|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Department of Data Science**  **Faculty of Computing and Information Technology**  **University of the Punjab**    **Semester Fall 2024**  **Mid Term Examination**   |  |  | | --- | --- | | **Course Code: DS-302** | **Title: Introduction to Data Science** | | **Shift / Section:** | **Max Time: 90 minutes** | | **Student ID:** | **Student Name:** | |  |

**Instructions:**

|  |
| --- |
| **Sheet No.:** |
| **Invigilator’s Signature** |
| **Date: 13 Nov 2024** |

1. Do not forget to pray before starting to attempt the paper. Trust me it helps.  
   Remember! SOMEONE is always with you (Be Relaxed), and HE is also watching you (Be Honest)
2. Question Paper is **SELF EXPLANATORY**. Write your assumption. In case of error in question, credit goes to the student.☺
3. Solve your paper using Black/Blue Pen only strictly in the given space.
4. Check the all the 10 page are present in answer sheet, if any page is missing, please inform the invigilator.
5. Attempt All Questions in a precise fashion.
6. No Answer sheet is required.

Good Luck

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Q3** | **Q4** | **Total** |
| 30 | 15 | 15 | 10 | 70 |
|  |  |  |  |  |

**Examiner’s Signature : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Marks: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DO NOT OPEN UNTIL YOU ARE TOLD TO DO SO.**

**Q#1: (30 marks)**

You are working as a data analyst for a company that wants to analyze employee payroll data. You need to process, summarize, and analyze various aspects of the employee payroll records.

1. **Task:** You are given a list of strings, where each string contains information about an employee's payroll in the format:

**“employee\_id:101,name:John,department:HR,salary:50000,bonus:2000"**

Write a function that converts each string into a dictionary with the following keys: "employee\_id", '"name"', "department", "salary", and "bonus”. Convert the salary and bonus to integers.

1. **Task:** Given a list of dictionaries from Part 1, write a function to calculate the total payroll for the company by summing up **‘salary + bonus’** for all employees.
2. **Task:** Write a function that finds the employee with the highest total compensation **(salary + bonus)**. Return the name of the employee and their total compensation.
3. **Task:** Write a function that returns a set of names of employees who belong to the '"HR"' department.
4. **Task:** Write a function that returns a list of employee names who received a bonus greater than $1500.
5. **Task:** Write a function that uses a lambda function to filter employees with a total compensation **(salary + bonus)** greater than **$60,000**.
6. **Task:** Modify your function from Part 1 to handle potential errors in the input format using a try-except block.If any error occurs (such as missing data or an invalid format), print an error message specifying the problem and skip that record.
7. **Task:** Write a function that uses the sorted() function to sort employees first by the length of their department name in ascending order.

**Q#2: (15 Marks)**

You’re working with a team on a collaborative project, and everyone is pushing updates to different branches on GitHub. To keep your work in sync with the latest changes from teammates, maintain a clean commit history, and fix any recent mistakes in your commits, you’ll need to complete a series of tasks.

1. **Task**: Your teammate has made updates to the main branch on GitHub. Before starting your own work, bring your local main branch up-to-date with the latest changes from GitHub.
2. **Task**: After making updates on your improve-cleaning-scripts branch, you notice a mistake in your last commit. Undo only the last commit without affecting the code changes, so you can make adjustments and recommit.
3. **Task**: Another teammate has pushed new changes to the data-cleaning branch on GitHub. Integrate these changes into your improve-cleaning-scripts branch without creating a merge commit, to keep the commit history clean.
4. **Task**: After completing your work, push the improve-cleaning-scripts branch to GitHub to share it with your team.
5. **Task**: Explain the difference between git merge and git rebase when integrating changes from one branch into another.
6. **Task**: Briefly describe the types of git reset. What is the key difference between them?

**Q#3: (15 Marks)**

You’re analyzing e-commerce data and want to gain insights into daily website visitors, cart values, and order counts over a 100-day period. To do this, you need to generate initial data arrays, perform basic operations, and aggregate the data for further analysis.

1. **Task**: Generate initial data arrays. Using NumPy, create three arrays of length 100 each to represent:
   * Daily visitors (integers between 500 and 1500)
   * Average cart value (random floating-point values between 20 and 100)
   * Daily total orders (integers between 50 and 300)
2. **Task**: Perform basic array operations. Calculate the mean cart value across all days and identify how many days had a cart value above this average.
3. **Task**: Create a data summary using boolean indexing. Using boolean indexing, create an array that flags days with over 1000 visitors.
4. **Task**: Find and display the count of days where both the visitors exceeded 1000 and the total orders were above 200.
5. **Task**: Perform data aggregation and slicing. Stack the daily visitors, average cart value, and daily orders arrays vertically to form a single 3x100 matrix.
6. **Task**: Calculate the row-wise average, representing the average visitors, cart value, and orders for each day.
7. **Task**: Extract a slice of this matrix for days 25 to 50 and calculate the mean for each metric within this range to analyze mid-period performance.

**Q#4: (10 Marks)**

Describe the stages of a Data Science project lifecycle in detail. For each stage, explain its primary purpose and provide a specific example of an activity or task that is typically carried out at that stage.