

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23

PART B

Roll No: C035	Name: Krisha Goti
Class: B	Batch: B1
Date of Experiment: 3/9/22	Date of Submission: 9/9/22
Grade	

B.1 Work done by student

CODE FOR SIMPLE LINEAR REGRESSION

```
#Krisha Goti C035
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from sklearn.metrics import mean_squared_error, mean_absolute_error
```

```
from sklearn import preprocessing
```

```
#Krisha Goti C035
```

```
df = pd.read_csv('USA_cars_datasets.csv')
```

```
df.head()
```

```
#Krisha Goti C035
```

```
df_ = df.loc[:, ['price', 'year']]
```

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23

```
df.plot(x='year',y='price',style='o')

plt.xlabel('year')

plt.ylabel('price')

plt.show()

x= pd.DataFrame(df['year'])

y= pd.DataFrame(df['price'])

from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x,y, train_size=0.7
,test_size=0.3, random_state=100)

x_train

y_train


#Krisha Goti C035

import statsmodels.api as sm

x_train_sm = sm.add_constant(x_train)

lr = sm.OLS(y_train, x_train_sm).fit()

lr.params

lr.summary()

plt.scatter(x_train, y_train)

plt.plot(x_train, -2.864e+06 + 1429.5888*x_train, 'r')

plt.xlabel('year')

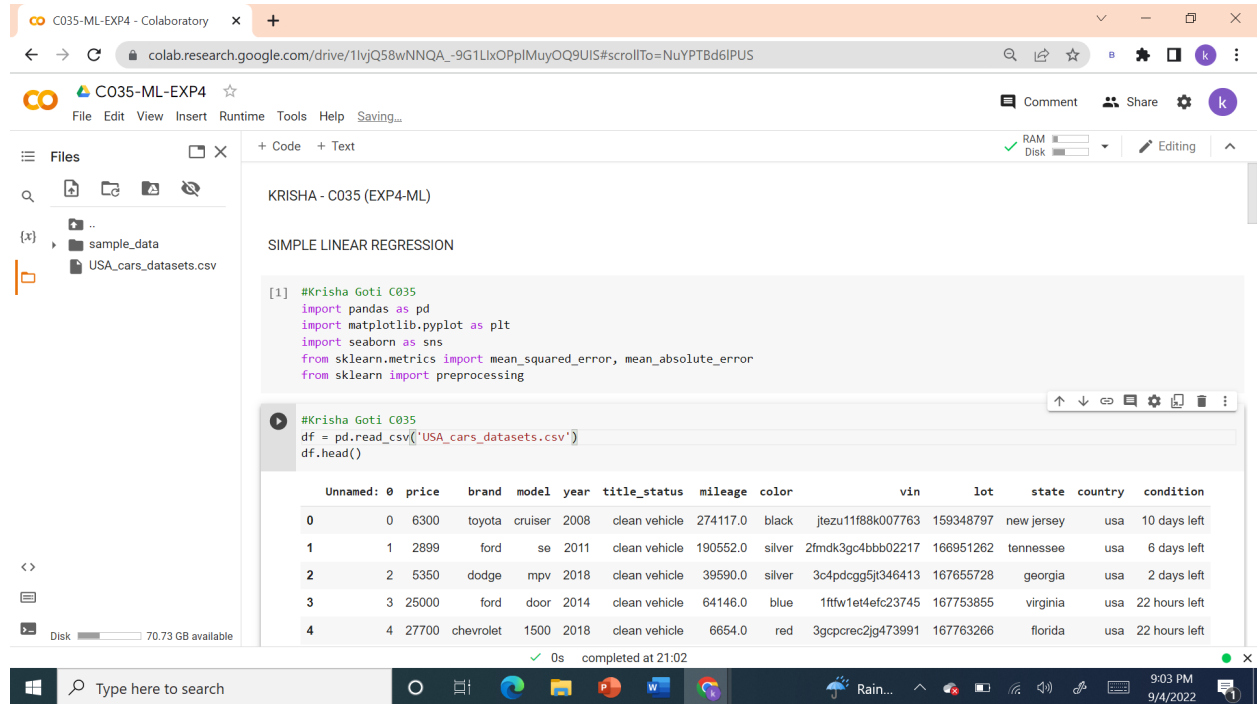
plt.ylabel('price')
```

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23

```
plt.show()
```

OUTPUT



035-ML-EXP4 - Colaboratory

colab.research.google.com/drive/1lvjQ58wNNQA_-9G1LlxOPpIMuyOQ9UIS#scrollTo=NuYPTBd6lPUS

C035-ML-EXP4

File Edit View Insert Runtime Tools Help Saving...

Files

- sample_data
- USA_cars_datasets.csv

Code

```
[1] #Krisha Goti C035
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn import preprocessing

#Krisha Goti C035
df = pd.read_csv('USA_cars_datasets.csv')
df.head()
```

	Unnamed: 0	price	brand	model	year	title_status	mileage	color	vin	lot	state	country	condition
0	0	6300	toyota	cruiser	2008	clean vehicle	274117.0	black	jtezu11f88k007763	159348797	new jersey	usa	10 days left
1	1	2899	ford	se	2011	clean vehicle	190552.0	silver	2fmdk3gc4bbb02217	166951262	tennessee	usa	6 days left
2	2	5350	dodge	mpv	2018	clean vehicle	39590.0	silver	3c4pdggg5j346413	167655728	georgia	usa	2 days left
3	3	25000	ford	door	2014	clean vehicle	64146.0	blue	1ftfw1et4efc23745	167753855	virginia	usa	22 hours left
4	4	27700	chevrolet	1500	2018	clean vehicle	6654.0	red	3gcpccrc2jg473991	167763266	florida	usa	22 hours left

0s completed at 21:02

9:03 PM 9/4/2022

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23

C035-ML-EXP4 - Colaboratory

colab.research.google.com/drive/1lvjQ58wNNQA_-9G1LxOPpIMuyOQ9UIS#scrollTo=NuYPTBd6IPUS

C035-ML-EXP4

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- USA_cars_datasets.csv

```
[35] #Krisha Goti C035
df_ = df.loc[:,['price','year']]

df.plot(x='year',y='price',style='o')
plt.xlabel('year')
plt.ylabel('price')
plt.show()
```

```
[6] #Krisha Goti C035
x= pd.DataFrame(df['year'])
y= pd.DataFrame(df['price'])
```

0s completed at 21:02

Type here to search

84°F

9:03 PM 9/4/2022

C035-ML-EXP4 - Colaboratory

colab.research.google.com/drive/1lvjQ58wNNQA_-9G1LxOPpIMuyOQ9UIS#scrollTo=NuYPTBd6IPUS

C035-ML-EXP4

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- USA_cars_datasets.csv

```
[25] #Krisha Goti C035
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, train_size=0.7 ,test_size=0.3, random_state=100)
```

```
#Krisha Goti C035
x_train
```

year
1816
751
545
198
1694
...
350
1930
79
1859
1544

1740 rows x 1 column

0s completed at 21:02

Type here to search

84°F

9:04 PM 9/4/2022

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23

The screenshot shows a Google Colab notebook titled 'C035-ML-EXP4'. The file explorer on the left shows a folder 'sample_data' containing a file 'USA_cars_datasets.csv'. The code cell [28] contains the following code:

```
[28] #Krisha Goti C035
y_train
```

The output of the code cell is a preview of the 'price' column from the 'USA_cars_datasets.csv' file. The data is displayed as a table with 1749 rows and 1 column.

price
1816
751
545
198
1694
...
350
1930
79
1859
1544
1749

The status bar at the bottom indicates '0s completed at 21:02'.

The screenshot shows the same Google Colab notebook. The code cell [29] contains the following code:

```
[29] #Krisha Goti C035
import statsmodels.api as sm
x_train_sm = sm.add_constant(x_train)

/usr/local/lib/python3.7/dist-packages/statsmodels/tsa/tsatools.py:142: FutureWarning: In a future version of pandas all arguments of concat
x = pd.concat(x[::order], 1)
```

The code cell [31] contains the following code:

```
[31] #Krisha Goti C035
lr = sm.OLS(y_train, x_train_sm).fit()
lr.params
```

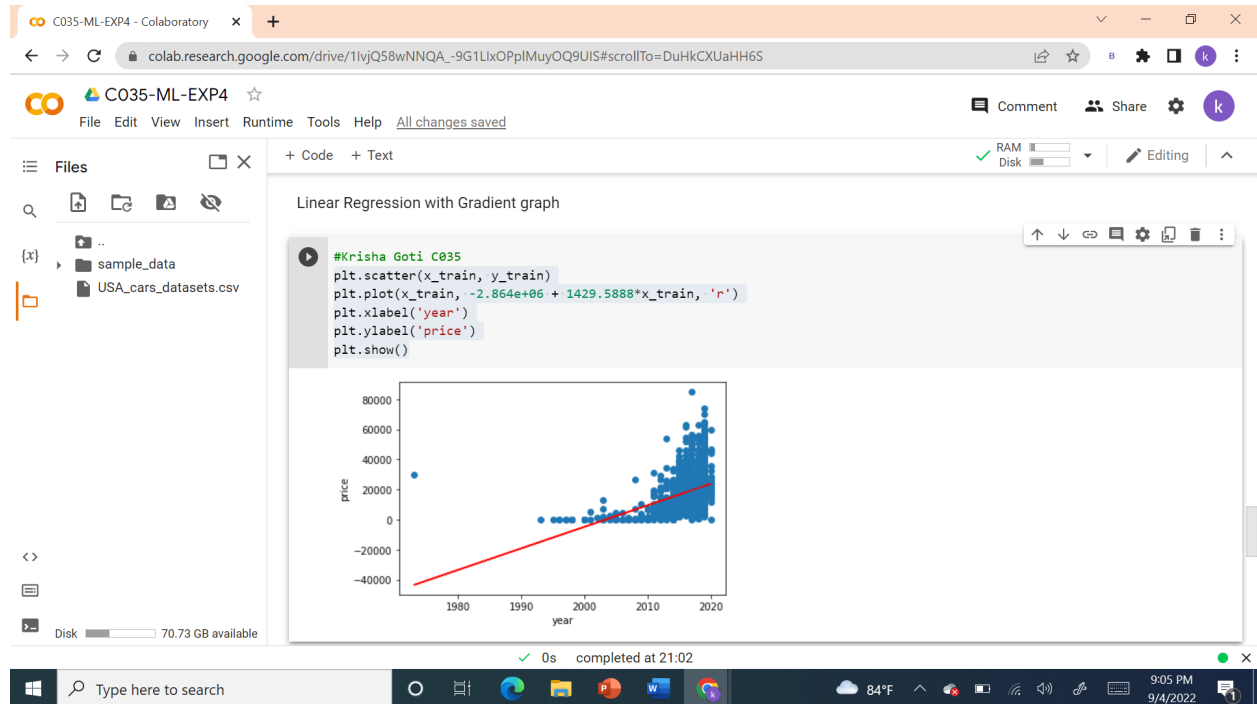
The output of the code cell [31] is the OLS regression results for the 'price' variable. The results are displayed as a table with 1749 rows and 1 column.

OLS Regression Results			
Dep. Variable:	price	R-squared:	0.163
Model:	OLS	Adj. R-squared:	0.162
Method:	Least Squares	F-statistic:	339.7
Date:	Sun, 04 Sep 2022	Prob (F-statistic):	1.87e-69
Time:	15:26:27	Log-Likelihood:	-18745.
No. Observations:	1749	AIC:	3.749e+04
Df Residuals:	1747	BIC:	3.751e+04

The status bar at the bottom indicates '0s completed at 21:02'.

SVKM'S NMIMS Deemed-to-be-University
Mukesh Patel School of Technology Management & Engineering
Department of Computer Engineering

Course Code	Computers	Program	B.Tech
Semester	IX	Year	V
Name of the Faculty	Artika Singh	Class	B and C
Course Title	Data Mining	Academic year	2022-23



B.2 Conclusion

After successfully completing this experiment, I am able to:

Apply simple regression on real world dataset and estimate its parameters. Also learnt some new python libraries for plotting a scatter plot. Learnt how to predict dependent variable for dataset.