DADS Experiment No: 3

Name: Abhishek S Waghchaure

PRN: **1032221714**

Dept: M Tech DSA

Aim:

Implement visualization and demonstrate Data Analytics using python on any real-world dataset (To Predict the Customer Churn for ABC Bank.).

Objective:

Visualize real world data

Dataset used:

This dataset is for ABC Multistate bank with following columns:

1.customer_id, unused variable.
2.credit_score, used as input.
3.country, used as input.
4.gender, used as input.
5.age, used as input.
6.tenure, From how many years he/she is having bank acc in ABC Bank.
7.balance, Account Balance.
8.products_number, Number of Product from bank.
9.credit_card, Is this customer have credit card ?
10.active_member, Is he/she is active Member of bank ?.
11.estimated_salary, used as input.
12.churn, used as the target. 1 if the client has left the bank during some period or 0 if he/she has not.

Sample data:

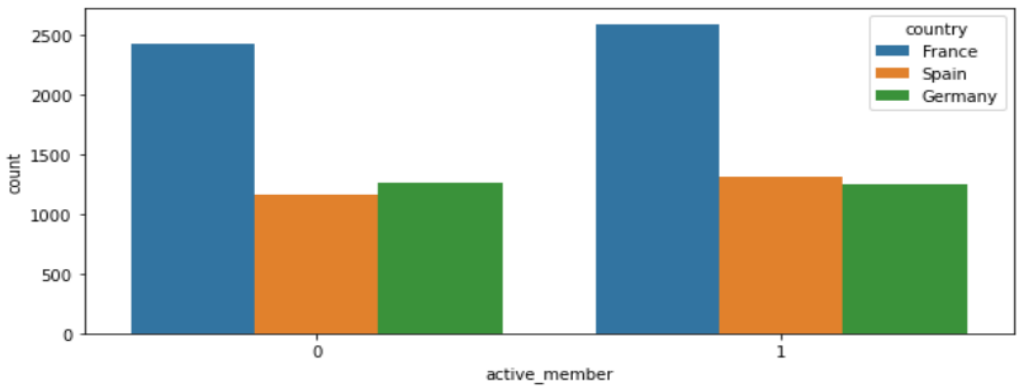
	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

Code & Visualization techniques for the dataset:

```
import numpy as np

import pandas as pd

import seaborn as sb
```

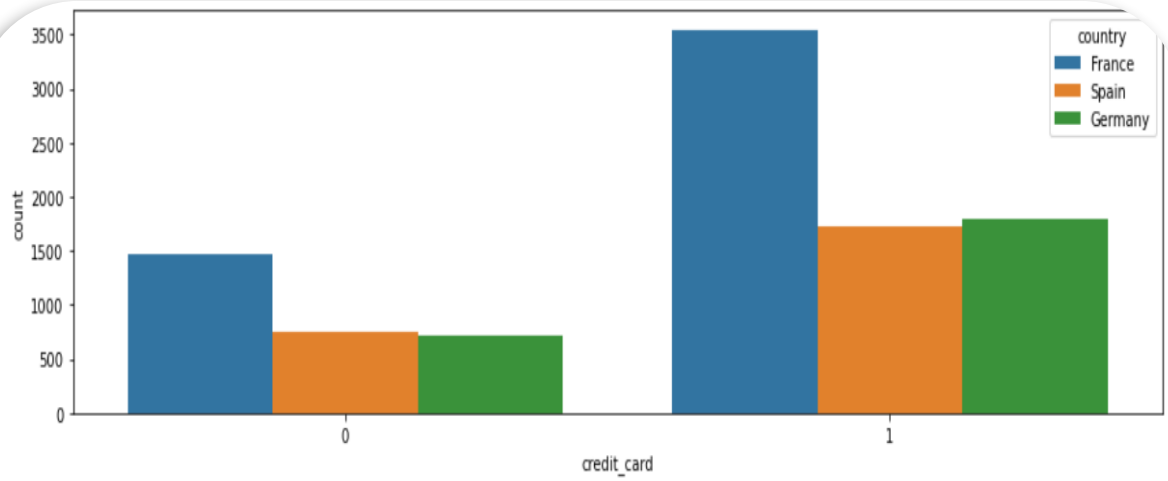
```
import matplotlib.pyplot as plt
df = pd.read_csv('bank.csv')
df.head()
df.info()

# count of active member by country
plt.figure(figsize = (10,4))
sb.countplot(x = 'active_member' , hue = 'country' , data = df)
```

```
Out[6]: <AxesSubplot:xlabel='active_member', ylabel='count'>
```



Count of active member by country

```
#count of credit card by country
plt.figure(figsize = (15,4))
sb.countplot(x = 'credit_card' , hue = 'country' , data = df)
```



Count of credit card by country

```
plt.figure(figsize = (25,8))
```

sb.scatterplot(x = 'credit_score' , y = 'estimated_salary' , hue = 'country' , size = 'estimated_salary' , style = 'gender' , data = df)

Credit Score by Estimated salary with hue(colours) as country

plt.figure(figsize = (25,8))

sb.scatterplot(x = 'credit_score' , y = 'age' , hue = 'gender' , size = 'age' , style = 'gender' , data = df)

plt.figure(figsize = (25,8))

sb.scatterplot(x = 'credit_score' , y = 'balance' , hue = 'country' , size = 'balance' , style = 'gender' , data = df)

df['start_age'] = df['age'] - df['tenure']

plt.figure(figsize = (15,7))

sb.kdeplot('start_age' , hue = 'gender' , shade = 'fill' , data = df.query('active_member == 1'))

df.query('credit_card == 1').groupby(['country' , 'gender'])['start_age'].mean().unstack().astype('int')

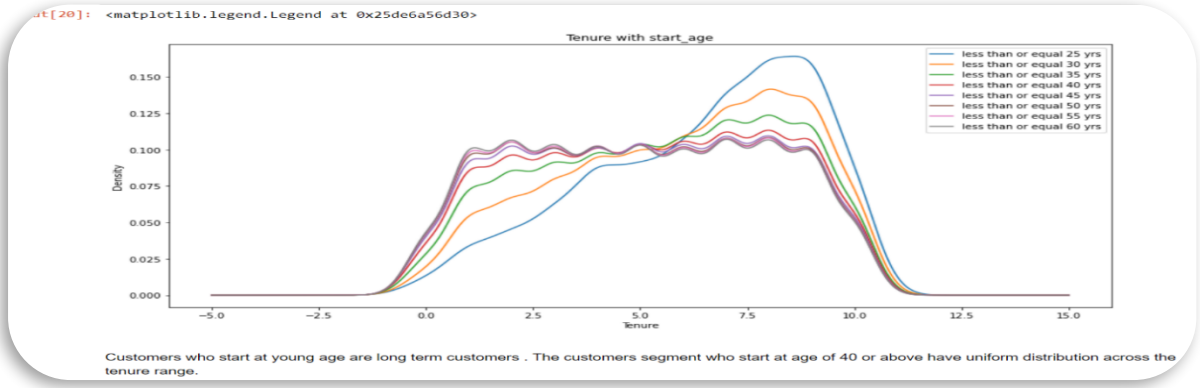
plt.figure(figsize = (15,7))

for age in range(25,65,5):

    df.query('start_age <= @age')['tenure'].plot(kind = 'kde' , title = 'Tenure with start_age', label = 'less than or equal {} yrs'.format(age))

plt.xlabel('Tenure')

plt.legend()



Customers who start at young age are long term customers . The customers segment who start at age of 40 or above have uniform distribution across the tenure range.

Conclusion:

From the above visualization's we can say that, the customers who start at young age are long term customers. The customers segment who starts at age of 40 or above have uniform distribution across the tenure range.

References:

1. https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset
2. https://www.geeksforgeeks.org/data-visualization-with-python/
3. https://www.analyticsvidhya.com/blog/2021/02/an-intuitive-guide-to-visualization-in-python/