**Introduction**

The goal of this project is to do an exploratory data analysis (EDA) to analyse and investigate datasets and summarize their main characteristics.

# **Data Description**

A dataset of 10,000 participants from four different age bands in five different locations must complete three different tasks. The score of each task for everyone was measured from 0 to 100 in percentage. The relationship of demographics and location with the scores and time taken to complete the tasks will be analysed.

**Features:**

10000 rows, 8 columns.

Gender: M or F

Location: A, B, C, D, E

Age.Band: U = (Under 16 (< 16), Y = Young Adult (16 – 39), M = Middle Aged (40 – 64), O = Older Adult (65+))

Completion.Time: Time (in seconds) to complete tasks A to C

Age: Numeric Value

Part.A.Score: Percentage (0 to 100)

Part.B.Score: Percentage (0 to 100)

Part.C.Score: Percentage (0 to 100)

# **Questions**

Q1. Is there a relationship between age and completion time?

Q2. Does age band matter regard scores?

Q3. In which location people take less time?

Q4. What is the mean percentage in every task?

**Needs**

1. Find any outliers in the dataset.

2. Find any relationship between the features.

3. What is the most interesting relationship?

**Tools**

* Technologies: Python, Jupyter notebook.
* Libraries: NumPy, Pandas, Matplotlib