**Data Citizenship Unit, Level 4**

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

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

* **download a dataset from a link given to them**
* **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 their interpretation**
* **create two visualisations.**
* **make recommendations based on their interpretation of the dataset.**

More information is available about the dataset at <https://www.nature.com/articles/sdata201919#MOESM36>

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

1. Download and open the North American zoo and aquarium species dataset as a CSV files

**Dataset:** <https://figshare.com/articles/AZA_MLE_Jul2018_csv/7539968> (short URL: <http://dataed.in/DCL4Edataset>)

Has the data been succesfully downloaded? (1 mark for successful download)

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

2. What does each column contain?

Looking for evidence that the learner understands what each column contains.
They understand the ranges of each age band (1 mark)

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

They have looked at the data dictionary. (1 mark)

**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.

Delete the six CI columns (Overall CI upper and lower, Male CI upper and lower, Female CI upper and lower) (1 mark)

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.

Filter the rows to show only ‘yes’ in each column, ensure ‘visible rows only’ is selected and delete the selected rows (**2 marks:** 2 marks for deleting all data, 1 mark for partial solution or needing assistance)

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 column and add a formula to subtract one column value (Male or Female MLE) from the other one. (**3 marks:** 3 marks for full solution, 2 marks for creating the new column and the formula without conditional formatting, 1 mark for partial solution or needing assistance)

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 a new table. Use advance filter to select unique values in the TaxonClass column. Add columns for the Overall, Male and Female MLEs, and use an averageif formula (or similar technique) to average the values in the main table into the new table. (**4 marks:** 4 marks for full solution. Partial marks could be awarded for: 1 mark for using getting the unique values from TaxonClass. 1 mark for each of the Overall, Male and Female columns. Other division of marks possible, particularly if assistance is required)

**Visualisations (9 marks):**

Learners can be presented with a new summarised dataset at this point, to address any challenges that may have been encountered by the manipulation activities.

When creating any plot, marks should be awarded for a title, axes labels, any required legends and putting the correct variable on the correct axis.

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.

Create a chart showing species with highest Overall Sample Size or lowest Overall Sample Size with 10-20 species listed. (3 marks: 3 marks for full solution, 2 or 1 mark if support required, graph not labelled, not all values displayed or other issues)

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

Create a chart showing Overall MLE for each taxonomical class (from task 7). (3 marks: 3 marks for full solution, 2 or 1 mark if support required, graph not labelled, not all values displayed, or other issues)

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

Create a chart showing difference between Male and Female MLE, or the difference between Female to Male MLE (from task 6). This column should have been sorted and the highest or lowest 10-20 species listed. (3 marks: 3 marks for full solution, 2 or 1 mark if support required, graph not labelled, not all values displayed, or other issues)

**Interpretation (3 marks):**

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

(1 mark for identifying chart type and giving a simple reason why that type was selected)

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?

(1 mark for identifying chart type and giving a simple reason why that type was selected)

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?

(1 mark for identifying chart type and giving a simple reason why that type was selected)

**Total marks available: 25 marks**

**Pass mark: 15 marks (60%)**