CRIME RATE PREDICTION IN CALIFORNIA

GROUP 8

CONTENTS

Alexandre Lazzari

Edward John Tagaca

Eric Shih

Hok Yin Cheung

Marisol Cornejo

Part 2

GOAL

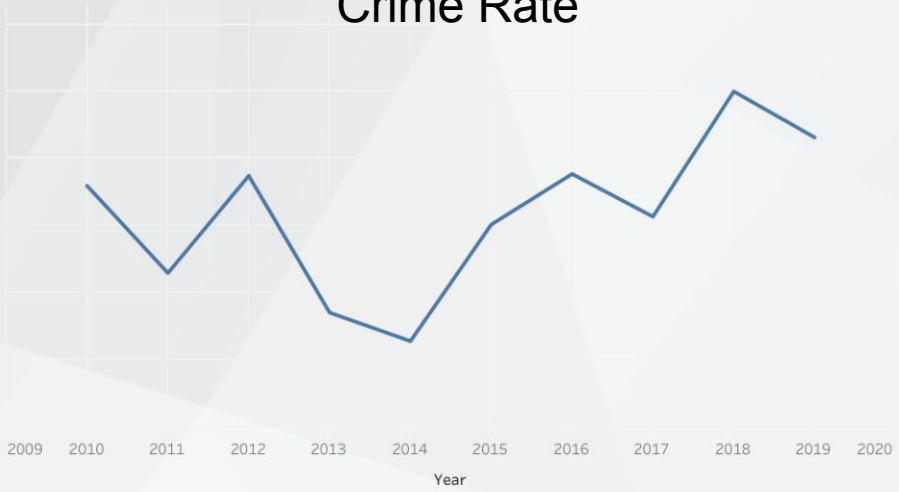
Goal

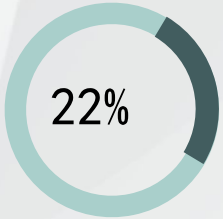
Our goal is to create a model to predict the crime rate based on the given data from the past (2010-2019).

Part 3

DATA

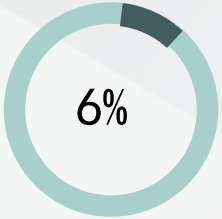
Data





Los Angeles

Los Angeles's crime rate is about 22%

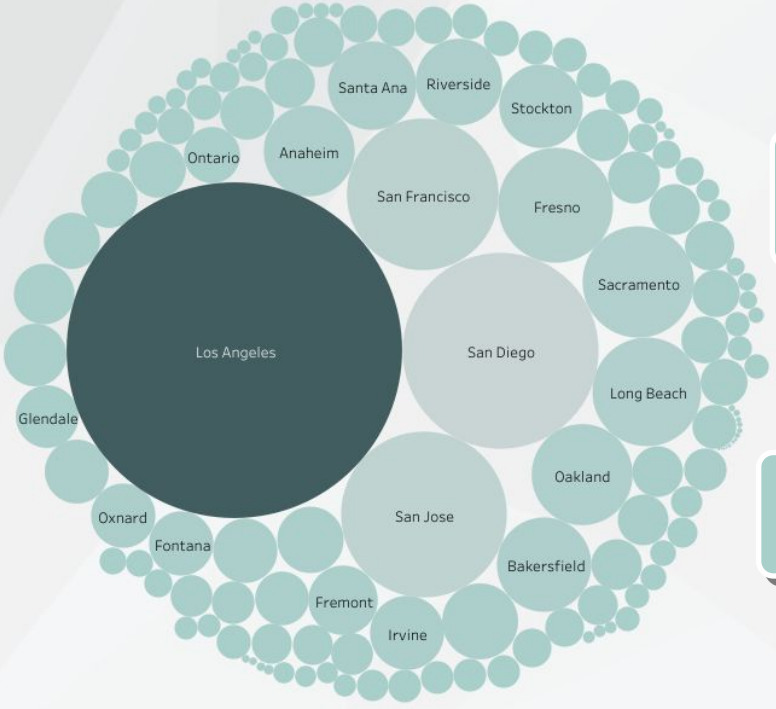


Oakland & San Francisco

Crime rate in Oakland & San Francisco are about 6%

Data

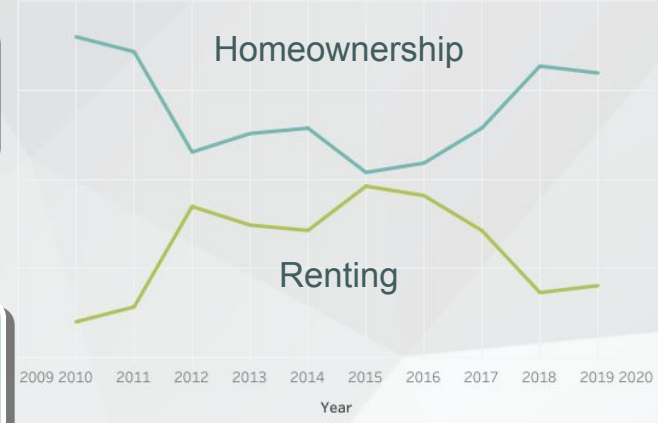
Total Population

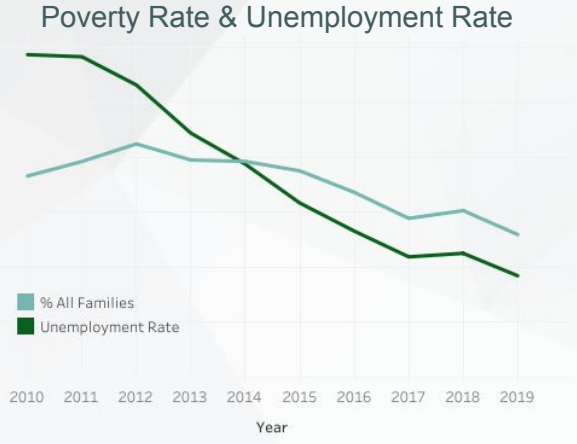


Largest population: Los Angeles

Homeownership & Renting show opposite trends

Poverty & Unemployment rates show decrease trend





Data

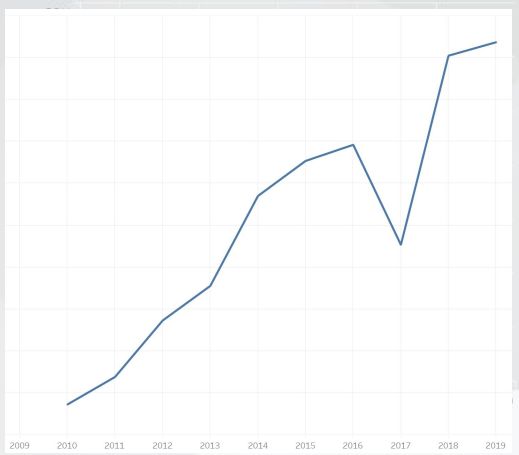
High school graduates, individuals with some college education (no degree), and individuals holding BA/BS degrees represent the largest segments.

Income levels are showing an upward trend

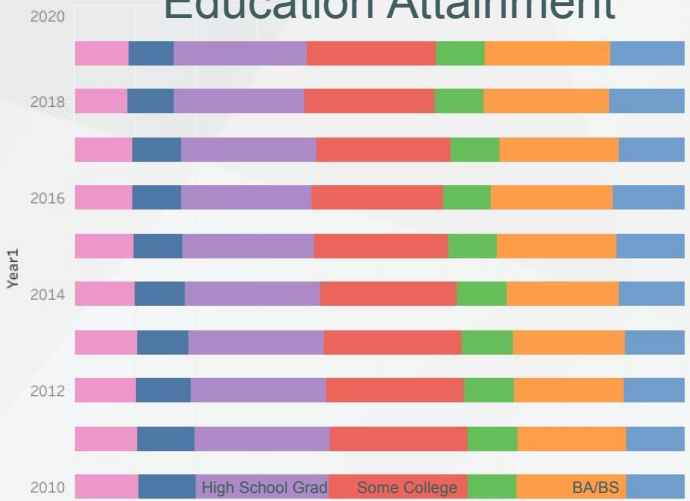
Majority of the population falls within the 25-54 age range

Average house price range from $148,000 to $2.5 million

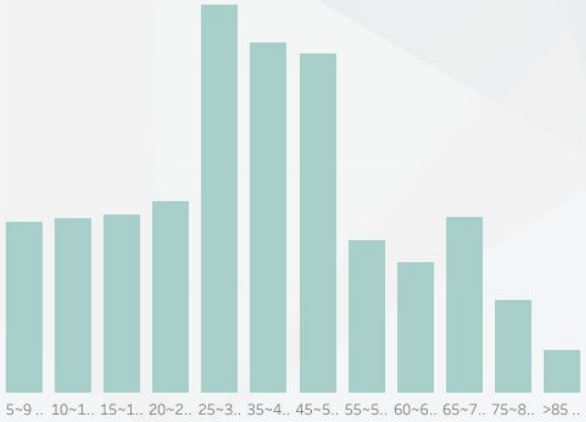
Income



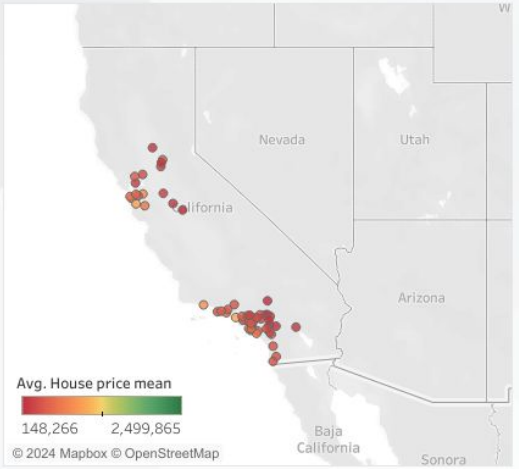
Education Attainment



Age



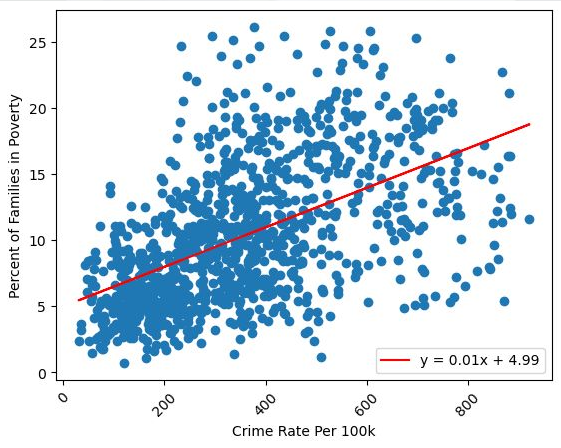
House Price



Part 4

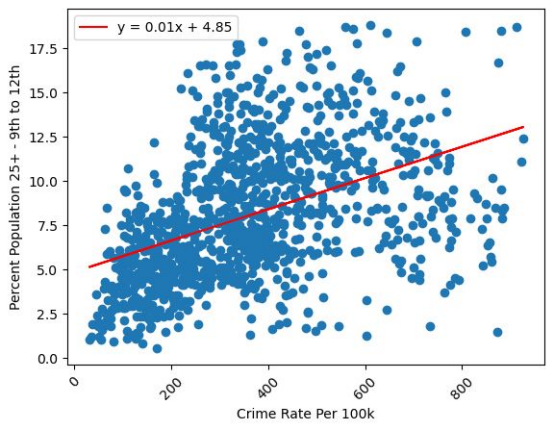
MACHINE LEARNING

Machine Learning - Linear Regression

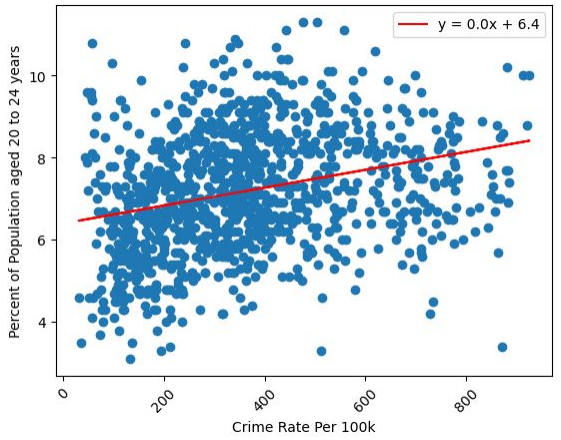


Poverty% & Crime Rate

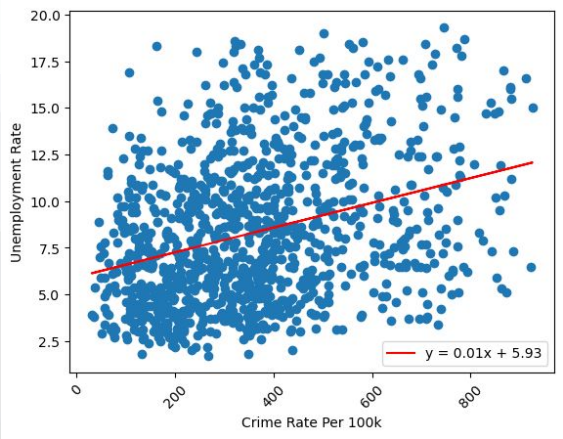
P-value: 1.76e-82



Population & Crime Rate
P-value: 2.18e-54

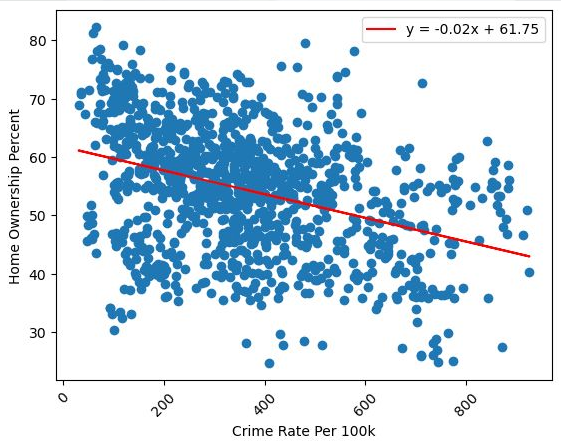


Age 20-24 & Crime Rate
P-value: 8.03e-22

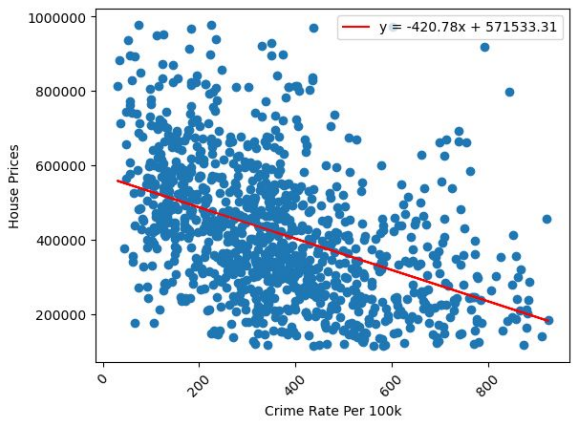


Unemployment & Crime Rate
P-value: 8.85e-29

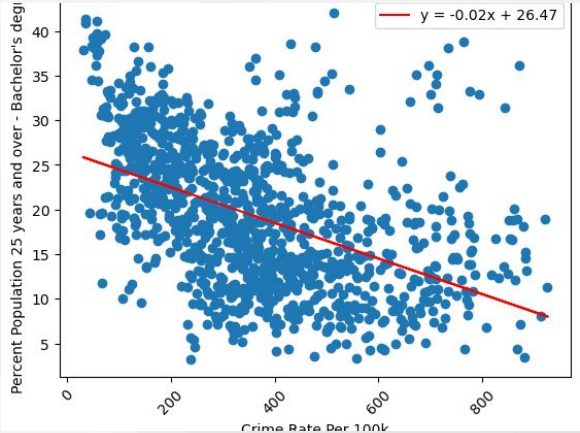
Machine Learning - Linear Regression



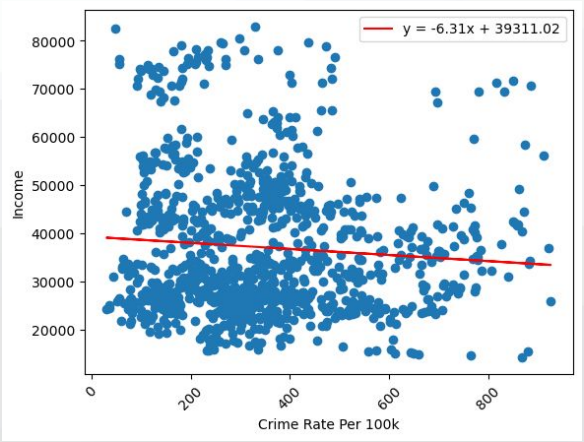
Homeowner
& Crime Rate
P-value: 1.74e-26



House Price & Crime Rate
P-value: 8.71e-50

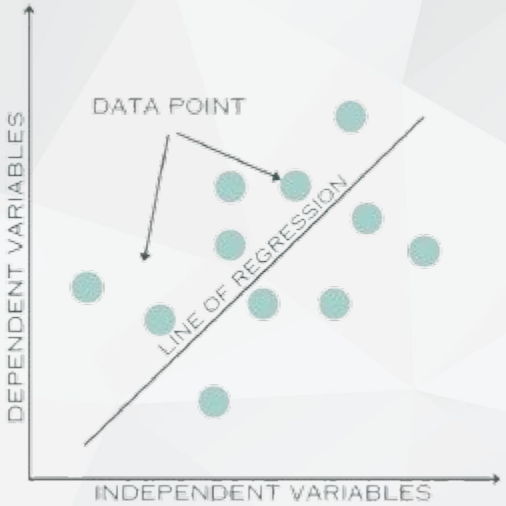


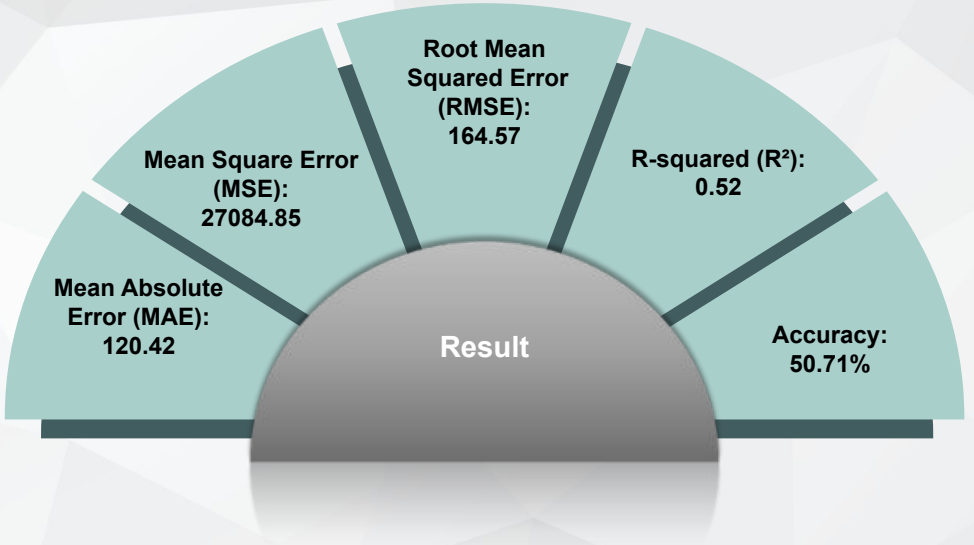
BA or BS's & Crime Rate
P-value: 1.01e-63



Income & Crime Rate
P-value: 0.01

Machine Learning - Linear Regression





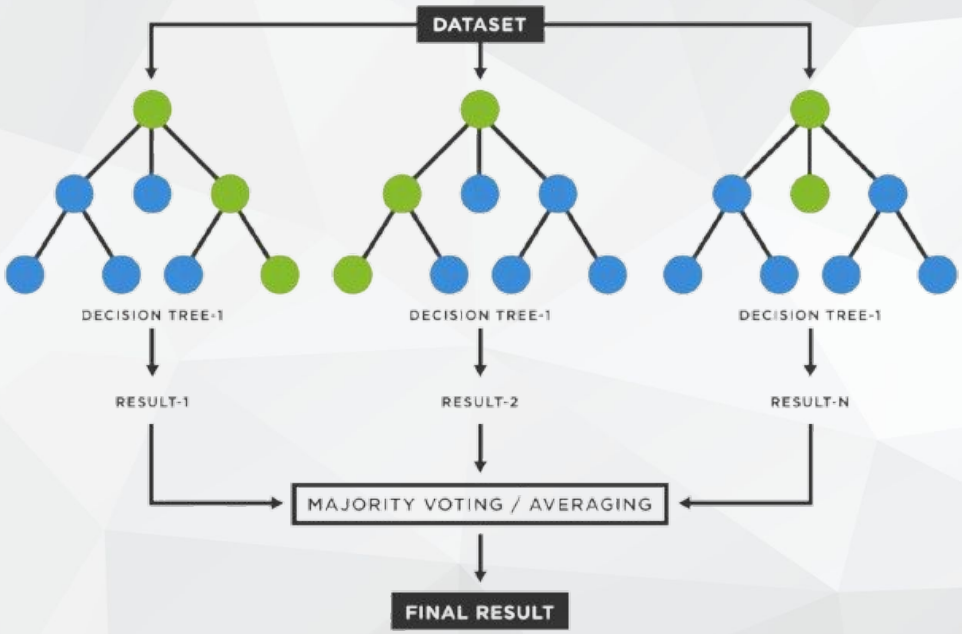
Select dependent and independent variable

Split data into traning and testing set

Apply Linear Regression and train the model

Prediction and result

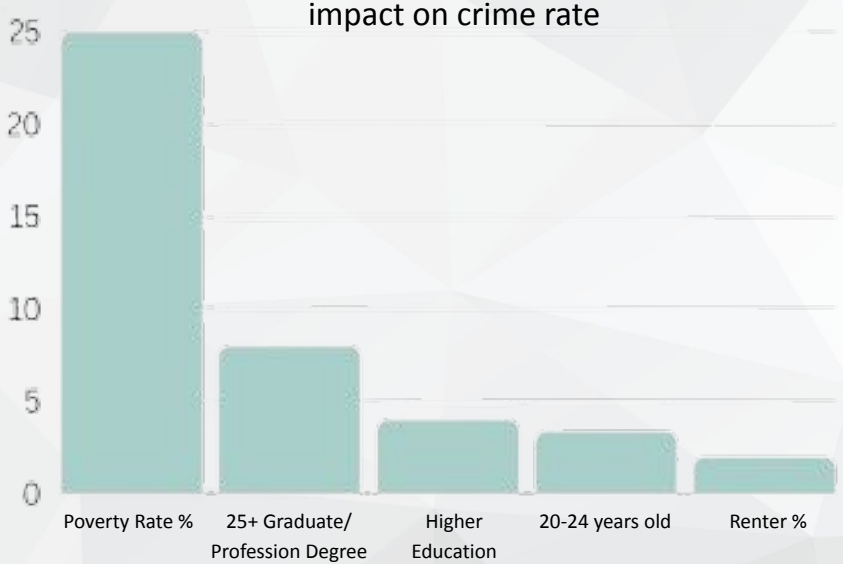
Machine Learning - Random Forest



Random Forest is a versatile and powerful that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classification or regression of the individual trees.

Machine Learning - Random Forest

Top 5 features significantly impact on crime rate



Mean Absolute Error (MAE): 77.54

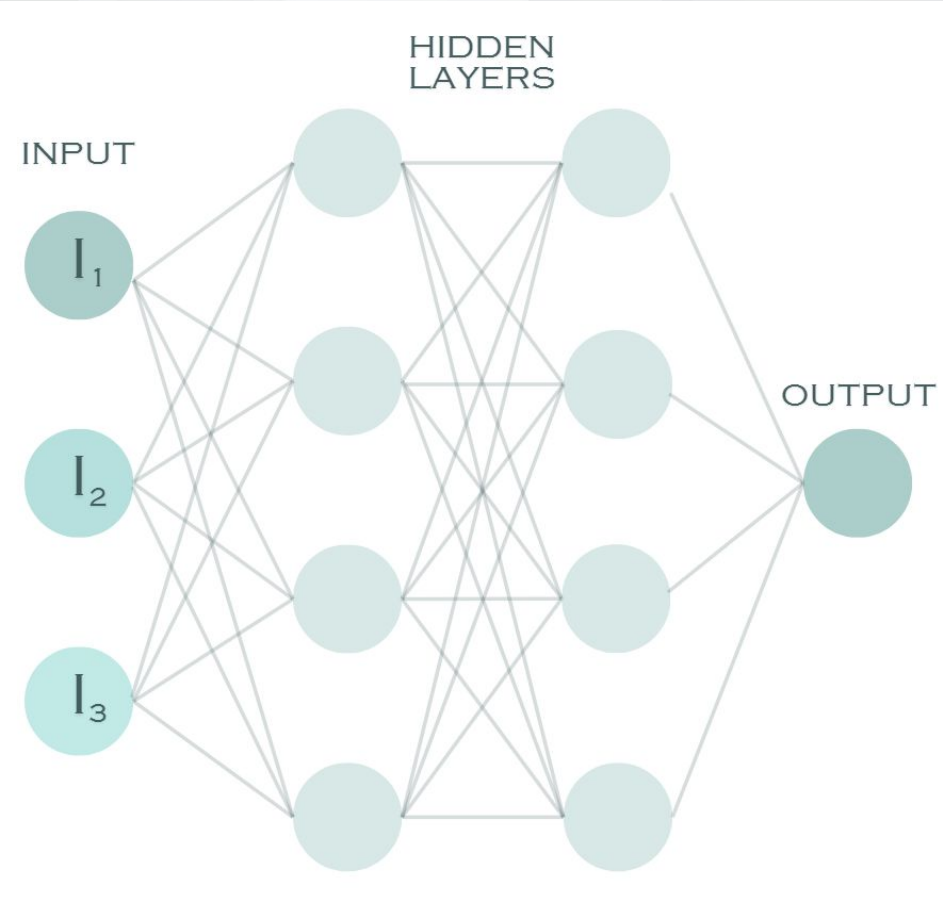
Mean Squared Error (MSE): 9397.71

Root Mean Squared Error (RMSE): 113.47

R-squared (R^2): 0.74

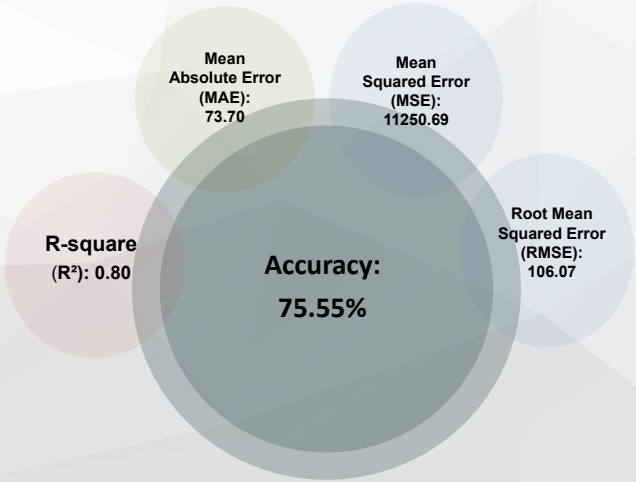
Accuracy: 69.71%

Machine Learning - Neural Network



Neural Network consisting of layers of neurons that can process information. It is designed to recongnize patterns and solve complex problems.

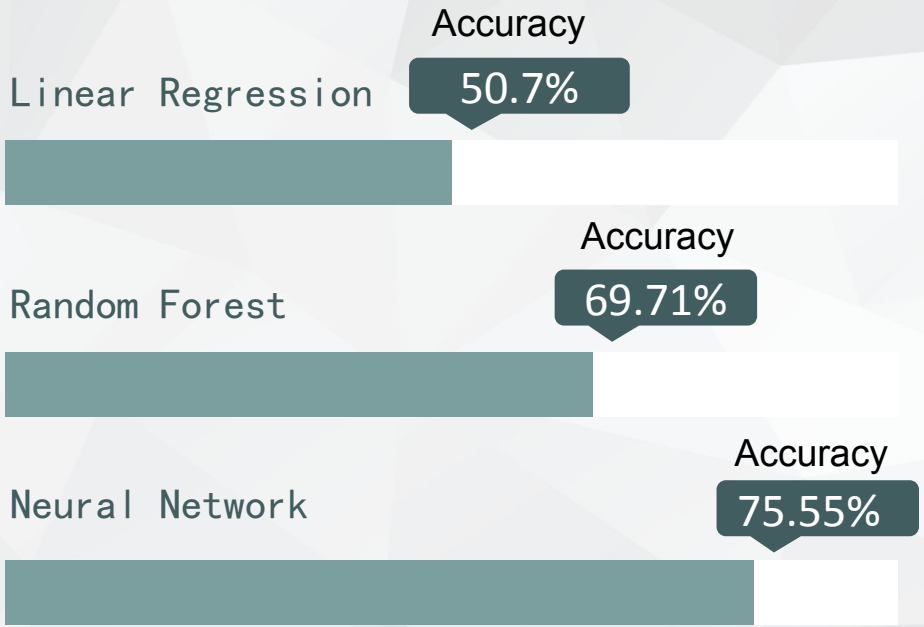
Neural Network had the best results from the three models tested.
We got an R-Squared value of 0.80.
We got an accuracy score of 75%.



Part 5

SUMMARY

Summary



Liner Regression
R-Squared: .52

Random Forest
R-Squared: .74

Neural Network
R-Squared: .80

Part 6

DEMO

THANK YOU