**ALY 6070 DATA MANAGEMENT AND BIG DATA**

**CAREER SATISFACTION ANALYSIS**

By

Pratibha Rai, Wanyang Cheng, Kaushik Bvk and Zhenhao Zong

Northeastern University

Analysis Report Prepared for

ALY 6110 81041 Data Management & Big Data

MAY 15, 2018

Instructor - Janos Mako

### **Table of Contents**

* Introduction
* Data Source
* Data Cleaning
* Methods
* Analysis and results
* Insights
* Conclusion
* References

### **Introduction**

A career can be defined as our occupational progress through our lifetime. It includes a pattern of work-related experiences and activities, such as job positions, duties, decisions and our subjective interpretations about work-related events. Being good at work gives a sense of achievement, a key ingredient of life satisfaction discovered from positive psychology.

It also gives the power to negotiate for the other components of a fulfilling job, such as the ability to work on meaningful projects, undertake engaging tasks and earn fair pay.  
Job satisfaction in any field of work depends a lot on how conducive the work environment is. The work itself, the pay and the scope for promotion are only some of the factors which have an impact on job satisfaction. While looking for the job, almost everyone thinks about the size of the company, salary, benefits, location, etc. Few people look for some additional things as well. They might think of few things upfront like how long the commute is, look and feel of the office, types of projects, etc. We can call these the secondary factors. More jobs mean more competition amongst companies to hire the best and brightest, so software engineering jobs pay well and often come with great benefits. The flip side of this is that many engineers receive multiple simultaneous offers. Making the right long-term career decision can be difficult when evaluating several opportunities. Being a software engineer is an excellent career choice for a person who excels at both left and right-brained thinking - analytic and numerical skills in addition to conceptual problem-solving skills. Software engineers are innate problem-solvers, good at collaboration and keen to see an issue through to successful completion. There are so many job roles but is everyone satisfied with their career? We are going to find that out with our analyses.

### **Data Source**

The original data set in the form of CSV file is retrieved from “Kaggle.com”, using data from Stack Overflow Developer Survey, which consists of 51392 rows and 154 columns. The Variables used in our analysis are Career Satisfaction, Job Satisfaction, Gender, Major Undergrad, Formal Education, Employment Status, Country, Race, Highest Education Parents and Salary.

### **Data Cleaning**

This dataset was provided in a clean format except for few missing values, Null Values and unfiltered variables. We have filtered the top number of Countries, Employment Status to Full-time and Part-time etc., to establish our analysis. Secondly, the part of regions (unfiltered) which belonged to countries are filtered into country itself. For example, the “American Samoa” are filtered into United States as they belong in United States territory etc. We have excluded the Null values for our Analyses.

### **Methods**

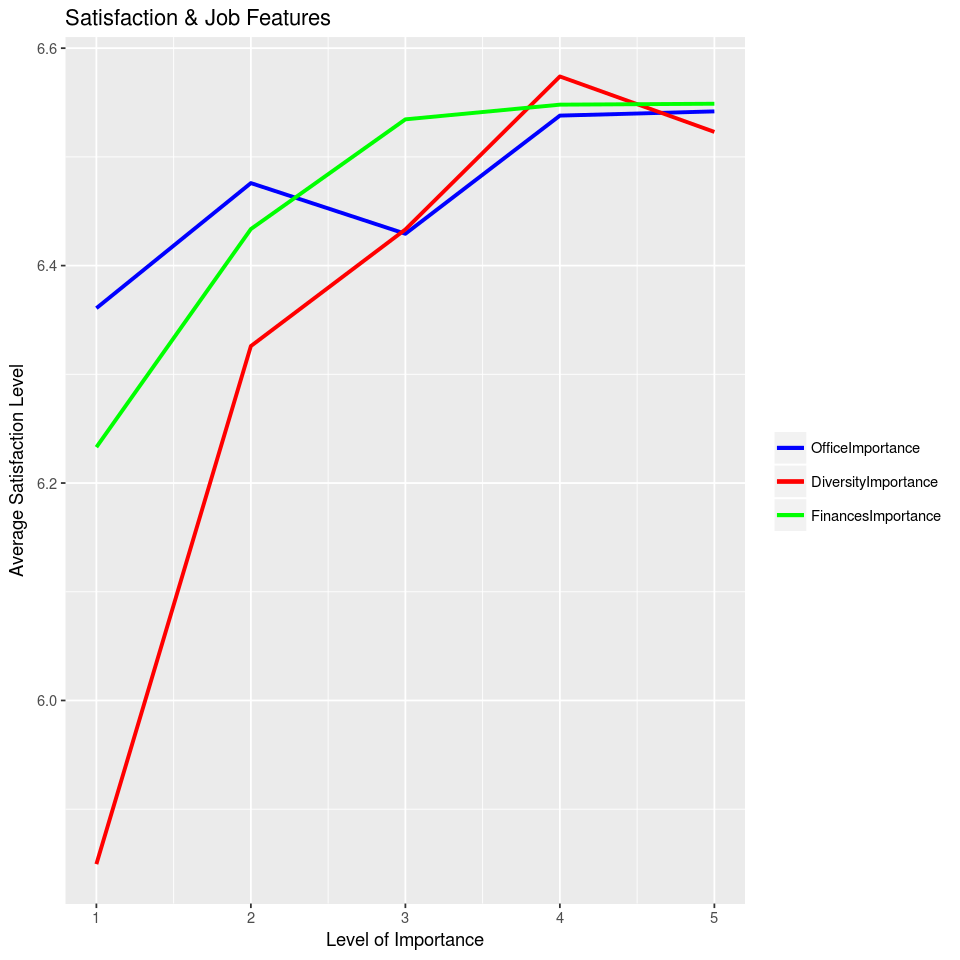
We began by exploring the dataset, then we created several dashboards in Tableau to understand the distribution of our dataset. Furthermore, we performed R programming in developing the Bar charts, Radar plots and Heat maps as well. After performing both R and Tableau, we also explored the use of Business Intelligence tool Qlik Sense to understand the analyses in depth by playing with variables and data sets.

### **Analysis and Results**

**VISUALIZATION & CODE SNIPPET**

For this analysis, we will use SparkR, SparkSQL, Qlik Sense, ggplot with R programming to gather some insights.

|  |
| --- |
| # Import SparkR library library(SparkR) library(ggplot2) library(reshape2)  # Load csv file dataPath <- "/FileStore/tables/survey\_results\_public.csv" career <- read.df(dataPath, source = "csv", header="true", inferSchema = "true")   # Display summary statistics summ = summary(career) showDF(summ)  # Creating temporary table to extract data createOrReplaceTempView (career, "career") |



**Fig 1. Satisfaction rate with Job features**

When people look for job, they might think of few thinks upfront like how long the commute is, look and feel of office, types of projects, etc. There are also some people who do not consider these factors while searching for a job. We will see how the job satisfaction vary across the people who evaluate these factors upfront vs those who don’t. There are five categories depending on how much importance, people place on these factors.

Here we are analyzing the average job satisfaction rating based on those employees who prefer to understand office place before joining the new job:

|  |
| --- |
| # Assess Job Office vs Job satisfaction query <- 'SELECT trim(AssessJobOffice) as level, AVG(JobSatisfaction) as OfficeImportance FROM career WHERE AssessJobOffice!= "" GROUP BY level ORDER BY OfficeImportance DESC' ma\_career <- sql(query) JobOffice <- collect(ma\_career) JobOffice |

         level          Office Importance  
1       Very important         6.541851  
2            Important     6.537969  
3   Not very important         6.475891  
4   Somewhat important         6.429268  
5 Not at all important         6.360825

Here we are focusing on the job diversity and are analyzing the satisfaction rating with respect to assess job diversity.

|  |
| --- |
| # Assess Job Diversity vs Job satisfaction  query <- 'SELECT trim(AssessJobDiversity) as level, AVG(JobSatisfaction) as DiversityImportance FROM career  WHERE AssessJobDiversity != "" GROUP BY level ORDER BY DiversityImportance DESC' ma\_career <- sql(query) JobDiversity <- collect(ma\_career) JobDiversity |

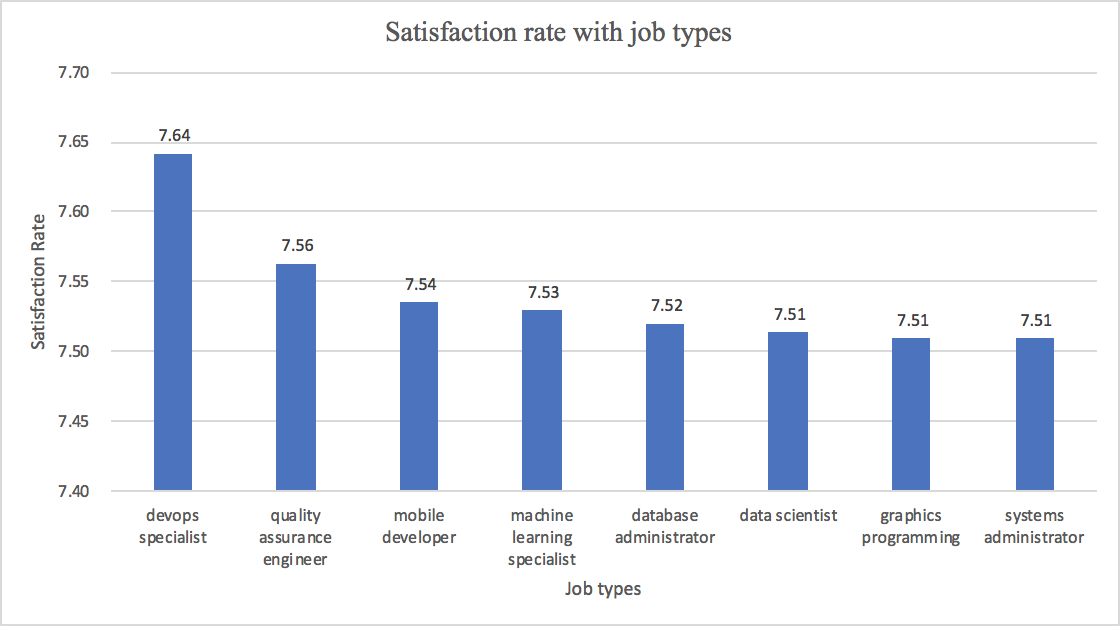
      level     Diversity Importance  
1            Important        6.573962  
2       Very important            6.522963  
3   Somewhat important            6.433248  
4   Not very important            6.325976  
5 Not at all important            5.849462

Here we are analysing the job satisfaction rate with respect to the finances.

|  |
| --- |
| # Assess Job Finances vs Job satisfaction  query <- 'SELECT trim(AssessJobFinances) as level, AVG(JobSatisfaction) as FinancesImportance FROM career  WHERE AssessJobFinances != "" GROUP BY level ORDER BY FinancesImportance DESC' ma\_career <- sql(query) JobFinances <- collect(ma\_career) JobFinances |

               level Finances Importance  
1       Very important           6.548881  
2            Important       6.548039  
3   Somewhat important           6.534527  
4   Not very important           6.433621  
5 Not at all important           6.232975

Opportunities for training and development are paramount in decisions regarding employee career choices. Everyone wants to do to their dream job for their satisfaction. So here we are analyzing the career satisfaction rate with respect to types of job.



**Fig 2. Career satisfaction rate with respect job types**

|  |
| --- |
| query <- 'SELECT split(DeveloperType, ";") as dev, CareerSatisfaction FROM career' query <- 'SELECT explode(split(lower(DeveloperType), ";")) as dev, CareerSatisfaction FROM career' ma\_career <- sql(query) r\_df <- collect(ma\_career) createOrReplaceTempView(ma\_career, "satisfaction") |

|  |
| --- |
| query <- 'SELECT trim(DevType) as dev\_new, AVG(CareerSatisfaction) as satisfactionrate FROM satisfaction WHERE lower(trim(dev))!= "other" GROUP BY dev\_new ORDER BY satisfactionrate Desc limit 8' ma\_dev <- sql(query) re\_df <- collect(ma\_dev) display(re\_df) |

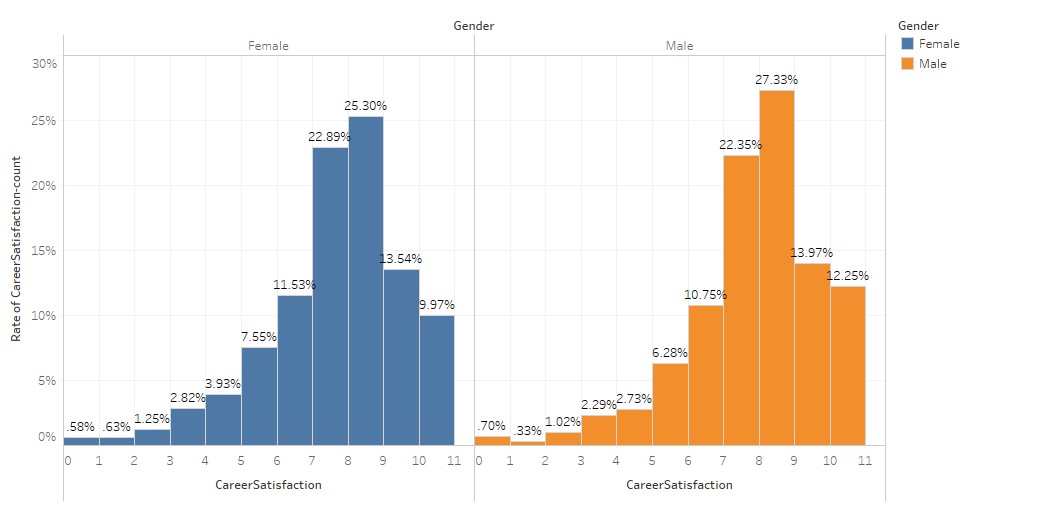
**Gender Analysis:**

1. **Career Satisfaction and Job Satisfaction with regards to Gender:**

Using filter function in excel to pull out the columns of Gender and Career satisfaction to

create a new data set and use tableau to analyze the data, then getting the distribution of

rate of career satisfaction between male and female which shown as following:

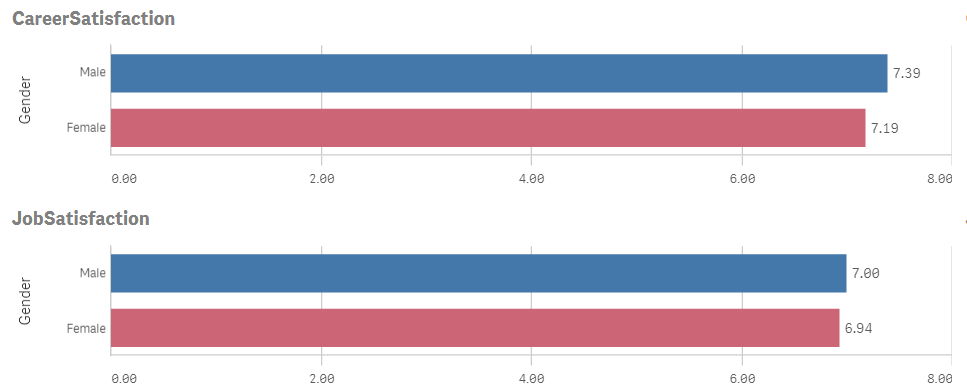


**Fig 3: Career Satisfaction vs Gender**

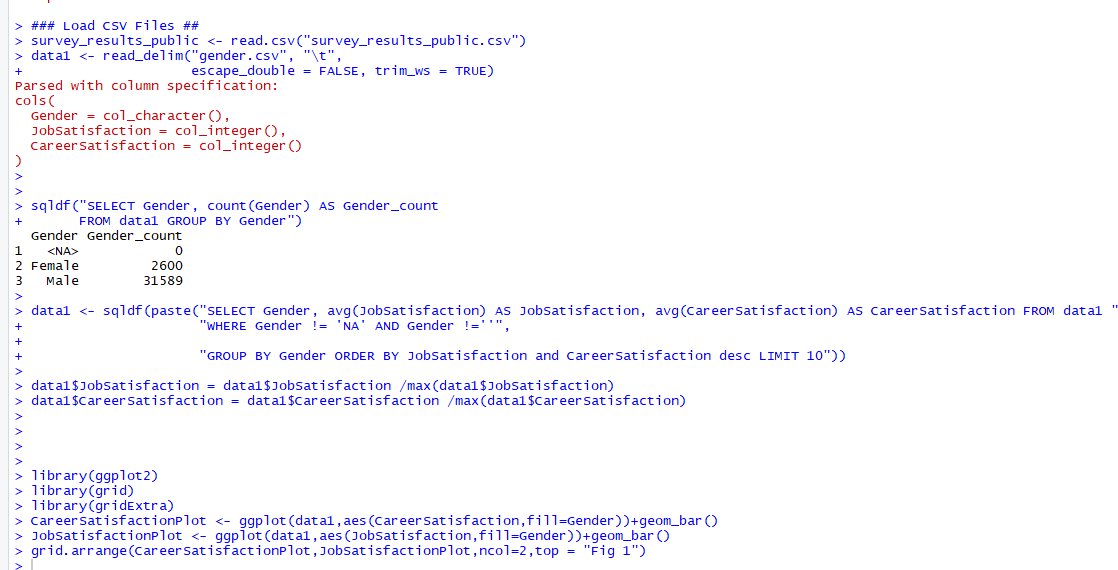
From the above graph we see that the cumulative values for Rate of Career Satisfaction in both Males and Females are almost similar with the max value at 8 rating for Females and max value at 8 rating for Males as well.

To show the exact average value for Career Satisfaction and Job Satisfaction for both females and Males, we have also used RSQL programming to get the analysis.

Using R programming and taking out the average values between Job Satisfaction and Career Satisfaction based on Gender, we could plot a bar chart which is shown below:



**Fig 4: Career and Job Satisfaction Average maximum rating for Males and Females**



From the above Bar chart and R console, we could analyse that the maximum Average Career Satisfaction rating for Males is 7.39 and for Females is 7.19, which is almost similar.

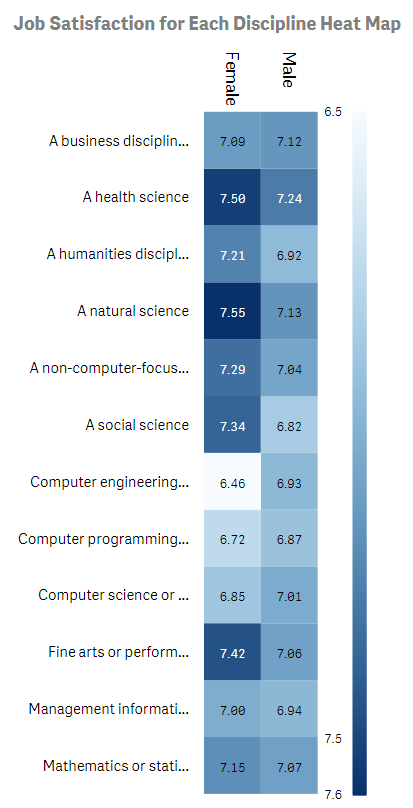
Similarly, the maximum Average Job Satisfaction rating for Males is 7.00 and for Females is 6.94, which is again almost similar.

From these two analyses we can say that both Job and Career Satisfaction does not actually depend on Gender, both Females and Males have equal amount of Satisfaction.

1. **Formal Education and Job Satisfaction rating based on Gender:**

We know that Formal Education is an important pathway for any individual to build a good career ahead.

Here we have analysed which Formal education has high job satisfaction rating for Males and Females using Heat Map chart as shown below.



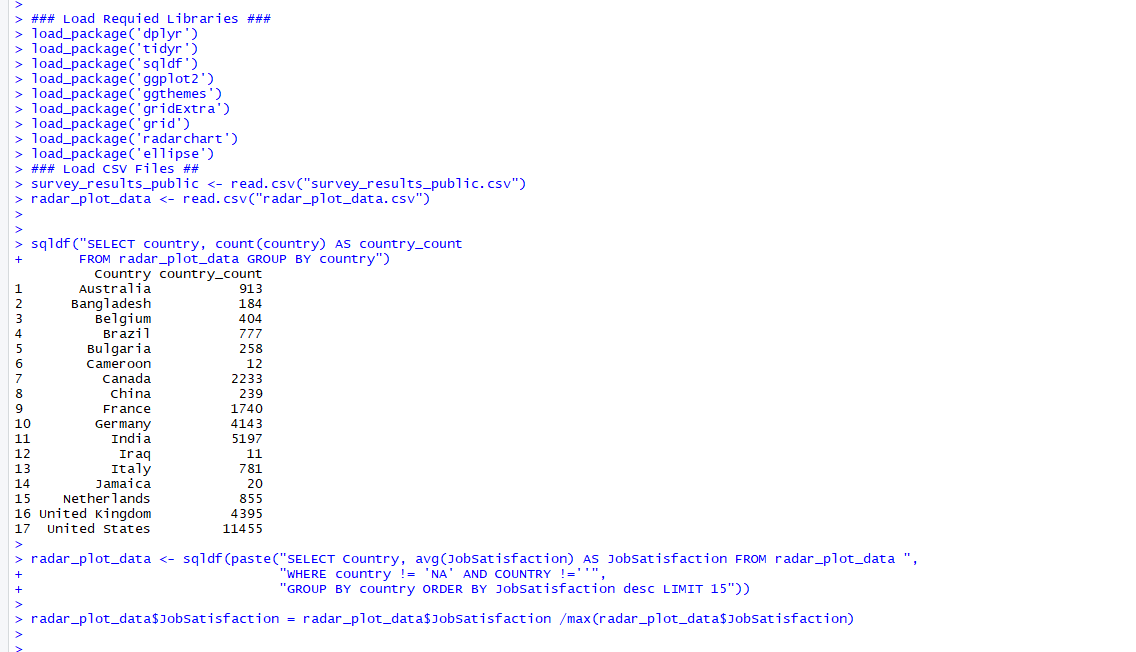
**Fig 5: Heat Map on Job Satisfaction for Formal Education**

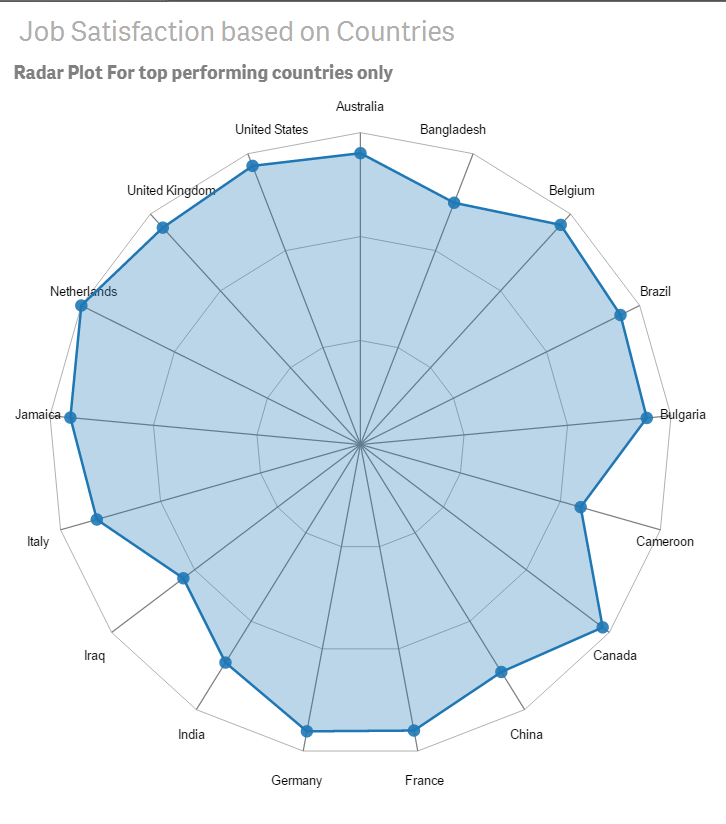
From the above analysis we see that Natural and health Science department have a higher correlation towards Job Satisfaction rating of 7.50 and 7.55 among women and Similarly Health Science and Mathematics or Statistics hold a higher Job Satisfaction rating of 7.13 and 7.07 among Men.

**Job Satisfaction rating based on Top Countries:**

Here, we are trying to find out which country has high Job Satisfaction rate and which country has low Satisfaction rate.

To analyse the above parameters, we took Top Countries, Job Satisfaction rating value and plotted them using a Radar Chart using R Programming.





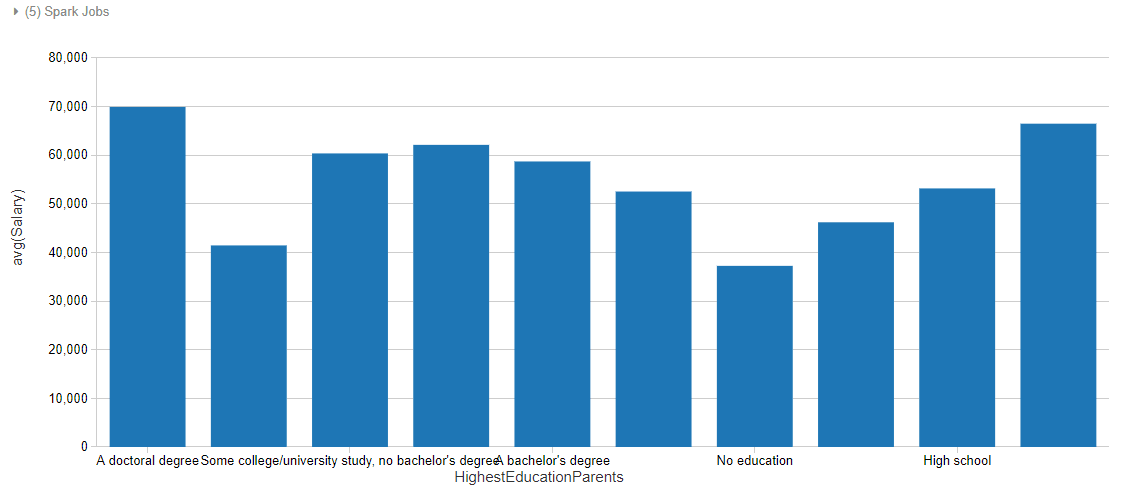
**Fig 6: Radar Chart for Job Satisfaction based on Top Countries**

From the above radar chart, we analyse that:

* Netherlands has the highest average Job Satisfaction rating with a value of 7.
* Top 4 performing countries with average high Job Satisfaction rating are Netherlands, United States, Canada and United Kingdom.
* Cameroon is an Outlier with a very low average Job Satisfaction rate of 5.
* This is one of many reasons why United States, United Kingdom, Canada, etc have high number of Immigrants and Students for higher education every year.

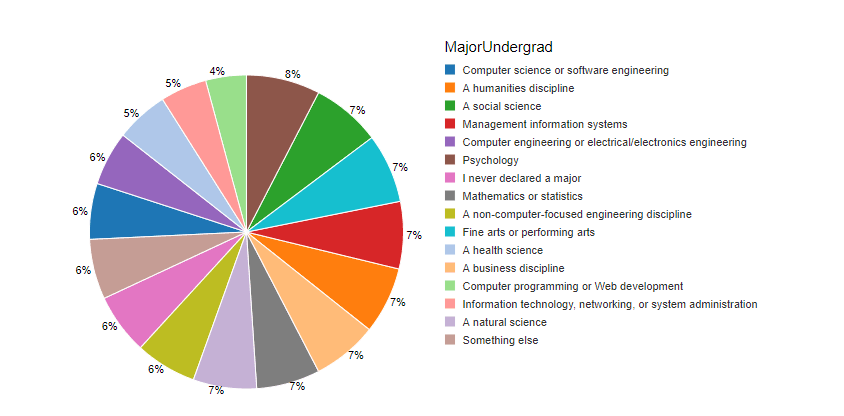
**Analysis on Major Undergrad and Salary:**

Here we are analysing the average Salaries for corresponding Major Undergrad and establishing a relationship on how this impacts on Career Satisfaction.



**Fig 7: Bar chart on Major Undergrad vs Salary**

|  |  |
| --- | --- |
| A doctoral Degree | $ 69934.70 |
| Primary Elementary School | S 41454.78 |
| Some college/university study, no bachelor’s degree | $ 60381.30 |
| A master’s degree | $ 62131 |
| I don’t know/not sure | $ 52539.18 |
| No education | $ 37254.59 |
| I prefer not to answer | $ 46216.74 |
| High school | $ 53178.63 |
| A professional degree | $ 66492.61 |



**Fig 8: Pie Chart for Major Undergrad**

From the above data, we can see that individuals who will go to universities and take majors usually get a good salary after they go into the society. We can see that students who comes from STEM major usually get a salary which is about $ 70000. Students who come from medical fields also have an average salary which is about 60000. Students who come from Business school have an average salary which is about 70000. Some students who choose arts or other major also have an average salary which is about 65000. There two exceptions that students coming from computer programming or web development and information technology only get salary which is below $ 50000. We can see that the difference of salary of individuals choosing different major not very large. No matter what major students choose, students usually can get higher salary that they want if they master their skills enough in their fields. These data also prove that education is one effective way to help individuals to improve income level in their future careers.

### Insights

* It helps improve job satisfaction, if one thinks about various aspects of work, such as commute, office, project, diversity, etc. upfront.
* Those with bachelor’s and master's degree have the higher employment rate in IT industry.
* Regardless of gender, men and women have similar career and Job Satisfaction.
* Top Performing Countries have higher Job Satisfaction rate than rest other Countries, which are attracting high number of Immigrants every year.
* Students pursuing their interests of Major Undergrad and advance onto the same field in Careers and Jobs, usually end up having more Job Satisfaction and eventually Career Satisfactions as well.

### Conclusion

We as a group have ingested the data sets, studied them, analysed and plotted various charts to establish our stats on how Career Satisfaction is rated and on what factors it depends on.

We demonstrated on how the variables Salary, Formal Education, Major Undergrad, etc have a positive correlation and may be the reason for an Individual’s Career to be successful.

We worked on analysing how each country shows its presence in providing high resources and Research through various means to make sure all the Individuals from various countries, who come to top performing Countries for Higher Education and Jobs learn, grasp and create a digitized educational society in learning new things and making a better Career for themselves.

### References

* Apache Spark. (n.d.). SparkR (R on Spark). Retrieved from <https://spark.apache.org/docs/latest/sparkr.html>
* Jonathan. (2017). Stack Overflow Developer Survey. Retrieved from <https://www.kaggle.com/stackoverflow/so-survey-2017/data>
* McKenna, L. (April 21, 2016). The Ever-Tightening Job Market for Ph.D.’s <https://www.theatlantic.com/education/archive/2016/04/bad-job-market-phds/479205/>