Analytics Assignment

Assignment Deliverables

1. Provide the code with readme with sample output and ppt explaining your approach.
2. Push code on git/bitbucket and attach link, ppt and sample input, output and your resume in the mail to summer@innovaccer.com with **HackerCamp Summer 2018 Submission - Analytics** as its subject.

Scenario

Variation in names leads to difficulty in identifying a unique person and hence deduplication of records is an unsolved challenge. The problem becomes more complicated in cases where data is coming from multiple sources. Following variations are same as Vladimir Frometa:

Vladimir Antonio Frometa Garo
Vladimir A Frometa Garo
Vladimir Frometa
Vladimir Frometa G
Vladimir A Frometa
Vladimir A Frometa G

Problem Statement

Train a model to identify unique patients in the sample dataset attached here. You are free to add more training data to make the model more accurate.

Additional Questions
If you have questions related to the assignment deliverables, please drop me an email on sourab.pareek@innovaccer.com and I will try to get back to you as soon as it allows.

Cheers, and good luck :)