### Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
```

### Import CSV File

```
Demo=pd.read_csv("/content/KPMG_Customer_Demo.csv")
Demo
```

| | customer_id | first_name | last_name | gender | past_3_years_bike_related_purchases | D |
|---|---|---|---|---|---|---|
| **0** | 1 | Laraine | Medendorp | F | 93 | 12/10/19 |
| **1** | 2 | Eli | Bockman | Male | 81 | 16/12/19 |
| **2** | 3 | Arlin | Dearle | Male | 61 | 20/01/19 |
| **3** | 4 | Talbot | NaN | Male | 33 | 03/10/19 |
| **4** | 5 | Sheila-kathryn | Calton | Female | 56 | 13/05/19 |
| **...** | ... | ... | ... | ... | ... | |
| **3995** | 3996 | Rosalia | Halgarth | Female | 8 | 09/08/19 |
| **3996** | 3997 | Blanch | Nisuis | Female | 87 | 13/07/20 |
| **3997** | 3998 | Sarene | Woolley | U | 60 | N |
| **3998** | 3999 | Patrizius | NaN | Male | 11 | 24/10/19 |
| **3999** | 4000 | Kippy | Oldland | Male | 76 | 05/11/19 |

4000 rows × 13 columns

## Pre Processing Data Information

### Data Information

```
Demo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4000 entries, 0 to 3999
Data columns (total 13 columns):
 #   Column                               Non-Null Count  Dtype  
---  ------                               --------------  -----  
 0   customer_id                          4000 non-null   int64  
 1   first_name                           4000 non-null   object 
 2   last_name                            3875 non-null   object 
 3   gender                               4000 non-null   object 
 4   past_3_years_bike_related_purchases  4000 non-null   int64  
 5   DOB                                  3913 non-null   object 
 6   job_title                            3494 non-null   object 
 7   job_industry_category                3344 non-null   object 
 8   wealth_segment                       4000 non-null   object 
 9   deceased_indicator                   4000 non-null   object 
 10  default                              3698 non-null   object 
 11  owns_car                             4000 non-null   object 
 12  tenure                               3913 non-null   float64
dtypes: float64(1), int64(2), object(10)
memory usage: 406.4+ KB
```

### Finding A Missing Values

```
missing_values = Demo.isnull().sum()
print("Missing Values:\n", missing_values)
```

```
Missing Values:
 customer_id                              0
first_name                                0
last_name                               125
gender                                    0
past_3_years_bike_related_purchases       0
DOB                                      87
job_title                               506
job_industry_category                   656
wealth_segment                            0
deceased_indicator                        0
default                                 302
owns_car                                  0
tenure                                   87
dtype: int64
```

### Finding The Data Types of Attribute

```
Demo.dtypes
```

```
customer_id                              int64
first_name                              object
last_name                               object
gender                                  object
past_3_years_bike_related_purchases      int64
DOB                                     object
job_title                               object
job_industry_category                   object
wealth_segment                          object
deceased_indicator                      object
default                                 object
owns_car                                object
tenure                                 float64
dtype: object
```

### GroupBy Gender

```
gender= Demo.groupby(['gender'])
gender.size()
```

```
gender
F            1
Femal        1
Female    2037
M            1
Male      1872
U           88
dtype: int64
```

### Replaceing Gender with wrong Entry

```
Demo['gender'] = Demo['gender'].replace(['F','Femal'],['Female','Female'])
Demo['gender'] = Demo['gender'].replace(['M'],['Male'])
```

### Drop The unrequied data

```
Demo = Demo.dropna()
```

### Check Duplicate Records from Data

```
duplicate_records = Demo[Demo.duplicated()]
print("Duplicate Records:\n", duplicate_records)
```

```
Duplicate Records:
 Empty DataFrame
Columns: [customer_id, first_name, last_name, gender, past_3_years_bike_related_purchases, DOB, job_title, job_industry_category, wealth_
Index: []
```

### Recheck Groupby Gender Size.

```
gender= Demo.groupby(['gender'])
gender.size()
```

```
gender
Female    1368
Male      1262
dtype: int64
```

### For Further Process Get again information of Data

```
Demo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2630 entries, 0 to 3996
Data columns (total 13 columns):
 #   Column                               Non-Null Count  Dtype  
---  ------                               --------------  -----  
 0   customer_id                          2630 non-null   int64  
 1   first_name                           2630 non-null   object 
 2   last_name                            2630 non-null   object 
 3   gender                               2630 non-null   object 
 4   past_3_years_bike_related_purchases  2630 non-null   int64  
 5   DOB                                  2630 non-null   object 
 6   job_title                            2630 non-null   object 
 7   job_industry_category                2630 non-null   object 
 8   wealth_segment                       2630 non-null   object 
 9   deceased_indicator                   2630 non-null   object 
 10  default                              2630 non-null   object 
 11  owns_car                             2630 non-null   object 
 12  tenure                               2630 non-null   float64
dtypes: float64(1), int64(2), object(10)
memory usage: 287.7+ KB
```

### Check Null Values in Data

```
Demo.isnull().sum()
```

```
customer_id                          0
first_name                           0
last_name                            0
gender                               0
past_3_years_bike_related_purchases  0
DOB                                  0
job_title                            0
job_industry_category                0
wealth_segment                       0
deceased_indicator                   0
default                              0
owns_car                             0
tenure                               0
dtype: int64
```

### Import new Library of datetime to get age from Born date

```
from datetime import datetime, date

born='1953-10-12'
print("Born :",born)

#Identify given date as date month and year
born = datetime.strptime(born, "%Y-%m-%d").date()

#Get today's date
today = date.today()

print("Age :",
      today.year - born.year - ((today.month,today.day) < (born.month,born.day)))
```

```
Born : 1953-10-12
Age : 69
```

### Convert List of born date into age and datatype also change Object into integer

```
from datetime import datetime, date
for i in Demo['DOB']:
    print(i)
    born = i
    born = datetime.strptime(born, "%d/%m/%Y").date()

    #Get today's date
    today = date.today()
    j = today.year - born.year - ((today.month, today.day) < (born.month, born.day))
    Demo['DOB'] = Demo['DOB'].replace([i],[j])
```

```
11/02/1961
17/12/1977
22/10/1967
29/01/1982
10/09/1980
27/10/1986
29/12/1959
21/05/1988
23/06/1959
26/12/1997
30/07/1963
21/01/1961
09/03/1976
04/01/1978
03/07/1976
04/09/1954
10/07/1986
25/11/1977
04/05/1960
07/06/1978
06/08/1973
18/11/1977
03/05/1978
02/04/1965
30/12/1997
04/05/1973
22/06/1986
19/12/2001
07/09/1968
18/06/1999
03/02/2000
05/08/1970
17/01/1957
27/04/1970
05/07/1992
18/04/1978
16/07/1977
27/11/1969
13/07/1963
11/08/1971
21/08/1965
02/03/1972
16/01/1964
10/09/1994
02/06/1962
16/02/1960
21/06/1998
14/07/1994
23/06/1999
05/03/1998
06/08/1985
02/04/1980
05/12/1974
07/04/1989
12/12/1975
09/08/1975
13/07/2001
```

**After Convert Checl the age on DOB Column.**

```
Demo
```

| | customer_id | first_name | last_name | gender | past_3_years_bike_related_purchases | DOB | job_titl |
|---|---|---|---|---|---|---|---|
| **0** | 1 | Laraine | Medendorp | Female | 93 | 69 | Executiv Secretar |
| **1** | 2 | Eli | Bockman | Male | 81 | 42 | Administrativ Office |
| **2** | 3 | Arlin | Dearle | Male | 61 | 69 | Recruitin Manage |
| **8** | 9 | Mala | Lind | Female | 97 | 50 | Busines System Developmer Analys |
| **9** | 10 | Fiorenze | Birdall | Female | 49 | 34 | Senior Qualit Enginee |
| **...** | ... | ... | ... | ... | ... | ... | . |
| **3992** | 3993 | Andi | Dumelow | Female | 6 | 48 | Libraria |
| **3993** | 3994 | Stephie | Byars | Female | 5 | 34 | Structura Analysi |

**Rename DOB column into Age**

```
Demo.rename(columns={'DOB':'Age'})
```

| | customer_id | first_name | last_name | gender | past_3_years_bike_related_purchases | Age | job_titl |
|---|---|---|---|---|---|---|---|
| **0** | 1 | Laraine | Medendorp | Female | 93 | 69 | Executiv Secretar |
| **1** | 2 | Eli | Bockman | Male | 81 | 42 | Administrativ Office |
| **2** | 3 | Arlin | Dearle | Male | 61 | 69 | Recruitin Manage |
| **8** | 9 | Mala | Lind | Female | 97 | 50 | Busines System Developmer Analys |
| **9** | 10 | Fiorenze | Birdall | Female | 49 | 34 | Senior Qualit Enginee |
| **...** | ... | ... | ... | ... | ... | ... | . |
| **3992** | 3993 | Andi | Dumelow | Female | 6 | 48 | Libraria |
| **3993** | 3994 | Stephie | Byars | Female | 5 | 34 | Structura Analysi Enginee |
| **3994** | 3995 | Rusty | Iapico | Male | 93 | 47 | Staff Scientis |
| **3995** | 3996 | Rosalia | Halgarth | Female | 8 | 48 | VP Produc Managemer |
| **3996** | 3997 | Blanch | Nisuis | Female | 87 | 22 | Statistician |

2630 rows × 13 columns

```
Demo.rename(columns={'DOB':'Age'}, inplace = True)
```

**Recheck it already change in data file or not.**

```
Demo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2630 entries, 0 to 3996
Data columns (total 13 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   customer_id                          2630 non-null   int64
 1   first_name                           2630 non-null   object
 2   last_name                            2630 non-null   object
 3   gender                               2630 non-null   object
 4   past_3_years_bike_related_purchases  2630 non-null   int64
 5   Age                                  2630 non-null   int64
 6   job_title                            2630 non-null   object
 7   job_industry_category                2630 non-null   object
```

```
 8   wealth_segment                       2630 non-null   object
 9   deceased_indicator                   2630 non-null   object
 10  default                              2630 non-null   object
 11  owns_car                             2630 non-null   object
 12  tenure                               2630 non-null   float64
dtypes: float64(1), int64(3), object(9)
memory usage: 287.7+ KB
```

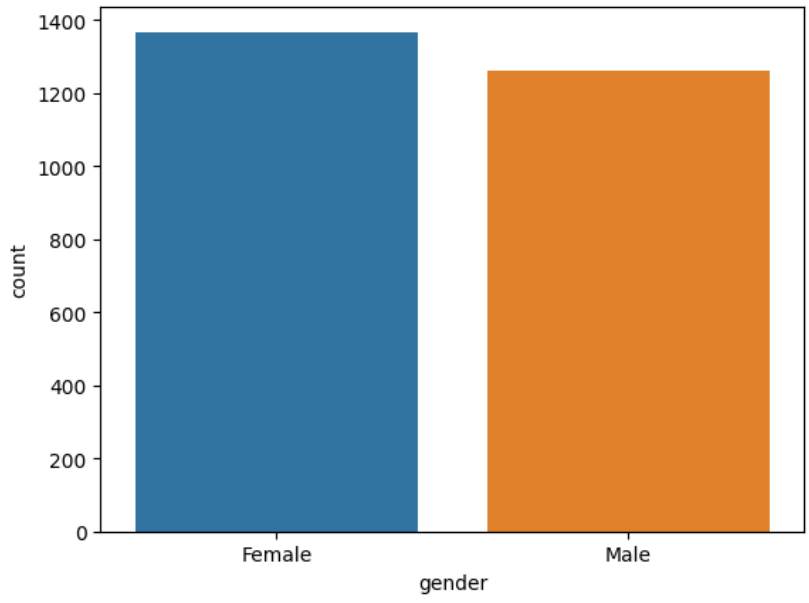
```
Demo= Demo.drop('customer_id', axis=1)
Demo
```

| | first_name | last_name | gender | past_3_years_bike_related_purchases | Age | job_title | job_indust |
|---|---|---|---|---|---|---|---|
| **0** | Laraine | Medendorp | Female | 93 | 69 | Executive Secretary | |
| **1** | Eli | Bockman | Male | 81 | 42 | Administrative Officer | Fina |
| **2** | Arlin | Dearle | Male | 61 | 69 | Recruiting Manager | |
| **8** | Mala | Lind | Female | 97 | 50 | Business Systems Development Analyst | |
| **9** | Fiorenze | Birdall | Female | 49 | 34 | Senior Quality Engineer | Fina |
| **...** | ... | ... | ... | ... | ... | ... | |
| **3992** | Andi | Dumelow | Female | 6 | 48 | Librarian | I |
| **3993** | Stephie | Byars | Female | 5 | 34 | Structural Analysis Engineer | N |
| **3994** | Rusty | Iapico | Male | 93 | 47 | Staff Scientist | N |
| **3995** | Rosalia | Halgarth | Female | 8 | 48 | VP Product Management | |
| **3996** | Blanch | Nisuis | Female | 87 | 22 | Statistician II | N |

2630 rows × 12 columns

```
sns.countplot(x=Demo['gender'])
```

```
<Axes: xlabel='gender', ylabel='count'>
```




```
sns.boxplot(data=Demo)
```

<Axes: >