IDS Assignment-1

Bharani Ujjaini Kempaiah
PES1201700005

Divya T Puranam
PES1201700749

About the Dataset and the Analysis

- We have carried out Exploratory Data Analysis on the dataset titled " World University Rankings".
- This dataset shows the ranking of universities across the world over a period of four years 2012-2015.
- The original dataset consists of 2200 rows and 14 columns.
- We also observed that the column titled 'broad_impact' was empty for the year 2012 and have suitably taken steps to counter the issue.
- Through visualizations, our insights into the data has been increased and some questions have been answered as well.
- Link to dataset: https://www.kaggle.com/mylesoneill/world-university-rankings

Data Cleaning

	world_rank
count	400.00000
mean	50.50000
std	28.90222
min	1.00000
25%	25.75000
50%	50.50000
75%	75.25000
max	100.00000

	broad_impact
count	400.000000
mean	70.600529
std	68.775163
min	1.000000
25%	25.750000
50%	54.000000
75%	89.125000
max	388.000000

- On observing the dataset, there was irregularity with the number of rows per year.
- Hence to make the data uniform, all the rows beyond 100 for a particular year were dropped, after which each year contains only the top 100 universities.
- The empty broad impact column for the years 2013,2013 has been suitably filled.
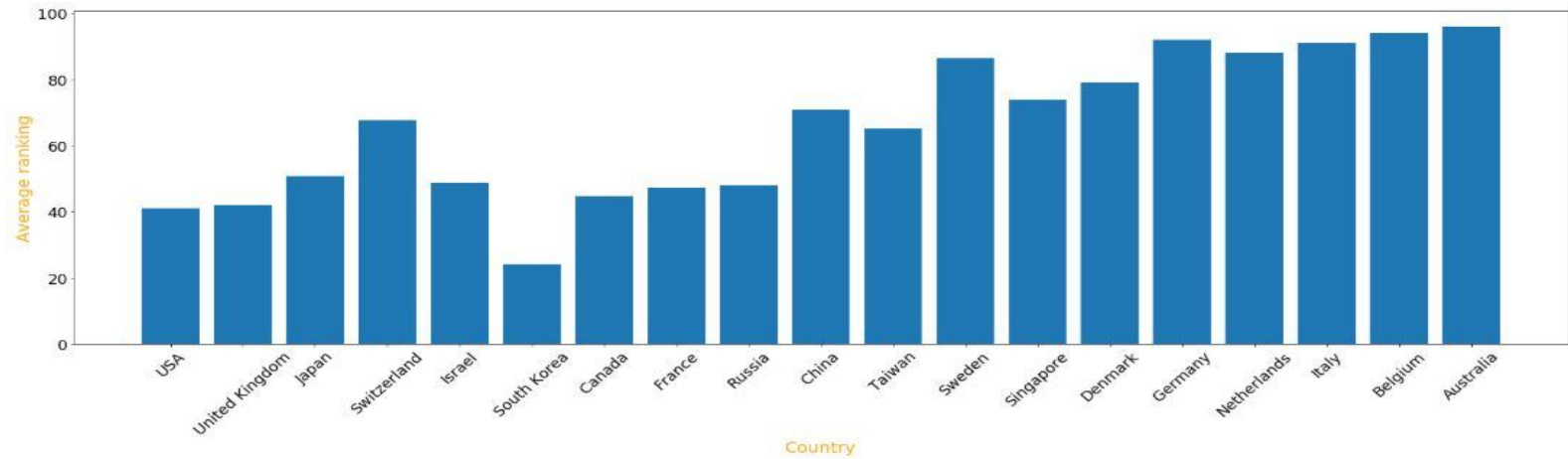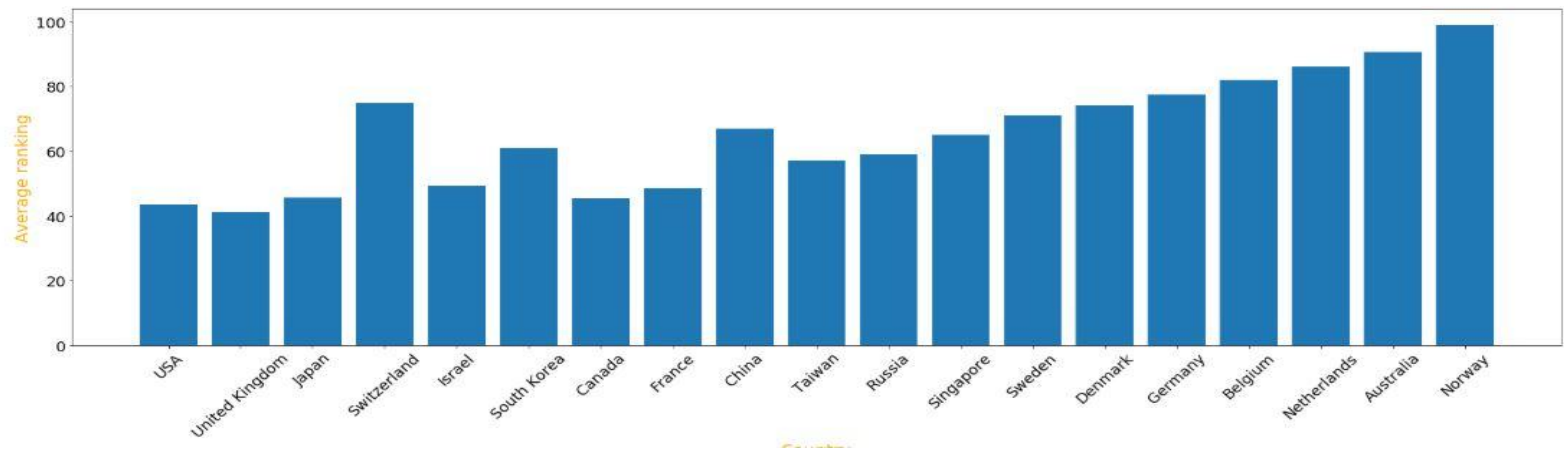
Visualization 1 : To determine the country-wise top 100 for all the years
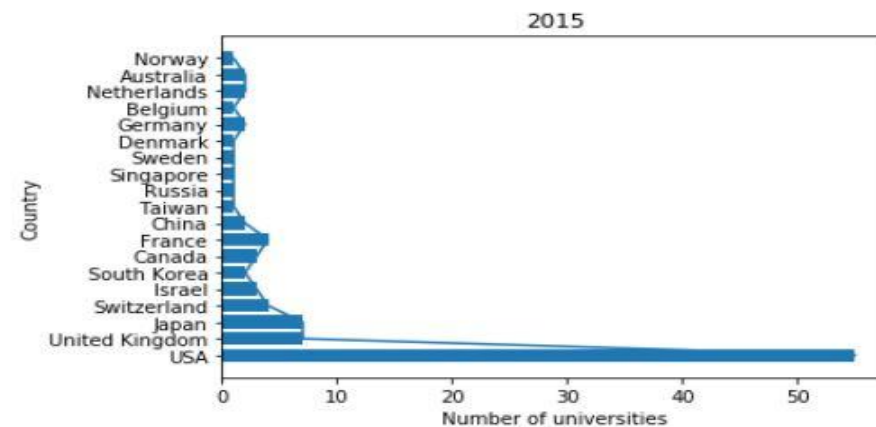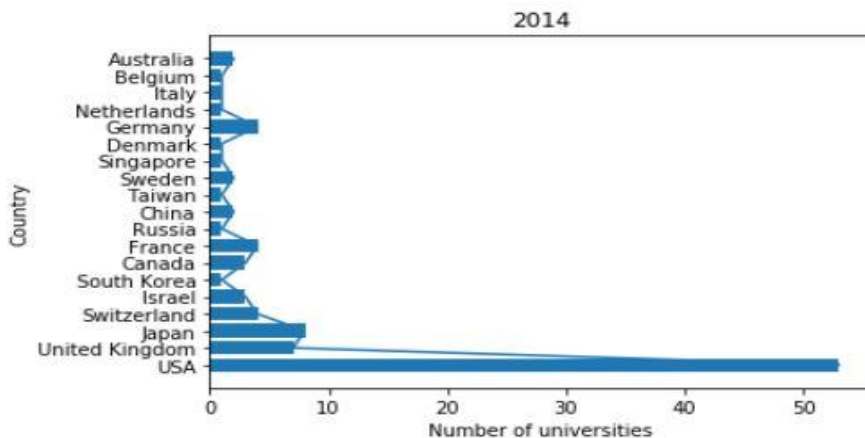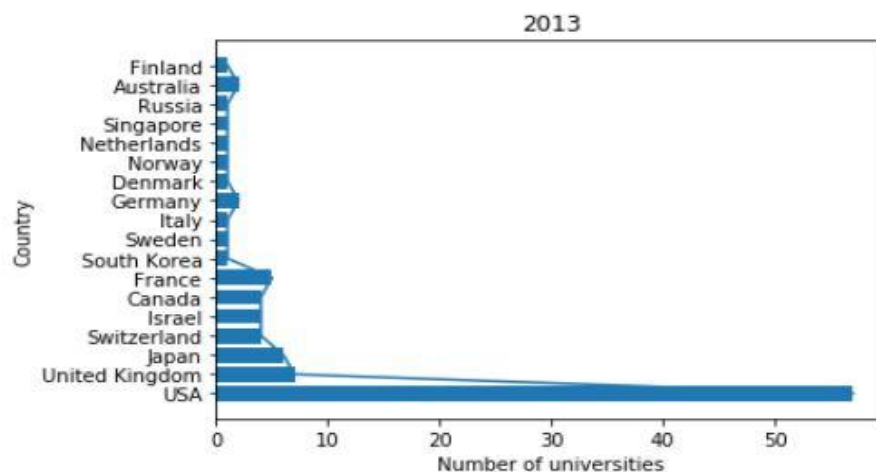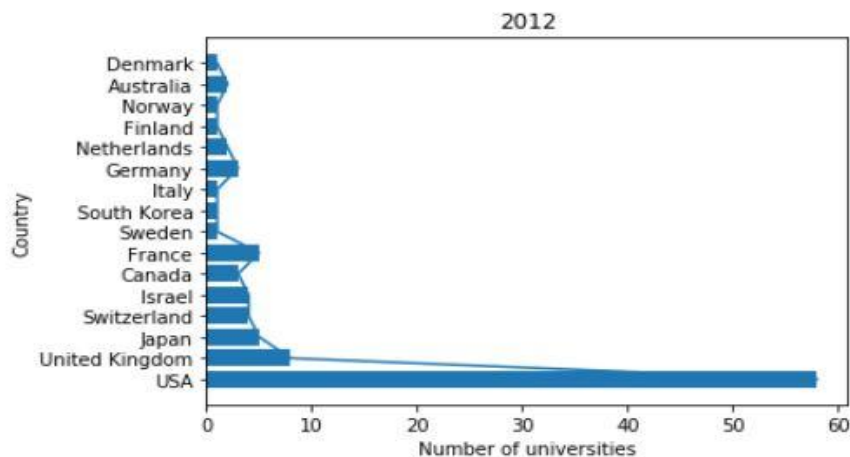
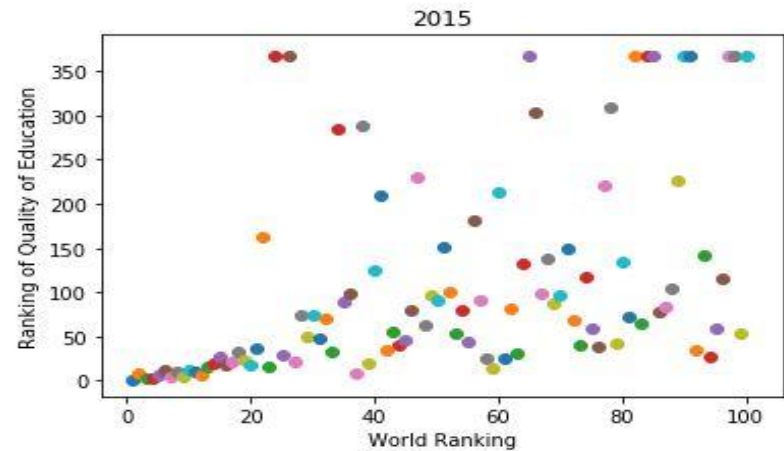# Visualization 2 – Average ranking country wise
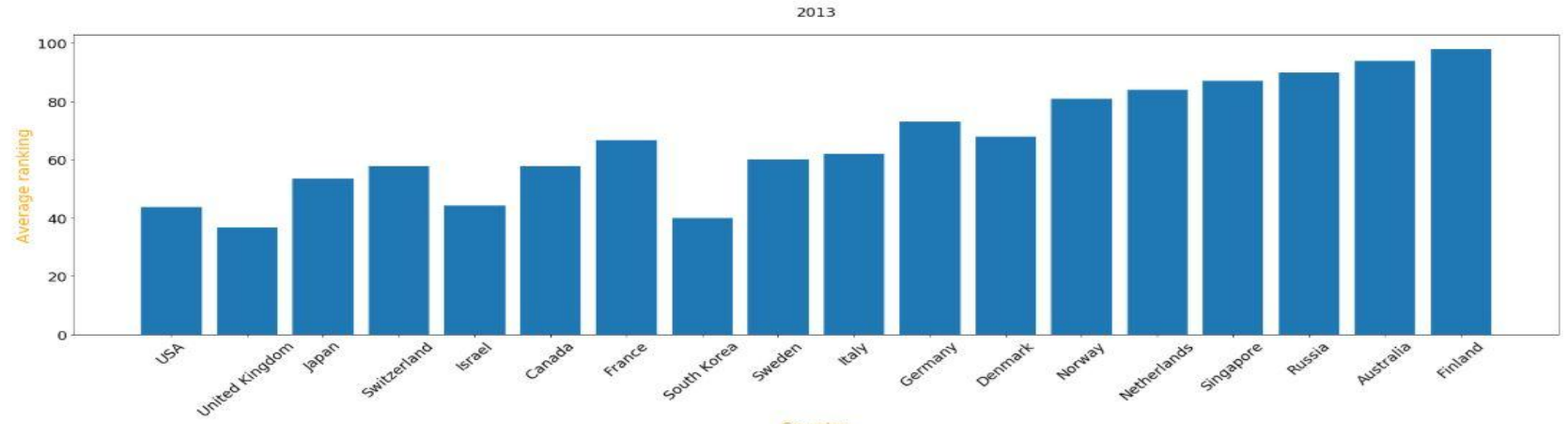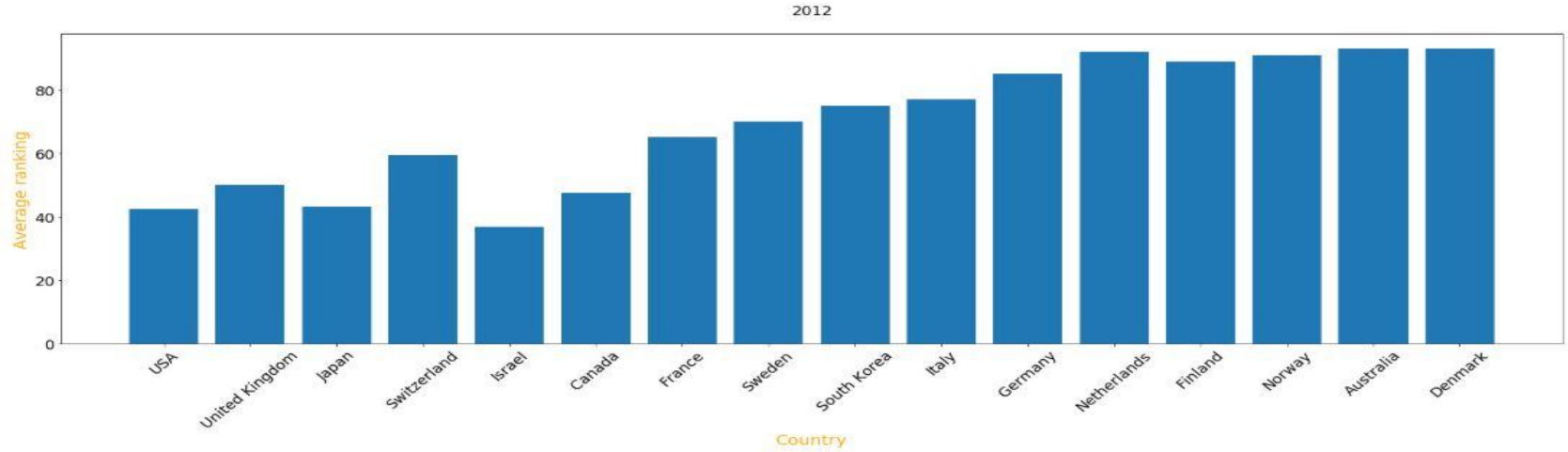
2014

2015

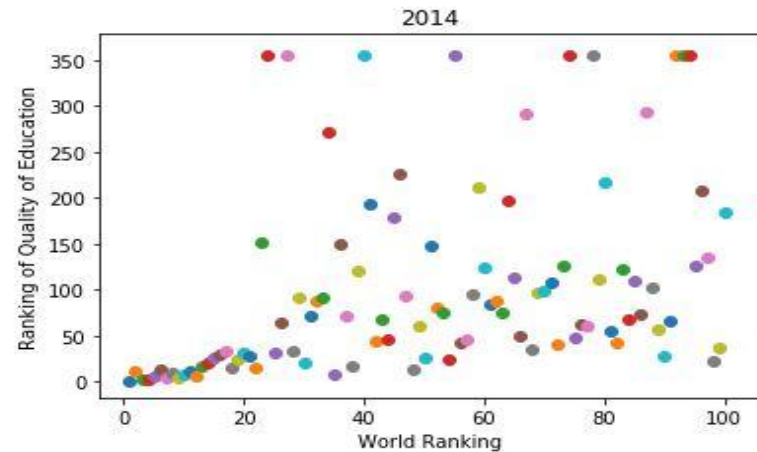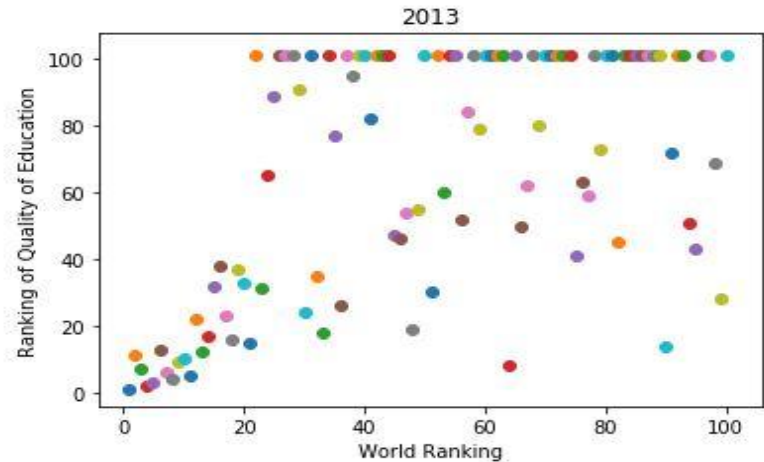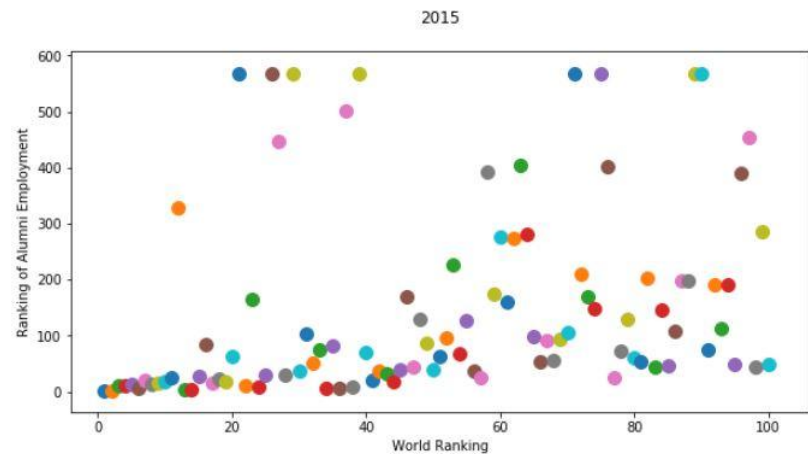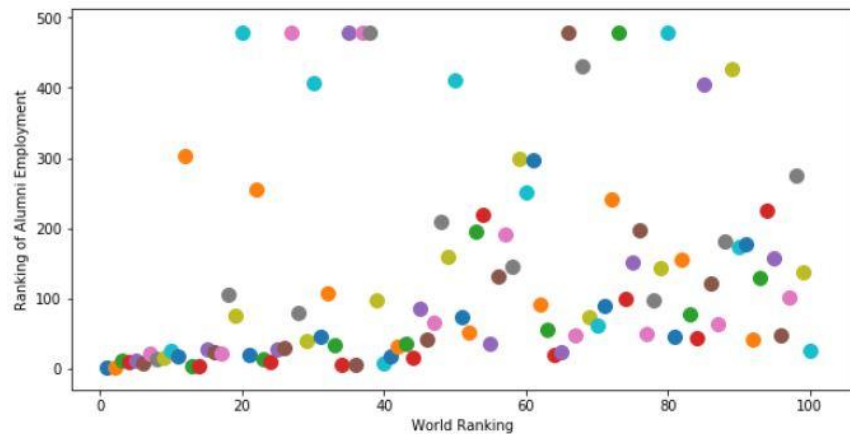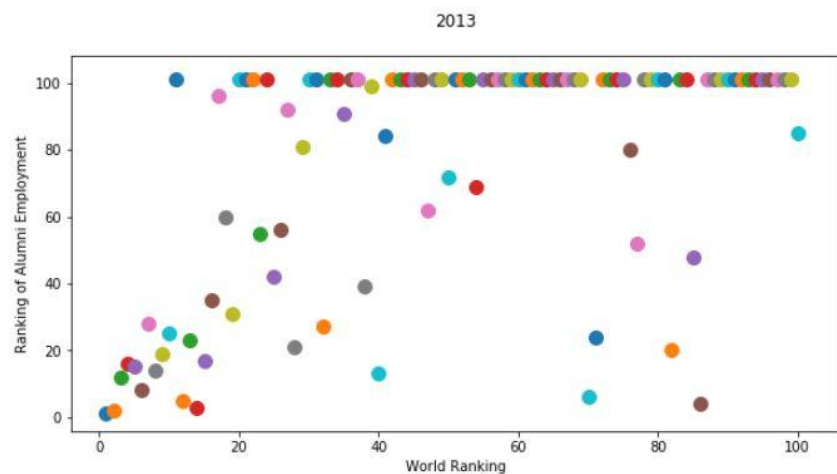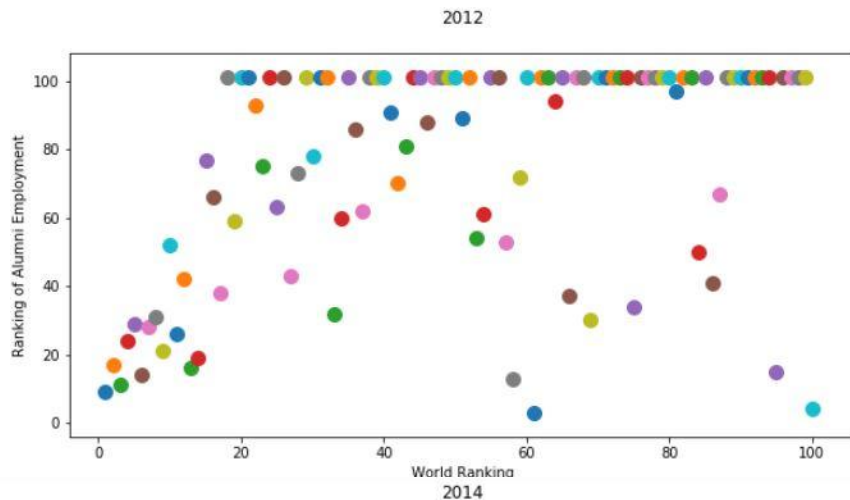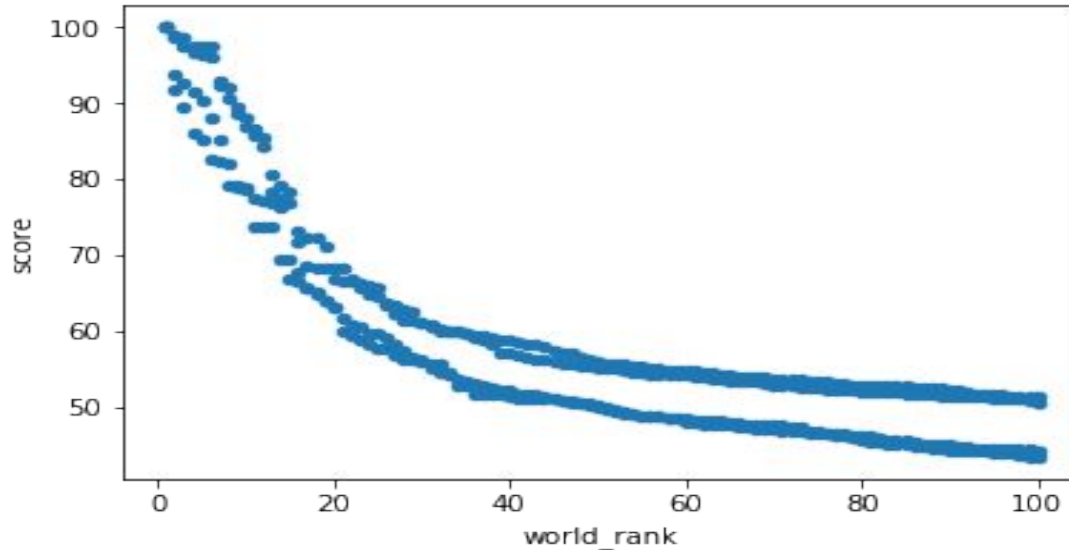# Visualization 3- World rank Vs Quality of education Ranking
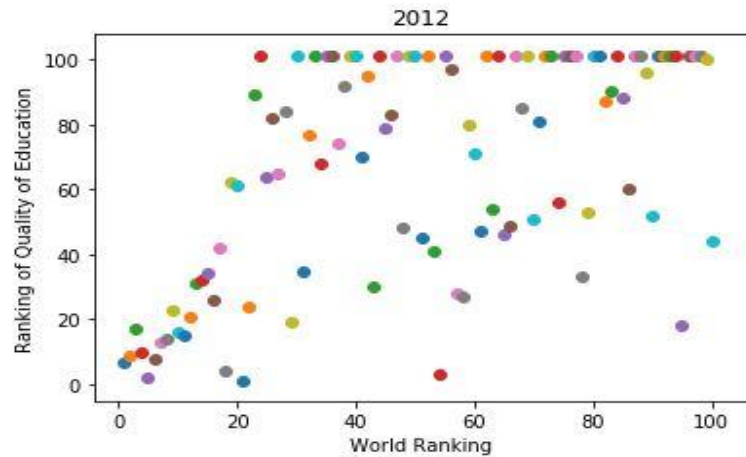
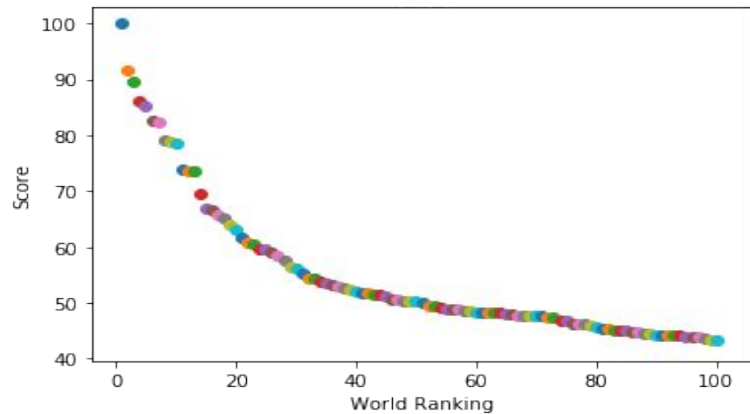# Visualization 4 – World rank vs Alumni employment

Visualization 5: Relationship between the world rank and score of the university

The following graph shows how the score varies almost linearly with the world rank of the university. It is seen that better the rank, the higher is the score.
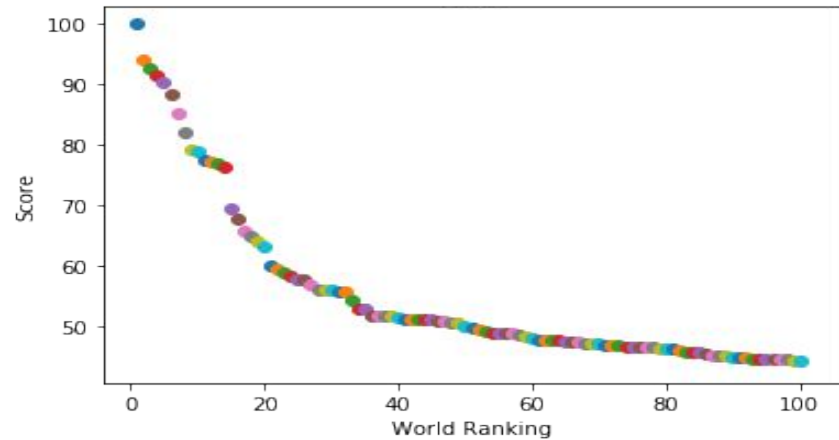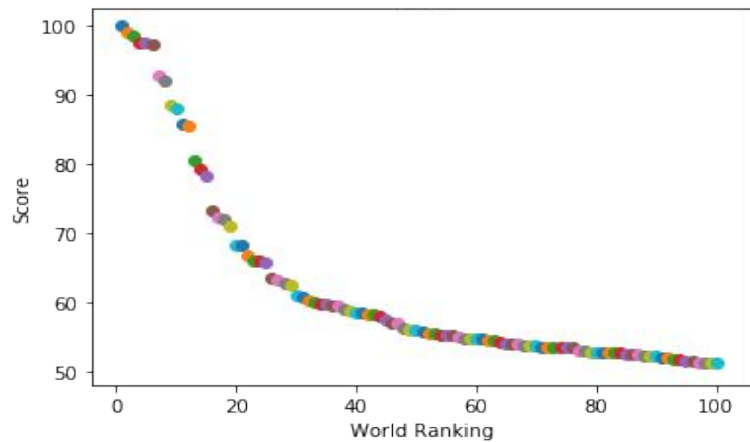
Visualisation 6 : World ranking vs Score per year

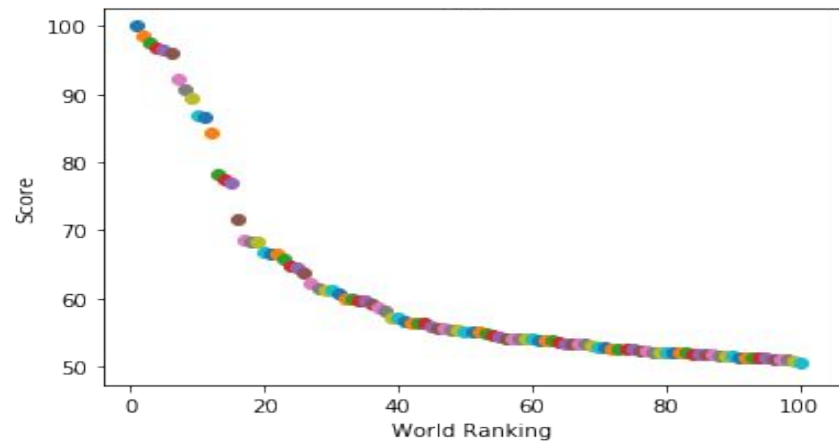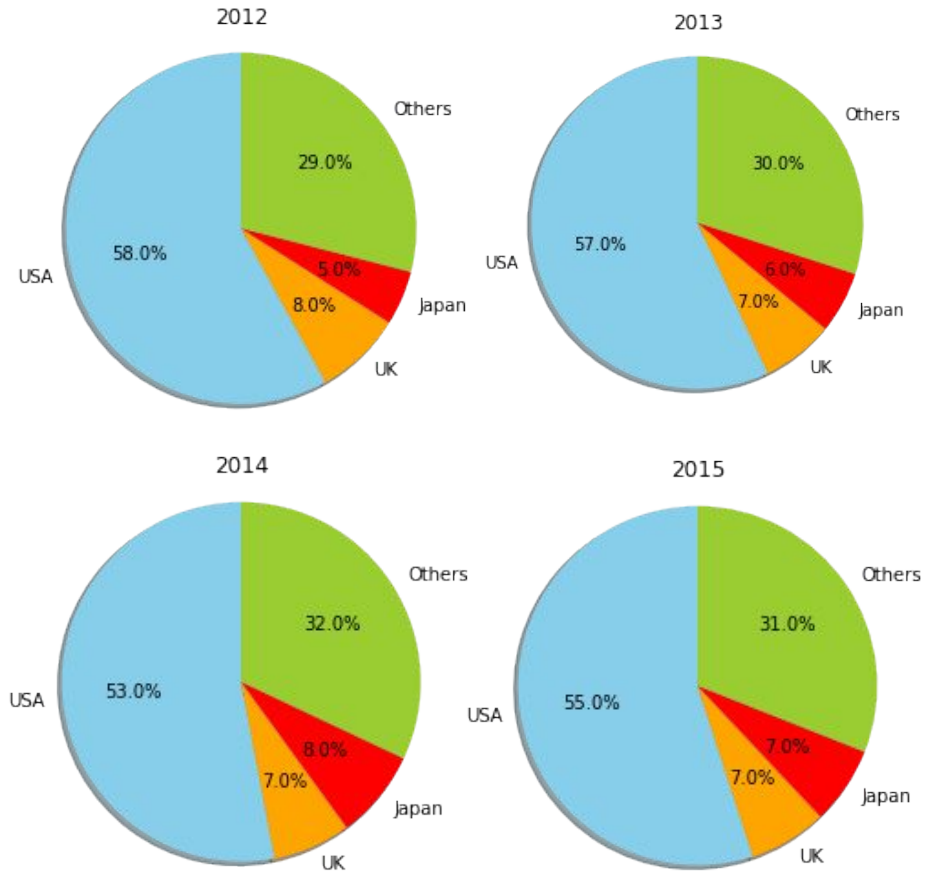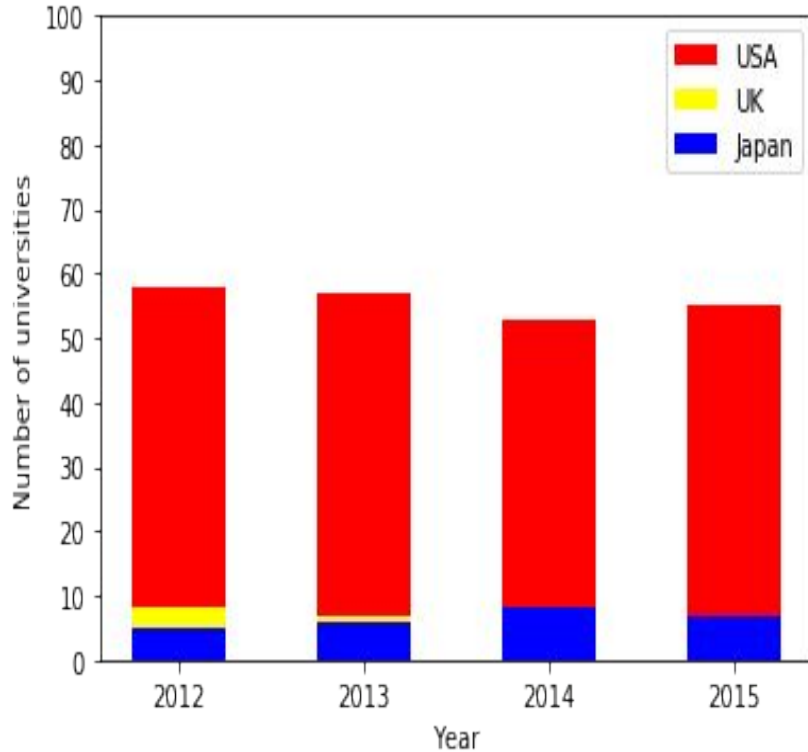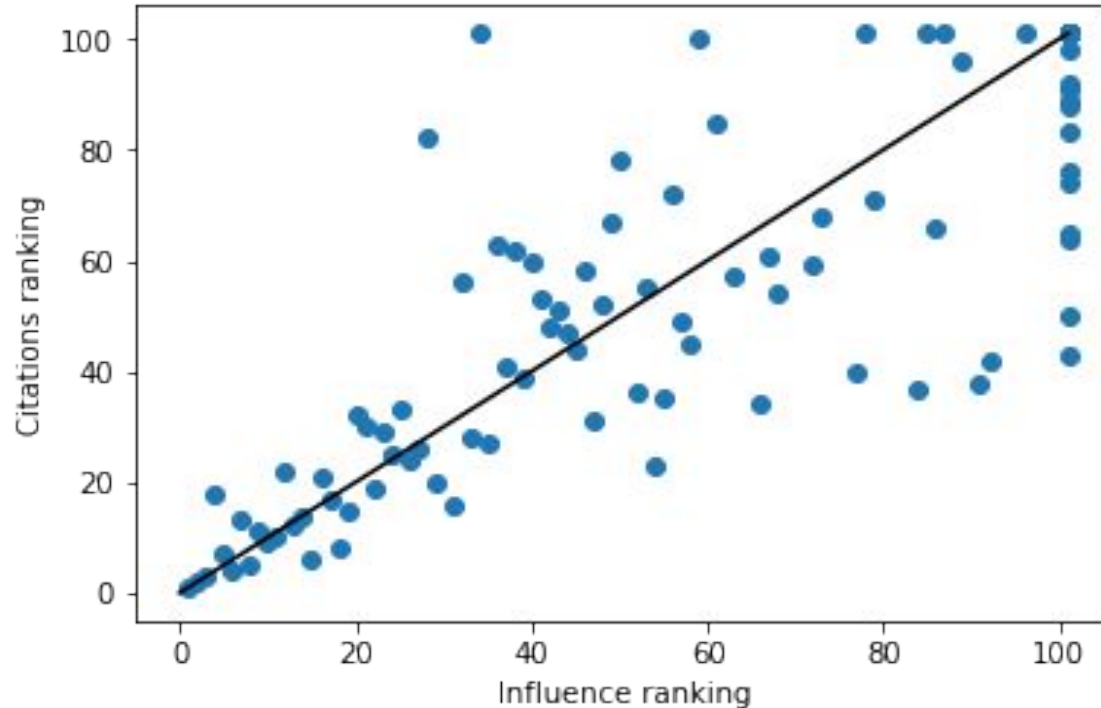# Visualization 7: Comparison between countries having most number of universities in the top 100

Visualisation 8: Citation vs Influence Ranking

From the graph, it is seen that the citations vary linearly with the influence ranking for the most part. That is, the higher the citation ranking of the university, the more influence and recognition it has.

Conclusion

➢ The world ranking of a university is a cumulative effect of the different parameters that are used in rating.
➢ We also observe that the countries with the best 100 universities are developed countries. This could be because of :
    Investment in Research and Development
    Better quality of Faculty
    Schools tend to be more holistic in their evaluation of application.
➢ A major portion of the best universities are in the US and our graphs do prove that the data is right.
➢ The US might be dominating now, but after a few years we are sure other developing countries like our own will definitely catch up.