

Content

- Introduction
- Load packages
- Read the data
- Data Explortion
- Data Cleansing && Analysis

1.Selecting Dataset From Kaagle

Introduction

Question2

A company's want's to know and predict cusotmers would leave their company in next 2 years? . Here main Objective is to build predictive model that predicts the prospects of future and present employee

- Education - Education of the employee
- Joining-Year - Joining Year of the employee
- City - Working location of the employee
- Payment-Tier - Type of payments
- Age - Age of the employee
- Gender - Gender of the employee
- Ever-Benched - Whether the employee has ever been on bench-means he/she wasnt working on any project
- Experience-In-Current-Domain - How much experience an employee has in current domain?
- Leave-Or-Not - Which employee will leave the organizaion

Load Packages

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import statistics as st
```

Read the Data

```
# reading data
data=pd.read_csv('../input/employee-future-prediction/Employee.csv')

#printing first five rows in the dataset
data.head()
```

Question 3

Data Exploration

```
# knowing statistics of the data
# basic stats of features
data.describe().style.background_gradient(cmap="Pastell")

# shape of the data
data.shape
```

Observations

- The dataset which consists of 4653 rows and 9 columns

```
data.columns
```

```
# check type of data
data.info()
```

Question 4

```
# let's check null or missing values in the dataset
```

```
data.isnull().sum()
```

```
sns.heatmap(data.isnull())
```

Observations:

- From above we can observe that there is no null or missing values in the dataset
- It's good sign towards good results

```
#checking Cor-relation
```

```
corr = data.corr().abs()
```

```
mask = np.triu(np.ones_like(corr, dtype=np.bool))
```

```
fig, ax = plt.subplots(figsize=(30, 14))
```

```
# plot heatmap
```

```
sns.heatmap(corr, mask=mask, annot=True, fmt=".2f", cmap='coolwarm',
            cbar_kws={"shrink": .8}, vmin=0, vmax=1)
```

```
# yticks
```

```
plt.yticks(rotation=0)
plt.show()
```

Question 5

I am going to analyze LeaveOrNot feature. Because A company's want's to know and predict cusotmers would leave their company in next 2 years? . Here main Objective is to analyzing and knowing the prospects of future and present employee

And looking to know which things going to effect company towards leaving company

Question 6

```
size = data['Gender'].value_counts()
colors = ['green', 'orange']
labels = "Male", "Female"
explode = [0, 0.01]

my_circle = plt.Circle((0, 0), 0.7, color = 'white')

plt.rcParams['figure.figsize'] = (9, 9)
plt.pie(size, colors = colors, labels = labels, shadow = True, explode
= explode, autopct = '%.2f%%')
plt.title('Distribution of Gender', fontsize = 20)
p = plt.gcf()
p.gca().add_artist(my_circle)
plt.legend()
plt.show()
```

Observations:

- Percentage of Males who were working in company is 59.70%
- Percentage of Females who were working in company is 40.30%
- Overall Employees most of them who were working in company they mostly Males only

In Which City Most of the Employees Living?

```
print('Employees city count',data['City'].value_counts())

plt.figure(figsize=(15,6))
sns.countplot(x='City',data = data,palette='magma')
plt.title('Distribution of Employees Living City', fontsize = 20)
plt.show()
```

Observations:

- Employees who were working in company most of the employees i.e 2228 are living in Bangalore only

- Rest of the employees i.e 1268 members living in Pune and 1157 members are living in New Delhi
- Finally my analysis here is Most of the employees are working in company most of them living in Bangalore

Question 8

```
plt.figure(figsize=(15,6))
sns.countplot(x='Age',data = data, hue =
'LeaveOrNot',palette='autumn')
plt.title("Employees Age VS Company Leaving Status")
plt.show()
```

Observations:

- Employees who existing from company most of their age is in between 26-28
- And Most of the employees who working in company their age is also in between 26-28
- Very least number of employees are working in company their age in between 22-23 and employees exiting rate is too high when we compare to working employees

```
plt.figure(figsize=(15,6))
sns.countplot(x='PaymentTier',data = data, hue =
'LeaveOrNot',palette='autumn')
plt.title("Employees PaymentTier VS Company Leaving Status")
plt.show()
```

Observations:

- From above we can observe that most of the employees who were working they mostly payment tier 3 employees
- Very least number of employees are Payment Tier1 Employees.
- More number of employees are exiting company they mostly were Payment tier 3.

```
plt.figure(figsize=(15,6))
sns.countplot(x='Education',data = data, hue =
'LeaveOrNot',palette='autumn')
plt.title("Employees Education VS Company Leaving Status")
plt.show()
```

Observations

- Most of the employees who were leaving company they mostly did their bachelors
- Very few employees are working in company who did their PHD
- When we compared to working and leaving percentage we can observe that who have PHD their percentage of leaving is when we compared to their working percentage

```
plt.figure(figsize=(15,6))
sns.countplot(x='ExperienceInCurrentDomain',data = data, hue =
'LeaveOrNot',palette='autumn')
plt.title("Employees Experience In Current Domain VS Company Leaving
```

```
Status")  
plt.show()
```

Observations:

- Most of the employees have 2 years experience in Their current domain and percentage of working in the company is high
- In the same way exiting from company percentage is also high i.e which employees have experience in 2 years.

```
plt.figure(figsize=(15,6))  
sns.countplot(x='EverBenched',data = data, hue =  
'LeaveOrNot',palette='autumn')  
plt.title("Employees Ever Benched VS Company Leaving Status")  
plt.show()
```

Observations:

- From above we can observe that Many of the employees working continuously on projects
- Leaving percentage of employees also high, who were working in company.

```
plt.figure(figsize=(15,6))  
sns.countplot(x='Gender',data = data, hue =  
'LeaveOrNot',palette='autumn')  
plt.title("Employees Gender VS Company Leaving Status")  
plt.show()
```

Observations:

- Female Leaving percentage in company is too high when compare to Males

```
plt.figure(figsize=(15,6))  
sns.countplot(x="JoiningYear",data = data, hue =  
'LeaveOrNot',palette='autumn')  
plt.title("Employees Joining Year VS Company Leaving Status")  
plt.show()
```

Observations:

- In 2017 More number employees were working when compared to other years
- In 2018 employees who were leaving percentage is too high