

PRE-PROCESS THE DATA

READ THE DATASETS

Date	31 October 2022
Team ID	PNT2022TMID40130
Project Name	Car Resale value Prediction

loading the data from csv file to pandas dataframe

```
car_dataset = pd.read_csv('/content/car data.csv')
```

#inspecting the first five rows of the dataframe

```
car_dataset.head()
```

output :

```
index,Car_Name,Year,Selling_Price,Present_Price,Kms_Driven,Fuel_Type,Seller_Type,Transmission,Owner
0,ritz,2014,3.35,5.59,27000,Petrol,Dealer,Manual,0
1,sx4,2013,4.75,9.54,43000,Diesel,Dealer,Manual,0
2,ciaz,2017,7.25,9.85,6900,Petrol,Dealer,Manual,0
3,wagon r,2011,2.85,4.15,5200,Petrol,Dealer,Manual,0
4,swift,2014,4.6,6.87,42450,Diesel,Dealer,Manual,0
```

#checking the number of rows and columns

```
car_dataset.shape
```

output :

```
(301,9)
```

```
#getting some information about dataset
```

```
car_dataset.info()
```

output :

```
<class 'pandas.core.frame.DataFrame'> RangeIndex:
```

```
301 entries, 0 to 300
```

```
Data columns (total 9 columns):
```

```
# Column      Non-Null Count  Dtype
---  -
0   Car_Name    301 non-null  object
1   Year         301 non-null  int64
2   Selling_Price 301 non-null  float64
3   Present_Price 301 non-null  float64
4   Kms_Driven    301 non-null  int64
5   Fuel_Type     301 non-null  object
6   Seller_Type   301 non-null  object
7   Transmission  301 non-null  object 8 Owner
   301 non-null  int64 dtypes: float64(2), int64(3),
   object(4) memory usage: 21.3+ KB
```

```
#checking the number of missing values
```

```
car_dataset.isnull().sum()
```

output :

```
Car_Name 0
```

```
Year 0
```

```
Selling_Price 0
```

```
Present_Price 0
```

```
Kms_Driven 0
```

```
Fuel_Type 0
```

```
Seller_Type 0
```

```
Transmission 0
```

```
Owner 0 dtype:
```

```
int64
```