DSBA/MBAD 6211: Group Project

Group Formation: **3-4 members max (Make sure that a majority is good with Python)**

Learning Objectives:
- To create an opportunity for students to work on real-world business analytics problems.
- To emphasize critical thinking, application of various analytics techniques, teamwork, and presentation skills.
- The dataset must be sizeable (# of records, # of variables)
- Project must provide business/economic/social insights

Grading:
- Novelty/creativity: originality of thought; surprising way of approaching the data
- Integration of multiple analytics techniques to gain meaningful insights into the data
- Integration of multiple data sets to yield new insights
- Clarity of the oral presentation

Project Deliverables:
- Submit everything via Canvas:
  - Original data sets
  - All the reports (in pdf format)
  - Presentation slides
  - Code

Project Deliverables:

Project proposal (10% -- 25p) deadline 2/9/24:

- 1 page (Time New Romans, single spacing, font 12)
- Sections to include**:**
  - project objectives,
  - data set,
  - proposed methods

Midterm report (20% -- 50p) deadline 3/15/24:

- 3 pages max (Time New Romans, single spacing, font 12)
- Sections to include**:**
  - Tasks completed,
  - Preliminary results/findings
  - Tasks still to complete
  - Any issues, etc

Final report (40% -- 100p)* deadline 5/6/24:

- 4 pages max (Time New Romans, single spacing, font 12)
- Sections to include:
  - Project summary (introduction and objectives)
  - Detailed analyses processes and steps (justification of method used and soundness of the results)
  - Results/ findings
  - Suggestions for additional work

Presentation (30% -- 75p)* deadline 5/6/24:

- Slides must be submitted the day before the final presentation and the submitted version will be used during the presentation time.
- Highlight selected analytics techniques and major findings.
- Each group member must present.
- Each group will have 15 minutes for the presentation.

*Final report and presentation grades will be weighted by peer review score.

Project Example 1: Post-Pandemic Airbnb Markets

Please choose **one city** from the following list and analyze Airbnb data in the selected city and explore and investigate the local Airbnb market in the past 12 months.

Cities:
- Asheville, NC
- Cambridge, MA
- Columbus, OH
- Denver, CO
- New Orleans, LA
- Portland, OR

Time Range
- All available data from the website (the last 12 months)

Data:
- http://insideairbnb.com/get-the-data.html
- Please download quarterly detailed listing, calendar, and review data.
- Additional data sources, such as census data and COVID-19 tracking data can be considered

Questions:
- Please explore and analyze the selected city's Airbnb market in the past 12 months. Please propose and design a comprehensive list of matrices, such as:
    - Market supply: the number of active listings
    - Market demand: occupancy rate, average monthly reviews per listing, etc.
    - Customer comments: topics, sentiments, etc.
- Do you observe any trend, seasonality and neighborhood differences in your local market?
- What factors affect Airbnb listings' performance?
    - Please propose a reasonable indicator(s) to quantify individual listings' performance.
        - For instance: average monthly occupancy rate, average monthly income, and/or average monthly reviews
    - Please identify suitable independent variables based on the available data sources.

Project Example 2: Analytics for Drug Safety and Compliance

FDA plays a critical role in protecting public health by ensuring the safety, efficacy, and security of drugs. Through its inspection and monitoring systems, the FDA identifies potential risks associated with drugs and takes necessary actions such as recalls. Understanding patterns and insights from this data can help in predicting future drug recalls, improving compliance, and overall better drug safety.

Data:
- https://datadashboard.fda.gov/ora/index.htm
- Use the API to fetch the latest datasets.

Questions:
- Analyze trends and common reasons for drug recalls.
- Investigate patterns and outcomes in inspection classifications.
- Evaluate compliance rates and reasons for non-compliance.
- Identify factors influencing drug safety and compliance.
- Predictive modeling of a drug being recalled, or a firm going through a drug recall.

Project Example 3: Kaggle Machine Learning & Data Science Survey (Adapted from a Kaggle Competition)

Since 2017, Kaggle set out to conduct an annual industry-wide survey that presents a truly comprehensive view of the state of data science and machine learning. The results include raw numbers about who is working with data, what's happening with machine learning in different industries, and the best ways for new data scientists to break into the field.

Please combine and dive deep into six years' survey data, and understand various stories of data scientists from around the world
- You can choose to drop the earliest survey data, if you think it was too different from later years.

Data:
- 2017 Kaggle Machine Learning & Data Science Survey:
  https://www.kaggle.com/kaggle/kaggle-survey-2017
- 2018 Kaggle Machine Learning & Data Science Survey:
  https://www.kaggle.com/kaggle/kaggle-survey-2018
- 2019 Kaggle Machine Learning & Data Science Survey: https://www.kaggle.com/c/kaggle-survey-2019/data
- 2020 Kaggle Machine Learning & Data Science Survey: https://www.kaggle.com/c/kaggle-survey-2020/data
- 2021 Kaggle Machine Learning & Data Science Survey: https://www.kaggle.com/c/kaggle-survey-2021/data
- 2022 Kaggle Machine Learning & Data Science Survey: https://www.kaggle.com/c/kaggle-survey-2022/data

Questions:
- Please provide overviews to understand data scientists from around the world
- What changes, if there is any, have you observed from the four years' survey data?
- Please explore regional differences in data scientist communities
- What is the value of higher education degrees in data science?
- What factors affect data scientists' salaries?
- Please identify a specific group of data scientists, and tell a story about them
  - A "story" could be defined any number of ways through a combination of both narrative text and data exploration
  - You can define the sub-community in any way, such as female data scientists, data scientists from a specific region, or data scientists from a specific field etc.