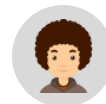


중고차 거래금액 예측 모델링

이수빈





1. 데이터 확인

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.7
1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.5
2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.5
3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.0
4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.7



2. 전처리-결측치 확인

```
train_orig.isnull().sum()
```

Unnamed: 0	0
Name	0
Location	0
Year	0
Kilometers_Driven	0
Fuel_Type	0
Transmission	0
Owner_Type	0
Mileage	2
Engine	36
Power	36
Seats	42
New_Price	5195
Price	0
dtype:	int64



2. 전처리-결측치 확인

```
train_orig.isnull().sum()
```

Unnamed: 0	0
Name	0
Location	0
Year	0
Kilometers_Driven	0
Fuel_Type	0
Transmission	0
Owner_Type	0
Mileage	2
Engine	36
Power	36
Seats	42
New_Price	5195
Price	0
dtype:	int64

```
train_orig.drop(columns=['Unnamed: 0', 'New_Price'], inplace=True)
train_orig.dropna('index', inplace=True)
pd.DataFrame(train_orig)
train_orig.set_index('Name', inplace=True)
train = train_orig.reset_index()
train.shape
#New_Price열에 결측치가 너무 많아서 삭제
(5975, 12)
```



3. Feature Engineering-Name

```
| name=train['Name']  
| len(np.unique(name))  
| #차종으로 하기는 너무 많다. 차종을 없애고 브랜드만 남겨두자  
| 1855
```

```
| for i in range(len(name)):  
|     name[i]=name[i].split(' ')[0]
```

```
| len(name.unique())  
| #브랜드가 31개다.  
| 31
```



3. Feature Engineering-단위삭제

```
for i in range(len(mileage)):
    mileage[i]=mileage[i].split(' ')[0]
    engine[i]=engine[i].split(' ')[0]
    power[i]=power[i].split(' ')[0]
```



3. Feature Engineering-데이터형식

```
print(train.dtypes)
```

Name	object
Location	object
Year	int64
Kilometers_Driven	int64
Fuel_Type	object
Transmission	object
Owner_Type	object
Mileage	object
Engine	object
Power	object
Seats	float64
Price	float64
dtype:	object



3. Feature Engineering-데이터형식

```
print(train.dtypes)
```

Name	object
Location	object
Year	int64
Kilometers_Driven	int64
Fuel_Type	object
Transmission	object
Owner_Type	object
Mileage	object
Engine	object
Power	object
Seats	float64
Price	float64
dtype:	object

```
train['Price']=train['Price'].astype(float)
train['Kilometers_Driven']=train['Kilometers_Driven'].astype(float)
train['Mileage']=train['Mileage'].astype(float)
train['Engine']=train['Engine'].astype(float)
train['Power']=train['Power'].astype(float)
#Object 자료형을 float로 바꾼다
```




3. Feature Engineering-데이터형식

```
print(train.dtypes)
```

Name	object
Location	object
Year	int64
Kilometers_Driven	int64
Fuel_Type	object
Transmission	object
Owner_Type	object
Mileage	object
Engine	object
Power	object
Seats	float64
Price	float64
dtype:	object

#카테고리별로 분류

```
train['Year']=pd.Categorical(train['Year'])  
train['Seats']=pd.Categorical(train['Seats'])
```



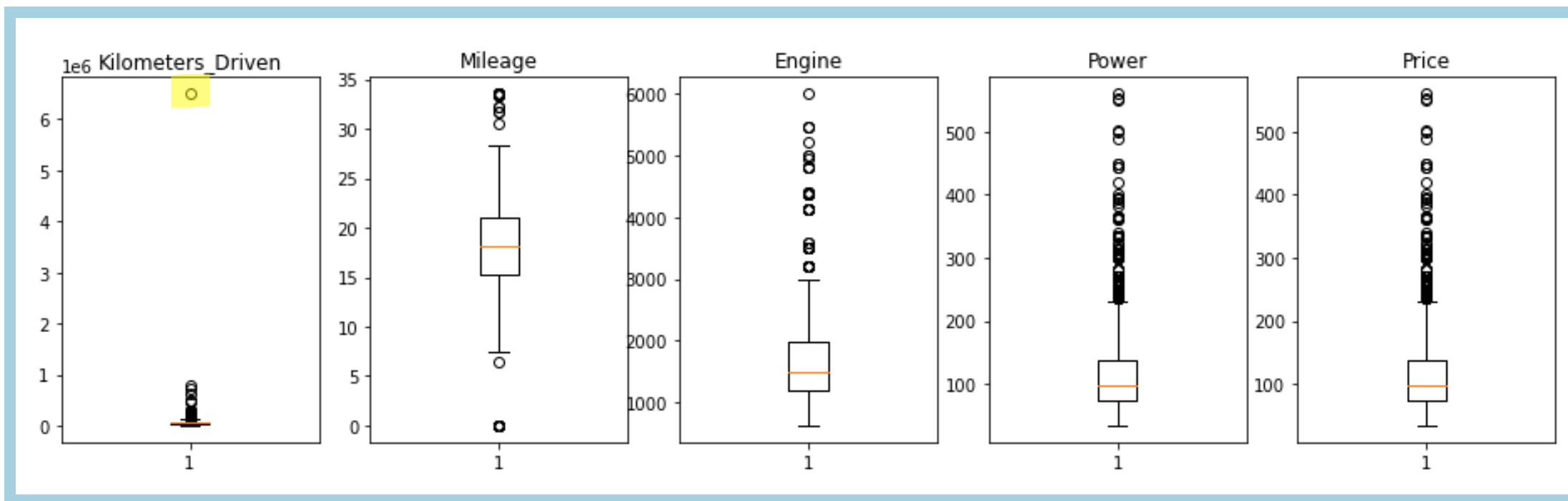
3. Feature Engineering-데이터형식

```
print(train.dtypes)
```

Name	object
Location	object
Year	int64
Kilometers_Driven	int64
Fuel_Type	object
Transmission	object
Owner_Type	object
Mileage	object
Engine	object
Power	object
Seats	float64
Price	float64
dtype:	object

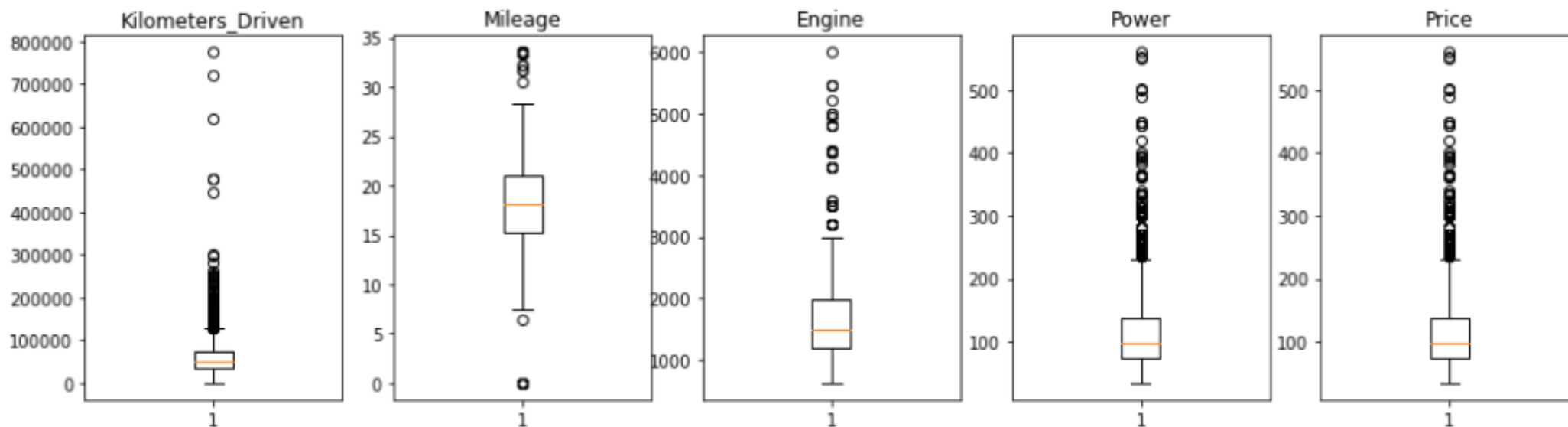
```
train=pd.get_dummies(train,prefix_sep='_',drop_first=True)
```

4. 이상치 확인





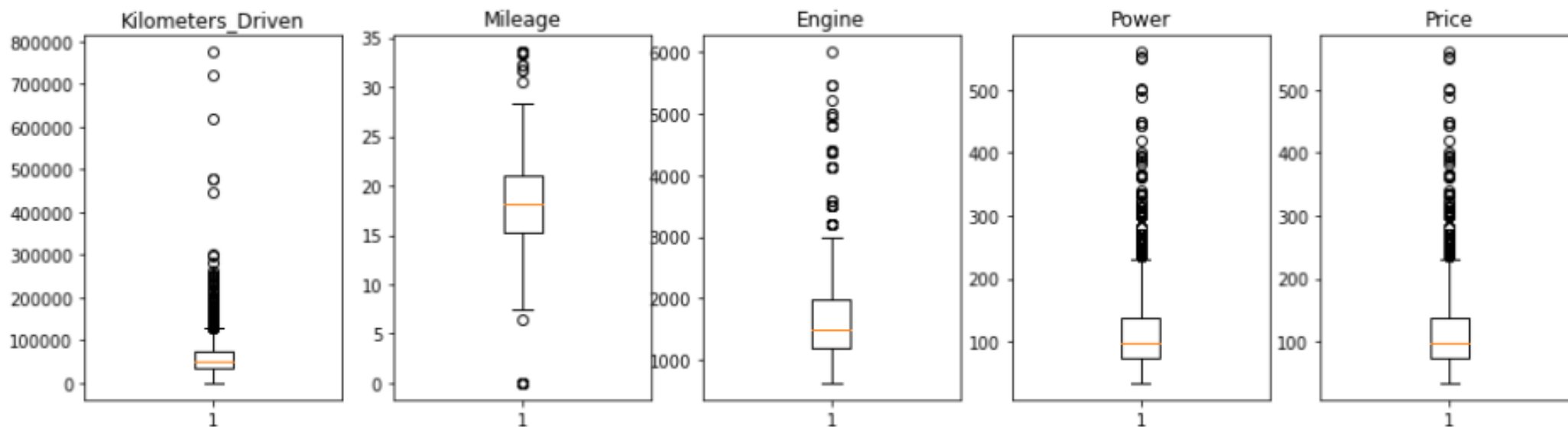
4. 이상치 확인



이상치 제거 후



4. 이상치 확인





5. 모델링- 선형회귀

```
x=train.values  
columns=train.columns  
  
scaler=preprocessing.MinMaxScaler()  
tmp=scaler.fit_transform(x)  
train=pd.DataFrame(tmp)  
train.columns=columns  
  
x=train.to_numpy()
```

```
#data split
```

```
x_train, x_test, y_train, y_test=train_test_split(x,y,train_size=0.8,random_state=1)
```

```
model=LinearRegression(fit_intercept=True,normalize=True,copy_X=True)  
model.fit(x_train,y_train)
```

```
LinearRegression(normalize=True)
```



5. 모델링 선형회귀

```
| model.score(x_train, y_train)
```

```
0.8029093360061492
```



5. 모델링 선형회귀

다음 할일 : 다른 모델링 할 예정입니다
\\(✿°▽°)ノ