IKEA PROFIT MODELING

Case 1 – Customer Order Analysis

ABSTRACT

A report on a predictive model of IKEA profits using a regression model and analysis based on customer order data provided by IKEA.

Harrison Michael
Prarthana Magon
Henry Harms
Micah Hertz

BNAD 277 Group 61

Table of Contents

Introduction

To determine what factors lead to the highest profits for IKEA, our team performed an in-depth visual and regression analysis. To conduct the visual analysis, Tableau was used. From our analysis in Tableau, we concluded that discounts could negatively impact profits for IKEA. Additionally, from our regression analysis model we determined that the south region, west region, discounts and sub-category of furniture were all significant in predicting the profits for IKEA. Further, we checked for common violations and made recommendations accordingly. We recommended that IKEA should focus on marketing office supplies and technology, keeping discounts below 25% or above 50%, and focusing on sales to the Central division.

Tableau

View exploratory Tableau infographic investigating trends of customer order data regions, profits, and product categories.

Statistical Analysis

Testing

Several tests were run to check the validity and reliability of the regression models. Firstly, an ANOVA joint test of significance was conducted to see if the variables were overall statistically significant, and with an F-stat of 422.95 and an estimated significance F stat of 0, the model was statistically significant as F-stat > significance F. Using the product order data from IKEA, a multitude of regression models were built using different combinations of quantitative and qualitative data variables, the latter of which were included through the creation of dummy variables in which qualities are indicated through the presence of the quality's predictive coefficient with a 1 or lack of the quality with no coefficient and a 0. By analyzing the Adjusted R-squared and Standard Errors of each of these models, the strongest model was selected for further analysis. Once determined, the strongest model's variables were individually tested for significance by comparing estimated p-values and the set confidence level of α = 0.10. Finally, residual plots and correlations between x-variables and regression residuals were investigated to look for common violations that may affect the validity of the model.

Regressions

Multiple regression models were run to find the best predictive power of the provided data. Each model version utilized the quantitative discount percentage, with qualitative variables of sales regions and product categories, as a baseline for predicting profits. The first model included consumer market segments as dummy variables, but these variables were found to be statistically insignificant, so they were not included in further models. The second model attempted to include shipping classes, but these were also found to be insignificant. Additionally, both models had low adjusted R-squared values around 0.06.

The next set of models sought to use more quantitative variables, so the quantity of products per order and revenue per order were included, resulting in an adjusted R-squared of 0.2799. The next model attempted to include the seasons in which orders were made by dividing the order months into three dummy variables for Winter, Summer, and Fall and a base case of Spring, but each variable was found to have p-values above our limit of 10%, restricting the data category from the regression. Finally, it was necessary to investigate other ways to represent discount percentages, as the residual plotting for the variable had signs of being non-linear or having endogeneity. Discount was broken down into dummy variables with a base case of no discount, where the variables represented orders with discounts under 25%, between 25% and 50%, and above 50%. This new interpretation of discount percentages gave an adjusted R-squared value of 0.2969, with an R-squared value of 0.2976. This, along with the regions, categories, quantities, and revenues were used to make up the final regression model and equation.

Estimated Regression Equation

$$\widehat{Profit} = 52.9738 - 21.2805(South_d) - 23.03(West_d) - 5.2487(East_d) - 17.5112(Furniture_d) + 4.5672(Technology_d) - 39.8948(Discount25_d) - 238.06539(Discount50_d) - 132.792(Discount50Plus_d) - 3.50156(Quantity) + 0.1892(Revenue)$$

$$R^2 = 0.2976, S_e = 196.4312$$

Figure 1 – Variables Legend

Variables Legend		
Name	**Description**	**Base Case**
$South_d$	dummy variable for Southern region	Central region
$West_d$	dummy variable for Western region	
$East_d$	dummy variable for Eastern region	
$Furniture_d$	dummy variable of an order being in the Furniture category	Office Supplies category
$Technology_d$	dummy variable of an order being in the Technology category	
$Discount25_d$	dummy variable of non-zero discounts under 25%	No discount (0%)
$Discount50_d$	dummy variable of discounts between 25% and 50%	
$Discount50Plus_d$	dummy variable of discounts above 50%	
Quantity	quantity of units in each order	
Revenue	revenue from each order	

Interpretations

The R-squared value of 0.2976 from the regression model signifies that the model is 29.76% the way towards perfectly predicting profits from product orders. The standard error of 196.43 means that the model's predictions are off from actual profits by $196.43 on average.

Coefficient Interpretations

- As revenue increases by $10, profit increases by $1.89 on average and all else constant
- As product quantity increases by 1, profit decreases by $3.50 on average and all else constant
- If an order has a discount that is equal to or under 25%, profit will be $39.89 less than an order with no discount, on average and all else constant
- If an order has a discount above 25% and under 50%, profit will be $238.07 less than an order with no discount, on average and all else constant
- If an order has a discount above 50%, profit will be $132.79 less than an order with no discount, on average and all else constant
- If an order is of a Technology product, it will have profits $4.57 higher than Office Supplies, on average and all else constant (statistically insignificant difference, see next section)
- If an order is of a Furniture product, it will have profits $17.51 lower than Office Supplies, on average and all else constant
- If an order is from the East region, it will have profits $5.25 lower than an order from the Central region, on average and all else constant (statistically insignificant difference, see next section)
- If an order is from the West region, it will have profits $23.03 less than an order from the Central region, on average and all else constant
- If an order is from the South region, it will have profits $21.28 less than an order from the Central region, on average and all else constant

Significance Testing

Figure 2 – Significance Testing of Regression Variables

P-values of Regression Model Variables			
Variable	**P-value**	**α**	**Significant**
South(d)	0.00096197	0.1	Yes
West(d)	0.000033605	0.1	Yes
East(d)	**0.34660969**	0.1	**No**
Furniture(d)	0.00095509	0.1	Yes
Technology(d)	**0.40040465**	0.1	**No**
Quantity	0.00010701	0.1	Yes
Revenue	0	0.1	Yes
Discount25(d)	1.60E-20	0.1	Yes
Discount50(d)	7.15E-130	0.1	Yes
Discount>50(d)	1.72E-68	0.1	Yes

The East region variable was found to be insignificant, but will be kept in the model to protect the interpretations of the rest of the dummy variables of order regions. This means that the coefficient of the East region variable cannot be interpreted, meaning there is not a significant difference between profits from sales in the East and Central regions. Similarly, the Technology product category was found to be insignificant, meaning there is no significant difference between the profits from Technology and Office Supply product orders.

Checks for Violations

An important element to building a strong regression model is making sure it does not suffer from a common violation that may occur in a regression. Firstly, heteroskedasticity, or changing variability, is evident through a residual plot of quantitative variables. The two numeric values, revenue and quantity, had residual plots that were mostly random and centered around the model's predictions. The plot for revenue does spread out at higher values, potentially indicating changing variability, but this may be due to large outlier from specific customers which is investigated in the Tableau. However, White's Standard Errors may be used to remedy this changing variability in future testing.

The next violation tests were done through correlation testing, where high correlation between the residuals and x-variables indicates endogeneity, or omitted variables, while high correlation between different x-variables signals multicollinearity. Two correlation tables were generated, where one included dummy variables and the other did not. While the table with dummies presented several errors as the inclusion of zeros meant correlations could not be checked, the values that were estimated showed very low correlations between residuals and variables, and correlations between variables all under the limit of ±0.80. Since correlation testing found that each correlation value was very low and did not test for statistical significance, variables can be generally trusted for predictions in this model.

Relevance

This regression model for predicting profits of IKEA sales revealed some interesting trends and findings that may prove useful to the decision making process of IKEA sales department. Based on the coefficients of the model, it predicts that increased quantity of products per order decreases profit by an average of $3.50 per additional item. Unsurprisingly, discounts do indeed decrease profits, but it was a surprise to find that discounts between 25% and 50% were the most reductive discount ranges of profit, with above 50% having a lesser effect, while discounts under 25% proved the strongest for retaining profits compared to having no discount at all. Testing also included regions and product categories. Within the regions, orders to the Central United States tend to be more profitable than orders to the West or South, while there is no significant difference in orders to the East. In the case of product categories, office supplies tend to be more profitable than furniture, while there is no significant difference between office supplies and technology.

Part of this testing process included utilizing different variables from the data to see if they had any effect on predicting profits. One model included the consumer market sections of orders, including Consumer, Corporate, and Home Office, but these categories did not test to be significant in predicting profit. Another model used the class of shipping per orders, with Standard, First Class, Second Class, and

Same Day. Similarly, these did not test to be statistically significant. Finally, a model was run dividing the dates of orders into four seasons, but these again were found to not have any predictive power.

Recommendations

Based on the relevant trends found through this regression model and analysis using order data to predict profits, there are some clear recommendations to IKEA to maximize their profits. Firstly, IKEA should focus on utilizing discounts of 25% or below or 50% or above, as these are the less destructive options of discounts compared to having no discount at all. Additionally, IKEA should focus on marketing office supplies and technology, as furniture sales tend to generate less profit. Finally, US marketing should focus more on the Central division, as it is more profitable compared to other US regions, as well as incentivizing smaller orders as larger orders decrease profits.

Conclusion

This report investigated customer order data from IKEA in order to build a predictive regression model of profits based on product, region, discount, and sales data from individual customer orders. Findings in this analysis include office supplies being the most profitable product category, discounts between 25% and 50% having the greatest detriment on profits, higher quantity orders decreasing profits per additional item, and the Central division being the most profitable region of US sales. These findings may be used in future decision-making processes to optimize profits from US product orders. As this model was far from a perfect predictor of profits, it is advised that further investigation be conducted utilizing more samples and additional types of data to build a stronger model. Additionally, White's Standard Errors should be implemented to account for changing variability found between predicted and actual order revenues. For any questions or interest in further studies, please contact ellersalesanalysis@ikea.com.

Appendix

Residual Plots

Figure 3 – Quantity Residual Plot

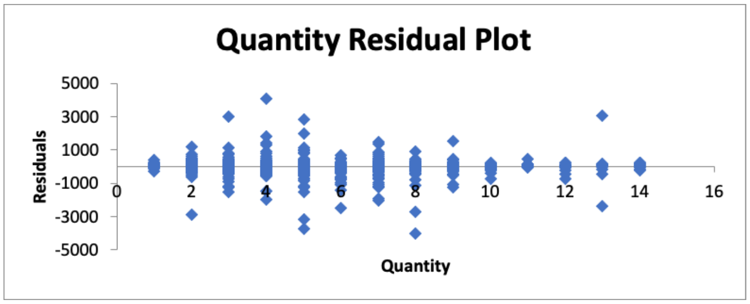
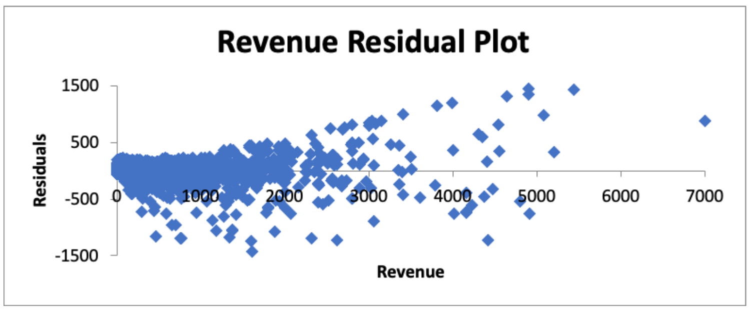


Figure 4 – Revenue Residual Plot



Regression Summary Output

Figure 5 – Regression Summary Output

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.545520085							
R Square	0.297592163							
Adjusted R Square	0.296888559							
Standard Error	196.431178							
Observations	9994							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	10	163197707.7	16319770.77	422.9540738	0			
Residual	9983	385196128.2	38585.20768					
Total	9993	548393836						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	52.97383269	6.079668609	8.713276348	3.42096E-18	41.05645629	64.89120909	42.97273967	62.97492572
South(d)	-21.2805115	6.443944127	-3.30240472	0.000961965	-33.91194137	-8.64908163	-31.88084004	-10.68018296
West(d)	-23.03000095	5.550146924	-4.149439874	3.36054E-05	-33.90940808	-12.15059383	-32.16002748	-13.89997442
East(d)	-5.248709188	5.576435353	-0.941230169	0.346609694	-16.17964693	5.682228559	-14.42198035	3.924561975
Furniture(d)	-17.51122834	5.299337968	-3.304418107	0.00095509	-27.89899934	-7.123457342	-26.22867257	-8.793784115
Technology(d)	4.567241022	5.431148154	0.840934714	0.400404651	-6.078904519	15.21338656	-4.367031786	13.50151383
Discount25(d)	-39.8948459	4.286690584	-9.306677287	1.59666E-20	-48.29762383	-31.49206797	-46.94647882	-32.84321297
Discount50(d)	-238.0653918	9.673997179	-24.60879277	7.1526E-130	-257.028377	-219.1024066	-253.9791779	-222.1516057
Discount>50(d)	-132.7919872	7.534310632	-17.62496845	1.71698E-68	-147.5607552	-118.0232191	-145.1859754	-120.3979989
Quantity	-3.501556065	0.903472533	-3.875664106	0.000107015	-5.272544408	-1.730567721	-4.987774053	-2.015338077
Revenue	0.189237084	0.003321846	56.96744715	0	0.182725596	0.195748572	0.183772627	0.194701542

Correlations

Figure 6 – Correlation Tables With and Without Dummy Variables

Correlation Test with Dummies

	Residuals	*South(d)*	*West(d)*	*East(d)*	*Furniture(d)*	*Technology(d)*	*Discount25(d)*	*Discount50(d)*	*Discount>50(d)*	*Quantity*	*Revenue*
Residuals	1										
South(d)	#DIV/0!	1									
West(d)	#DIV/0!	#DIV/0!	1								
East(d)	#DIV/0!	#DIV/0!	#DIV/0!	1							
Furniture(d)	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1						
Technology(d)	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1					
Discount25(d)	-0.0061005	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1				
Discount50(d)	0.0017912	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.185212854	1			
Discount>50(d)	0.00662626	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.238607693	-0.072433923	1		
Quantity	4.1963E-16	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.01020778	0.003194447	-0.00571486	1	
Revenue	1.8282E-16	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.013525095	0.012387568	0.00976448	0.20079477	1

Correlation Test Only Quantitative Variables

	Residuals	*Quantity*	*Revenue*
Residuals	1		
Quantity	4.1963E-16	1	
Revenue	1.8282E-16	0.20079477	1