

# Assignment 1 : CS4038 DM Winter 2019

**Step 1:** Download the world-happiness dataset from Kaggle.com from the link

<https://www.kaggle.com/unsdsn/world-happiness>

**Step 2:** Download the customized world development indicators data uploaded in EduServer.

Original data is from <https://databank.worldbank.org>

**Step 3:** Merge both the datasets, suggest doing the following for achieving this:

- a) Bring happiness data into the format  
Country Code/Name, Year, Happiness Score (ignore error & the split up scores)
- b) Bring Development Indicators data into the format  
Country Code/Name, Year, Indicator 1 Value, Ind 2 Value, .... Ind 35 Value
- c) Join the datasets on Country Code/Name, Year

**Step 4:** Create a model which can predict the happiness index from the selected indicators.

**Step 5:** Analyze various indicators and see which of them can predict the happiness score better, you can use last year's data as test data.

**Note:** You can even try to predict a happiness class (Poor, Average, Good, Best) instead of the predicting the score.

**Step 6:** Summarize your findings as a small one/two page report.