```
! pip install pycaret
```

```
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in /usr/lo
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/di
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3
Requirement already satisfied: deprecated>=1.2.13 in /usr/local/lib/python3
Requirement already satisfied: scikit-base<0.6.0 in /usr/local/lib/python3.
Requirement already satisfied: Flask<2.3.0,>=1.0.4 in /usr/local/lib/python
Requirement already satisfied: Werkzeug<2.3.0 in /usr/local/lib/python3.10/
Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/li
Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/li
Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.
Requirement already satisfied: typing-extensions>=4.1.1 in /usr/local/lib/
Requirement already satisfied: retrying in /usr/local/lib/python3.10/dist-p
Requirement already satisfied: ansi2html in /usr/local/lib/python3.10/dist-
Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.10/di
Requirement already satisfied: wrapt<2,>=1.10 in /usr/local/lib/python3.10/
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/di
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/d
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10
Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.
Requirement already satisfied: itsdangerous>=2.0 in /usr/local/lib/python3.
Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.10/dist
Requirement already satisfied: pyzmq<25,>=17 in /usr/local/lib/python3.10/d
Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dis
Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/di
Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.1
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.
Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.1
Requirement already satisfied: jupyter-server>=1.8 in /usr/local/lib/python
Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/pytho
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-pac
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.
Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/pytho
Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-p
```

```
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/pytho
Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.1
Requirement already satisfied: websocket-client in /usr/local/lib/python3.
Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/di
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/d
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/d
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist
```

```
import pandas as pd
import numpy as np
from scipy import stats
from pycaret.regression import models,setup, compare_models, evaluate_model, pre
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, ca
```

```
train = pd.read_csv('/content/drive/MyDrive/lab-2/train.csv')
test = pd.read_csv('/content/drive/MyDrive/lab-2/test.csv')
```

```
train.head()
```

	id	store_sales(in millions)	unit_sales(in millions)	total_children	num_children_at_home
0	0	8.61	3.0	2.0	2.0
1	1	5.00	2.0	4.0	0.0
2	2	14.08	4.0	0.0	0.0
3	3	4.02	3.0	5.0	0.0
4	4	2.13	3.0	5.0	0.0

```
train.shape
```

```
(360336, 17)
```

```
train.columns
```

```
Index(['id', 'store_sales(in millions)', 'unit_sales(in millions)',
       'total_children', 'num_children_at_home', 'avg_cars_at
home(approx).1',
       'gross_weight', 'recyclable_package', 'low_fat', 'units_per_case',
       'store_sqft', 'coffee_bar', 'video_store', 'salad_bar',
'prepared_food',
       'florist', 'cost'],
      dtype='object')
```

```
train.describe()
```

	id	store_sales(in millions)	unit_sales(in millions)	total_children	num_chil
count	360336.000000	360336.000000	360336.000000	360336.000000	
mean	180167.500000	6.337376	3.043881	2.456482	
std	104020.187637	3.307980	0.784676	1.488992	
min	0.000000	0.510000	1.000000	0.000000	
25%	90083.750000	3.720000	3.000000	1.000000	
50%	180167.500000	5.780000	3.000000	2.000000	
75%	270251.250000	8.400000	4.000000	4.000000	
max	360335.000000	22.920000	6.000000	5.000000	

```
print(len(train))
train = train.drop_duplicates()
print(len(train))
```

```
360336
360336
```

```
correlations = train.corr()
income_correlation = correlations['cost']
income_correlation_ranked = income_correlation.abs().sort_values(ascending=False
```

```
income_correlation_ranked
```

```
cost                            1.000000
florist                         0.110414
video_store                     0.106786
prepared_food                   0.098843
salad_bar                       0.098810
coffee_bar                      0.052086
store_sqft                      0.049201
avg_cars_at home(approx).1      0.027098
unit_sales(in millions)         0.026509
store_sales(in millions)        0.012387
total_children                  0.007482
low_fat                         0.001975
num_children_at_home            0.001727
recyclable_package              0.001455
id                              0.000611
units_per_case                  0.000180
gross_weight                    0.000116
Name: cost, dtype: float64
```

```
top_features = income_correlation_ranked[1:-3]
selected_features = top_features.index.tolist()
train = train[selected_features + ['cost']]
test = test[selected_features]
```

```
train.head()
```

	florist	video_store	prepared_food	salad_bar	coffee_bar	store_sqft	h
0	0.0	0.0	0.0	0.0	0.0	36509.0	
1	0.0	0.0	0.0	0.0	1.0	28206.0	
2	0.0	0.0	0.0	0.0	1.0	21215.0	
3	0.0	0.0	0.0	0.0	1.0	21215.0	
4	1.0	1.0	1.0	1.0	1.0	27694.0	

```
reg = setup(train,target="cost",session_id=123,memory=False)
```

	Description	Value
0	Session id	123
1	Target	cost
2	Target type	Regression
3	Original data shape	(360336, 14)
4	Transformed data shape	(360336, 14)
5	Transformed train set shape	(252235, 14)
6	Transformed test set shape	(108101, 14)
7	Numeric features	13
8	Preprocess	True
9	Imputation type	simple
10	Numeric imputation	mean
11	Categorical imputation	mode
12	Fold Generator	KFold
13	Fold Number	10
14	CPU Jobs	-1
15	Use GPU	False
16	Log Experiment	False
17	Experiment Name	reg-default-name
18	USI	1194

```
models()
```

ID	Name	Reference	Turbo
lr	Linear Regression	sklearn.linear_model._base.LinearRegression	True
lasso	Lasso Regression	sklearn.linear_model._coordinate_descent.Lasso	True
ridge	Ridge Regression	sklearn.linear_model._ridge.Ridge	True
en	Elastic Net	sklearn.linear_model._coordinate_descent.Elast...	True
lar	Least Angle Regression	sklearn.linear_model._least_angle.Lars	True
llar	Lasso Least Angle Regression	sklearn.linear_model._least_angle.LassoLars	True
omp	Orthogonal Matching Pursuit	sklearn.linear_model._omp.OrthogonalMatchingPu...	True
br	Bayesian Ridge	sklearn.linear_model._bayes.BayesianRidge	True
ard	Automatic Relevance Determination	sklearn.linear_model._bayes.ARDRegression	False
par	Passive Aggressive Regressor	sklearn.linear_model._passive_aggressive.Passi...	True
ransac	Random Sample Consensus	sklearn.linear_model._ransac.RANSACRegressor	False
tr	TheilSen Regressor	sklearn.linear_model._theil_sen.TheilSenRegressor	False
huber	Huber Regressor	sklearn.linear_model._huber.HuberRegressor	True
kr	Kernel Ridge	sklearn.kernel_ridge.KernelRidge	False
svm	Support Vector Regression	sklearn.svm._classes.SVR	False
	K Neighbors		

```
# best = compare_models()


# best


xgboost = create_model("xgboost")
lightgbm = create_model("lightgbm")
gbr = create_model("gbr")
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	24.0064	794.1742	28.1811	0.1153	0.3006	0.2762
1	24.1754	803.3188	28.3429	0.1129	0.3027	0.2786
2	24.0972	801.1558	28.3047	0.1048	0.3035	0.2796
3	24.0313	799.1892	28.2699	0.1107	0.3029	0.2788
4	23.9901	794.1148	28.1800	0.1094	0.3011	0.2765
5	24.0624	798.1615	28.2518	0.1101	0.3015	0.2771
6	24.1828	805.0244	28.3730	0.1078	0.3033	0.2791
7	24.0655	797.5125	28.2403	0.1080	0.3035	0.2802
8	24.1248	801.2208	28.3058	0.1058	0.3012	0.2761
9	24.0647	799.2228	28.2705	0.1105	0.3018	0.2773
Mean	24.0801	799.3095	28.2720	0.1095	0.3022	0.2780
Std	0.0621	3.3723	0.0596	0.0030	0.0011	0.0014

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	24.2432	800.1166	28.2863	0.1087	0.3020	0.2793
1	24.3800	807.4705	28.4160	0.1083	0.3038	0.2814
2	24.2769	803.4234	28.3447	0.1023	0.3043	0.2819
3	24.3039	804.4828	28.3634	0.1049	0.3043	0.2822
4	24.1956	798.2353	28.2531	0.1048	0.3022	0.2793
5	24.3026	803.2576	28.3418	0.1044	0.3028	0.2802
6	24.3961	808.4463	28.4332	0.1040	0.3042	0.2817
7	24.2292	800.0525	28.2852	0.1052	0.3043	0.2824

7	24.2292	800.0525	28.2852	0.1052	0.3043	0.2824
8	24.3591	805.8085	28.3868	0.1007	0.3023	0.2790
9	24.3472	805.5850	28.3828	0.1035	0.3033	0.2809
Mean	24.3034	803.6879	28.3493	0.1046	0.3033	0.2808
Std	0.0640	3.1880	0.0562	0.0023	0.0009	0.0012

Fold	**MAE**	**MSE**	**RMSE**	**R2**	**RMSLE**	**MAPE**
0	24.9540	835.6777	28.9081	0.0691	0.3084	0.2877
1	25.0439	839.9818	28.9824	0.0724	0.3095	0.2890
2	24.9296	834.6230	28.8898	0.0674	0.3099	0.2896
3	24.9514	836.3991	28.9206	0.0693	0.3099	0.2897
4	24.8721	830.6970	28.8218	0.0684	0.3081	0.2872
5	24.9691	836.7638	28.9269	0.0671	0.3087	0.2879
6	25.0577	841.9485	29.0163	0.0668	0.3102	0.2896
7	24.8865	832.7636	28.8576	0.0686	0.3101	0.2901
8	24.9989	837.0093	28.9311	0.0658	0.3078	0.2865
9	24.9806	838.5326	28.9574	0.0668	0.3092	0.2884
Mean	24.9644	836.4396	28.9212	0.0682	0.3092	0.2886
Std	0.0571	3.1317	0.0541	0.0018	0.0008	0.0012

```
stacked_model = stack_models(estimator_list=[xgboost, lightgbm])
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	23.9474	791.7085	28.1373	0.1180	0.2999	0.2753
1	24.0903	799.5542	28.2764	0.1170	0.3018	0.2774
2	24.0177	797.6735	28.2431	0.1087	0.3027	0.2783
3	24.0054	797.0660	28.2324	0.1131	0.3023	0.2781
4	23.9233	791.7165	28.1375	0.1121	0.3005	0.2755
5	24.0149	796.0427	28.2142	0.1125	0.3009	0.2763
6	24.1178	801.4582	28.3100	0.1117	0.3025	0.2780
7	23.9827	793.6069	28.1710	0.1124	0.3026	0.2789
8	24.0782	798.9592	28.2659	0.1083	0.3006	0.2752
9	24.0355	797.1680	28.2342	0.1128	0.3013	0.2767
Mean	24.0213	796.4954	28.2222	0.1127	0.3015	0.2770
Std	0.0586	3.1068	0.0551	0.0029	0.0009	0.0013

```
tuned_model = tune_model(stacked_model)
```

Initiated	. . . . .	19:26:14
Status	. . . . .	Searching Hyperparameters
Estimator	. . . . .	Linear Regression

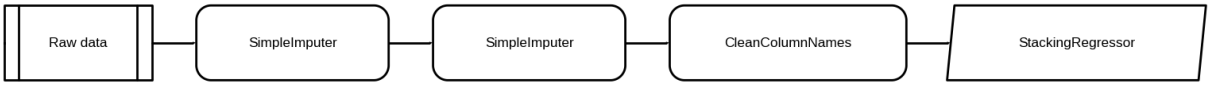
Processing: 0% 0/7 [00:00<?, ?it/s]

```
Fitting 10 folds for each of 2 candidates, totalling 20 fits
```

```
evaluate_model(stacked_model)
```

Plot Type:

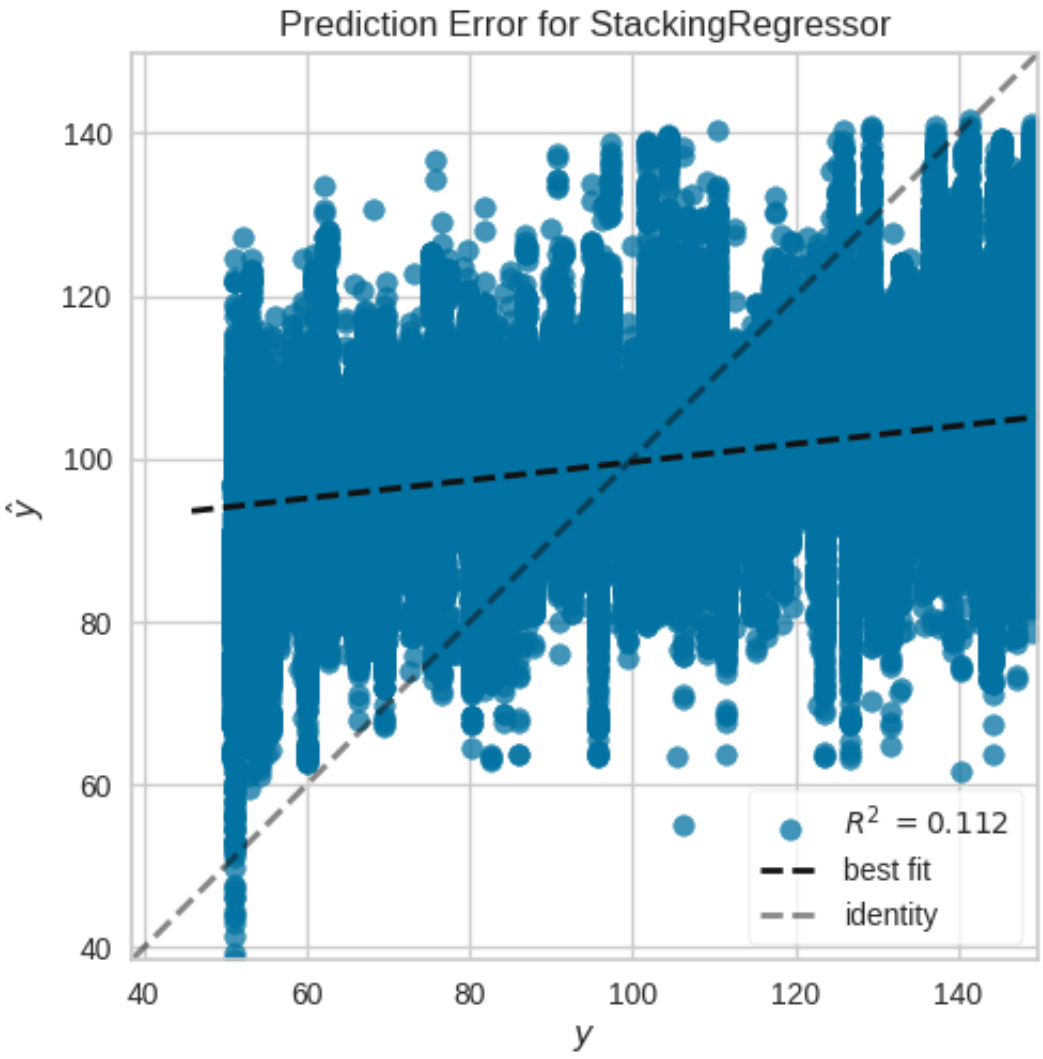
Pipeline Plot	Hyperparameters	Residuals	Prediction Error
Cooks Distance	Feature Selection	Learning Curve	Manifold Learning
Validation Curve	Feature Importance	Feature Importanc...	Decision Tree
Interactive Residuals			




```
predictions = predict_model(stacked_model, data=test)
```

```
# model_tuned = tune_model(stacked_model)
```

```
plot_model(stacked_model,plot="error")
```




```
test_pred= predictions[["prediction_label"]]
```

```
test_pred.rename(columns={'prediction_label': 'cost'}, inplace=True)
```

```
test_pred.head()
```

	cost
0	97.458560
1	97.837415
2	97.381243
3	99.794896
4	79.360686

```
og_test = pd.read_csv('/content/drive/MyDrive/lab-2/test.csv')
```

```
result_df = pd.concat([og_test['id'], test_pred['cost']], axis=1)
```

```
result_df.head()
```

	id	cost
0	360336	97.458560
1	360337	97.837415
2	360338	97.381243
3	360339	99.794896
4	360340	79.360686

```
result_df.to_csv('190249E_test_output.csv', index=False)
```