Reflections:

This unit is one among the most useful one, that I have completed as part of my course. I would consider this unit as the most industry focused.

What I liked ?

- Inclusion of AWS Sagemaker
- Git
- Encouragement to explore projects and enabling students to push their boundaries.

Challenges faced by students:

- Frequent switching between python and R for various courses
- Final assessment data size vs 2hr AWS sessions, uploading to us-east-1 instance takes a lot of time.

Suggestions:

- The version control system should be the first lesson and students needs to maintain a gitlab or github repo for dsts-lab, so that they can practice. Git takes a lot of practice.
- Azure offers more to students as uni uses Microsoft Products.
- DSTS could focus more on cloud computing (Azure or AWS), pyspark and MLflow/airflow (used in industry).
- Data ingestion and ETL is not taught in any other units in UC, this unit could focus more on this area.
- Data serialisation, deserialization, IO buffers could be discussed in lectures
- Tableau could be part of Data Exploration and visualisation unit, currently being taught by Shuang
- PCA, feature engineering and other basic data science concepts are already covered in other units in data science course (PRML by Dr. Dharmendra, Programming for DS by Ram, Soft Computing, all these units discusses these topics).

Thanks,

Ajul