**Data Citizenship Unit, Level 4**

**Practice / Exemplar Assessment on Zoo and Aquarium animals**

**In this assessment you will need to:**

* **download a dataset from a link given to you**
* **carry out simple summaries of the data**
* **interpret the data to answer some questions or solve a problem with the dataset.**
* **choose the best types of visualisation to use to show your interpretation**
* **create two visualisations.**
* **make recommendations based on your interpretation of the dataset.**

**Access the Dataset (1 mark):**

1. Download and open the North American zoo and aquarium species dataset

**Excel Dataset (with data dictionary):** <http://dataed.in/DCL4Eexcel>

**Original Dataset (CSV without data dictionary):** <http://dataed.in/DCL4Edataset>

**Understanding the dataset (2 marks):**

**Extra practice task:** Data quality check – Does the Overall Sample Size column equal the sum of the Male and Female Sample Size columns?

Create a new column and give it the heading like Calculated Sample Size. In the first empty cell of that column (row 2) type ‘=’ then select the Male Sample Size column entries. Add a ‘+’ to your formula then select the Female Sample Size column entries. You can select column entries by clicking on the top value of the column then click Ctrl-Shift-↓ (down arrow).

2. What does each column contain?

3. Do a data quality check and find out if there any data quality issues

**Manipulations (10 marks):**

4. You will not need data on confidence limits (the CI columns). Remove all of the columns that are not required for your analysis.

**Extra practice task:**Create a column to calculate the difference between the original dataset column ‘Overall Sample Size’ and the column you have just created with the Male and Female sample sizes added together. You could use conditional formatting to colour this to highlight greater differences.

5. Remove the rows of poor quality data where there is a ‘yes’ in the Male Data Deficient or Female Data Deficient columns. Then delete these two columns as they should be no data stored here now.

6. Find out if males or females of these species live longer on average. Calculate the difference between the Male and Female MLE (median life expectancies) columns. Use conditional formatting to colour the cells differently depending on the values.

7. The TaxonClass variable shows the type of species. Create another data table that summarises the sample size and MLE data about each unique Taxonclass value.

**Extra practice task:**What is the longest lived species? Create a graph about the longest or shortest living species. Sort your data by Overall MLE and select the top 10 or 15.

**Extra practice task:**Do mammals live longer on average than birds? Create a column to calculate the totals of Overall Sample Size, Male Sample Size and Female Sample Size for each of the TaxonClass values.

**Extra practice task:**

Which TaxonClass type has there been most of in North American zoos and aquariums? Create a chart using the summary of Overall Sample Size that you did earlier.

**Visualisations (9 marks):**

8. What is the most popular species? What are the least popular? Create a chart with the most popular or the least popular creatures in zoos and aquariums.

**Extra practice task:**Pick a species where there are at least four species varieties (such as penguins or monkeys) and summarise the range of Overall, Male and Female MLEs. Do male or female of that species live longer?

9. Create a chart showing which TaxonClass species type lives longest on average?

10. Create a graph showing the species that have the greatest difference between the male and female life expectancies

**Interpretation (3 marks):**

11. What type of graph did you choose to show the most common species? Why did you choose this?

12. What type of graph did you choose to show the average life expectancies of different types of species (taxonomical classes)? Why did you choose this type of graph?

13. What type of graph did you choose to show the differences between male and female average life expectancies? Why did you choose this?

Assessment Total = 25 marks