|  |  |  |  |
| --- | --- | --- | --- |
|  | **PES University, Bengaluru**  (Established under Karnataka Act No. 16 of 2013) | | **UE20CS901** |
| **JANUARY/ FEBRUARY 2021: END SEMESTER ASSESSMENT (ESA)**  **M TECH DATA SCIENCE AND MACHINE LEARNING\_ SEMESTER I**  **UE20CS901 - Python for Data Science** | | | |
| Time: 2 Hrs | | Answer All Questions | Max Marks: 60 |

|  |  |  |  |
| --- | --- | --- | --- |
| **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) | What are the differences between the lambda and def with respect to Python functions? | 2 |
| b) | What are some of the advantages of using list comprehensions? (Write at least 2) | 2 |
| c) | Explain the “Break” statement with examples. | 2 |
| d) | What is Boolean in Python? | 2 |
| e) | Explain negative indexing in Python with example. | 2 |
|  | | | |
| 2 | a) | What is the use of Shape() function in Numpy array? | 2 |
| b) | What is the difference in Pandas series and Numpy array? | 2 |
| c) | Define the different ways in which a DataFrame can be created in Pandas? | 2 |
| d) | Which function can be used to remove a column from DataFrame? | 2 |
| e) | How to Rename the index or columns of a Pandas DataFrame? | 2 |
|  | | | |
| **SECTION B – 20 MARKS** | | | |
| 3 | a) | Using the list of tree names = ["Mango tree", "Coconut tree", "papaya tree" , "Apple tree", "Banana tree", "Blackberry tree"] answer the below questions.  Step1: Using tree names remove items at indexes [2,3,4] by replacing with an empty list.(2 Mark)  Step2: Add the items[Neem Tree, Peepal Tree] starting at index [3]. (3 Mark)  Step3: Write a function determine how many times a given letter "n" occurs in a string= (" Banana tree") (5 Marks) | 10 |
| 3 | b) | i). One way to solve math equations is by trial and error. Computers can help with that. The equation 21𝑥2−𝑥3+21904=0 has an integer solution between 1 and 100. Print that x value which satisfies this equation (5 Marks)  ii). Write a program that asks the user to enter a string of at least six characters. Then print out the following: (5 Marks)  (a) Every character of the string except the last one  (b) Every character of the string except the first and last  (c) If the string contains a lowercase letter ‘a’, print out the index of the first ‘a’. Otherwise print that there is no lowercase letter ‘a’.  (d) The string in all caps  (e) The string with every space replaced with an underscore ( \_ )  Sample Input/Output:  Enter a string: ABCDE aGHI  a. Every character of the string except the last one: ABCDE aGH  b. Every character of the string except the first and last: BCDE aGH  c. index of the first a : 6  d. The string in all caps: ABCDE AGHI  e. The string with every space replaced with an underscore: ABCDE\_aGHI | 10 |
|  |  |  |  |
| **SECTION C– 20 MARKS** | | | |
| 4 | a) | **kc\_house\_data\_mini** The dataset consists of house prices from King County an area in the US State of Washington, this data also covers Seattle   * id * date * price * bedrooms * bathrooms * sqft\_living * sqft\_lot * sqft\_above * sqft\_basement * yr\_built * yr\_renovated   Read the Dataset 'kc\_house\_data\_mini.csv' and answer the following questions.  A. Convert date column to datetime object (1 Mark)  B. Create a new column for Apartment\_age in years of an apartment as difference between the year in which apartment is built and date column in dataset. (2 Marks)  C. Find out 5 most recently renovated apartments (2 Marks)  D. Count the number of apartments for each bedroom’s category. Plot it using count plot. (2 Marks)  E. What is mean price for the 3 Bedroom apartment with at least 3 bathrooms. (3 Marks) | 10 |
| 4 | b) | Read the dataset blockbusters and answer the below questions  The highest grossing movies of each year are not necessarily the best movies of that year. And more often than not, the ones scooping up all the awards don't end up as the highest grossing ones or box-office hits.  \* Main\_Genre - The most relevant genre of the film.  \* Genre\_2 - The most relevant sub-genre of the film.  \* Genre\_3 - The lesser relevant sub-genre of the film.  \* imdb\_rating - The IMDb rating of the film, scraped from the official IMDb website.  \* length - The length of the film in minutes.  \* rank\_in\_year - The rank of the film (1 to 10) in that year, according to higher gross.  \* rating - The maturity index of a film, namely G, PG, PG-13 and R, given by Motion Picture Association of America.  \* studio - The studio house that distributes the film worldwide. While cleaning the data, I replaced every studio name with its \* parent studio to avoid sparsity. All the studios which are not present now, or have been bought by other studios, are also \* replaced.  \* title - Obviously, the name of the film.  \* worldwide\_gross - The total gross of the film, Domestic and Worldwide, in dollars, from Box Office Mojo.  \* year - The year of release for the film.  A. Identify the number of blockbuster movies under the Main Genre Comedy. (1 Mark)  B. List all the names of all the blockbuster movies with imdb ratings above 8. (1 Mark)  C. Display the details of all the movies which were produced by Walt Disney Pictures and suitable to be viewed by General Audiences. (3 Marks)  D. Plot a bar graph to show the total worldwide gross for year. (5 Marks) |  |