|  |  |  |  |
| --- | --- | --- | --- |
| A close up of a sign  Description automatically generated | | | |
| Data Science Academy Bootcamp | | | |
| **Final exam** | | | |
| **Name, Surname:** |  | Date: |  |
| **EXAM:**  Data Science with Python | |  |  |
| ***Instructions:*** You will be given ***2 hours*** to complete examination. There are 2 parts and 4 types of questions including: True/False, Multiple choice, Fill in the blanks and Practical.  **First Part** will last ***40 minutes***  and you must write answers directly on word document. Internet is not permitted.  **Second Part** is practical question and will last ***1 hour and 20 minutes.*** It is your responsibility to submit your exam answers electronically to [homework@dsa.az](mailto:homework@dsa.az) with **”Python Exam”** subject.  Before you leave exam, you **MUST** contact mentor to confirm that you have submitted your examination.  Good Luck!   |  | | --- | | **PART 2: Practical** | | **Directions: Read each question carefully and provide responses as script including codes, results, and interpretation.** | | 1. In local restaurant, a food server recorded the following data on all customers they served during an interval of two and a half months in early 1990. The restaurant, located in a suburban shopping mall, was part of national chain and served a varied menu. In observance of local law, the restaurant offered to seat in a non-smoking section to patrons who requested it. Each record includes a day and time, and taken together, they show the server’s work schedule. Import **Tips** (Tips.csv) dataset and answer following questions: 2. Call our dataset and get its first 8 and last 4 rows, show number of observations. (2 points) 3. What is max, min and mean of **size**? Round it. (2 points) 4. Report number of null values for each column. (3 points) 5. Remove columns with null values. 6. What is the most frequent day of given tips? (2 points) 7. Sort values by **total bill** and **tip** at restaurant in descending order. (2 points) 8. Group by data by **time** and **sex** and find mean. (5 points) 9. Which customers pay more than others? 10. Filter **total bill** greater than 20 and **tip** less than 3. (4 points) 11. Filter male customers with **size** equal to 2. (4 points) 12. Find most frequent **gender** using “if” function. (3 points) 13. Find number of bills greater than daily average. (5 points) |  |  | | --- | | **PART 2: Practical** | | **Directions: Read each question carefully and provide answers on script with your full name. send all answers in existing script you created for previous part.** | | 1. Write a Python program that takes input from the user and removes the characters which have even index values. (4 points) 2. Suppose you are given a list of words as **Wordlist.** Write pyhton code that will wirte different words in paragraph by repeating words three times. For example, if **Wordlist** is [‘Jose’, ‘Sue’, ‘Ivan’], then your code should print. (3 points)   Jose Jose Jose  Sue Sue Sue  Ivan Ivan Ivan   1. Do the following tasks:   **Note:** You can do all these tasks in one class, you do not have to create new classes for each task:   1. Create a “def function” that takes 2 arguments, and returns a value to the main method. (1 point) 2. Create a “def function” that accepts different values as parameters and returns a list. (2 points) 3. Create a “Pandas Dataframe”, use visualization techniques: seaborn and matplotlib. (2 points) 4. Create a “Pandas Dataframe”, use indexing, slicing, cleaning operations. (3 points) |  |  | | --- | | **PART 2: Practical** | | **Directions: Read Each question carefully and provide responses as script and separate report** | | 1. There has been a revenue decline in the Portuguese Bank and they would like to know what actions to take. After investigation, they found that the root cause was that their customers are not investing enough for long term deposits. So, the bank would like to identify existing customers that have a higher chance to subscribe for a long-term deposit and focus marketing efforts on such customers.   The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be subscribed (“yes”) or not (“no”) subscribed.  **Portuguese\_banking.csv** consists of 32951 observations and 15 features without the target feature.  **Goal**: The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).  **Dataset description:**  **Age:** customer age **Housing:** has housing loan?  **Job:** type of job **Loan:** has personal loan  **Marital:** marital status **Contact:** contact communication type  **Education:** education level **day of week:** last contact day of the week  **Default:** has credit in default? **campaign:** number of contacts performed during this  **Month:** last contact month of year campaign and for this client  **Duration:** last contact duration, in seconds **previous:** number of contacts performed before this campaign  **Pdays:** number of days that passed by after and for this client  the client was last contacted from a previous **y-** has the client subscribed a term deposit?  campaign  **Poutcome:** outcome of the previous marketing campaign   1. Import data (Portuguese\_banking.csv) and get familiarized with it. (3 points) 2. Show distribution of age columns. Get visual to check if outliers exist in ‘age’ column or not. (3 points) 3. Prepare data for modelling (7 points) 4. Fit data into different models- Logistic Regression, SVM, Naïve Bayes, Decision Tree, Random Forest. (5 points) 5. Print detailed reports regarding accuracy metric and confusion matrix for each model, check overfitting. (5 points) | | | | |