

# **IBM DATA SCIENCE SPECIALIZATION COURSE**

# **ANALYZING EDUCATION ABROAD**

A Project Report

Submitted by **Glenys Abigail**

On **20<sup>th</sup> August, 2019**

## Table of Contents

<b>1.</b>	<b>Introduction</b> Background Problem Interest	4
<b>2.</b>	<b>Data Acquisition and Cleaning</b> Data Sources Data Cleaning and Formatting	5
<b>3.</b>	<b>Methodology</b> Explanatory Data Analysis Further Analysis using FourSquare Data	7
<b>4.</b>	<b>Results</b>	12
<b>5.</b>	<b>Discussions</b>	13
<b>6.</b>	<b>Conclusion</b>	13
<b>7.</b>	<b>Appendix</b>	14

# **Introduction**

## ***Background***

India is known for its cultured society along with its goal-oriented and ambitious citizens. One of the biggest trends in the education sector right now is to pursue a Master's degree in a foreign country. The very first batch of brave students who took up the challenge of studying abroad, stand as an inspiration to other. Their success soon becomes a legendary story that is often repeated which in turn aspires others to replicate it as their own tale.

There are several reasons why students prefer studying abroad. A few of those reasons are: exposure, creating new networks, living independently, experiencing new cultures and so on. According to one of the leading newspapers in India, between Oct 2012 and Feb 2013, 5,600 student visas were issued in India for students. There has also been a 50% growth in the total number of visas offered since the previous year. These facts state clearly that the trend is moving in an uphill direction.

Foreign institutions offer numerous courses under one field of study with various kinds of specializations and practical experience than the ones in India. They have unique teaching methods and a syllabus that is structured based on the student's needs and preferences. Concepts such as a double degree is new and exciting to Indian students and the absence of the constant pressure to be on the top stimulates their learning experience.

## ***Problem***

Though the students benefit from pursuing their education abroad, India stands to lose its large number of student population, that is not replaced by the incoming foreign students. It causes a problem since most students who are successful in another country tend to settle down and start a new life there, which does not benefit India.

## ***Interest***

Knowing why students are unable to find the right course and university in India could help the Government to make the necessary changes and update the education system. This could in turn help build the country with the assistance of well-learned students.

## Data Acquisition and Cleaning

### *Data Sources*

The data for this project was sourced from various sites hosted by the Government of India (<https://www.mea.gov.in/lok-sabha.htm?dtl/29313/QUESTION+NO2755+INDIAN+STUDENTS+ABROAD>) as well as other Non-Governmental Organizations (<https://www.iie.org/Research-and-Insights/Project-Atlas/Explore-Data/India/Outbound-Mobility---Past-Years>). The data retrieved from the first site consists of the number of foreign students studying in India between the years 2014 – 2017. The data sourced from the next website contains the various countries Indian students have migrated to, which is also ranked in ascending order. This will help build a reliable database with which data analysis can be performed.

Foursquare API was used for further data analysis towards the end. The data extracted from this was used to analyse one of the possible reasons for students' preference to study abroad.

### *Data Cleaning and Formatting*

The first step was to import all the necessary libraries used for the data analysis. This was followed by converting the acquired data which was in tables into data frames. The tables were organized and inserted with the correct indices. The data frames were put through the basic formatting and cleaning processes to bring it to the expected format. Data visualization is an important part of the analysis so all the values were converted to the necessary datatype in order to plot them.

Below are the final data frames:

:

	Year	Number of Students
1	2014	65602
2	2015	74061
3	2016	72542
4	2017	69322

*Fig. 1*

	Rank	Number of Students
Destination		
United States	1	97613
Australia	2	25562
United Kingdom	3	19604
Canada	4	13626
New Zealand	5	10255

Fig. 2

Fig. 1 shows the number of foreign students studying in India between the years 2014 – 2017 Fig. 2 shows the countries where Indian students chose to pursue their education in, along with the number of students.

The FourSquare data collected was the details regarding the universities in New York, USA (one of the four states with the most number of universities) and Rajasthan, India (state with most universities in India). These data were extracted and converted into a data frame. This was followed by formatting and cleaning the data.

The final data frames are as below:

	name	categories	address	city	country	distance	formattedAddress	lat	lng	state	id
0	Pace University	University	1 Pace Plz	New York	United States	224	[1 Pace Plz (btwn Nassau & Gold St), New York,...	40.710720	-74.005743	NY	49e7c7c6f964a5201b651fe3
1	New York University	University	25 W 4th St	New York	United States	2162	[25 W 4th St, New York, NY 10012, United States]	40.729518	-73.993122	NY	5058de7be4b007870dd7ad7a
2	Columbia University	University	116th St & Broadway	New York	United States	11137	[116th St & Broadway, New York, NY 10027, Unit...	40.807282	-73.962849	NY	4227a500f964a520c5111fe3
3	Pace University - William St	University	163 William St	New York	United States	298	[163 William St (at Ann St), New York, NY 1003...	40.710052	-74.006267	NY	4cd03c7ca03a9eb0b458b103
4	Pace University Gym	College Gym	1 Pace Plz	New York	United States	143	[1 Pace Plz (btwn Nassau & Gold St), New York,...	40.711573	-74.005255	NY	4cc0aa4522ce4688e6c63847

Fig. 3

	name	categories	address	city	country	distance	formattedAddress	lat	lng	state	id
0	university Staff Colony, Jodhpur	None	NaN	NaN	India	94576	[India]	26.267945	73.037710	NaN	50268a84e4b0a90de240b15f
1	SPUP - Sardar Patel University of Police, Secu...	University	NaN	NaN	India	83891	[India]	26.397253	73.063681	NaN	55a0e3d7498e818b3cb719eb
2	university Staff Colony, Jodhpur	None	NaN	NaN	India	95675	[India]	26.261844	73.029049	NaN	50268b0ce4b08ea913aca0ed
3	National Law University	Law School	NLU	Jodhpur	India	86116	[NLU (Nau meel road), Jodhpur, Rajasthan, India]	26.361829	73.063791	Rajasthan	4e09fbd52271dfa46bc0244f
4	M D S University	University	NaN	Ajmer	India	97704	[Ajmer, Rajasthan, India]	26.499952	74.686970	Rajasthan	4e38e72014959f8577a3df50

Fig. 4

Fig. 3 shows the details of the universities in New York

Fig. 4 shows the details of the universities in Rajasthan

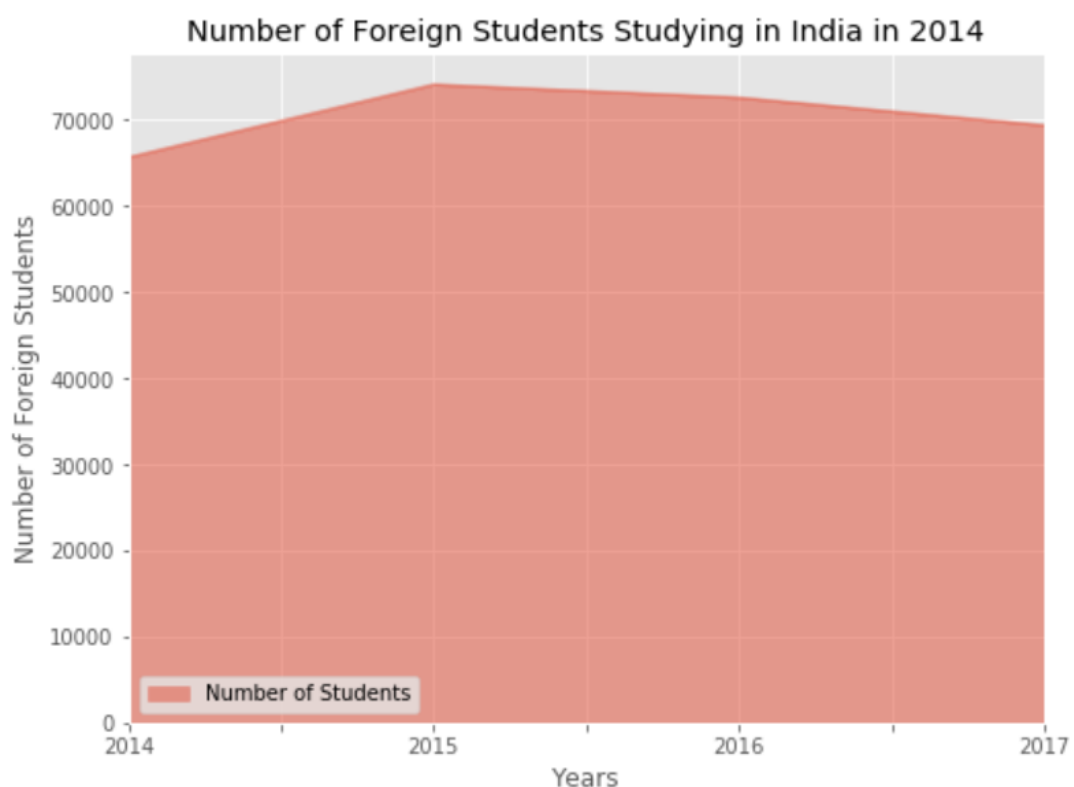
## Methodology

### *Explanatory Data Analysis*

To understand the data better, data visualization techniques were used to display the data. There are several techniques to be used. Some of them are: Bar Charts, Pie Charts, Histograms, Word Clouds, Box Plots and so on.

For the data used, bar charts, area plots and pie charts worked best among the rest.

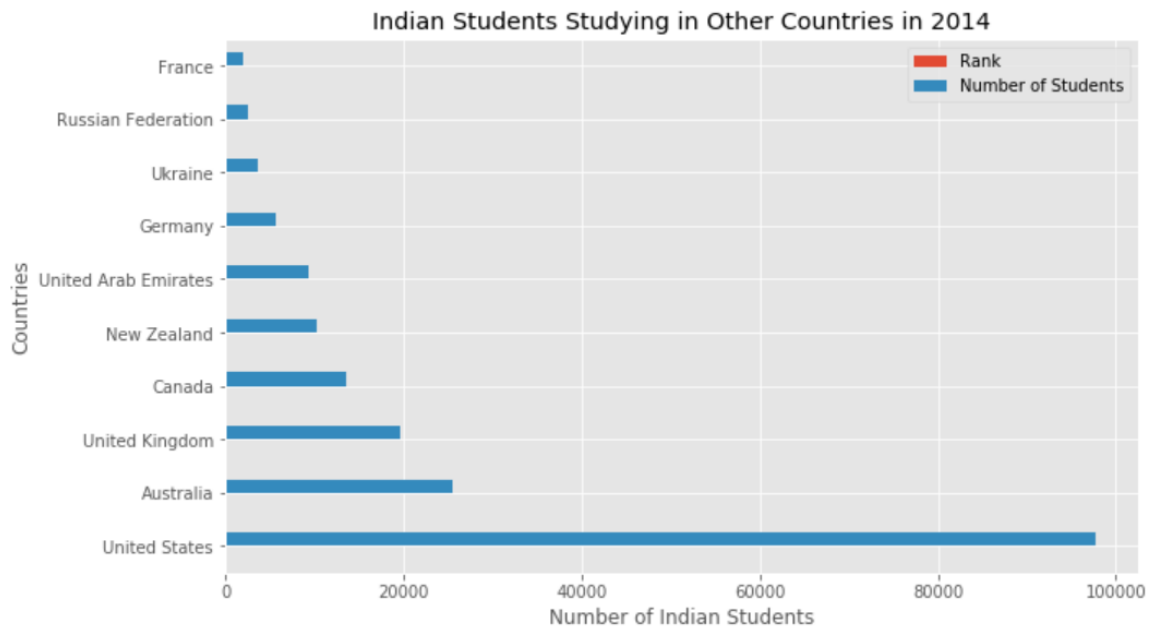
The first visualization of the data was done using an area plot. This figure displayed the years from 2014 to 2017 on the X-axis and the number of foreign students studying in India on the Y-axis.



*Fig. 5*

The figure above shows an increase with a slight peak in the numbers between 2014 and 2015. This is followed by a descent as it reaches 2017. From this it was inferred that the number of foreign students opting to study in India is on a constant decline.

The next visualization is of the various countries Indian students migrated to for their education in 2014. The data consisted of 10 countries, along with the number of students residing in each country. The countries were ranked from 1 to 10, 1 being the country with the most number of Indian students.

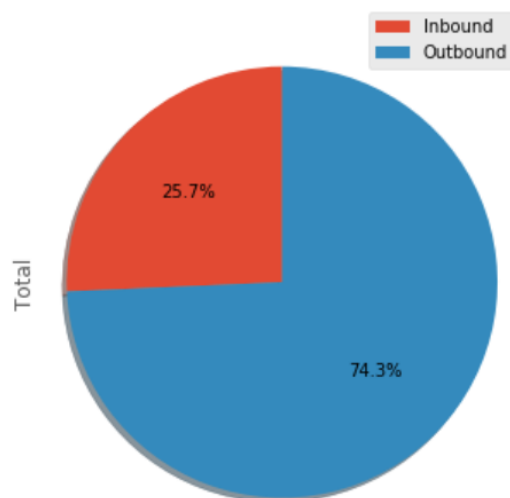


*Fig. 6*

The figure above displays the number of students on the X-axis and the countries on the Y-axis. It is clear that the United States of America has the most number of Indian students, followed by Australia and the United Kingdom. The conclusion drawn from this was that the United States of America had clearly won the race and had no other country to qualify as its competitor.

A pie chart was used to compare the number of Indian students studying abroad with the number of foreign students studying in India in the year 2014.

Comparing Number of Inbound and Outbound Students (India - 2014)



*Fig. 7*



The above figure shows that there is a great difference between the number of Indian students studying abroad and the number of foreign students studying in India in 2014. The difference between both the results is not equal (i.e. 50%), it is extremely unbalanced as 25% and 74%. From this it was concluded that the major difference proves that there exists a problem which is the cause for students to prefer studying in other countries than in India.

As an added measure, the data from the first data frame (which consisted of the number of foreign students studying in India from 2014 to 2017) was used to construct a simple linear regression model. Though the data available proved to be insufficient, the model was built and fitted using the available data to predict the number of students studying in India in a given year. The number of students acted as the dependent or response variable (Y), and the years acted as the independent variable (X).

Linear Function:

$$Y = a + bX$$

a is the intercept

b is the slope

The intercept and co-efficient derived from the model created is displayed bellowed:

```

1rm.intercept_
2]: -1872761.8000000007

1rm.coef_
3]: array([964.1])

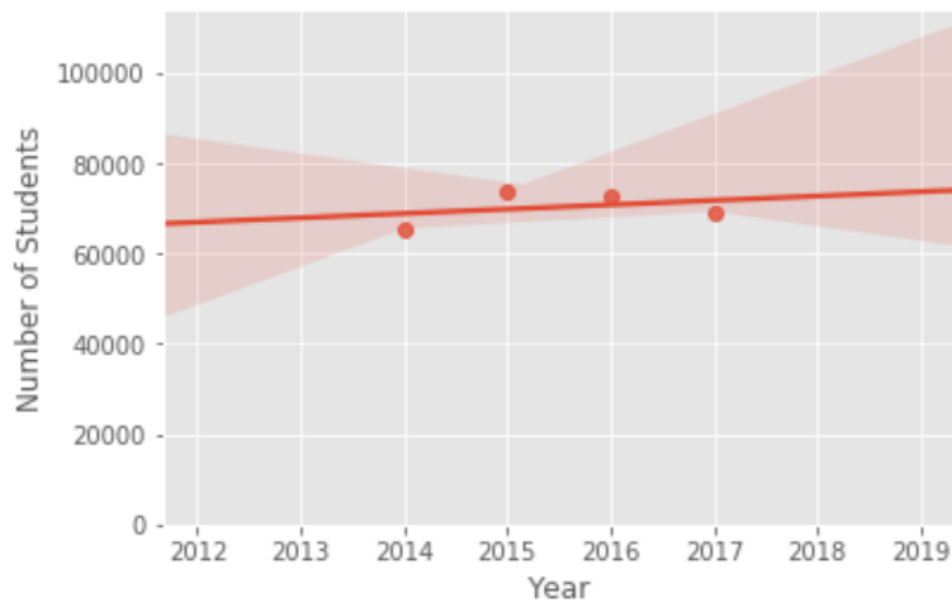
```

*Fig. 8*

Therefore, the estimated linear model derived was

$$\text{Number of Students} = -1872761.80 + 964.1 * \text{Year}$$

The next step was to visualize the model in order to evaluate it. In simple linear regression, one of the best ways is to visualize the fit of the model is by using regression plots. This plot will show a combination of scattered data points (a scatter plot), as well as the fitted linear regression line going through the data. This will give a reasonable estimate of the relationship between the two variables, the strength of the correlation, as well as the direction (positive or negative correlation).



*Fig. 9*

The slope does not appear to increase nor decrease, this makes it difficult to determine if the number of foreign students studying in India increases or decreases as the years increase. There were not too many plots scattered around the regression line. This was due to the insufficient amount of data used to build the model. However, it was inferred that the scattered plots that did exist, existed near the line which states that a linear model could work for this.

### ***Further Analysis Using FourSquare Data***

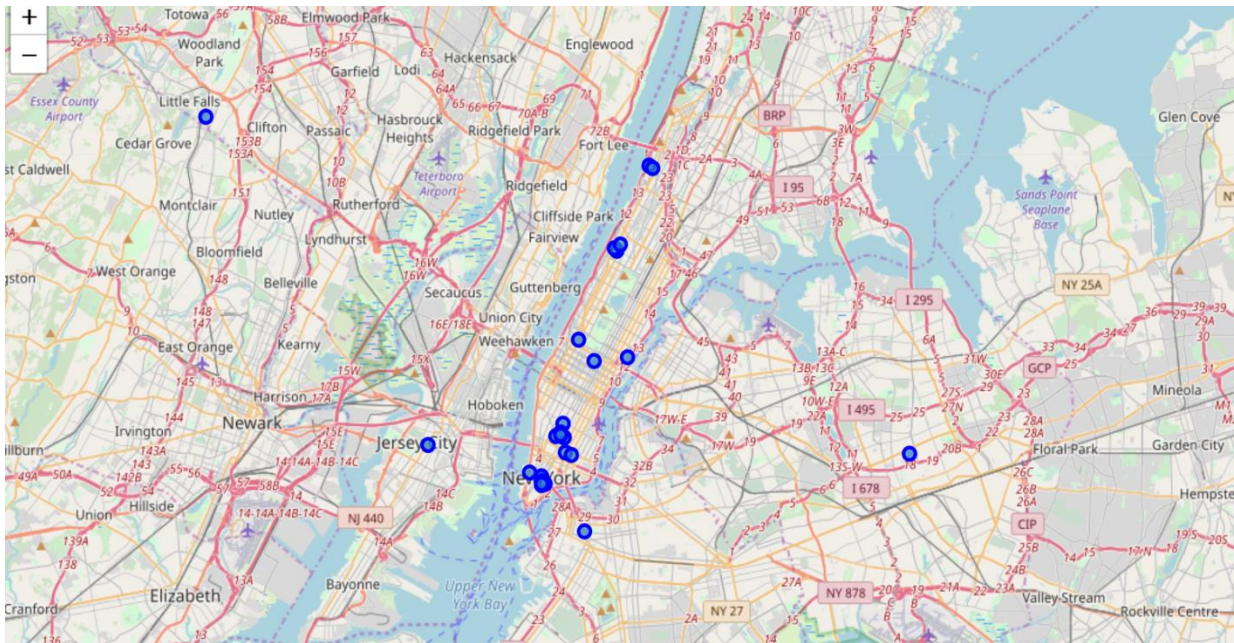
Once the data was analysed, used to build a linear regression model and then analysed again, the Foursquare data was used for some side-analysis. The data retrieved using Foursquare was the number of universities in New York, USA and Rajasthan, India. These two states were chosen for the following reasons:

New York is one of the four states in the United States of America that has the most number of universities. Since the USA stood to be the number one country where Indian students migrated to, this state was chosen. The data collected from this state using FourSquare was the details regarding the universities in New York.

Rajasthan was chosen since it is the state which has the most number of universities in India. The data collected using FourSquare was the details regarding the universities in the state. This data was compared with the data regarding the universities in New York.

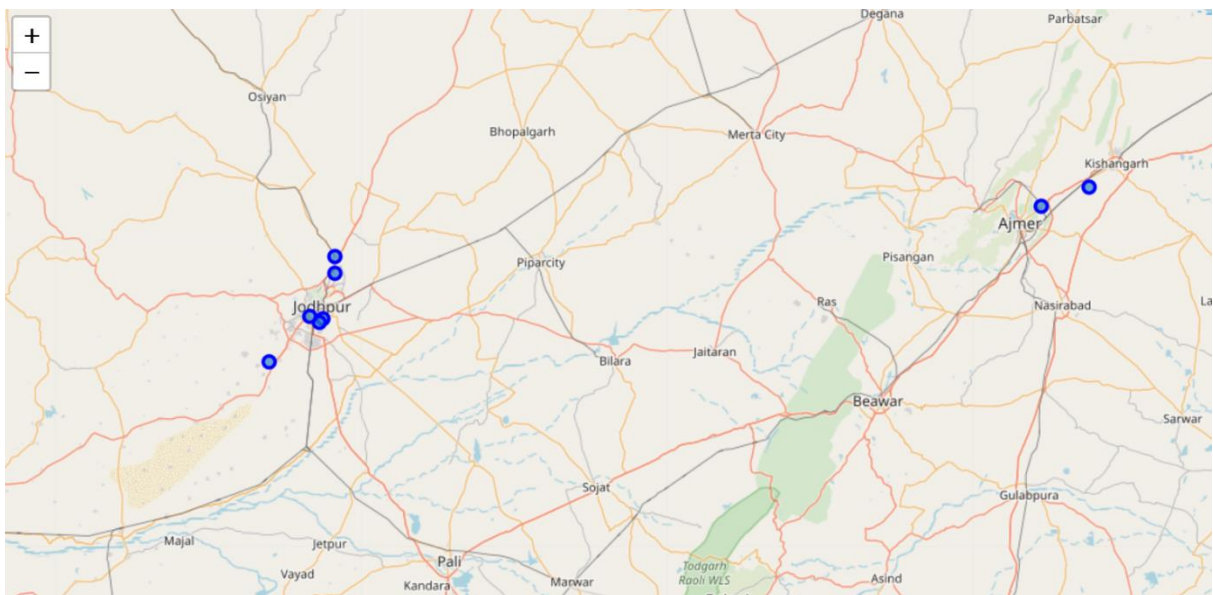
Once the data was ready to be analysed, it was plotted on a world map using Folium, with the blue dots representing the universities.

The first map is of New York and its universities,



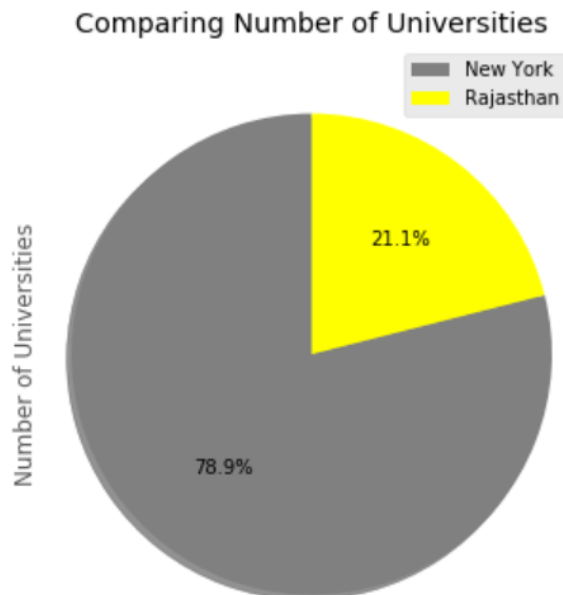
*Fig. 10*

The next map show Rajasthan and its universities,



*Fig. 11*

The final analysis was done based on the number of universities in New York and Rajasthan, that was retrieved using FourSquare. The number of universities in New York returned as 30 and for Rajasthan as 8. Using this, a pie chart was created to further analyse this newfound data.



*Fig. 12*

The above figure shows the drastic difference between the two results. The number of universities in New York was more than half of the number of universities in Rajasthan. The visualization of this data helped conclude that the number of universities available can be one of the reasons for students preferring to study in other countries.

## Results

From the analysis of the above data we can conclude the following results:

- The number of foreign students studying in India is far less compared to the number of Indian students studying in other countries. This could be due to several reasons as discussed in the Introduction section, but on further analysis using FourSquare it was discovered that the number of universities available in India could be one of the reasons students choose to go abroad. The universities abroad are more and offer a lot of different courses, compared to the ones offered in India.
- It was also inferred that the United States of America was the country that had the most number of Indian students in 2014. This led to the comparison of the number of universities in New York and Rajasthan, which supported the statement that the number of universities available could be one of the reasons for more number of outbound students from India.
- Using linear regression to build a model, it was concluded that even though the available data for the model was insufficient, the intercept and slope derived can be

used to calculate the number of foreign students studying in India with the right inputs. There was a slight peak in between the years 2014 and 2015 however, the numbers seemed to decrease slightly after 2015.

## **Discussions**

The points to be noted and discussed from this analysis are as follows:

- The data procured from the various sources regarding the number of inbound and outbound were insufficient for the building of reliable models and future predictions. Hence, for better predictions more data ranging over more than 4 years is required.
- The data procured from FourSquare also seemed to require an update as 30 universities in New York and 8 universities in Rajasthan is very less compared to the universities that exist in those two states now (in 2019).
- The number of universities is one of the many details of the data that could be analysed to discover more findings about why students prefer universities abroad.

## **Conclusion**

The analysis can be concluded by grouping together several points of importance such as the fact that even though the times have changed and new countries have risen with better opportunities, from this analysis of the data that has been sourced, we have inferred that Indians prefer studying abroad, especially in the United States of America. It can also be noted that India must be put under the spotlight to improve its education system in order to make it more attractive to foreigners as well as its own citizens.

## Appendix

### Web References:

- <https://www.mea.gov.in/lok-sabha.htm?dtl/29313/QUESTION+NO2755+INDIAN+STUDENTS+ABROAD>
- <https://www.iie.org/Research-and-Insights/Project-Atlas/Explore-Data/India/Outbound-Mobility---Past-Years>
- <https://www.mapsofindia.com/my-india/education/why-indian-students-want-to-study-in-foreign-universities>
- <https://www.ugc.ac.in/stats.aspx>
- <https://www.worldatlas.com/articles/which-state-has-the-most-colleges.html>
- <https://timesofindia.indiatimes.com/city/jaipur/state-tops-in-number-of-universities-third-in-colleges-in-country/articleshow/62387076.cms>