**Employee Retention**

**Goal**

Employee turnover 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.

t 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

## Hints:

What are the main factors that drive employee churn? Do they make sense? Explain your findings.

What might you be able to do for the company to address employee Churn, what would be follow-up actions?

If you could add to this data set just one variable that could help explain employee churn, what would that be?

Your output should be in the form a a jupyter notebook and pdf output of a jupyter notebook in which you specify your results and how you got them.

**Data**

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 o 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

13