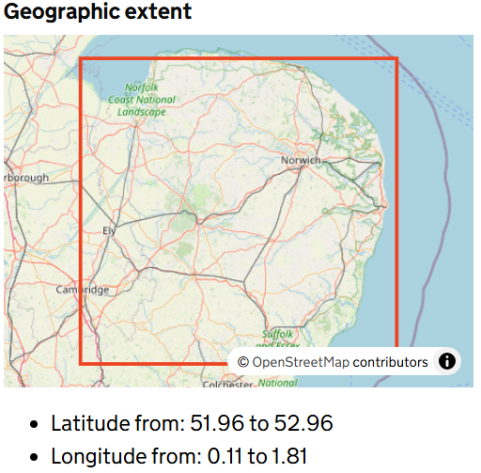
# Analysis of Great Crested Newt - Risk Zones (Norfolk & Suffolk)

This dataset was obtained from environment.data.gov.uk and was selected because it was recently revised on 5 June 2024, ensuring that the information is relatively up to date.



This dataset covers the **Norfolk and Suffolk regions**, representing the geographical area of analysis

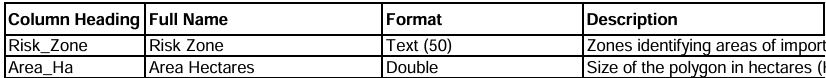
There are several ways to analyse this dataset, such as conducting a comparative analysis with previous versions to detect spatial or ecological changes, using GIS software for mapping and spatial exploration, applying Python libraries like GeoPandas for geospatial data analysis, or even employing machine learning models to predict potential risk zones.

For the purpose of this project, I will carry out a simple analysis using Python, as it aligns with one of the skills highlighted in the job description.

## Dataset preprocessing

According to the dataset’s metadata in the pdf, there are two main attributes:

* **Risk\_Zone:** A text field indicating the classification of areas based on their importance for Great Crested Newts (e.g. Red, Amber, Green, or White zones).
* **Area\_Ha:** A numeric field showing the area size of each polygon in hectares.  
  These fields form the basis of the spatial and quantitative analysis conducted in this project.



Metadata of the dataset

Next, we load the required libraries and import the dataset.

A black screen with orange and white text

AI-generated content may be incorrect.

Then we check a summary of the data to see what preprocessing steps need to be taken

A screen shot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

These columns don’t contribute to the analysis according to the metadata as well the summary and therefore will be removed.

The data was cleaned in the following steps in order to maintain consistency and accuracy in the analysis.

1. **Standardising column names:**

****

1. **Selecting relevant columns:**



1. **Removing invalid or missing entries:**



Rows without a valid risk zone were removed to maintain data accuracy and ensure all entries used in the analysis have complete classification information.

## Statistics Summary

The following code calculates the total area covered by each risk zone.  
This process is similar to performing **aggregation and grouping in SQL**, a skill which was mentioned in the job description and I’m familiar with.

A computer screen with text

AI-generated content may be incorrect.

This step provides a clear summary of how much land falls under each risk category (Amber, Green, Red, etc.), giving an overview of the spatial distribution of Great Crested Newt habitats across Norfolk and Suffolk.

A black screen with white text

AI-generated content may be incorrect.

## Visualisation

To better understand the spatial distribution and proportions of each risk zone, two types of visualisations were created.

**Spatial Map:**  
The first plot visualises the *Great Crested Newt Risk Zones* across **Norfolk and Suffolk**. Each zone (Green, Amber, and Red) is displayed in distinct colours using the *Set2* colormap. This helps to clearly distinguish between areas of different conservation or development risk levels.  
Turning off the axes makes the map cleaner and more presentation-ready.

A computer screen shot of text

AI-generated content may be incorrect.

A map of a green and blue area

AI-generated content may be incorrect.

Spatial plot

**Bar Chart:**

The second visualisation presents the *percentage of total area* occupied by each zone. It provides a quick overview of how much land falls under each risk category.  
For example, from the chart it’s clear that **Green Zones** make up the majority of the total area, followed by **Amber**, while **Red Zones** cover a relatively small fraction.

A green and orange squares

AI-generated content may be incorrect.

Bar Chart

Together, these visualisations communicate both **geographic context** and **quantitative insights.**

## Conclusion

At first glance, the Red Zones cover a relatively small proportion of the total area, suggesting that only a limited number of sites in Norfolk and Suffolk fall under the highest risk category for *Great Crested Newt* conservation. However, this observation is only a preliminary insight.

A more thorough analysis could involve comparing this dataset with **previous versions** to identify trends or changes over time, such as whether high-risk areas are expanding or shrinking. Additionally, integrating **other environmental datasets** for example, habitat quality, land use, or development activity would provide deeper context and help validate whether these zones are consistent with broader ecological patterns.

This kind of follow-up work would give a more complete picture of how conservation and development priorities are evolving across the region.

You can view the full code, visuals, and documentation here:🔗 [**GitHub Repository – newtAnalysis**](https://github.com/msheikh1/newtAnalysis)