Maria Ali

Assignment 3- Machine Learning Course

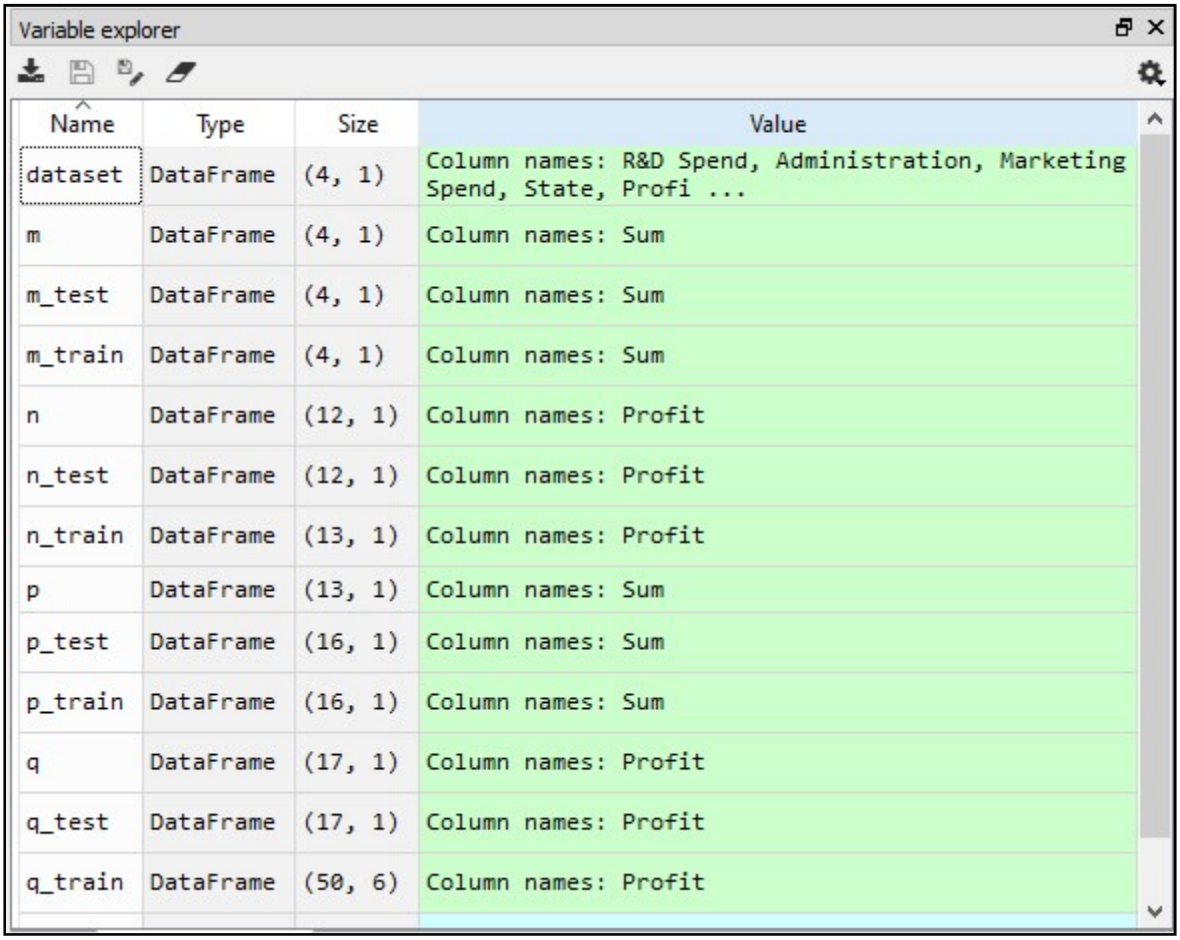
Question No. 1: Take 50 startups of any two countries and find out which country is going to provide best profit in future.

Using Decision Tree Regression

Countries selected:

1. California
2. Florida

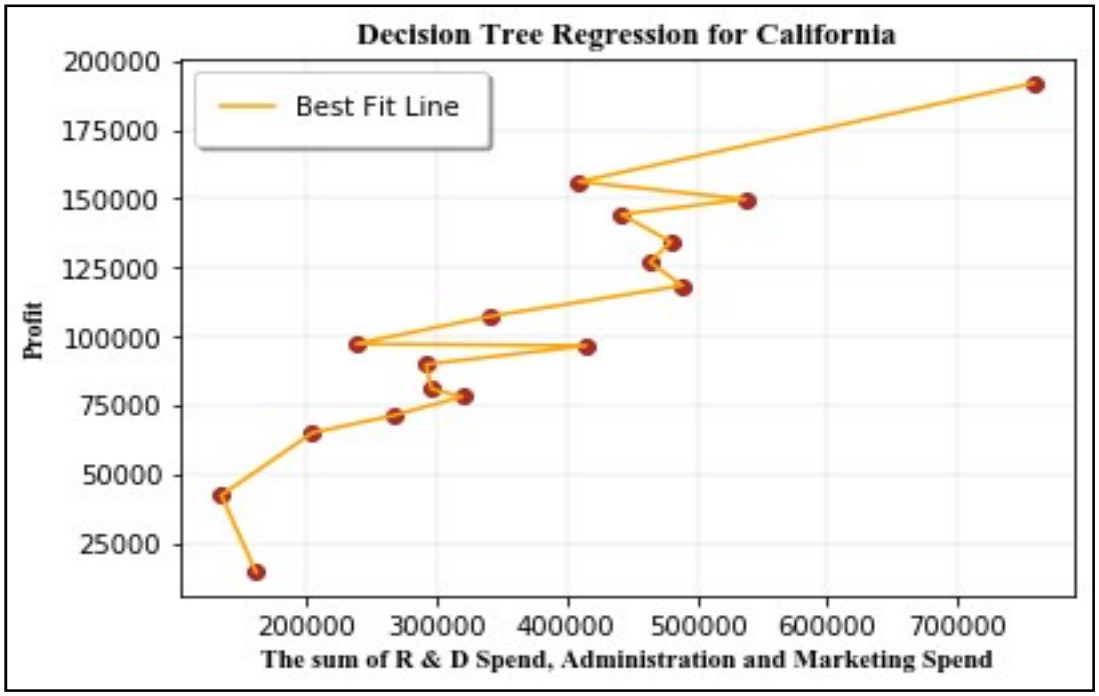
Variable Explorer



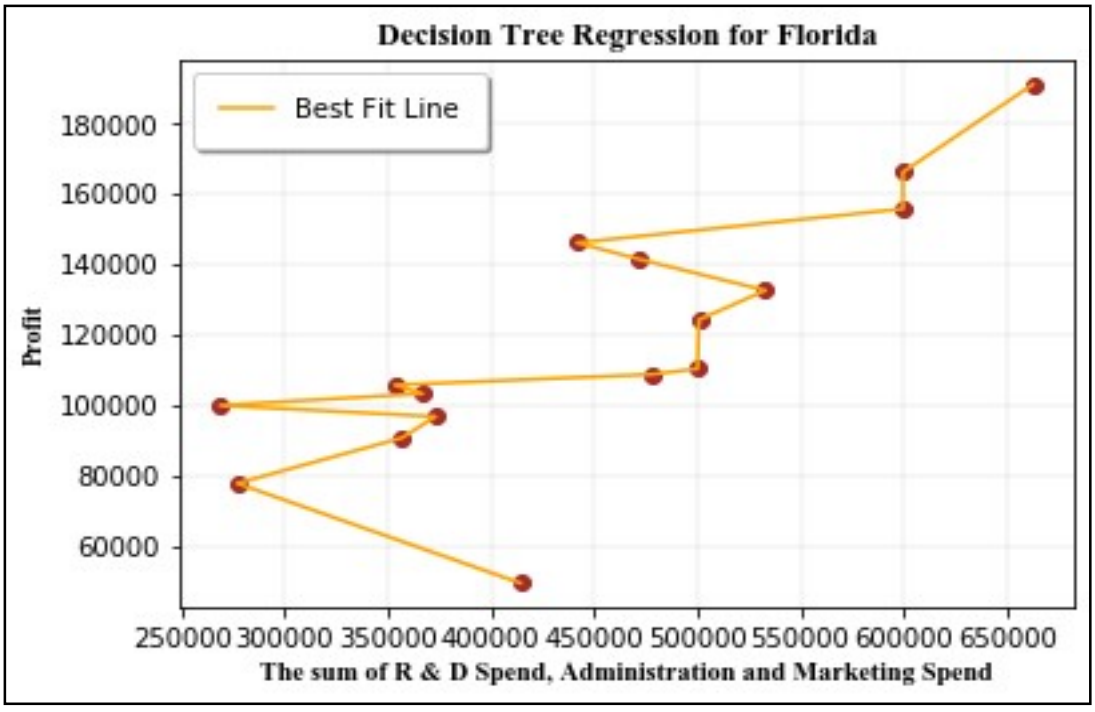
Name	Type	Size	Value
dataset	DataFrame	(4, 1)	Column names: R&D Spend, Administration, Marketing Spend, State, Profi ...
m	DataFrame	(4, 1)	Column names: Sum
m_test	DataFrame	(4, 1)	Column names: Sum
m_train	DataFrame	(4, 1)	Column names: Sum
n	DataFrame	(12, 1)	Column names: Profit
n_test	DataFrame	(12, 1)	Column names: Profit
n_train	DataFrame	(13, 1)	Column names: Profit
p	DataFrame	(13, 1)	Column names: Sum
p_test	DataFrame	(16, 1)	Column names: Sum
p_train	DataFrame	(16, 1)	Column names: Sum
q	DataFrame	(17, 1)	Column names: Profit
q_test	DataFrame	(17, 1)	Column names: Profit
q_train	DataFrame	(50, 6)	Column names: Profit

Plots

For California



For Florida



Result: As shown in graph, California is going to provide best profit in future.

Question No. 2: Annual temperature between two industries is given. Predict the temperature in 2016 and 2017 using the past data of both industries.

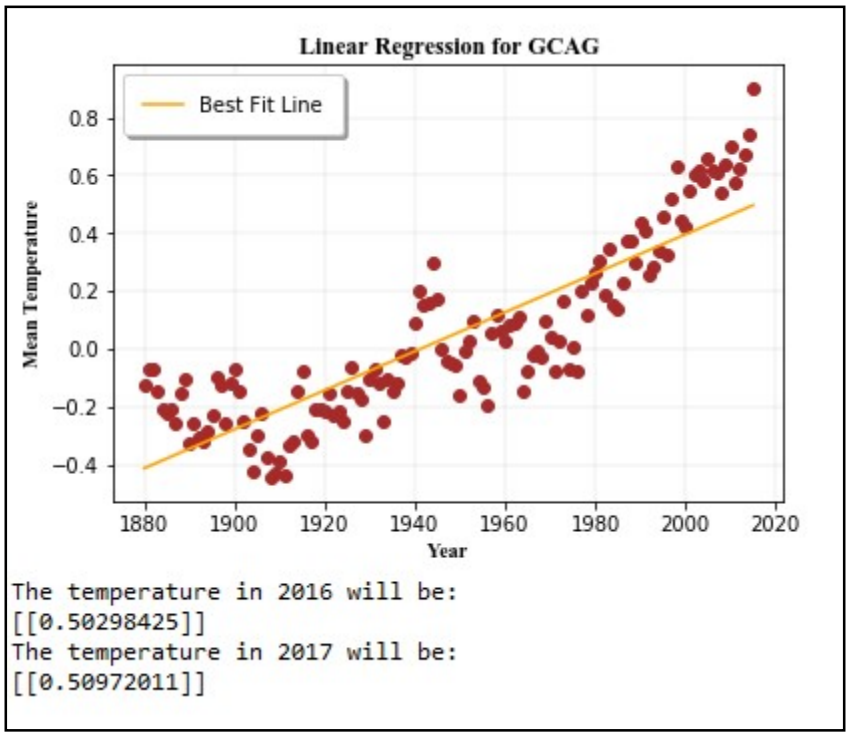
Variable Explorer

Variable explorer

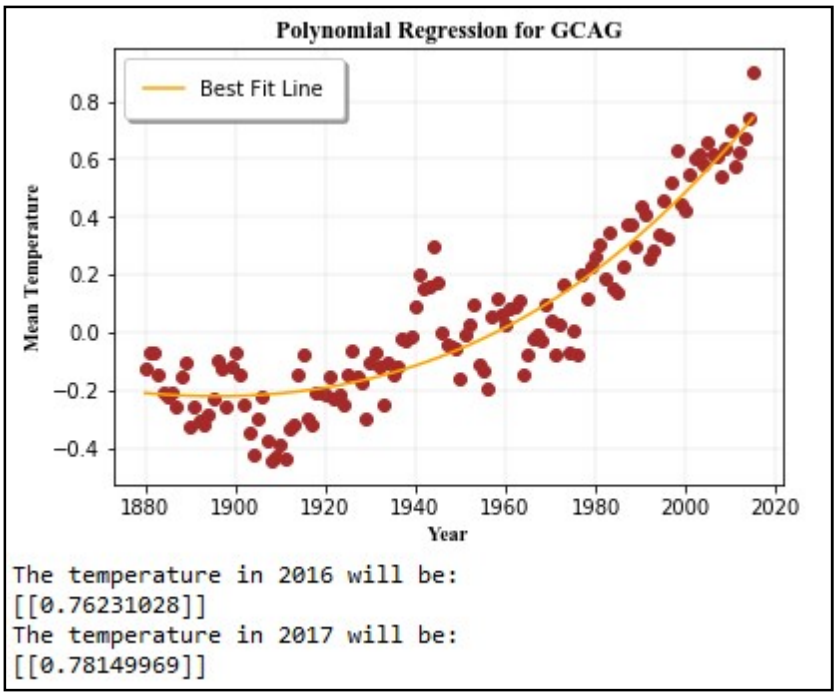
Name	Type	Size	Value
dataset	DataFrame	(272, 3)	Column names: Source, Year, Mean
m	DataFrame	(136, 1)	Column names: Year
m_poly	float64	(136, 5)	[[1.00000000e+00 2.01500000e+03 4.060… 1.6485 ...
n	DataFrame	(136, 1)	Column names: Mean
p	DataFrame	(136, 1)	Column names: Year
p_poly	float64	(136, 5)	[[1.00000000e+00 2.01500000e+03 4.060… 1.6485 ...
q	DataFrame	(136, 1)	Column names: Mean
temperature2016	float64	(1, 1)	[[0.50298425]]
temperature2017	float64	(1, 1)	[[0.50972011]]
temperaturee2016	float64	(1, 1)	[[0.76231028]]
temperaturee2017	float64	(1, 1)	[[0.78149969]]
temperatureee2016	float64	(1, 1)	[[0.49777778]]
temperatureee2017	float64	(1, 1)	[[0.50477625]]
temperatureeee2016	float64	(1, 1)	[[0.78885745]]
temperatureeee2017	float64	(1, 1)	[[0.81039365]]
tnrfont	dict	1	{'fontname':'Times New Roman'}

Plots For GCAG

Prediction of temperature in 2016 and 2017 using Linear Regression



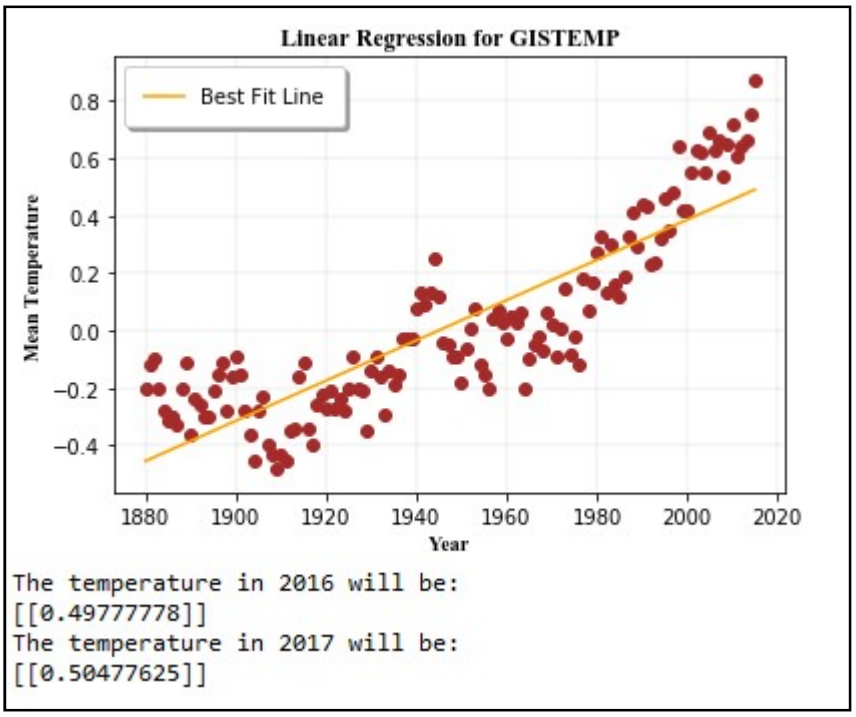
Prediction of temperature in 2016 and 2017 using Polynomial Regression



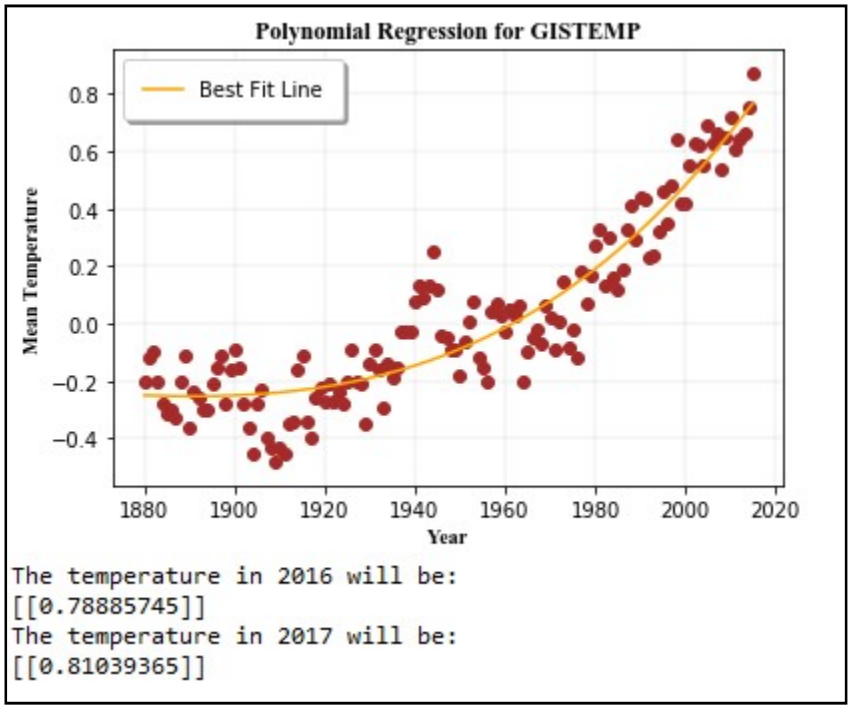
Result: Temperatures predicted with Polynomial Regression are more accurate.

Plots For GISTEMP

Prediction of temperature in 2016 and 2017 using Linear Regression



Prediction of temperature in 2016 and 2017 using Polynomial Regression



Result: Temperatures predicted with Polynomial Regression are more accurate.

Question No. 3: Data of global production of CO2 of a place is given between 1970s to 2010. Predict the CO2 production for the years 2011, 2012 and 2013 using the old data set.

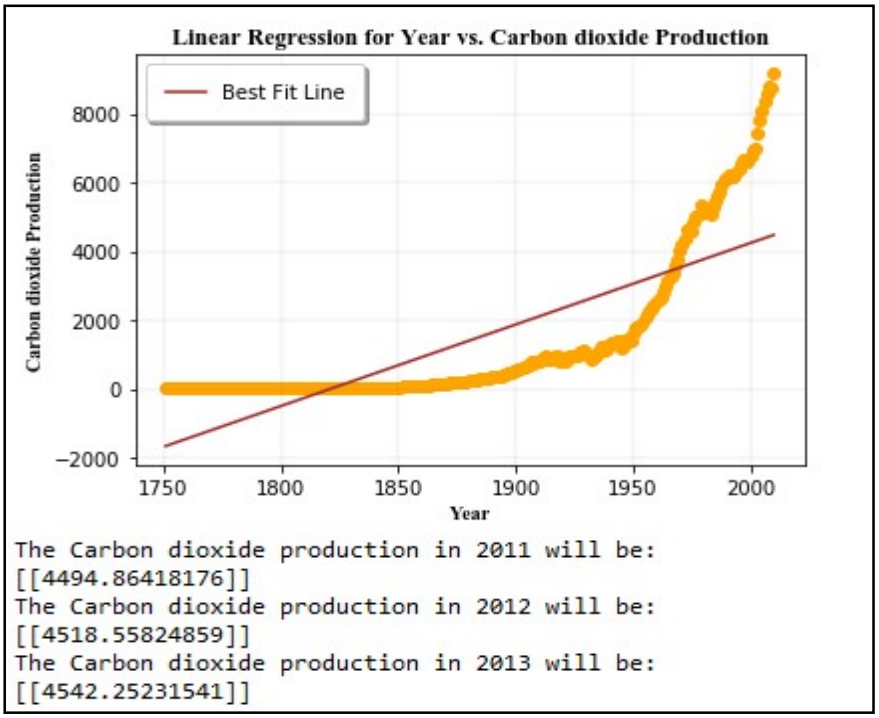
Variable Explorer

Variable explorer

Name	Type	Size	Value
dataset	DataFrame	(260, 8)	Column names: Year, Total, Gas Fuel, Liquid Fuel, Solid Fuel, Cement, ...
p	int64	(260, 1)	[[1751] [1752]
p_grid	float64	(2590, 1)	[[1751.] [1751.1]
p_poly	float64	(260, 5)	[[1.00000000e+00 1.75100000e+03 3.06600100e+06 5.… 9.4003 ...
q	int64	(260, 1)	[[3] [3]
tnrfont	dict	1	{'fontname':'Times New Roman'}
y2011	float64	(1, 1)	[[4494.86418176]]
y2012	float64	(1, 1)	[[4518.55824859]]
y2013	float64	(1, 1)	[[4542.25231541]]
y_2011	float64	(1, 1)	[[4494.86418176]]
y_2012	float64	(1, 1)	[[4518.55824859]]
y_2013	float64	(1, 1)	[[4542.25231541]]

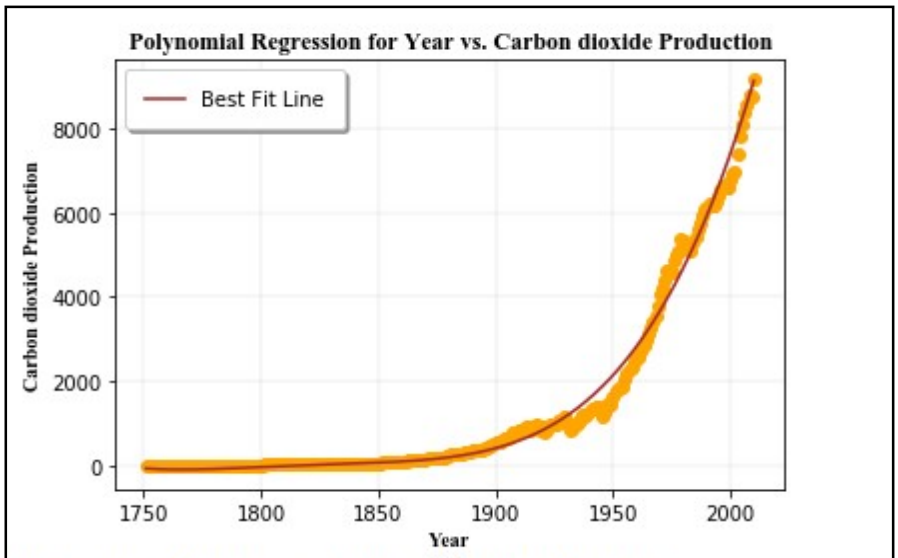
Plots

Carbon dioxide production for the years 2011, 2012 and 2013 using Linear Regression:




```
The Carbon dioxide production in 2011 will be:
[[4494.86418176]]
The Carbon dioxide production in 2012 will be:
[[4518.55824859]]
The Carbon dioxide production in 2013 will be:
[[4542.25231541]]
```

Carbon dioxide production for the years 2011, 2012 and 2013 using Polynomial Regression:




```
The Carbon dioxide production in 2011 will be:
[[4494.86418176]]
The Carbon dioxide production in 2012 will be:
[[4518.55824859]]
The Carbon dioxide production in 2013 will be:
[[4542.25231541]]
```

Result: Carbon dioxide production predicted with Polynomial Regression is more accurate.

Question No. 4: Housing price according to the ID is assigned to every-house. Perform future analysis where when ID is inserted the housing price is displayed.

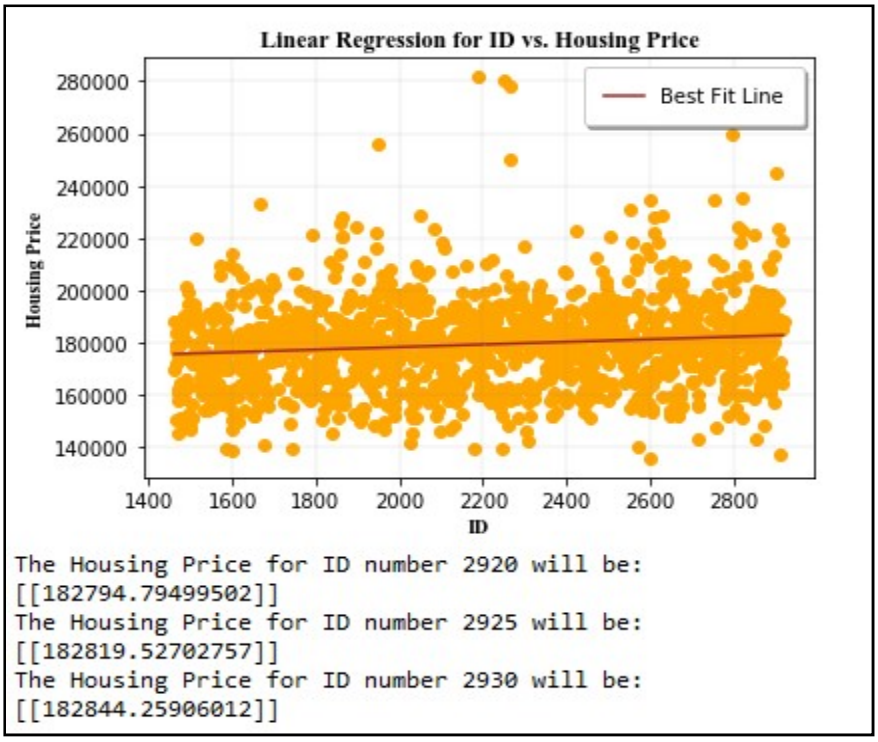
Variable Explorer

Variable explorer

Name	Type	Size	Value
dataset	DataFrame	(1459, 2)	Column names: Id, SalePrice
hp2920	float64	(1, 1)	[[182794.79499502]]
hp2925	float64	(1, 1)	[[182819.52702757]]
hp2930	float64	(1, 1)	[[182844.25906012]]
hp_2920	float64	(1, 1)	[[182794.79499502]]
hp_2925	float64	(1, 1)	[[182819.52702757]]
hp_2930	float64	(1, 1)	[[182844.25906012]]
p	int64	(1459, 1)	[[1461] [1462]]
p_grid	float64	(14580, 1)	[[1461.] [1461.1]]
p_poly	float64	(1459, 5)	[[1.00000000e+00 1.46100000e+03 2.13452100e+06 3… 4.5561 ...
q	float64	(1459, 1)	[[169277.0524984] [187758.39398877]]
tnrfont	dict	1	{'fontname':'Times New Roman'}

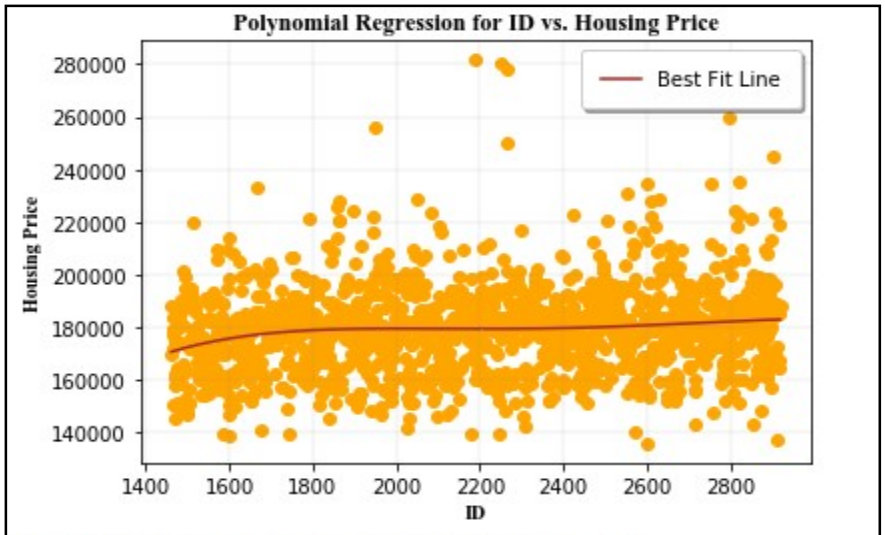
Plots

Future analysis for housing price using Linear Regression:




```
The Housing Price for ID number 2920 will be:
[[182794.79499502]]
The Housing Price for ID number 2925 will be:
[[182819.52702757]]
The Housing Price for ID number 2930 will be:
[[182844.25906012]]
```

Future analysis for housing price using Polynomial Regression:




```
The Housing Price for ID number 2920 will be:
[[182794.79499502]]
The Housing Price for ID number 2925 will be:
[[182819.52702757]]
The Housing Price for ID number 2930 will be:
[[182844.25906012]]
```

Result: Future analysis of housing price done with Polynomial Regression is more accurate.

Question No. 5: Data of monthly experience and income distribution of different employs is given. Perform regression.

Variable Explorer

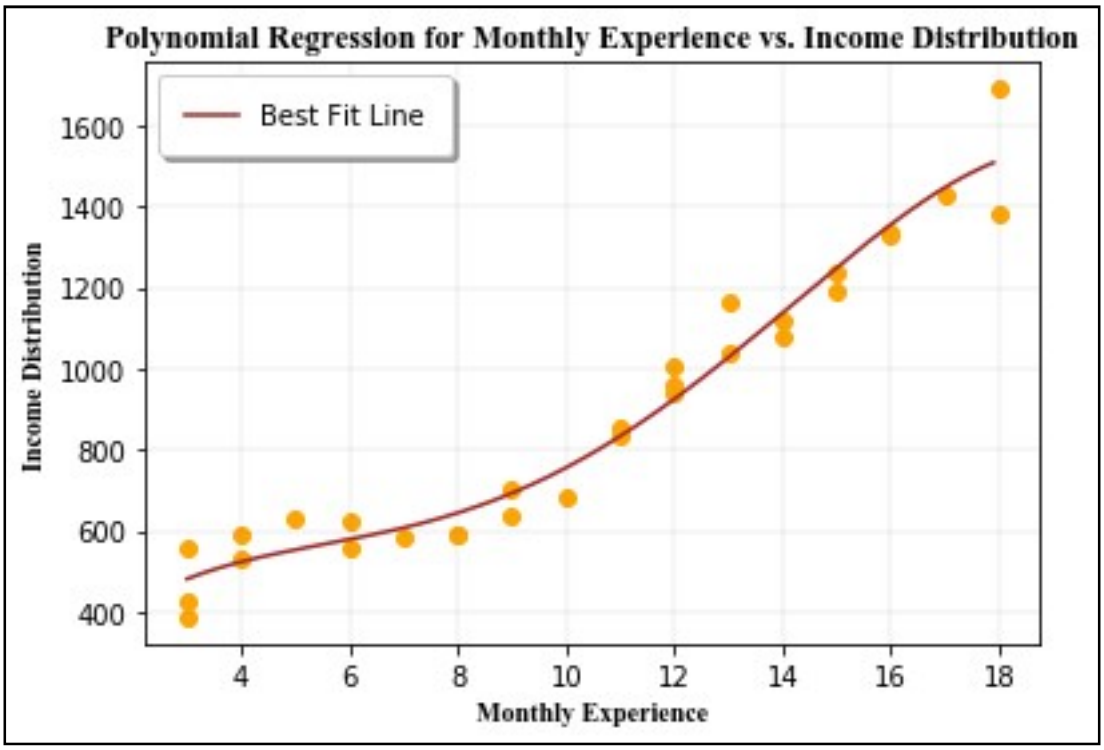
Variable explorer

Name	Type	Size	Value
dataset	DataFrame	(30, 2)	Column names: MonthsExperience, Income
p	int64	(30, 1)	[[3] [3]]
p_grid	float64	(168700, 1)	[[3.] [3.01]]
p_poly	float64	(30, 5)	[[1.00000e+00 3.00000e+00 9.00000e+00 2.70000e+… [1.000 ...
q	int64	(30, 1)	[[424] [387]]
tnrfont	dict	1	{'fontname':'Times New Roman'}

Plots

Linear Regression:

Polynomial Regression:



Decision Tree Regression: