**She Code Africa Data Science Technical Assessment**

The dataset downloaded was a database containing a collection of job postings, some of which are fraudulent. The data contains textual & meta info about these jobs. The task was to study the details of the dataset and find out which industry has the highest number of fake job postings and then graphically display your answer.

Problem Statement: Which industry has the highest number of fake job postings?

Approach;

* The dataset contained 17,881 rows and 8 columns. It was then cleaned by removing the unnecessary columns and only three(3) columns left.
* Then ,filtering the rows which have (1) as their fraudulent number, as that shows they are positive for fake postings.
* After filtering for positive fake postings, the columns reduced to 866 columns.
* The empty values in the industry column was then replaced with ‘Unknown’, so as to be able to quantify it.
* Next, a pivot table and chart was created for visualization of the data.

Result;

* The Unknown values were a lot and that shows the number of empty cells present in the data.
* The industry with the highest number turned out to be the Oil and Energy Industry. The visual can be found below.

