# Exercise 6.1

1. **Data sources** –

* **cwurData -**

Source: Kaggle

Owner-Meils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill. It is open source and can be trusted for further evaluation and analysis. This is the Center for World University Rankings ranking(CWUR) of world universities.

* **education\_expenditure\_supplementary\_data –**

Source: Kaggle

Owner-Meils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill. The data is from an open source and can be trusted as used for further evaluation and analysis. The data shows the name of the countries and expenditure incurred in education.

* **educational\_attainment\_supplementary\_data**

Source: Kaggle

Owner-M.eils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill. The data is from an open source and can be trusted as used for further evaluation and analysis.

* **school\_and\_country\_table–**

Source: Kaggle

Owner-Meils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill.. The data is from an open source and can be trusted as used for further evaluation and analysis**.** The data shows school names and their location.

* **shanghaiData–**

Source: Kaggle

Owner-Meils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill. The data is from an open source and can be trusted as used for further evaluation and analysis**.** This is the shanghai ranking(ARWU) of world universities.

* **timesdata–**

Source: Kaggle

Owner-Meils O’Neill

Data has been taken from Kaggle, owned by Meils O’Neill. The data is from an open source and can be trusted as used for further evaluation and analysis. This is the times ranking for the world universities.

1. **Data Collection Method –**

The data has been collected by 3 fully independent organizations - CWUR, ARWU, and Shanghai Ranking.

Shanghai Ranking (ARWU), Times Higher Education (THE), and Center for World University Rankings (CWUR) are three of the most well-known university ranking systems. Each ranking uses different methodologies and focuses on various aspects of university performance.

The data is an open source data.

1. **Data overview-**

**cwurData.csv** - CWUR data wherein world rankings are given by the Centre of World University Rankings.

**education\_expenditure\_supplementary\_data.csv** – Government expenditure on different types of universities.

**educational\_attainment\_supplementary\_data.csv –** This is the age-wise country data.

**school\_and\_country\_table.ccsv-** This data is about the schools or colleges and their base countries.

**ShanghaiData -**  Shanghai Data is the world rankings given by ShanghaiRanking Consultancy and is world wide popular.

**Times Data -** Times Data is the world rankings given by Times Higher Education and is worldwide popular.

1. **Data Inconsistency checks-**

Expenditure data had too many blank columns which were removed and for some values, the mean value was assigned to them.

In Shanghai dataset, the mean value was assigned to the ‘Award’ column.

No duplicate columns were present in any datasets.

**limitations of the data set**

The data has been collected and published by various consultancies; thus, we consider this data accurate and trustworthy. But the data made me question some elements which can be valuable-

* 1. The data is incomplete as it does not include all the universities in the world especially those in developing nations.
  2. The parameters of measuring could be different for all the nations. Hence a same parameter might not be able to justify the data.
  3. The data is biased as it only includes Universities of Developed nations.
  4. The expenditure data could be biased as many nations might show inaccurate figures.
  5. Many variables are blank in the data which can limit the analysis.

**Data Relevance**

All the datasets are highly relevant for addressing the project's objective and testing the stated hypotheses related to nations working on their education systems. The diverse datasets provide sufficient information from which this analysis can be made.

**Key Questions**

1. What are the key parameters to define the University rankings?
2. Which country has the most Universities in the top 500?
3. Which country has the most expenditure on education(Sector- wise)
4. How have the rankings changed over time? Are there significant trends or patterns in the performance of different universities or regions?
5. Top 5 universities per sector
6. What is average years of studying for female students in different countries?