HW2.ipynb contains the entirety of the code for this assignment.

The only possible algorithms to discuss here are how I webscraped from abs.gov.au.

1. Get data from link
2. Parse data from link using BeautifulSoup
3. Search1 = Data.Find_All(divs with chart/table pairs)
4. Body = Search1[Relevant table].Find_All(table wrapper)
5. For Row in Body:
    a. Interpret age range using Row head
    b. Assign value = data from Row in 2022 column
    c. Modify dataframe such that null values in new smoke column are now value when age is within range
6. Return modified dataframe

Sample output:

Linear CV and accuracy:

[0.45408804 0.38711907 0.34660319 0.57199746 0.29997448]

0.4421630779582859

Logistic CV and accuracy:

[0.80246914 0.77777778 0.81481481 0.85802469 0.76397516]

0.8121137206427689

KNN CV and accuracy:

[0.69135802 0.62345679 0.69753086 0.74074074 0.70807453]

0.6555555555555556

Logistic regression performed the best.