BYGB/ISGB 7967, Fall 2017, DATA MINING FOR BUSINESS – PROJECT PROPOSAL

1. You work in groups
2. Reports are by 5pm due date. Penalty for late submission is 10% for each day delay.
3. Each report must have a cover page with title, names, class, section, date and name of the course.
4. The group is responsible for the project, not individual group members. The report and presentation will be graded based on the '4 Cs.'
5. Each group will submit one copy of your Phase 1 report, which may or may not lead to a follow-up discussion with the instructor.

Completeness – is work complete?
Correctness – is the design technically sound? Is correct terminology used?
Consistency – is the report cohesive, or does it appear choppy? Is there flow and continuity?
Communicability – is report formatted professionally? Does report communicate design in easily understood language?

Phase 1: DESIGN PROPOSAL

1. Choose an application domain and find a publically available BIG data set -- This is better to be related to your current job or career of interest.
2. Useful data sources: UCI Machine Learning Repository, IPUMS, data.gov, other open datasets.
3. Identify and formulate a data-mining problem
4. Choose and apply appropriate data-mining technique(s) to solve this problem. You can compare the performance of multiple data-mining techniques.
5. Your design proposal is in the form of power point slides and a short executive summary (5 pages maximum, double-spaced).
6. Components of proposal:
    A. Title/ team members (1 slide)
    B. Abstract (1 slide)
    C. Introduction: what is the topic, why it is important and interesting, & brief background material (2-3 slides)
    D. Data Description: data source, processing, descriptive stats, variables, etc. (1-2 slides)
    E. Problem Statement: what exactly do you propose to do? (1 slide)
    F. Methodology: outline steps to execute your project –data collection, data preprocessing, variable selection, model building, and evaluation. Have a diagram to show your overall methodology (1 diagram + 5 slides)
    G. Expected results & conclusions (1 slide)
    H. Project timeline (1 slide)
    I. Executive summary (5 pages maximum, double-spaced)

Note: The group should consider multiple options for data sets/problem statement and choose one collaboratively. Get prior approval from instructor before drafting proposal.