**BShelly\_Fletcher**

**5/30/2023**

Section 6 Homework – Capstone development

Are job seekers in the Information Technology, specifically Data Analysis field affected in a negative way by false job postings? Within the past few years the Data Analysis field has taken off, and will only continue to grow exponentially in the years to come:

I want to use the data I’ve found to identify whether the false ads are created to obtain personal data from potential phishing or cybersecurity attacks?

* Are they created by recruiters to gauge the response received by potential employees for future opportunities?
* How can information obtained from job seekers responses be used in potential phishing or cybersecurity attacks.
* What can job seekers be on the lookout for when applying to online jobs.

1. Datasets:

This information will be identified through analyzing several datasets that contain job postings from several websites. The data I will be analyzing will be the posting sites, the job title and the locations of these jobs.

* dataset\_EMSCAD.csv (This is the primary dataset I will use to compare against the other datasets. This data set contains information pulled from multiple job positing sites into one .csv
* fake\_job\_postings.csv (This dataset was obtained from Kaggle.com international job postings 9/2021)
* gsearch\_jobs.csv( This dataset contains data pulled from Google search results for Data Analyst positions in the US)

1. The Approach:

I plan on analyzing data sets listed above to extract/compare first the Data Analysis jobs posted worldwide via dataset 1/dataset 2

Once I’ve extracted and compared the worldwide job results, I then plan on pulling in the 3rd dataset to compare the number of international Data Analysis jobs vs. those in the Domestic United States.

I want the data to show, if it is at all possible, the companies to lookout for, what region of the world/country these offers may originate from and which job posting sites have the most reports of false job postings.

I will use SQL and or Python to join the datasets, once complete I will use Tableau as a data visualization tool.