### 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
```

### Load Customer Address File

```
Add=pd.read_csv('/content/KPMG_Customer_Address.csv')
Add
```

| | customer_id | address | postcode | state | country | property_valuation |
|---|---|---|---|---|---|---|
| **0** | 1 | 060 Morning Avenue | 2016 | New South Wales | Australia | 10 |
| **1** | 2 | 6 Meadow Vale Court | 2153 | New South Wales | Australia | 10 |
| **2** | 4 | 0 Holy Cross Court | 4211 | QLD | Australia | 9 |
| **3** | 5 | 17979 Del Mar Point | 2448 | New South Wales | Australia | 4 |
| **4** | 6 | 9 Oakridge Court | 3216 | VIC | Australia | 9 |
| **...** | ... | ... | ... | ... | ... | ... |
| **3994** | 3999 | 1482 Hauk Trail | 3064 | VIC | Australia | 3 |
| **3995** | 4000 | 57042 Village Green Point | 4511 | QLD | Australia | 6 |
| **3996** | 4001 | 87 Crescent Oaks Alley | 2756 | NSW | Australia | 10 |
| **3997** | 4002 | 8194 Lien Street | 4032 | QLD | Australia | 7 |

### Data Describe

```
Add.describe()
```

| | customer_id | postcode | property_valuation |
|---|---|---|---|
| **count** | 3999.000000 | 3999.000000 | 3999.000000 |
| **mean** | 2003.987997 | 2985.755939 | 7.514379 |
| **std** | 1154.576912 | 844.878364 | 2.824663 |
| **min** | 1.000000 | 2000.000000 | 1.000000 |
| **25%** | 1004.500000 | 2200.000000 | 6.000000 |
| **50%** | 2004.000000 | 2768.000000 | 8.000000 |
| **75%** | 3003.500000 | 3750.000000 | 10.000000 |
| **max** | 4003.000000 | 4883.000000 | 12.000000 |

### Data Information

```
Add.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3999 entries, 0 to 3998
Data columns (total 6 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   customer_id         3999 non-null   int64 
 1   address             3999 non-null   object
 2   postcode            3999 non-null   int64 
 3   state               3999 non-null   object
 4   country             3999 non-null   object
 5   property_valuation  3999 non-null   int64 
dtypes: int64(3), object(3)
memory usage: 187.6+ KB
```

### Check Null Values

```
Add.isnull().sum()
```

```
customer_id           0
address               0
postcode              0
state                 0
country               0
property_valuation    0
dtype: int64
```

### Check Customer ID with GroupBy

```
customer_id = Add.groupby(['customer_id'])
customer_id.size()
```

```
customer_id
1       1
2       1
4       1
5       1
6       1
       ..
3999    1
4000    1
4001    1
4002    1
4003    1
Length: 3999, dtype: int64
```

### Check Customer Address by GroupBy

```
address = Add.groupby(['address'])
address.size()
```

```
address
0 3rd Road                1
0 American Ash Parkway    1
0 Arapahoe Court          1
0 Bay Center              1
0 Blaine Point            1
                         ..
9985 Barnett Circle       1
9987 Stuart Alley         1
9989 Toban Way            1
999 Huxley Hill           1
9993 Mccormick Street     1
Length: 3996, dtype: int64
```

### Check State by GroupBy

```
state = Add.groupby(['state'])
state.size()
```

```
state
NSW               2054
New South Wales     86
QLD                838
VIC                939
Victoria            82
dtype: int64
```

### Replace Some state with Exact State

```
Add['state'] = Add['state'].replace(['NSW','VIC'] , ['New South Wales','Victoria'])
```

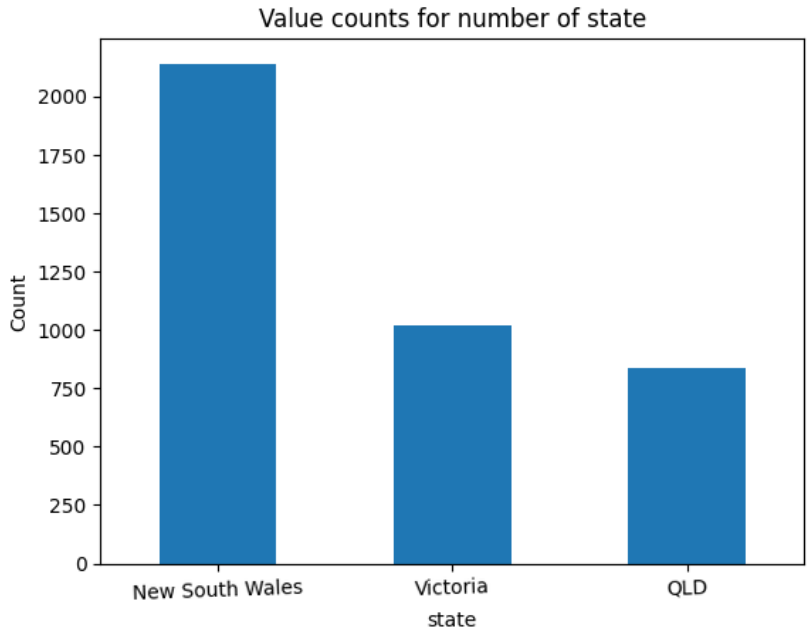
### Recheck State by GroupBy

```
state = Add.groupby(['state'])
state.size()
```

```
state
New South Wales    2140
QLD                 838
Victoria           1021
dtype: int64
```

### Create Bar chart of State

```
Add.state.value_counts().plot(kind="bar")
plt.title("Value counts for number of state")
plt.xlabel("state")
plt.xticks(rotation = 2)
plt.ylabel("Count")
plt.show()
```




### Check Property Valuation with Describe

```
Add['property_valuation'].describe()
```

```
count    3999.000000
mean        7.514379
std         2.824663
min         1.000000
25%         6.000000
50%         8.000000
75%        10.000000
max        12.000000
Name: property_valuation, dtype: float64
```

### Create Hist Graph to check Property Valuation

```
plt.hist(Add['property_valuation'], bins = 10)
plt.show()
```


## Check Skewness and Kurtosis of property Valuation

```
print(f"Skewness: {Add['property_valuation'].skew()}")
print(f"Kurtosis: {Add['property_valuation'].kurt()}")
```

```
Skewness: -0.6314510405576405
Kurtosis: -0.3446494068812771
```

## Check Correction of Data



```
Add.corr()
```

```
<ipython-input-59-89908575a4c5>:1: FutureWarning: The default value of numeric_only in DataFra
  Add.corr()
```

| | customer_id | postcode | property_valuation |
|---|---|---|---|
| **customer_id** | 1.000000 | 0.011396 | -0.012073 |
| **postcode** | 0.011396 | 1.000000 | -0.508392 |
| **property_valuation** | -0.012073 | -0.508392 | 1.000000 |