

```
#Importing Libraries

In [10]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('/content/Churn_Modelling.csv')
df

Out[10]:
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	10134
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	11254
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	11393
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	9382
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	7908
...
9995	9996	15606229	Objiaiku	771	France	Male	39	5	0.00	2	1	0	9627
9996	9997	15569892	Johnstone	516	France	Male	35	10	57369.61	1	1	1	10169
9997	9998	15584532	Liu	709	France	Female	36	7	0.00	1	0	1	4208
9998	9999	15682355	Sabbatini	772	Germany	Male	42	3	75075.31	2	1	0	9288
9999	10000	15628319	Walker	792	France	Female	28	4	130142.79	1	1	0	3819

10000 rows x 14 columns

```
In [12]: df.shape
```

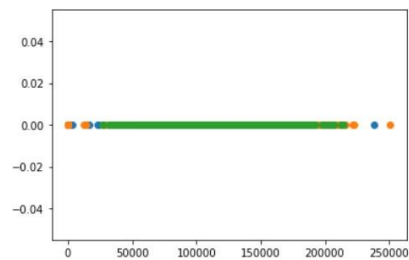
```
Out[12]: (10000, 14)
```

Univariate , Bivariate and Multivariate analysis

Univariate analysis

```
In [14]: df_france=df.loc[df['Geography']=='France']  
df_spain=df.loc[df['Geography']=='Spain']  
df_germany=df.loc[df['Geography']=='Germany']
```

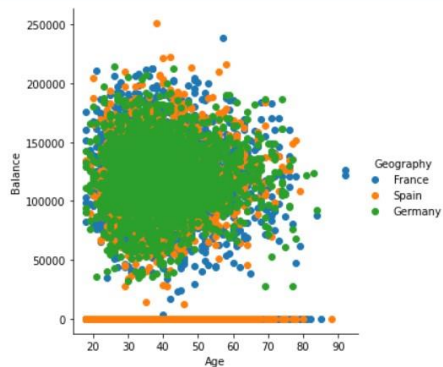
```
In [15]: plt.plot(df_france['Balance'],np.zeros_like(df_france['Balance']),'o')  
plt.plot(df_spain['Balance'],np.zeros_like(df_spain['Balance']),'o')  
plt.plot(df_germany['Balance'],np.zeros_like(df_germany['Balance']),'o')  
plt.xlabel('Age')  
plt.show()
```



Bivariate Analysis

```
In [16]: sns.FacetGrid(df,hue="Geography",size=5).map(plt.scatter,"Age","Balance").add_legend();  
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size` parameter has been renamed to `height`;  
please update your code.  
warnings.warn(msg, UserWarning)
```

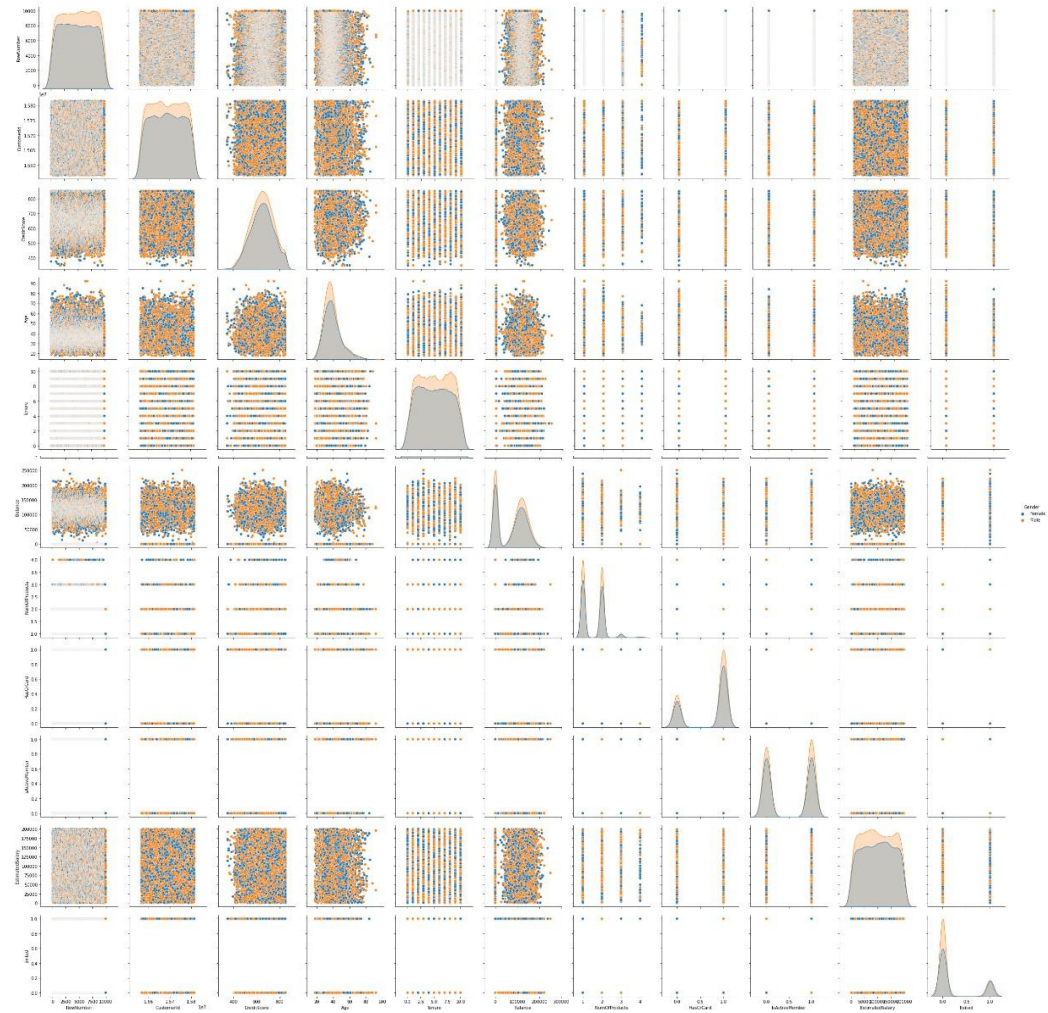


Multivariate Analysis

```
In [17]: sns.pairplot(df, hue="Gender", size=3)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: The `size` parameter has been renamed to `height`  
; please update your code.  
warnings.warn(msg, UserWarning)
```

```
Out[17]: <seaborn.axisgrid.PairGrid at 0x7fc904178f10>
```



Descriptive Statistics

In [18]: `df.head()`

Out[18]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10

In [19]: `df.mean()` # Get the mean of each column

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
"""Entry point for launching an IPython kernel.

Out[19]:

RowNumber	5.000500e+03
CustomerId	1.569094e+07
CreditScore	6.505288e+02
Age	3.892180e+01
Tenure	5.012800e+00
Balance	7.648589e+04
NumOfProducts	1.530200e+00
HasCrCard	7.055000e-01
IsActiveMember	5.151000e-01
EstimatedSalary	1.000902e+05
Exited	2.037000e-01
dtype:	float64

```
plt.vlines(norm_data.median(), # Plot red line at median
            ymin=0,
            ymax=0.4,
            linewidth=2.0,
            color="red");
```

