Ignite – Anonymise sensitive data in CSV files using python before storing in the cloud

Problem Statement:

Sensitive C3 level data needs to be anonymised/encrypted before we can upload to GCP (for data analysis purposes).

Requirements (scope):

Must

1. Have CSV file with random data
2. Create a button to upload the csv file that needs to be encrypted
3. Read the contents of csv file
4. Anonymise contents of csv file using python functions
5. Write the anonymised contents to new file
6. If data is the same, provide it with the same anonymised code
7. Convert first name, last name, employer id, email and location

Could

1. Unanonymise the CSV File
2. Anonymise only specific columns in the csv file (look into packages)
3. User friendly GUI
4. Give users option of naming new anonymised file
5. Explain reasoning behind the algorithm used for anonymization
6. Presentation

Timescale:

1. Small – 10 mins
2. Medium – 3 hours
3. Medium – 1 hour
4. Medium – 3 hours
5. Medium – 1 hour
6. Large – 4 hour
7. Medium – 2 hours
8. Laaaarge – dunno
9. Large – 9 hours
10. Medium – 1 hour
11. Large – 4 hours
12. Medium – 1 hour
13. Presentation – 3 hours