**Job Dissatisfaction - Analyzing Employee Exit Surveys**

In this project, we are going to analyze two employee exit surveys done by the [Department of Education, Training and Employment](https://en.wikipedia.org/wiki/Department_of_Education_and_Training_(Queensland) (DETE) and the [Technical and Further Education](https://en.wikipedia.org/wiki/Technical_and_further_education) (TAFE) institute in Queensland, Australia.

TAFE Queensland is the largest, most experienced training and educator provider in the state. They deliver practical, industry-relevant training across a range of industries from entry-level certificates to bachelor’s degrees, at more than 60 locations in Queensland. They offer a range of services to students, staff, and the local community.

**Business Questions:**

We will analyze both surveys and answer the following questions:

* Are employees who only worked for the institutes for a short period of time resigning due to some kind of dissatisfaction? What about employees who have been there longer?
* Are younger employees resigning due to some kind of dissatisfaction? What about older employees?
* How many people in each career stage resigned due to some kind of dissatisfaction?
* How many people in each age group resigned due to some kind of dissatisfaction?
* Did more employees in the DETE survey or TAFE survey end their employment because they were dissatisfied in some way?

**Data Cleaning:**

Here´s the workflow followed to answer the above questions:

1. initial data sets exploration and identifying the columns needed for this project.
2. Each dataset contains many of the same columns, but the column names are different. Hence keeping the column names same in both Tefe and Dete datasets
3. Date Columns:

Check for the anomalies in dates containing columns; verifying that the years in the “cease\_date” and “dete\_start\_date” columns make sense. Since the “cease\_date” is the last year of the person's employment and the “dete\_start\_date” is the person's first year of employment, it wouldn't make sense to have years after the current date. Given that most people in this field start working in their 20s, it's also unlikely that the “dete\_start\_date” was before the year 1940. If we have many years higher than the current date or lower than 1940, we wouldn't want to continue with our analysis, because it could mean there's something very wrong with the data. If there are a small number of values that are unrealistically high or low, we can remove them. However, there do not appear to be any major issues with the values. The span of the cease years for both datasets is a bit different:

* + - * DETE: 2006 - 2014
      * TAFE: 2009 – 2013

1. The length of time an employee spends in a workplace is referred to as their Years of service. The TAFE dataset contains a column called “institute\_service”. Unfortunately, the DETE dataset does not have such a column. We do, however, have the needed data to create this column. It should contain the difference between the “cease\_date” and the “dete\_start\_date” columns.
2. filtered only the answers from the respondents who resigned, leaving out the respondents who left the job due to other reasons (i.e. retirement, expired contracts, etc.)
3. Identifying Dissatisfied employees: There are multiple columns/answers that indicate an employee resigned because they were dissatisfied. If the employee indicated any of the factors below to cause them to resign, we'll mark them as dissatisfied in a new column.
4. TAFE dataset:
   1. Contributing Factors. Dissatisfaction
   2. Contributing Factors. Job Dissatisfaction
   3. Contributing Factors. Interpersonal conflict
   4. Main Factor. Which of these was the main factor for leaving?
5. DETE dataset:
   1. job\_dissatisfaction
   2. dissatisfaction\_with\_the\_department
   3. physical\_work\_environment
   4. lack\_of\_recognition
   5. lack\_of\_job\_security
   6. work\_location
   7. employment\_conditions
   8. work\_life\_balance
   9. workload
6. Clean the 'institute\_service' column: As it contains values in a couple of different formats, to analyze the data, we'll convert these numbers into categories. We'll base our analysis on this [article](https://www.businesswire.com/news/home/20171108006002/en/Age-Number-Engage-Employees-Career-Stage), which makes the argument that understanding employee's needs according to career stage instead of age is more effective. We'll use the slightly modified definitions below:

* New: Less than 3 years at a company
* Experienced: 3-6 years at a company
* Established: 7-10 years at a company
* Veteran: 11 or more years at a company

**Data Analysis:**

*< Please check the Excel file “EmployeeExitSurvey\_analysis” for data analysis, report and observations >*

Thank You!