**Introduction**

* **Intent of the application:** Explore relationships within the California Housing dataset utilizing exploratory analysis techniques and regression models.
* **Dataset to be used:** California housing dataset from sklearn package
* **Use case:**  A group of students have decided to move from Florida to California after graduating college. They would like to know where the best place to live would be, so they have gathered data on California Districts to run analysis methods on. They would like to understand what district would be best for their income and age group before making the move out to California.

**Dataset Analysis**

* **Define variables**:
  + MedInc
  + HouseAge
  + AveRooms
  + AveBedrms
  + Population
  + AveOccup
  + Latitude
  + Longitude
* **Define labels:**
  + 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

**Inputs**

* **Data import:**
  + California housing dataset

**Proposed Libraries**

* **Libraries**
  + Sklearn
  + Pandas
  + pandas\_profiling
  + Numpy
  + seaborn
* **Library source**
  + Scikit-learn.org
  + Pandas.pydata.org
  + <https://github.com/ydataai/pandas-profiling>
  + Numpy.org
  + Seaborn.pydata.org

**Proposed Solution**

* Import the dataset and explore and display the features of the dataset.
* Explore the dataset using tools and libraries available in Python.
* Create visualizations of the dataset
* Perform regression techniques such as linear, ridge, and lasso by use of train-validation sets.

**Proposed Outputs**

* Linear regression model
* Ridge regression model
* Lasso regression model

**Proposed Visualization**

* Scatterplots
* 3D scatterplots of median incomes vs longitude vs latitude
* Histograms
* Pandas\_profiling
* Pairplot

**Conclusions**

* In our analysis, we will find that median income will be correlated to latitude and longitude. With our models we will be able to find the best one to predict certain outcomes.