

# **Readme**

## **Data Source**

The datasets of this paper are abstracted from the 2016 General Social Survey (GSS) on Canadians at Work and Home. The 2016 GSS on Canadians at Work and Home was conducted from August 2nd to December 23rd 2016. The target population includes all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada. Data are subject to both sampling and non-sampling errors.

The General Social Survey (GSS) is an annual survey of Canadians that aims to gather data on living conditions and well-being as well as provide information on social policy issues of current or emerging interest. The GSS was established in 1985 and conducts telephone surveys across the ten provinces. Core content is designed to measure changes in society related to living conditions, and to supply data to inform policy issues.

## **work\_ethic\_dataset**

The work\_ethic\_dataset is extracted from the part of “Life at work” in the whole database of 2016 GSS. There are 8 variables in this dataset which are all related to how employees feel about their job in some specific scenario. All that information is collected by survey questions and the responses are recorded by degree level from 0 to 10, where 0 being ‘completely disagree’ and 10 being ‘completely agree’ to the related survey statement.

## **income\_dataset**

The Income\_dataset is extracted from the part of “Life at work” in the whole database of 2016 GSS. There are 2 variables in this dataset. The name of the first variable is TTLINCG2 which represents the personal income group (before tax) of all research objectives. The name of the second variable is FAMINCG2 which represents the family income group (before tax) of all research objectives. The data of all those variables are recorded according by the frequency of answer categories which are Less than \$25,000, \$25,000 to \$49,999, \$50,000 to \$74,999, \$75,000 to \$99,999, \$100,000 to \$124,999, \$125,000 or more, Valid skip, Don’t know, Refusal, and Not stated.

## **after\_cleaned\_dataset**

We combined the work\_ethic\_dataset and income\_dataset and removed the FAMINCG2 variable. Then a data cleaning process has been made to get the final dataset which is after\_cleaned\_dataset. This dataset contains 9 variables and excludes any missing value, which is ready for further analysis and model building.

