# COMPANY X EMPLOYEE ATTRITION PROBLEM

ANALYSIS, INTERPRETATION, PREDICTION

# Content

- Introduction
- Data Exploration
- Model Development
- Interpretation
- Prediction

# Introduction

Employees are the greatest assets of a company and Employee Attrition is a crucial problem that creates vacuum and poses human resources management challenges for the company. The goal of this project is to use Data Analysis techniques and Business intelligence tool to understand Company X Attrition case, to determine the reason why Employees are leaving, and predict the Employees that are prone to leave the Company next.

I shall use the Random Forest Classifier – a Machine Learning Algorithm to analyze the datasets. This ensemble learning combines multiple decision trees or more learning models to gain average and build better predictive result, rather than relying on one decision tree. Thus, I am enthusiasm about the accuracy of the outcome of the prediction.

# Data Exploration

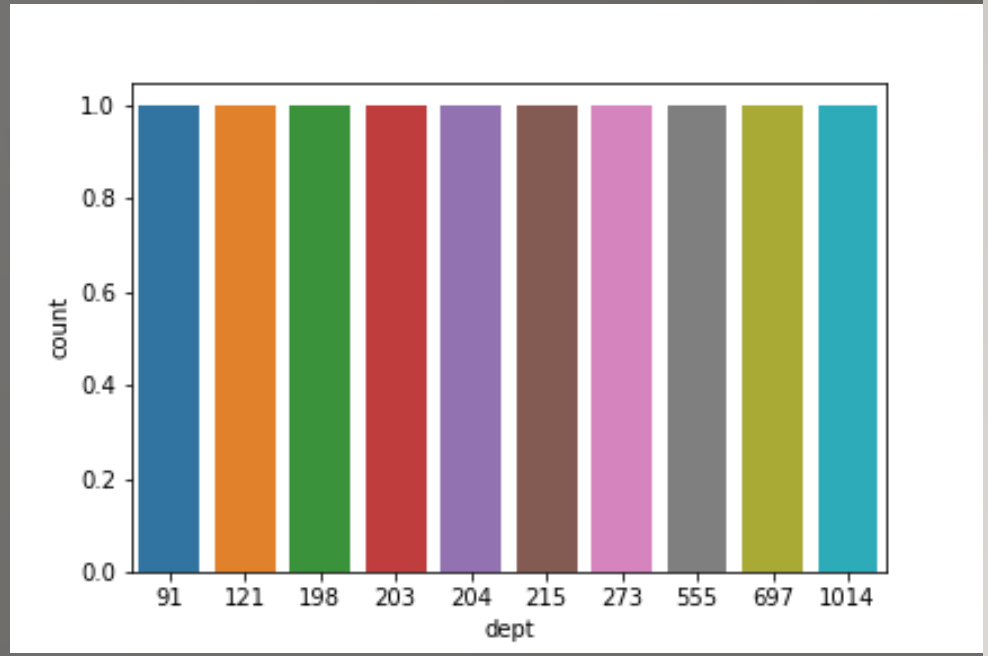
A critical look at our data set using pandas, numpy and scikit learn, our model clearly predict high employee dissatisfaction in some of the variables presented. According to our model, salary variable tends to have higher impact, and/or is a determinant to who leaves, why, and when the employee leaves Company X. Our model confirms accuracy of 98% . It also help to understand that great relationship exist between the datasets. Visualizing the dataset of those that already left the company help me to make informed predictions about the existing employee because of great similarity in the structure of the data. It is discover that majority of the ex-employees are from Sales, Support, and Technical department due to low salary .Lets gets to dive with Python…importing a few Libraries:

# Data Exploration

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Importing Dataset
ex_emp = pd.read_csv("Ex_Employees.csv")
print (ex_emp.head())
exi_emp = pd.read_csv("Employee_existing.csv")
print(exi_emp.head())
comb = ex_emp.append(exi_emp)
print(comb.head())
```

# Data Exploration

```
print(comb.drop("Emp ID", 1))
print(ex_emp.drop("Emp ID", 1))
print(exi_emp.drop("Emp ID", 1))
print(ex_emp['dept'].value_counts())
print(ex_emp['salary'].value_counts())
sns.countplot(ex_emp['dept'].value_counts())
plt.savefig('ex_emp_dept.png')
```



# Data Exploration

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Importing Dataset
ex_emp = pd.read_csv("Ex_Employees.csv")
print (ex_emp.head())
exi_emp = pd.read_csv("Employee_existing.csv")
print(exi_emp.head())
comb = ex_emp.append(exi_emp)
print(comb.head())
```