Data exploration analysis on job trends using R

**ABSTRACT**— Today’s constantly changing and fluctuating corporate environment requires professionals and students to constantly be acquainted with the ever-changing job trends and requirements and job market in order to prepare well for the jobs being demanded. And when it comes to the information technology(IT) jobs sector it becomes more important to be updated with the new technologies and newer job requirements as they are changing very rapidly. Recently the demand of data science related job positions has seen a huge growth and this is all due to the recent *data explosion* incurred by the industries and organizations globally. The need to harness and utilize the amount of information hidden inside these huge datasets for effective decision-making has become the need of the hour. And this is where a data analyst or a data scientist comes into play. They are domain experts who have the skillset and knowledge to extract hidden meaning from data.

This paper will illustrate the use of data mining and advanced data analysis techniques such as data aggregation, summarization and reduction etc along with data visualization to understand and analyze the job trends in United states of America(USA) specially for data science related job positions from year 2011 to 2016. The raw data was collected from the US Office of Foreign Labor Certification which generates data about the immigration programs including the H1-B visa. The raw data was unclean and messy so a set of transformations were performed for making data more suitable for fast and rapid analysis and exploration. The data was then analyzed and visualized to find answers about the job trends in USA over time.

**INTORDUCTION**—The H1-B visa is an employment based, non-immigrant visa category for temporary foreign workers in the United States. The first step of the H1B application process is taken by the US employer(firm) to file the work visa application on behalf of a foreign worker [2]. The employment letter should include the necessary details such as dates of employment, job position, salary offered, detailed description, contact information etc. The next step is to start preparing the petition by the sponsoring employer and file it at the USCIS(United States Citizenship and Immigration Services) office. Processing times of H1B application depends and is subject to locations. The world is creating data at a tremendous pace of and around 5 zettabytes of data is available today and this figure is projected to grow to over 44 zettabytes by 2020. This data comes from a wide variety of sources including government, customers, web logs, social media etc and in a variety of different formats such as structured, semi-structured and unstructured. Such vast amounts of data stores enormous opportunities and hidden useful information, but only for organizations that have the will and the talent to extract meaning and utilize this information. This talent shortage is pervasive across sectors, not just in companies who are easily identified as technology companies.[7] The requirement and need of data-fluent talent has become one of the definitive jobs issues globally, and touches all levels across all industries [6]. This is the reason data science related job positions have seen such a huge increase in the past few years and this is what this paper tries to find out and answer.

LITERATURE REVIEW: (5-6 paras-each 250-300 words)

PROPOSED METHOD:

This paper utilizes the H-1B petition disclosure data and various data analysis techniques to analyze the employers with the most H1-B applications issued, recent explosive growth in data science related job positions and relationship between salaries offered etc from 2011 to 2016. This paper answers interesting questions such as –“Who are the top US employers who issued the most H1B visa applications”, ”What are the annual wages of the employees of several technical job positions”, “What are the most preferred worksites” etc, and same questions asked for data science related job positions too. The Top ten list of companies which issue most H1B visa applications are dominated by the Indian-origin IT companies with most applications issued in year 2016. *Infosys* leads the pack followed by *Wipro, Tata Consultancy* *services, Deloitte* etc were the next top three companies which issued most H1B visa applications and are of Indian-origin. There has been a big increase in demand of data scientist, data-engineer and machine learning engineer from 2011 to 2016. The highest annual salary amongst the above job positions is for machine learning engineer. The annual salaries have been somewhat fluctuating for data scientists and machine learning engineers, but for a data-engineer it has increased over time form 2011-2016. The highest number of H1B applications were issued for data scientist job title and least for machine learning engineer. The top five companies which issue most H1B applications for data-science related job positions are *Amazon, Facebook, Microsoft, LinkedIn* and *Uber* whereas companies which pay highest annual salaries for data-science job positions are *Netflix, Apple, Airbnb, Twitter* and *Paypal*.

RESULTS AND EVALUATIONS:

CONCLUSION:

REFRENCES:

1. <https://www.bls.gov/web/empsit/ceshighlights.pdf>
2. <http://www.immi-usa.com/h1b-application-process-step-by-step-guide/>
3. <http://www.mrrc.isr.umich.edu/publications/papers/pdf/wp285.pdf>
4. Lahey, Joanna N. (2008). “Age, women, and hiring: an experimental study,” Journal of Human Resources,
5. [Polanyi's Paradox and the Shape of Employment Growth  
   David H. Autor .Federal Reserve Bank of St. Louis: Economic Policy Proceedings, Reevaluating Labor Market Dynamics. 2015.](https://economics.mit.edu/files/11650)

1. <https://www.stemconnector.com/wp-content/uploads/2016/12/Big-Data-1.pdf>