Artificial Intelligence 2

Assignment 1, Matthew O'Neill / C11354316

# Outliers

There were a number of outliers in the provided dataset that could cause issues should we only be taking the mean into account; the mode and median values come in useful here.

# Missing Values

With non-numeric values a value of ‘?’ was assumed to signify a missing value. For numerical (continuous features) a value of 0 was taken to be a missing value.

# Addressing Quality Issues

Missing values should be ignored while calculating and mode and median employed in addition to simply mean.

# Continuous Features

Continuous features identified in the given data set were as follows:

* Age
* Fnlwgt
* Education Number
* Capital Gain
* Capital Loss
* Hours per Week

# Categorical Features

Categorical features identified in the given data set were as follows:

* ID
* Work Class
* Education
* Occupation
* Marital Status
* Relationship
* Race
* Sex
* Native Country
* Target