

# Exploratory Data Analysis

---

- StudentID: 21700567
- Name: Juchan Lee
- 1st Major: ICT Convergence
- 2nd Major: Data Science

*Brief summary of your proposed project idea.*

## 1. Data overview

Descriptives statistics on overall data (sample size, number of variables, data type, data range, distribution, etc.)

The data contains the results of General Social Survey(GSS) conducted in United States. The data consists of 59600 rows and 11232 columns. The initial data has some problems to explore data. For instance, there are some unnecessary variables which contains labels of each numerical vairables (e.g., "NUMBER OF HOURS WORKED LAST WEEK\_labels" refers to the labels of "NUMBER OF HOURS WORKED LAST WEEK), data has two data types: float64 and object, however, most of object type variables contains numerical values, the actual column names are stored in the first row. Thus, proper preprocessing is needed.

## 2. Univariate analysis

Presentation of key variables from various aspects

### 2.1 (ex) Variable 1

~~

 BTS Google Trend

Figure 1. BTS Google Trend

### 2.2 (ex) Variable 2

~~

### 3. Multivariate analysis

Presentation of hidden patterns between variables (correlation, clustering, etc.)

#### 3.1 (ex) Correlation

~~

#### 3.2 (ex) Clustering

~~

### 4. Suggestion

Based on the insights you obtained from the previous stages, propose the potential project idea.