HEALTH AND LIFESTYLE

ARTIFICIAL INTELLIGENCE AND IN PLANNING AND DECISION MAKING
COURSE PROJECT WORK

USING LINEAR REGRESSION AND CLUSTERING

DATA SET AND STEPS

- Getting data from Kaggle
- Making sure only needed features are present
- Making sure only numeric values are in the dataset
- Making sure there are no Null values

CORRELATION MATRIX

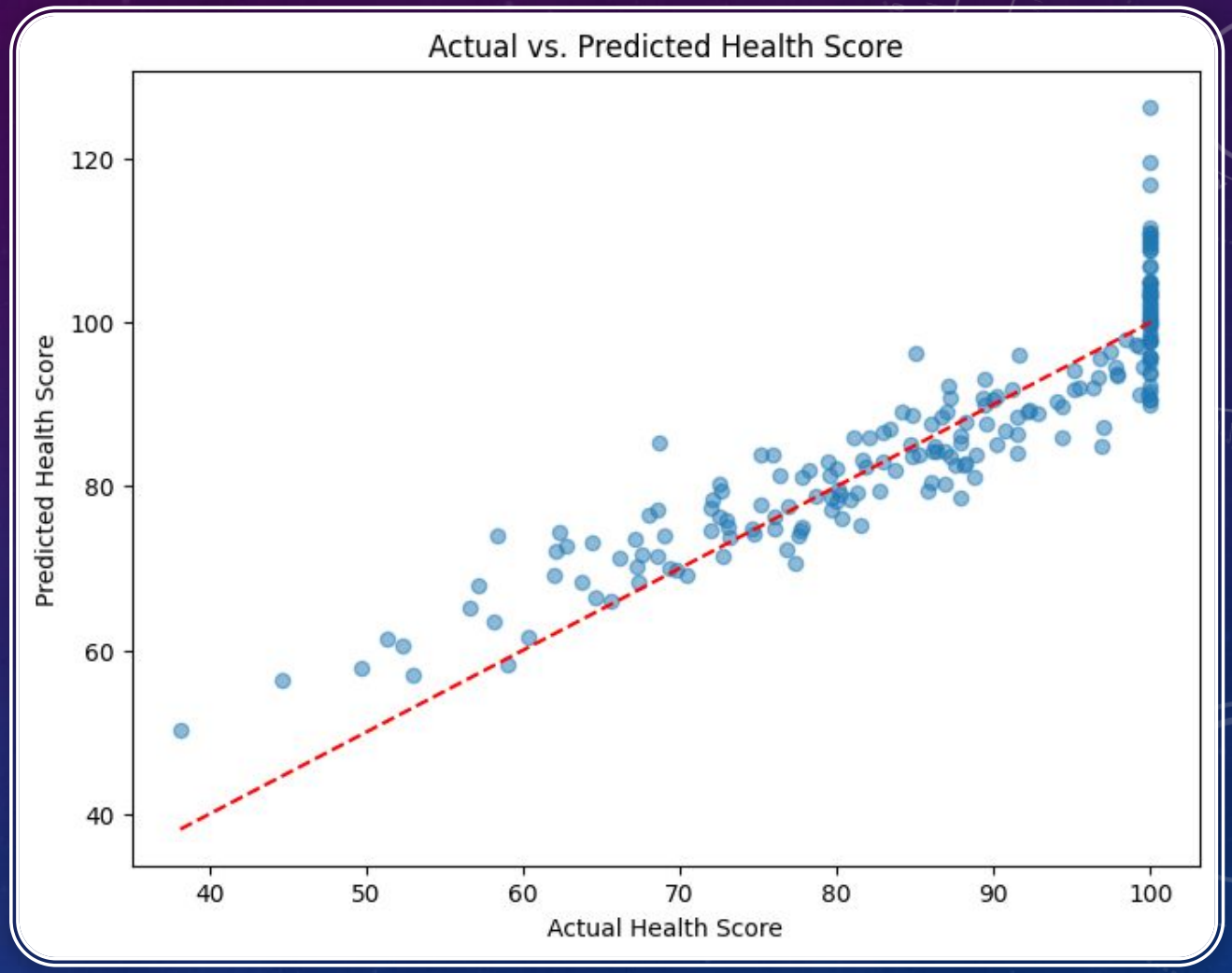
- No correlation between different features -> They are independent
- Slight correlations between exact features and the dependent variable – Health Score

TRAINING THE MODEL

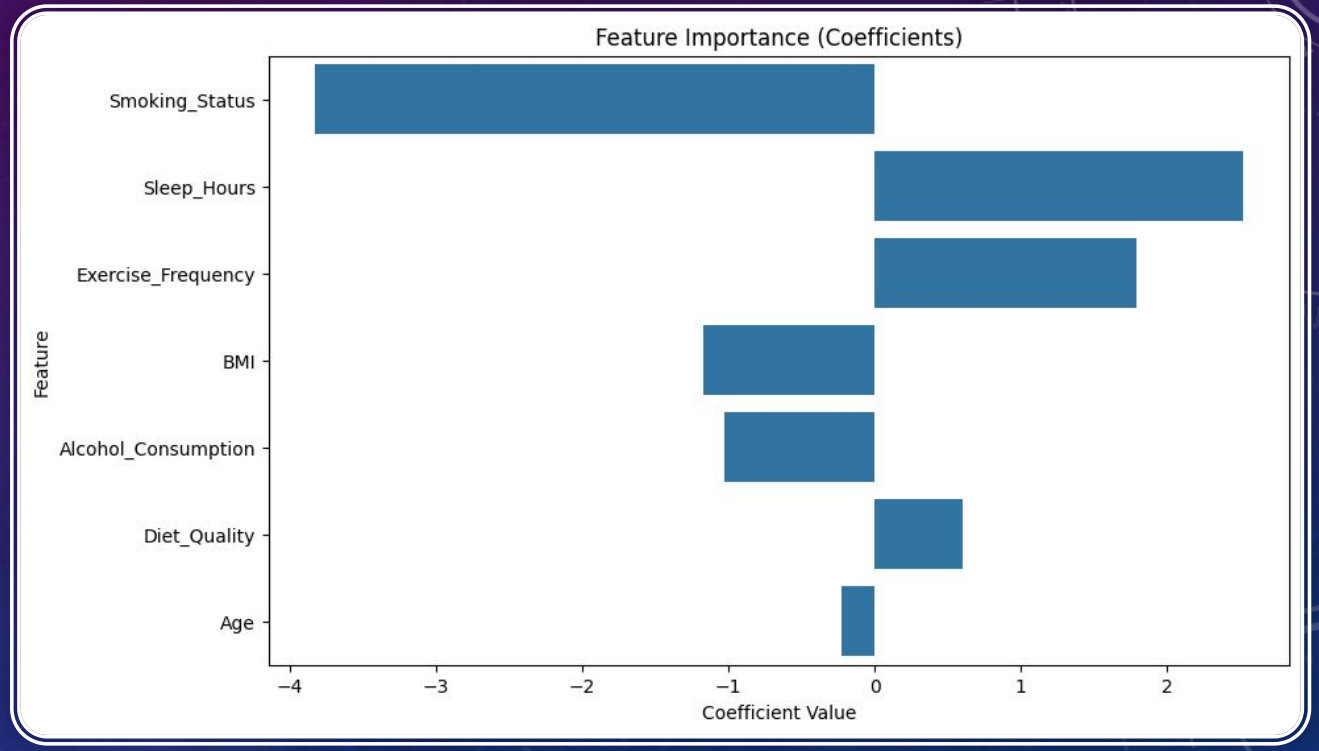
- Separating the target variable from the rest
- Training the model with a 80% training and 20% test distribution
- Determining mean squared error and R^2 values
- Determining the accuracy of the model (80.90%)

COMPARING ACTUAL DATA WITH PREDICTION

- Red diagonal line showing the case for perfect prediction
- The linear regression model makes a good job predicting the actual health score

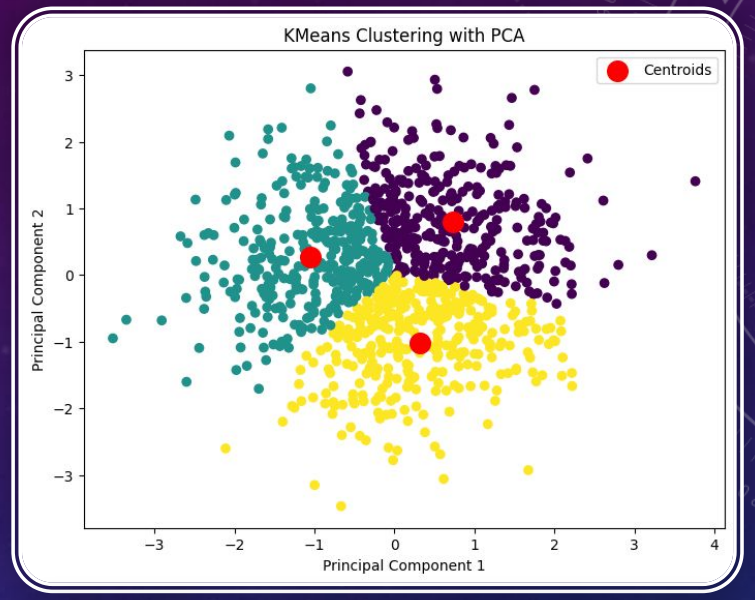


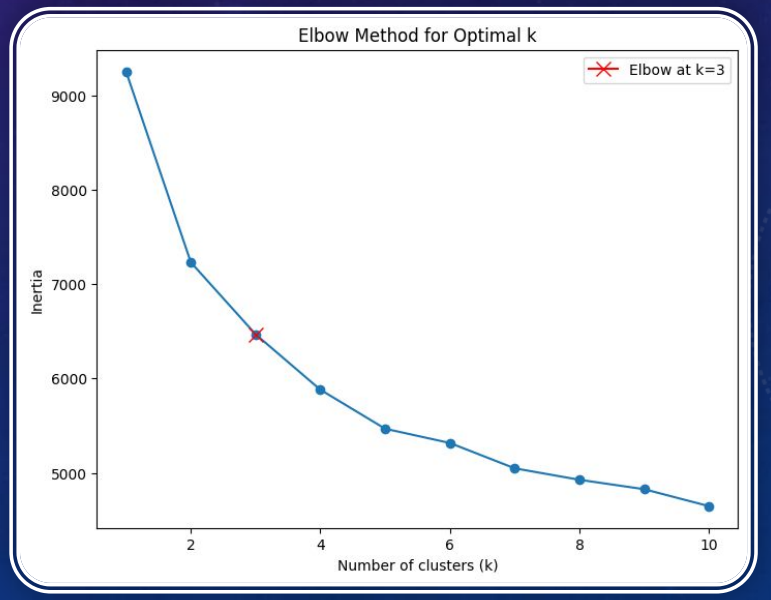
FEATURE IMPORTANCE



CLUSTERING AND ELBOW METHOD

- Using K-means method to cluster the data.
- PCA to reduce the feature numbers
- Clustering seems arbitrary





THANK YOU FOR YOUR ATTENTION

- Made by Lantos Sebestyén Fehér
- Source code on Google Collab