# Data Migration - Take Home Test Instructions

## Overview

Create a github repo where you will store both the python scripts and the outputted files.

You will submit this github repo as the work.

## Python

### Objective:

You have been provided with [a CSV file](https://docs.google.com/spreadsheets/d/1txqq8CEvrYgrV7Ask4kelhVxAB1-pld3Bs3NthJT-c0/edit?gid=2049450242#gid=2049450242) containing contact information. Your task is to create a

Python script that processes this file according to the requirements listed below and outputs a new CSV file with the specified changes applied. You should use pandas or similar packages to implement your script.

### Instructions:

1. Read the file

2. Perform Data Transformation Requirements:

* Remove Duplicate Rows:
  + Ensure the output file contains only unique rows.
* Modify Column Headers:
  + Prefix all column headers with “Contact: “
  + Example: The column header "First Name" should be changed to "Contact: First Name".
* Standardize Name Formatting:
  + Convert all names to proper cases (e.g., "JOHN DOE" should become "John Doe").
* Reformat Dates:
  + Change the format of the "Birthday" column from DD/MM/YYYY to MM/DD/YYYY.
* Generate Unique IDs:
  + Create a unique numeric ID for each contact after duplicates are removed.
  + IDs should range between 1 and 1,000,000 and must remain strictly numeric.
* Adjust the Assigned Column:
  + Replace the initials in the "Assigned" column with their full names using the following legend:
    - GM = Gabe Michel
    - AA = Aaron Artsen
    - BL = Bond Liver
    - IC = Individual Contributor
    - TM = Tim Mint
    - Blanks or empty values should be assigned to "Gabe Michel" as the default.

3. Output File:

* Save the transformed data into a new CSV file.

### Deliverables

1. A Python script that implements the above transformations.
2. The new CSV file generated by the script.
3. Submit your Python script and the output CSV file in a **github project**.

## Excel

### Objective

Exhibit the ability to use excel commands to manipulate data

### Instructions

* Use the following file: [Link](https://docs.google.com/spreadsheets/d/1g5HlQ35pyhLjp_fsdiK09SrwVMZSu5P003mb6rLg-sc/edit?gid=568348716#gid=568348716)
* Download it and save as an excel file.
* Using excel commands like VIF and VLOOKUP, do the following
  + Combine first, middle, and last names (columns B, C, and D) into a new column in the following format: Last, First Middle
  + There should be no space before the last name or after the middle name.
  + There should be one space after the comma.
  + There should be one space between the first and middle names
  + Split “Address” field (column F) into 5 different columns (street1, street2, city, state, and zip)
    - There should be no spaces before or after the contents of each cell
  + Clean the “Cases” field (column O) of all excess spaces
  + Convert “Birthday” field (column L) to date format of M/d/YYYY, where the month is represented by a single digit, the day is represented by a single digit and the year is 4 digits long
* Save the excel file and add it to the github repo that you created