**Overview**

This is the time to dive into the data science process using techniques we’ve learned, and really put your skills on display via a self-directed project.

It is entirely up to you to select both the the data you will use and question to answer.

**Make sure that your check the quality of your data before you invest a lot of time cleaning and working with it!!**

**Dataset**

* You will be using a dataset of your own choice.
* You can use API's, webscraping and available open data repositories to get your data. Some suggestions are:
  + [UCI ML Repo](https://archive.ics.uci.edu/ml/index.php)
  + [Kaggle](https://www.kaggle.com/datasets)
  + [MLdata.io](https://www.mldata.io/datasets/)
  + [Wolfram Data Repository](https://datarepository.wolframcloud.com/)
  + [World Resources Institute](https://datasets.wri.org/dataset)
  + [Wikipedia - Kaggle ML](https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research)
  + [Spotify API](https://developer.spotify.com/documentation/web-api/)
  + [Yelp API](https://www.yelp.com/fusion)
* When you choose your dataset please be aware of the following pitfalls:
  + Before spending too much time for cleaning the dataset make sure that it's good for the problem you would like to solve. So first make plan and then invest time and effort.
  + **Stay aware from time series data. We didn't studied the techniques for handling time series data.**
  + Make sure that you find a dataset that is not studied extensively online. For example: No Titanic, House Price, Iris, MNIST etc. datasets.
  + If you don't have a compelling reason don't work with a small dataset (at least 1000 rows and 10 features.)

**Deadlines:**

1. Submit the completed deliverables on December 8th at 9:00 AM.

**Deliverables**

For this project there will be two deliverables:

1. A Github repo that contains your project.
2. A Jupyter Notebook (I will refer to it as technical notebook or report).

**Github Guidelines**

* Put an attention grabbing title to the repo.
* Github should include a license.
* ReadMe should be included.
* ReadMe should include these sections:
  + Overview
  + Goals
  + Data
  + Table of Content
  + Packages and Software

**Report Guidelines**

This will a Jupyter notebook and it will be uploaded to your Github repo.

Should include the following the sections:(Make sure that each of them explained in a separate section)

* [ ] Frame the problem and look at the big picture
* [ ] Getting the data
* [ ] Explore the data to gain insights
* [ ] Prepare the data to better expose the underlying data patterns to ML algorithms
* [ ] Explore many different models and short-list the best ones
* [ ] Fine-tune your models
* [ ] Present your solution
* [ ] Limitations and later work
* [ ] References and contributions

For more details about the each section and what you might want to include in each section please take a look at this [check-list](https://github.com/ageron/handson-ml/blob/master/ml-project-checklist.md).