**H1B Petitions SQL Project**

**Purpose**

As an international student, my goal is to analyze H1B petition trends across industries, workspace durations, salaries, job titles, and other variables. I'm particularly focused on all job titles, aiming to provide insights into salary distribution, top hiring states, and key companies involved in H1B visa sponsorships for the year 2015. The aim is to capture a holistic view of the H1B visa landscape during this period and extract valuable insights from the available data.

**Gathering data and Transformation**

Originally, this dataset I downloaded included over 1,000,000 rows of data and thus caused many errors when trying to transform the data. I realized that there were two major transformations that I needed to make

1. Limit the dataset to 2015 values only and 10,000 rows of data
2. Create two Excel spreadsheets: general and employ with a unique identifier case\_id.

**Tools used for analysis:**

* Excel: Pivot table, data visualization, data cleaning
* SQL: JOIN, SUBQUERIES, aggregated functions

**Questions answered:**

* What are the top three states with Data Engineer job titles in 2015?
* What are the top 10 salaries of the sample of H1B petitions in 2015? Where are the visas applied to? What companies are applying for these H1Bs?
* By how much do managerial positions make on average more than analyst positions?
* Provide the average wage for all the states
* The highest wage for the above query is Maine with an average salary of 141627.73.

Which company is applying for such a high salary petition? List the top companies and the wages.

**What to improve on in the future**

* Be comfortable working with large datasets.
* Incorporate more functions. Such as ranking using DENSE RANK and creating tables
* Commenting while coding, allows for better flow during the analysis process