* Open Anaconda Prompt & then by using following Command open

-Jupyter notebook

* > jypter notebook

|  |  |
| --- | --- |
| In [1] : | import pandas as pd  from pathlib import Path  import fuzzymatcher  import recordlinkage |
| In [2] : | hospital\_accounts = pd.read\_csv('https://github.com/chris1610/pbpython/raw/master/ data/hospital\_account\_info.csv') |
| In [3] : | hospital\_accounts.head() |
| In [4] : | hospital\_reimbursement = pd.read\_csv('https://raw.githubusercontent.com/chris1610/pbpython/master/  data/hospital\_reimbursement.csv') |
| In [5] : | hospital\_reimbursement.head() |
| In [6] : | left\_on=["Facility Name","Address", "City","State"]  right\_on=["Provider Name","Provider Street Address","Provider City","Provider State"] |
| In [7] : | matched\_results = fuzzymatcher.fuzzy\_left\_join(hospital\_accounts,hospital\_reimbursement, left\_on, right\_on,left\_id\_col='Account\_Num',right\_id\_col='Provider\_Num') |
| In [8] : | matched\_results.head( ) |
| In [9] : | cols = [ "best\_match\_score", "Facility Name", "Provider Name", "Address", "Provider Street Address","Provider City", "City", "Provider State", "State"] |
| In [10] : | matched\_results[cols].sort\_values(by=['best\_match\_score'], ascending=False).head(5) |
| In [11] : | matched\_results[cols].sort\_values(by=['best\_match\_score'], ascending=True).head(5) |
| In [12] : | matched\_results[cols].query("best\_match\_score <= .80").sort\_values(  by=['best\_match\_score'], ascending=False).head(5) |
| In[13] | matched\_results[cols].query("best\_match\_score <= 1").sort\_values(by=['best\_match\_score'], ascending=False).head(10) |