BUS111 Business Analytics
Fall 2018

Team Project: Investigating Data & Producing Managerially Relevant Decisions

Project Descriptions

Introduction: It is common in business analytics to start with data previously collected and examine it in order to extract managerially relevant observations and improve decisions. In case a prior team has worked on the problem, it is necessary to (a) work through it to understand their approach and (b) attempt to improve their model performance. It is also commonplace to do such work in a team setting.

Objectives:

- Apply your business analytics knowledge to understand a descriptive or predictive model created by others.
- Think critically about a model for the purpose of improving it.
- Leverage R, RStudio, GitHub and similar tools to collaborate on a project.
- Develop and present a clear and persuasive narrative explaining the business value and operation of your revised model.

Summary: Working in teams, you have selected a large dataset from the options below. Your assignment is to prepare an RMarkdown document and also deliver a PowerPoint-supported presentation in class that:

a. Explains the practical questions posed by the supplier of the data (your *client)*,
b. Explains the essential content of the dataset so that the class understands it,
c. Explains the model chosen by your team, and in case a prior team has worked on it (in the form of a Kaggle data kernel), explain the model used by them and how you improved it to answer the original questions.
d. Explains the conclusions that the client should draw from your analysis.

Datasets: You can use one of the following four datasets for your project –

a. Synthetic Financial Data from PaySim for Fraud Detection (https://www.kaggle.com/ntnu-testimon/paysim1/home)

b. Analyzing Black Friday Purchases for customers and product categories (https://www.kaggle.com/mehdidag/black-friday/home)

c. Analyzing the viability of Kickstarter projects (https://www.kaggle.com/kemical/kickstarter-projects/home)

d. 2015 Flight Delays and Cancellations – which airline to fly on? (https://www.kaggle.com/usdot/flight-delays/home)

Specific requirements and guidelines:

- You must use one of the datasets that your team selected. Consult with me if you decide to change datasets. You are permitted to supplement it with other publicly available data if you deem it appropriate.
- You should create a private GitHub repository to share the Project with your team members, and use RStudio (and/or SourceTree) to track versions and contributions of different team members. That project will contain your R Markdown code and knitted files.
- Since you are choosing Kaggle datasets, you must first get the original Kaggle-supplied kernel to run. As part of your report and presentation, you should clearly cite the Kaggle kernel(s) your team used, and interpret the output of their code.
- You must, in some fashion, make a *substantial modification* to the original code (if any) to better respond to the supplier's questions. This might involve selecting different explanatory features, transforming data to accommodate non-linearity, using a different modeling approach, creating a more informative visualization, or other similar changes. Improvements should go beyond simple cosmetic changes or dropping a variable from the model. If you're uncertain whether a particular change is substantial, speak with me.
- Your R Markdown file should contain relevant code as well as narrative explaining the different parts of your investigation, explaining what each code chunk does and why you did created the chunk of code. As with earlier assignments, knit the markdown file to a Word doc, then **resave** that document as a pdf for upload to LATTE. **Do not** make changes to the knitted Word file. The RMarkdown code should correspond directly to the knitted outpout.
- ***The first sentence of the markdown document should say: "The URL for our Team GitHub repository is {your unique URL}." The document should also include the URL to the kernel and Kaggle dataset.***
- The PowerPoint presentation should include highlights of your work that clearly address items *a–d* listed in the Summary above. Note that R Studio also has the ability to create presentation slides much like creating a Notebook or Markdown document. If your team wants to investigate that capability, go for it!

Deliverables:

There is a LATTE Assignment page, and each group must submit the following:

1. Your knitted RMarkdown Doc/PDF
2. A pdf of your PowerPoint slides (or equivalent). Let's avoid the Windows/Mac compatibility issues. No Prezi please.

When I visit your Team's GitHub repository, I should find code with multiple versions, the raw data, and the original R kernel (if any). IF you used additional data, I should also see that in your repo.

For detailed description of these datasets, please read the associated webpages.