**Assignment 3 – Baby Project  
Jin Hyung Lee – A01365180 – BABI4005**

Github Link - <https://github.com/ejh6/JinTest>

# Data Source

The dataset used for this assignment was obtained from Government of Canada “**Proactive Disclosure - Travel Expenses - Proactive Disclosure - Travel Expenses”.** It contains information on Government officials traveling expenses. The dataset was downloaded on the 28th of January.

# Reflection Questions

## What do your three files look like online? Which one looks best?

* .ipynb files appears as an interactive Jupyter Notebook with code cells and outputs.
* .html file appears big blog of html code, outputting visuals separately
* .md file is simple and clean showing visuals elements, with ‘style’ code occasionally appearing in between the code.
* **Best Looking File:** The .ipynb appears the best, with no other code showing, showing the visuals nicely on the underneath the code.

## Do you think there are any problems with your dataset? Explain.

* Some organizations names are too long, making the table hard to read in the visualization
* Various of missing or incomplete data in some categories
* Outliers with expenses, distorting the overall analysis

To fix this issue, perhaps using the automation cleaning we learned in class will help narrow down the issues with this dataset.

* Standardizing long organization names for better readability.
* Filling or removing missing values to ensure data consistency.
* Identifying and handling outliers to prevent skewed results.

## What Questions do you have about the data?

* Are there any trends over time, such as seasonal spending patterns?
* Are there certain organizations that consistently have higher expenses than others?

## Is there a problem you would want to solve? Something to predict?

* Predict future expenses based on past trends
* Is there a spending habit based on their position?