General Assembly

Final Project Guidelines

DS-SF-42

Objective

- Conduct a project in which you apply machine learning to a dataset of your choosing.
- Your goal is to solve a problem and/or advance the understand of an issue. What are you trying to do predict? Why is that important? What relationships and other phenomenon are you trying to better understand
- Apply the tools and concepts you've learned in this course.
- Can either be classification or regression. Or can do an unsupervised learning project in which you use clustering.

Presentation Expectations

- You will be allotted up to and no more than 8 minutes for your final project presentation.
- Conduct your presentation as you would if you're doing this at work.
- Assume your audience is data science knowledgeable, so you can be technical but don't be *too* technical.

Outline

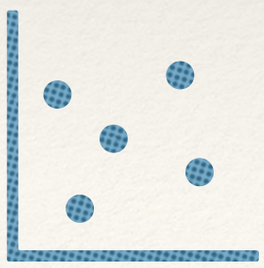
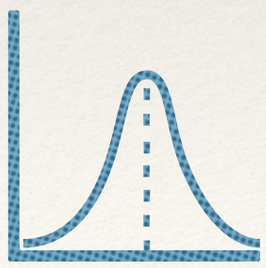
- I. Introduction
- II. Exploratory Data Analysis
- III. Modeling Process
- IV. Modelings Results
- V. Conclusions/Further Work

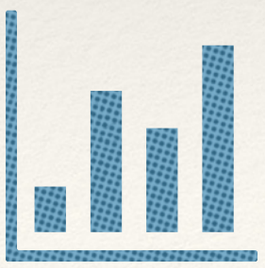
I. Introduction

- What is the problem/issue you addressed?
- State your objective
- Tell us about the dataset. Where did it come from? How did you it acquire it?
- Provide context about this dataset. What is the value?
- Why did you choose this project/data?

II. Exploratory Data Analysis (EDA)

- CHARTS CHARTS CHARTS!!
- Use this section to further acquaint the audience with your data.
- Tell us what you learned from your EDA process.
- How did this part of the project help inform you which features are important?



III. Modeling Process

- Walk us through the modeling process. Which models did you choose?
- Tell us what you did to improve your models. Did you drop or add features? Transform your data in any other way.
- What are the best features from your dataset?

IV. Modeling Results

- What is the best model you derived? Which algorithm did you use? And with which parameters?
- What was the final set of features you used?
- Which metric did you use to evaluate your model? (accuracy, recall, precision, roc_auc, etc…)
- How well did your model perform? Whatever metric you use, make sure it is cross-validated. Use charts to show your results, definitely include a ROC curve if you're doing classification.

V. Conclusion and Further Work

- Tell us what you learned about the issue you're project is about.
- Did you experience a grande revelation about your data? If so tell us.
- Was your hypothesis correct? Tell us why or why not
- Is there any unfinished business? If you had more time or data, what would you do?

Guidelines

- You're going to have a lot more information than can be fit into a 8 minute project. Think critically about the content you want to include and *not* include as well.
- No code! It doesn't happen in the real world, so please don't put it in your presentation. And no screenshots of pandas tables.
- We're 16 days away from final presentation day, so please budget your time wisely.
- Use Stewart and I as much as possible for help. Come to office hours.

MOST IMPORTANTLY

HAVE FUN!!!

???QUESTIONS???