

```
In [7]: import pandas as pd
```

```
In [11]: pip install xlrd
```

```
Collecting xlrd
  Downloading xlrd-2.0.1-py2.py3-none-any.whl (96 kB)
----- 0.0/96.5 kB ? eta -:--:--
----- 0.0/96.5 kB ? eta -:--:--
----- 10.2/96.5 kB ? eta -:--:--
----- 41.0/96.5 kB 388.9 kB/s eta 0:00:01
----- 81.9/96.5 kB 573.4 kB/s eta 0:00:01
----- 96.5/96.5 kB 501.7 kB/s eta 0:00:00

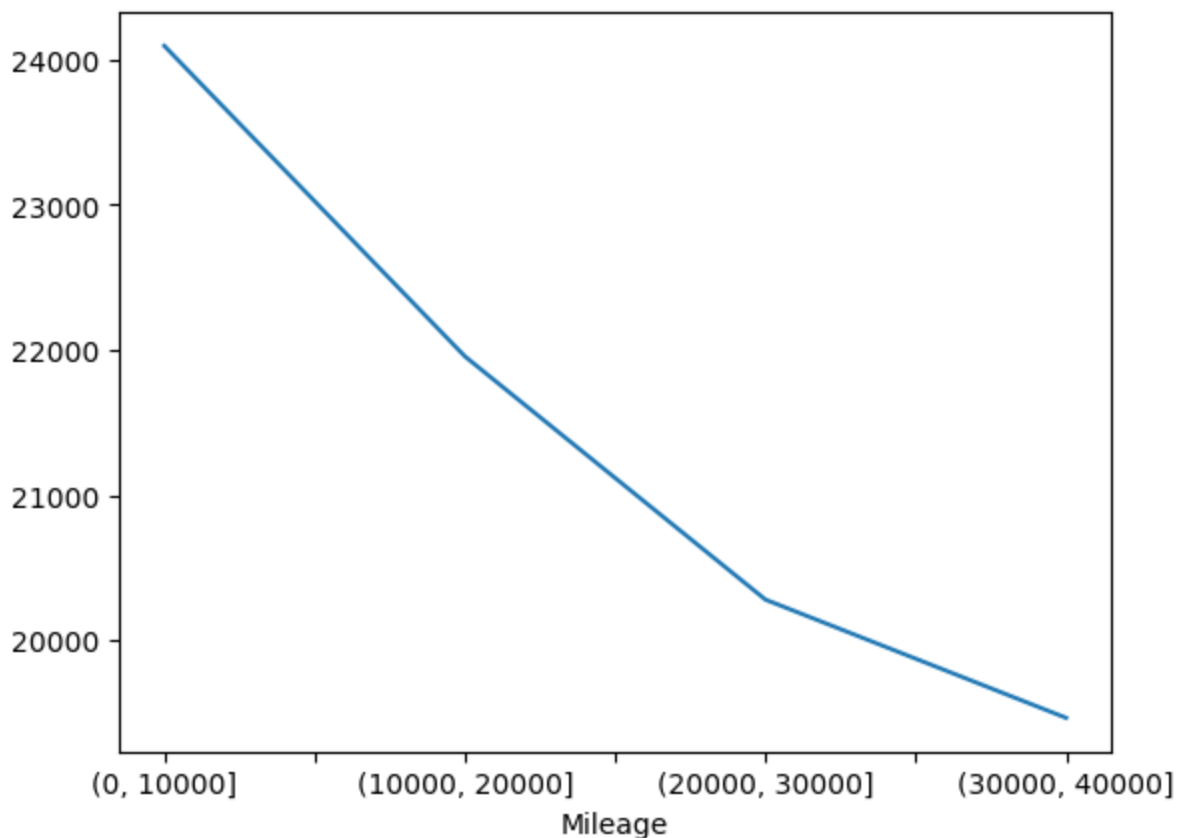
Installing collected packages: xlrd
Successfully installed xlrd-2.0.1
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: import pandas as pd
df = pd.read_excel(r'C:\Users\SAKTHI\Downloads\milestone\github\cars.xls')
```

```
In [14]: %matplotlib inline
import numpy as np
df1=df[['Mileage','Price']]
bins = np.arange(0,50000,10000)
groups = df1.groupby(pd.cut(df1['Mileage'],bins)).mean()
print(groups.head())
groups['Price'].plot.line()
```

	Mileage	Price
Mileage		
(0, 10000]	5588.629630	24096.714451
(10000, 20000]	15898.496183	21955.979607
(20000, 30000]	24114.407104	20278.606252
(30000, 40000]	33610.338710	19463.670267

Out[14]: <Axes: xlabel='Mileage'>



```
In [15]: import statsmodels.api as sm
from sklearn.preprocessing import StandardScaler
scale = StandardScaler()
X = df[['Mileage', 'Cylinder', 'Doors']]
Y = df['Price']
X[['Mileage', 'Cylinder', 'Doors']] = scale.fit_transform(X[['Mileage', 'Cylinder', 'Doors']])
X=sm.add_constant(X)
print(X)
est = sm.OLS(Y,X).fit()
print(est.summary())
```

```
      const  Mileage  Cylinder  Doors
0      1.0 -1.417485   0.52741  0.556279
1      1.0 -1.305902   0.52741  0.556279
2      1.0 -0.810128   0.52741  0.556279
3      1.0 -0.426058   0.52741  0.556279
4      1.0  0.000008   0.52741  0.556279
..      ...      ...      ...      ...
799    1.0 -0.439853   0.52741  0.556279
800    1.0 -0.089966   0.52741  0.556279
801    1.0  0.079605   0.52741  0.556279
802    1.0  0.750446   0.52741  0.556279
803    1.0  1.932565   0.52741  0.556279
```

[804 rows x 4 columns]

#### OLS Regression Results

```
=====
Dep. Variable:          Price      R-squared:            0.360
Model:                  OLS        Adj. R-squared:         0.358
Method:                 Least Squares    F-statistic:         150.0
Date:                  Tue, 16 Jan 2024    Prob (F-statistic):    3.95e-77
Time:                  16:26:51      Log-Likelihood:        -8356.7
No. Observations:      804          AIC:                  1.672e+04
Df Residuals:          800          BIC:                  1.674e+04
Df Model:               3
Covariance Type:        nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	2.134e+04	279.405	76.388	0.000	2.08e+04	2.19e+04
Mileage	-1272.3412	279.567	-4.551	0.000	-1821.112	-723.571
Cylinder	5587.4472	279.527	19.989	0.000	5038.754	6136.140
Doors	-1404.5513	279.446	-5.026	0.000	-1953.085	-856.018

```
=====
Omnibus:                157.913    Durbin-Watson:           0.069
Prob(Omnibus):           0.000     Jarque-Bera (JB):         257.529
Skew:                    1.278     Prob(JB):                 1.20e-56
Kurtosis:                4.074     Cond. No.                  1.03
=====
```

#### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

C:\Users\SAKTHI\AppData\Local\Temp\ipykernel\_5548\2703298333.py:6: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
X[['Mileage', 'Cylinder', 'Doors']] = scale.fit_transform(X[['Mileage', 'Cylinder', 'Doors']].values)
```

```
In [17]: Y.groupby(df.Doors).mean()
```

```
Out[17]: Doors
2      23807.135520
4      20580.670749
Name: Price, dtype: float64
```

```
In [18]: scaled = scale.transform([[45000,0,4]])
scaled = np.insert(scaled[0],0,1)
print(scaled)
predicted = est.predict(scaled)
print(predicted)

[ 1.          3.07256589 -3.7995081    0.55627894]
[-4577.08180073]
```

```
In [ ]:
```