

What is the problem you want to solve?

Using DLP techniques to fuzzy match file store addresses to ERP master addresses.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

My client is a natural gas company. All service installation records are stored and maintained in a file store with address info. However, no system or field linkage is built between the file store and ERP application. The address is entered manually each time the file added to the file store. The only way to find the records for an address in ERP is to manually search the file store using the address.

What data are you using? How will you acquire the data?

- Data set 1 is the list of addresses from a legacy system: the data will be massaged and merged insider the internal database and then be imported into the Python notebook.
- Data Set 2 is the list of addresses from address master table: the data will be massaged and merged insider the internal database and then be imported into the Python notebook.
- Data set 3 USPS Suffix data: download from USPS website, manually format and reshape the data, then import into the Python notebook.

Important Files

Capstone Project 2 Proposal
CapstoneProject2_DataWranglingAndEDA

Project 2 Milestone Report 1
CapstoneProject2_MachineLearning
Project 2 Milestone Report 2