Exercise 0 Dow Jones Dataset Analysis.

The chosen dataset from our group it is the Dow Jones dataset which contains relevant data about the stock market prices, and it has a couple of attributes which we can apply data preprocessing and filtering for it to be adequate to generate a Machine Learning model, or perform data analysis on a given requirement. I want to start by describing the dataset which contains a total sample of 751, and it has 14 attributes. This dataset has a mix of datatypes, for example we have the stock attribute which is nominal, and the other 13 are continuous numerical data.

An interesting data which I have made some analysis would be the closing price which has the following information: a mean of 53.33$, mode = 20$, min= 10$ and max = 170$, these values were abstracted from the closing column to better understand what is the data range, learning what values are most repeated on the column, better visualizing the mean value of the data which we are handling, visualizing what price parameters have in its extremums like the minimum value and the maximum, in this way by taking these pieces of information apart from the whole column makes the attribute closer to human perception and readability so you can apply the given logic based on the requirement on in.

The preprocessing steps that I have taken to properly work on this data set are: removing the US dollar Symbol out of the price values, converting the string price values into floats, dropping null values, turning the data attributes from string data types to Date data type for Pandas. The tools that were used for the data analysis used to analyze this data set its Jupyter notebook and pandas library.

As a conclusion for this paper we can subs tract that the dataset that we are working on has in total 14 attributes, a mix of nominal and continuous, we analyzed an attribute and

described the properties of its distribution, and explained the data processing that took place