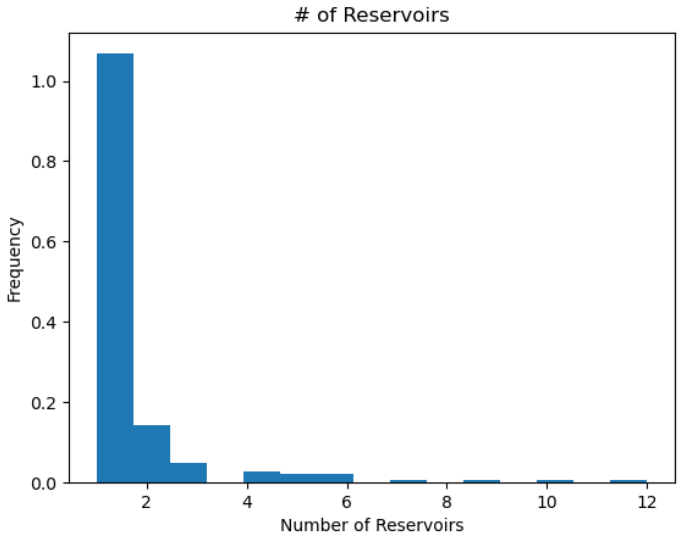
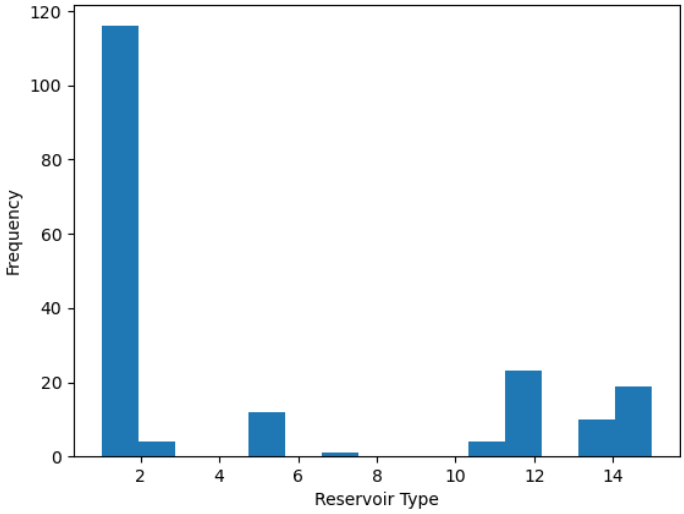
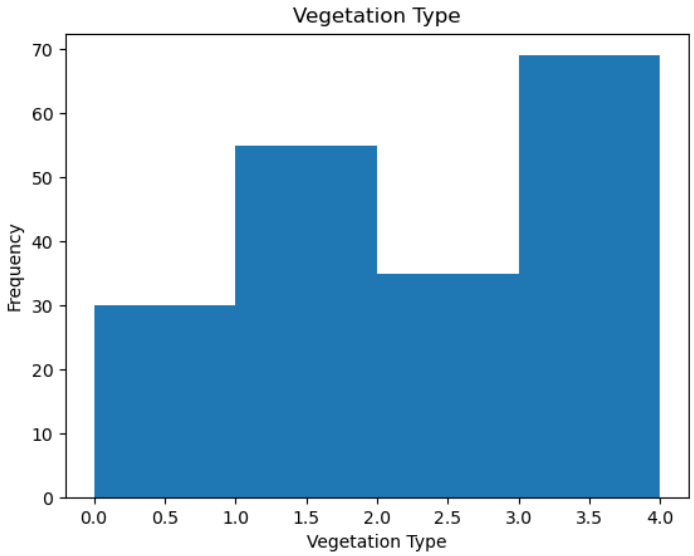
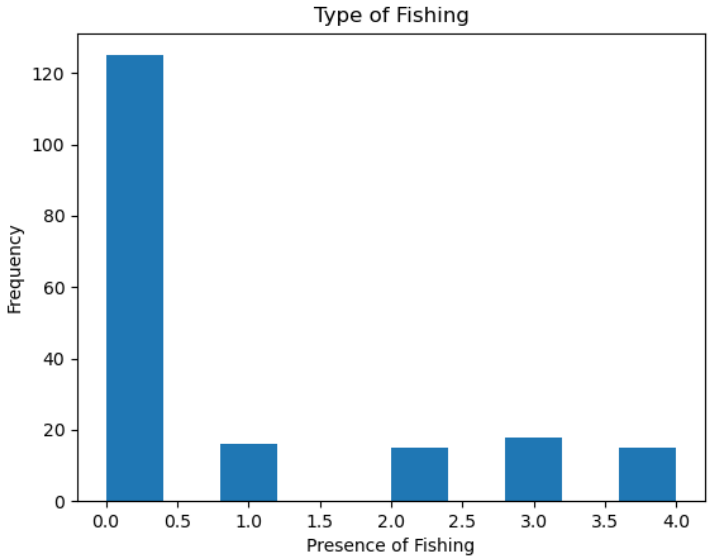
We did Reservoir Size (m^2) already. Run quick checks on each attribute, talk about which distributions to use, then maybe explore some multivariate ideas.

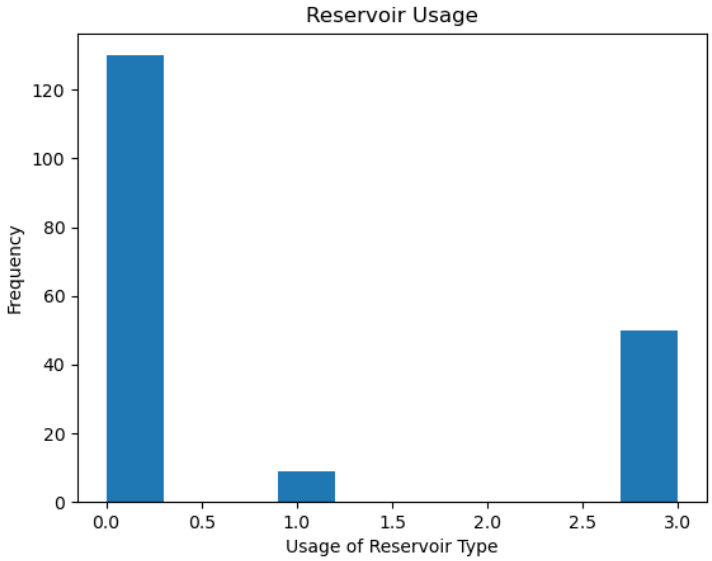


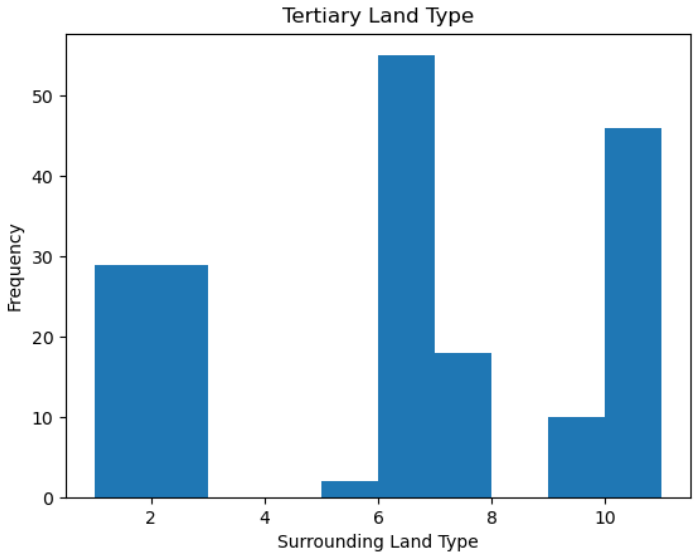


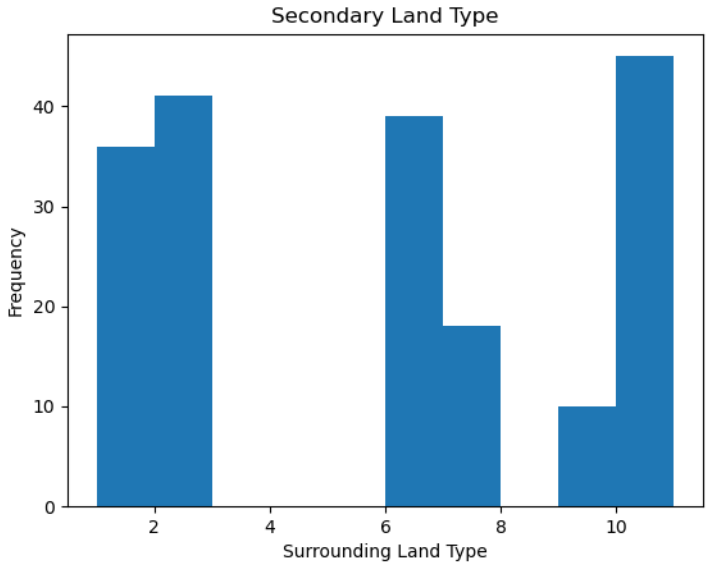
This one is actually great for multivariate, a bit unfortunate that most of our data is in Type 1, but that may actually be beneficial and highlight something valuable if we check just the 1s.

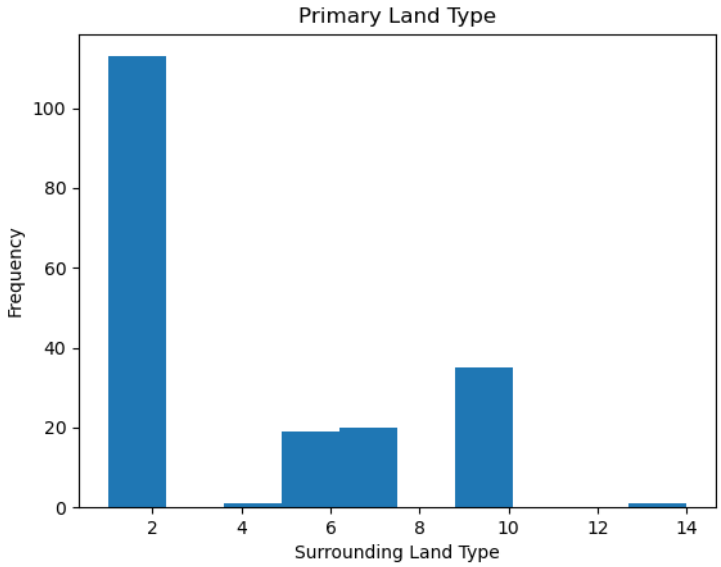

Vegetation Type
Frequency
0
10
20
30
40
50
60
70
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Vegetation Type


Type of Fishing
Frequency
Presence of Fishing
0
20
40
60
80
100
120
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Reservoir Usage
Frequency
Usage of Reservoir Type
0
20
40
60
80
100
120
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Tertiary Land Type
Frequency
0
10
20
30
40
50
2
4
6
8
10
Surrounding Land Type





I recognize these histograms are both inaccurate (in some cases) and useless.
I also didn't find a method for ordinal data, or try out power/student t/beta/etc distributions. This dataset is quite complex in comparison to Iris, and would take at least another week to explore.
Out of time. I will double the time invested on the next project (and hopefully take half the time for the implementation part - for GMM this time)

While this is unfinished for now, I will come back to it on my own time.

- **First to Iris for the GMM**
- **Then to explore more in detail on this set.**