2018 DEVELOPER SURVEY ANALYSIS

Nitish Jaiswal

March 11th, 2020

Contents

Background of Problem and Stakeholders

Every year, Stack Overflow conducts a survey of developers from all over the globe in which it asks them subjective and objective questions about their work and home lives. Some of the questions are directly related to work, i.e. type of developer, years of coding experience, size of company you work in, undergrad major, full time/part time employment, salary and job/career satisfaction, whereas others aren't directly related but can contribute to work-life balance and career satisfaction, i.e. hours spent outside or exercising, time they wake up, if they code as a hobby or not, whether they contribute to open source projects, their views on AI's future and their thoughts on ethics at work.

Since career satisfaction is an amalgamation of the many factors listed above, I wanted to see how each of them contributed to it, and which ones were the most significant when it comes to career satisfaction as well as dissatisfaction. This data can be used by companies to increase employee satisfaction, reduce turnover and hence increase overall productivity. It can also be used by employees to understand and work on things that could improve their job/career satisfaction. Even though this survey is conducted all over the world, about 70% of the responses come from European or North American (mostly America and Canada) countries. Hence, the data is skewed towards developers from developed countries.

Data and description of features

The data was obtained from Kaggle: https://www.kaggle.com/stackoverflow/stack-overflow-2018-developer-survey. At first, I checked the number of respondents in the survey and the various questions that were asked. I picked about 20 questions (features) which would give us insight into a good mix of subjective (personal opinions on AI, Ethics and its implications) as well as objective (hours spent exercising, Undergrad Major, salary, years of experience) measures as mentioned above.

The survey had a total of 98,855 respondents from all over the world. However, there were a lot of questions to which respondents didn't provide a response to. As a result, about 69,000 responses had at least one null value in them. There were about 28,000 rows with no null values. I couldn't drop all the rows with any null values as this would introduce a bias into my analysis where I only use data from surveyors who answered all the questions. Since the features are highly personal and not interdependent, I couldn't replace the missing values accurately predict them without adding bias to the data. As a result, I dropped only the datapoints with null values from the following features:
'Employment','ConvertedSalary','EthicalImplications','AIFuture','YearsCoding','EthicsReport','EthicsChoice', 'CompanySize', 'HoursOutside', 'Exercise', 'CareerSatisfaction', 'DevType', 'FormalEducation', 'YearsCodingProf', 'HopeFiveYears', 'LastNewJob', 'JobSearchStatus', 'CommunicationTools', 'TimeFullyProductive', 'UpdateCV', 'OperatingSystem', 'HoursComputer', 'SkipMeals', 'Gender', 'Age', 'Country', 'LanguageWorkedWith', 'LanguageDesireNextYear', 'IDE'.

To check how well different countries and continents are represented, I created a table with the most represented countries.

	Country	NumResponses	Continent
1	United States	12626	North America
2	India	4932	Asia
3	United Kingdom	3588	Europe
4	Germany	3507	Europe
5	Canada	1862	North America
6	France	1312	Europe
7	Russian Federation	1279	Asia
8	Australia	1170	Oceania
9	Brazil	1169	Sount America
10	Netherlands	978	Europe

Table 1: 10 Most Represented Countries

As mentioned above, the data is skewed since about 70% of the responses come from European or North American (mostly America and Canada) countries. Also, about two-thirds of the responses came from the top 10 most represented countries shown above.

Continent	NumResponses
Europe	18995
North America	15063
Asia	10497
Sount America	1945
Oceania	1481
Africa	926
Other	36

Table 2: Most Represented Continents

Statistical Analysis

Firstly, I performed some statistical analysis to check if the two populations, ones who filled the responses completely and ones who left some entries blank, were different. After converting Career Satisfaction to numerical values and performing a T-Test, I found that the more questions the respondent answers, the more likely they are to have a higher career satisfaction. Using the T-Test, it was found that the population of respondents who didn't reply to multiple questions are different from the ones who did. Some results from the statistical analysis are below:

Respondents with multiple empty responses:

Mean Career Satisfaction: 1.24 (scale of -3 to 3)

Standard Deviation of career satisfaction: 1.68

Respondents with all filled responses:

Mean Career Satisfaction: 1.14

Standard Deviation of career satisfaction: 1.63

P-value: 8.97e-24

Data Exploration and Visualization

I explored the features individually to see what kinds of patterns and correlations emerge between them and career satisfaction. I used both data visualization to see if there were any clear patterns, and also checked the correlations matrix to see if any features were correlated.

Career Satisfaction

About 40% of the people surveyed were moderately satisfied, and on a whole, about 75% of the people were satisfied to some extent with their jobs. About 17% were dissatisfied and the rest were indifferent.

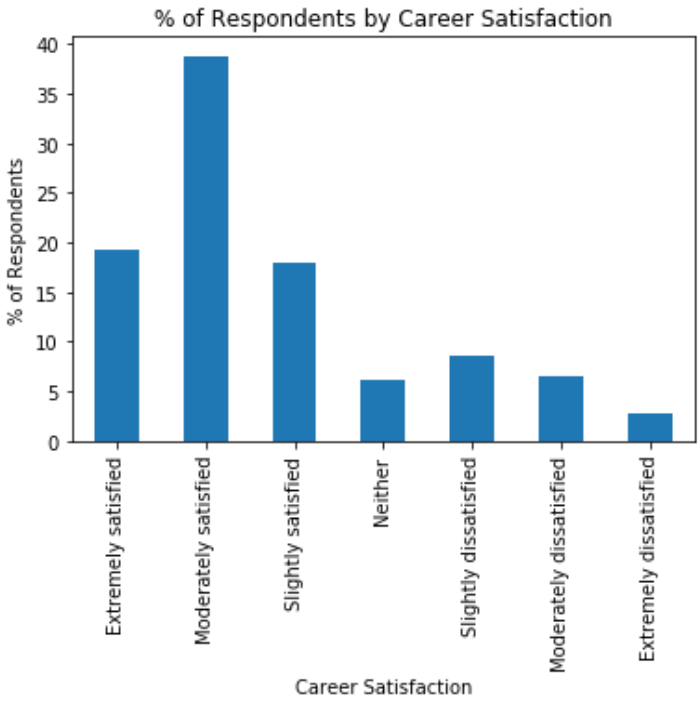

Figure 1: Percent of respondents by Career Satisfaction

Next, I wanted to see which countries have the happiest developers and which ones have the least happy ones. If we take a look at the most satisfied countries, we can see that most of them are European/North American countries.

Country

A lot of the happiest countries are Western Democracies and Scandinavian countries. A lot of the countries which exercise the most also tend to be happier as we'll see when we look at hours of exercise by country.

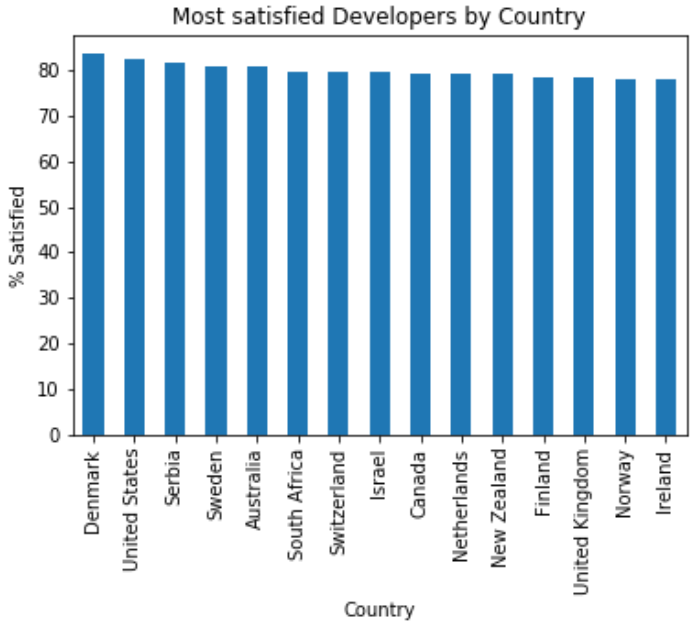

Figure 2: Most Satisfied Developers by Country

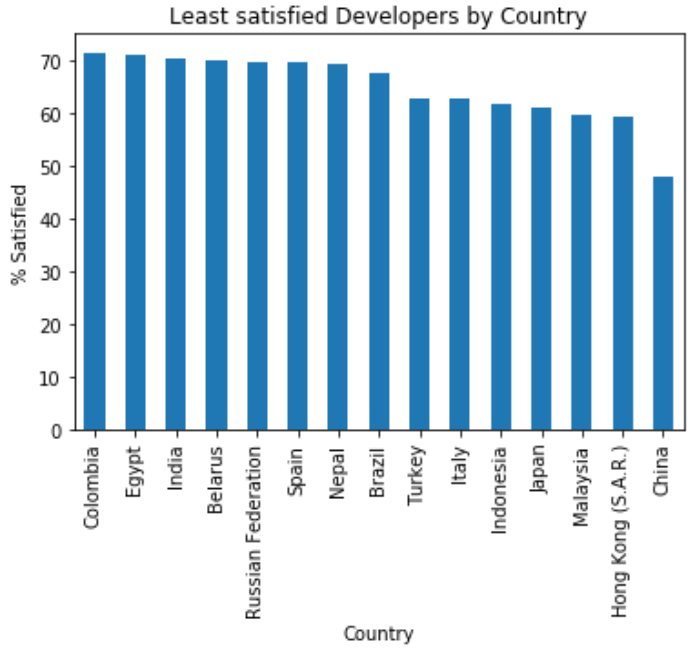

Figure 3: Most Satisfied Developers by Country

Exercise

Moderate exercise about 3-4 times a week seems to be optimum for overall satisfaction. This has also been proven by a lot of research. The developers who don't exercise are the least satisfied.

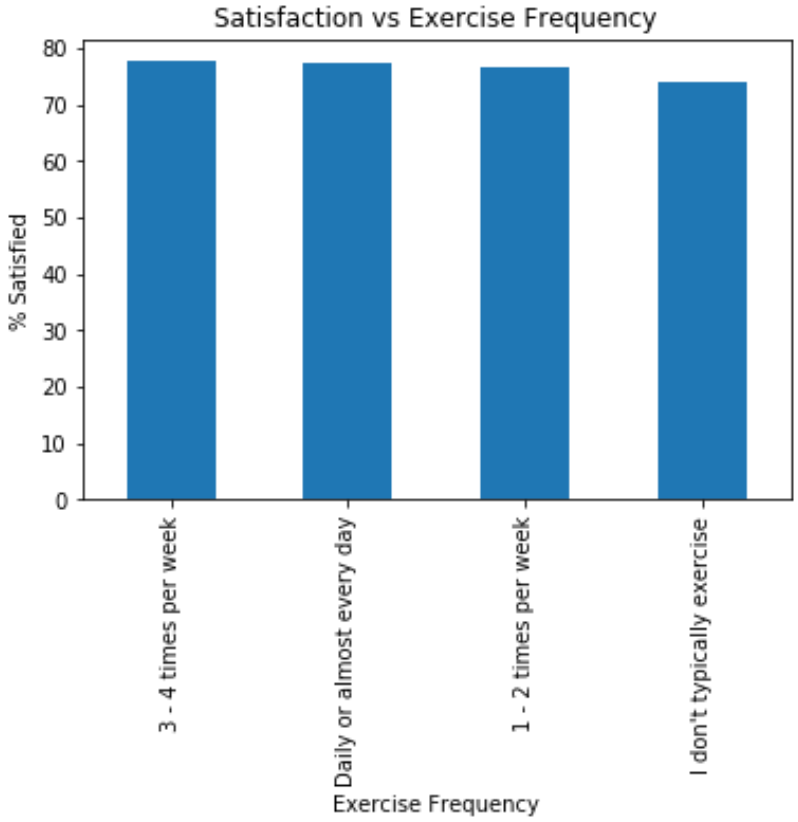

Figure 4: Satisfaction vs Exercise Frequency

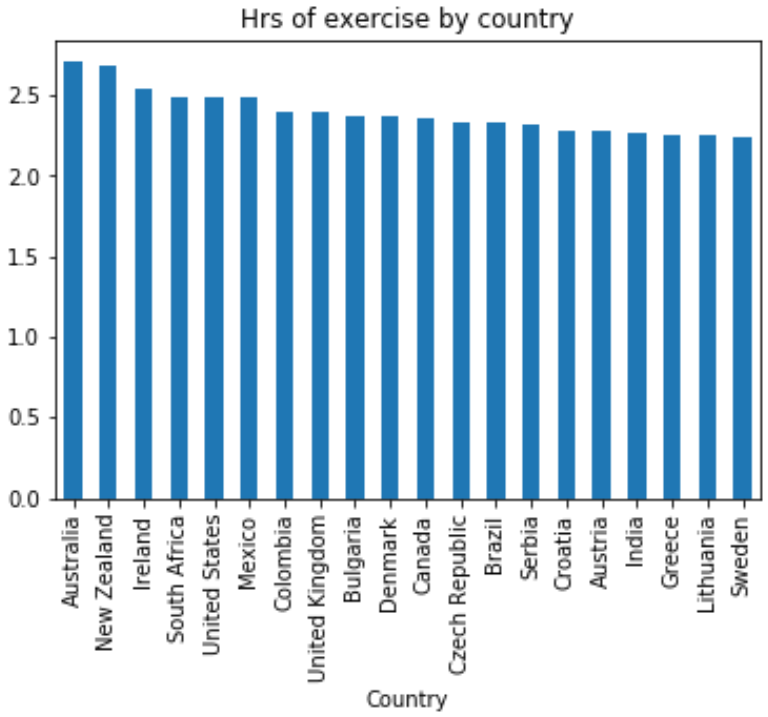

Figure 5: Exercise Frequency by Country

Most of the countries that exercise more also have a higher percentage of developers that are satisfied with their careers. Oceanic countries' developers exercise a lot more than others.

Undergrad Major

Only about 64% of current developers majored in computer science, computer engineering, or software engineering. About 27% of developers switched from other STEM fields, and the remaining 9% from non-STEM fields which shows that there aren't any significant barriers to entry if you want to become a developer, especially if you're from a STEM field. Coding bootcamps now-a-days offer a job guarantee as long as you have the basic skills (which you can learn easily) and are able to apply yourself during the bootcamps. Below is a breakdown of developers by Undergrad Major.

	Percentage
Computer science, computer engineering, or software engineering	63.7
Another engineering discipline (ex. civil, electrical, mechanical)	8.8
Information systems, information technology, or system administration	8.2
A natural science (ex. biology, chemistry, physics)	3.9
Mathematics or statistics	3.6
Web development or web design	3.1
A business discipline (ex. accounting, finance, marketing)	2.4
A humanities discipline (ex. literature, history, philosophy)	2.0
A social science (ex. anthropology, psychology, political science)	1.7
Fine arts or performing arts (ex. graphic design, music, studio art)	1.4
I never declared a major	0.9
A health science (ex. nursing, pharmacy, radiology)	0.3

Table 3: Developers by Undergrad Major

Coding as a hobby

We can see a clear relationship below. Developers who code as a hobby are more likely to be satisfied with their career. However, it's interesting to see that people who are extremely dissatisfied are more likely to code as a hobby than those who are moderately or slightly dissatisfied.

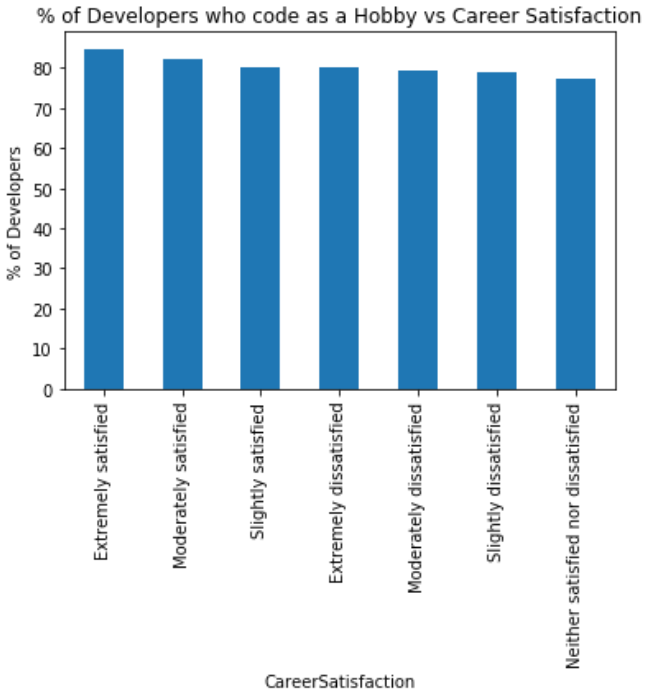

Figure 6: Hobby vs Career Satisfaction

Company Size

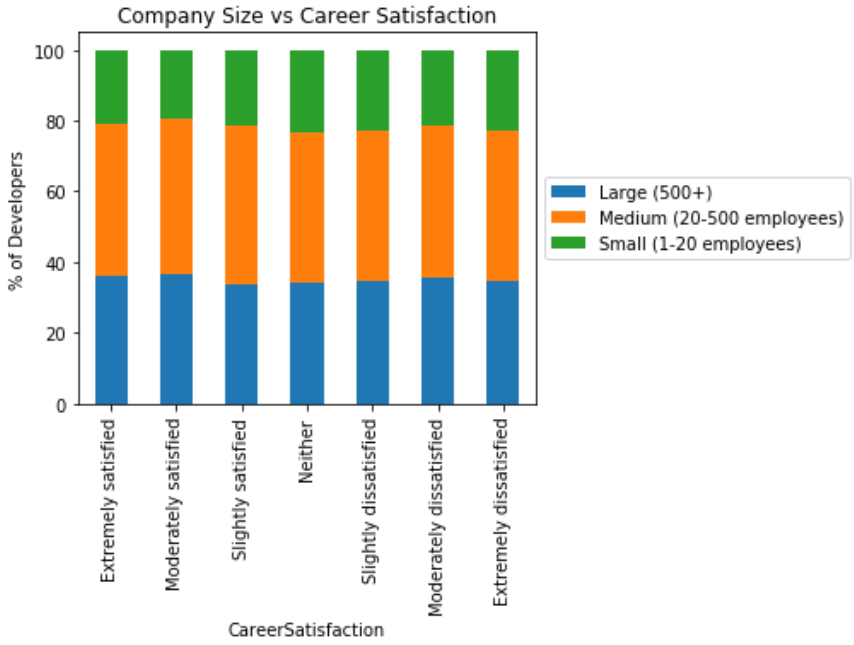

Figure 7: Company Size vs Career Satisfaction

Visually, company size doesn't seem to play a role in career satisfaction.

Company Size	% Satisfied	% Ext. Satisfied	% Mod. Satisfied	% Sl. Satisfied	% Dissatisfied
Large (500+)	43.390481	14.775412	14.928295	13.686774	42.777432
Medium (20-500 employees)	43.634357	14.151735	14.542219	14.940403	42.304536
Small (1-20 employees)	40.464418	13.765189	12.747640	13.951589	44.077975

Table 4: Company Size vs Career Satisfaction

The table confirms that Figure 7 doesn't show any conclusive evidence that company size doesn't seem to affect an employee's career satisfaction.

Salary

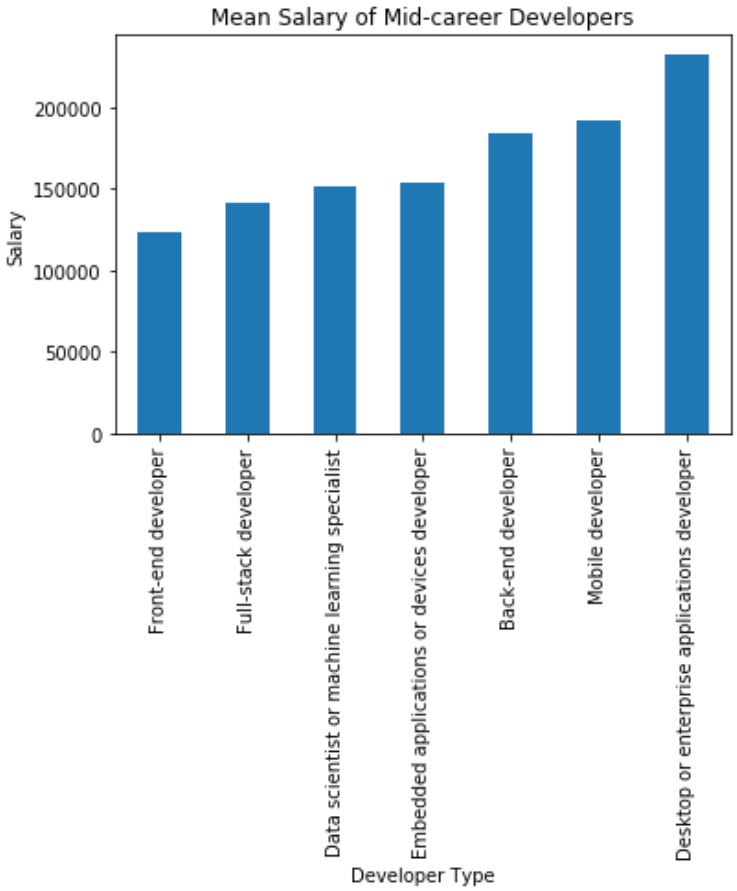

Figure 8: Mean Salary of Mid-career Developers

Since salary varies a lot depending on country, I only chose respondents from the United States to avoid any bias/skew in the data. Desktop/enterprise applications developer are the highest paid while Front end developers come in last.

Views on AI's future

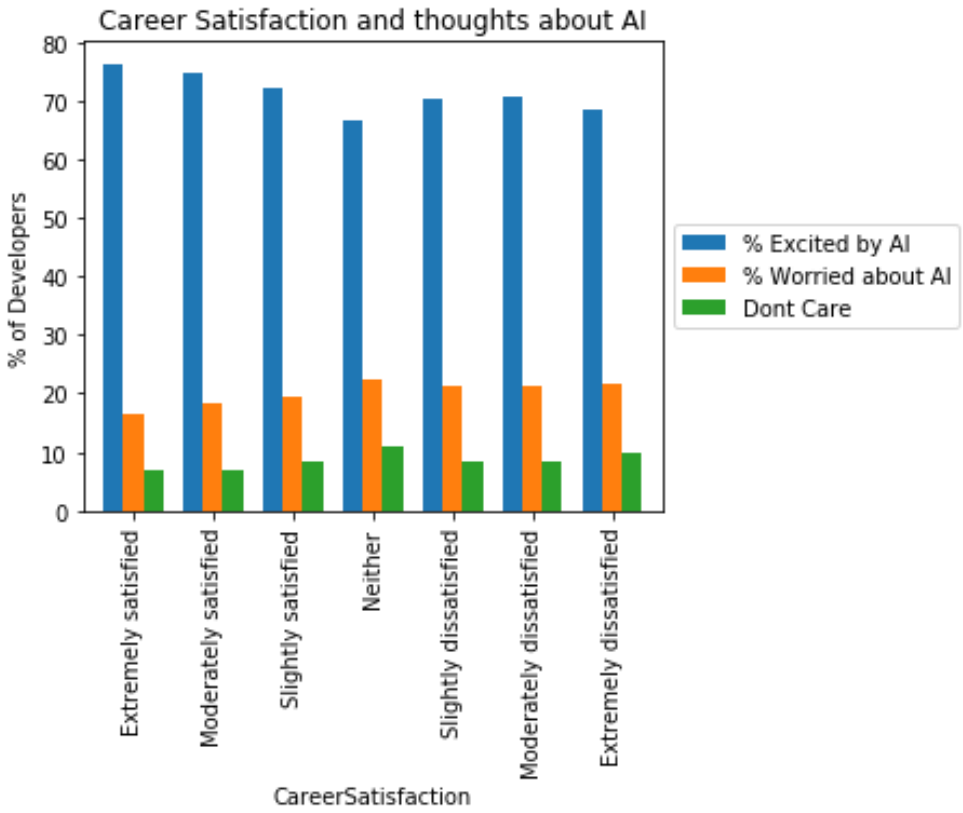

Figure 9: Career Satisfaction vs Thoughts about AI

People with higher career satisfaction also tend to be more excited about possibilities of AI, and those who are dissatisfied tend to be more worried by it. Those who are neither satisfied nor dissatisfied are least excited by AI.

Considering Ethics at work

People who are more satisfied with their career tend to consider ethical implications of their code more than those who aren't.

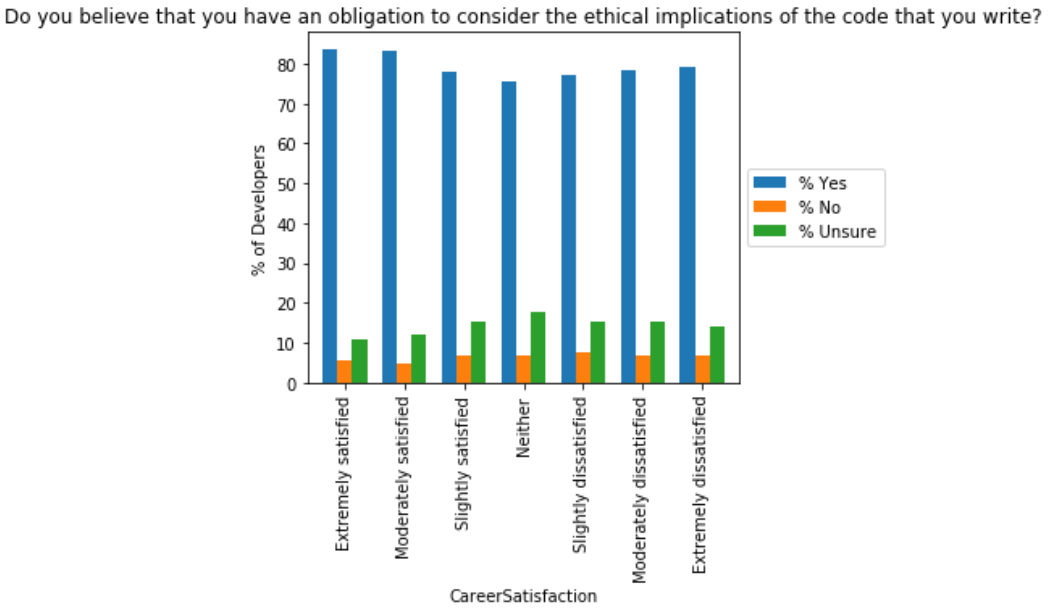

Figure 10: Considering Ethical Implications a Work

Calling out unethical code

People who are either extremely satisfied or dissatisfied are more likely to publicly call out unethical code. But people who are satisfied tend to call out unethical code within their company more often than those who aren't.

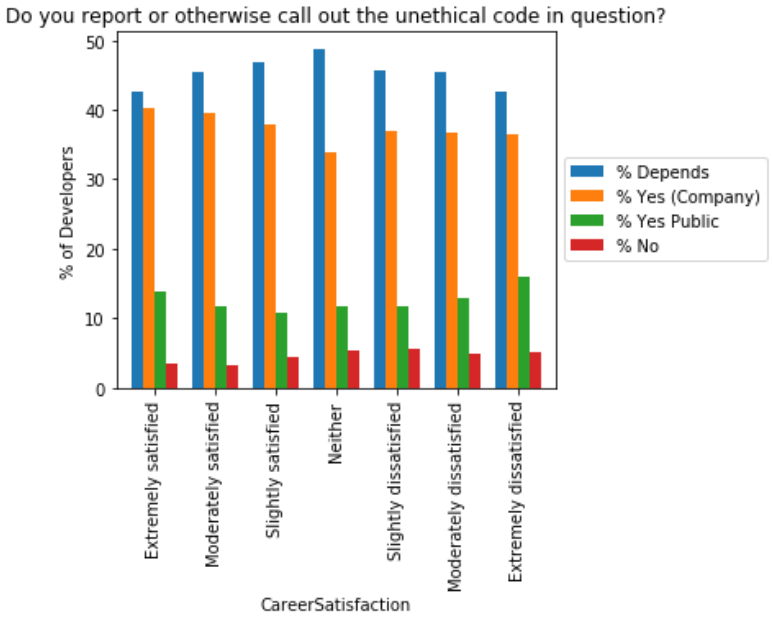

Figure 11: Calling out unethical code

Developer Type

Among non-managerial roles, data scientist or ML specialists are the most satisfied with their jobs whereas data/business analysts are the least satisfied. This is interesting because to many people, they're very similar jobs, but in fact they work with data very differently and do very different things with it.

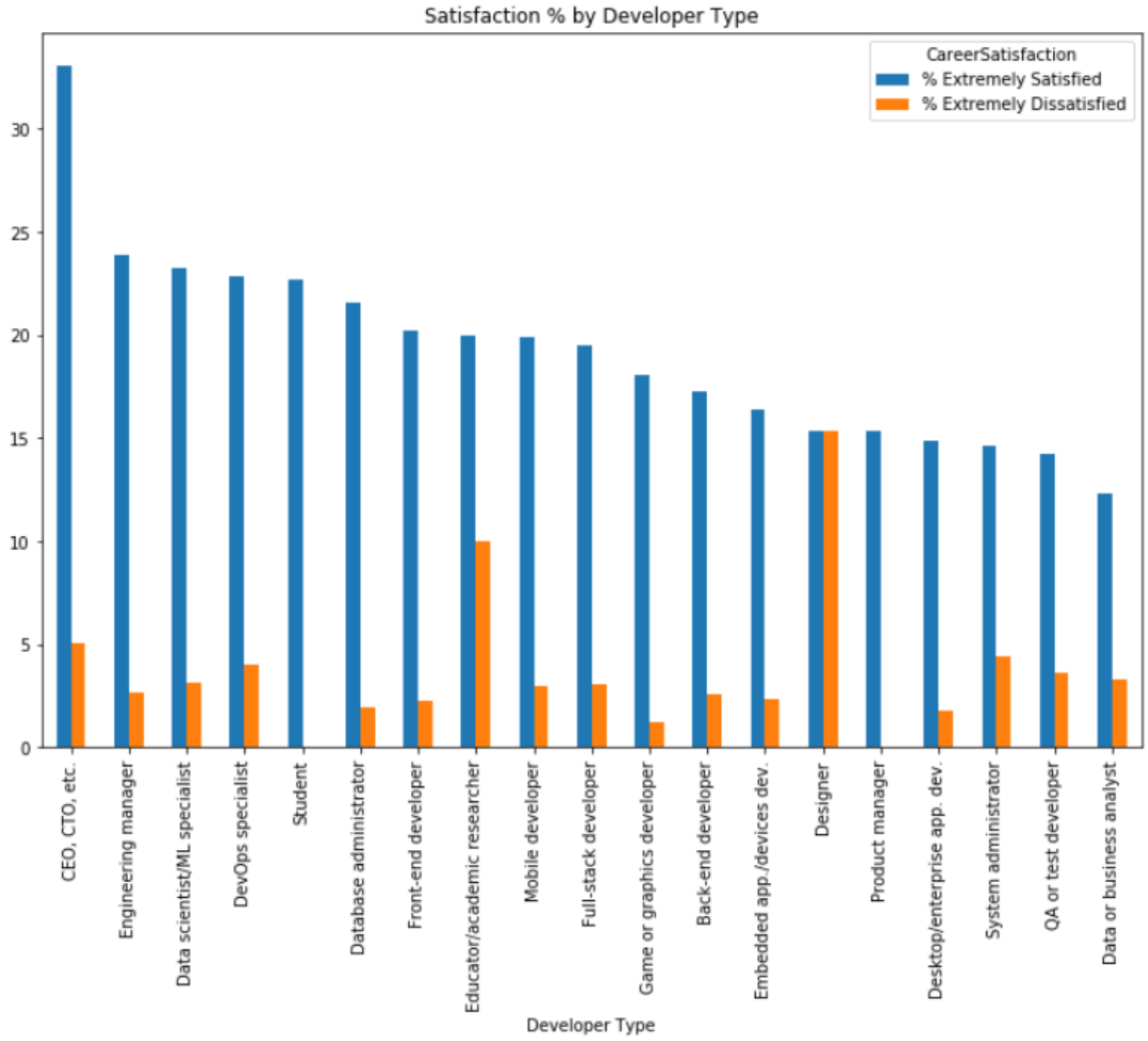

Figure 12 Satisfaction by Developer Type

Feature Selection

To see which features correlate strongly with career satisfaction, I created a correlation matrix with all the features. None of the features seemed to be correlated strongly. The only 2

features which are at least slightly correlated (0.22) are "Open source" and "Hobby". This makes sense because people who program as a hobby are also more likely to contribute to open source projects as a part of that hobby.

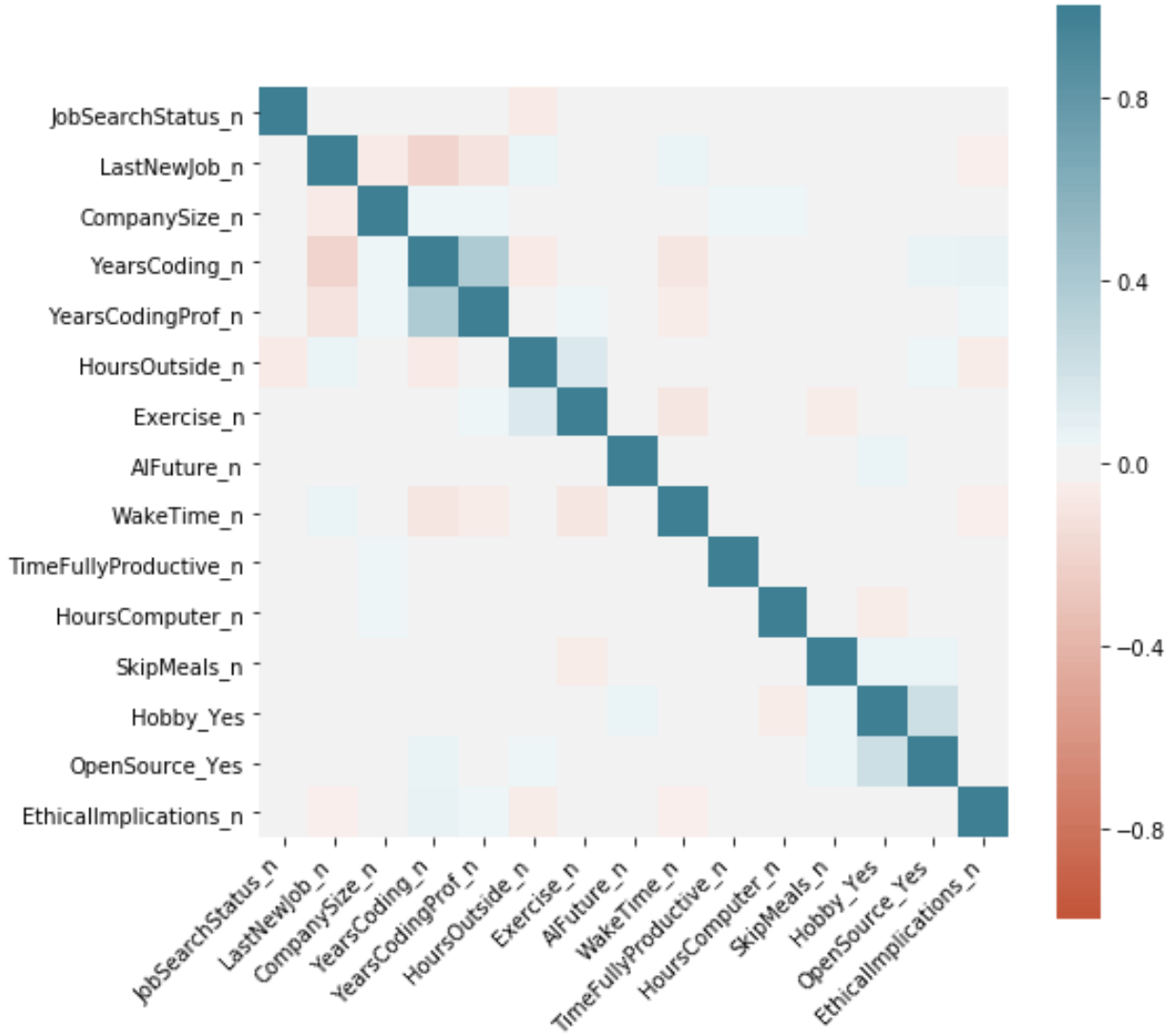

Figure 13: Correlation Matrix

I then used the feature_selection method in sklearn to find which features were the most significant when it came to career satisfaction and dissatisfaction. Below are the features most significant towards career satisfaction:

1. Job search status

2. Contributing to open source projects

3. Switching from another career to CS

4. Working as an engineering manager or other functional manager

5. Being a student

6. Years of coding

7. Being a CEO, CTO, etc.

8. Coding as a hobby

9. Having a Professional degree (JD, MD, etc.)

10. Being a designer

Machine Learning

After trying multiple ML algorithms and after tuning their hyperparameters, we can see that they are not able to predict the target variable. There could be multiple reasons for this. Firstly, there wasn't enough data (only 5,000 data points). Secondly, job satisfaction depends on a lot of factors that interact in complex ways that the survey probably couldn't capture. Most importantly, we could figure out the most influential features and hence the client can use this information to increase their employee satisfaction.