in the field of statistics that uses mathematical tools and techniques to make forecasts and projections by analysing a given dataset is known as _____.
9. Identify the function used in Python to write records stored in a data frame to an SQL database? (2 marks)
10. The statistical measurement used to validate a hypothesis against observed data is referred to as _____. (2 marks)
11. Using the statement “**import numpy as np**”, write a Python code to create an array object called “**obj**” shown below? (2 marks)

```
[[1, 2]
 [3, 4]]
```

12. Identify the chart shown in the figure below as used in data visualisation. (2 marks)



13. The mathematical science and methods of collecting, organizing and analyzing data in such a way that meaningful inferences can be drawn from them is called: (2 marks)
14. The feature that enables you to extract portions of arrays, strings, tuples and lists to generate new ones in python programming is known as: (2 marks)
15. In linear regression visualisation, the distance between an actual, observed value and the value predicted by the regression line is referred to as? (2 marks)
16. A Python based open-source library for solving mathematical, scientific, engineering, and technical problems which allows users to manipulate the data and visualize it and usually depends on numpy library is known as? (2 marks)
17. There exists a huge number of data sources available on the internet today for free use by anyone for their needs such as models testing and experimentation. State the name of the data source that is openly accessible, exploitable, and editable and shared by anyone. (2 marks)
18. The process consisting of a number of steps in which the raw data are transformed and processed in order to produce data visualisations to make predictions by use of analytical and statistical tools is referred to as: (2 marks)

19. State the name given to the Python programming data structure which is based on the concept of Key-Value pairs? (2 marks)
20. Rewrite a python statement that is equivalent to: **“from matplotlib import pyplot as plt”**, in relations to Python data visualisation. (2 marks)

SECTION II (60 MARKS)

21. Create a word processing document named “Crops” and use the word processor document to save your answers to questions (a) to (d).

The table shown below contains the comparison of the produce of two crops; tea and avocados for six consecutive years in tons in one acre of land in Meru County. Study the tables to answer the questions that follow.

Year	Tea	Avocados
2011	2000	4000
2012	1900	3800
2013	2300	2700
2014	2500	2800
2015	2345	3950
2016	2390	4200

- (a) Using Python programming language create lists of year, tea and avocados and display their values on the console screen. (5 marks)
- (b) Use the list created in question (a) above to create a well labelled bar graph plot with X and Y axis and titles to show the production of tea against the years. (5 marks)
- (c) Use the list created in question (a) above to create a well labelled line graph. Use the default line style, colour and a circle marker to show the production of avocados against the years. (5 marks)
- (d) Use the list created in question (a) above to create a well labelled stacked bar graph, containing a title to show the production of tea and avocados against the years. (5 marks)

Upload Crops document.

(Total: 20 marks)

22. Create a word processing document named “Library” and use the word processor document to save your answers to questions (a) to (d).

- (a) Import the Python library for seaborn using “sns” as an alias and load the seaborn dataset called (“tips” in an object called “custtips” and display the dataset on the console screen. (4 marks)
- (b) Use the dataset in question (a) above to create bar plot to visualise the average day-wise bill amount (Y-Axis) of tips by customers for different days (X-Axis) of the week with X and Y axes clearly labeled. (4 marks)
- (c) Use the dataset in question (a) above to create bar plot to visualise the bill amount (Y-Axis) of tips by customers for different days (X-Axis) of the week labeled. (2 marks)
- (d) Write user friendly Python programming scripts to answer the following questions:
- (i) Given a list of numbers 5, 6, 7, 8, 9 create a tuple called “numbers” and display on the console screen. (2 marks)
- (ii) Write a python function to calculate the total of the elements of the tuple created in part d (i). (2 marks)

- (iii) Display the last element of the tuple “numbers”. (2 marks)
- (iv) Display the elements of the tuple created in d (i) in reverse. (2 marks)
- (v) Using a relevant tuple function, display size of the tuple created in d (i). (2 marks)

Upload Library document.

(Total: 20 marks)

23. Create a word processing document named “Random” and use the word processor document to save your answers to questions (a) to (e) below.

- (a) Import Python numpy and matplotlib libraries, and assign them aliases “np” and “plt” respectively. (2 marks)
- (b) Use the numpy to create an object called “myrange” to store random numbers in the range of 50 to 80 with a step up of 0.05 and display the values on the console screen. (2 marks)
- (c) Convert the random numbers created in question (b) above and display the values on the console screen. (2 marks)
- (d) Visualise using a line graph, the conversion to the random numbers created in question (b) above to its equivalent cosines numbers, with X-axis labeled as "Random Numbers" and Y-axis labeled as "Cosines" with the title as "Cosine of random cosine numbers". (4 marks)
- (e) (i) Create the dataset shown below and save it on your PC’s drive C as “examanalysis.csv”. (4 marks)

Reg. No	Mean	Grade
001	56	C
002	78	A
003	67	B
004	82	A
005	45	D

- (ii) Use pandas’ relevant function to load the dataset created in e (i) into an object called exams. (3 marks)
- (iii) Display the contents of the object created in part e (ii). (1 mark)
- (iv) Create a histogram to visualise grades scored by the students in the class. (2 marks)

Upload Random document.

(Total: 20 marks)

.....