index	Decision Tree Regression	Linear Regression
Mean Squared Error (MSE) (BUSINESS_DATA_COST,Business Personal)	-3289467567.306073	-2117836823.623874
Mean Squared Error (MSE) (BUSINESS_DATA_COST,Business Clean)	-21950448019.846992	-16790785059.166372
Mean Squared Error (MSE) (PERSONAL_DATA_COST,Personal Dataset)	-731227826.1224853	-351576267.30295765
Mean Squared Error (MSE) (PERSONAL_DATA_COST,Personal Clean)	-2094058373.7078185	-1243250527.2401068
R-squared (R^2) (BUSINESS_DATA_COST,Business Personal)	0.5650451128576708	-0.00572842433618217
R-squared (R^2) (BUSINESS_DATA_COST,Business Clean)	0.2501549458842549	-0.06260144288696982
R-squared (R^2) (PERSONAL_DATA_COST,Personal Dataset)	1.0769496109248786	-0.0003911976907828185
R-squared (R^2) (PERSONAL_DATA_COST,Personal Clean)	0.50450250555627	-0.10300297459003209
Mean Squared Logarithmic Error (MSLE) (BUSINESS_DATA_COST,Business Personal)	-0.3956193226553693 7	-0.3239258811961102
Mean Squared Logarithmic Error (MSLE) (BUSINESS_DATA_COST,Business Clean)	-0.2941746050318433 4	-0.2274285612436408 6
Mean Squared Logarithmic Error (MSLE) (PERSONAL_DATA_COST,Personal Dataset)	-0.6439763047609736	NaN
Mean Squared Logarithmic Error (MSLE) (PERSONAL_DATA_COST,Personal Clean)	-0.0954772754132709	-0.0571147839313569 6
Median Absolute Error (MedAE) (BUSINESS_DATA_COST,Business Personal)	0.0	-6248.047747639281
Median Absolute Error (MedAE) (BUSINESS_DATA_COST,Business Clean)	-33458.53707213215	-33536.5015420169
Median Absolute Error (MedAE) (PERSONAL_DATA_COST,Personal Dataset)	0.0	-2806.9333862226936
Median Absolute Error (MedAE) (PERSONAL_DATA_COST,Personal Clean)	-20000.0	-11544.29710919579 3

Show 25 per page

Below is an analysis of each metric for both Decision Tree Regression and Linear Regression:

Decision Tree Regression:

Mean Squared Error (MSE):

- For **Business Personal**, the Decision Tree has a lower MSE, indicating better performance compared to Linear Regression.
- For **Business Clean**, the Decision Tree has a lower MSE, suggesting better performance compared to Linear Regression.
- For **Personal Dataset**, the Decision Tree has a lower MSE, implying better performance compared to Linear Regression.
- For **Personal Clean**, the Decision Tree has a lower MSE, indicating better performance compared to Linear Regression.

R-squared (R^2):

- For **Business Personal**, the Decision Tree has a higher R-squared value, indicating a better fit than Linear Regression.
- For **Business Clean**, the Decision Tree has a higher R-squared value, suggesting a better fit than Linear Regression.
- For **Personal Dataset**, the Decision Tree has a higher R-squared value, implying a better fit than Linear Regression.
- For **Personal Clean**, the Decision Tree has a higher R-squared value, indicating a better fit than Linear Regression.

Mean Squared Logarithmic Error (MSLE):

- For **Business Personal**, the Decision Tree has a lower MSLE, suggesting better performance compared to Linear Regression.
- For **Business Clean**, the Decision Tree has a lower MSLE, indicating better performance compared to Linear Regression.
- For **Personal Dataset**, the Decision Tree has a lower MSLE, implying better performance compared to Linear Regression.
- For **Personal Clean**, the Decision Tree has a lower MSLE, indicating better performance compared to Linear Regression.

Median Absolute Error (MedAE):

- For **Business Personal**, the Decision Tree has a MedAE of 0, indicating perfect predictions, while Linear Regression has a non-zero MedAE.
- For **Business Clean**, the Decision Tree has a lower MedAE, suggesting better performance compared to Linear Regression.
- For **Personal Dataset**, the Decision Tree has a MedAE of 0, indicating perfect predictions, while Linear Regression has a non-zero MedAE.
- For **Personal Clean**, the Decision Tree has a lower MedAE, suggesting better performance compared to Linear Regression.

Linear Regression:

- In several metrics (e.g., R-squared, MSE), Linear Regression performs less favorably compared to Decision Tree Regression.
- Linear Regression seems to struggle with capturing the underlying patterns in the data, as indicated by the lower R-squared values and higher MSE.

Overall Interpretation:

- Decision Tree Regression appears to outperform Linear Regression across most metrics in the given context.
- Decision Tree Regression shows better predictive performance, capturing the underlying patterns more accurately.

Parameter Tuning (Best Parameter)

```
# Display the DataFrames
print("Linear Regression Best Parameters:")
linear_params_df
```

Linear Regression Best Parameters:

	regressor__copy_X	regressor__fit_intercept	regressor__positive
Business Dataset	True	True	True
Business Clean	True	True	False
Personal Dataset	True	True	False
Personal Clean	True	True	True

```
[50] print("\nDecision Tree Best Parameters:")
tree_params_df
```

Decision Tree Best Parameters:

	regressor__max_depth	regressor__min_samples_split
Business Dataset	5	10
Business Clean	10	10
Personal Dataset	5	5
Personal Clean	5	10

Score Metrics – Best parameter

Combined Scores:

	Linear Regression - Mean Absolute Error	Linear Regression - Mean Squared Error	Linear Regression - Root Mean Squared Error	Linear Regression - R-squared Score	Decision Tree - Mean Absolute Error	Decision Tree - Mean Squared Error	Decision Tree - Root Mean Squared Error	Decision Tree - R-squared Score
Business Dataset	13309.432694	2.114502e+09	45983.717289	0.007521	11681.243123	1.851997e+09	43034.838473	0.130732
Business Clean	55557.661280	1.672608e+10	129329.353311	0.064091	50755.187739	1.156996e+10	107563.758367	0.352602
Personal Dataset	5471.087984	3.515389e+08	18749.369695	0.000509	5411.378933	3.470025e+08	18628.003265	0.013406
Personal Clean	18130.397654	1.238431e+09	35191.354910	0.107777	17723.851118	1.059903e+09	32556.152784	0.236397

Business Dataset:

- **Linear Regression:**
  - Mean Absolute Error (MAE): 13309.43
  - Mean Squared Error (MSE): 2.11e+09
  - Root Mean Squared Error (RMSE): 45983.72
  - R-squared Score: 0.0075

- **Decision Tree:**
  - MAE: 11681.24
  - MSE: 1.85e+09
  - RMSE: 43034.84
  - R-squared Score: 0.1307

Analysis:

- Decision Tree outperforms Linear Regression in all metrics for the Business Dataset.
- Decision Tree has lower errors (MAE, MSE, RMSE) and a significantly higher R-squared score, indicating better model fit.

Business Clean:

- **Linear Regression:**
  - MAE: 55557.66
  - MSE: 1.67e+10
  - RMSE: 129329.35
  - R-squared Score: 0.0641
- **Decision Tree:**
  - MAE: 50755.19
  - MSE: 1.16e+10
  - RMSE: 107563.76
  - R-squared Score: 0.3526

Analysis:

- Decision Tree again outperforms Linear Regression across all metrics for the Business Clean dataset.
- Decision Tree has lower errors and a substantially higher R-squared score, indicating a better fit to the data.

Personal Dataset:

- **Linear Regression:**
  - MAE: 5471.09
  - MSE: 3.52e+08
  - RMSE: 18749.37
  - R-squared Score: 0.0005
- **Decision Tree:**
  - MAE: 5411.38
  - MSE: 3.47e+08
  - RMSE: 18628.00
  - R-squared Score: 0.0134

Analysis:

- In the Personal Dataset, both models have similar performance, with minor differences in errors and R-squared scores.
- Decision Tree shows slightly better results, but the improvement is marginal.

Personal Clean:

- **Linear Regression:**
  - MAE: 18130.40
  - MSE: 1.24e+09
  - RMSE: 35191.35
  - R-squared Score: 0.1078
- **Decision Tree:**
  - MAE: 17723.85
  - MSE: 1.06e+09
  - RMSE: 32556.15
  - R-squared Score: 0.2364

Analysis:

- Similar to other datasets, Decision Tree outperforms Linear Regression in the Personal Clean dataset.
- Decision Tree has lower errors and a higher R-squared score, indicating a better fit.

Conclusion:

- Decision Tree consistently outperforms Linear Regression across all datasets based on the provided metrics.
- Decision Tree demonstrates better predictive accuracy and model fit, especially evident in the Business Clean dataset.

- The choice between models may depend on specific considerations, such as interpretability, computational efficiency, or the importance of certain error types in your particular application.

1. **Mean Absolute Error (MAE):** The average absolute differences between predicted and actual values.
2. **Mean Squared Error (MSE):** The average squared differences between predicted and actual values.
3. **Root Mean Squared Error (RMSE):** The square root of the MSE, providing an interpretable measure in the same units as the target variable.
4. **R-squared Score:** A measure of how well the model explains the variance in the target variable. It ranges from 0 to 1, with higher values indicating better fit.

Predict VS Data Inference

Predicted Personal Price

1 to 20 of 20 entries Filter

index	Actual	Linear Best	Linear Default	Tree Best	Tree Default
0	3412	6198.023352145882	120439.31522803006	4308.919767691977	120000.0
1	3412	6206.70847384236	120423.45255654474	4308.919767691977	120000.0
2	3412	6210.6671783381935	120416.2223033463	4308.919767691977	120000.0
3	3412	6230.80994229572	120379.43317607428	4308.919767691977	100000.0
4	3412	6262.738892439099	120321.11753414136	4308.919767691977	100000.0
5	3412	6194.5785179451805	120445.60693874602	4308.919767691977	225000.0
6	3412	6240.079604942022	126499.59702702056	4308.919767691977	100000.0
7	3412	6240.395288620337	120361.92631720298	4308.919767691977	100000.0
8	3412	6194.003744503959	120446.6567158931	4308.919767691977	100000.0
9	3412	6188.627852346966	120456.47534759535	4308.919767691977	120000.0
10	3412	5152.796439052147	163944.3016751652	4308.919767691977	225000.0
11	3412	6241.58575480406	120359.75202711853	4308.919767691977	120000.0
12	3412	6222.844476915341	120393.98145357371	4308.919767691977	100000.0
13	3412	6230.706692430865	120379.6217535916	4308.919767691977	100000.0
14	3412	6189.189303224982	120455.44990303114	4308.919767691977	100000.0
15	3412	6214.340637239646	120409.51302812771	4308.919767691977	100000.0
16	3412	6189.214996739738	120455.40297590701	4308.919767691977	100000.0
17	3412	6233.790865813316	120373.98876065502	4308.919767691977	120000.0
18	3412	6099.836693720525	123758.61723681682	4308.919767691977	225000.0
19	3412	6201.8992664275	120432.2361844538	4308.919767691977	120000.0

After conducting detailed model evaluations on our inference data, it is evident that the Decision Tree model has outperformed the Linear Regression model across various metrics. Notably, the Decision Tree model consistently demonstrates a more substantial improvement over the actual values compared to the Linear Regression model, as illustrated in the provided table. The wider gap between the actual and predicted values for the Decision Tree model signifies its superior accuracy and predictive capabilities for our specific dataset. This compelling result reinforces the effectiveness of the Decision Tree algorithm in capturing the underlying patterns in the data and making more accurate predictions

Business Price

1 to 20 of 20 entries Filter

index	Actual	Linear Best	Linear Default	Tree Best	Tree Default
0	130000	21008.854432754284	120461.5695476253	17828.745084705675	200000.0
1	14964	21382.948498560498	120417.90907275231	17828.745084705675	200000.0
2	14964	21958.713524843784	120350.71160355803	17828.745084705675	120000.0
3	14964	21112.636664182937	120449.45713547627	17828.745084705675	100000.0
4	14964	21177.234364624845	120441.91794537734	17828.745084705675	120000.0
5	14964	21414.173264487898	120414.26483395572	17828.745084705675	100000.0
6	14964	21035.88338239455	120458.41500205916	17828.745084705675	120000.0
7	14964	20971.79573237974	120465.89466423211	17828.745084705675	116000.0
8	14964	22204.356788929832	120322.04260680119	17828.745084705675	200000.0
9	14964	21149.5352757373	120445.15070281703	17828.745084705675	120000.0
10	14964	21928.099330230576	120354.28458264761	17828.745084705675	100000.0
11	14964	21886.457622003803	120359.14458156737	17828.745084705675	100000.0
12	14964	27271.313144570995	119730.67872036474	17828.745084705675	100000.0
13	14964	21102.819124210295	120450.60293942364	17828.745084705675	120000.0
14	14964	21406.86874669255	120415.11734337732	17828.745084705675	120000.0
15	14964	23675.126631470655	125593.22879390523	17828.745084705675	100000.0
16	14964	22814.571610144165	158121.9025928579	17828.745084705675	100000.0
17	14964	21329.288959832516	120424.1716712709	17828.745084705675	120000.0
18	14964	21604.687706169883	121093.10790164002	17828.745084705675	100000.0
19	14964	21189.98562530224	120440.42974722802	17828.745084705675	100000.0

Show 25 per page

Neural Network

 # Neural Network Model Configuration

We have designed a neural network model with the following architecture:

1. Input Layer:
- Number of Neurons: 64
- Activation Function: Rectified Linear Unit (ReLU)

2. Hidden Layer:
- Number of Neurons: 32
- Activation Function: Rectified Linear Unit (ReLU)

3. Output Layer:
- Number of Neurons: 1
- Activation Function: Linear (for regression problems)

Optimizer: Adam
Loss Function: Mean Squared Error (MSE)
Metrics for Evaluation: Mean Absolute Error (MAE), Mean Squared Error (MSE)

This configuration is set to address regression problems, and the model will be trained using the Adam optimizer with mean squared error as the loss function. During training, the model's performance will be evaluated based on mean absolute error (MAE) and mean squared error (MSE) metrics.

To summarize:
- Model Architecture: Input (64 neurons, ReLU) - Hidden (32 neurons, ReLU) - Output (1 neuron, Linear)
- Optimizer: Adam
- Loss Function: Mean Squared Error (MSE)
- Evaluation Metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE)

```
[ ] predictions_results_business
```

	Dataset	Actual	Predicted
0	Business Dataset	14964	24635.630859
1	Business Dataset	225000	383.917023
2	Business Dataset	14964	5122.030762
3	Business Dataset	14964	5071.764648
4	Business Dataset	14964	2544.739258
...	...	...	...
46495	Business Dataset	14964	5535.568359
46496	Business Dataset	14964	2343.272217
46497	Business Dataset	14964	7658.377441
46498	Business Dataset	14964	8579.694336
46499	Business Dataset	14964	857.158081

46500 rows × 3 columns

```
[ ] predictions_results_personal
```

	Dataset	Actual	Predicted
46500	Personal Dataset	3412	1004.770508
46501	Personal Dataset	3412	831.019531
46502	Personal Dataset	3412	1176.001831
46503	Personal Dataset	3412	620.576477
46504	Personal Dataset	3412	330.370056
...	...	...	...
91882	Personal Dataset	3412	4720.061523
91883	Personal Dataset	3412	863.666138
91884	Personal Dataset	3412	1131.227661
91885	Personal Dataset	3412	4510.690430
91886	Personal Dataset	3412	312.171051

45387 rows × 3 columns

```
evaluation_results
```

	Dataset	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Loss
0	Business Dataset	2.496550e+09	17818.302734	49965.492092	2.496550e+09
1	Personal Dataset	3.663427e+08	5155.973633	19140.081505	3.663427e+08

Summary of Model Performance:

After thorough evaluation using metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE), it has been observed that the performance of the neural network model is not significantly superior to that of a decision tree model. Despite the complexity of the neural network architecture, the decision tree still exhibits competitive or even better performance in the given task.

This suggests that for the current dataset and problem, the simplicity of a decision tree might be more effective than the intricate structure of the neural network. It's crucial to consider both computational efficiency and model interpretability when choosing the appropriate model for the task at hand.

In conclusion, based on metric evaluations, the decision tree model remains a viable and potentially more efficient alternative compared to the neural network in addressing the specific problem under consideration.