Background and Introduction

- Overall goal: To identify which courses fit the Data and Information Literacy initiatives from the CMU Core Competencies Initiative
- Data Literacy: Evaluating and using existing data, Generating and evaluating new data
- Information Literacy: Identification and navigation of the information landscape, Critical evaluation of information, Information engagement in and across communities of practice, Shaping the information ecosystem
- Dataset: CMU courses from 2019-2022 - course name, department, ID, description, etc.

Data Processing

- To classify courses, it is important to analyze the description of courses.
- We eventually found some “uninformative” descriptions:
  - Empty string, N/A
  - tba, tbd, etc.
  - no description, etc.
  - fall and spring: 0 units
  - course description will be available on heinz college website
  - course description coming soon
  - see university of pittsburgh course catalog for description

- To contrast, there is an informative description:
  - “No course description provided.”

- We also noticed duplicates might cause bias in our modeling. Thus, the final step to the data preprocessing is to filter out courses with the same course ID and course descriptions.

- We found 8698 distinct courses, and after filtering out other uninformative courses, we finally found 7562 distinct courses with informative course descriptions.

Conclusion

- We were able to map information and data literacy skills (as identified by the campus-wide working group) to CMU courses via text mining and topic modeling.
- Using keywords and phrases from the competency definitions, we identified course topics with positive and negative prevalences to information and data literacy, and specified course levels and course departments that contribute the most to CMU students’ information and data literacy development.
- We hope the results from the course landscape scan would help with better course design at CMU in the future.