# Introduction:

The purpose of this document is to propose ideas for the Capstone Project for Data Science Intensive Workshop.

A Capstone Project has to be completed as a part of Data Science Workshop. Coming up with ideas for a Capstone Project has been slightly more challenging than I thought. This matter was discussed initially with the mentor in the weekly calls and agreed upon to spend some time to look at some data-sets. The mentor had suggested the following datasets.

* <https://github.com/caesar0301/awesome-public-datasets>
* <https://docs.google.com/spreadsheets/d/1wZhPLMCHKJvwOkP4juclhjFgqIY8fQFMemwKL2c64vk/edit#gid=0>

I was not sure if Kaggle challenges were accepted as a Capstone Project or not. However, the mentor confirmed that they are accepted as Capstone Project.

After considering various data-sets and some of the Kaggle competitions, I would like to propose the following ideas for my potential Capstone Project.

# Predict whether a mobile ad will be clicked

This is a completed [Kaggle competition](https://www.kaggle.com/c/avazu-ctr-prediction). The underlying idea is to predict if a mobile ad will be clicked or not. Click-Through-Rate (CTR) metric is used for evaluating ad performance and CTR systems are widely used by internet economy. This problem has been solved before, but would be refining for coming days. For learners, it provides great opportunity to try a real world problem.

# Predict Facebook check ins

This is a completed [Kaggle competition](https://www.kaggle.com/c/facebook-v-predicting-check-ins). This competition was used as Facebook recruitment exercise. The scope of the project seems to be fitting with Capstone Project requirement. This is not solving a real problem but it provides great learning opportunity in a controlled environment. It would be good to learn and solve an interview problem.

# Classify species of Fish

This is a [currently active](https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring) competition at Kaggle. The idea of competition is noble. Currently, the compliance activities are manual and error prone. By automating the image processing and classifying the species of fish, the compliance process can be more accurate which will help saving threatened marine ecosystem. Also, this project involves image process and probably deep learning which is where I want to focus in future. The scope of the project seems to be more than the expected Capstone Project.