23 Feb 15:00 - 16:15(Front + Back)
Important dates:
Week 9: Interim check-in
Week 12: (3/4 Monday) Video submission
Week 13: (14/4 Friday) Report submission

Main target of the project:
The task in this project is to analyse the similarities and differences between data science departments using unsupervised learning techniques.

Front end:
The front end team should present their product to prospective students who may want to use it to make their decision between universities or programs.

Backend:
The modelling team should direct their findings to University administrators who would want to use the results to identify gaps/strengths in their program for future improvements. University administrators are typically former academics, so they are usually comfortable with technical details. However they are extremely busy.

Tasks:
1. [Front + Back] Data collection: Web scraping
2. [Back] Data cleaning and modelling:
3. [Front] Design webpage
4. [Front] Implementate webpage
5. [Front + Back] compose up and conduct UAT

Notes (Ideas for what to include):
- Starting pay
- Module feedback/Introduction(key word of the content)
- Match list of skills that jobs require to course content
- Similarity score for each module, show differences

Key metrics:
NUS vs NTU vs SMU [comparing two at a time eg. apple website]
1. Percentage of computing and Maths: For individual module level, determine by module code
2. Post graduate route: Starting pay, jobs for different unis
3. [Module level]: Mapping between modules of each school, compare by content similarities (module description) score

Goals to meet before next meeting:
Data collection: focus on modules related to DS (not on course)
- For each module: department, difficulty(1k/2k/3k/4k), module description(to extract key information), student feedback(key info), is core module of which major, prerequisites

Schools: NUS (melody), NTU (micole), SMU (ernest), SIT(liuchen)(yu jie), SUTD (boyuan)(byeo), SUSS (ying xin), focus on DSA **only**

Possible way to collect data: school web, reddit, interview,

5 Mar 21:00 - 22:15(Front + Back)
Standardize format of uni courses:
- CSV file
- **Mod code, mod name, mod level (1/2/3/4/5)**, MCs, **department**, workload, preclusions, prerequisites, semester offered (still want?) (1/2/1,2), S/U option available(still want?) (0/1), **mod description**, final(T/F), project(T/F) , **URL Link, topics**

Wed (8th Mar): self research due
Sun (12th Mar): interview due (email/ask freshie) & job research
Front end: framework and skeleton for the web (not urgent)

Backend inputs
1. Needs of job market (whether the course is aligned to job description)
   a. Score on whether university course meets the needs and teaches the skills required for jobs, recommendation system
   b. Link: https://www.sciencedirect.com/science/article/pii/S2666920X21000369
   c. Model to implement: Name-Entity Recognition to identify names of technical skills from text

   Possible way to present this:
   ● Multiple/single selection for skills and match with school course
   ● Roadmap/mindmap kind of thing(demo: https://www.imda.gov.sg/cwp/assets/imtalent/skills-framework-for-ict/index.html)

2. Topic Modelling
   a. Percentage of stats/cs/maths/UEs for each school's course
   b. Model to implement: Latent Dirichlet Allocation (unsupervised)

   Possible way to present this:
   ● Selection based on school to show the model
   ● Maybe can compare unis → similarity scores

Data we need:
- Specific skills needed for certain jobs
- Job sites for job descriptions
  - Linkedin
- Starting pay
  - https://docs.google.com/spreadsheets/d/1QtC8efWw0mVkGXW4QA9bX4f0nJhGbmFqCfVLSumMZ0I/edit#gid=446994348
  - https://app.techinterviewhandbook.org/offers?companyId=&companyName=&jobTitleId=&sortDirection=&sortType=&yoeCategory=

Github repo structure
- Front end, back end separated

- Commit the project meeting record regularly

Next meeting: Sunday 9am (tentative)
- Scrape job sites