

```
In [2]: #!/usr/bin/python

"""
    Starter code for exploring the Enron dataset (emails + finances);
    loads up the dataset (pickled dict of dicts).
    The dataset has the form:
    enron_data["LASTNAME FIRSTNAME MIDDLEINITIAL"] = { features_dict }
    {features_dict} is a dictionary of features associated with that person.
    You should explore features_dict as part of the mini-project,
    but here's an example to get you started:
    enron_data["SKILLING JEFFREY K"]["bonus"] = 5600000

"""

import pickle

enron_data = pickle.load(open("../final_project/final_project_dataset.pkl", "r"))
```

The Enron fraud is a big, messy and totally fascinating story about corporate malfeasance of nearly every imaginable type. The Enron email and financial datasets are also big, messy treasure troves of information, which become much more useful once you know your way around them a bit. We've combined the email and finance data into a single dataset, which you'll explore in this mini-project.

Getting started:

Clone this git repository: <https://github.com/udacity/ud120-projects>
(<https://github.com/udacity/ud120-projects>)

Open the starter code: datasets_questions/explore_enron_data.py

For each person, how many features are available?

```
In [4]: len(enron_data['THE TRAVEL AGENCY IN THE PARK'])
```

```
Out[4]: 21
```

The "poi" feature records whether the person is a person of interest, according to our definition.

How many POIs are there in the E+F dataset?

In other words, count the number of entries in the dictionary where `data[person_name]["poi"]==1`

```
In [10]: count = 0
for x in enron_data:
    if x['poi'] == True:
        count += 1
print(count)
```

```
TypeErrorTraceback (most recent call last)
<ipython-input-10-e897dd1cbe66> in <module>()
      1 count = 0
      2 for x in enron_data:
----> 3     if x['poi'] == True:
      4         count += 1
      5 print(count)
```

TypeError: string indices must be integers, not str

We compiled a list of all POI names (in ../final_project/poi_names.txt) and associated email addresses (in ../final_project/poi_email_addresses.py).

How many POI's were there total?

(Use the names file, not the email addresses, since many folks have more than one address and a few didn't work for Enron, so we don't have their emails.)

35

As you can see, we have many of the POIs in our E+F dataset, but not all of them. Why is that a potential problem?

We will return to this later to explain how a POI could end up not being in the Enron E+F dataset, so you fully understand the issue before moving on.

Less data

Like any dict of dicts, individual people/features can be accessed like so:

```
enron_data["LASTNAME FIRSTNAME"]["feature_name"]
```

or, sometimes

```
enron_data["LASTNAME FIRSTNAME MIDDLEINITIAL"]["feature_name"]
```

What is the total value of the stock belonging to James Prentice?

```
In [17]: enron_data["PRENTICE JAMES"]['total_stock_value']
```

```
Out[17]: 1095040
```

Like any dict of dicts, individual people/features can be accessed like so:

```
enron_data["LASTNAME FIRSTNAME"]["feature_name"]
```

How many email messages do we have from Wesley Colwell to persons of interest?

```
In [19]: enron_data["COLWELL WESLEY"]["from_this_person_to_poi"]
```

```
Out[19]: 11
```

at's the value of stock options exercised by Jeffrey K Skilling?

```
In [21]: enron_data["SKILLING JEFFREY K"]["exercised_stock_options"]
```

```
Out[21]: 19250000
```

```
In [ ]:
```