1. Steps to run Biogen-

* Sagemaker -> Training -> Biogen -> Antony -> Biogen\_Loop.py

1. To run in background-

* Open terminal inside training instance
* cd SageMaker/Antony
* nohup python Biogen\_Loop.py &
* tail -f nohup.out to see the running flow

1. Finish running for both Ongoing and Historic Input Data
2. Output files will be stored as Generic and Result Output -

s3://pas-biogen-sagemaker/Antony/Ongoing\_Output/General/{chunk\_id}\_generic.csv

s3://pas-biogen-sagemaker/Antony/Ongoing\_Output/Result/{chunk\_id}\_result.csv

1. Go to EMR Cluster-

* Notebook -> anto -> /Biogen\_Final\_Consolidation/
* Hist\_2019\_spark.ipynb

Consolidates all 2019 output into a single data, adds demographic features

* Hist\_2020\_spark.ipynb

Consolidates all 2020 output into a single data, adds demographic features

* Hist\_2021\_spark.ipynb

Consolidates all 2021 output into a single data, adds demographic features

* Ongoing\_Final\_Consolidation.ipynb

Fetches all ongoing data stored in s3 bucket and adds the AD flag and anchor features

* Hist\_Final\_Consolidation.ipynb

Fetches all ongoing data stored in s3 bucket and adds the AD flag and anchor features

1. Final Output

s3://pas-biogen-sagemaker/Antony/Ongoing\_Output/Alzheimer\_Historical

s3://pas-biogen-sagemaker/Antony/Ongoing\_Output/Alzheimer\_Ongoing