Software Engineer Intern - Fuzzy Join

PARIS, FRANCE . ENGINEERING – INTERNSHIP

Dataiku's mission is big: to enable all people throughout companies around the world to use data by removing friction surrounding data access, cleaning, modeling, deployment, and more. But it's not just about technology and processes; at Dataiku, we also believe that people (including our people!) are a critical piece of the equation.

Today Dataiku Data Science Studio allows users to analyze & transform large data sets using visual recipes (prepare, filter, join, stack, …). Currently users can perform joins between 2 datasets, but the operation only joins rows where values for a given column are identical in both datasets.

We are looking for a **software engineer intern** for a 4 to 6 months internship in our Paris office. The **objective of the internship** is to improve this join recipe to perform fuzzy joins, that is to perform joins matching not just on values identical on both sides, but on rows roughly equivalent on both sides.
- We could consider that 2 numeric values match if thy are equal within some tolerance (+/- 0.5 for example)
- We could consider that 2 strings match if their Levenshtein distance is less than 5
- We could consider that 2 rows match if their longitude & latitude represents positions that are within 10 meters
- ...

City	Population
Paris	2 420 069
Saint-Étienne	172 023
Aix-en-Provence	141 545

City	Area (km^2)
Paris	105,40
Saint etienne	19,97
Aix	186,08
Le Havre	46,95

City	Population	Area (km^2)
Paris	2 420 069	105,40
Saint-Étienne	172 023	19,97
Aix-en-Provence	141 545	186,08
Le Havre		46,95

```
-----------------------------                -----------------------------
| City              | Population |           | City          | Area (km^2) |
|-------------------|------------|           |---------------|-------------|
| Paris             |  2 420 069 |           | Paris         |      105,40 |
| Aix-en-Provence   |    141 545 |           | Saint etienne |       19,97 |
| Saint-Étienne     |    172 023 |           | Aix           |      186,08 |
-----------------------------                | Le Havre      |       46,95 |
                                             -----------------------------

        ------------------------------------------------
        | City              | Population | Area (km^2) |
        |-------------------|------------|-------------|
        | Paris             |  2 420 069 |      105,40 |
        | Saint-Étienne     |    172 023 |       19,97 |
        | Aix-en-Provence   |    141 545 |      186,08 |
        | Le Havre          |            |       46,95 |
        ------------------------------------------------
```

You are the ideal recruit if you:

- are **autonomous**, to drive your subject.
- have **no fear** to dive deep into source code.
- are **eager to learn** new things.
- have a **good knowledge of a programming language** (Python, Java, C#, Javascript, R, Ruby, <insert your favorite language here>).
- have a basic knowledge of **Web development**

Benefits:

As an intern, you'll join the **R&D team of a startup**, composed of 30 engineers passionate about software development and developing an amazing product. Yes, we love our product. No, we are not biased ;-).

We have **great office space in Paris** (Gare de Lyon) with free breakfast, coffee, video games, yoga and much more.

The monthly remuneration will be:

- 1000 euros for students in 4th year of university / engineering school.
- 1400 euros for students in 5th years of university / engineering school.

Remuneration additionally includes 50% reimbursement of public transit pass.

A full-time contract position might be offered at the end of the internship.