Assignment 7 – Apache Beam
DS 561 – Cloud Computing

Vishwas Bhaktavatsala
U74206902
vishwasb@bu.edu
GitHub Url: https://github.com/vishwas21/DS561-vishwas-assignments

---

Command used to run the Dataflow:
(Can be run by installing all the dependencies in the requirements.txt)

Outputs:

Outgoing File:


```
[('4257.html', 249), ('8540.html', 249), ('9800.html', 249), ('5580.html', 249), ('4975.html', 249)]
```

Incoming File Count:


```
[('5984.html', 188), ('1912.html', 166), ('5789.html', 163), ('7885.html', 162), ('7231.html', 162)]
```