1. Sample Exercise Instructions

2. Locate and download data sets you have been given access to

3. Open the data sets and review the structure of the data. (column and row headings).

4. From data source 1 ("UFOGB_Details.txt")
   copy all the data and paste into a new document

5. Use the following appropriate column headings
   - ID
   - SIGHTING_DATE
   - CITY_STATE
   - COUNTRY_CODE
   - SHAPE
   - DURATION
   - DATE_POSTED

6. Save this file to create an appropriately named separate file or data source

7. Combine and merge your new data set created in step 4, with data source 2 ("UFOGB_Comments.txt")

8. Use the following appropriate column headings
   - ID
   - COMMENTS

9. Ensure your merged file has only unique columns across the two datasets. Save this file to create an appropriately named separate file or data source of the merged data

10. Some new sightings have just come through on another file ("UFO_latest.txt") open this file and append it to your merged dataset. Check the columns match and the correct number of rows and columns are added to your merged dataset.

11. Save and create an appropriately named separate version file or data source of your appended dataset.

12. With all three text files now merged into one dataset we can start cleansing the data in the following steps.

13. Please make sure all dates and times if available are formatted as follows
   12 June 2022 5pm

14. Cleanse the following columns
   a. SHAPE
      i. Missing values replaced with "Not Provided"
      ii. Amend casing to capitalise each word
   b. COMMENTS – replace non printable characters
      i. &#44 with [,] (comma)
      ii. &#39 with ['] (single quote)
      iii. &#33 with [!] (exclamation mark)
      iv. &#8220 with ["] (double quote)
      v. &#8230 with […] (horizontal ellipsis …)

15. Split the CITY_STATE columns into CITY & COUNTRY

   Before
   CITY_STATE
   london (uk/englnd)
   newcastle (uk/England)

Cardiff (uk/wales)

After
CITY
London   England
Newcastle England
Cardiff   Wales

16.   Check for misspellings, correct any issues and ensure you only can filter on the following countries
- England
- Scotland
- Wales
- Northern Ireland

17.   Format your data to ensure it looks consistent and follows any rules for grammar and appearance. For example capitalisation of words. Numeric data justification.

18.   We now do some basic visualisation and statistical analysis to answer the following questions

19.   How many sightings have there been for each of the countries?