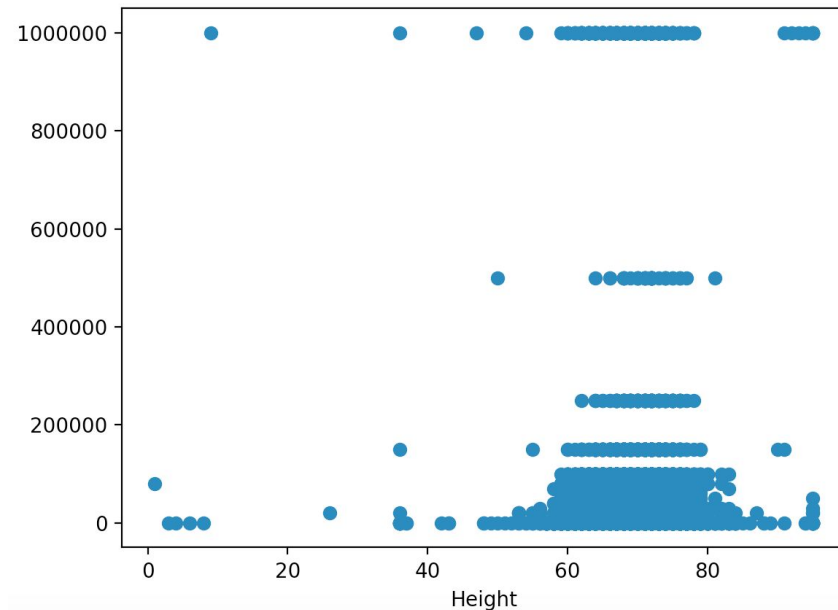
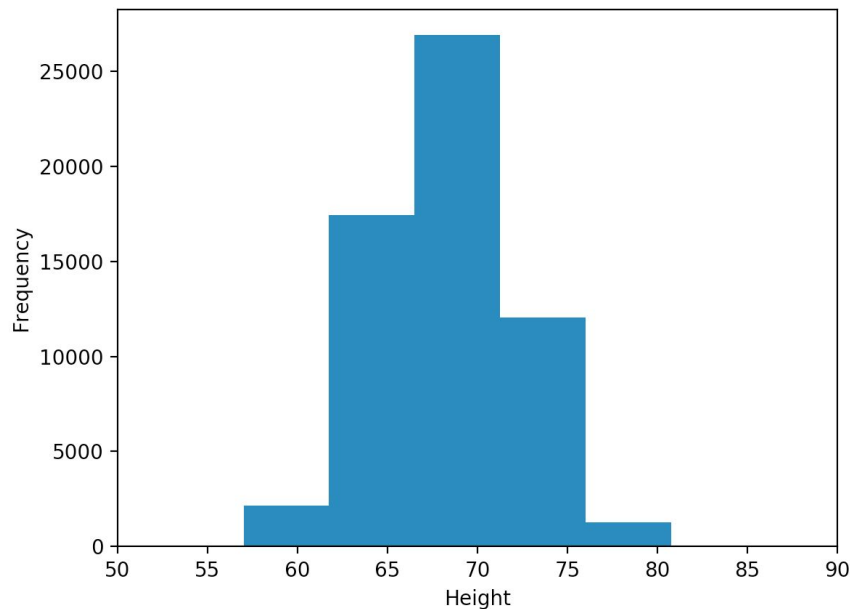


# Machine Learning Fundamentals - Capstone project -

Codecademy - Cohort Sep-18 2018

Repo: <https://github.com/sdurighello/python-capstone-starter>

# Two graphs containing exploration of the dataset



Do taller people get better pay? Visually i can't see any correlation, so i would say no

# Statement of your question

## Question 1

Can I predict the income based  
Education level?

I expected education to be a driver of  
income.

## Question 2

Can I predict the type of job based on  
income and education?

Common believe is that doctors and  
lawyers are paid more.

# Explanation of new column - Education level

- Created an ordinal mapping from least educated to highest education:
  - -1: NaN
  - 0: no education
  - 1: high school or below bachelor
  - 2: bachelor
  - 3: master (including law and medicine school)
  - 4: Phd or higher
- Mapped this value to column 'education' and added a new column 'education\_level'

# Explanation of new column - Job labels

- Created qualitative labelling for each job value from 1 to 18
- Used the label 0 for 'other' and 'NaN'
- Mapped this values to column 'job' and created new column 'job\_label'

# Comparison between two regressions: MLR v KNR

Question 1: Predict income based on education level

## **Multiple Linear regression**

Fit execution time: 0.0006s

Accuracy: 6.566051019196717e-05

## **K-Neighbors Regressor**

Fit execution time: 0.022s

Accuracy: -0.191

# Comparison between two classifications: K-NB v SVM

Question 2: Predict job label given education level and income

## K-Nearest Neighbors

Fit execution time: 0.015s

Accuracy: 0.24

Precision: 0.16

Recall: 0.25

F1-score: 0.18

## Support Vector Machines

Fit execution time: 1.613s

Accuracy: 0.29

Precision: 0.16

Recall: 0.29

F1-score: 0.2

# Conclusions

## Question 1

Can I predict the income based Education level?

I expected education to be a driver of income.

## Question 2

Can I predict the type of job based on income and education?

Common believe is that doctors and lawyers are paid more.

The model I used showed no prediction ability with very poor correlation.

In my sample i've removed all rows that had no income. This might have introduced bias since most probably higher earners are more comfortable in sharing their income and would be over-represented.