

# Minimum Viable Product (MVP)

- The goal of this project is to use data from the company to create a machine learning model to predicts the prospects of future and present employee.
- To start exploring this goal, I did some EDA for data like:

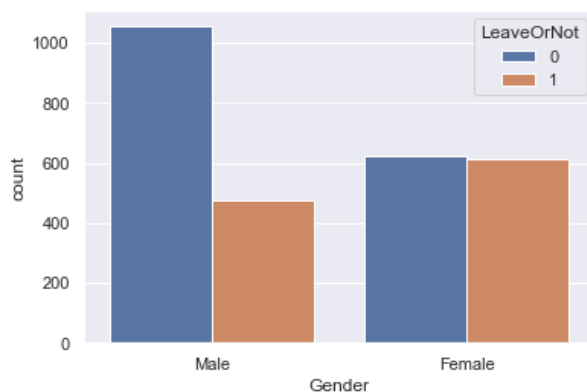
```
df.shape  
(4653, 9)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4653 entries, 0 to 4652  
Data columns (total 9 columns):  
#   Column                                Non-Null Count  Dtype  
---  ---                                -  
0   Education                             4653 non-null   object  
1   JoiningYear                           4653 non-null   int64  
2   City                                  4653 non-null   object  
3   PaymentTier                           4653 non-null   int64  
4   Age                                   4653 non-null   int64  
5   Gender                                4653 non-null   object  
6   EverBenched                           4653 non-null   object  
7   ExperienceInCurrentDomain              4653 non-null   int64  
8   LeaveOrNot                             4653 non-null   int64  
dtypes: int64(5), object(4)  
memory usage: 327.3+ KB
```

```
df.groupby(['Gender', 'LeaveOrNot'])['Gender'].count()
```

```
Gender  LeaveOrNot  
Female  0          621  
        1          614  
Male    0         1055  
        1          474  
Name: Gender, dtype: int64
```



```
df.groupby(['Education', 'LeaveOrNot'])['Education'].count()
```

```
Education  LeaveOrNot  
Bachelors  0          1232  
           1          739  
Masters    0          328  
           1          309  
PHD        0          116  
           1           40  
Name: Education, dtype: int64
```

