Address Deduplication using record linkage

Synopsys:

Precisely is a market leader in providing data that drive trust.

But to achieve a satisfactory level of integrity from data sourced from different vendor is not easy.

It requires a sleuth of cleansing, standardization and linking operation steps to achieve data integrity.

This means a lot of manual effort, extra cost and increased time to market for clean product.

This project aims at applying machine learning and industry standard data cleansing means to reduce the net cost accrued in standardizing address data.

The major steps involved in the project are as follows: -

1. Acquisition of Raw data
2. Understanding the data
3. Data cleansing and standardization
4. Creating record linkage i.e. pairing one record with other based on matching attributes
5. Creating a model from linked records
6. Apply the model to generate more matched data from unforeseen data.