Employee Retention

Goal

Employee turn-over is a very costly problem for companies. The cost of replacing an employee if often larger than 100K USD, taking into account the time spent to interview and find a replacement, placement fees, sign-on bonuses and the loss of productivity for several months.

It is only natural then that data science has started being applied to this area. Understanding why and when employees are most likely to leave can lead to actions to improve employee retention as well as planning new hiring in advance. This application of DS is sometimes called *people analytics* or *people data science* (if you see a job title: people data scientist, this is your job).

In this challenge, you have a data set with info about the employees and have to predict when employees are going to quit by understanding the main drivers of employee churn.

Challenge Description

We got employee data from a few companies. We have data about all employees who joined from 2011/01/24 to 2015/12/13. For each employee, we also know if they are still at the company as of 2015/12/13 or they have quit. Beside that, we have general info about the employee, such as avg salary during her tenure, dept, and yrs of experience.

As said above, the goal is to predict employee retention and understand its main drivers. Specifically, you should:

- Assume, for each company, that the headcount starts from zero on 2011/01/23. Estimate employee headcount, for each company, on each day, from 2011/01/24 to 2015/12/13. That is, if by 2012/03/02 2000 people have joined company 1 and 1000 of them have already quit, then company headcount on 2012/03/02 for company 1 would be 1000. **You should create a table with 3 columns: day, employee_headcount, company_id.**
- What are the main factors that drive employee churn? Do they make sense? Explain your findings.
- If you could add to this data set just one variable that could help explain employee churn, what would that be?

Data

We have 1 table downloadable by clicking **here**.

The table is:

```
"employee_retention" - comprehensive information about employees.
```

Columns:

- **employee_id** : id of the employee. Unique by employee per company
- **company_id** : company id.
- **dept** : employee dept
- **seniority** : number of yrs of work experience when hired
- **salary**: avg yearly salary of the employee during her tenure within the company
- **join_date**: when the employee joined the company, it can only be between 2011/01/24 and 2015/12/13
- **quit_date**: when the employee left her job (if she is still employed as of 2015/12/13, this field is NA)

Example

```
    Let's now check the characteristics of the employee in first row.
```

head(employee_retention, 1)

Field	Value	Description
employee_id	13201	unique identifier of the employee
company_id	7	she works for company 7
dept	customer_service	she works in the customer service dept
seniority	28	she had 28 yrs of work experience when she was hired by company 7
salary	89000	her yearly salary is 89K USD
join_date	2014-03-24	she joined company 7 on March, 24 2014
quit_date	2015-10-30	she quit her job at company 7 on Oct, 30 2015. That is, she worked there for ~19 months