**Capstone Project - Python Fundamentals**

**Dataset Explanations**

**Attributes:**

1. id - Assigned number for the project head who will be in charge of the project.
2. name - The person handling the project.
3. gender – Male(M), Female(F).
4. city - Locations of the project.
5. age - Number of years the project will be active.
6. status - Status of the project.
7. designation level - Position of the project head.
   * Excessive failures indicate designation grades to reduce.
   * A person with a good reputation means a very high chance to increase his designation.

**Designation Scale:**

* 1 - Highest.
* 2, 3 - Mid positions, with 4 being the least.
* If anyone crosses 4, they lose eligibility for heading the project.

**Dataframes:**

**Employee DataFrame:**

| **ID** | **Name** | **Gender** | **City** | **Age** |
| --- | --- | --- | --- | --- |
| A001 | John Alter | M | Paris | 25 |
| A002 | Alice Luxumberg | F | London | 27 |
| A003 | Tom Sabestine | M | Berlin | 29 |
| A004 | Nina Adgra | F | New York | 31 |
| A005 | Amy Johny | F | Madrid | 30 |

**Seniority Level DataFrame:**

| **ID** | **Designation Level** |
| --- | --- |
| A001 | 2 |
| A002 | 2 |
| A003 | 3 |
| A004 | 2 |
| A005 | 3 |

**Project DataFrame:**

| **ID** | **Project** | **Cost** | **Status** |
| --- | --- | --- | --- |
| A001 | Project 1 | 1002000 | Finished |
| A002 | Project 2 | 2000000 | Ongoing |
| A003 | Project 3 | 4500000 | Finished |
| A004 | Project 4 | 5500000 | Ongoing |
| A005 | Project 5 |  | Finished |
| A002 | Project 6 | 680000 | Failed |
| A005 | Project 7 | 400000 | Finished |
| A003 | Project 8 | 350000 | Failed |
| A001 | Project 9 |  | Ongoing |
| A003 | Project 10 | 300000 | Finished |
| A001 | Project 11 | 2000000 | Failed |
| A004 | Project 12 | 1000000 | Ongoing |
| A004 | Project 13 | 3000000 | Finished |
| A005 | Project 14 | 200000 | Finished |

**Problems**

**Task 1:** Create three dataframes in Python based on the tables above and save them as .csv files. Use the saved .csv files for all subsequent tasks.

**Task 2:** The “Cost” column in the Project dataframe has missing values. Use a running average to fill these missing values using a "For" loop.

**Task 3:** Split the “Name” column in the Employee dataframe into two new columns, “First Name” and “Last Name,” and remove the older “Name” column.

**Task 4:** Join all three dataframes into one single dataframe called "Final."

**Task 5:** Add a new "Bonus" column in the Final dataframe. Give a 5% bonus based on the project cost to employees who have finished projects.

**Task 6:** Demote the designation level by 1 for employees whose project status is "Failed." Remove employee records whose designation level exceeds 4.

**Task 7:** Add "Mr." or "Mrs." to the first name column and drop the gender column.

**Task 8:** Promote the designation level by 1 for employees older than 29 using an IF condition.

**Task 9:** Add the total cost of all projects for each employee into a new dataframe called "TotalProjCost," with columns for ID, First Name, and Total Cost.

**Task 10:** Print all the employee details whose city name contains the letter “o.”

**Performance Evaluation:** You need to complete all the given tasks using Python libraries like Pandas and NumPy. The goal is to evaluate your data manipulation and analysis skills for real-world data science challenges.

**Requirements or Deliverables:**

* Submit a .ipynb file containing your code and outputs.
* Include comments explainations.