**Data Citizenship Unit, Level 4**

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

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

Try selecting all the columns and doing Ctrl-T to tell Excel to make this a table. That will shade the headings and rows to make it more readable, and it will set up variable names for your columns.

* **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

Try looking at the headings and the first few rows. Guess what data you think is stored there. If there is a data dictionary then look at that to help you understand the data stored in the table. Use the arrows next to each column heading to sort some of the data such as sample size. What are the largest and smallest values?

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

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

Skim through the table and look for missing data. Are there any missing values?
Can you see numbers where the rest of a column is text? Can you see text in a numerical column? Are there any strange or corrupted values?

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

2. What does each column contain?

Select the column heading, right-click and select **Delete**. You can select multiple columns by holding down the Ctrl key then clicking on the column headings.

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.

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.

**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.

**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).

Create a new column to calculate the difference in Male and Female MLE (median life expectancies). This can a formula for either Male MLE – Female MLE or the other way round. You could use conditional formatting to colour this to highlight greater differences. Make sure this is coloured the way you expect.

Use the filter button on the Male Data Deficient column heading to unselect the ‘blanks’, which will just show the rows where there is poor quality data for Male MLE. So that we don’t delete the rows in between the ones that are visible, click ‘Find and Select’ and then ‘Go to Special’. Then select the ‘Visible Cells Only’ option. You can now right-click on the rows and select ‘Delete Row’

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.

Create a new table at the right of your sheet. Copy the headings (TaxonClass, Overall MLE, Male MLE, Female MLE) over from the first table or type them in. In the new TaxonClass column copy unique values from the original TaxonClass column by using Advanced Filter, select ‘Copy to a new location’, in the ‘Copy to’ box select the new column, and tick the ‘Unique records only’ box.

Summarise the Overall MLE column. In the new column type **=averageif(TaxonClass,N2,Overall\_MLE)** \*change N2 to be the cell you want the data to be put.Do the same for the Male MLE and Female MLE columns, changing Overall\_MLE to be Male\_MLE or Female\_MLE and N2 to be the new columns (like O2 and P2)

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.

Create another new table at the right of your sheet. Copy the headings (TaxonClass, Overall Sample Size, Male Sample Size, Female Sample Size) over from the first table or type them in. In the new TaxonClass column copy unique values from the original TaxonClass column by using Advanced Filter, select ‘Copy to a new location’, in the ‘Copy to’ box select the new column, and tick the ‘Unique records only’ box.

Summarise the Overall Sample Size column. In the new column type **=sumif(TaxonClass,N2,Overall\_Sample\_Size)** \*change N2 to be the cell you want the data to be put. Do the same for the Male Sample Size and Female Sample Size columns, changing Overall\_ Sample\_Size to be Male\_Sample\_Size or Female\_Sample\_Size and N2 to be the new columns (like O2 and P2)

**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.

**Visualisations (9 marks):**

**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.

Sort your data by Overall Sample Size and select the top 10 or 15. Choose a chart type that you think works best for this data.

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:**

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.

You have a table that has this summarised. Select the TaxonClass and Overall MLE columns and create a chart, choosing a chart type that you think works best for this data.

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

**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?

You have a column that has this calculated. You could create a graph showing species where there are lots more males than females or species with lots more females than males. You are not being expected to do complex calculations to work out the greatest difference with both.

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

Think about all the types of graphs you have been taught about:

Bar charts, column charts (or horizontal bar charts), grouped bar charts and stacked bar charts, histograms, dot plots, line graphs, scatter plots, pie charts and donut charts.

What would work well for the data you have here?

**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 greatest differences between male and female average life expectancies? Why did you choose this?

Assessment Total = 25 marks