What Twitter Says About Swiss Happiness

Nooshin Mirzadeh Amir Shaikhha Asmae Tounsi
EPFL

Happiness is the Key

Happier environment → More successes

Happy people*:

- are less likely to get sick
- make more money
- are more productive at work
- are more generous
- are more creative

How to raise happiness in our environment?

- First step: Identify the happiness situation

→ Measure happiness in our environment

→ Social network data helps us for the first estimation

* The Greater Good Science Center at the UC Berkeley

People express their feelings in the social networks

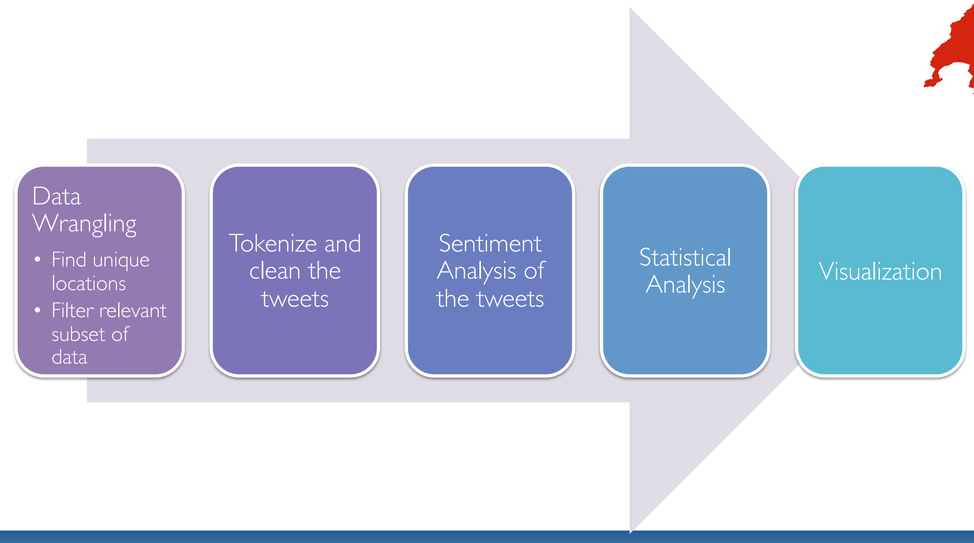
Use Social Network Data to Identify Happiness

Twitter is an appropriate social network for mining the happiness!

- Tweets are short and usually represents users' feelings

Analyze the tweets in Switzerland's cantons over time

- How locations and time affect the happiness?



Build a system to analyze the tweets in Switzerland in a time window

Data Wrangling and Sentiment Analysis

Divide Switzerland to its cantons, and analyze tweets in the cantons

- First challenge: Big Data! We need to use large-scale platforms, e.g., Spark

Challenges in extracting the locations provided by users

- There is no geo-located information → need to identify cantons based on names
- There are more than 21000 unique locations!
- Some locations are too general or meaningless, e.g., Switzerland, From Everywhere!
- Some locations include some random characters, e.g., i❤clivio, ★ * .* * ✵ !
- Some locations are in other languages, e.g., المملكة العربية السعودية, スイス

Challenges in sentiment analysis

- Switzerland is a multilingual country: three official languages
- Various people from other countries increase the available languages
- The available resources in non-English languages are limited
- Typos and multi-lingual tweets

Our approach:

Analyze three languages in *meaningful locations*: English, German, French

Raw data has its own challenges; we need to clean the available data

Statistical Analysis: Can We Rely on the Data?

Compare each population of tweets with all the data

- Group the data by cantons

For each Single day

- Compare the tweets of each canton to the tweets of all the day
- Not all the cantons are represented in a single day
- Some resulting p values make no sense

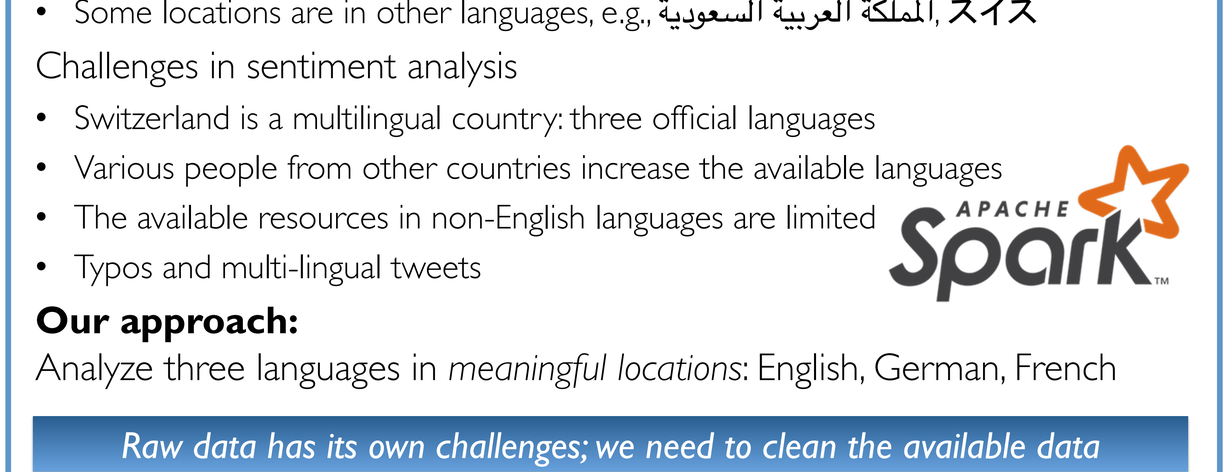
For each Month

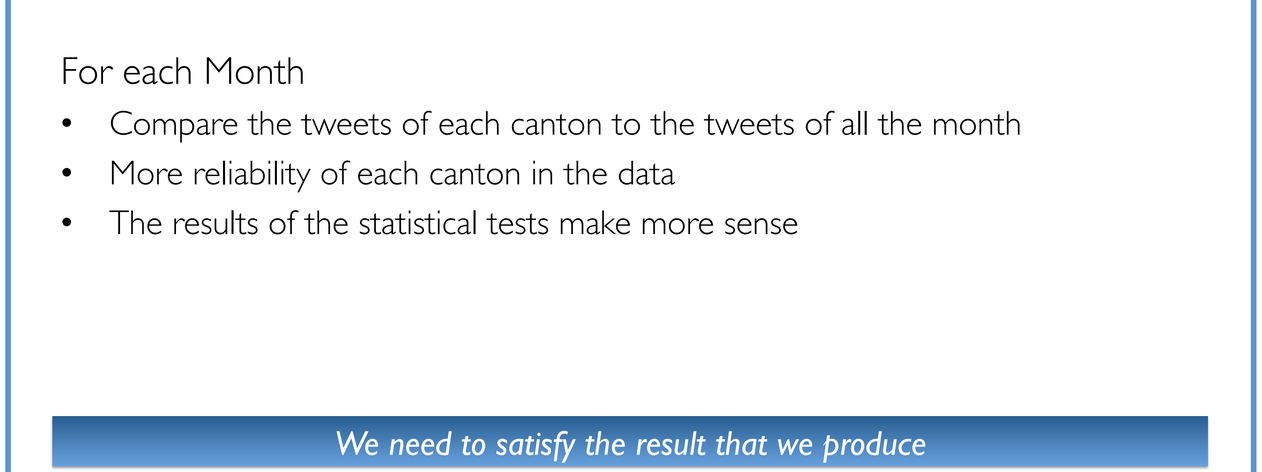
- Compare the tweets of each canton to the tweets of all the month
- More reliability of each canton in the data
- The results of the statistical tests make more sense

We need to satisfy the result that we produce

Visualization

- Interactive Switzerland map showing the happiness of each canton per day
- Using p-value to hide the happiness information of statistically insignificant cantons
- More detailed information of each canton
- Number of tweets per language per canton
- Aggregating the information for each month





ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE