**Assignment #3**

**Support Vector Machine**

**Written Responses**

1. In you written response writes a paragraph explaining your findings about each column.

Many of the columns are int64, meaning that almost all the columns are numerical, the bare column is stored as an object instead of a numerical value, because it contains non-numerical values like ‘?’, so, Pandas assumes that the whole column is a String. The Max

No missing Values are reported, the ID column has a very large range, it has 8 columns, the range 1 to 10 suggest they might be ordinal features (ratings or severity levels).

1. what are the key insights and findings from the plots

A graph of different colors

AI-generated content may be incorrect.

This bar chart represents the distribution of the classes in this case the 10 columns in the dataset, the similar values of each features means that there are no missing values or imbalances, The different color represent the different categories, providing a different distinction

A graph of a number of bars

AI-generated content may be incorrect.

This histogram displays the distribution of the Bare Nuclei Class, this one is highly right-skewed, so, most of the values are concentrated on the lower end (between 1 and 2), the there Is another increase in the 10, suggesting there is high amount of bare nuclei values which may correlate with malignant tumors, the rest of the values are imbalanced, indicating there are few bare nuclei counts.

A screenshot of a computer screen

AI-generated content may be incorrect.

Finally, this graph represents the distribution of the classes in the dataset, where: Class 2 = Benign Tumors and Class 4 = Malign Tumors.

In this graph we can observe that there is a high count in Benign Tumors than Malign ones.