

H1B visa applications Analysis

Project Author: - Sandeep Kumar Bhandoria

Objective: The H1B is an employment-based, non-immigrant visa category for temporary foreign workers in the United States. For a foreign national to apply for H1B visa, an US employer must offer a job and petition for H1B visa with the US immigration department. This is the most common visa status applied for and held by international students once they complete college/ higher education (Masters, Ph.D.) and work in a full-time position.

DATA WE HAVE:

1	CERTIFIED-WITHDRAWN	UNIVERSITY OF MICHIGAN	BIOCHEMISTS AND
	BIOPHYSICISTS POSTDOCTORAL RESEARCH FELLOW	N	36067 2016 ANN ARBOR,
	MICHIGAN	-83.7430378	42.2808256
2	CERTIFIED-WITHDRAWN	GOODMAN NETWORKS, INC.	CHIEF EXECUTIVES
	CHIEF OPERATING OFFICER	Y	242674 2016 PLANO, TEXAS -96.6988856
	33.0198431		

TECHNOLOGY WE USED:

- 1) HDFS
- 2) Map Reduce programming in Java
- 3) HIVE
- 4) PIG
- 5) Sqoop
- 6) Mysql
- 7) Excel

SOFTWARE WE USED:

- 1) Virtual Machines
- 2) Eclipse
- 3) Ubuntu

PROJECT DESCRIPTION:

THE DATASET DESCRIPTION IS AS FOLLOWS:

THE COLUMNS IN THE DATASET INCLUDE:

- **CASE_STATUS:** STATUS ASSOCIATED WITH THE LAST SIGNIFICANT EVENT OR DECISION. VALID VALUES INCLUDE "CERTIFIED," "CERTIFIED-WITHDRAWN," "DENIED," AND "WITHDRAWN".

CERTIFIED: EMPLOYER FILED THE LCA, WHICH WAS APPROVED BY DOL

CERTIFIED WITHDRAWN: LCA WAS APPROVED BUT LATER WITHDRAWN BY EMPLOYER

WITHDRAWN: LCA WAS WITHDRAWN BY EMPLOYER BEFORE APPROVAL

DENIED: LCA WAS DENIED BY DOL

- **EMPLOYER_NAME:** NAME OF EMPLOYER SUBMITTING LABOUR CONDITION APPLICATION.
- **SOC_NAME:** THE OCCUPATIONAL NAME ASSOCIATED WITH THE SOC_CODE. SOC_CODE IS THE OCCUPATIONAL CODE ASSOCIATED WITH THE JOB BEING REQUESTED FOR TEMPORARY LABOUR CONDITION, AS CLASSIFIED BY THE STANDARD OCCUPATIONAL CLASSIFICATION (SOC) SYSTEM.
- **JOB_TITLE:** TITLE OF THE JOB
- **FULL_TIME_POSITION:** Y = FULL TIME POSITION; N = PART TIME POSITION
- **PREVAILING_WAGE:** PREVAILING WAGE FOR THE JOB BEING REQUESTED FOR TEMPORARY LABOUR CONDITION. THE WAGE IS LISTED AT ANNUAL SCALE IN USD. THE PREVAILING WAGE FOR A JOB POSITION IS DEFINED AS THE AVERAGE WAGE PAID TO SIMILARLY EMPLOYED WORKERS IN THE REQUESTED OCCUPATION IN THE AREA OF INTENDED EMPLOYMENT. THE PREVAILING WAGE IS BASED ON THE EMPLOYER'S MINIMUM REQUIREMENTS FOR THE POSITION.
- **YEAR:** YEAR IN WHICH THE H1B VISA PETITION WAS FILED

- WORKSITE: CITY AND STATE INFORMATION OF THE FOREIGN WORKER'S INTENDED AREA OF EMPLOYMENT
 - LON: LONGITUDE OF THE WORKSITE
 - LAT: LATITUDE OF THE WORKSITE
-
-

1 a) Is the number of petitions with Data Engineer job title increasing over time?

Yes. In this data Data Engineer increasing overtime.

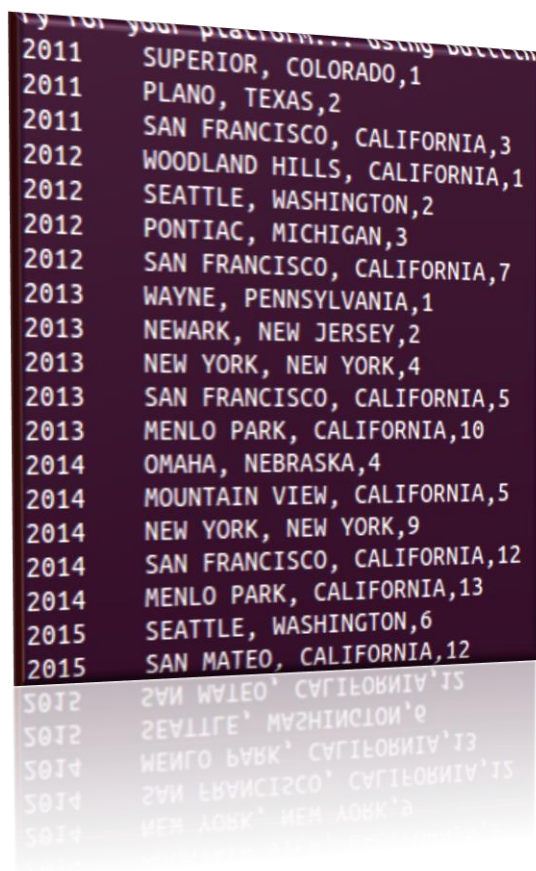


1 b) Find top 5 job titles who are having highest growth in applications.

This report shows that Top 5 job having highest growth in applications.

PROGRAMMER ANALYST	249038
SOFTWARE ENGINEER	121307
COMPUTER PROGRAMMER	70570
SYSTEMS ANALYST	61965
SOFTWARE DEVELOPER	42907
SOFTWARE DEVELOPER	45201

2 a) Which part of the US has the most Data Engineer jobs for each year?



2011	SUPERIOR, COLORADO,1
2011	PLANO, TEXAS,2
2011	SAN FRANCISCO, CALIFORNIA,3
2012	WOODLAND HILLS, CALIFORNIA,1
2012	SEATTLE, WASHINGTON,2
2012	PONTIAC, MICHIGAN,3
2012	SAN FRANCISCO, CALIFORNIA,7
2013	WAYNE, PENNSYLVANIA,1
2013	NEWARK, NEW JERSEY,2
2013	NEW YORK, NEW YORK,4
2013	SAN FRANCISCO, CALIFORNIA,5
2013	MENLO PARK, CALIFORNIA,10
2014	OMAHA, NEBRASKA,4
2014	MOUNTAIN VIEW, CALIFORNIA,5
2014	NEW YORK, NEW YORK,9
2014	SAN FRANCISCO, CALIFORNIA,12
2014	MENLO PARK, CALIFORNIA,13
2015	SEATTLE, WASHINGTON,6
2015	SAN MATEO, CALIFORNIA,12

This output result shows that Worksite of US has most Data Engineer jobs and group by worksite for each year.

2 b) find top 5 locations in the US who have got certified visa for each year.

NEW YORK, NEW YORK	2016	34639
NEW YORK, NEW YORK	2015	31266
NEW YORK, NEW YORK	2014	27634
NEW YORK, NEW YORK	2012	23737
NEW YORK, NEW YORK	2013	23537
NEW YORK, NEW YORK	2011	23172

In this Scenario We first find case_status = Certified or Certified_Withdrawn both and then order by location as descending.

This output shows that top 5 locations in the US have got certified visa for particular year.

3) Which industry has the most number of Data Scientist positions?

STATISTICIANS	249	2016
OPERATIONS RESEARCH ANALYSTS	178	2016
COMPUTER AND INFORMATION RESEARCH SCIENTISTS	172	2016
STATISTICIANS	166	2015
COMPUTER AND INFORMATION RESEARCH SCIENTISTS	161	2015
MATHEMATICIANS	82	2016
OPERATIONS RESEARCH ANALYSTS	75	2015
Statisticians	74	2014
Computer and Information Research Scientists	74	2014
COMPUTER OCCUPATIONS, ALL OTHER	69	2016
COMPUTER OCCUPATIONS, ALL OTHER	52	2015
Computer and Information Research Scientists	46	2013
COMPUTER SYSTEMS ANALYSTS	43	2016
SOFTWARE DEVELOPERS, APPLICATIONS	40	2016
SOFTWARE DEVELOPERS, APPLICATIONS	39	2015
Operations Research Analysts	36	2014
MATHEMATICIANS	26	2015
MATHEMATICIANS	26	2015
Operations Research Analysts	30	2014
SOFTWARE DEVELOPERS, APPLICATIONS	30	2012
SOFTWARE DEVELOPERS, APPLICATIONS	40	2010
SOFTWARE DEVELOPERS, APPLICATIONS	40	2010
COMPUTER OCCUPATIONS, ALL OTHER	40	2010

4) Which top 5 employers file the most petitions each year?

```
((TATA CONSULTANCY SERVICES LIMITED,2011),5416)
((MICROSOFT CORPORATION,2011),4253)
((DELOITTE CONSULTING LLP,2011),3621)
((WIPRO LIMITED,2011),3028)
((COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION,2011),2721)
((INFOSYS LIMITED,2012),15818)
((WIPRO LIMITED,2012),7182)
((TATA CONSULTANCY SERVICES LIMITED,2012),6735)
((DELOITTE CONSULTING LLP,2012),4727)
((IBM INDIA PRIVATE LIMITED,2012),4074)
((INFOSYS LIMITED,2016),25352)
((CAPGEMINI AMERICA INC,2016),16725)
((TATA CONSULTANCY SERVICES LIMITED,2016),13134)
((WIPRO LIMITED,2016),10607)
((IBM INDIA PRIVATE LIMITED,2016),9787)
((INFOSYS LIMITED,2015),33245)
((TATA CONSULTANCY SERVICES LIMITED,2015),16553)
((WIPRO LIMITED,2015),12201)
((IBM INDIA PRIVATE LIMITED,2015),10693)
((IBM INDIA PRIVATE LIMITED,2012),10003)
((MICROSOFT CORPORATION,2012),15501)
((DELOITTE CONSULTING LLP,2012),10223)
((INFOSYS LIMITED,2012),33542)
((IBM INDIA PRIVATE LIMITED,2012),10101)
```

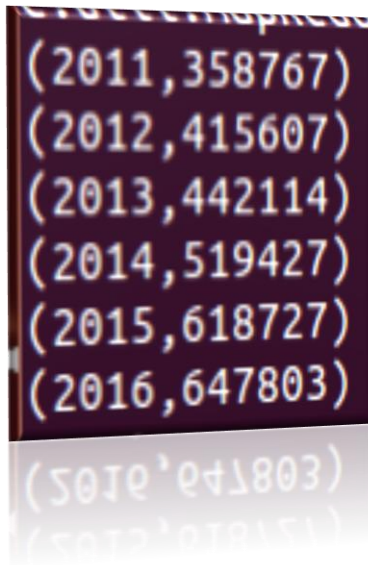
In this case study , we partitioned data by year and group by employer_name & year. After we count employer_name for particular year.

5) Find the most popular top 10 job positions for H1B visa applications for each year?

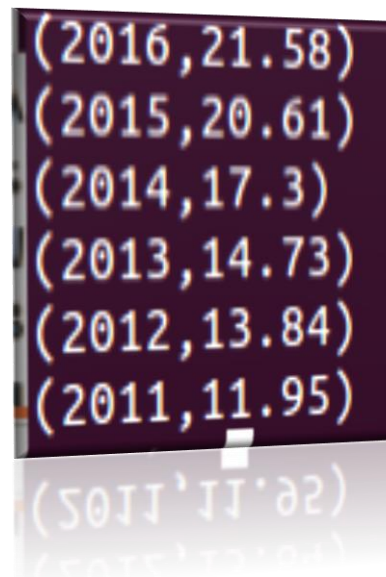
```
PROGRAMMER ANALYST 2011
SOFTWARE ENGINEER 2011
COMPUTER PROGRAMMER 2011
SYSTEMS ANALYST 2011
BUSINESS ANALYST 2011
COMPUTER SYSTEMS ANALYST 2011
ASSISTANT PROFESSOR 2011
PHYSICAL THERAPIST 2011
SENIOR SOFTWARE ENGINEER 2011
SENIOR CONSULTANT 2011
PROGRAMMER ANALYST 2012
SOFTWARE ENGINEER 2012
COMPUTER PROGRAMMER 2012
SYSTEMS ANALYST 2012
BUSINESS ANALYST 2012
COMPUTER SYSTEMS ANALYST 2012
SOFTWARE DEVELOPER 2012
PHYSICAL THERAPIST 2012
ASSISTANT PROFESSOR 2012
SENIOR CONSULTANT 2012
PROGRAMMER ANALYST 2013
SOFTWARE ENGINEER 2013
SOFTWARE ENGINEER 2013
PROGRAMMER ANALYST 2013
SENIOR CONSULTANT 2013
PHYSICAL THERAPIST 2013
SOFTWARE DEVELOPER 2013
```

this output result shows that top 10 job positions for particular year. We count job_title for each year and then order by job_title.

6) Find the percentage and the count of each case status on total applications for each year. Create a graph depicting the pattern of All the cases over the period of time.

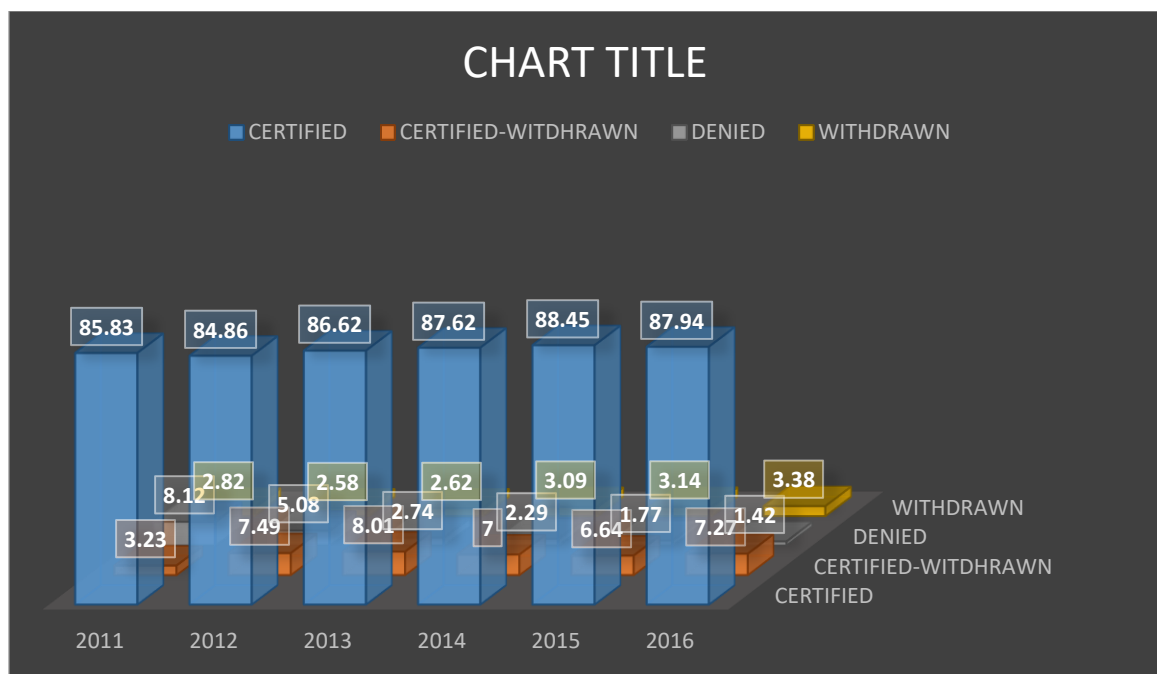


(A)



(B)

There are total Two output result. Picture A shows Count of Case_status for each year. And Picture B shows that Percentage of Case_status for each year.

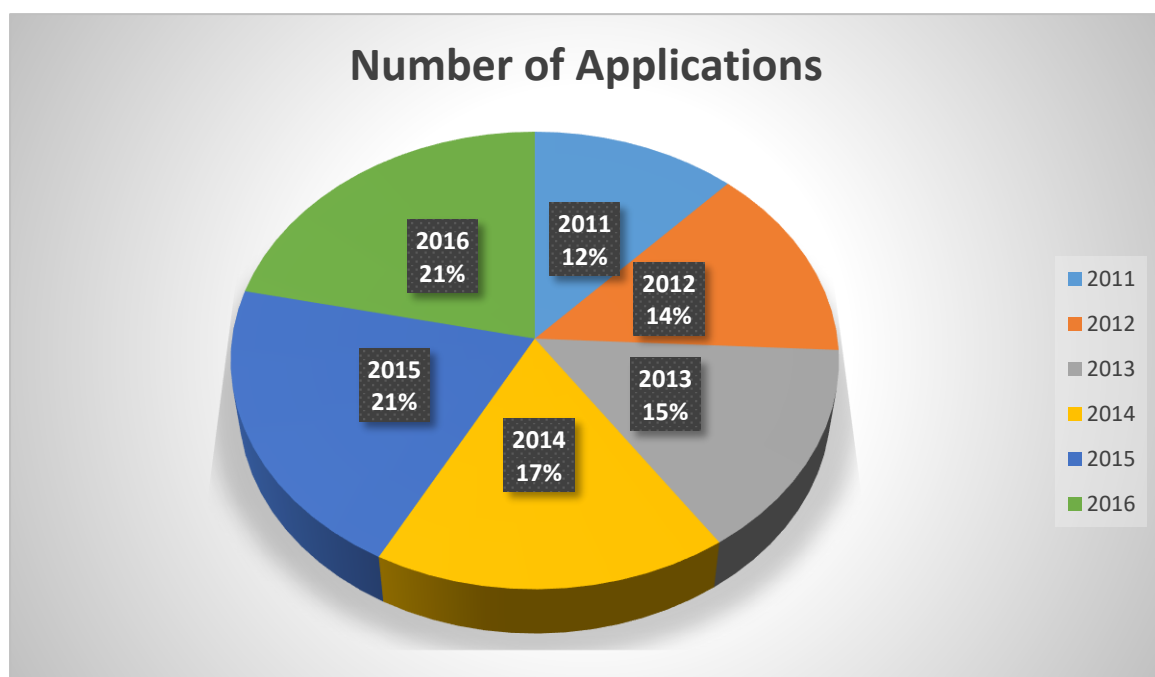


This Graph shows all

7) Create a bar graph to depict the number of applications for each year.

This results shows simple count of applications for each year.

2011	358767
2013	442114
2015	618727
2012	415607
2014	519427
2016	647803



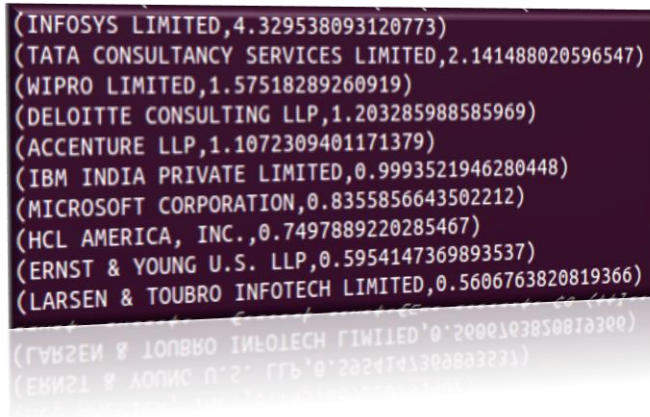
8) Find the average Prevailing Wage for each Job for each Year (take part time and full time separate). Arrange the output in descending order.

This records shows the average Prevailing Wage for particular job for each year. There are also Part time and Full time job. So, output will show both avg. Prevailing Wage as descending order.

PROFESSOR OF LAW	24120.0	80329.6	2015	
PROFESSOR OF LAW	24120.0	45968.0	2016	
LEAD TECHNICIAN	24117.5	20758.0	2016	
STUDIO ASSOCIATE	24107.0	38521.6	2015	
STUDIO ASSOCIATE	24107.0	62249.25		2016
PRIMARY TEACHER	24087.0	34361.6	2011	
PRIMARY TEACHER	24087.0	41995.2	2012	
PRIMARY TEACHER	24087.0	41297.5	2016	
BOOKKEEPING, ACCOUNTING, AND AUDITING CLERKS		24065.6	39062.4	2015
CURATOR OF REPTILE LAGOON	24045.0	35093.07		2016
ASIA PACIFIC STOCKBROKER	24045.0	26416.0	2016	
COACH (TRAMPOLINE, TUMBLING, CIRQUE)	24040.0	24044.8	2012	
PROGRAM DEVELOPMENT ADVISOR	24030.0	20363.2	2016	
ASSISTANT ROWING COACH	24025.0	34110.0	2016	
DANCER	24024.0	29660.8	2012	
DIRECTOR OF JUNIOR TENNIS	24017.0	37544.0	2013	
BUSINESS COMMUNICATIONS SPECIALIST	24003.0	49247.47		2015
INTERNATIONAL ADMISSIONS ASSOCIATE	24003.0	77396.8	2015	
BEHAVIORAL SPECIALIST	24003.0	24918.0	2016	
PUBLIC HEALTH EDUCATOR	24000.0	37266.67		2015
FEATURE JOURNALIST	23982.4	36275.2	2013	
INSTRUCTOR/COORDINATOR	23982.0	48235.2	2012	
ASSISTANT WOMEN'S SOCCER COACH	23952.86		21084.27	2014
ASSISTANT WOMEN'S SOCCER COACH	23952.86		21520.65	2016
ASSISTANT WOMEN'S SOCCER COACH	53025.80		51250.92	5010
ASSISTANT WOMEN'S SOCCER COACH	53025.80		51084.51	5014
INSTRUCTIONAL COORDINATOR	53085.0	48532.5	5015	
LEARNING TECHNOLOGIST	53085.4	30512.5	5013	
LIBRARY ASSISTANT	54000.0	31500.01		5012
LIBRARY ASSISTANT	54000.0	34210.0	5010	
LIBRARY ASSISTANT	54000.0	34210.0	5010	

9) Which are employers along with the number of petitions who have the success rate more than 70% in petitions and total petitions filed more than 1000?

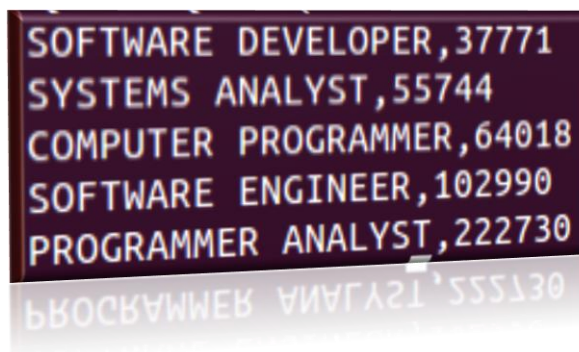
Here result shows that success rate of employees.



Employer	Success Rate
(INFOSYS LIMITED,4.329538093120773)	4.329538093120773
(TATA CONSULTANCY SERVICES LIMITED,2.141488020596547)	2.141488020596547
(WIPRO LIMITED,1.57518289260919)	1.57518289260919
(DELOITTE CONSULTING LLP,1.203285988585969)	1.203285988585969
(ACCENTURE LLP,1.1072309401171379)	1.1072309401171379
(IBM INDIA PRIVATE LIMITED,0.9993521946280448)	0.9993521946280448
(MICROSOFT CORPORATION,0.8355856643502212)	0.8355856643502212
(HCL AMERICA, INC.,0.7497889220285467)	0.7497889220285467
(ERNST & YOUNG U.S. LLP,0.5954147369893537)	0.5954147369893537
(LARSEN & TOUBRO INFOTECH LIMITED,0.5606763820819366)	0.5606763820819366

10) Which are the job positions along with the number of petitions which have the success rate more than 70% in petitions and total petitions filed more than 1000?

In this Output, there are number petitions for particular job positions .



Job Position	Number of Petitions
SOFTWARE DEVELOPER	37771
SYSTEMS ANALYST	55744
COMPUTER PROGRAMMER	64018
SOFTWARE ENGINEER	102990
PROGRAMMER ANALYST	222730

11) Export result for question no 10 to MySql database.