**Data Science – Homework 4**

**Submission deadline: Sunday, 22 August 2021**

**Lead Instructor: Polly**

**DESCRIPTION**

1. Use graphs that we build together during session 16.
2. These graphs visualise analytics data for the Iris flowers classification.
3. Your task will be to provide synopsis description for each of the graphs mentioned below.

**WHAT WE ARE LOOKING FOR:**

1. Your ability to interpret analysis results.
2. Your ability to interpret and explain diagrams.

**TASK**

1. GRAPH 1: “Create a graph to find relationship between the sepal length and width.” Describe and interpret graph results.

Sepals are leaf-like parts that enclose the flower bud, protecting it from weather or injuries while the flower is developing. One can assume that the bigger the sepals, the greater protection they will offer a developing flower.

Graph 1 presents the bivariate relationship between the sepal length and width of three species of the flowering plant, iris. The dataset includes information from 50 of each species. Both axes have undefined measurements; we will make assumptions that they are defined in centimetres.

The Setosa specie length range is between 4.3 -5.8cm while its width range is 2.3-4.4cm. The graph presents a strong positive correlation between the two variables; as the length increases so does that width. The average length and width are 5cm and 3.4 respectively. The Versicolor specie has a length range of 4.9-7cm and width range of 2-3.4cm. The graph presents that the two variables have a normal distribution. As the length increases, the width does. However once the maximum width is achieved 6cm length, the width begins to decrease. The average length and width are 5.9cm and 2.8cm respectively. This shows that the average length Versicolor sepals are larger than Setosa, however Setosa have an average larger width. This difference could be due to the method of flowering, the positioning of the petals that they provide support or even the environment that they protect from. The Virginica specie has the largest ranges out of the three types; its length range is 4.9-7.9cm and width range is 2.2-3.8cm. This specie also has a positive correlation with an increase length correlation with an increase in the width. The average length and width are 6.6am and 3cm respectively. However there is more bunching around the lower length range of 6-7cm. With further statistical analysis, may present length of 7.5cm and above as outliers.

1. GRAPH 2: “Create a jointplot to describe individual distributions on the same plot between Sepal length and Sepal width.” Describe and interpret graph results.

Graph 2 is a jointplot of bivariate relationship with the sepal length and width. It presents the collective information for all three species. We can see that the is a slight skewness in the sepal length to the lower range with peaks at 5cm and 6.3cm. Whereas there is a normal distribution for the sepal width, with a single peak at 3cm. The heaviest bunching is at length width; 5cm and 3.4cm; and 6cm and 2.7cm. This is likely due to the Versicolor overlapping with the lower range of Setosa and the upper range of Virginica.