**BNP Data Scientist Technical Exam**

**Getting Started**

1. Set up [git](http://help.github.com/) and create a repository. We'll leave the details for you to work out.
2. Work on your solution to the problem below locally.
3. When you are satisfied with your results, email [Nelson.Claros@us.bnpparibas.com](mailto:Nelson.Claros@us.bnpparibas.com) with a zipped copy of your cloned repo. **Don't include any files generated by your solution**. We should be able to run any scripts you submit that generate additional files.

**Compose a Data Summarization Script**

You have to write this script in Python.

Please create a script “summarize-enron.py” that can be run from the [Unix] command line in the format:

> python summarize-enron.py enron-event-history-all.csv

The Enron event history (.csv, adapted from the widely-used publicly available data set) is attached to this email. The columns contain:

* **time** - time is Unix time (in milliseconds)
* **message identifier**
* **sender**
* **recipients** - pipe-separated list of email recipients
* **topic** - always empty
* **mode** - always "email"

Your script should produce three outputs:

1. A .csv file with three columns---"person", "sent", "received"---where the final two columns contain the number of emails that person sent or received in the data set. This file should be sorted by the number of emails sent.
2. A PNG image visualizing the number of emails sent over time by some of the most prolific senders in (1). There are no specific guidelines regarding the format and specific content of the visualization---you can choose which and how many senders to include, and the type of plot---but you should strive to make it as clear and informative as possible, making sure to represent time in some meaningful way.
3. A visualization that shows, for the same people, the number of unique people/email addresses who contacted them over the same time period. The raw number of unique incoming contacts is not quite as important as the relative numbers (compared across the individuals from (2) ) and how they change over time.

**Assessment**

Your solution will be assessed based on:

* attention to detail
* completion of the tasks
* algorithm efficiency
* code readability
* adherence to common coding practices that best enable sharing, re-using, and extending the code.