

A Network tour of Data Science

Project proposal

Romain Choukroun / SCIPER: 203917

Matthias Leroy / SCIPER: 213994

Alain Milliet / SCIPER: 228408

November 28, 2017

1 Questions

Wouldn't it be interesting to know what is the "hottest" field of computer science to work in right now? Is there an "average" type of developer? Do you pronounce "Gif" or "Jif"? What are the preferred operating systems by type of job?

2 Dataset

The dataset provided by StackOverflow on Kaggle seems to be a great start. It contains about fifty thousand answers from a sample of the active StackOverflow population about a lot of questions, namely 154. This means that we would have a tremendous insight into what makes a programmer unique, but also can help us to answer a lot of interesting questions.

3 Project

For this project we will focus on the salary side of the job, looking at all of the variables to determine which are the most impactful in terms of salary, but also which ones we can have an effect on in order to earn more or to achieve maybe another goal, such as a higher rank in our company.

Considering the amount of information contained in the dataset, we are highly open to suggestions as to what we could complement our analysis with. All the tools learned in the class can help us understand the data but also maybe derive a more important goal as to what we should analyze later on. In fact out of the 50000 people, about 12000 answered the question about the salary, which is to be expected. But in fact there are more axes to explore, such as "What makes people happy with their current position?", and much more.

We plan on fixing the goal after taking a hard look at the data, and by discussing with you guys.