Final Project 2: Exploratory Data Analysis on University Rankings

Yvonne Gan • 19.09.2020

Part A: Data Understanding and Exploration

Exploratory Data Analysis

Key elements:

- Inspect
- Understand the dataset
- Exploring the data
- Experiments of the data

Next Step?

Methods I used for EDA on the dataset:

- .head()
- .tail()
- .shape
- .describe()
- .isnull().sum()
- .value_counts()
- .isnull().any(axis=1)
- .dtypes
- .info()
- .sort_values()

Conclusion of Data Exploration:

Outcome

- Understood the type of date
- Overview on data usage
- Explored methods for other parts
- Learnt from mistakes

Exploration do not stopped here but continue throughout the analysis

Part B: 6 Questions

Question 01:

1. Which country (throughout all years) has the most number universities within top 100 of world_rank?
    - Created the count_100 variable
    - The variable contains - filtered of dataset for top 100(inclusive) world_rank across the years

	world_rank	institution	country	national_rank	quality_of_education	alumni_employment	quality_of_faculty	publications	influence	citations	broad
0	1	Harvard University	USA	1	7	9	1	1	1	1	
1	2	Massachusetts Institute of Technology	USA	2	9	17	3	12	4	4	
2	3	Stanford University	USA	3	17	11	5	4	2	2	
3	4	University of Cambridge	United Kingdom	1	10	24	4	16	16	11	
4	5	California Institute of Technology	USA	4	2	29	7	37	22	22	
...	...	...	...	...	...	...	...	...	...	...	
1295	96	University of Zurich	Switzerland	4	115	389	85	56	46	68	
1296	97	Arizona State University	USA	55	367	455	66	143	85	76	
1297	98	Yonsei University	South Korea	2	367	45	218	140	326	321	
1298	99	University of Oslo	Norway	1	53	286	40	106	134	101	
1299	100	Kyushu University	Japan	7	367	49	218	152	182	264	

400 rows × 14 columns

Question 01:

2. What are the top 10 countries by value of count_100?
   - To get the count of countries from the top 100 universities across all years
   - Value counts by the “country” column and from the output show the top 10 countries

USA	223
United Kingdom	29
Japan	26
France	18
Switzerland	16
Israel	14
Canada	13
Germany	11
Australia	8
Netherlands	6

Question 01:

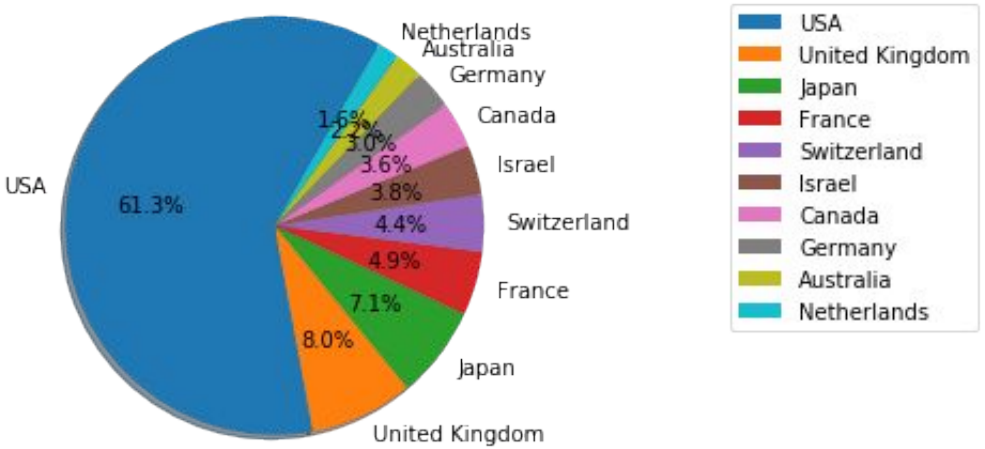
3. What is count_100 for each of these countries?
    - An overview of the top 100 universities under these countries
    - Groupby “country” followed by “institution”, then do the count by “institution”

country	institution	
Australia	Australian National University	1
	University of Melbourne	2
	University of Queensland	1
	University of Sydney	4
Belgium	Katholieke Universiteit Leuven	2
Canada	McGill University	4
	University of Alberta	1
	University of British Columbia	4
	University of Toronto	4
China	Peking University	2
	Tsinghua University	2
Denmark	University of Copenhagen	4
Finland	University of Helsinki	2
France	Mines ParisTech	1
	Paris Diderot University - Paris 7	1
	Pierre-and-Marie-Curie University	4
	University of Paris-Sud	4
	École Polytechnique	4

Question 02:

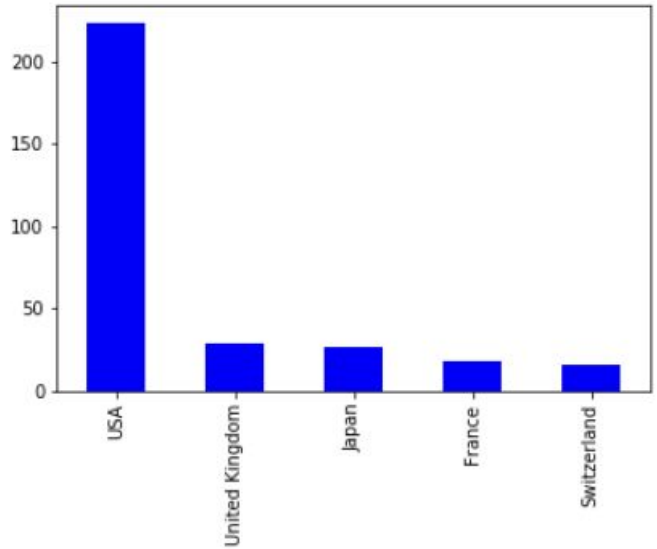
- What is the percentage composition of each countries' count_100 out of the 10 countries?
  - Using previous statement on top 10 countries with largest amount of top 100 university and rename the count column to “count_100”
  - Formula: (Count of top 10 countries with top 100 university/ Sum of top 100 university in top 10 countries)*100

	count_100	perc
USA	223	61.263736
United Kingdom	29	7.967033
Japan	26	7.142857
France	18	4.945055
Switzerland	16	4.395604
Israel	14	3.846154
Canada	13	3.571429
Germany	11	3.021978
Australia	8	2.197802
Netherlands	6	1.648352



Question 03:

- Plot a graph showing the top 5 countries from Q1 and their counts
  - From Q1 the list of Top 100 Universities, get the top 5 countries and plot a graph
  - Chosen “Bar Graph” - simple and easy overview



Question 04:

- Groupby year and country showing the count of universities with world_rank within top 100 for that country and year.
  - Groupby year → country → count the numbers of universities (Pic 01)
- Sort the year in ascending order while count_top100 in descending order.
  - Sort the year then counts to get table as Pic 02

```
year  country
2012  Australia          2
      Canada             3
      Denmark            1
      Finland            1
      France             5
                        ..
2015  Sweden             1
      Switzerland        4
      Taiwan             1
      USA               55
      United Kingdom     7
```

Pic 01

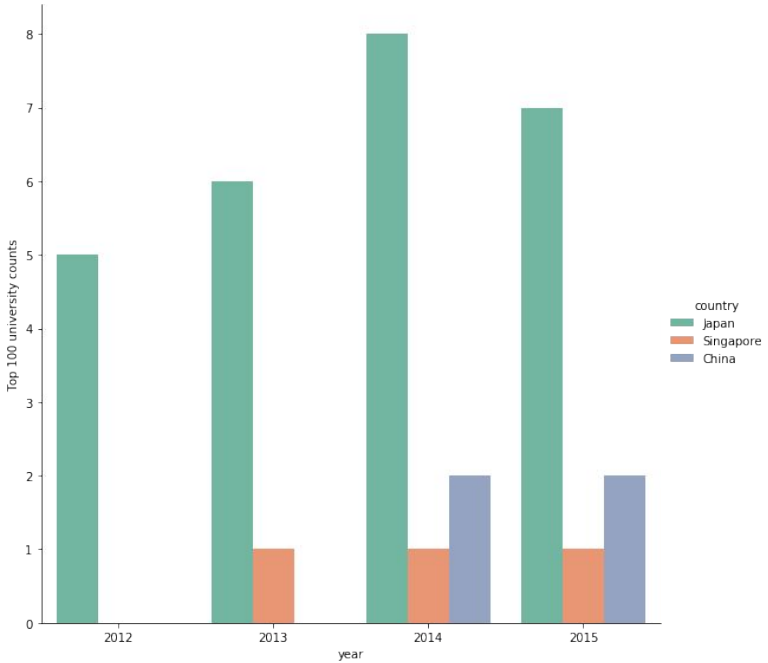
	year	country	count_top100
14	2012	USA	58
15	2012	United Kingdom	8
4	2012	France	5
8	2012	Japan	5
6	2012	Israel	4
...	...	...	...
63	2015	Norway	1
64	2015	Russia	1
65	2015	Singapore	1
67	2015	Sweden	1
69	2015	Taiwan	1

72 rows × 3 columns

Pic 02

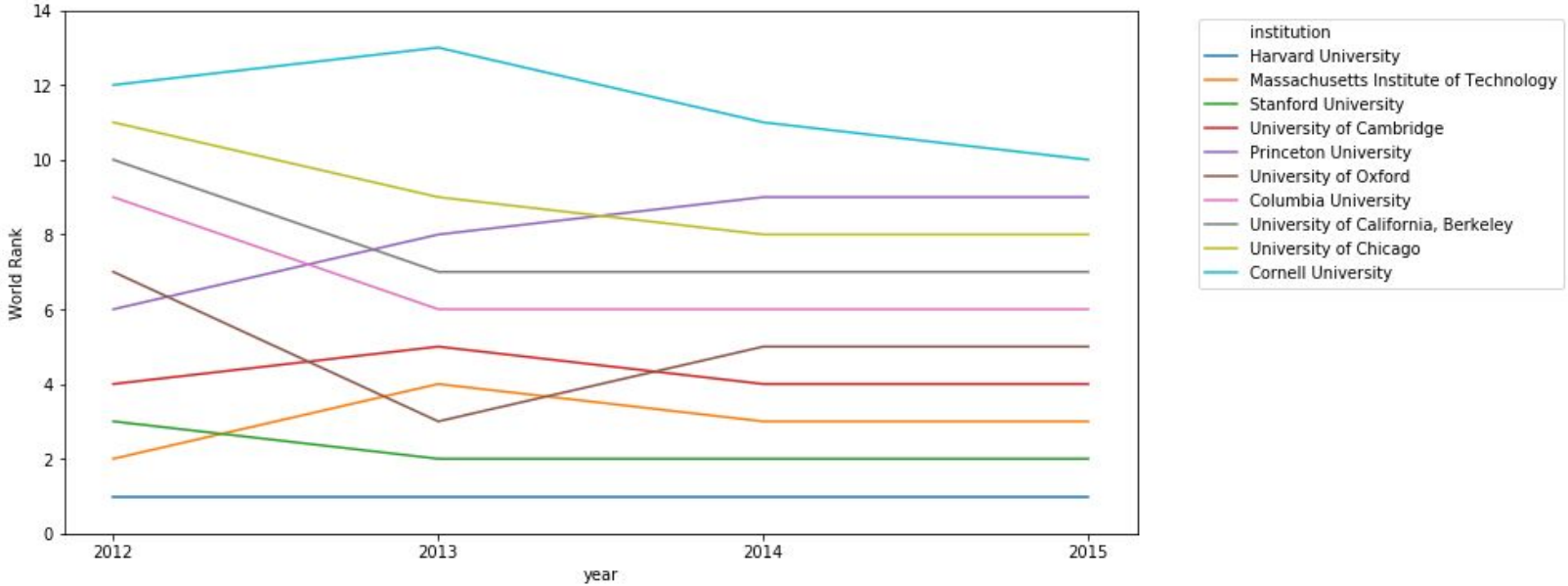
Question 05:

- The trend for the count of top 100 universities for Japan, China and Singapore for years 2012 - 2015



Question 06:

- For top 10 universities in 2015, what was the trend for world_rank for those universities across all the years? Which universities generally improved, and which ones declined in world_rank?



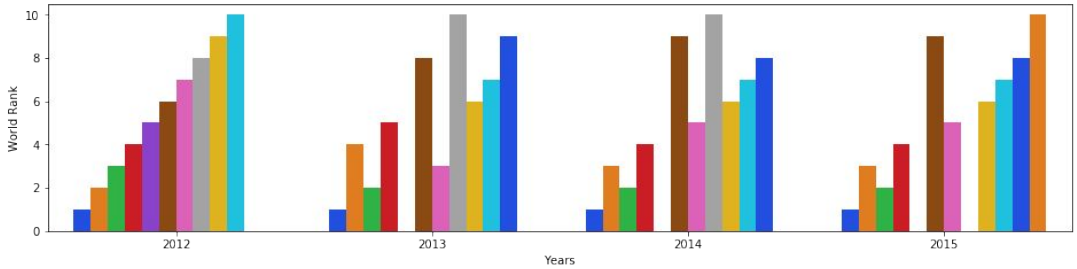
Part C: Open-ended Question

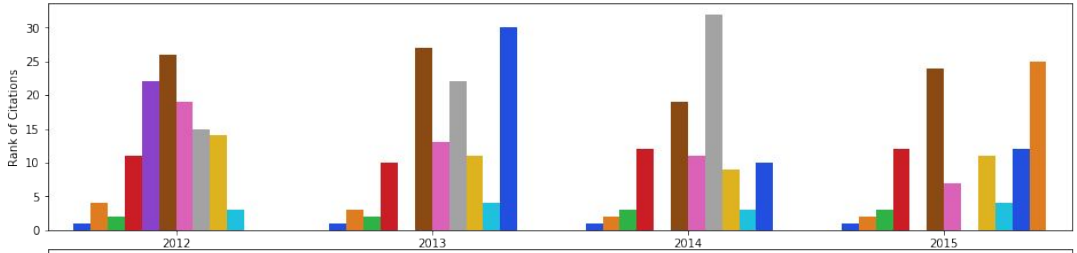
Choice: Option 01

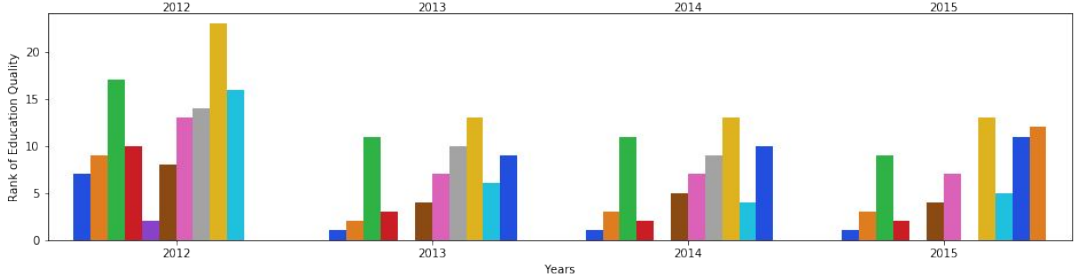
What are some of the factors (variables) that may be related to the world_rank and/or score? Is there any trends indicating shift in importance of factors influencing world_rank over the years or between countries?

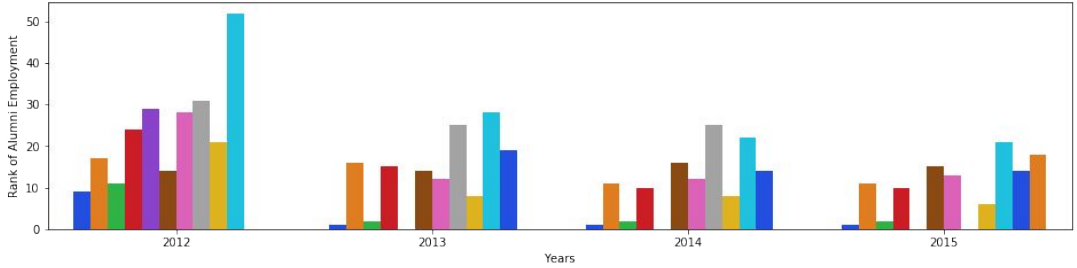
- Large Dataset, find it hard to focus on the analysis
- Decided to narrow it down to Top 10 World Rankings across all years
- It may not be excellent analysis, but it’s able to do some study on the patterns or trends

Analysis based on Universities in Top 10s between 2012 to 2015

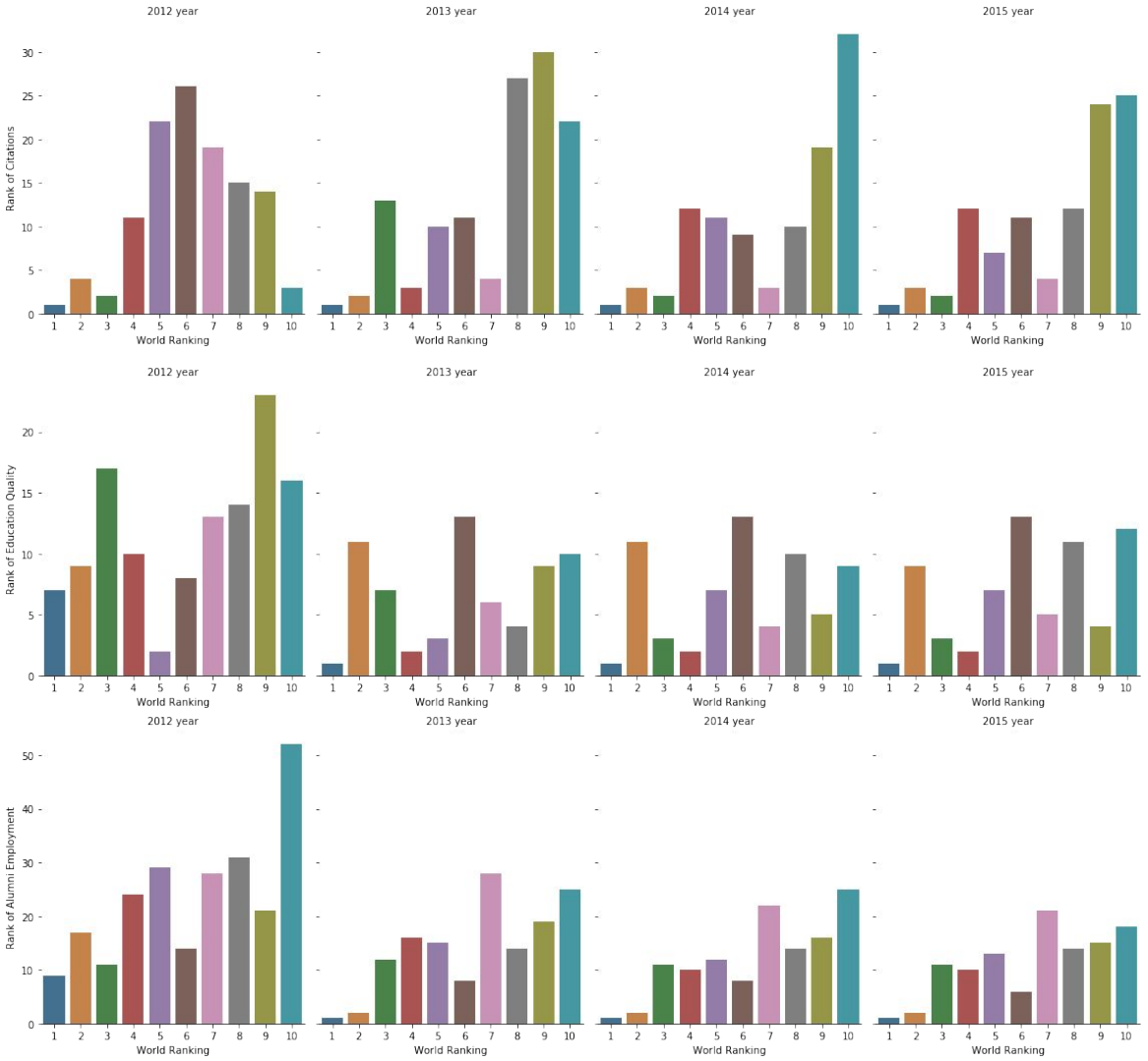




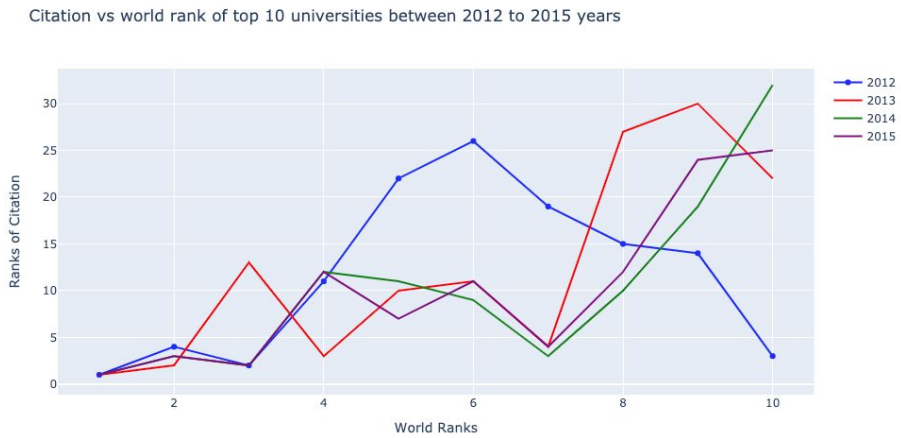


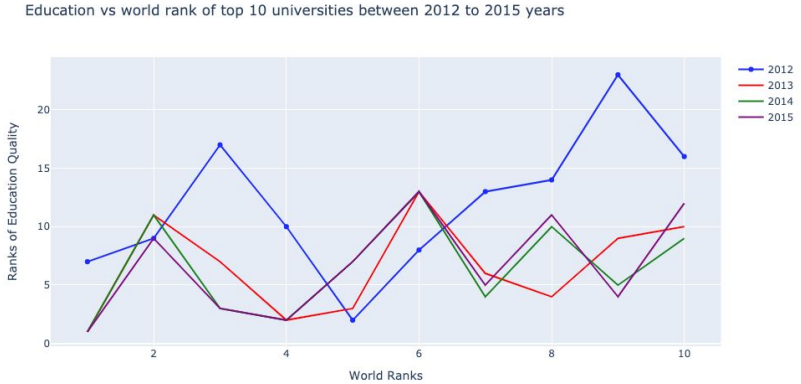


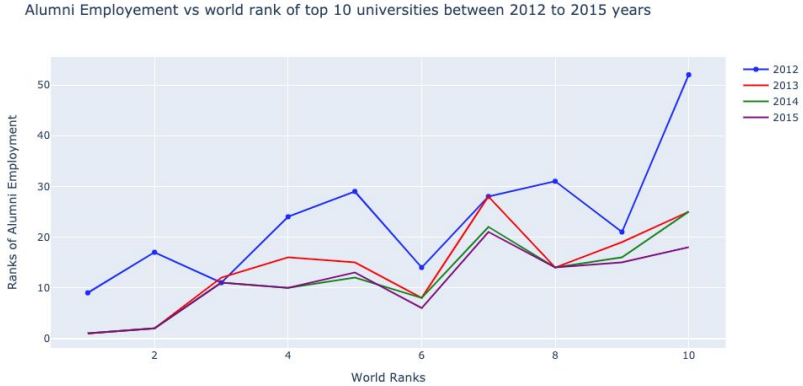
Analysis based on World Ranking between 2012 to 2015



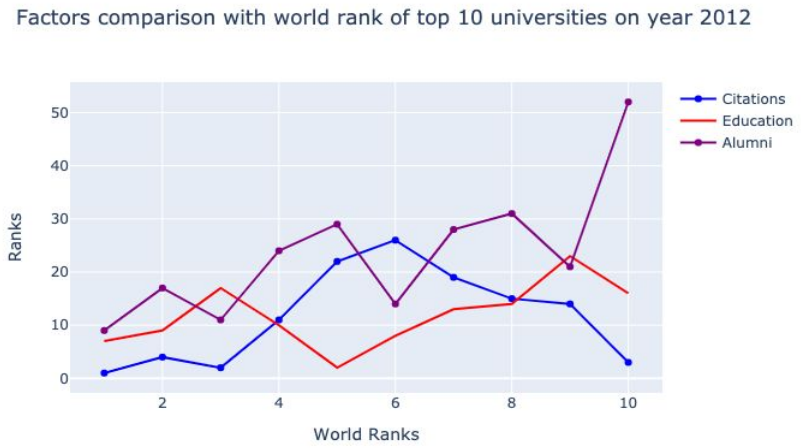
Analysis of Factors vs World Rank of Top 10 Universities between 2012 to 2015

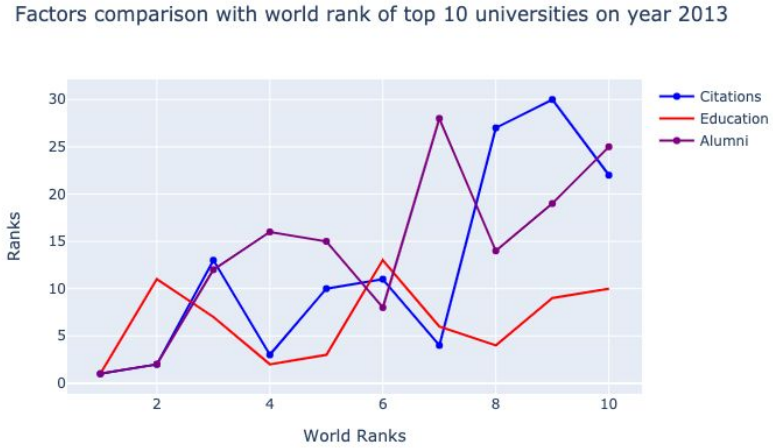


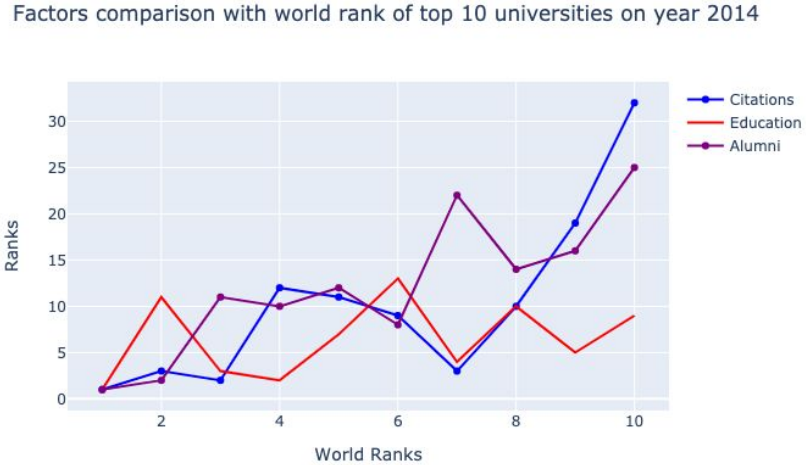


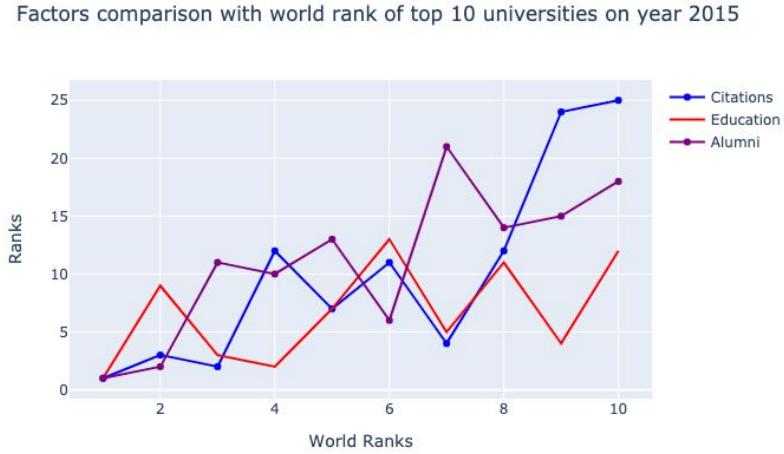


Factors comparison with World Rank of Top 10 Universities based on yearly









Overview

- Simple yet complex dataset
- Numeric data, however these integers it’s meant for rankings hence lesser mathematical methods can be rely on to assess the data.

- End of Presentation -

Thank you for your time!