## CS4431 HW6

## Team: Peter Strimbu, Megan Hoeksema, Samantha Reno, Patrick Ayers

## 2022-07-20

# Week 6 Design Document

## Introduction

**Intent of the application**

* The intent is to create successful regression models using various techniques to explore and identify influential variables in the California Housing Dataset.
* The main variable of interest will be median income.

**Dataset to be used, including source**

* California Housing Dataset
* Source: sklearn packages
* Original collection through the 1990 U.S. Census Bureau

**Use case**

* A group of young entrepreneurs are interested in exploring the effects of different housing variables on median income using various regression techniques.

## Dataset Analysis

**Define labels**

* There are 8 continuous variables plus the target in the California housing dataset.
* These include:
  + MedInc: Median income in block group
  + HouseAge: Median house age in block group
  + AveRooms: Average number of rooms per household
  + AveBedrms: Average number of bedrooms per household
  + Population: Block group population
  + AveOccup: Average number of household members
  + Latitude: Block group latitude
  + Longitude: Block group longitude
* The target variable is the median house value for California districts.

**Define variables**

* The target variable is the median house value for California districts, expressed in hundreds of thousands of dollars ($100,000).

## Inputs

## Proposed Libraries

**Libraries**

**Library source**

## Proposed Solution

## Proposed Outputs

## Proposed Visualization

## Conclusions