|  |  |  |  |
| --- | --- | --- | --- |
|  | **PES University, Bengaluru**  (Established under Karnataka Act No. 16 of 2013) | | **UE20CS901** |
| **March/April 2024: END SEMESTER ASSESSMENT (ESA)**  **M TECH DATA SCIENCE AND MACHINE LEARNING\_ SEMESTER I**  **UE20CS901 - Python for Data Science** | | | |
| Time: 3 Hrs | | Answer All Questions | Max Marks: 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| **INSTRUCTIONS** | | | |
| * All questions are compulsory. * Section A should be handwritten in the answer script provided and signed at the end of the same. * Section B and C are coding questions which have to be answered in the system. | | | |
|  | | | |
| **SECTION A – 20 MARKS** | | | |
| 1 | a) | How do you count the occurrences of a particular element in the list? a\_list = ["a", "b", a"] | 2 |
| b) | What is Global Keyword in Python? | 2 |
| c) | How can you get a random number in python? | 2 |
| d) | What do you know about the list and dictionary comprehensions? Explain with an example. | 2 |
| e) | Explain negative indexing in Python with example | 2 |
|  | | | |
| 2 | a) | Which function can be used to remove duplicates in DataFrames? | 2 |
| b) | How do you identify missing values and deal with missing values in Dataframe? | 2 |
| c) | Explain different ranking methods for Series. | 2 |
| d) | What is GroupBy in Pandas? | 2 |
| e) | Which plot is used to visualize distribution of data? Which methods are available for it? | 2 |
|  | | | |
| **SECTION B – 40 MARKS** | | | |
| 3 | a) | Happy Number  A happy number is a number defined by the following process:  Starting with any positive integer, replace the number by the sum of the squares of its digits, and repeat the process until the number either equals 1 (where it will stay), or it loops endlessly in a cycle which does not include 1.  For example: 7\*7 -> 49, 4\*4 + 9\*9 -> 97  7 => 49 => 97 => 130 => 10 => 1 => Happy!  Another example: 2 => 4 => 16 => 37 => 58 => 89 => 145 => 42 => 20 => 4 => Loop encountered => Not Happy! | 10 |
| b) | Write a program that asks the user to enter a list of test scores. Then create and print a list of lists,where the first list contains all the As (90 or above), the second all the Bs (80–89), etc. (where Cs are 70–79, Ds are 60–69, and Fs are 0–59) ( 5 marks ) | 5 |
| c) | Write a program that asks the user to enter a date in the format of a month name followed by a day number. The program should print out whether the date is real. For instance, January 32 is not a real date, and neither is February 30. Assume that the user enters the full month name and that the month name is valid. To validate the day number, use a dictionary that contains the number of days in each month. (5 Marks) | 5 |
| d) | A superstore has multiple items to sell. These items are categorised as category A and category B. For a festival time, the superstore decided to declare some discount, as given below.  for category A items, it charges Rs.15 per item if you buy less than 30 quantities of an item. If you buy between 30 and 99 quantity of an item, the cost is Rs.12 per item. If the buying quantity is 100 or more for an item, the cost is Rs.10 per item.  For category B items, it charges Rs.30 per item if you buy less than 50 quantity of an item. If you buy between 50 and 99 quantity of an item, the cost is Rs.25 per item. If the buying quantity is 100 or more for an item, the cost is Rs.20 per item. Further, it also gives a discount of 5% if the total bill exceeds 1000 and 10% if the bill exceeds 1500.  Write a program that repeatedly asks the user to enter no of items and their category. If the user enters 'STOP' it shows the user has done with the purchase, now calculate the total payable amount, and display it to the user(10 Marks)  Sample Input/Output:  Enter no of items and its category: 23,A  Enter no of items and its category: 56, b  Enter no of items and its category: 67,b  Enter no of items and its category: 110,a  Enter no of items and its category: STOP Total amount payable is : 2808.0 | 10 |
| e) | Write a program to answer following: Have the function UsernameValidation(`str`) take the `str` parameter being passed and determine if the string is a valid username according to the following rules:  #1. The username is between 4 and 25 characters.  #2. It must start with a letter.  #3. It can only contain letters, numbers, and the underscore character.  #4. It cannot end with an underscore character  #If the username is valid then your program should return the string `true`, otherwise return the string `false`.  \*\*\*Sample output:\*\*\*  aa\_ =>False  uaa\_\_hello\_worldW =>True | 10 |
|  |  |  |  |
| **SECTION C – 40 MARKS** | | | |  |
| 4 | a) | Use the countries dataset to answer the following questions  1. What is overall life expectancy across the world.(2 marks)  2. Create a dataframe containing 10 countries with highest population( 2 marks )  3. Add a new column in countries\_df to record overall GDP per country product of population & per capita Gdp.( 2 marks )  4. Create a dataframe containing 10 countries with lowest gdp per capita, among the countries with population greater than 100 million.(3 marks)  5. Create a data frame that counts number of countries in each continent. (2 marks )  6.How does the distribution of population vary across different continents,? (3 marks)  7.Categorize countries based on life expectancy. Use apply function and create a new column to store the different categories (3 marks)  8.Calculate the average population for each continent where the average life expectancy is greater than 75.) ( 3 marks) | 20 |
| 4 | b) | **Use the adult dataset to answer the following questions**  1.What is the proportion of German citizens (native-country feature)? ( 2marks)  2.Which occupation was the most common, illustrate with a appropriate graph? (2 marks)  3. What are mean value and standard deviation of the age of those who receive more than 50K per year (salary feature) and those who receive less than 50K per year? (2 marks)  4. Is it true that people who receive more than 50k have at least high school education? (education - Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters or Doctorate feature (2marks)  5. Among whom the proportion of those who earn a lot(>50K) is more: among married or single men (marital-status feature)? Consider married those who have a marital-status starting with Married (Married-civ-spouse, Married-spouse-absent or Married-AF-spouse), the rest are considered bachelors. (4 marks)  6. What is the maximum number of hours a person works per week (hours-per-week feature)? How many people work such a number of hours and what is the percentage of those who earn a lot among them? (4 marks)  7. Count the average time of work (hours-per-week) those who earning a little and a lot (salary) for each country (native-country). (4 marks) | 20 |