Davis Adams

1/19/2023

Machine Learning

Lab 1

Who you worked with?

Andrew and Duc

What you did?

Imported CSV

Read CSV into Pandas

Used head to view rows (0-4)

Found 207 missing datapoints on total\_bedrooms

Looked at stats of median age, bedrooms, etc…

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 20640 entries, 0 to 20639

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 longitude 20640 non-null float64

1 latitude 20640 non-null float64

2 housing\_median\_age 20640 non-null float64

3 total\_rooms 20640 non-null float64

4 total\_bedrooms 20433 non-null float64

5 population 20640 non-null float64

6 households 20640 non-null float64

7 median\_income 20640 non-null float64

8 median\_house\_value 20640 non-null float64

9 ocean\_proximity 20640 non-null object

dtypes: float64(9), object(1)

memory usage: 1.6+ MB

Chart, histogram, box and whisker chart

Description automatically generated

What questions you have about the lab?

How can missing data affect machine learning?