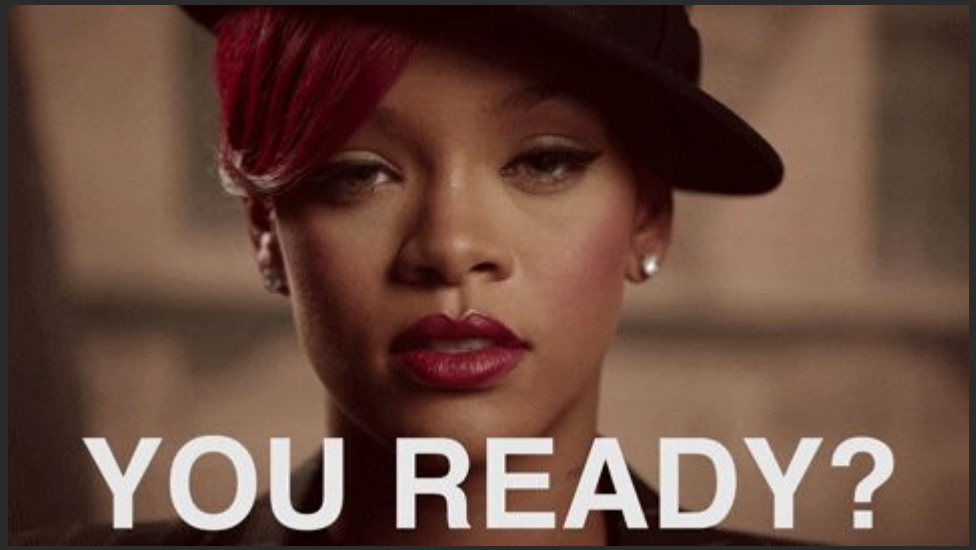
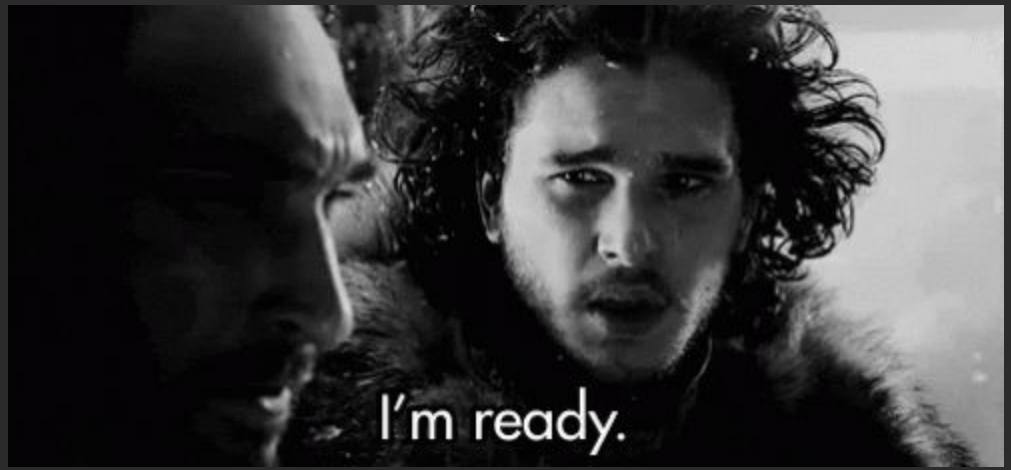
Module 3 Project Intro



Amber Yandow
05/18/20

// FLATIRON SCHOOL

If this is you… proceed to deliverables



Deliverables

What is the project?

Your goal is to identify a problem that you can solve with supervised learning(classification) using a dataset from the curated list or a dataset that is pre-approved by Amber. You must clean, explore, model and interpret your findings.

What do you turn in?

- ***Non-technical presentation(pdf)***
- ***Technical Notebook(jupyter) including detailed README***
- ***Blog post***
- ***Recorded non-technical walk-thru(youtube)***

Specifics

What dataset should you choose?

1. *Chicago Car Crash Data- predict the cause of a car accident. BEWARE this is a multi-classification problem(over 40 classes)*
2. *Terry Stops Data - predict whether an arrest was made or not. Students have struggled with this one in the past.*
3. *Customer Churn Data - predict whether a customer is likely to churn. Popular dataset to work with.*
4. *Tanzanian Water Well Data - Ternary classification for predicting the condition of a water well. Students have enjoyed this one.*
5. *Pick your data! - Must work for classification; can't be a dataset we've used before; should be at least 1000 rows/10columns; send to me for approval*

Non-tech Specifics

Non-Tech Presentation

Should Include:

- ***A slide for each of the following: Problem/business question, methodology, results, recommendations, future work and thank you slide***
- ***Visuals should be relevant to the business question, properly labeled without too much text***
- ***You will have to talk about about some technical terms, try to be concise.***
- ***Keep it between 5-8 minutes***

Technical Specifics

Technical Presentation

Should Include:

- ***A well documented and orderly notebook with comments and docstrings where appropriate. It should follow your chosen methodology(OSEMN, CRISP-DM, ROSE-MED)***
- ***Have an intro/conclusion***
- ***Interesting visuals that properly show the relationships of predictors***
- ***An X-factor, something not taught in the curriculum. This can be a python package or method, a modeling technique, web-scraping/API calls.***

Resources

For finding data:

- UCI Machine Learning Datasets Repository
- Kaggle Datasets
- Awesome Datasets Repo on Github
- New York City Open Data Portal
- Inside AirBNB

Previous Projects:

- ***Google Doc with previous student's Mod3 Projects***

Questions?