```
In [19]:  import pandas as pd
```

```
In [20]:  data=pd.read_excel('company.xlsx' )
          data
```

Out[20]:

	company	fuel_type	number_of_door	body_style	number_of_cylinder
0	audi	gas	2	sedan	five
1	audi	gas	4	sedan	five
2	bmw	gas	2	sedan	four
3	bmw	gas	2	sedan	four
4	bmw	gas	2	sedan	five
5	mercedes_benz	diesel	4	sedan	six
6	mercedes_benz	diesel	3	wagon	six

finding the duplicate data

```
In [21]:  data.duplicated()
```

```
Out[21]:  0    False
          1    False
          2    False
          3     True
          4    False
          5    False
          6    False
          dtype: bool
```

```
In [22]:  data.duplicated().sum()
```

Out[22]: 1

Removing the duplicate data from the data set

```
In [23]:  data.drop_duplicates()
```

Out[23]:

	company	fuel_type	number_of_door	body_style	number_of_cylinder
0	audi	gas	2	sedan	five
1	audi	gas	4	sedan	five
2	bmw	gas	2	sedan	four
4	bmw	gas	2	sedan	five
5	mercedes_benz	diesel	4	sedan	six
6	mercedes_benz	diesel	3	wagon	six

```
In [24]:  data.drop_duplicates(subset=['company'])
```

Out[24]:

	company	fuel_type	number_of_door	body_style	number_of_cylinder
0	audi	gas	2	sedan	five
2	bmw	gas	2	sedan	four
5	mercedes_benz	diesel	4	sedan	six

```
In [25]:  data.drop_duplicates(subset=['fuel_type','body_style'])
```

Out[25]:

	company	fuel_type	number_of_door	body_style	number_of_cylinder
0	audi	gas	2	sedan	five
5	mercedes_benz	diesel	4	sedan	six
6	mercedes_benz	diesel	3	wagon	six

```
In [26]:  data.drop_duplicates(subset=['fuel_type','body_style'],keep='last')
```

Out[26]:

	company	fuel_type	number_of_door	body_style	number_of_cylinder
4	bmw	gas	2	sedan	five
5	mercedes_benz	diesel	4	sedan	six
6	mercedes_benz	diesel	3	wagon	six

```
In [ ]:
```