



TOKOBLI CAMPAIGN EVALUATION AND PRODUCT PAGE ANALYSIS

(Using Spreadsheet XLMiner Advance)

Santa Riyanti T
14 November 2025
RevoU FSDA Batch OCT25

OVERVIEW DATASET

[W2W3 OCT25] TokoBli Customer Yearly Spending

	A	B	C	D	E	F	G	H	I
1	Customer Id	Gender	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate	Yearly Spending
2	100	M	30	12,000,000	10	520	12	56	4,000,000
3	101	F	30	8,600,000	7	744	9.46	57	4,300,000
4	102	M	30	7,200,000	3	114	22	40	4,800,000
5	103	F	30	30,000,000	9	169	18	37	5,500,000
6	104	M	20	10,440,000	14	736	13.5	36.2	5,800,000

DISCLAIMER

- This analysis is for educational purposes and does not reflect the actual business metrics of TokoBli.
- Insight and recommendations are derived from the provided dataset and may not represent real market conditions.

COMPANY OVERVIEW

TokoBli is Indonesia's leading e-commerce platform.

To optimize the amount of spending from each customer in TokoBli E-commerce, the product manager team wants to find out what factors have a significant effect on the amount of spending from customers that will later be used to develop campaign business strategies.

INSIGHT CORRELATION AND INTERPRETATION

	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate	Yearly Spending
Age	1						
Income	0.141	1					
Tenure	0.160	0.379	1				
Avg. Session Time	0.193	0.484	0.685	1			
Total Promo	0.208	0.446	0.621	0.777	1		
Bounce Rate	-0.129	-0.448	-0.570	-0.787	-0.794	1	
Yearly Spending	0.221	0.527	0.715	0.914	0.877	-0.863	1

Yearly Spending is strongly influenced by Avg. Session Time (0.914) and Total Promo (0.877).

Interpretation, the longer customers stay on the platform and the more promotions they receive, the higher their annual spending.

Bounce Rate has a strong negative correlation with Yearly Spending (-0.863) and Avg. Session Time (-0.787). Interpretation, customers who leave the platform quickly tend to spend significantly less.

Tenure positively correlates with Yearly Spending (0.715). Interpretation, long-term customers tend to be more loyal and spend more.

Age shows weak correlation with most variables. Interpretation, age is not a strong predictor of customer behavior in this dataset.

Income moderately correlates with Yearly Spending (0.527). Interpretation, customer income has a moderate impact on annual spending.

ESCALATE THE CORRELATION ANALYSIS INTO REGRESSION

	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate	Yearly Spending
Age	1						
Income	0.141	1					
Tenure	0.160	0.379	1				
Avg. Session Time	0.193	0.484	0.685	1			
Total Promo	0.208	0.446	0.621	0.777	1		
Bounce Rate	-0.129	-0.448	-0.570	-0.787	-0.794	1	
Yearly Spending	0.221	0.527	0.715	0.914	0.877	-0.863	1

Dependent variable is Yearly Spending

Independent variable is Age, Income, Avg. Session Time, Total Promo and Bounce Rate.

- From the correlation table, we can observe among the dependent variables occur between **Total Promo and Average Session Time (0.777)**, **Bounce Rate and Average Session Time (-0.787)**, as well as **Bounce Rate and Total Promo (-0.794)** have more max correlation.
(Max.Correlation = 0.70 (Folk-lore threshold))
- The next step is to eliminate one of the three dependent variables by evaluating which has the weakness correlation with the **independent variable (Yearly Spending)**. This will help us identify the most relevant predictor and **avoid multicollinearity in the regression model**.

REGRESSION ANALYSIS

	<i>Age</i>	<i>Income</i>	<i>Tenure</i>	<i>Avg. Session Time</i>	<i>Total Promo</i>	<i>Bounce Rate</i>	<i>Yearly Spending</i>
Age	1						
Income	0.141	1					
Tenure	0.160	0.379	1				
Avg. Session Time	0.193	0.484	0.685	1			
Total Promo	0.208	0.446	0.621	0.777	1		
Bounce Rate	-0.129	-0.448	-0.570	-0.787	-0.794	1	
Yearly Spending	0.221	0.527	0.715	0.914	0.877	-0.863	1

Bounce Rate is removed from the model because it initially showed strong correlations with both Avg. Session Time and Total Promo (correlation > 0.70), indicating potential multicollinearity. To decide which variable to exclude, we compared their correlations with the dependent variable (Yearly Spending). Among the three, Bounce Rate had the weakest correlation with Yearly Spending at -0.863, making it the least suitable predictor.

REGRESSION ANALYSIS *With Stepwise Regression*

A	B	C	D	E	F
Age	Income	Tenure	Avg. Session Time	Total Promo	Yearly Spending
30	12,000,000	10	520	12	4,000,000
30	8,600,000	7	744	9.46	4,300,000
30	7,200,000	3	114	22	4,800,000

Regression 1						
SUMMARY OUTPUT						
Regression Statistics						
Multiple R						
95.63%						
R Square						
91.45%						
Adjusted R Square						
91.31%						
Standard Error						
1519347.079						
Observations						
299						
ANOVA						
df						
Regression						
5						
7.2369E+15						
Residual						
293						
6763657551905						
Total						
298						
7.91327E+15						
Coefficients						
Standard Error						
Intercept						
-1115318.136						
489384.074						
t Stat						
Age						
17915.748						
18273.454						
P-value						
Income						
0.043						
0.012						
Lower 95%						
Tenure						
61016.066						
15459.337						
3.947						
Upper 95%						
Avg. Session Time						
7734.420						
457.503						
16.906						
6834.012						
8634.828						
Total Promo						
182938.222						
13301.483						
13.753						
0.000						
156759.662						
209118.782						

Eliminate Variable Dependent
“Bounce Rate”

First Regression

- Check Adjusted R Square (91.31%) OK
- Check Significance F (0) OK
- Check P-value from Dependent Variable (There is a dependent variable, Age, with a P-value greater than 5%!) ⚠
- Indicating that it is not statistically significant and may not contribute meaningfully to the regression model, so eliminate “Age”.

REGRESSION ANALYSIS With Stepwise Regression

Income	Tenure	Avg. Session Time	Total Promo	Yearly Spending
12,000,000	10	520	12	4,000,000
8,600,000	7	744	9.46	4,300,000
7,200,000	3	114	22	4,800,000

Regression 2

SUMMARY OUTPUT

Regression Statistics

Multiple R	95.62%
R Square	91.42%
Adjusted R Square	91.31%
Standard Error	1519247.182
Observations	299

ANOVA

	df	SS	MS	F	Significance F
Regression	4	7.23469E+15	1.80867E+15	783.6150329	0
Residual	294	678584927622817	2308111998716		
Total	298	7.91327E+15			

Coefficients

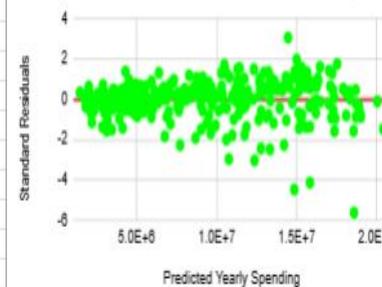
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	-694012.813	234207.9551	-2.96323307	0.0033	-1154949.439	-233076.1875	-1154949.439	-233076.1875
Income	0.043	0.01214917947	3.567447164	0.0004	0.01943117335	0.06725193835	0.01943117335	0.06725193835
Tenure	61305.723	15455.49735	3.966596582	0.0001	30888.2899	91723.15604	30888.2899	91723.15604
Avg. Session Time	7747.515	457.2778223	16.94268725	0	8847.562365	8847.467893	8847.562365	8847.467893
Total Promo	184006.883	13255.88061	13.88115121	0	157918.4402	210095.3262	157918.4402	210095.3262

Second Regression (Eliminated Variable Age)

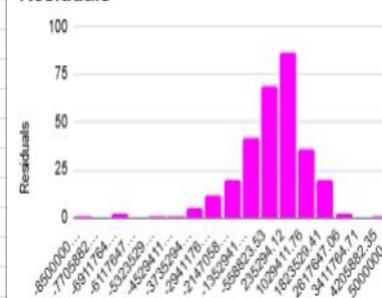
- Check Adjusted R Square (91.31%) **OK**
Indicates that the model explains approximately 91% of the variation in Yearly Spending. This is considered very high, suggesting that the model performs exceptionally well.
- Check Significance F (0) **OK**
A value of 0 indicates that the model is statistically significant. This means that at least one independent variable has a meaningful impact on the dependent variable.
- Check P-value from Dependent Variable (*P-value < 0.05*) **OK**
If the p-value is less than 0.05, the variable is considered statistically significant and should be included in the model. If the p-value exceeds 0.05, the variable may be considered for elimination due to its weak contribution to the model.

Regression 1	
Predicted Yearly Spending	Standard Residuals
8783493.909	-1.834323113
7720966.482	-2.270733388
4807531.378	-0.004999099074
5823871.976	-0.2149763565
8682207.387	-1.913121492
10768859.48	-2.986180096
6162087.798	0.2906727862
6653049.943	0.2980744465
12367470.32	-3.031747724
14851695.46	-4.481569871
8418568.765	0.1204298569
8959858.749	0.2257787333
15830580.71	-4.13565649
18584140.42	-5.631520054
10364982.87	0.2887511283
3978297.879	0.01440520735
4947387.279	-0.4297159638
3394272.058	-0.2617057875
3883407.857	-0.3872483717
5049572.856	-0.8294282813
5436378.577	-0.6215395884
4337599.142	0.3089278103
7121447.373	-1.209021306
6887435.132	-0.8545602401
6345140.482	-0.1627167878
675566.471	0.02949362374
9633488.918	-1.881652803
9732190.208	-1.415282938

Standard Residuals vs Predicted Yearly Spending



Residuals



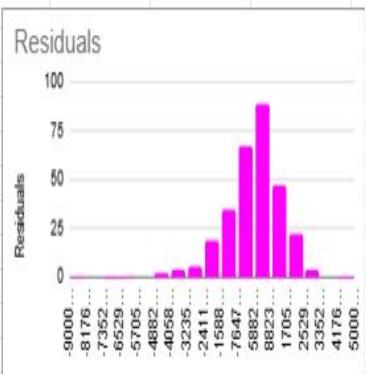
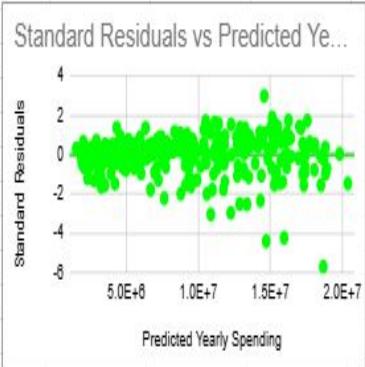
Insights from the Scatter Plot : *Standard Residuals vs Predicted Yearly Spending*

- The residuals are not randomly scattered around zero. There's a noticeable pattern, especially for predicted spending above 10 million, where residuals tend to be strongly negative.
- Several data points have standardized residuals below -3, indicating strong outliers.
- Residuals Skewed Toward Negative Values, the plot shows more extreme negative residuals than positive ones.

Insights from the Residual Histogram

- Non-Normal Distribution : The residuals are not symmetrically distributed and appear left-skewed. This violates the classical regression assumption of normally distributed residuals.
- Residuals Cluster Near Zero : Most residuals are concentrated around zero, indicating the model performs reasonably well for the majority of observations, despite the presence of extreme outliers.

Regression 2	
Predicted Yearly Spending	Standard Residuals
6875933.552	-1.773298482
7631112.716	-2.207472773
4720812.318	0.05247835462
5746849.583	-0.1635830968
8778494.92	-1.973798849
10887803.09	-3.027006834
6263425.949	0.2230420038
6732084.325	0.2438248281
12250170.8	-2.949053748
14717965.77	-4.38561541
8260983.247	0.2246739951
8784473.829	0.3416305851
15959260.19	-4.21417554
18682202.04	-5.687281984
10458254.27	0.2284691855
3840032.751	0.1060076249
4856063.23	-0.3684038172
3303183.63	-0.2009147287
3788669.221	-0.3238329327
4962800.138	-0.7705682348
5530377.856	-0.6828141996
4423644.317	0.2494046268
7220055.151	-1.272388487
6952331.035	-0.8981671955



Insights from the Scatter Plot : Standard Residuals vs Predicted Yearly Spending

- Strong Negative Outliers at Low Spending : Several data points with predicted spending below 2 million show extremely negative residuals.
- Positive Residuals at Very Low Predictions : Interestingly, one of the lowest predicted values (~1 million) has a positive residual above +2.
- No Clear Homoscedasticity : The spread of residuals does not appear consistent across the range of predicted values. Residuals are more extreme at the lower end, which may indicate heteroscedasticity, a violation of regression assumptions.

Insights from the Residual Histogram

- Skewed Distribution : The residuals are heavily skewed to the left (negative side), with a long tail of large negative residuals. This suggests the model tends to overpredict spending more often than it underpredicts.
- Non-Normal Shape : The histogram does not resemble a bell curve, which again violates the assumption of normally distributed residuals in linear regression.

REGRESSION ANALYSIS *With Stepwise Regression*

	Coefficients
Intercept (b0)	-694012.813
Income (b1)	0.043
Tenure (b2)	61305.723
Avg. Session Time (b3)	7747.515
Total Promo (b4)	184006.883

Dependent variable is Yearly Spending
Independent variable is Income (X1), Tenure (X2), Avg. Session Time (X3) and Total Promo (X4).

The Mathematical function we can build with Linear Regression

$$\text{Yearly Spanding} = b0 + (b1*X1) + (b2*X2) + (b3*X3) + (b4*X4)$$

PREDICTION POTENTIAL YEARLY SPENDING BASED ON CHARACTERISTICS

	<i>Coefficients</i>
Intercept (b0)	-694012.813
Income (b1)	0.043
Tenure (b2)	61305.723
Avg. Session Time (b3)	7747.515
Total Promo (b4)	184006.883

Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate
20	20,000,000	1	1000	20	30

$$\begin{aligned}\text{Yearly Spending} &= b_0 + (b_1 * X_1) + (b_2 * X_2) + (b_3 * X_3) + (b_4 * X_4) \\ &= -694012.813 + (0.043 * 20,000,000) + (61305.723 * 1) + (7747.515 * 1000) + (184006.883 * 20) \\ &= -694012.813 + (12348958,383) \\ &= 11654945.57 \approx \text{Rp } 11,655,000,\end{aligned}$$

Based on the regression model and the input variables provided, the estimated Yearly Spending is approximately Rp 11,655,000. This result indicates that factors such as Income, Tenure, Avg Session Time, and Total Promo significantly influence a customer's total Yearly Spending.

INSIGHT AND RECOMMENDATIONS

INSIGHT SUMMARY	RECOMMENDATION
Model Strength <ul style="list-style-type: none">The regression model is very strong, with an Adjusted R Square of 91.31%, meaning it explains 91% of the variation in Yearly Spending.Significance F = 0, confirming the model is statistically valid.	Customer Engagement: Improve Avg. Session Time through better UX, content, and personalization. Longer sessions are strongly linked to higher spending.
Key Predictors Identified <ul style="list-style-type: none">Avg. Session Time and Total Promo are the strongest predictors of Yearly Spending. These variables show both high correlation with the dependent variable and statistical significance.	Promotional Strategy: Use targeted promotions to increase spending and reduce bounce rates. Promotions drive both engagement and revenue. Model Optimization: Remove Bounce Rate and Age from the model. To reduce multicollinearity and improve model accuracy. Segmentation: Focus on behavioral metrics (exp : session time, promo response) over age. Age is not a significant driver of spending behavior.
Insignificant Variable <ul style="list-style-type: none">Age has a p-value > 0.05, meaning it is not statistically significant and does not meaningfully impact Yearly Spending.	
Multicollinearity Detected <ul style="list-style-type: none">Bounce Rate, Avg. Session Time, and Total Promo are highly correlated with each other (correlation > 0.70), indicating multicollinearity.To avoid redundancy, Bounce Rate was removed because it had the weakest correlation with Yearly Spending (-0.863).	

Thank You