Analysis Of Enron Dataset

The Enron Corpus is a large database of over 600,000 emails generated by 158 employees

This dataset was collected and prepared by the CALO Project (A Cognitive Assistant that Learns and Organizes). It contains data from about 150 users, mostly senior management of Enron, organized into folders. The corpus contains a total of about 0.5M messages. This data was originally made public, and posted to the web, by the Federal Energy Regulatory Commission during its investigation.

18 insider was involved

35 total poi was there

Concepts that we have learned till now; Project Outcome:

1)loops and data structures

2) modules

3) How to read pickel file

4) How to plot data: (matplotlib)

5) Basic of machine leaning: (Library: sklearn)

>>Naive Bayes Classifier

>>Support Vactor Machine

>>Regression

>>Outliers

Resources:

1.<https://www.udacity.com/>course/viewer#!/c-ud120/l-2269418543/e-3026708756/m-3022188629 : Udacity course

2) https://www.cs.cmu.edu/~./enron/ : Enron Dataset

Output

