

## Capstone Project: Machine Learning Overview

Framing my project as a machine learning problem, the main question I am attempting to answer is:

- Based on several census survey questions, what is the income level of each survey participant?

This problem is a classic example of supervised learning. The form of supervised learning used will be classification. Participants' income level will be classified using several factors within the dataset. 'Income' is the dependent variable within the dataset, and will be used for training my model.

The main features that I will use are age, gender, and educational attainment. From my previous statistical analysis report, these three predictors are highly correlated to income level. Therefore, it would be ideal to include these variables within my model.

The machine learning technique I will use is logistic regression. This algorithm is an excellent solution for my problem and its dataset. I will evaluate the success of my machine learning technique through cross validation. Comparison between my model's output and the results from the original table is crucial for evaluating the performance of my algorithm.