Submission Deadline: Tuesday, March 4th, at 6:00 PM via iCollege

Project Overview:

For the midterm project, each group may select one of two available datasets:

1. FDD Item 5 Dataset (relatively easier)
2. FDD Item 12 Dataset

Each group is required to submit both:

* A written report (minimum of five pages)
* A presentation deck (approximately 15 slides)
* A python/Jupyter Notebook file

Each group should submit only one set of results. Individual submissions from group members are not required, as this would lead to duplicate submissions.

**Grading Criteria**

1. Coding Component (60 points)

Each group must complete the following tasks:

Select one dataset from the two provided options. (Note: The Item 5 dataset is slightly less complex.)

Choose one file from the selected dataset and generate three word cloud visualizations using: (10 points) (Use any one file among 50 files)

1. Bag of Words (BoW)
2. N-Gram
3. Term Frequency-Inverse Document Frequency (TF-IDF)

Apply at least one frequency-based method and one embedding method, conduct topic modeling, and provide an interpretation of the results. (10 points) **(Use all 50 files)**

Question Answering Tasks: (40 points) (Use all 50 files)

* + For the Item 5 dataset: Extract the initial fee from each file.
  + For the Item 12 dataset: Analyze whether the franchisee is offered an exclusive territory.

The final grading for this section will be based on the accuracy of the question-answering results.

2. Report Requirements (25 points)

Clearly explain the task and provide a summary of the dataset, including a description of the variables, their dimensions, and data types. (10 points)

Present and interpret the word cloud visualizations and topic modeling results. (10 points)

Display the question-answering results in a structured table and report the accuracy. (5 points)

3. Presentation (10 points)

Each group will be randomly asked two questions during their presentation, with each question worth 5 points.

4. Formatting Requirements (5 points)

Use your business question as the title of both your report and presentation. (1 point)

Include your group number and all group members' names (ordered alphabetically by last name) on the first slide and the first page of your report. (1 point)

Minimize the use of direct Python code screenshots in your slides and report. Instead, present outputs in tables and visualizations. (3 points)

Please ensure that your submission meets all the outlined requirements. Let me know if you have any questions.