



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

# **A Network Tour of Data Science**

Prof. Pierre Vanderghenst & Prof. Pascal Frossard

## A Stack Overflow recommender system

Romain Choukroun, Matthias Leroy, Alain Milliet, Hector Parmantier

January 24, 2018

# Outline

- Data exploration
  - Statistics
  - Correlations
- Recommender system
  - Students
  - Professionals

# Data exploration

- Survey
- 12'000 Professionals
- 3'000 Students
- A lot of features



Good for Data Science

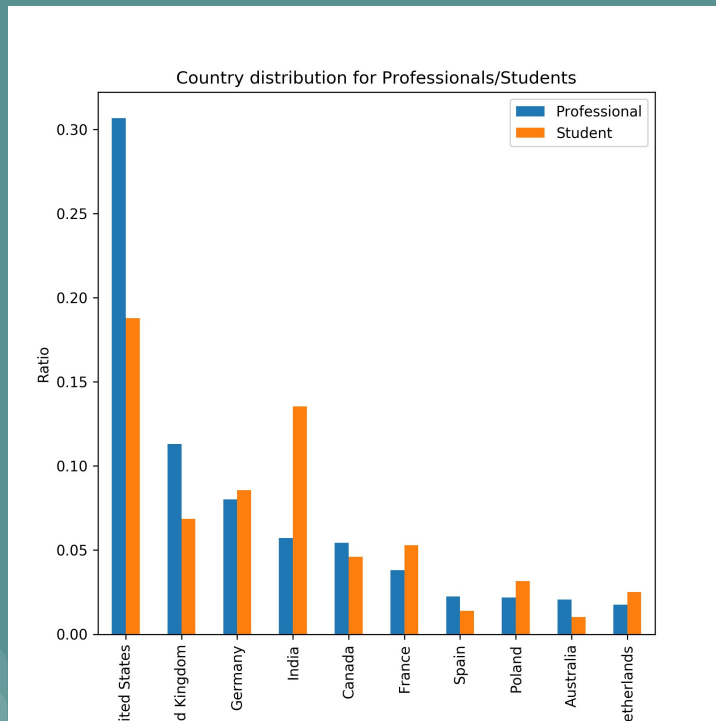
# Data exploration

- Basic statistics



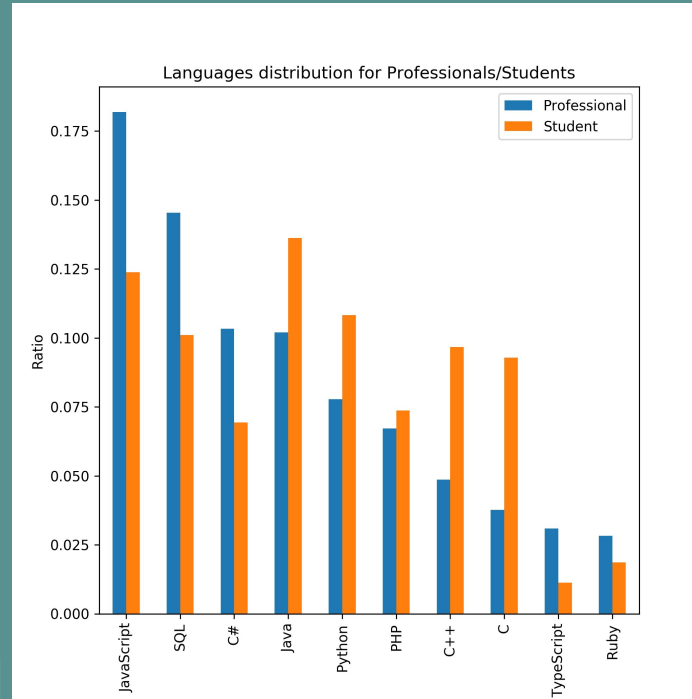
# Data exploration - Statistics

- Country



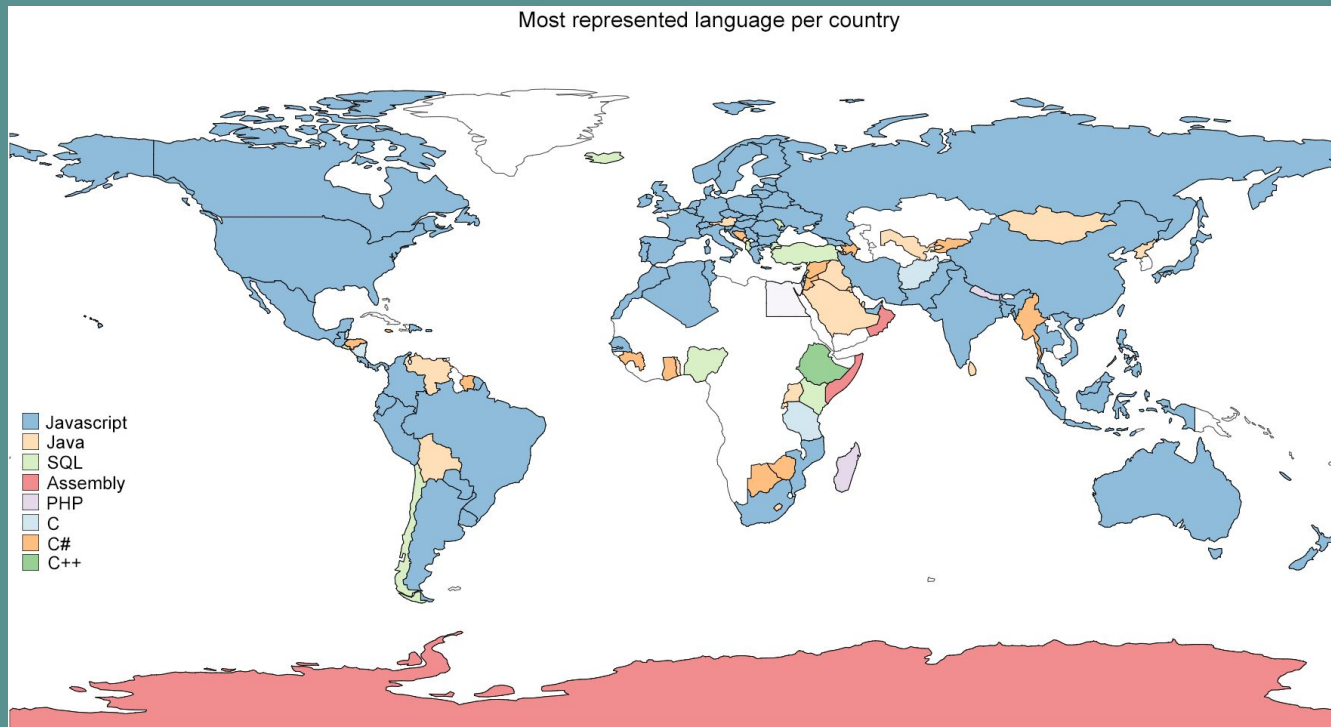
# Data exploration - Statistics

- Programming languages



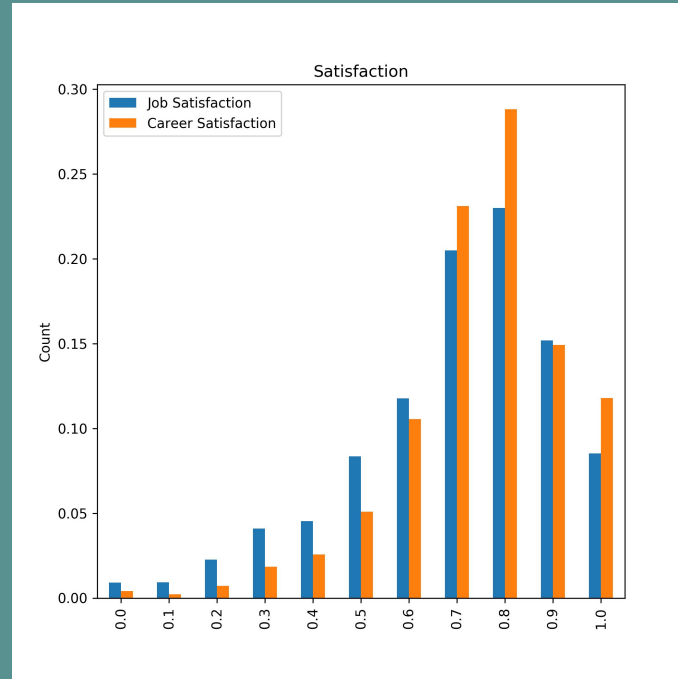
# Data exploration - Statistics

- Programming language per country



# Data exploration - Statistics

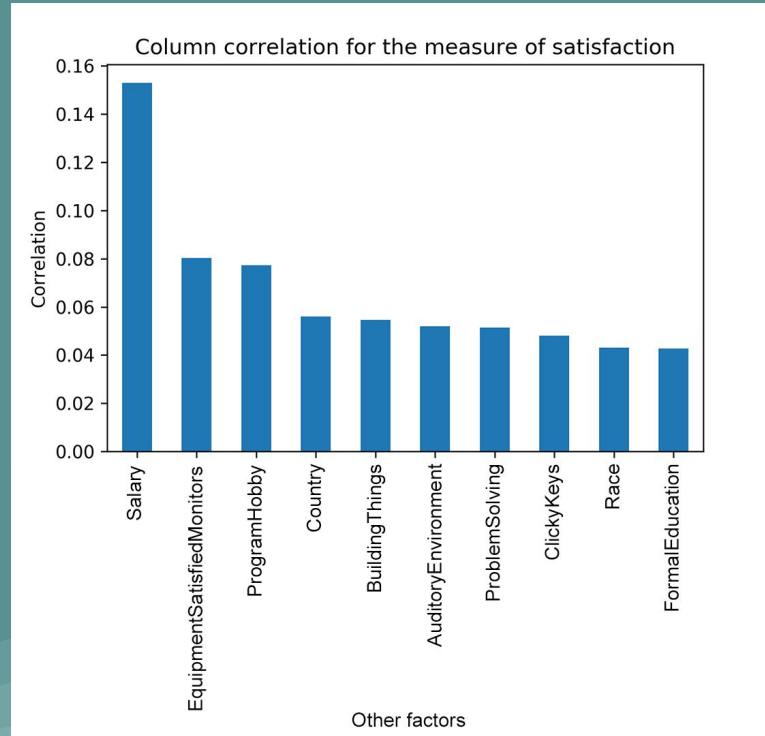
- Job satisfaction





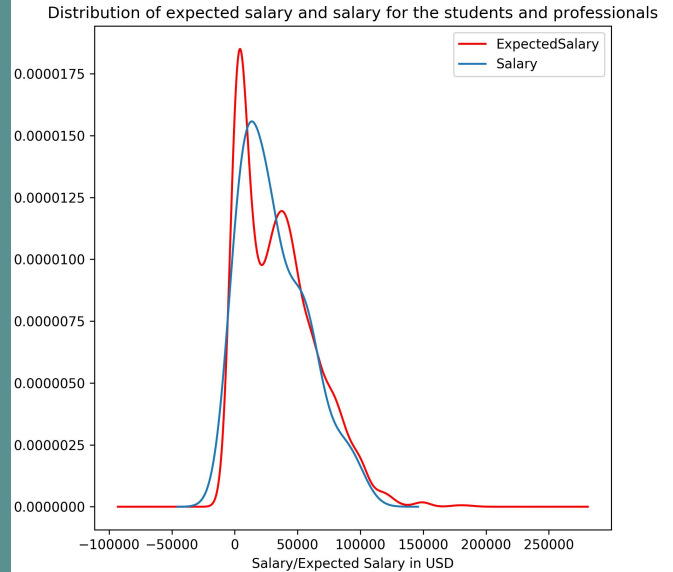
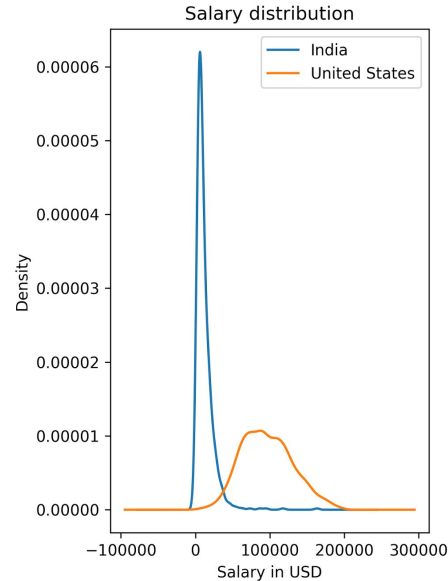
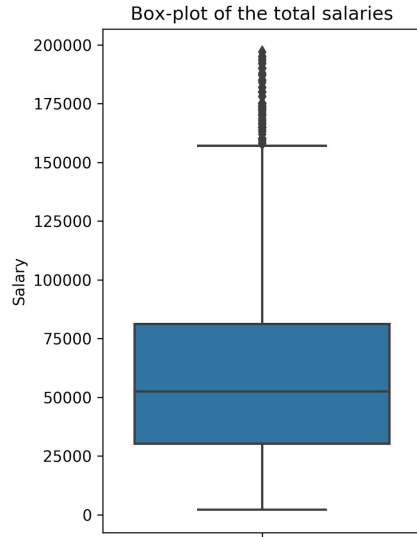
# Data exploration - Happiness

- What affects job satisfaction ?



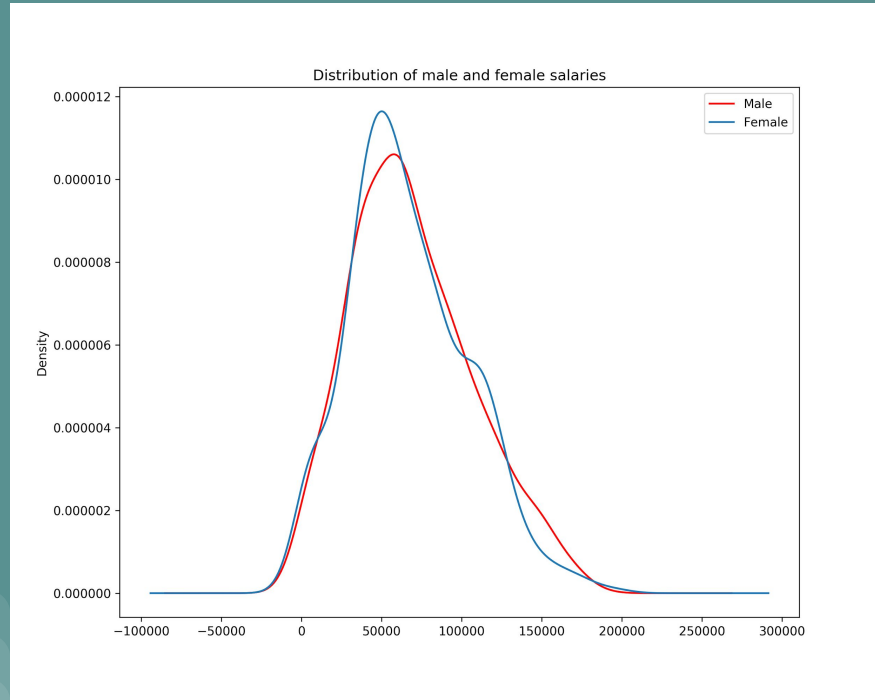
# Data exploration - Statistics

- Salary



# Data exploration - Gender gap

- Does the gender influences salary?



# Recommender system - Overview

- Recruitment tool:
  - define technical requirements
  - return top candidates match
- Graph-based systems:
  - student/professional distinction
  - profile-specific features

# Recommender system - Graph construction

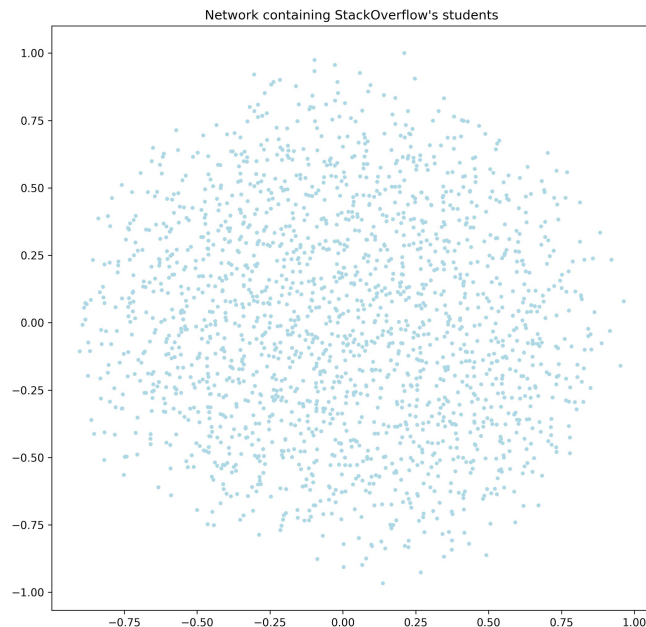
- Candidates as nodes
- Adjacency matrix:
  - Connect nearest neighbors
  - From distance to similarity (rbf kernel)

# Recommender system - Graphs

- Network with professionals

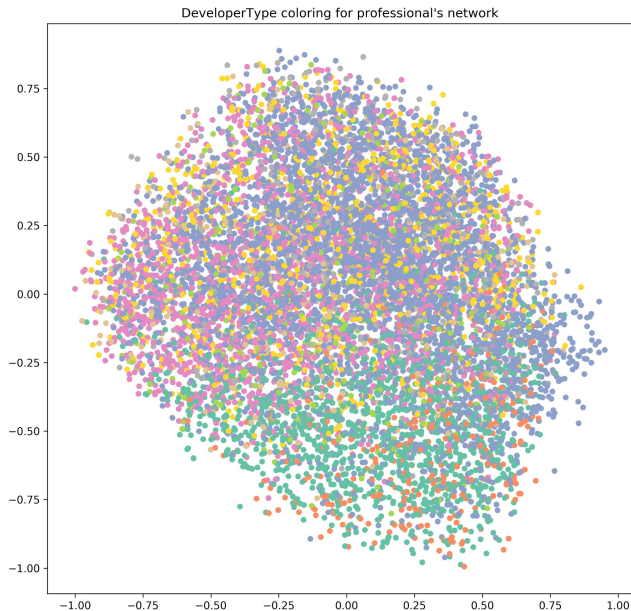


- Network with students



# Recommender system - Feature Visualization

- Distribution
- Emphasize clusters
- Target features examples:
  - DeveloperType
  - FormalEducation
  - WantWorkedLanguages
  - CompanyType



# Recommender system - Query example

- Do you program as a hobby or contribute to open source projects?: **Yes, both**
- In which country do you currently live?: **France**
- Are you currently enrolled in a formal, degree-granting college or university program?: **No**
- Which of the following best describes the highest level of formal education that you've completed?: **Secondary school**
- How long has it been since you first learned how to program?: **2 to 3 years**
- Suppose you could choose your own working hours for an 8-hour day. What time would you start work for the day? Please adjust the slider to the hour nearest your ideal start time. The box next to the slider will display your selection using a 24-hour clock: **10:00 AM**
- Which of the following languages have you done extensive development work in over the past year, and which do you want to work in over the next year?: **C#, Java**
- Which of the following languages have you done extensive development work in over the past year, and which do you want to work in over the next year?: **C++, Python**
- Suppose you're about to start a few hours of coding and have complete control over your auditory environment (music, background noise, etc.). What would you do?: **Turn on some music**

Respondent	ProgramHobby	Country	University	FormalEducation	YearsProgram	WorkStart	HaveWorkedLanguage	WantWorkLanguage	AuditoryEnvironn
51377	Yes, I program as a hobby	Germany	Yes, full-time	Secondary school	6 to 7 years	6:00 AM	C; C#; Java; Python; SQL; Swift	C++; Go; Java; JavaScript; Python; SQL; Swift	Turn on some m
50258	Yes, I program as a hobby	United Kingdom	No	Secondary school	2 to 3 years	10:00 AM	Python	Java	Turn on some m
28610	Yes, both	Germany	No	Secondary school	1 to 2 years	10:00 AM	JavaScript	JavaScript	Turn on some m
33330	Yes, I program as a hobby	Canada	No	Secondary school	2 to 3 years	10:00 AM	C; C++	C; C++	Turn on some m
1397	Yes, I program as a hobby	Spain	No	Secondary school	2 to 3 years	10:00 AM	C; C++; C#	Assembly	Turn on some m
27641	Yes, I program as a hobby	China	No	Secondary school	2 to 3 years	9:00 AM	Python	C#; Python	Turn on some m



# Conclusion

- Broad understanding of developer population
- Useful tool for recruiter

## Future work

- Try different distances between entities

# Thanks