BUAN 690 Applied Analytics Practicum Project

Introduction

The purpose of this course project is to give you an opportunity to utilize data analysis knowledge and skills using Python, and so integrate Python applications for effective presentation of data analysis in decision making process. The scope and contents of the project are flexible and are largely based on your interests and course objectives. The project should demonstrate your understanding and mastery of the programming tools, and analytical methodology taught in classes, as well as your ability to apply them. Recommended project topics are Stock Market Analysis and Prediction, Sentiment Analysis of Financial Markets, Visualization of Financial Data

Data Selection

You may use any preexisting archival data. **The dataset should have at least 100 observations.**
Some popular common sources of datasets are

- UC Irvine Machine Learning Repository - http://archive.ics.uci.edu/ml/index.php
- U.S. Government's Open Data - https://www.data.gov/
- Kaggle Datasets - https://www.kaggle.com/datasets

The suggested structure of course project report

- Title and Authors
- Research questions : State research questions you want to answer with your data and why these questions are of interest or are important to answer. State any hypotheses you have about what you expect to find and why.
- Motivation and background of your research questions : Explain your motivation for your research questions and why they are worth for your analysis. Briefly describe what research has been done before on this question.
- Dataset : Explain the data set you use for your project (ex : source and information about the dataset). You are not allowed to use a dataset that was used for lecture and homework in any course you took in the program. Describe all variables in your dataset (ex: level of measurement, qualitative or quantitative, unit of measurement, etc.)
- Data Analysis: Statistical or Machine Learning Analysis with appropriate Exploratory Data Analysis are required. State what methods you use for your data analysis and why you use them. You also must provide programming codes for your analysis with appropriate comments and explanation. Present all outputs, tables, plots, and calculations with appropriate interpretation and comprehensive explanation.
- Results and Conclusions : Present, interpret and discuss your analysis results and answers for your research questions. Make conclusions about your research questions and implications if any. Evaluate the strengths and weaknesses of your project. Discuss the limitations of your project and the direction of your future projects, etc.
- References : Cite any sources or information you used for your project.

BUAN 690 Applied Analytics Practicum Project

Timeline

- Group Choice (9/23)
  - Form your group for Course Project of up to 4 members and report on the Group List that is posted on the BB.
- Project Data Selection (10/7)
  - Select your project data and report on the Group List that is posted on the BB.
- Midterm report (10/28)
  - Submit your report about the progress of the project such as research goals, methodology, algorithms used for the project, and experiment setting, etc
- Project Presentation (TBA in the week of 12/02)
  - Summarize your project report as a Power Point file and present your findings in the class. You should follow the structure of the report that is described below.
- Final Report (Saturday, 12/02, 11:59 PM)
  - Submit one copy of the final project report per each group.
  - It should show all participating members' names.
  - It should be a MS-Word file that includes all works such as **programming codes, plots, tables, and detailed analysis** with **appropriate interpretation**.
  - There is no limitation on the number of pages for the report.

Grading of the final report

The point distribution for the project presentation and report will be as follows:

- **Research Question, Data selection (15%)** - Quality of research ideas and data selected. Is the project idea of theoretically interest? Are the selected data related to the research questions being asked? Does it appear that time and effort went into the planning of the project?
- **Statistical\Machine Learning Data Analysis and Programming Procedures (30%)** - Are the analyses of data accurate? Is the adopted methodology statistically meaningful and appropriate? Are statistical analysis procedures carried out and interpreted correctly? How sophisticated are the Python programming skills for the data analyses?
- **Data Display and Visualization (30%)** - Includes appropriate, well labeled, accurate displays (graphs and tables) of the data. Is the presentation of the data clear and concise? Is the visualization of analysis effective? Are graphs, tables, analyses, and any attachments neatly presented? Is the visualization of analysis appropriate for your research questions?
- **Conclusion and Reflection on process (15%)** : Conclusion includes a clear answer to the research question that is consistent with the data analysis and the method of data collection. Gives a good overall picture of the project—what went well and what didn't—and includes ideas for further study. Are implications of the findings thought through? Are the limitations of the study recognized? Is the project persuasive?
- **Overall Presentation (10%)** - Attractive, well-organized, well written project report.