Size of Purchase	Company Size	% Imports	Distance	Central Purchaser
27.9	25.6	41	18	1
89.6	109.8	16	75	0
12.8	39.4	29	14	0
34.9	16.7	31	117	0
408.6	278.4	14	209	1
173.5	98.4	8	114	1
105.2	101.6	20	75	0
510.6	139.3	17	50	1
382.7	207.4	53	35	1
84.6	26.8	27	15	1
101.4	13.9	31	19	0
27.6	6.8	22	7	0
234.8	84.7	5	89	1
464.3	180.3	27	306	1
309.8	132.6	18	73	1
294.6	118.9	16	11	1

Sales	Hours Worked	Number of Customers
15.6	44	54
15.7	43	52
15.4	41	55
14.3	41	55
11.8	40	39
9.7	40	28
9.6	40	37
10.2	38	58
11.3	38	67
14.3	32	186
14.8	37	226

Sales	Number of Employees
20.2	120
24.3	122
28.6	127
33.7	135
35.2	142
35.9	156
36.3	155
36.2	167
36.5	183
36.6	210

Size of Purchase	Company Size	% Imports	Distance	Central Purchaser
27.9	25.6	41	18	1
89.6	109.8	16	75	0
12.8	39.4	29	14	0
34.9	16.7	31	117	0
408.6	278.4	14	209	1
173.5	98.4	8	114	1
105.2	101.6	20	75	0
510.6	139.3	17	50	1
382.7	207.4	53	35	1
84.6	26.8	27	15	1
101.4	13.9	31	19	0
27.6	6.8	22	7	0
234.8	84.7	5	89	1
464.3	180.3	27	306	1
309.8	132.6	18	73	1
294.6	118.9	16	11	1

#1: "Full" Model

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.879085014
R Square	0.772790461
Adjusted R Square	0.690168811
Standard Error	94.43477087
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	333650.4689	83412.61723	9.353365087	0.001515424
Residual	11	98097.18545	8917.92595		
Total	15	431747.6544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*	
Intercept	-1.778759417	69.21609014	-0.025698641	0.979957993	-154.1223467	150.5648278	-154.1223467	150.5648278	
Company Size	1.373539646	0.441244957	3.1128733	0.009874722	0.402366043	2.344713249	0.402366043	2.344713249	
% Imports	-0.320557877	2.064733706	-0.155253859	**0.879433437**	-4.865006124	4.223890371	-4.865006124	4.223890371	<---------- Variable with the highest p-value. Insignificant. To be dropped from the model
Distance	0.111014834	0.378919917	0.292977037	0.774992657	-0.72298228	0.945011948	-0.72298228	0.945011948	
Central Purchaser	110.4337388	57.45373817	1.922133221	0.080859506	-16.02108629	236.8885639	-16.02108629	236.8885639	

Summary:

Strength of the model:
The R Square is 0.7728. The relatively high R Squared is indicating that approximately 77.28% of the variability in the size of purchases can be explained by the model.

ANOVE Table Insights:
F-statistic is 9.35. This indicates that the overall model is statistically significant (P-value = 0.0015), meaning at least one predictor contributes to explaning the variability in purchase size.

The Standard Error is 94.43. This is an error that is not negligible but also not exessively large. For small purchase size (e.g. 12.8, 27.6), the standard error exceeds the value itself. This suggests poor accuracy for predictions involving smaller purchases. For larger purchase size (e.g. 510.6), the standard error is smaller relative to the magnitude, making the model more reliable for such cases.

Significant Predictor Variable:
Among the 4 predictor variables, "company size" is significant at the 5% significance level. The coefficient of 1.374 indicates that every $1 million sales in company size would increase the size of purchase by $1374, on average, ceteris paribus.

At 10% significance level, "central purchaser" would be a significant predictor. The coefficient of 110.434 indicates that companies with a central purchaser would purchase $110,434 more than companies withut a central purchaser, on average, ceteris paribus.

"% Imports" and "Distance" are not among the significant predictor variables.

#2: "Full" Model - % Imports

Size of Purchase	Company Size	Distance	Central Purchaser
27.9	25.6	18	1
89.6	109.8	75	0
12.8	39.4	14	0
34.9	16.7	117	0
408.6	278.4	209	1
173.5	98.4	114	1
105.2	101.6	75	0
510.6	139.3	50	1
382.7	207.4	35	1
84.6	26.8	15	1
101.4	13.9	19	0
27.6	6.8	7	0
234.8	84.7	89	1
464.3	180.3	306	1
309.8	132.6	73	1
294.6	118.9	11	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.878801791
R Square	0.772292588
Adjusted R Square	0.715365735
Standard Error	90.51341567
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	333435.5134	111145.1711	13.56640228	0.000366694
Residual	12	98312.141	8192.678417		
Total	15	431747.6544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*	
Intercept	-10.19581789	41.24365383	-0.247209375	0.808925025	-100.05802	79.66638422	-100.05802	79.66638422	
Company Size	1.370177955	0.422412965	3.243602948	0.007038269	0.449819168	2.290536743	0.449819168	2.290536743	
Distance	0.123093227	0.355448189	0.346304273	**0.73511033**	-0.651361848	0.897548301	-0.651361848	0.897548301	<---------- Variable with the highest p-value. Insignificant. To be dropped from the model
Central Purchaser	110.9294421	54.98290177	2.017526149	0.066573838	-8.868009715	230.7268939	-8.868009715	230.7268939	

Summary:

Strength of the model:
The R Square dropped slightly from 0.7728 to 0.7723. The adjusted R Square, however, improved slightly from 0.6902 to 0.7154. This indicates a beterr fit after exclusing an irrelevant variable.

ANOVE Table Insights:
F-statistic is 13.57. This indicates that the overall regression model is statistically significant (P-value = 0.00037). This means at least one of the predictors contributes to explaining the variability in purchase size.

The Standard Error is 90.51, which is slightly lower than the previous model's 94.43. This indicates that the predictions in this model are slightly more precise. Although this value remains large for small purchase size, this model is more reliable than the previous model.

Significant Predictor Variable:
Among the 3 predictor variables, "company size" remains significant at the 5% significance level. The coefficient of 1.371 indicates that every $1 million sales in company size would increase the size of purchase by $1371, on average, ceteris paribus.

At 10% significance level, "central purchaser" remains as a significant predictor. The coefficient of 110.929 indicates that companies with a central purchaser would purchase $110,929 more than companies withut a central purchaser, on average, ceteris paribus.

"Distance" remains as an insignificant predictor variable.

#3: "Full" Model - % Imports - Distance

Size of Purchase	Company Size	Central Purchaser
27.9	25.6	1
89.6	109.8	0
12.8	39.4	0
34.9	16.7	0
408.6	278.4	1
173.5	98.4	1
105.2	101.6	0
510.6	139.3	1
382.7	207.4	1
84.6	26.8	1
101.4	13.9	0
27.6	6.8	0
234.8	84.7	1
464.3	180.3	1
309.8	132.6	1
294.6	118.9	1

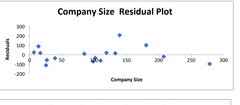
SUMMARY OUTPUT

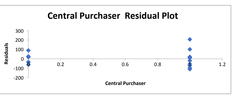
Regression Statistics	
Multiple R	0.877506072
R Square	0.770016906
Adjusted R Square	0.734634891
Standard Error	87.39594317
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	332452.9929	166226.4964	21.76294699	7.09616E-05
Residual	13	99294.66148	7638.050883		
Total	15	431747.6544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-7.771082382	39.24508251	-0.198014169	0.84609681	-92.55492855	77.01276379	-92.55492855	77.01276379
Company Size	1.450820591	0.340290254	4.263479708	0.000923776	0.715668193	2.185972989	0.715668193	2.185972989
Central Purchaser	109.4070292	52.91920754	2.067435139	**0.059203564**	-4.917968076	223.7320265	-4.917968076	223.7320265



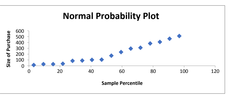


RESIDUAL OUTPUT

Observation	*edicted Size of Purcha*	*Residuals*
1	138.776954	-110.876954
2	151.5290185	-61.9290185
3	49.3912489	-36.5912489
4	16.45762149	18.44237851
5	505.5443993	-96.94439934
6	244.396693	-70.89669298
7	139.6322897	-34.43228965
8	303.7352551	206.8647449
9	402.5361374	-19.83613738
10	140.5179387	-55.91793867
11	12.39532383	89.00467617
12	2.094497636	25.50550236
13	224.5204509	10.27954912
14	363.2188994	101.0811006
15	294.0147572	15.78524281
16	274.1385151	20.46148491

PROBABILITY OUTPUT

Percentile	*Size of Purchase*
3.125	12.8
9.375	27.6
15.625	27.9
21.875	34.9
28.125	84.6
34.375	89.6
40.625	101.4
46.875	105.2
53.125	173.5
59.375	234.8
65.625	294.6
71.875	309.8
78.125	382.7
84.375	408.6
90.625	464.3
96.875	510.6



Summary:

Strength of the model:
The R Square dropped slightly from 0.7723 to 0.7700. The adjusted R Square, however, improved slightly from 0.7154 to 0.7346. This shows further improvement compared to previous models, which have an adjusted R square of 0.6902 and 0.7154, respectively).

ANOVA Table Insights: <---------- **Model to keep!**
F-statistic is 21.76. This indicates that the overall regression model is highly statistically significant (P-value = 7.0962e-05). This suggests that the included predictors contribute significantly to explaining variability in purchase size.

Note: We decided to keep this model because the p-value of "Central Purchaser" is only slightly above the 5% significant level, but this model has a higher R-squared and Adj. R-Squared compared to the model below (i.e. dropping "Central Purchaser".

Also, we did not introduce non-linearity/binary variable/interaction variables into this model, since that didnt seem to be what the question is asking for. Please let us know if we should still attempt those in an interview, even if the question did not ask for it. Thanks!

The Standard Error is 87.40. This is the lowest among all models, indicating further improvement in prediction precision.

Significant Predictor Variable:
Among the 2 predictor variables, "company size" remains significant at the 5% significance level. The coefficient of 1.451 indicates that every $1 million sales in company size would increase the size of purchase by $1451, on average, ceteris paribus.

At 10% significance level, "central purchaser" remains as a significant predictor. The coefficient of 109.407 indicates that companies with a central purchaser would purchase $109,407 more than companies withut a central purchaser, on average, ceteris paribus.

Residual Plot Analysis:
Company Size Residual Plot:
- Residuals show slight heteroscedasticity, with larger residuals observed at higher company sizes.
- The spread of residuals is not entirely random, suggesting the model may not fully capture all relationships.
- Given the small sample size (n=16), these patterns could be due to random variation rather than systematic issues.

Central Purchaser Residual Plot:
- Residuals cluster at two distinct values (0 and 1), as expected for a binary variable.
- Variation in residual spread is observed between the two groups, but no clear violations of regression assumptions are evident.
- Interpretation is limited by the binary nature of this predictor.

Normal Probability Plot:
- The plot indicates minor deviations from normality, especially in the tails.
- A slight S-shaped pattern suggests potential skewness.
- With a small sample size (n=16), these deviations are not severe enough to invalidate the analysis.

#2: "Full" Model - % Imports - Distance - Central Purchaser

Size of Purchase	Company Size
27.9	25.6
89.6	109.8
12.8	39.4
34.9	16.7
408.6	278.4
173.5	98.4
105.2	101.6
510.6	139.3
382.7	207.4
84.6	26.8
101.4	13.9
27.6	6.8
234.8	84.7
464.3	180.3
309.8	132.6
294.6	118.9

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.833306933
R Square	0.694400445
Adjusted R Square	0.672571905
Standard Error	97.07945609
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	299805.7632	299805.7632	31.811585	6.09665E-05
Residual	14	131941.8911	9424.420795		
Total	15	431747.6544			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	24.31628087	40.0396127	0.607305596	0.55337131	-61.56014744	110.1927092	-61.56014744	110.1927092
Company Size	1.818195309	0.322364997	5.640175973	6.09665E-05	1.126791156	2.509599463	1.126791156	2.509599463

Summary:

Strength of the model:
The R square decreased from 0.7700 to 0.6944, while the adjusted R square fell from 0.7436 to 0.6726. These indicate that the model's ability to explain variability in purchase size is reduced after removing "Central Purchaser".

ANOVE Table Insights:
The Standard Error increased significantly from 87.40 to 97.08, indicating a decrease in prediction precision, compared to the previous model.

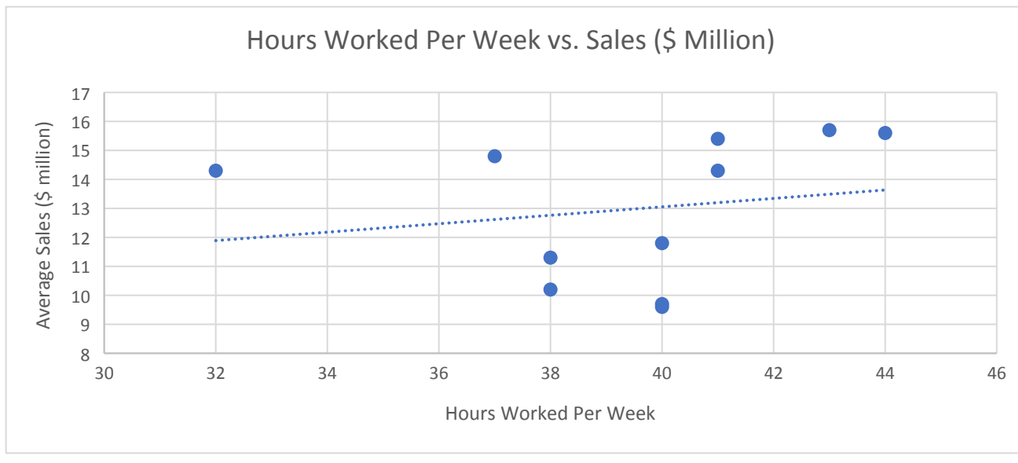
Significant Predictor Variable:
"Company Size" remains a strong anf significant predictor. Its contribution to explaining the purchase size is unaffected by removing "Central Purchase".

Sales	Hours Worked	Number of Customers
15.6	44	54
15.7	43	52
15.4	41	55
14.3	41	55
11.8	40	39
9.7	40	28
9.6	40	37
10.2	38	58
11.3	38	67
14.3	32	186
14.8	37	226

Step 1: Examining Relationships

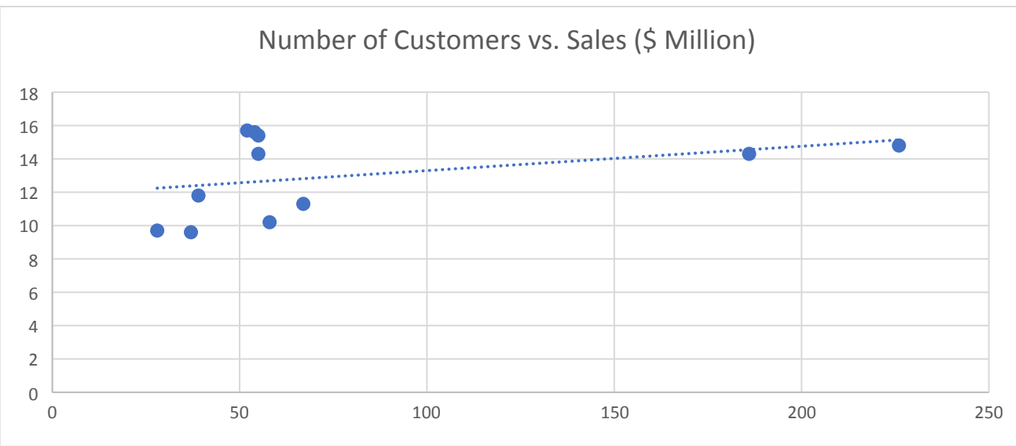
	Sales	*Hours Worked*	*Number of Customers*
Sales	1		
Hours Worked	**0.190562614**	1	
Number of Customers	**0.383615714**	-0.69842519	1

Hours Worked	Sales
44	15.6
43	15.7
41	15.4
41	14.3
40	11.8
40	9.7
40	9.6
38	10.2
38	11.3
32	14.3
37	14.8



<-------------- **Weak positive relationship, same conclusion from correlation test**

Number of Customers	Sales
54	15.6
52	15.7
55	15.4
55	14.3
39	11.8
28	9.7
37	9.6
58	10.2
67	11.3
186	14.3
226	14.8

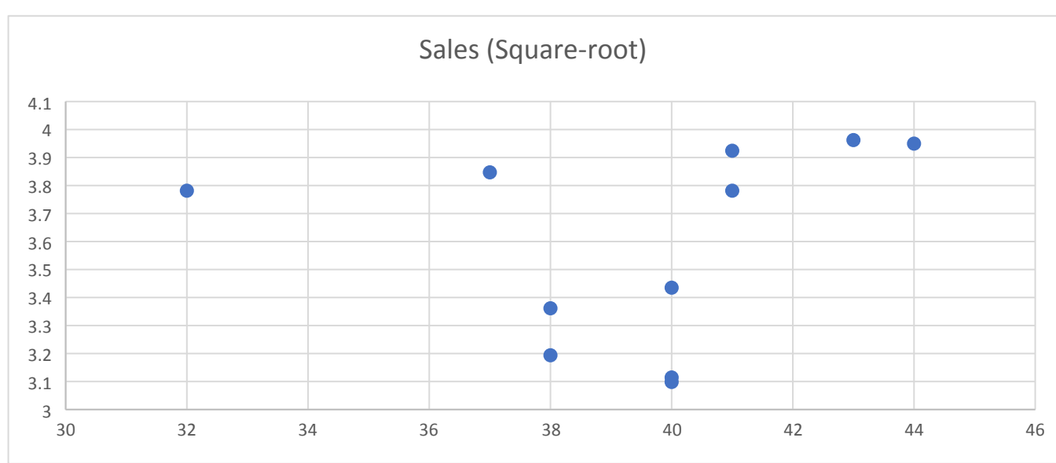


<-------------- **Relatively weak positive relationship, same conclusion from correlation test**

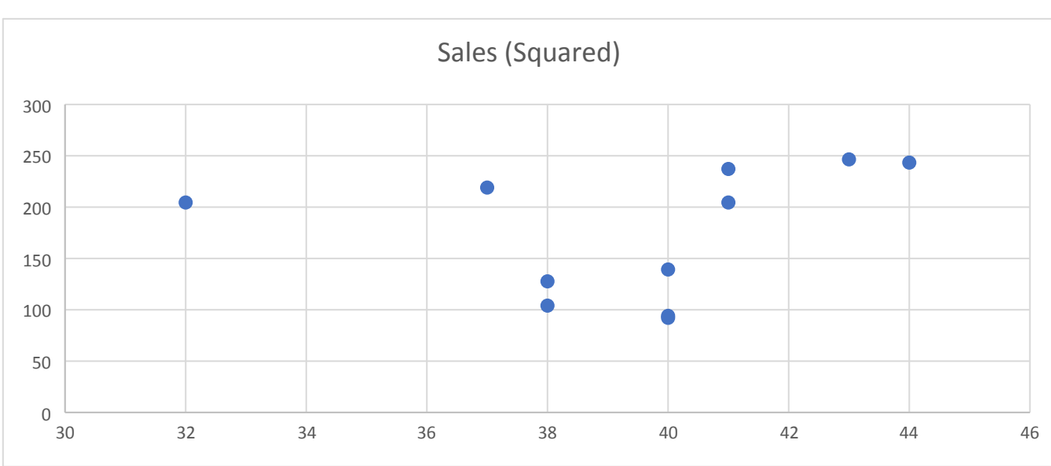
Step 2: Introducing Tukey's four quadrant approach in effort to linearize trend in data

Transforming dependent variable - Hours Worked vs Sales

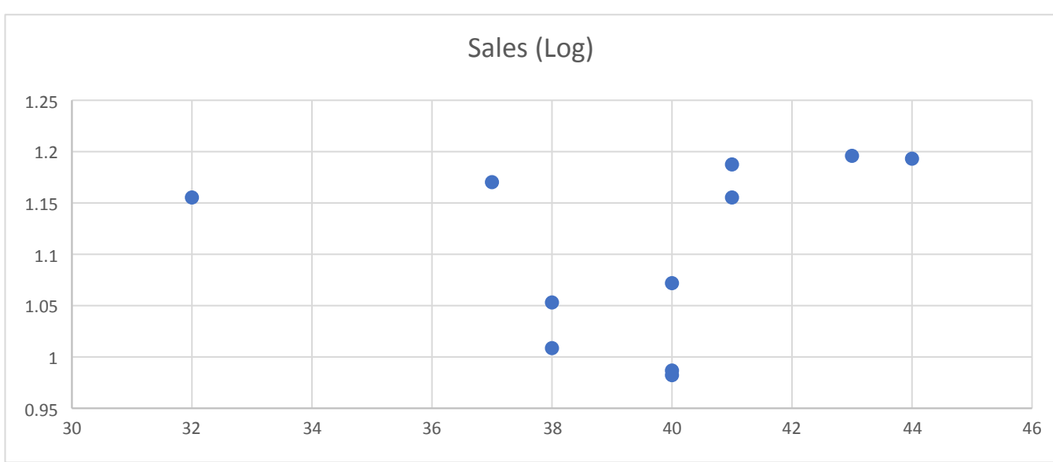
Hours Worked	Sales (Square-root)	Sales
44	3.949683532	15.6
43	3.962322551	15.7
41	3.924283374	15.4
41	3.78153408	14.3
40	3.435112807	11.8
40	3.1144823	9.7
40	3.098386677	9.6
38	3.193743885	10.2
38	3.361547263	11.3
32	3.78153408	14.3
37	3.847076812	14.8



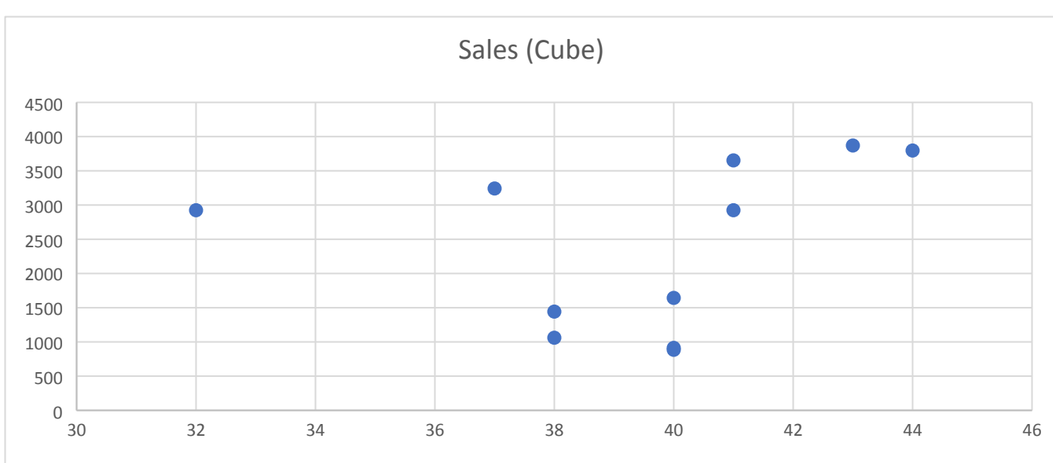
Hours Worked	Sales (Squared)	Sales
44	243.36	15.6
43	246.49	15.7
41	237.16	15.4
41	204.49	14.3
40	139.24	11.8
40	94.09	9.7
40	92.16	9.6
38	104.04	10.2
38	127.69	11.3
32	204.49	14.3
37	219.04	14.8



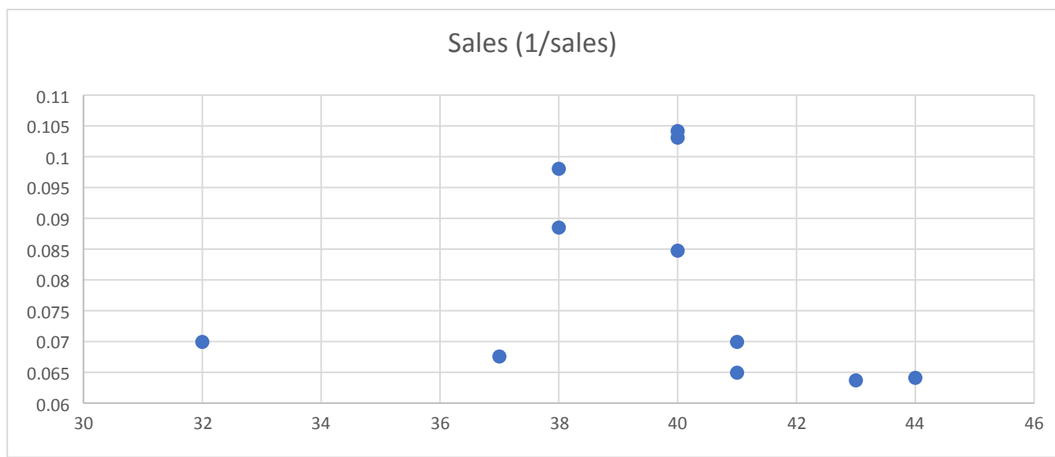
Hours Worked	Sales (Log)	Sales
44	1.193124598	15.6
43	1.195899652	15.7
41	1.187520721	15.4
41	1.155336037	14.3
40	1.071882007	11.8
40	0.986771734	9.7
40	0.982271233	9.6
38	1.008600172	10.2
38	1.053078443	11.3
32	1.155336037	14.3
37	1.170261715	14.8



Hours Worked	Sales (Cube)	Sales
44	3796.416	15.6
43	3869.893	15.7
41	3652.264	15.4
41	2924.207	14.3
40	1643.032	11.8
40	912.673	9.7
40	884.736	9.6
38	1061.208	10.2
38	1442.897	11.3
32	2924.207	14.3
37	3241.792	14.8

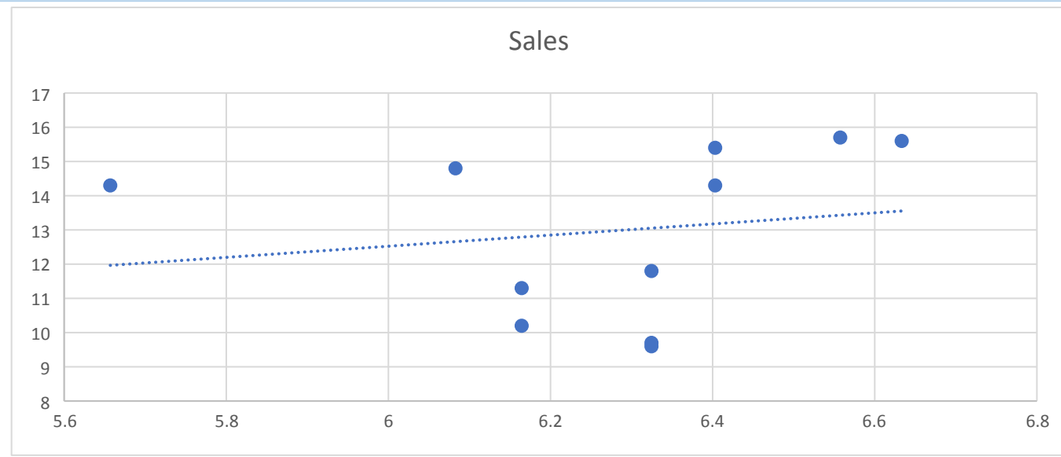


Hours Worked	Sales (1/sales)	Sales
44	0.064102564	15.6
43	0.063694268	15.7
41	0.064935065	15.4
41	0.06993007	14.3
40	0.084745763	11.8
40	0.103092784	9.7
40	0.104166667	9.6
38	0.098039216	10.2
38	0.088495575	11.3
32	0.06993007	14.3
37	0.067567568	14.8

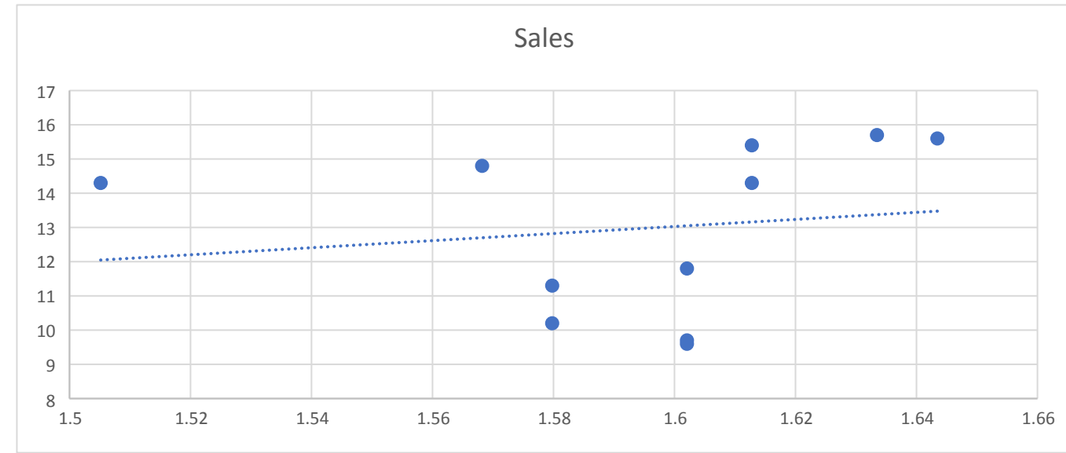


Transforming independent variable- Hours Worked vs Sales

Hours Worked	Hours Worked (Square-	Sales
44	6.633249581	15.6
43	6.557438524	15.7
41	6.403124237	15.4
41	6.403124237	14.3
40	6.32455532	11.8
40	6.32455532	9.7
40	6.32455532	9.6
38	6.164414003	10.2
38	6.164414003	11.3
32	5.656854249	14.3
37	6.08276253	14.8

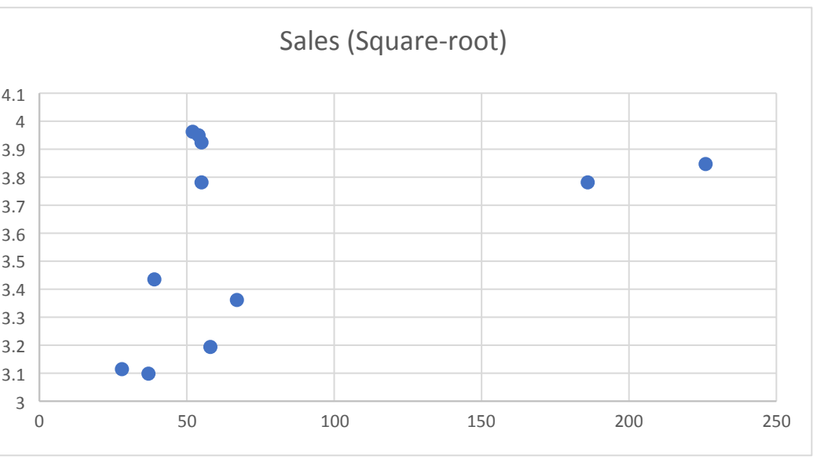


Hours Worked	Hours Worked (log)	Sales
44	1.643452676	15.6
43	1.633468456	15.7
41	1.612783857	15.4
41	1.612783857	14.3
40	1.602059991	11.8
40	1.602059991	9.7
40	1.602059991	9.6
38	1.579783597	10.2
38	1.579783597	11.3
32	1.505149978	14.3
37	1.568201724	14.8

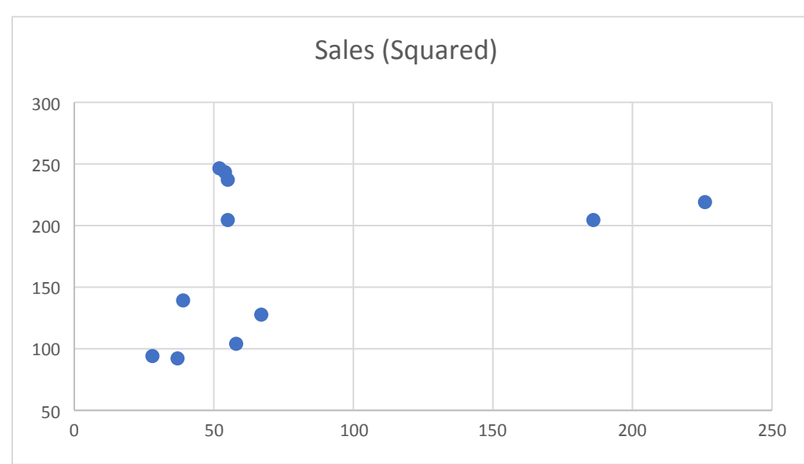


Transforming dependent variable - Number of Customers vs Sales

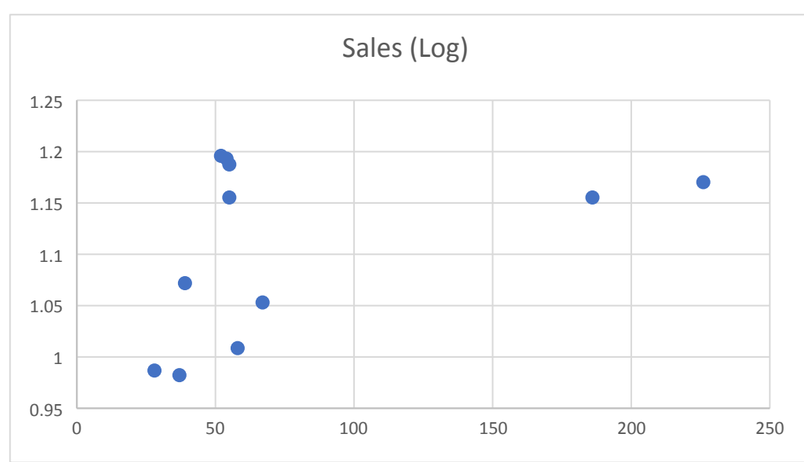
Number of Customers	Sales (Square-root)	Sales
54	3.949683532	15.6
52	3.962322551	15.7
55	3.924283374	15.4
55	3.78153408	14.3
39	3.435112807	11.8
28	3.1144823	9.7
37	3.098386677	9.6
58	3.193743885	10.2
67	3.361547263	11.3
186	3.78153408	14.3
226	3.847076812	14.8



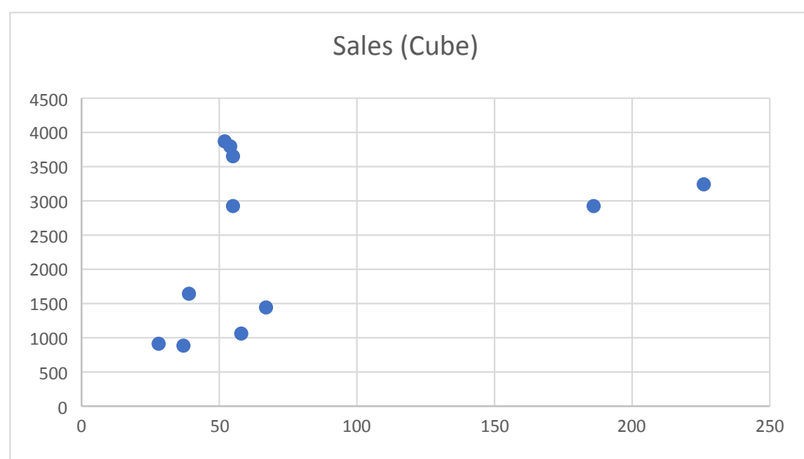
Number of Customers	Sales (Squared)	Sales
54	243.36	15.6
52	246.49	15.7
55	237.16	15.4
55	204.49	14.3
39	139.24	11.8
28	94.09	9.7
37	92.16	9.6
58	104.04	10.2
67	127.69	11.3
186	204.49	14.3
226	219.04	14.8



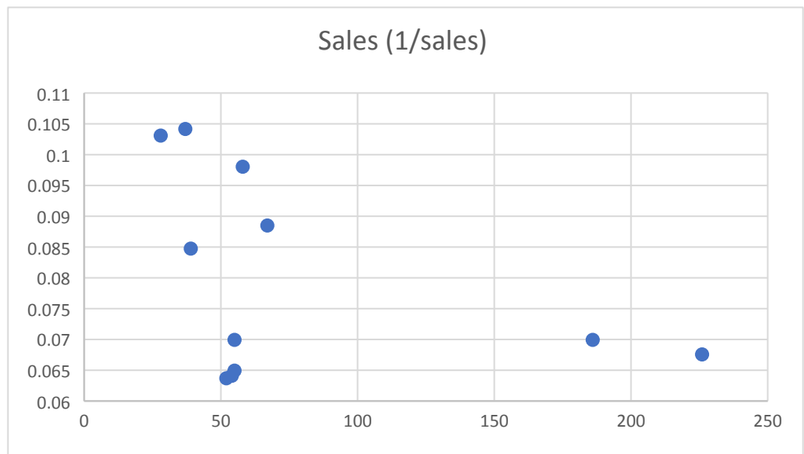
Number of Customers	Sales (Log)	Sales
54	1.193124598	15.6
52	1.195899652	15.7
55	1.187520721	15.4
55	1.155336037	14.3
39	1.071882007	11.8
28	0.986771734	9.7
37	0.982271233	9.6
58	1.008600172	10.2
67	1.053078443	11.3
186	1.155336037	14.3
226	1.170261715	14.8



Number of Customers	Sales (Cube)	Sales
54	3796.416	15.6
52	3869.893	15.7
55	3652.264	15.4
55	2924.207	14.3
39	1643.032	11.8
28	912.673	9.7
37	884.736	9.6
58	1061.208	10.2
67	1442.897	11.3
186	2924.207	14.3
226	3241.792	14.8

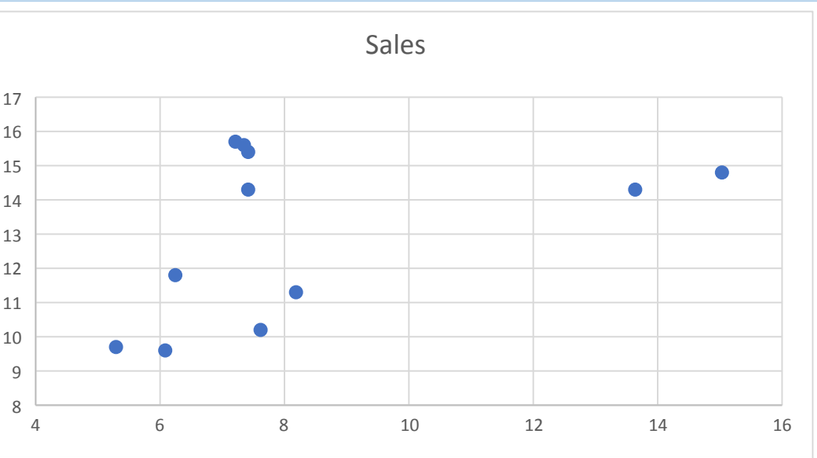


Number of Customers	Sales (1/sales)	Sales
54	0.064102564	15.6
52	0.063694268	15.7
55	0.064935065	15.4
55	0.06993007	14.3
39	0.084745763	11.8
28	0.103092784	9.7
37	0.104166667	9.6
58	0.098039216	10.2
67	0.088495575	11.3
186	0.06993007	14.3
226	0.067567568	14.8

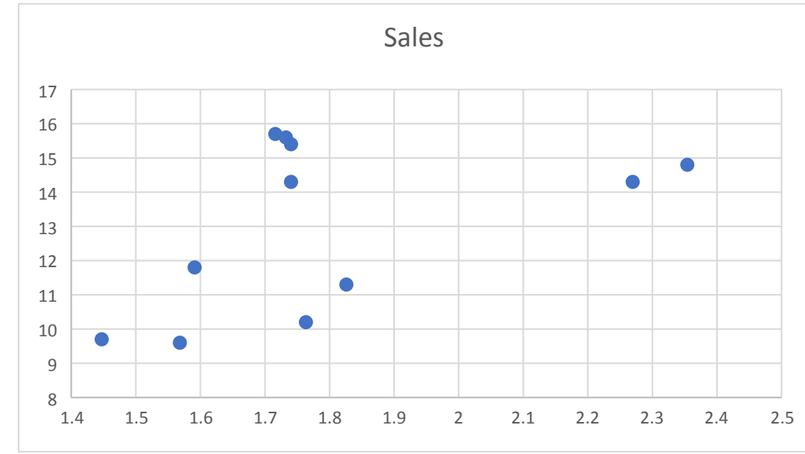


Transforming independent variable- Hours Worked vs Sales

Number of Customers	Number of Customers (Square-root)	Sales
54	7.348469228	15.6
52	7.211102551	15.7
55	7.416198487	15.4
55	7.416198487	14.3
39	6.244997998	11.8
28	5.291502622	9.7
37	6.08276253	9.6
58	7.615773106	10.2
67	8.185352772	11.3
186	13.6381817	14.3
226	15.03329638	14.8



Number of Customers	Number of Customers (log)	Sales
54	1.73239376	15.6
52	1.716003344	15.7
55	1.740362689	15.4
55	1.740362689	14.3
39	1.591064607	11.8
28	1.447158031	9.7
37	1.568201724	9.6
58	1.763427994	10.2
67	1.826074803	11.3
186	2.269512944	14.3
226	2.354108439	14.8



Step 3: Introducing Binary Variables (Tukey's approach)

Transforming the independent and dependent variables doesn't seem to help in linearizing the trend. In that case, we introduce binary variables.

Hours Worked	Sales	Hours_High
44	15.6	1
43	15.7	1
41	15.4	1
41	14.3	1
40	11.8	0
40	9.7	0
40	9.6	0
38	10.2	0
38	11.3	0
32	14.3	0
37	14.8	0

Median 40

Transforming the independent and dependent variables doesn't seem to help in linearizing the trend. In that case, we introduce binary variables.

Number of Customers	Sales	Customer_High
54	15.6	0
52	15.7	0
55	15.4	0
55	14.3	0
39	11.8	0
28	9.7	0
37	9.6	0
58	10.2	1
67	11.3	1
186	14.3	1
226	14.8	1

Median 55

Step 4: Conducting Stepwise Regression

Step 4.1: "Full" model

Stepwise Regression: Backward Elimination;
Adding non-linearity to the original independent variables

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	mber of Customer	Hours_High	Customer_High	hours*customers
15.6	44	54	1936	2916	1	0	2376
15.7	43	52	1849	2704	1	0	2236
15.4	41	55	1681	3025	1	0	2255
14.3	41	55	1681	3025	1	0	2255
11.8	40	39	1600	1521	0	0	1560
9.7	40	28	1600	784	0	0	1120
9.6	40	37	1600	1369	0	0	1480
10.2	38	58	1444	3364	0	1	2204
11.3	38	67	1444	4489	0	1	2546
14.3	32	186	1024	34596	0	1	5952
14.8	37	226	1369	51076	0	1	8362

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.978658861
R Square	0.957773165
Adjusted R Square	0.859243884
Standard Error	0.926930915
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	58.46421542	8.352030774	9.720695776	0.044220504
Residual	3	2.577602766	0.859200922		
Total	10	61.04181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-413.1799896	752.8893731	-0.548792431	0.621	-2809.209993	1982.850014	-2809.209993	1982.85001
Hours Worked	18.43262273	33.65820051	0.547641361	0.622	-88.68279314	125.5480386	-88.68279314	125.548039
Number of Customers	1.072684096	1.696612738	0.632250408	0.572	-4.326694845	6.472063037	-4.326694845	6.47206304
(Hours Worked)^2	-0.199266106	0.37126146	-0.536727152	0.629	-1.380785768	0.982253556	-1.380785768	0.98225356
(Number of Customers)^2	-0.000696841	0.000617	-1.129402852	0.341	-0.00266041	0.001266728	-0.00266041	0.00126673
Hours_High	1.21740404	2.40166307	0.506900429	0.647	-6.425759723	8.860567802	-6.425759723	8.8605678
Customer_High	0.473686673	5.247913928	0.090261898	**0.934**	-16.22751762	17.17489096	-16.22751762	17.174891
hours*customers	-0.022547212	0.041741771	-0.540159444	0.627	-0.155388156	0.110293732	-0.155388156	0.11029373

<-------------- **Variable with the highest p-value. Insignificant. To be dropped from the model**

Step 4.2: "Full" model - "Customer_High"

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	mber of Customer	Hours_High	hours*customers
15.6	44	54	1936	2916	1	2376
15.7	43	52	1849	2704	1	2236
15.4	41	55	1681	3025	1	2255
14.3	41	55	1681	3025	1	2255
11.8	40	39	1600	1521	0	1560
9.7	40	28	1600	784	0	1120
9.6	40	37	1600	1369	0	1480
10.2	38	58	1444	3364	0	2204
11.3	38	67	1444	4489	0	2546
14.3	32	186	1024	34596	0	5952
14.8	37	226	1369	51076	0	8362

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.97860027
R Square	0.957658488
Adjusted R Square	0.894146221
Standard Error	0.803835004
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	58.45721533	9.742869221	15.07832307	0.010159184
Residual	4	2.584602857	0.646150714		
Total	10	61.04181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-353.00711	303.4252338	-1.163407228	0.309	-1195.450615	489.4363951	-1195.450615	489.436395
Hours Worked	15.74296328	13.57190431	1.159967159	0.311	-21.938684	53.42461056	-21.938684	53.4246106
Number of Customers	0.937537184	0.691948559	1.35492324	0.247	-0.983620006	2.858694373	-0.983620006	2.85869437
(Hours Worked)^2	-0.169695978	0.151477162	-1.12027434	0.325	-0.590264003	0.250872048	-0.590264003	0.25087205
(Number of Customers)^2	-0.000674259	0.000489101	-1.378567505	0.240	-0.002032222	0.000683704	-0.002032222	0.0006837
Hours_High	1.245990994	2.064534713	0.603521455	**0.579**	-4.486076303	6.978058291	-4.486076303	6.97805829
hours*customers	-0.019111882	0.014865454	-1.285657477	0.268	-0.060385	0.022161235	-0.060385	0.02216124

<-------------- **Variable with the highest p-value. Insignificant. To be dropped from the model**

Step 4.3: "Full" model - "Customer_High" - "Hours_High"

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	mber of Customer	hours*customers
15.6	44	54	1936	2916	2376
15.7	43	52	1849	2704	2236
15.4	41	55	1681	3025	2255
14.3	41	55	1681	3025	2255
11.8	40	39	1600	1521	1560
9.7	40	28	1600	784	1120
9.6	40	37	1600	1369	1480
10.2	38	58	1444	3364	2204
11.3	38	67	1444	4489	2546
14.3	32	186	1024	34596	5952
14.8	37	226	1369	51076	8362

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.976628327
R Square	0.95380289
Adjusted R Square	0.90760578
Standard Error	0.750993421
Observations	11

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	58.22186259	11.64437252	20.64637567	0.002369702
Residual	5	2.819955596	0.563991119		
Total	10	61.04181818			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-503.5082254	161.4969826	-3.117756241	0.026318893	-918.6494354	-88.36701538	-918.6494354	-88.367015
Hours Worked	22.37113246	7.44962651	3.00298712	0.029996136	3.221257867	41.52100705	3.221257867	41.521007
Number of Customers	1.294050132	0.336647582	3.843931167	0.012075339	0.428669973	2.159430291	0.428669973	2.15943029
(Hours Worked)^2	-0.242365802	0.085865035	-2.822636744	0.036993667	-0.463088901	-0.021642703	-0.463088901	-0.0216427
(Number of Customers)^2	-0.000940319	0.000197913	-4.751171492	0.005099215	-0.001449071	-0.000431567	-0.001449071	-0.0004316
hours*customers	-0.026554515	0.007755129	-3.42412291	0.018754346	-0.04648971	-0.006619321	-0.04648971	-0.0066193

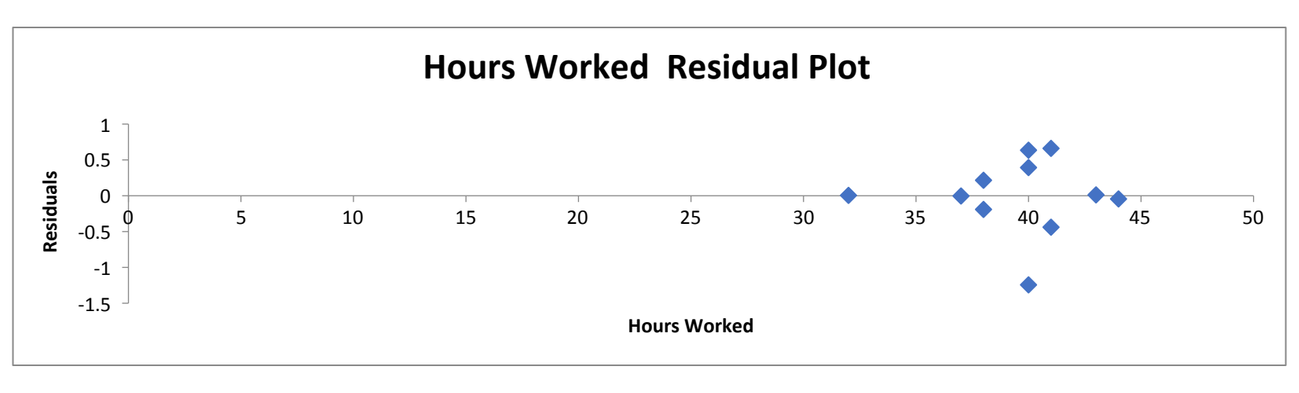
<-------------- **All independent variables are now significant at 5% significance level. We can stop here! :)**

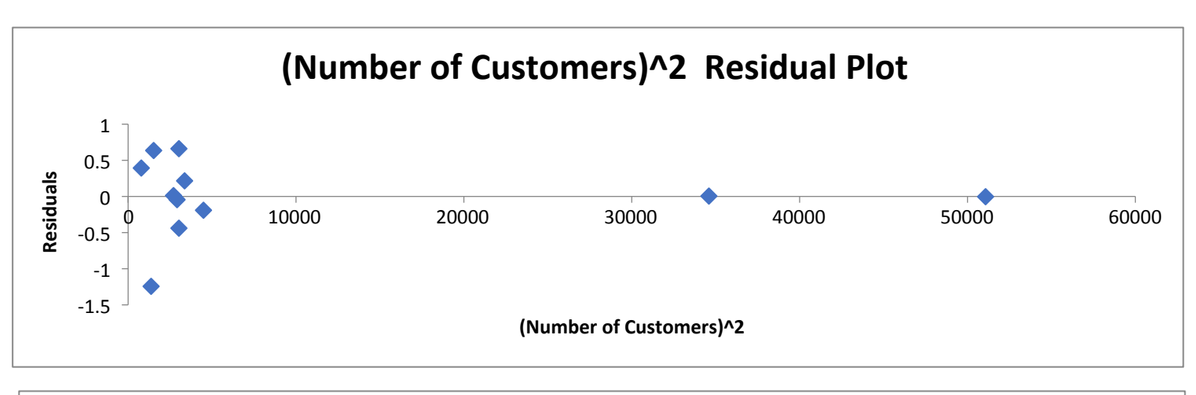
RESIDUAL OUTPUT

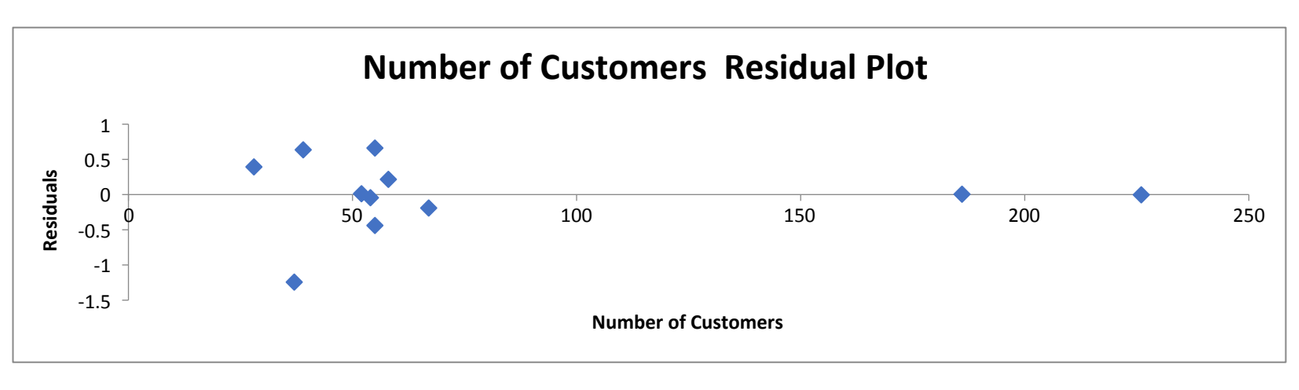
Observation	*Predicted Sales*	*Residuals*
1	15.64461805	-0.044618046
2	15.68818988	0.011810117
3	14.73915204	0.660847962
4	14.73915204	-0.439152038
5	11.16447518	0.63552482
6	9.306925473	0.393074527
7	10.84366461	-1.243664609
8	9.984112443	0.215887557
9	11.49106068	-0.191060684
10	14.29501017	0.004989831
11	14.80363944	-0.003639437

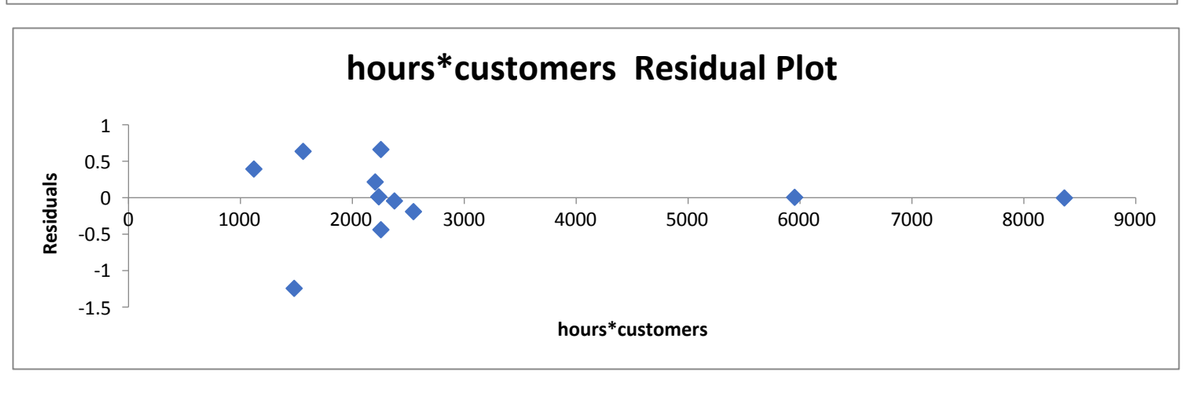
PROBABILITY OUTPUT

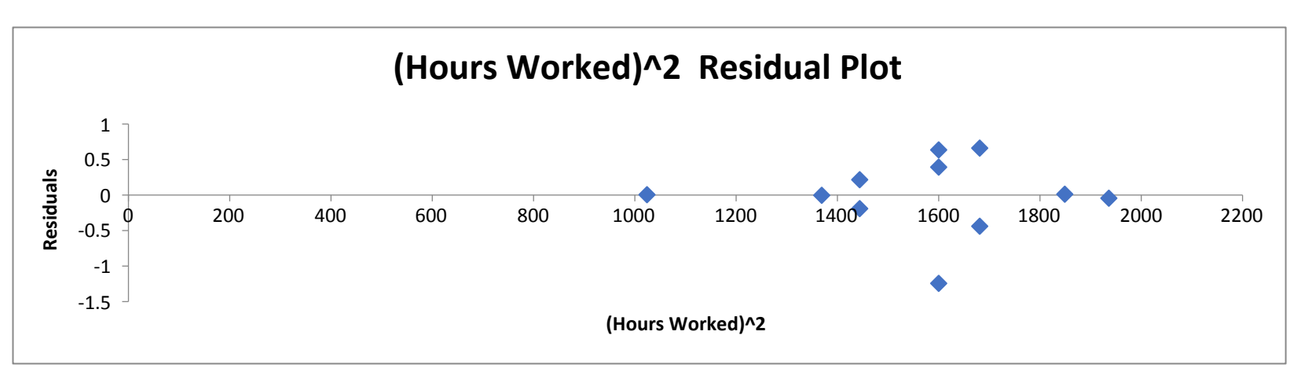
Percentile	*Sales*
4.545454545	9.6
13.63636364	9.7
22.72727273	10.2
31.81818182	11.3
40.90909091	11.8
50	14.3
59.09090909	14.3
68.18181818	14.8
77.27272727	15.4
86.36363636	15.6
95.45454545	15.7

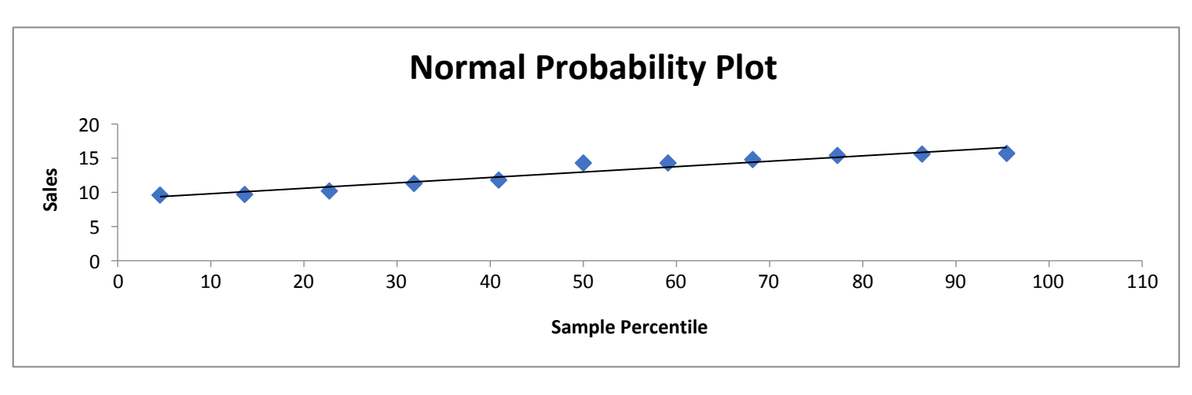












Summary:

Overall Model Fit:

R Square (0.954): The model explains 95.4% of the variation in sales, a significant improvement from the previous model.

Adjusted R Square (0.908): Still very high, confirming the model's robustness even with additional predictors.

Significance F (0.00237): The overall model is statistically significant, as the p-value is well below 0.05.

Predictors and Coefficients:

Hours Worked: Significant (p = 0.030). A unit increase in hours worked increases sales by ~22.37 units, assuming all else is constant.

Number of Customers: Significant (p = 0.012). A unit increase in the number of customers increases sales by ~1.29 units, holding other factors constant.

(Hours Worked)^2: Significant (p = 0.037). Suggests a diminishing return to hours worked, indicating a quadratic relationship.

(Number of Customers)^2: Significant (p = 0.005). Suggests diminishing returns as the number of customers increases.

hours*customers: Significant (p = 0.019). Indicates that the interaction between hours worked and number of customers negatively impacts sales. This might occur if overworking employees or overserving customers reduces effectiveness.

Insights from Predictors:

Linear Effects: Hours Worked and Number of Customers positively influence sales on their own.

Quadratic Effects: Both predictors exhibit diminishing returns, meaning excessive increases in hours or customers start reducing sales effectiveness.

Interaction Effect: The negative interaction term implies that the simultaneous increase in hours worked and customers reduces sales, potentially reflecting operational inefficiencies or overextension.

Residual Plot Analysis:

Hours Worked Residual Plot:
- Residuals display a slight fan-shaped pattern, indicating potential heteroscedasticity (increasing spread with hours worked).
- Several potential outliers are visible, especially around the 35–40 hour range.
- The observed pattern could reflect random variation due to the small sample size rather than a systematic model issue.

Number of Customers Residual Plot:
- Residuals cluster in the 0–100 customer range, with scattered points beyond.
- A few outliers are present, particularly at lower customer counts.
- The uneven spread may reflect natural variability in customer numbers rather than model misspecification.

Squared Terms Residual Plots:
Hours Worked^2:
- Residuals show no clear pattern but include some clustering.
- One notable outlier is observed around the 1500 mark.
- The random scatter suggests the quadratic term captures non-linear relationships effectively.

Number of Customers^2:
- Most residuals cluster at lower values, with decreasing spread at higher values.
- A few influential points at very low values could be driving the observed pattern.
- The slight heteroscedasticity may not be problematic given the small sample size.

Hours × Customers Interaction Plot:
- Residuals appear relatively random, with no clear pattern indicating good model fit.
- One notable outlier is observed around the 2000 mark.
- The scatter suggests the interaction term is appropriately capturing joint effects between the variables.

Conclusion:

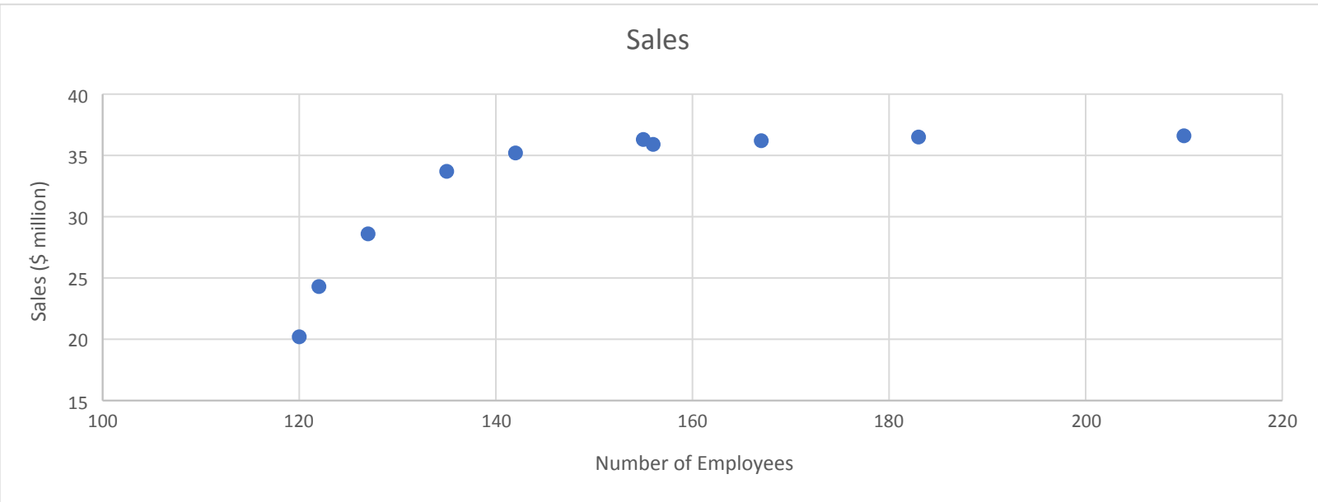
Given that none of the attempt to linearizing is successful, we went ahead and suspect that there is non-linearity in the data. We then introduced square(^2) to hours worked and numbers of customers, adding binary variables, as well as introduced interaction terms. Using stepwise regression, we dropped variables that arent significant in the models.

These outliers, while present, don't appear to severely impact the model's overall performance given the high R-squared value (0.954) and significant p-values for all predictors. However, investigating these outlying observations might reveal important operational insights or potential data collection issues.

Number of Employees	Sales
120	20.2
122	24.3
127	28.6
135	33.7
142	35.2
156	35.9
155	36.3
167	36.2
183	36.5
210	36.6

Step 1: Examining Relationships

	Number of Employees	*Sales*
Number of Employees	1	
Sales	**0.735891267**	1

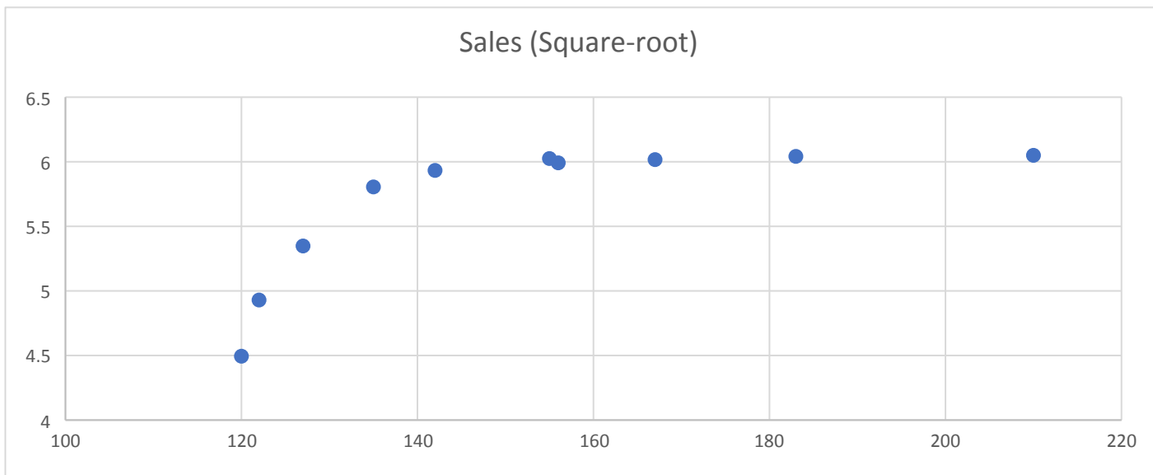


<------------- **Positive and diminishing**

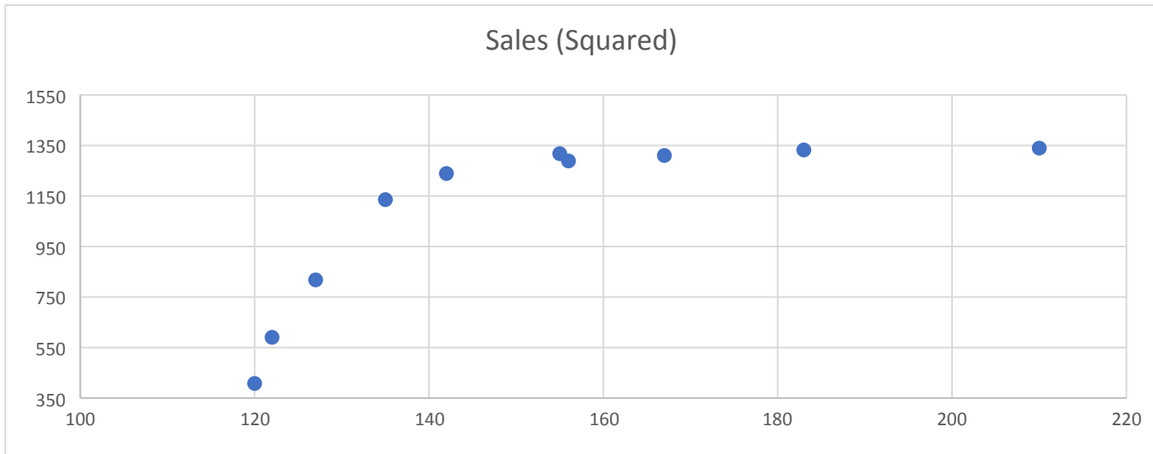
Step 2: Introducing Tukey's four quadrant approach in effort to linearize trend in data

Transforming dependent variable - Number of Employees vs Sales

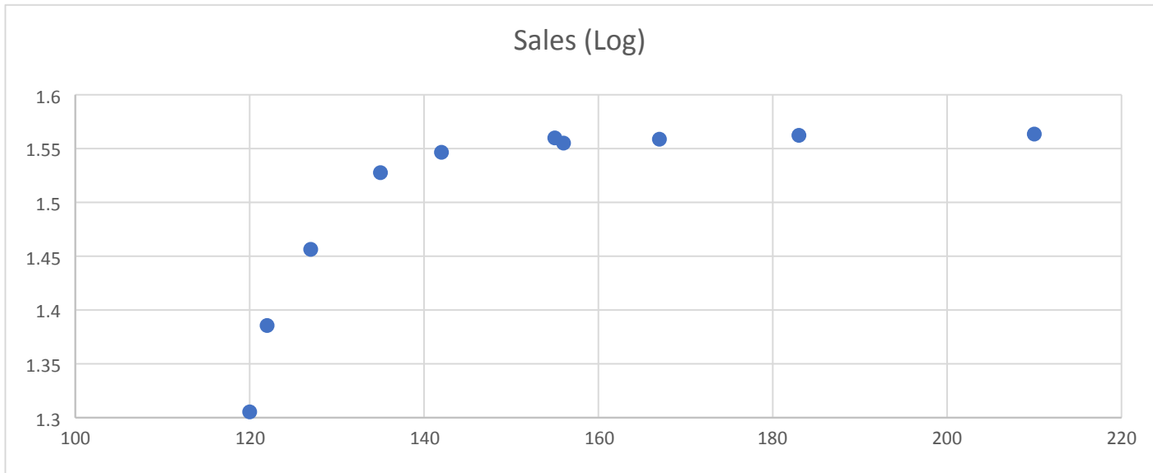
Number of Employees	Sales (Square-root)	Sales
120	4.494441011	20.2
122	4.929503018	24.3
127	5.347896783	28.6
135	5.805170109	33.7
142	5.93295879	35.2
156	5.991660872	35.9
155	6.024948133	36.3
167	6.016643583	36.2
183	6.041522987	36.5
210	6.049793385	36.6



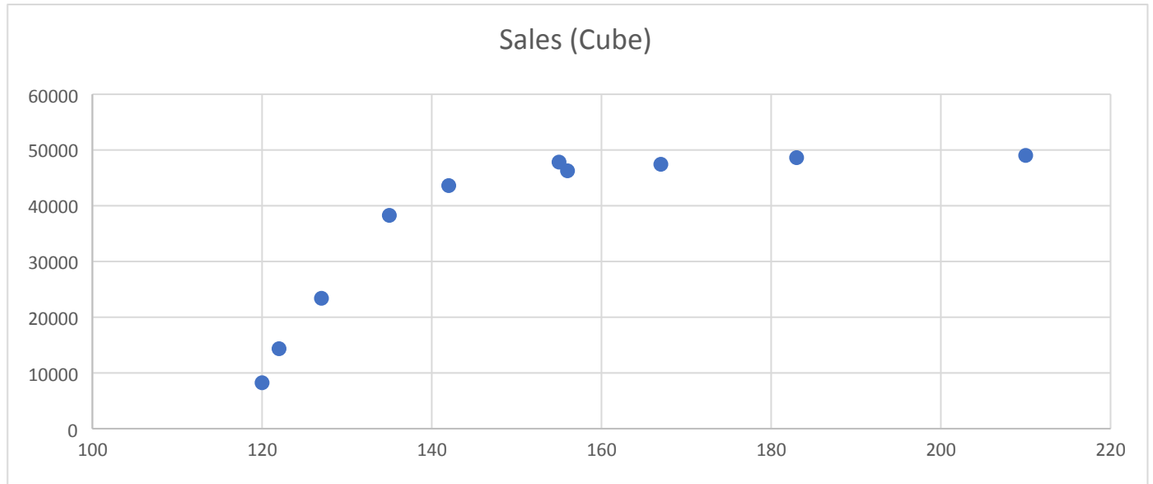
Number of Employees	Sales (Squared)	Sales
120	408.04	20.2
122	590.49	24.3
127	817.96	28.6
135	1135.69	33.7
142	1239.04	35.2
156	1288.81	35.9
155	1317.69	36.3
167	1310.44	36.2
183	1332.25	36.5
210	1339.56	36.6



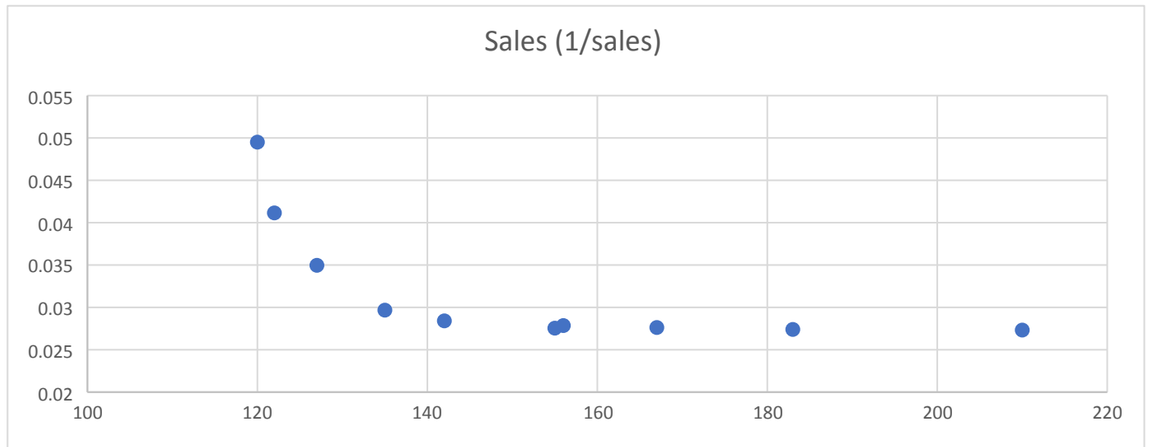
Number of Employees	Sales (Log)	Sales
120	1.305351369	20.2
122	1.385606274	24.3
127	1.456366033	28.6
135	1.527629901	33.7
142	1.546542663	35.2
156	1.555094449	35.9
155	1.559906625	36.3
167	1.558708571	36.2
183	1.562292864	36.5
210	1.563481085	36.6



Number of Employees	Sales (Cube)	Sales
120	8242.408	20.2
122	14348.907	24.3
127	23393.656	28.6
135	38272.753	33.7
142	43614.208	35.2
156	46268.279	35.9
155	47832.147	36.3
167	47437.928	36.2
183	48627.125	36.5
210	49027.896	36.6

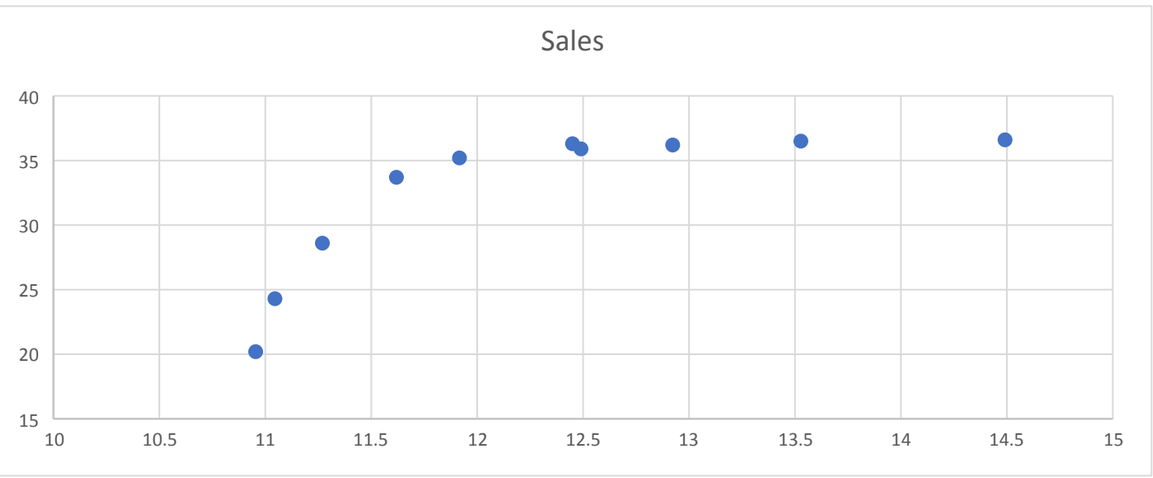


Number of Employees	Sales (1/sales)	Sales
120	0.04950495	20.2
122	0.041152263	24.3
127	0.034965035	28.6
135	0.029673591	33.7
142	0.028409091	35.2
156	0.027855153	35.9
155	0.027548209	36.3
167	0.027624309	36.2
183	0.02739726	36.5
210	0.027322404	36.6

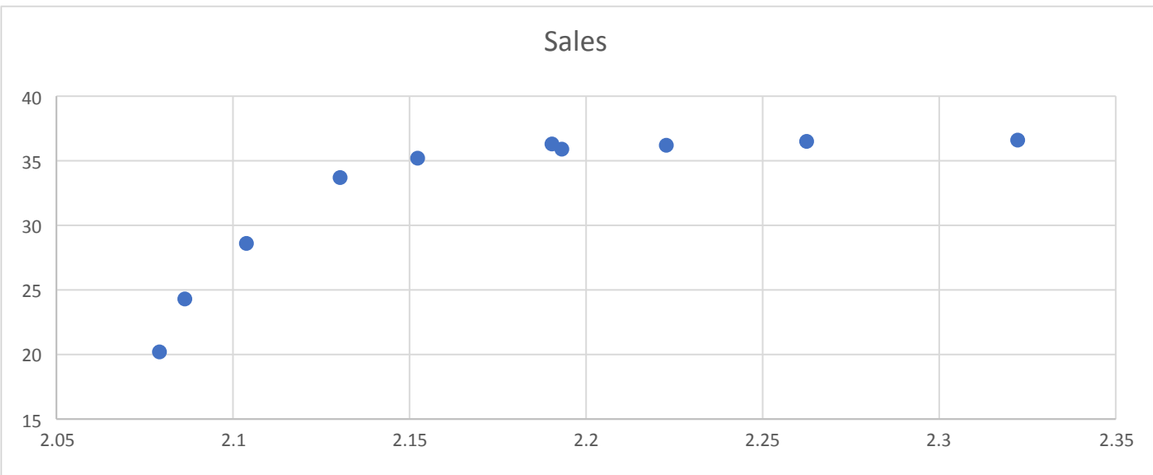


Transforming independent variable- Hours Worked vs Sales

Number of Employees	Number of Employees (Square-root)	Sales
120	10.95445115	20.2
122	11.04536102	24.3
127	11.26942767	28.6
135	11.61895004	33.7
142	11.91637529	35.2
156	12.489996	35.9
155	12.4498996	36.3
167	12.92284798	36.2
183	13.52774926	36.5
210	14.49137675	36.6



Number of Employees	Number of Employees (log)	Sales
120	2.079181246	20.2
122	2.086359831	24.3
127	2.103803721	28.6
135	2.130333768	33.7
142	2.152288344	35.2
156	2.193124598	35.9
155	2.190331698	36.3
167	2.222716471	36.2
183	2.26245109	36.5
210	2.322219295	36.6



Step 3: Introducing Binary Variables (Tukey's approach)

Transforming the independent and dependent variables doesn't seem to help in linearizing the trend. In that case, we introduce binary variables.

Sales	Number of Employees	High_employee
20.2	120	0
24.3	122	0
28.6	127	0
33.7	135	0
35.2	142	0
35.9	156	1
36.3	155	1
36.2	167	1
36.5	183	1
36.6	210	1

median 148.5

Step 4: Conducting Stepwise Regression

Step 4.1: "Full" model

Sales	Number of Employees	High_employee	(Number of Employees)^2
20.2	120	0	14400
24.3	122	0	14884
28.6	127	0	16129
33.7	135	0	18225
35.2	142	0	20164
35.9	156	1	24336
36.3	155	1	24025
36.2	167	1	27889
36.5	183	1	33489
36.6	210	1	44100

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.959100926
R Square	0.919874587
Adjusted R Square	0.879811881
Standard Error	2.050165565
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	289.5259269	96.50864231	22.96086983	0.0010909
Residual	6	25.21907306	4.203178844		
Total	9	314.745			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-142.2284392	31.40294831	-4.529142862	0.003978772	-219.06869	-65.388193	-219.06869	-65.388193
Number of Employees	2.040301961	0.382680831	5.331602195	0.001775428	1.1039157	2.97668822	1.1039157	2.97668822
High_employee	-6.207323478	2.954711759	-2.100822004	**0.080386276**	-13.437243	1.02259574	-13.437243	1.02259574
(Number of Employees)^2	-0.005547515	0.0010872	-5.102573046	0.002215794	-0.0082078	-0.0028872	-0.0082078	-0.0028872

<------------- **Variable with the highest p-value. Insignificant. To be dropped from the model**

Step 4.2: "Full" model - "High_employee"

Sales	Number of Employees	(Number of Employees)^2
20.2	120	14400
24.3	122	14884
28.6	127	16129
33.7	135	18225
35.2	142	20164
35.9	156	24336
36.3	155	24025
36.2	167	27889
36.5	183	33489
36.6	210	44100

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.927866529
R Square	0.860936295
Adjusted R Square	0.821203808
Standard Error	2.500560101
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	270.9753943	135.4876971	21.6683213	0.00100288
Residual	7	43.76960573	6.252800818		
Total	9	314.745			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-94.15577085	26.23101363	-3.589482747	0.008863252	-156.18226	-32.12928	-156.18226	-32.12928
Number of Employees	1.474009371	0.331304562	4.449106771	0.002974843	0.69059857	2.25742017	0.69059857	2.25742017
(Number of Employees)^2	-0.004086056	0.001019048	-4.009680874	0.005126387	-0.0064957	-0.0016764	-0.0064957	-0.0016764

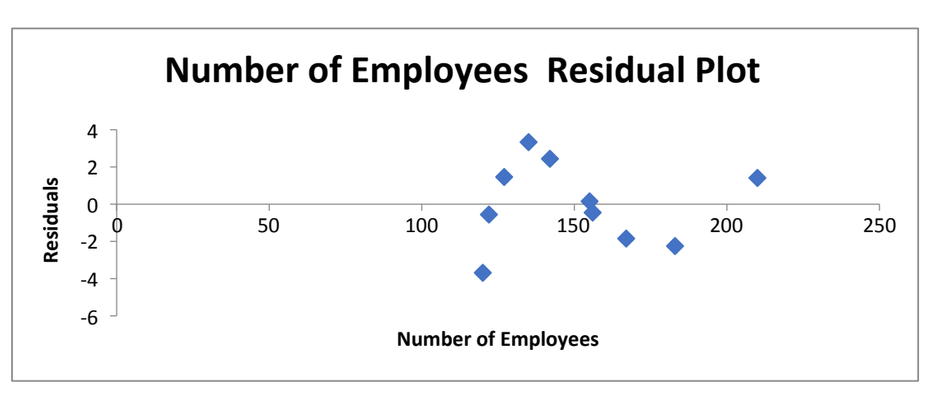
<------------- **All independent variables are now significant at 5% significance level. We can stop here! :)**

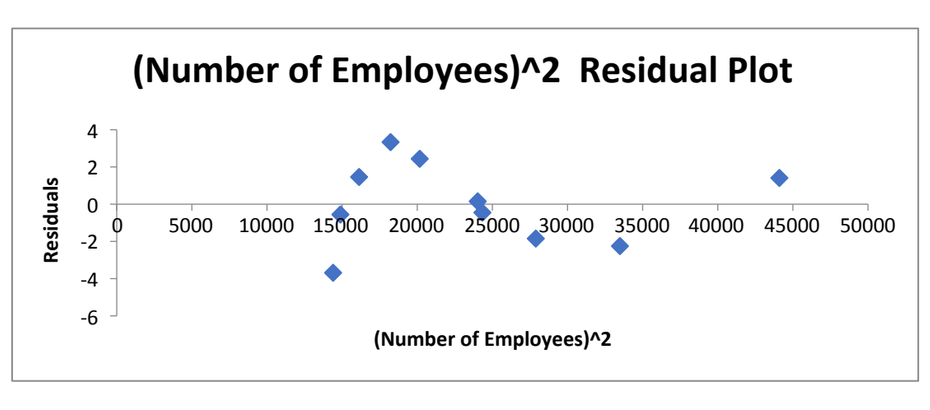
RESIDUAL OUTPUT

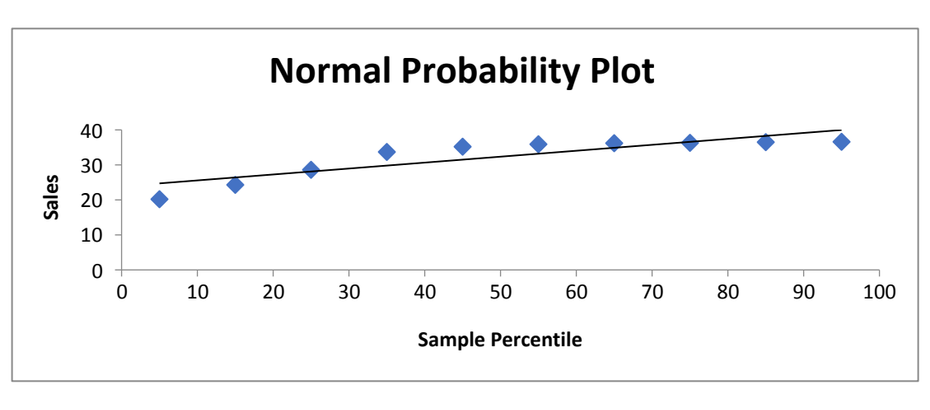
Observation	*Predicted Sales*	*Residuals*
1	23.88614282	-3.686142824
2	24.85651031	-0.556510313
3	27.13941706	1.460582937
4	30.36711801	3.332881993
5	32.76232042	2.43767958
6	36.35142469	-0.451424692
7	36.14817883	0.151821167
8	38.04776971	-1.847769706
9	38.75000431	-2.250004311
10	35.19111383	1.40888617

PROBABILITY OUTPUT

Percentile	*Sales*
5	20.2
15	24.3
25	28.6
35	33.7
45	35.2
55	35.9
65	36.2
75	36.3
85	36.5
95	36.6







Step 4.3: Introducing cubic model to the "Full" model to eliminate trend in the residual plots

Step 4.3.1: "Full" cubic model

After checking the residual plots, it seems like "patterns" still present. We went ahead and introduce cubic term to see if it will fix the problem:

Sales	Number of Employees	High_employee	(Number of Employees)^2	(Number of Employees)^3
20.2	120	0	14400	1728000
24.3	122	0	14884	1815848
28.6	127	0	16129	2048383
33.7	135	0	18225	2460375
35.2	142	0	20164	2863288
35.9	156	1	24336	3796416
36.3	155	1	24025	3723875
36.2	167	1	27889	4657463
36.5	183	1	33489	6128487
36.6	210	1	44100	9261000

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.991487697
R Square	0.983047853
Adjusted R Square	0.969486136
Standard Error	1.033015335
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	309.4093966	77.35234915	72.48697397	0.00012937
Residual	5	5.335603413	1.067120683		
Total	9	314.745			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-519.1176516	88.73425465	-5.850250883	0.0021	-747.21631	-291.01899	-747.21631	-291.01899
Number of Employees	9.3984636	1.715500453	5.478555011	0.0028	4.9886293	13.8082979	4.9886293	13.8082979
High_employee	-3.296015267	1.634433117	-2.016610672	**0.0998**	-7.4974593	0.90542882	-7.4974593	0.90542882
(Number of Employees)^2	-0.052370148	0.010860995	-4.821855205	0.0048	-0.0802892	-0.0244511	-0.0802892	-0.0244511
(Number of Employees)^3	9.66393E-05	2.23879E-05	4.316575254	0.0076	3.9089E-05	0.00015419	3.9089E-05	0.00015419

<------------- **Variable with the highest p-value. Insignificant. To be dropped from the model**

Step 4.3.2: "Full" cubic model - "High_employee"

Sales	Number of Employees	(Number of Employees)^2	(Number of Employees)^3
20.2	120	14400	1728000
24.3	122	14884	1815848
28.6	127	16129	2048383
33.7	135	18225	2460375
35.2	142	20164	2863288
35.9	156	24336	3796416
36.3	155	24025	3723875
36.2	167	27889	4657463
36.5	183	33489	6128487
36.6	210	44100	9261000

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.984509995
R Square	0.969259931
Adjusted R Square	0.953889897
Standard Error	1.269861078
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	305.0697171	101.6899057	63.06166316	6.2805E-05
Residual	6	9.675282945	1.612547158		
Total	9	314.745			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-570.59532	104.4678531	-5.461922527	0.002	-826.21895	-314.97169	-826.21895	-314.97169
Number of Employees	10.56748571	1.984774228	5.324275963	0.002	5.71091813	15.4240533	5.71091813	15.4240533
(Number of Employees)^2	-0.060752789	0.012334624	-4.925386307	0.003	-0.0909345	-0.0305711	-0.0909345	-0.0305711
(Number of Employees)^3	0.000115269	2.50686E-05	4.598167867	0.004	5.3929E-05	0.00017661	5.3929E-05	0.00017661

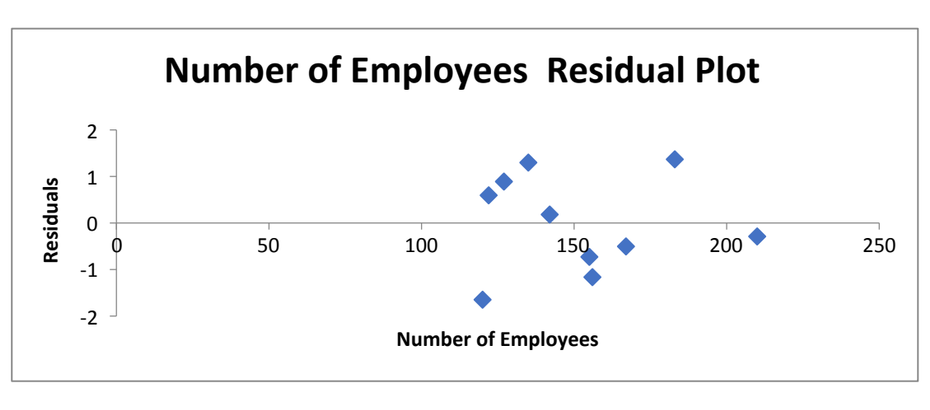
<------------- **All independent variables are now significant at 5% significance level. We can stop here! :)**

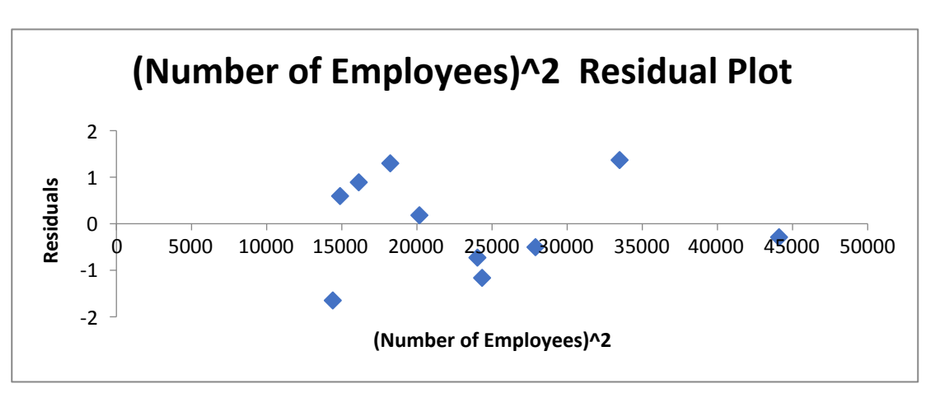
RESIDUAL OUTPUT

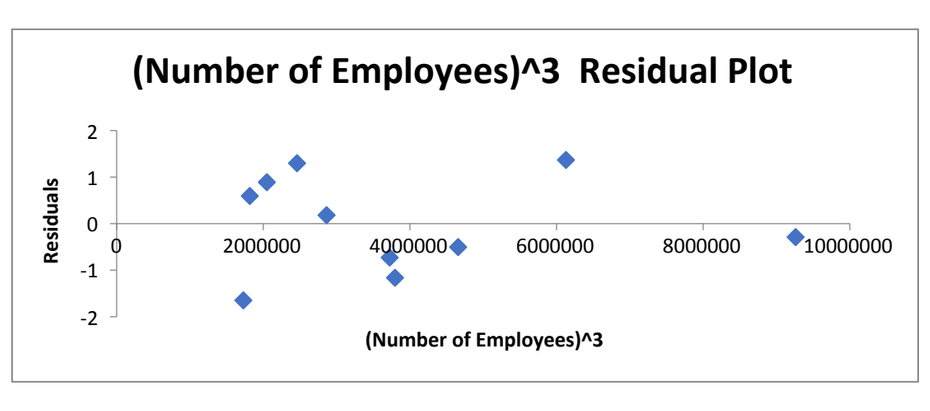
Observation	*Predicted Sales*	*Residuals*
1	21.84846282	-1.648462824
2	23.70527753	0.594722475
3	27.70967198	0.890328021
4	32.40181495	1.298185054
5	35.01812848	0.181871519
6	37.06347075	-1.163470752
7	37.02833917	-0.728339173
8	36.703593	-0.503592998
9	35.13191547	1.368084528
10	36.88932585	-0.289325851

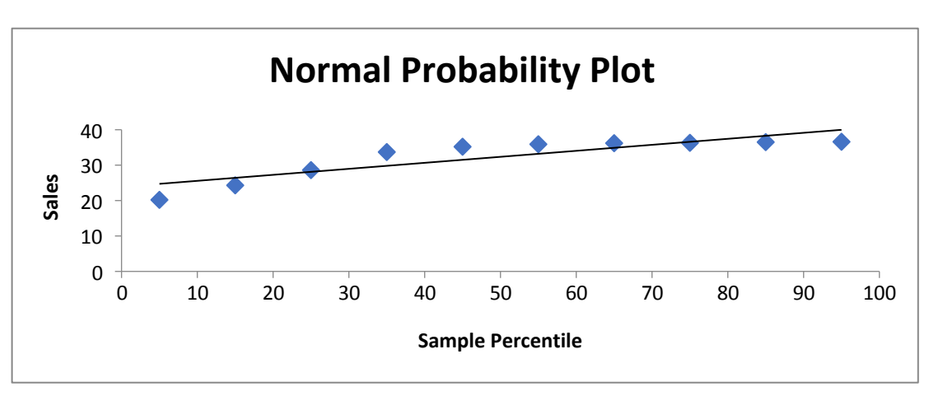
PROBABILITY OUTPUT

Percentile	*Sales*
5	20.2
15	24.3
25	28.6
35	33.7
45	35.2
55	35.9
65	36.2
75	36.3
85	36.5
95	36.6









Summary:

Overall Model Fit:

R Square (0.9693): The cubic model explains 96.93% of the variation in sales, representing a significant improvement from the quadratic model (R Square = 0.8609).

Adjusted R Square (0.9539): Still very high, confirming the robustness of the model even with the inclusion of additional predictors.

Significance F (6.28E-05): The overall model is statistically significant, as the p-value is well below 0.05.

Predictors and Coefficients:

Number of Employees: Significant (p=0.0018). A unit increase in employees initially increases sales by ~10.57 units, on average, ceteris paribus.

(Number of Employee)^2: Significant (p=0.0026). Suggests a diminishing return to adding employees, indicating a quadratic relationship.

(Number of Employee)^3: Significant (p=0.0037). Captures the more nuanced plateauing trend in sales as employee numbers increase further.

Insights from Predictors:

Linear Effects: The number of employees positively influences sales at first.

Quadratic Effects: The negative quadratic term indicates diminishing returns, meaning increasing employees beyond a certain point yields smaller gains in sales.

Cubic Effects: The positive cubic term reflects the leveling off or slight increase in sales after a plateau, showing a nonlinear pattern.

Conclusion:

Initial attempts at linear and quadratic modeling indicated some residual patterns, suggesting non-linearity in the relationship. Introducing cubic terms significantly improved the model fit, as evidenced by higher R Square and adjusted R Square .

Model Selection:

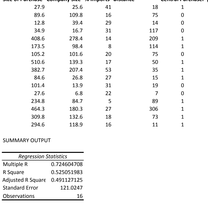
The cubic model (^3) was chosen as it balances capturing the complex patterns in the data while avoiding overfitting (e.g. introducing higher-order terms like ^4). The problem of still having some sort of patterns in the residual plots could also probably be fixed by adding more variables into the regression models. We are unable to do that here since the variables are given to us for analysis purpose.

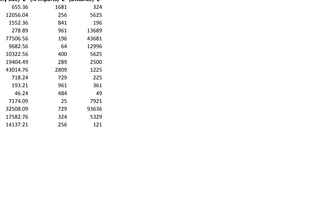
Management Recommendations:

- The relationship between employees and sales shows diminishing returns, plateauing as the number of employees grows. This suggests potential operational inefficiencies or limits to growth through hiring alone.
- To address this, management should focus on enhancing employee productivity or identifying external factors influencing sales.
- Avoid overstaffing, as adding more employees beyond a certain point may not yield significant sales growth.

Size of Purchase	Company Size	% Imports	Distance	Central Purchaser
27.9	25.6	41	18	1
89.6	109.8	16	75	0
12.8	39.4	29	14	0
34.9	16.7	31	117	0
408.6	278.4	14	209	1
173.5	98.4	8	114	1
105.2	101.6	20	75	0
510.6	139.3	17	50	1
382.7	207.4	53	35	1
84.6	26.8	27	15	1
101.4	13.9	31	19	0
27.6	6.8	22	7	0
234.8	84.7	5	89	1
464.3	180.3	27	306	1
309.8	132.6	18	73	1
294.6	118.9	16	11	1

Company Size	% Imports	Distance	Central Purchaser	Size of Purchase
25.6	41	18	1	27.9
109.8	16	75	0	89.6
39.4	29	14	0	12.8
16.7	31	117	0	34.9
278.4	14	209	1	408.6
98.4	8	114	1	173.5
101.6	20	75	0	105.2
139.3	17	50	1	510.6
207.4	53	35	1	382.7
26.8	27	15	1	84.6
13.9	31	19	0	101.4
6.8	22	7	0	27.6
84.7	5	89	1	234.8
180.3	27	306	1	464.3
132.6	18	73	1	309.8
118.9	16	11	1	294.6





Size of Purchase	Company Size	% Imports	Distance	Central Purchaser	(Company Size)^2	(% Imports)^2	(Distance)^2
27.9	25.6	41	18	1	655.36	1681	324
89.6	109.8	16	75	0	12056.04	256	5625
12.8	39.4	29	14	0	1552.36	841	196
34.9	16.7	31	117	0	278.89	961	13689
408.6	278.4	14	209	1	77506.56	196	43681
173.5	98.4	8	114	1	9682.56	64	12996
105.2	101.6	20	75	0	10322.56	400	5625
510.6	139.3	17	50	1	19404.49	289	2500
382.7	207.4	53	35	1	43014.76	2809	1225
84.6	26.8	27	15	1	718.24	729	225
101.4	13.9	31	19	0	193.21	961	361
27.6	6.8	22	7	0	46.24	484	49
234.8	84.7	5	89	1	7174.09	25	7921
464.3	180.3	27	306	1	32508.09	729	93636
309.8	132.6	18	73	1	17582.76	324	5329
294.6	118.9	16	11	1	14137.21	256	121

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.724604708
R Square	0.525051983
Adjusted R Square	0.491127125
Standard Error	121.0247
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	226690	226689.9622	15.47691012	0.001497927
Residual	14	205057.7	14646.97801		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	112.061197	38.2197	2.932084765	0.010926451	30.09029546	194.0364985	30.09029546	194.0364985
(Company Size)^2	0.005954497	0.001514	3.934070427	0.001497927	0.002708424	0.009201516	0.002708424	0.009201516

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.04532209
R Square	0.002054092
Adjusted R Square	-0.069227759
Standard Error	175.4302476
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	886.8493	886.8493314	0.028816477	0.867632009
Residual	14	430860.8	30775.77179		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	211.4001433	62.12355	3.402898675	0.004288095	78.15838195	344.6419046	78.15838195	344.6419046
(% Imports)^2	-0.010858909	0.063968	-0.169754167	0.867632009	-0.148057563	0.126339745	-0.14805756	0.126339745

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.502199549
R Square	0.252204387
Adjusted R Square	0.198790415
Standard Error	151.8596631
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	108888.7	108888.6527	4.721693153	0.047442316
Residual	14	322859	23061.35726		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	161.5199549	42.6882	3.783714279	0.002014576	69.9628701	253.0770397	69.9628701	253.0770397
(Distance)^2	0.001506823	0.001514	2.172956732	0.047442316	4.54447E-05	0.006968201	4.54447E-05	0.006968201

Size of Purchase	Company Size	% Imports	Distance	Central Purchaser	(Company Size)^2	(Distance)^2
27.9	25.6	41	18	1	655.36	324
89.6	109.8	16	75	0	12056.04	5625
12.8	39.4	29	14	0	1552.36	196
34.9	16.7	31	117	0	278.89	13689
408.6	278.4	14	209	1	77506.56	43681
173.5	98.4	8	114	1	9682.56	12996
105.2	101.6	20	75	0	10322.56	5625
510.6	139.3	17	50	1	19404.49	2500
382.7	207.4	53	35	1	43014.76	1225
84.6	26.8	27	15	1	718.24	225
101.4	13.9	31	19	0	193.21	361
27.6	6.8	22	7	0	46.24	49
234.8	84.7	5	89	1	7174.09	7921
464.3	180.3	27	306	1	32508.09	93636
309.8	132.6	18	73	1	17582.76	5329
294.6	118.9	16	11	1	14137.21	121

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.907233971
R Square	0.814928622
Adjusted R Square	0.691547704
Standard Error	94.22439719
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	351843.5	58640.58686	6.604980998	0.006503224
Residual	9	79904.13	8878.237027		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.40866535	105.7483	0.098428674	0.9237	-228.8106186	249.6279493	-228.810619	249.6279493
Company Size	2.466237923	1.073492	2.297390723	0.0472	0.0378294	4.894646445	0.0378294	4.894646445
% Imports	-0.708251864	2.405565	-0.294422248	**0.7751**	-6.150018024	4.733514295	-6.15001802	4.733514295
Distance	-0.968617412	1.179022	-0.821542937	0.4326	-3.635751067	1.698516243	-3.63575107	1.698516243
Central Purchaser	88.03994882	59.43409	1.481303792	0.1727	-46.40931263	222.4892103	-46.409313	222.4892103
(Company Size)^2	-0.003845081	0.003843	-1.000556475	0.3432	-0.012538424	0.004848259	-0.01253842	0.004848259
(Distance)^2	0.003726085	0.003645	1.022220243	0.3334	-0.004519682	0.011971852	-0.00451968	0.011971852

Size of Purchase	Company Size	Distance	Central Purchaser	(Company Size)^2	(Distance)^2
27.9	25.6	18	1	655.36	324
89.6	109.8	75	0	12056.04	5625
12.8	39.4	14	0	1552.36	196
34.9	16.7	117	0	278.89	13689
408.6	278.4	209	1	77506.56	43681
173.5	98.4	114	1	9682.56	12996
105.2	101.6	75	0	10322.56	5625
510.6	139.3	50	1	19404.49	2500
382.7	207.4	35	1	43014.76	1225
84.6	26.8	15	1	718.24	225
101.4	13.9	19	0	193.21	361
27.6	6.8	7	0	46.24	49
234.8	84.7	89	1	7174.09	7921
464.3	180.3	306	1	32508.09	93636
309.8	132.6	73	1	17582.76	5329
294.6	118.9	11	1	14137.21	121

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.901746133
R Square	0.813146088
Adjusted R Square	0.719719131
Standard Error	89.81856068
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	351073.9	70214.78319	8.703548957	0.002073576
Residual	10	80673.74	8067.373843		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-15.32624939	56.73596	-0.27013288	0.7926	-141.7418553	111.0893565	-141.741855	111.0893565
Company Size	2.5264433	1.00456	2.514974747	0.0307	0.288143897	4.764742703	0.2881439	4.764742703
Distance	-0.801858725	0.785714	-0.813497987	**0.4349**	-2.998166679	1.394449237	-2.99816668	1.394449237
Central Purchaser	89.54018025	56.44642	1.586286331	0.1438	-36.2302753	215.3106358	-36.230275	215.3106358
(Company Size)^2	-0.004163769	0.003515	-1.184593189	0.2636	-0.011995535	0.003667996	-0.01199554	0.003667996
(Distance)^2	0.003261241	0.003132	1.041133073	0.3222	-0.003716832	0.010239315	-0.00371683	0.010239315

Size of Purchase	Company Size	Central Purchaser	(Company Size)^2	(Distance)^2
27.9	25.6	1	655.36	324
89.6	109.8	0	12056.04	5625
12.8	39.4	0	1552.36	196
34.9	16.7	0	278.89	13689
408.6	278.4	1	77506.56	43681
173.5	98.4	1	9682.56	12996
105.2	101.6	0	10322.56	5625
510.6	139.3	1	19404.49	2500
382.7	207.4	1	43014.76	1225
84.6	26.8	1	718.24	225
101.4	13.9	0	193.21	361
27.6	6.8	0	46.24	49
234.8	84.7	1	7174.09	7921
464.3	180.3	1	32508.09	93636
309.8	132.6	1	17582.76	5329
294.6	118.9	1	14137.21	121

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.894863697
R Square	0.800783336
Adjusted R Square	0.728337776
Standard Error	88.42681646
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	345735.3	86433.83345	11.05390672	0.000756034
Residual	11	86012.32	7819.301869		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-37.3675931	49.07561	-0.76142904	0.4624	-145.3822781	70.64709192	-145.382278	70.64709192
Company Size	2.318014092	0.956285	2.423979244	0.0338	0.213245815	4.422782369	0.2132458	4.422782369
Central Purchaser	98.42160159	54.52217	1.805159963	0.0985	-21.58132867	218.4245535	-21.581329	218.424535
(Company Size)^2	-0.003882292	0.003444	-1.127372571	0.2836	-0.011461745	0.003697161	-0.01146175	0.003697161
(Distance)^2	0.000883665	0.001108	0.797777953	**0.4419**	-0.001554273	0.003321603	-0.00155427	0.003321603

Size of Purchase	Company Size	Central Purchaser	(Company Size)^2
27.9	25.6	1	655.36
89.6	109.8	0	12056.04
12.8	39.4	0	1552.36
34.9	16.7	0	278.89
408.6	278.4	1	77506.56
173.5	98.4	1	9682.56
105.2	101.6	0	10322.56
510.6	139.3	1	19404.49
382.7	207.4	1	43014.76
84.6	26.8	1	718.24
101.4	13.9	0	193.21
27.6	6.8	0	46.24
234.8	84.7	1	7174.09
464.3	180.3	1	32508.09
309.8	132.6	1	17582.76
294.6	118.9	1	14137.21

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.888399918
R Square	0.789254414
Adjusted R Square	0.736568017
Standard Error	87.07703131
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	340758.7	113586.2473	14.98023142	0.000213536
Residual	12	90988.91	7582.409383		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-37.4932023	48.32625	-0.775835158	0.4529	-142.7870515	67.80064692	-142.787052	67.80064692
Company Size	2.368031374	0.939661	2.520089931	0.0269	0.320684938	4.415377811	0.3206849	4.415377811
Central Purchaser	98.81882283	53.68159	1.84061696	0.0905	-18.15701936	215.794665	-18.157019	215.794665
(Company Size)^2	-0.003517787	0.003361	-1.046612179	**0.3159**	-0.010841034	0.00380546	-0.01084103	0.003805546

Size of Purchase	Company Size	Central Purchaser
27.9	25.6	1
89.6	109.8	0
12.8	39.4	0
34.9	16.7	0
408.6	278.4	1
173.5	98.4	1
105.2	101.6	0
510.6	139.3	1
382.7	207.4	1
84.6	26.8	1
101.4	13.9	0
27.6	6.8	0
234.8	84.7	1
464.3	180.3	1
309.8	132.6	1
294.6	118.9	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.877506072
R Square	0.770016906
Adjusted R Square	0.734634891
Standard Error	87.39594317
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	332453	166226.4964	21.76294699	7.09616E-05
Residual	13	99294.66	7638.050883		
Total	15	431747.7			

	Coefficients	*Standard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	7.737082382	39.24508	-0.198014169	0.84609681	-92.55492855	77.01276379	-92.554926	77.01276379
Company Size	1.450820591	0.94029	4.263479708	0.000923776	0.715668193	2.185972989	0.7156682	2.185972989
Central Purchaser	109.4070292	52.91921	2.067435139	0.059203564	-4.917968076	223.7320265	-4.917968	223.7320265

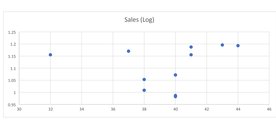
Sales	Hours Worked	Number of Customers
15.6	44	54
15.7	43	52
15.4	41	55
14.3	41	55
11.8	40	39
9.7	40	28
9.6	40	37
10.2	38	58
11.3	38	67
14.3	32	186
14.8	37	226

Hours Worked	Sales
44	15.6
43	15.7
41	15.4
41	14.3
40	11.8
40	9.7
40	9.6
38	10.2
38	11.3
32	14.3
37	14.8

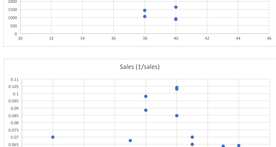
Number of Customers	Sales
54	15.6
52	15.7
55	15.4
55	14.3
39	11.8
28	9.7
37	9.6
58	10.2
67	11.3
186	14.3
226	14.8

Transforming dependent variable - Hours Worked vs Sales

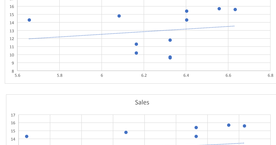
Hours Worked	Sales (Square-root)	Sales
44	3.949683532	15.6
43	3.962322551	15.7
41	3.924283374	15.4
41	3.78153408	14.3
40	3.435112807	11.8
40	3.1144823	9.7
40	3.098386677	9.6
38	3.193743885	10.2
38	3.361547263	11.3
32	3.78153408	14.3
37	3.847076812	14.8



Hours Worked	Sales (Squared)	Sales
44	243.36	15.6
43	246.49	15.7
41	237.16	15.4
41	204.49	14.3
40	139.24	11.8
40	94.09	9.7
40	92.16	9.6
38	104.04	10.2
38	127.69	11.3
32	204.49	14.3
37	219.04	14.8

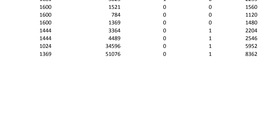


Hours Worked	Sales (Log)	Sales
44	1.193124598	15.6
43	1.195899652	15.7
41	1.187520721	15.4
41	1.155336037	14.3
40	1.071882007	11.8
40	0.986771734	9.7
40	0.982271233	9.6
38	1.008600172	10.2
38	1.053078443	11.3
32	1.155336037	14.3
37	1.170261715	14.8



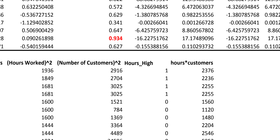
Hours Worked	Sales (Cube)	Sales
44	3796.416	15.6
43	3869.893	15.7
41	3652.264	15.4
41	2924.207	14.3
40	1643.032	11.8
40	912.673	9.7
40	884.736	9.6
38	1061.208	10.2
38	1442.897	11.3
32	2924.207	14.3
37	3241.792	14.8

Hours Worked	Sales (1/sales)	Sales
44	0.064102564	15.6
43	0.063694268	15.7
41	0.064935065	15.4
41	0.06993007	14.3
40	0.084745763	11.8
40	0.103092784	9.7
40	0.104166667	9.6
38	0.098039216	10.2
38	0.088495575	11.3
32	0.06993007	14.3
37	0.067567568	14.8

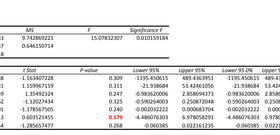


Transforming Independent variable- Hours Worked vs Sales

Hours Worked	Hours Worked (Square-	Sales
44	6.633249581	15.6
43	6.557438524	15.7
41	6.403124237	15.4
41	6.403124237	14.3
40	6.32455532	11.8
40	6.32455532	9.7
40	6.32455532	9.6
38	6.164414003	10.2
38	6.164414003	11.3
32	5.656854249	14.3
37	6.08276253	14.8



Hours Worked	Hours Worked (log)	Sales
44	1.643452676	15.6
43	1.633468456	15.7
41	1.612783857	15.4
41	1.612783857	14.3
40	1.602059991	11.8
40	1.602059991	9.7
40	1.602059991	9.6
38	1.579783597	10.2
38	1.579783597	11.3
32	1.505149978	14.3
37	1.568201724	14.8

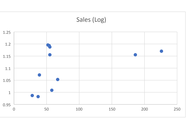


Transforming the independent and dependent variables doesn't seem to help in linearizing the trend. In that case, we introduce binary variables.

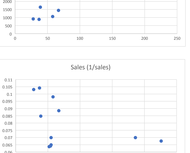
Hours Worked	Sales	Hours_High
44	15.6	1
43	15.7	1
41	15.4	1
41	14.3	1
40	11.8	0
40	9.7	0
40	9.6	0
38	10.2	0
38	11.3	0
32	14.3	0
37	14.8	0
Median	40	

Transforming dependent variable - Number of Customers vs Sales

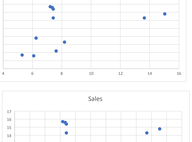
Number of Customers	Sales (Square-root)	Sales
54	3.949683532	15.6
52	3.962322551	15.7
55	3.924283374	15.4
55	3.78153408	14.3
39	3.435112807	11.8
28	3.1144823	9.7
37	3.098386677	9.6
58	3.193743885	10.2
67	3.361547263	11.3
186	3.78153408	14.3
226	3.847076812	14.8



Number of Customers	Sales (Squared)	Sales
54	243.36	15.6
52	246.49	15.7
55	237.16	15.4
55	204.49	14.3
39	139.24	11.8
28	94.09	9.7
37	92.16	9.6
58	104.04	10.2
67	127.69	11.3
186	204.49	14.3
226	219.04	14.8



Number of Customers	Sales (Log)	Sales
54	1.193124598	15.6
52	1.195899652	15.7
55	1.187520721	15.4
55	1.155336037	14.3
39	1.071882007	11.8
28	0.986771734	9.7
37	0.982271233	9.6
58	1.008600172	10.2
67	1.053078443	11.3
186	1.155336037	14.3
226	1.170261715	14.8



Number of Customers	Sales (Cube)	Sales
54	3796.416	15.6
52	3869.893	15.7
55	3652.264	15.4
55	2924.207	14.3
39	1643.032	11.8
28	912.673	9.7
37	884.736	9.6
58	1061.208	10.2
67	1442.897	11.3
186	2924.207	14.3
226	3241.792	14.8

Number of Customers	Sales (1/sales)	Sales
54	0.064102564	15.6
52	0.063694268	15.7
55	0.064935065	15.4
55	0.06993007	14.3
39	0.084745763	11.8
28	0.103092784	9.7
37	0.104166667	9.6
58	0.098039216	10.2
67	0.088495575	11.3
186	0.06993007	14.3
226	0.067567568	14.8

Transforming Independent variable- Hours Worked vs Sales

Number of Customers	Number of Customers (Square-root)	Sales
54	7.348469228	15.6
52	7.211102551	15.7
55	7.416198487	15.4
55	7.416198487	14.3
39	6.244997998	11.8
28	5.291502622	9.7
37	6.08276253	9.6
58	7.615773106	10.2
67	8.185352772	11.3
186	13.6381817	14.3
226	15.03329638	14.8

Number of Customers	Number of Customers (log)	Sales
54	1.73233976	15.6
52	1.716003344	15.7
55	1.740362689	15.4
55	1.740362689	14.3
39	1.591064607	11.8
28	1.447158031	9.7
37	1.568201724	9.6
58	1.763427994	10.2
67	1.826074803	11.3
186	2.269512944	14.3
226	2.354108439	14.8

Transforming the independent and dependent variables doesn't seem to help in linearizing the trend. In that case, we introduce binary variables.

Number of Customers	Sales	Customer_High
54	15.6	0
52	15.7	0
55	15.4	0
55	14.3	0
39	11.8	0
28	9.7	0
37	9.6	0
58	10.2	1
67	11.3	1
186	14.3	1
226	14.8	1
Median	55	

Stepwise Regression: Backward Elimination;
Adding non-linearity to the original independent variables

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	(Number of Customers)^2	Hours_High	Customer_High	hours*customers
15.6	44	54	1936	2916	1	0	2376
15.7	43	52	1849	2704	1	0	2236
15.4	41	55	1681	3025	1	0	2255
14.3	41	55	1681	3025	1	0	2255
11.8	40	39	1600	1521	0	0	1560
9.7	40	28	1600	784	0	0	1120
9.6	40	37	1600	1369	0	0	1480
10.2	38	58	1444	3364	0	1	2204
11.3	38	67	1444	4489	0	1	2546
14.3	32	186	1024	34596	0	1	5952
14.8	37	226	1369	51076	0	1	8362

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.978658861
R Square	0.957773165
Adjusted R Square	0.850243884
Standard Error	0.926930955
Observations	11

ANOVA

	df	SS	MS	F	Significance F
Regression	7	58.46421542	8.352030774	9.720695776	0.044220504
Residual	3	2.577602766	0.859200922		
Total	10	61.04181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	-413.1799896	752.8893731	-0.548792431	0.621	-2809.209993	1982.850014	-2809.209993	1982.85001	
Hours Worked	18.43262273	33.65820051	0.547641361	0.622	-88.68279314	125.5480386	-88.68279314	125.548039	
Number of Customers	1.072684096	1.696612738	0.632250408	0.572	-4.326698845	6.472063037	-4.326698845	6.47206304	
(Hours Worked)^2	-0.139304106	0.37126146	-0.536727152	0.629	-1.320726768	0.982253356	-1.380785768	0.98225336	
(Number of Customers)^2	-0.000696841	0.000617	-1.129402852	0.341	-0.002656041	0.001260728	-0.002656041	0.00126073	
Hours_High	1.217404804	2.401656307	0.506900429	0.647	-6.425759723	8.860568782	-6.425759723	8.86056878	
Customer_High	0.473686673	5.247913928	0.090261898	**0.934**	-16.22751762	17.17489096	-16.22751762	17.174891	<-- Variable with the highest p-value. Insignificant. To be dropped from the model
hours*customers	-0.022547212	0.041741771	-0.540159444	0.627	-0.155388156	0.110293732	-0.155388156	0.11029373	

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	(Number of Customers)^2	Hours_High	hours*customers
15.6	44	54	1936	2916	1	2376
15.7	43	52	1849	2704	1	2236
15.4	41	55	1681	3025	1	2255
14.3	41	55	1681	3025	1	2255
11.8	40	39	1600	1521	0	1560
9.7	40	28	1600	784	0	1120
9.6	40	37	1600	1369	0	1480
10.2	38	58	1444	3364	0	2204
11.3	38	67	1444	4489	0	2546
14.3	32	186	1024	34596	0	5952
14.8	37	226	1369	51076	0	8362

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.97866027
R Square	0.957658488
Adjusted R Square	0.894146221
Standard Error	0.803835004
Observations	11

ANOVA

	df	SS	MS	F	Significance F
Regression	6	58.45725153	9.742869221	15.07832307	0.010159184
Residual	4	2.584602857	0.646150714		
Total	10	61.04181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	-353.007711	303.4252338	-1.163407228	0.309	-1195.450615	489.4361951	-1195.450615	489.436395	
Hours Worked	15.74296328	13.57190431	1.159967159	0.311	-21.9386844	53.42461056	-21.9386844	53.4246106	
Number of Customers	0.937537184	0.691948559	1.35492324	0.247	-0.983620006	2.858694373	-0.983620006	2.85869437	
(Hours Worked)^2	-0.169695978	0.151477162	-1.12027434	0.325	-0.590264003	0.250872048	-0.590264003	0.25087205	
(Number of Customers)^2	-0.000674259	0.000489101	-1.378567505	0.240	-0.002032222	0.000683704	-0.002032222	0.0006837	
Hours_High	1.245990994	2.064534713	0.603521455	**0.579**	-4.486076303	6.978058291	-4.486076303	6.97805829	<-- Variable with the highest p-value. Insignificant. To be dropped from the model
hours*customers	-0.019111882	0.014865454	-1.285657477	0.268	-0.060385	0.022161235	-0.060385	0.02216124	

Sales	Hours Worked	Number of Customers	(Hours Worked)^2	(Number of Customers)^2	hours*customers
15.6	44	54	1936	2916	2376
15.7	43	52	1849	2704	2236
15.4	41	55	1681	3025	2255
14.3	41	55	1681	3025	2255
11.8	40	39	1600	1521	1560
9.7	40	28	1600	784	1120
9.6	40	37	1600	1369	1480
10.2	38	58	1444	3364	2204
11.3	38	67	1444	4489	2546
14.3	32	186	1024	34596	5952
14.8	37	226	1369	51076	8362

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.970628327
R Square	0.953802289
Adjusted R Square	0.90760578
Standard Error	0.750993421
Observations	11

ANOVA

	df	SS	MS	F	Significance F
Regression	5	58.22186259	11.64437252	20.64637567	0.002369702
Residual	5	2.819955596	0.563991119		
Total	10	61.04181818			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	-503.5082254	161.4969826	-3.117756241	0.026	-918.6494354	-88.36701538	-918.6494354	-88.367015	
Hours Worked	22.37113246	7.44962651	3.00239712	0.030	3.221257867	41.52100705	3.221257867	41.521007	
Number of Customers	1.294050132	0.364547582	3.849331167	0.012	0.428669973	2.159430291	0.428669973	2.1594303	<-- All independent variables are now significant at 5% significance level. We can stop here! :)
(Hours Worked)^2	-0.242365802	0.085865035	-2.822636744	0.037	-0.463088901	-0.021642703	-0.463088901	-0.0216427	
(Number of Customers)^2	-0.000940319	0.000137913	-4.751171492	0.005	-0.001449071	-0.000431567	-0.001449071	-0.0004316	
hours*customers	-0.026554515	0.007755129	-3.424212291	0.019	-0.046648971	-0.006619321	-0.046648971	-0.0066193	

Summary:

Overall Model Fit:

R Square (0.954): The model explains 95.4% of the variation in sales, a significant improvement from the previous model.

Adjusted R Square (0.908): Still very high, confirming the model's robustness even with additional predictors.

Significance F (0.00237): The overall model is statistically significant, as the p-value is well below 0.05.

Predictors and Coefficients:

Hours Worked: Significant (p = 0.030). A unit increase in hours worked increases sales by ~22.37 units, assuming all else is constant.

Number of Customers: Significant (p = 0.012). A unit increase in the number of customers increases sales by ~1.29 units, holding other factors constant.

(Hours Worked)^2: Significant (p = 0.037). Suggests a diminishing return to hours worked, indicating a quadratic relationship.

(Number of Customers)^2: Significant (p = 0.005). Suggests diminishing returns as the number of customers increases.

hours*customers: Significant (p = 0.019). Indicates that the interaction between hours worked and number of customers negatively impacts sales. This might occur if overworking employees or overserving customers reduces effectiveness.

Insights from Predictors:

Linear Effects: Hours Worked and Number of Customers positively influence sales on their own.

Quadratic Effects: Both predictors exhibit diminishing returns, meaning excessive increases in hours or customers start reducing sales effectiveness.

Interaction Effect: The negative interaction term implies that the simultaneous increase in hours worked and customers reduces sales, potentially reflecting operational inefficiencies or overextension.

Conclusion:

Given that none of the attempt to linearizing is successful, we went ahead and suspect that there is non-linearity in the data. We then introduced square(^2) to hours worked and numbers of customers, as well as introduced interaction terms. Using stepwise regression, we dropped variables that arent significant in the models.