BAIS:6100 Text Analytics

Final Test

Kang-Pyo Lee

Course Schedule

Week	Date	Topics	Due
1	Jan 28	Introduction to Text Analytics Introduction to Python, Jupyter Notebook, and UI Interactive Data Analytics Service (IDAS)	
2	Feb 4	Module 1. Python Basics for Text Processing, Part 1 : Strings, Collections, Built-in Functions, Flow Control, and User-Defined Functions	
3	Feb 11	Module 2. Python Basics for Text Processing, Part 2 : Files, Dataframes, and Pattern Matching Using Regular Expressions	HW 1
4	Feb 18	Module 3. Basic Natural Language Processing (NLP) Techniques : Tokenization, Part-of-Speech Tagging, Stemming, Lemmatization, N-grams, Noun Phrase Extraction, Language Detection and Translation, and Gender Prediction Module 4. Keyword Analysis and Visualization	HW 2
5	Feb 25	Test 1	HW 3 (Feb 24)
6	Mar 4	Modules 5 & 6. Text Data Collection Using Twitter APIs and Web Scraping Group Project Announcement	
7	Mar 11	Module 7. Document-Term Representation Module 8. Text Classification	HW 4
8	Mar 18	Module 9. Text Clustering and Topic Modeling	Project Proposal
9	Mar 25	Module 10. Text Similarity Module 11. Keyword Network Analysis	
10	Apr 1	Test 2	HW 5 (Mar 31)
11	Apr 8	Group Project Presentations and Course Wrap-Up	Project Deliverables

Final Test

- You can ask the instructor questions for clarification using the chat box
- Leave the speaker on during the whole test in case the instructor may have something to announce
- You may leave the Zoom session when you get finished, but you cannot resubmit it to ICON once you have left the meeting
- There will be no official break during the test; feel free to use the bathroom as needed

Group Project

- **Submission**
  - **All deliverables must be placed in your shared group folder on IDAS by 6 pm on Thu, Apr 8 (Do NOT submit anything to ICON)**
  - **Create a text file named *readme.txt* in the shared group folder, which states**
    - **the paths to and names of the deliverables including the three Jupyter notebooks, the folder with +3,000 HTML files, and the CSV file**
    - **if any, the list of Python packages that were used in the project but not covered in class**
- **You are NOT allowed to make any change to the deliverables after the deadline**

Group Project

- **Presentation**
  - **Each of the 10 groups will have 10 minutes for presentation (try not to exceed 10 minutes)**
  - **Presentation order will be randomly determined**
  - **Have one of the group members take the role of sharing the screen on Zoom and controlling the presentation, while all members speak in turn**
  - **You do not have to present every little detail; try to focus on the data, outcome, and the issues/challenges**
  - **Every student is expected to watch all other presentations**
- **The presentations will be followed by a brief course wrap-up by the instructor**

For presentation

- No PowerPoint slides or Word report

- Only Jupyter notebooks!