ISQS 6350 Multivariate Analysis

Group Project Instruction

For the project, you will be required to find a dataset then ask some questions about the data and perform multivariate data analysis techniques to answer those questions. You will work in groups of three to four. Each group member should contribute to the project and should be familiar with all aspects of the project. A group grade will be assigned. Additionally, feedback from team members will be used to adjust the group grade for individuals.

Objectives

By the end of this project, students will be able to:

1. Describe how to use multivariate analysis method on a real dataset.
2. Perform basic data cleaning techniques on a real dataset.
3. Generate visualizations for multivariate data.
4. Create insights about the data by performing dimension reduction techniques.
5. Create insights by performing cluster analysis techniques to discover structures and patterns in high-dimensional data.
6. Develop a confirmatory factor analysis to discover the latent variables behind a set of variables.

Milestones

- **Project Introduction:**
    - Create a team of three to four students.
    - Find a dataset of your interest. This data set must mostly include numerical data and multiple variables (columns).
    - Read the project rubric, and make a plan about how you want to break the work between team members through the semester.

- Write a brief report to introduce your dataset and cite where you found it. The team must clearly explain the motivation behind studying the data. This is related to item 1 (Introduction) in the project rubric (page 4). This will be the introduction of your project and you can revise it in your final report.
- Follow the project format guideline (page 3).
- Submit your introduction draft as a e-copy: .docx or .pdf
- This update will not be graded, but your report might get accepted or rejected. If your dataset is not appropriate for multivariate data analysis, then you have a second chance for submission.

- **Project Presentation:**
  - You must submit your presentation .ppt or .pdf file by 11:59 pm of the night before the presentation.
  - You have 10 minutes to present your work. There will be 5 minutes for Q&A and transition.
  - All team members must participate in the presentation.
  - Your presentation grade counts for 20% of your project.
- **Project Final Report:**
  - Construct your final draft report based on the items listed in the rubric (page 4).
  - Follow the project format guideline (page 3).
- **Team Peer Evaluation**
  - Each team member evaluates his teammates in terms of communication, reliability, and contribution.

Project format guideline:

1. You are graduate students! That means that you should be proficient writers. Your submission should be completely free of grammatical errors, misspellings, and misused words.
2. Create a cover page and make an interesting title for your project.
3. Start each section (or subsection) with an introductory paragraph. Do not start sections with core reports, charts, or tables.
4. The tone and style should be suitable for a business report or academic paper. There should be a balance between text, charts, and tables.
5. Use 12-point font, 1-inch margins, and 1.5 or double spacing.
6. Include tables or charts in the body of the paper, not at the end. All charts and tables must be captioned with an appropriate description.
7. Every code or large data tables must go to the appendix.
8. Font and formats must be consistent in the entire report.
9. Provide citations for any paraphrased factual claims or resources used to support your arguments. Plagiarism will not be tolerated.
10. The author of each section must be mentioned. For example, *this section is authored by Ryan Miller*.

Project Report Rubric **Evaluator Name: …….**

	Unacceptable (0-5)	Below Average (6-9)	Average (10-12)	Excellent (13-15)	Grade
1- Introduction, The Problem Context	Team did not introduce the dataset and the problem context.	Team did not clearly describe the observations and variables of the dataset. Team did not clearly describe the motivation behind studying the data.	Team clearly described the observations and variables of the dataset but did not clearly describe the motivation behind studying the data.	Team clearly described the dataset and clearly described the motivation behind studying the data. Team provided scholarly citations or quantitative facts to describe the motivation.	
2- Data cleaning and visualization	Team did not describe their data cleaning and outlier removal process. Team did not present any visualization.	Team did not clearly describe their data cleaning and outlier removal process. Team did not present sufficient visualization.	Team did clearly describe their data cleaning and outlier removal process. Team did present sufficient visualization, but they did not extract new insight from the data.	Team clearly described their data cleaning and outlier removal process. Team presented insightful visualizations motivating to do further exploratory or confirmatory analysis.	
3- Dimension Reduction Analysis	Team did not apply any dimension reduction analysis.	Team did not apply dimension reduction analysis sufficiently and correctly.	Team applied dimension reduction analysis correctly but did not discuss the motivation sufficiently. Also, they did not provide insights into the results.	Team applied dimension reduction analysis correctly and discussed the motivation behind that. Also, they provided interesting insights into the results.	
4- Cluster Analysis	Team did not apply any cluster analysis.	Team did not apply clustering analysis correctly.	Team applied cluster analysis correctly but did not discuss the motivation sufficiently. Also, they did not provide insights into the results.	Team applied cluster analysis correctly and discussed the motivation behind that. Also, they provided interesting insights into the results.	
5- Confirmatory Factor Analysis (CFA)	Team did not apply CFA.	Team did not apply CFA correctly.	Team applied CFA correctly but did not discuss the motivation sufficiently. Also, they did not provide insights into the results.	Team defined appropriate latent variables and applied CFA correctly and discussed the motivation and provided insights into the results.	
6- Conclusion	Not done.	Team did not present a clear conclusion.	Team presented some meaningful findings but they did not discuss the pros and cons of their study.	Team presented meaningful findings and discussed the pros and cons of their study. Team suggested some future works for post analysis.	
	Unacceptable (0-4)	**Below Average (5-6)**	**Average (7-8)**	**Excellent (9-10)**	
7- Write-up and format	Very sloppy.	Team followed a few items of project format guideline.	Team followed the project format guideline partially.	Team followed the formatting guideline completely.	

Comments: