**To execute ETL process.**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Activity:**

# **For ETL:**

1. Go through the tutorial provided by RapidMiner
2. Extract data from 2 to 3 heterogeneous sources such as excel, MYSQL, Postgres etc.
3. Download any data set from [*https://w*](http://www.kaggle.com/datasets)*ww.k*[*aggle.com/datas*](http://www.kaggle.com/datasets)*e*[*ts*](http://www.kaggle.com/datasets) or similar website
4. Apply five different transformations and filters to the data with specific requirement
5. Prepare a report for the activities 2 and 4 (ETL part) with steps and visualisations applied.



**Importing data set (any dataset)**

[**https://www.kaggle.com/datasets/shivkumarganesh/world-happiness-report-**](https://www.kaggle.com/datasets/shivkumarganesh/world-happiness-report-)[**20152022?resource=download**](https://www.kaggle.com/datasets/shivkumarganesh/world-happiness-report-20152022?resource=download)

1. **Check Basic Statistics**
2. **Check Happiness rank by region.**
   1. **Data visualization using median aggregation (pie chart and Bar chart)**

**3. Apply various charts styles to check for following data analysis**

# **3.1 Happiness rank vs. country –Scatter chart**

**3.2 Happiness score vs. country**

**3.3 Happiness score vs. region**

* 1. **Happiness rank vs. Economy**

**Write comment on each analysis.**

1. **Basic display using RapidMiner**
2. **Apply correlation matrix**
3. **Data normalization**