Instructions for assignment

Please use the link to download the data set employees-turnover.csv:

<https://drive.google.com/drive/folders/1yCqG4x3z6SXttr1esje3-zFg9QU7osJY>

Imagine, as an HR analyst you are trying to understand what the key drivers are for employee turnover in your organization. You want to ensure that you attract and retain your top talent and take remedial action if it looks like you’re going to lose good people

You have been given a data set about all your past and current employees, so you have access to some demographic information like age, gender and marital status along with their compensation and performance related details. But the field that you should be most interested about is attrition which mean whether they’re a current employee or they’ve left.

instruction:

1. Import the data in your **jupyter notebook environment.**
2. Perform preprocessing on the data
3. From the past year you have noticed that many of your employees are leaving the company, so your main objective is to find out the factors which are responsible for this employee turnover. Do the **exploratory data analysis** of the data and find some of the useful insights from the data

(minimum **four insights** from the data)

1. Indicate your insights using the **python visualization tools** (matplotlib, seaborn)
2. Build a ML model which can predict whether an employee is going to leave the company in future or not by using **logistic regression, Decision tree, KNN, SVM and random forest.**

**Note – You have to submit your assignment as jupyter notebook file(.ipynb) in the google form below.**

**Last date for submission of assignment is 18 july.**

Link for submission:

<https://docs.google.com/forms/d/e/1FAIpQLSegvhqazJwMJwYfjFgU1kw-qW4bp9XXC7IgUDkWwBF2DXiuQg/viewform?usp=sf_link>

In case of any doubt please write in the discussion forum.