PRESENTED BY:

Prashanth Srinivasan

Sriram Shanmugam

Jeshurun Paul S

TEXT MINING

- Text mining is done for almost every column in the dataset presented as every column had valuable information in text. Hence the Abstract column is taken initially and Natural Language Processing has been applied for Topic modelling using Latent Dirichlet Allocation (LDA) method.
- In this process initially the number of topics under which all the abstracts has to be classified is specified as five. So that all the abstracts fall under five topics which are obtained by checking the high frequency words.
- Hence there are five topics which are named from Topic-0 to Topic-4, and the topics are to be generalised from the list of top 10 most occurring words under each topic. The list of words for each topic are as follows.

THE TOP 10 WORDS FOR TOPIC - 0

['francis', 'taylor', 'ha', 'group', 'paper', 'industry', 'technology', 'study', 'literature', 'research']

THE TOP 10 WORDS FOR TOPIC - 1

['2019', 'decision', 'elsevier', 'model', 'algorithm', 'ha', 'different', 'paper', 'based', 'result']

THE TOP 10 WORDS FOR TOPIC - 2

['study', 'francis', 'taylor', 'manufacturing', '2019', 'paper', 'industry', 'research', 'data', 'technology']

THE TOP 10 WORDS FOR TOPIC - 3

['data', 'performance', 'network', 'elsevier', 'time', 'result', 'model', 'based', 'paper', 'proposed']

THE TOP 10 WORDS FOR TOPIC - 4

['ha', 'business', 'study', 'information', 'paper', 'thing', 'iot', 'data', 'internet', ‘based’]

The topic numbers for each abstract is attached along with this file as an excel sheet.

TEXT SUMMARIZATION

Summarization is the task of condensing a piece of text to a shorter version, reducing the size of the initial text while at the same time preserving key informational elements and the meaning of content.

In this section, we have mined the Abstract into sentences and tokens and performed a function of sentence similarity and created a matrix which is a set of numbers showing the similarity between sentences. Further we have ranked the sentence based on its relevancy and sorted it in order.

We then print out the 2 sentence which is ranked the highest. This gives the overall summary of the abstract. We can change the number of sentence which should be selected for summarization as per our convenience.

Following is an example of the text summarization:

Abstract - 1 :

The explosion of reusable web services (e.g., open apis, open data sources, and cloud/iot services), has become a new opportunity for modern service-composition based applications development. however, this enormous growth of web services increases the difficulty of selecting the best suitable web services for a particular application. hence, the design of an effective and efficient web service recommendation, primarily based on user feedback, has become a challenge. in the

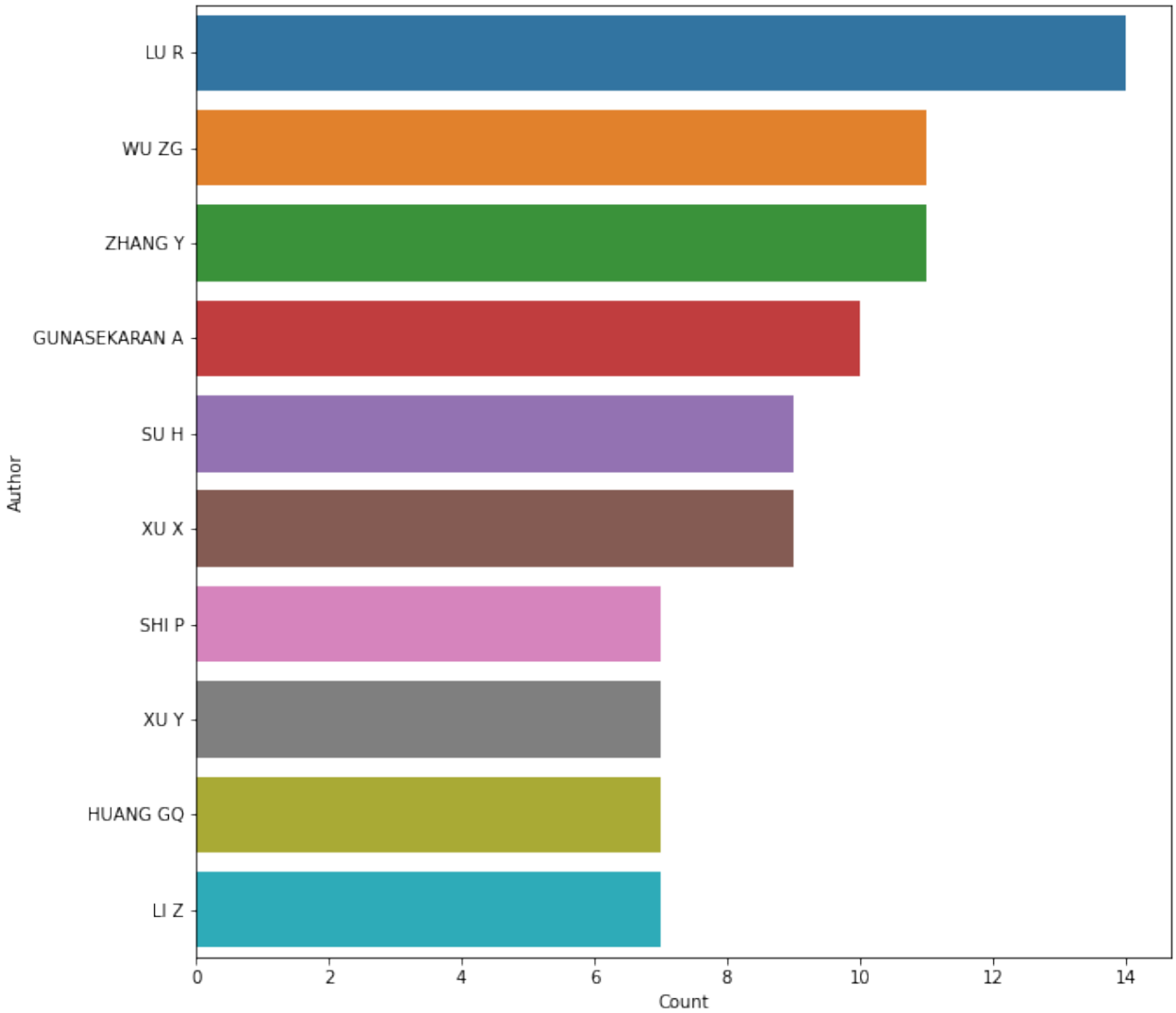
mashup-api recommendation scenario, the most available feedback is the implicit invocation data, i.e., the binary data indicating whether or not a mashup has invoked an api. various efforts are exploiting potential impact factors, such as the invocation context, to augment the implicit invocation data with the aim to improve service recommendation performance. one significant factor affecting the context of web service invocations is geographical location, but it has been given less attention in the implicit-based service recommendation. in this paper, we propose a probabilistic matrix factorization based recommendation approach, which considers geographic location information in the derivation of the preference degree underlying a mashup-api interaction. the geographic information, which is integrated with functional descriptions, complements the mashup-api invocation data input for our matrix factorization model. we demonstrate the effectiveness of our approach by conducting extensive experiments on a real dataset crawled from programmable web. the evaluation results show that augmenting the implicit data with geographical location information increases the precision of api recommendation for mashup services. 2020

Summary Output :

Various efforts are exploiting potential impact factors, such as the invocation context, to augment the implicit invocation data with the aim to improve service recommendation performance. the evaluation results show that augmenting the implicit data with geographical location information increases the precision of api recommendation for mashup services.

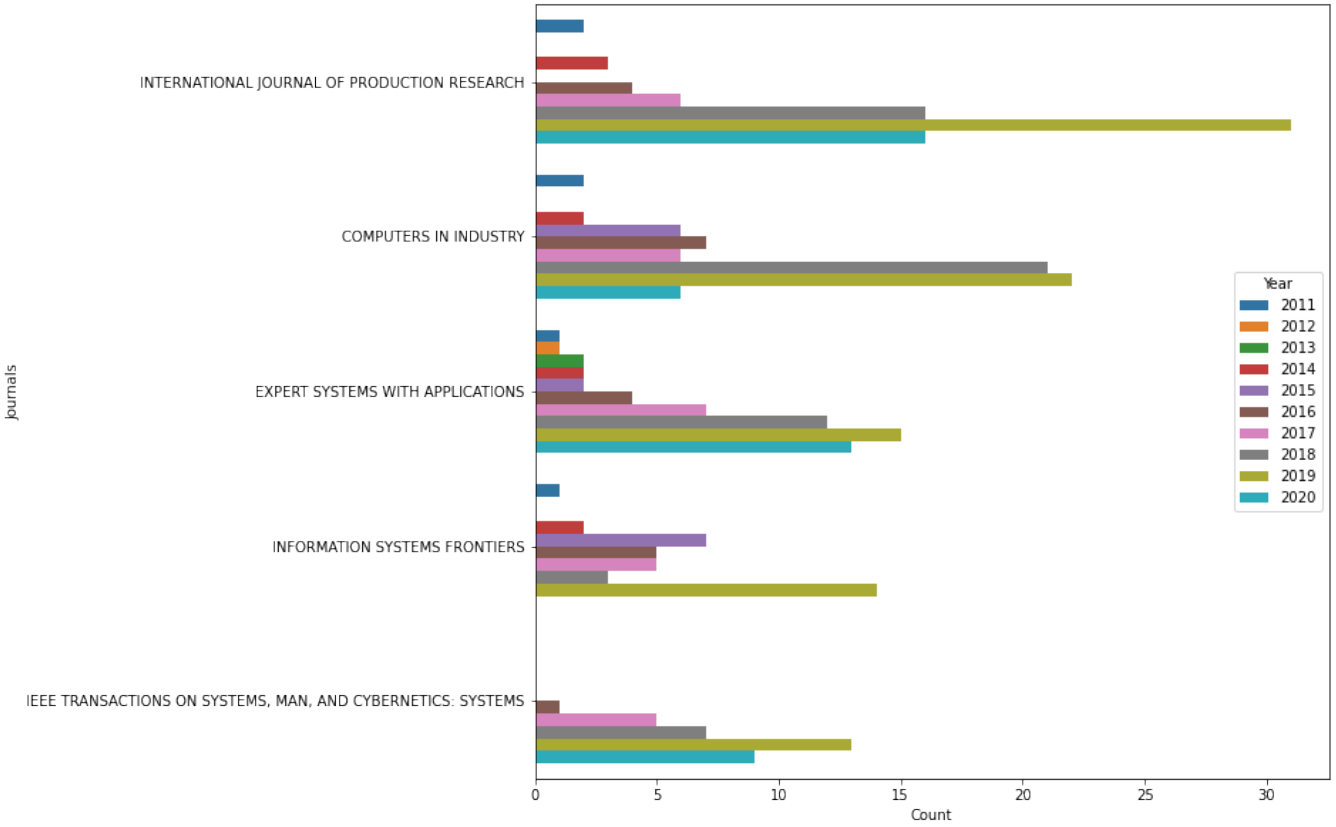
TOP 10 AUTHORS AND THEIR CONTRIBUTIONS

In this section the top 10 authors who have contributed most in this collection of articles are displayed in the form of a combined Bar graph where the X-axis denotes the number of papers published and the Y-axis denotes the names of the top 10 authors based on their contributions.



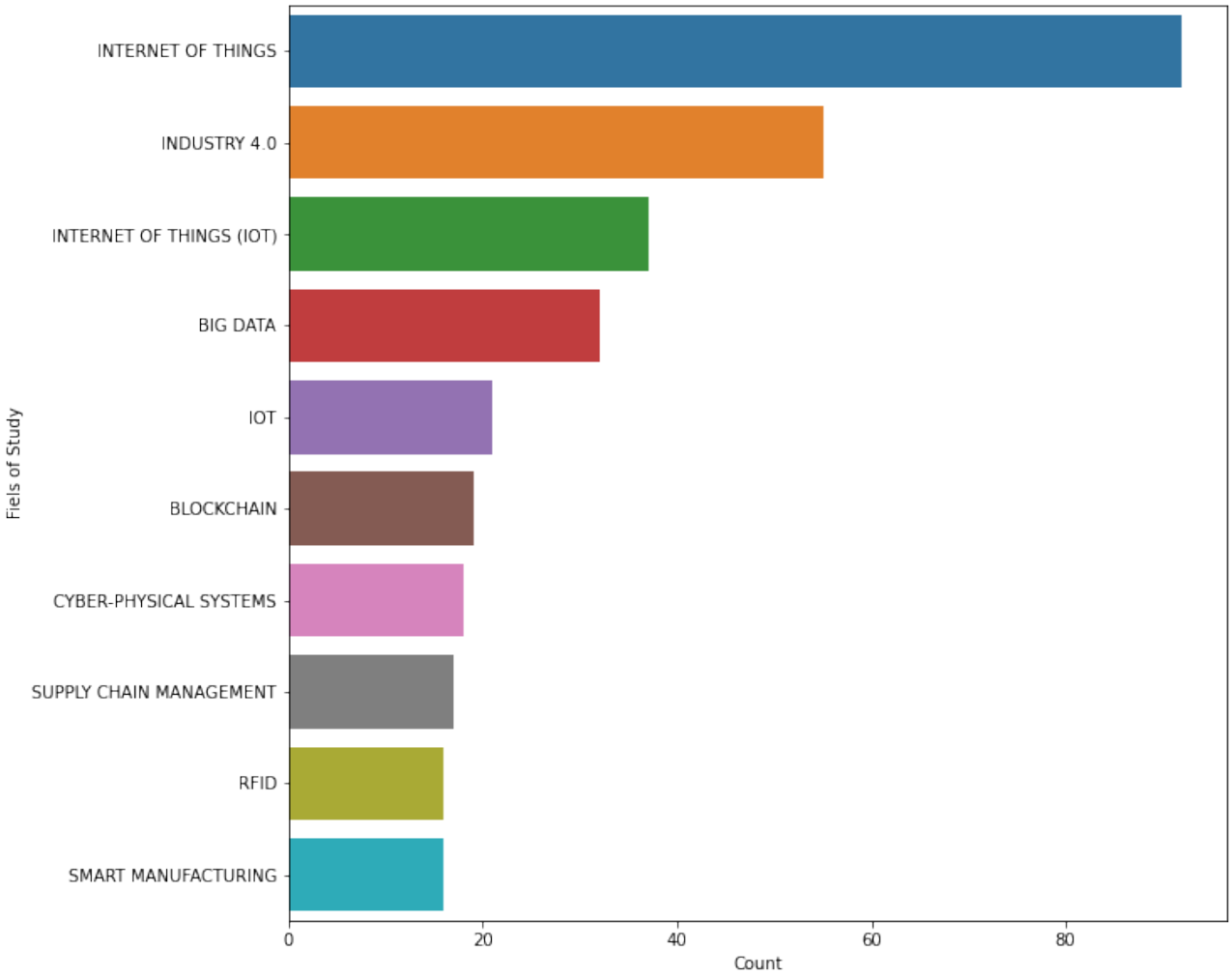
THE TOP FIVE PUBLICATIONS

In this section the top five journals or publications in which most of the papers from this collection of papers are published are displayed in the form of a combined bar graph for all the years from 2011 to 2020 where the X-axis denotes the number of papers published and the Y-axis denotes the names of the top 5 publications in which the papers are published.



THE TOP 10 FIELDS OF STUDY

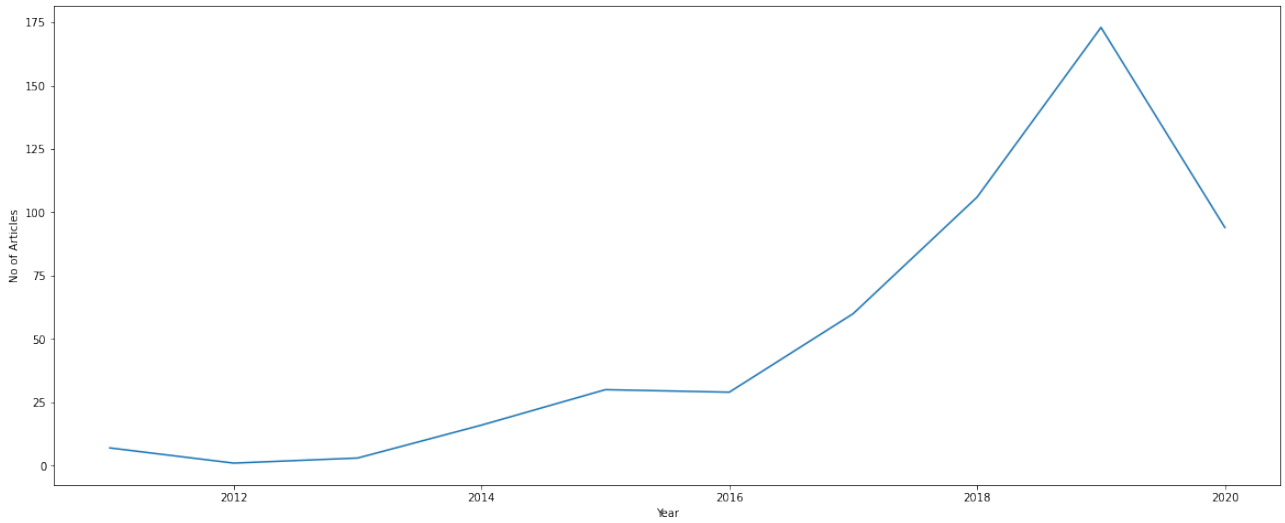
In this section the top ten fields of study where most of the papers have concentrated from the collection of papers provided are listed and displayed in the form of a Bar graph where the X-axis denotes the number of papers published and the Y-axis denotes the names of the top 10 fields of study in which the papers are published.



TREND ANALYSIS

NUMBER OF PAPERS PUBLISHED OVER THE YEARS

The number of papers published overall from all the countries over the years 2011 to 2020 has been displayed in the form of a line graph to analyse the trend.



It has been observed that more number of papers are published as the years progress and the highest number of papers are published in the year 2019.

In the upcoming sections we will see the trend analysis of both the publications and the country wise contribution for the research papers over the years in a detailed manner.

TREND ANALYSIS FOR PUBLICATIONS OVER THE YEARS

The top five Publications in which most papers are published are mined from the SO column of the dataset and their trend is observed over the years, The top five Publications are International Journal of Production Research, Computers in Industry, Expert Systems with Applications, Information Systems Frontier, and IEEE transaction on systems, man and cybernetics: system.

PUBLICATION	NO. OF .PAPERS
INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH	78
COMPUTERS IN INDUSTRY	72
EXPERT SYSTEMS WITH APPLICATIONS	59
INFORMATION SYSTEMS FRONTIERS	37
IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS	35

- These five Publications are the ones in which most papers are published. And their trend over the years from 2011 to 2020 are seen as follows

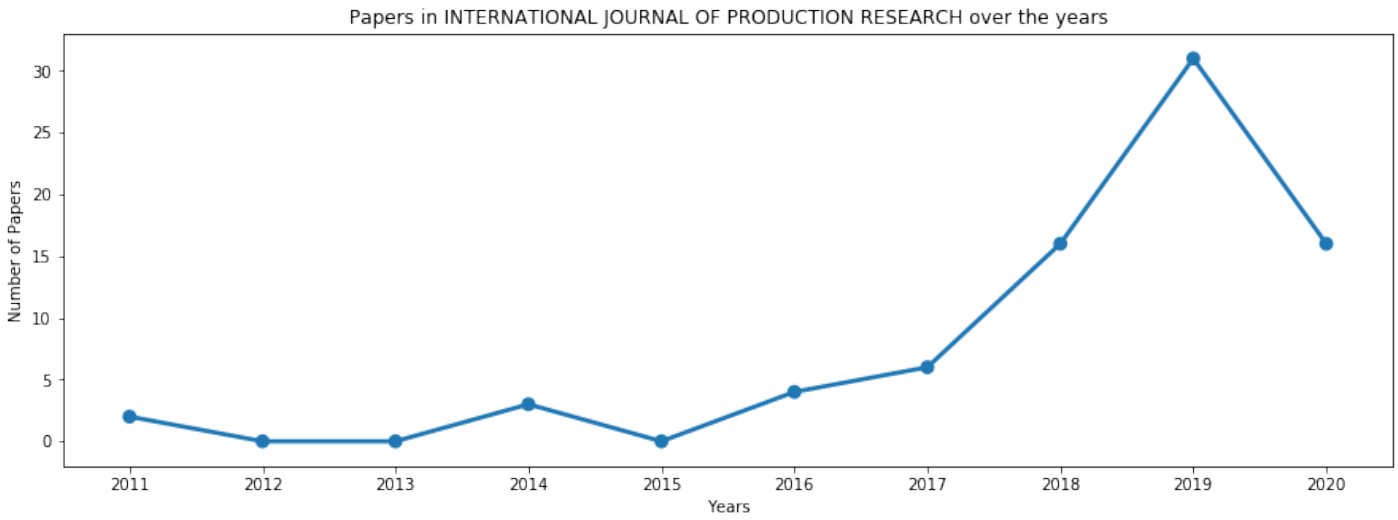
- It is observed that every Publication initially in the years 2011 to 2015 show very less publications of research papers and they gradually increase after that giving rise to an upward trend

- The trend over the years 2011 to 2020 are shown for each Publication in the following Pages.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH:

	Number of Papers	Year
0	16.0	2020
1	31.0	2019
2	16.0	2018
3	6.0	2017
4	4.0	2016
5	0.0	2015
6	3.0	2014
7	0.0	2013
8	0.0	2012
9	2.0	2011

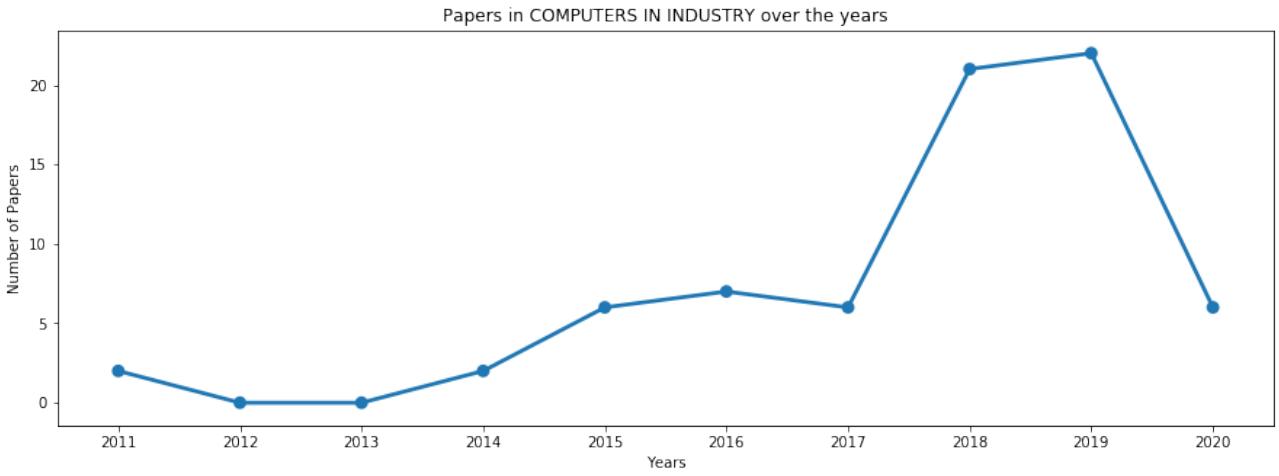
The trend over various years are shown in the line graph below:



COMPUTERS IN INDUSTRY

	Number of Papers	Year
0	6	2020
1	22	2019
2	21	2018
3	6	2017
4	7	2016
5	6	2015
6	2	2014
7	0	2013
8	0	2012
9	2	2011

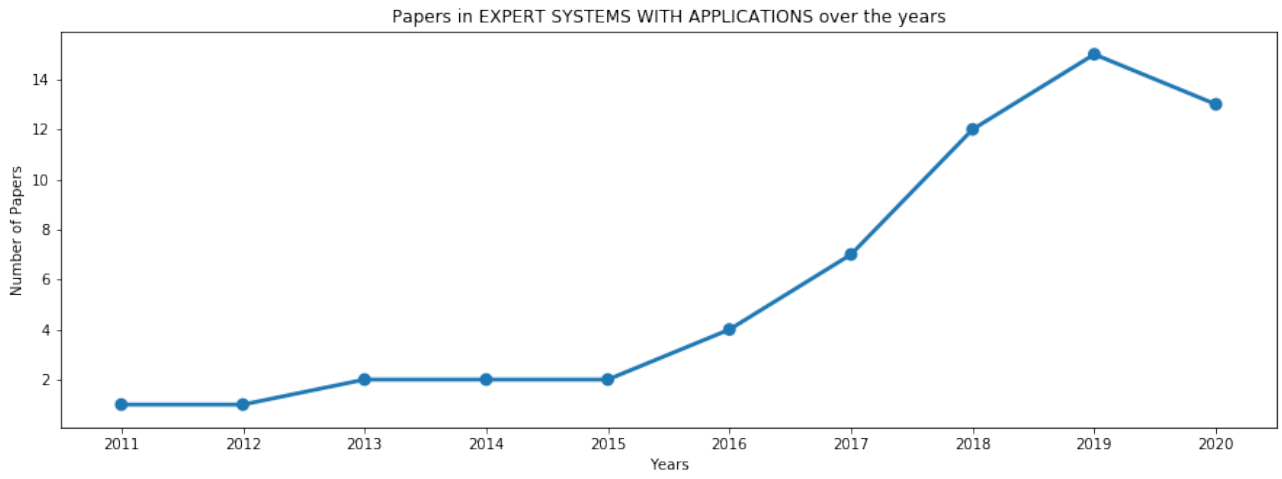
The trend over various years are shown in the line graph below:



EXPERT SYSTEMS WITH APPLICATIONS

	Number of Papers	Year
0	13	2020
1	15	2019
2	12	2018
3	7	2017
4	4	2016
5	2	2015
6	2	2014
7	2	2013
8	1	2012
9	1	2011

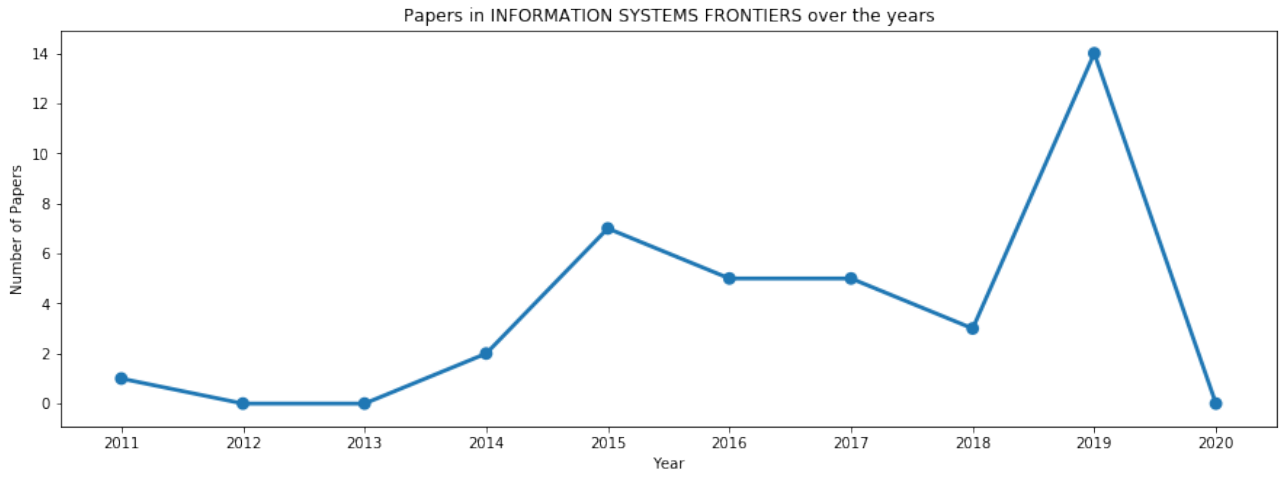
The trend over various years are shown in the line graph below:



INFORMATION SYSTEMS FRONTIERS

	Number of Papers	Year
0	0	2020
1	14	2019
2	3	2018
3	5	2017
4	5	2016
5	7	2015
6	2	2014
7	0	2013
8	0	2012
9	1	2011

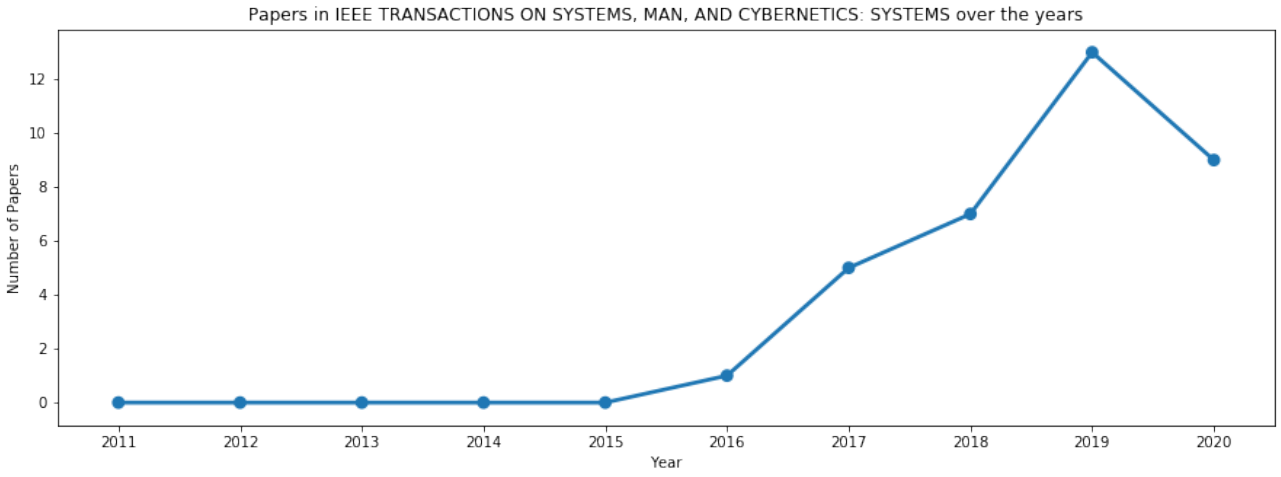
The trend over various years are shown in the line graph below:



IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS

	Number of Papers	Year
0	9	2020
1	13	2019
2	7	2018
3	5	2017
4	1	2016
5	0	2015
6	0	2014
7	0	2013
8	0	2012
9	0	2011

The trend over various years are shown in the line graph below:



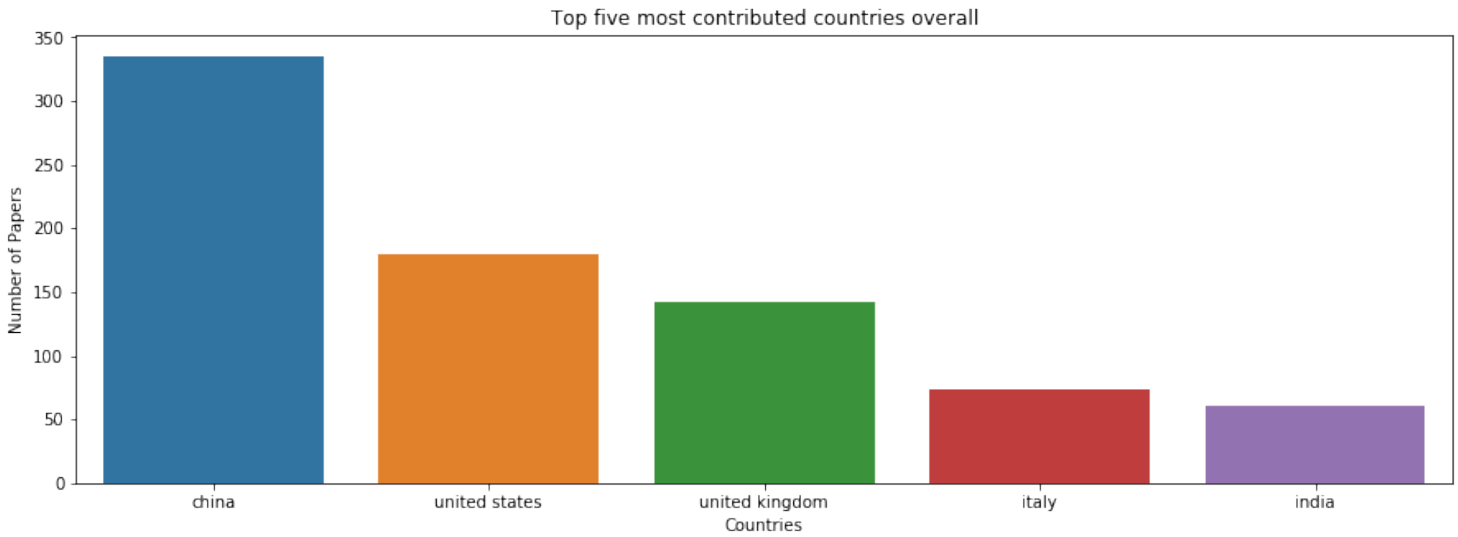
TOP RESEARCH PAPER CONTRIBUTIONS - COUNTRY WISE FOR DIFFERENT YEARS.

OVERALL

The overall contributions of Research Papers from all the countries are mined from the ‘C1’ column which has listed the Author affiliations. The contributions of top five countries are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	335
1	united states	180
2	united kingdom	142
3	italy	74
4	india	51

The top five countries that has contributed the most overall are China, United States, United Kingdom, Italy and India. They are plotted in the form of a Bar Graph for Visualization.

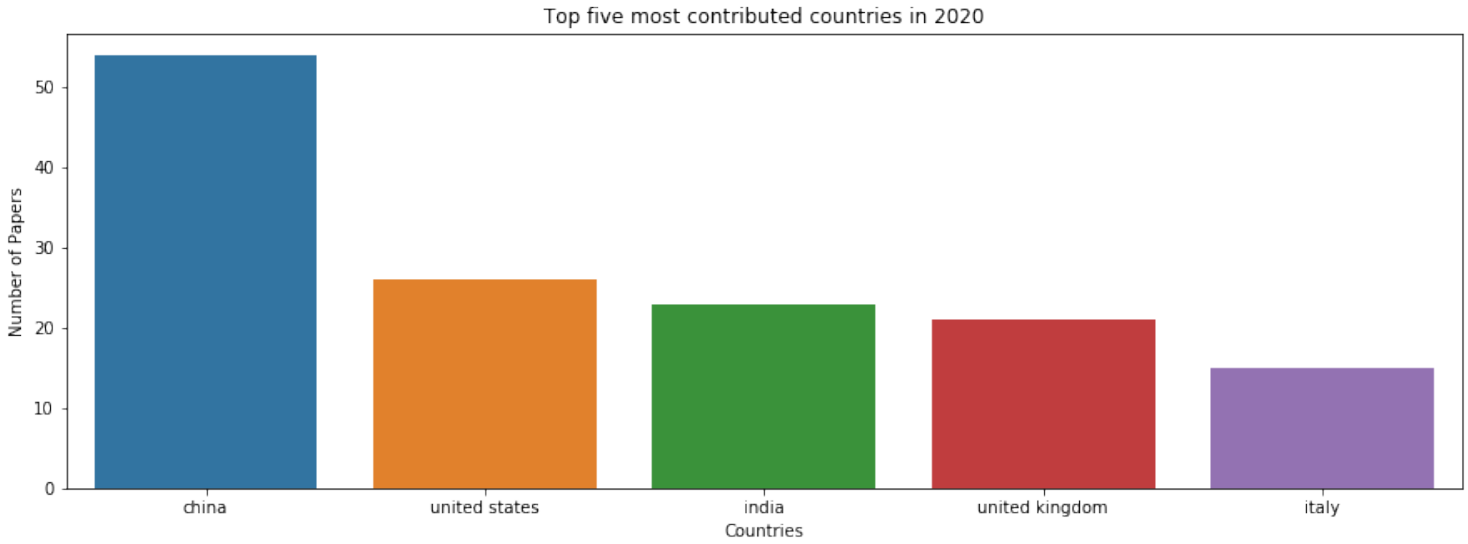


YEAR 2020

The contributions of top five countries in the year 2020 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	54
1	united states	26
2	india	23
3	united kingdom	21
4	italy	15

The top five countries that has contributed the most in the year 2020 are China, United States, India, United Kingdom and Italy. They are plotted in the form of a Bar Graph for Visualization.

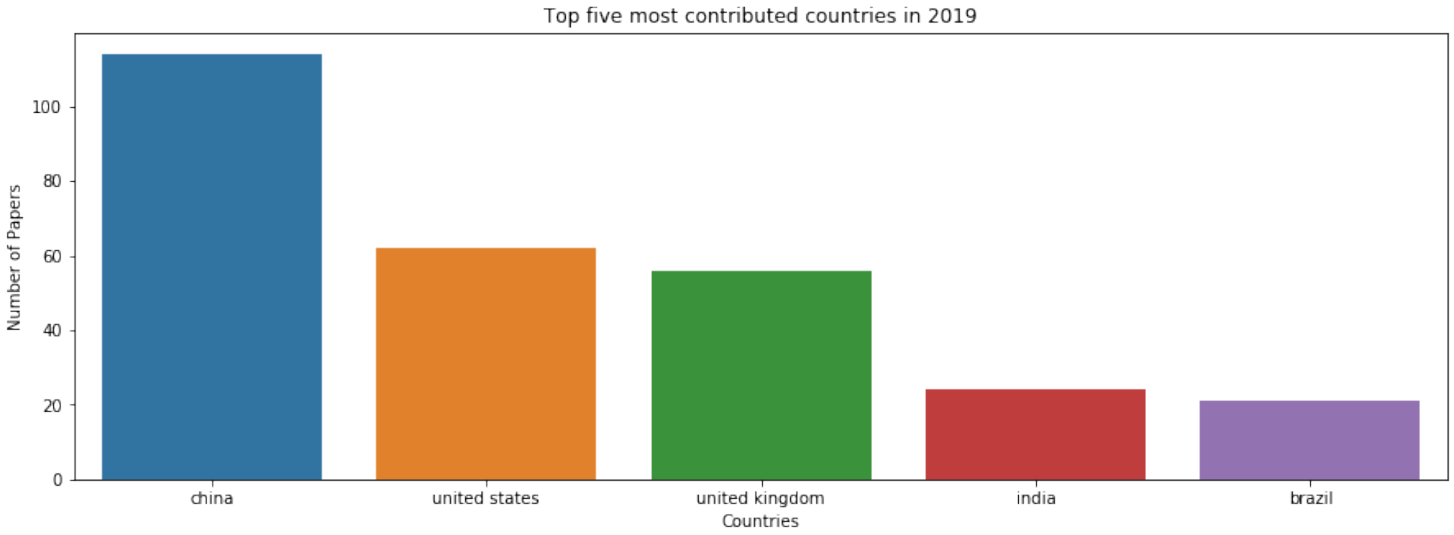


YEAR 2019

The contributions of top five countries in the year 2019 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	114
1	united states	62
2	united kingdom	56
3	india	24
4	brazil	21

The top five countries that has contributed the most in the year 2019 are China, United States, United Kingdom, India and Brazil. They are plotted in the form of a Bar Graph for Visualization.

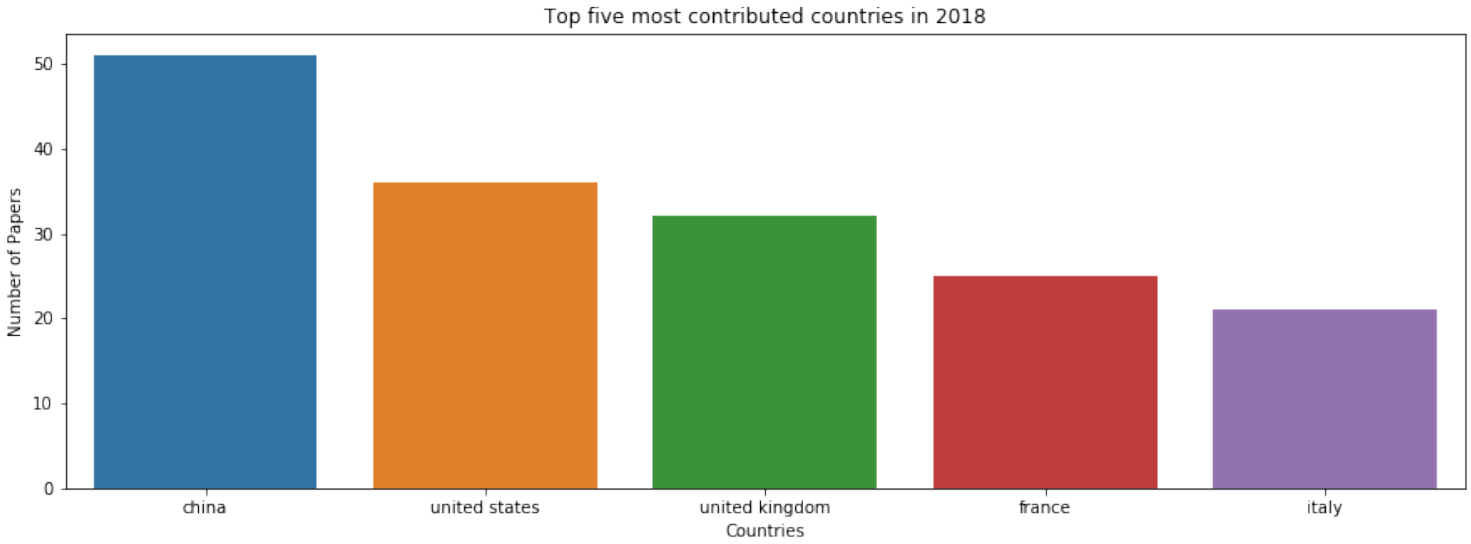


YEAR 2018

The contributions of top five countries in the year 2018 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	51
1	united states	36
2	united kingdom	32
3	france	25
4	italy	21

The top five countries that has contributed the most in the year 2018 are China, United States, United Kingdom, France and Italy. They are plotted in the form of a Bar Graph for Visualization.

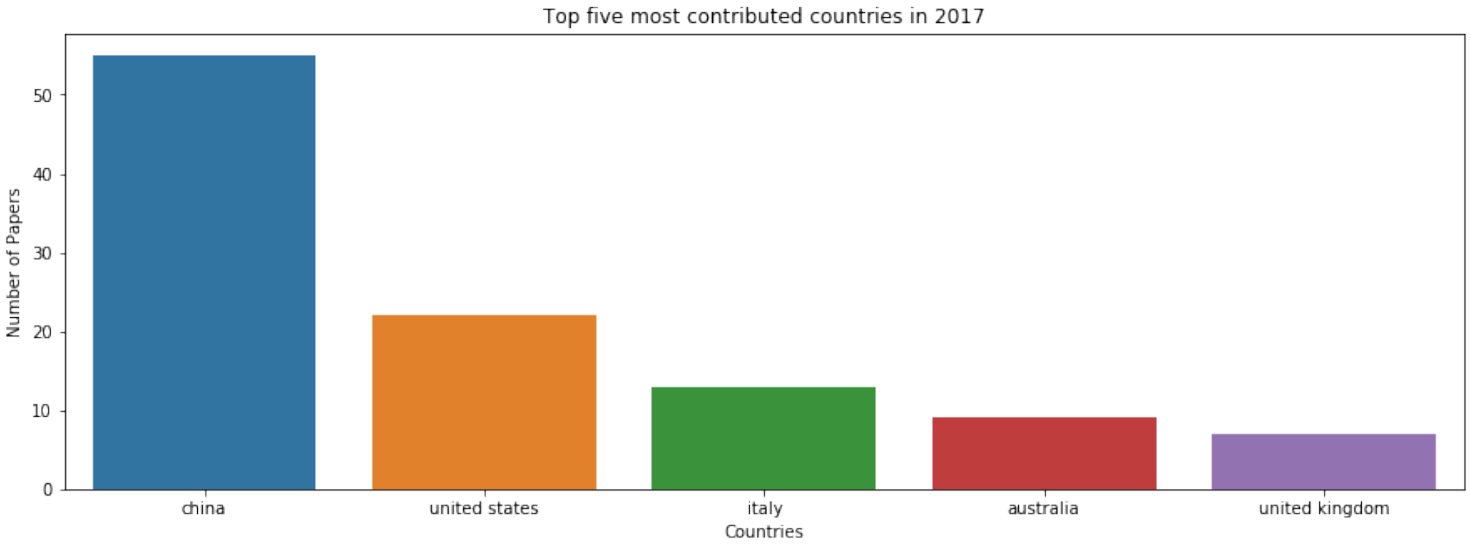


YEAR 2017

The contributions of top five countries in the year 2017 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	55
1	united states	22
2	italy	13
3	australia	9
4	united kingdom	7

The top five countries that has contributed the most in the year 2017 are China, United States, Italy, Australia and United Kingdom. They are plotted in the form of a Bar Graph for Visualization.

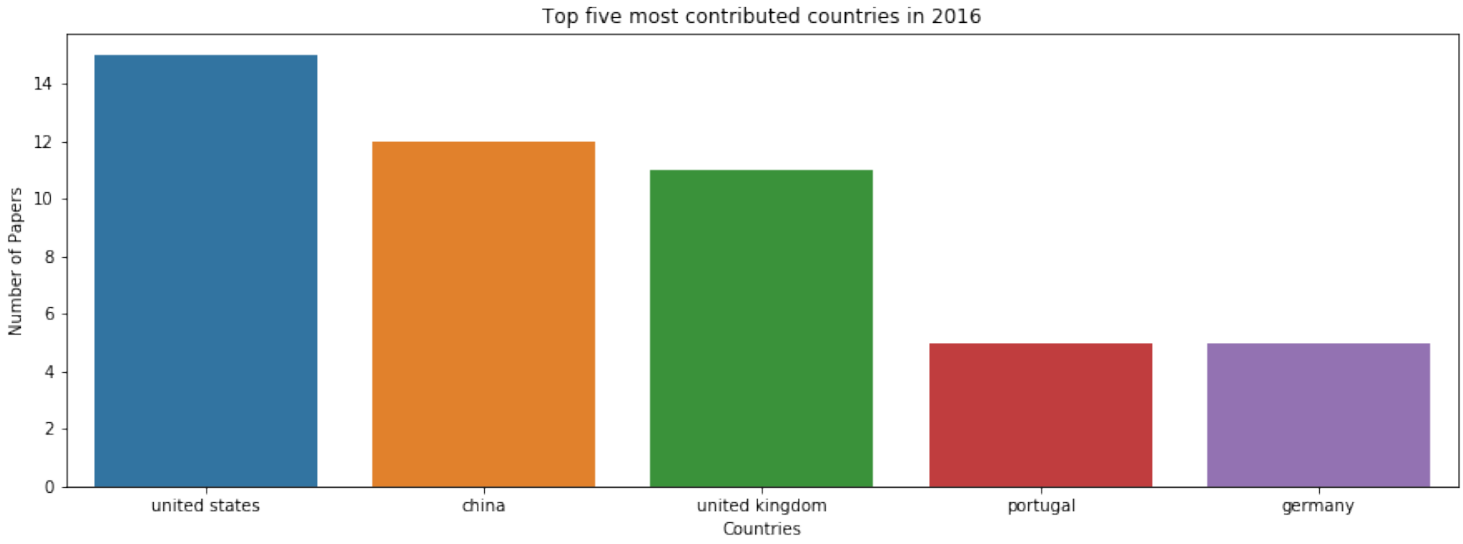


YEAR 2016

The contributions of top five countries in the year 2016 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	united states	15
1	china	12
2	united kingdom	11
3	portugal	5
4	germany	5

The top five countries that has contributed the most in the year 2016 are United States, China, United Kingdom, Portugal and Germany. They are plotted in the form of a Bar Graph for Visualization.

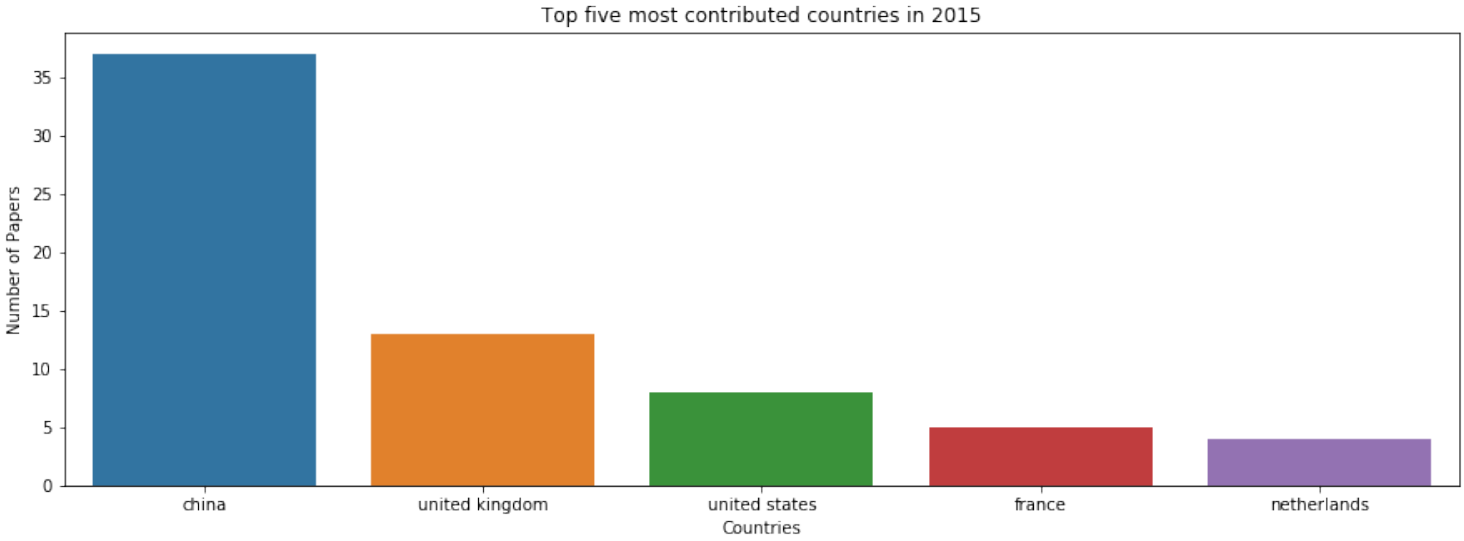


YEAR 2015

The contributions of top five countries in the year 2015 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	china	37
1	united kingdom	13
2	united states	8
3	france	5
4	netherlands	4

The top five countries that has contributed the most in the year 2015 are China, United Kingdom, United States, France and Netherlands. They are plotted in the form of a Bar Graph for Visualization.

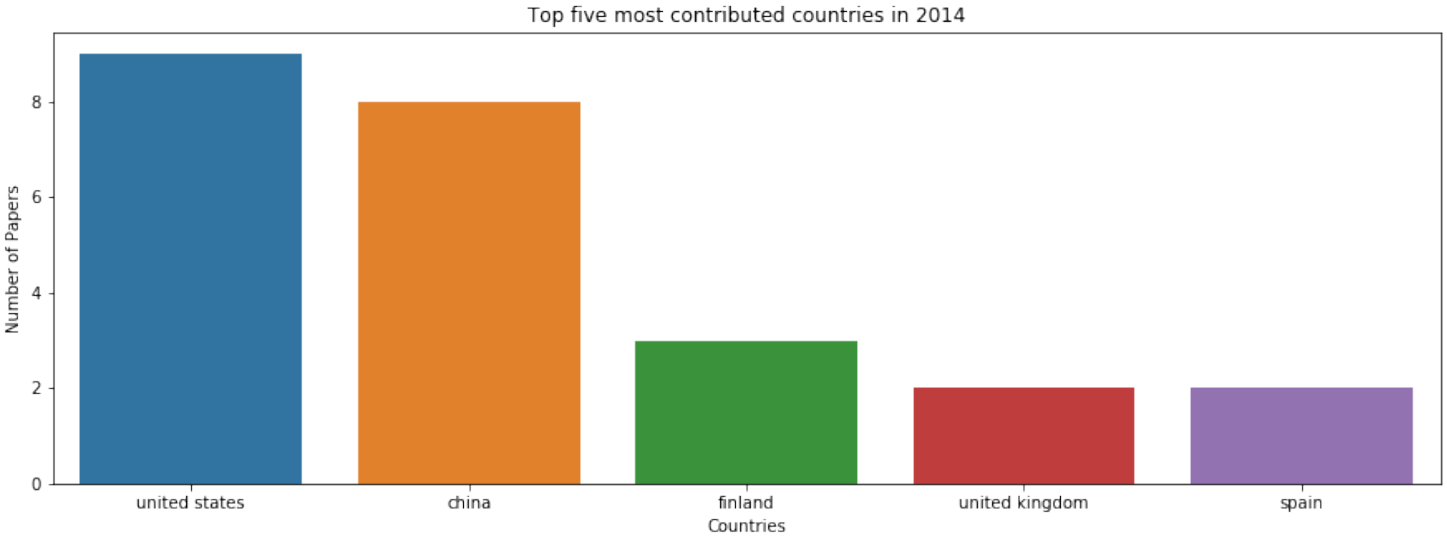


YEAR 2014

The contributions of top five countries in the year 2014 are sorted and listed below in the form of a table.

	Country	Number of Papers
0	united states	9
1	china	8
2	finland	3
3	united kingdom	2
4	spain	2

The top five countries that has contributed the most in the year 2014 are United States, China, Finland, United Kingdom and Spain. They are plotted in the form of a Bar Graph for Visualization.

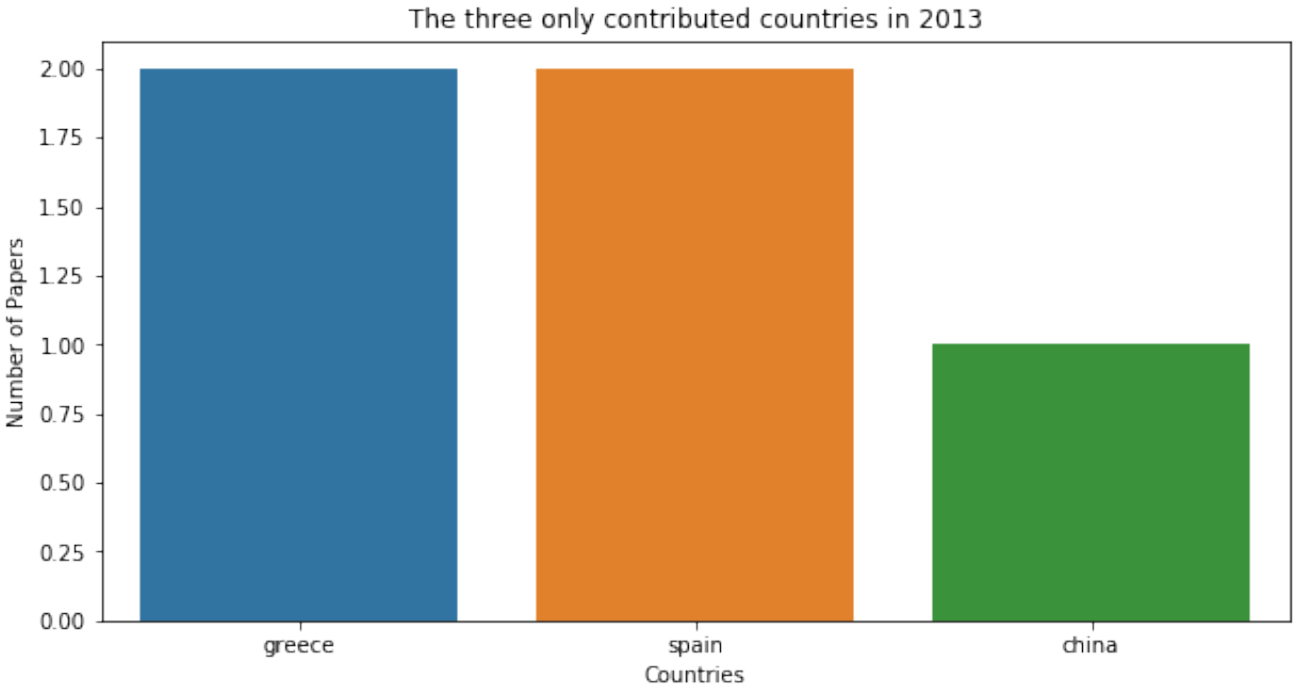


YEAR 2013

The contributions of the only three countries in the year 2013 are listed below in the form of a table.

	Country	Number of Papers
0	greece	2
1	spain	2
2	china	1

The only three countries that has contributed in the year 2013 are Greece, Spain and China. They are plotted in the form of a Bar Graph for Visualization.

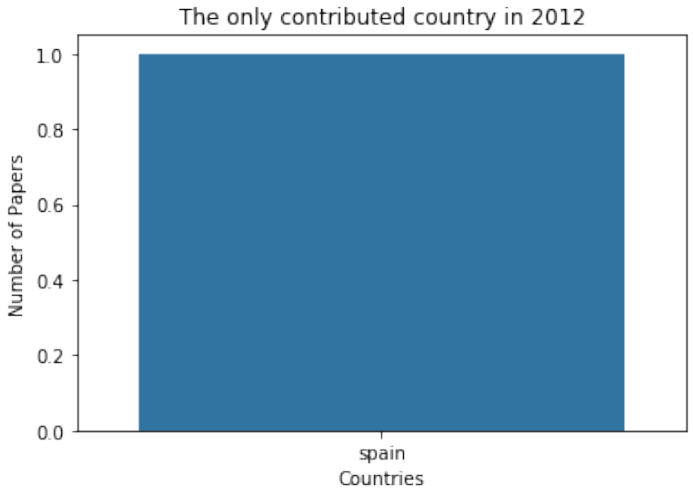


YEAR 2012

Only one country has contributed a paper and it is from Spain.

	Country	Number of Papers
0	spain	1

The Bar Chart Visualization for this year is as follows.



YEAR 2011

The contributions of the top five countries in the year 2011 are listed below in the form of a table.

	Country	Number of Papers
0	china	3
1	netherlands	2
2	united states	2
3	greece	2
4	australia	2

The top five countries that has contributed the most in the year 2011 are China, Netherlands, United States, Greece and Australia. They are plotted in the form of a Bar Graph for Visualization.