Predicting Catalog Demand

REVIEW | HISTORY

Meets Specifications

Hey!

Wonderful attempt! I appreciate your efforts on the same, especially the well laid and thorough explanation- great job indeed!! I really liked your business decision making skills explored in the project along with data analysis. You are one step closer to graduating from the Nanodegree 🎉👏🏻

Let us know if you have any queries or issues on https://knowledge.udacity.com/. I have included certain remarks in the reviews in order to help you. I hope they'll be useful. Don't worry, we want the best from you; hence gave you all the possible areas of improvement.

If you are satisfied, please feel free to rate the review and provide your feedback 😄

Wishing you all the best for future endeavors. Good luck 😊👍🏻

Business and Data Understanding

✓ The section is written clearly and is concise. The section is written in less than 500 words.

1. What decisions need to be made?
   The management is supposed to decide whether or not to send this year's catalog out to the 250 new customers from their mailing list.
   The management wants to send the catalog out to these new customers only if it could result in generation of profit in excess of $10,000.

2. What data is needed to inform those decisions?

   We need to calculate the **expected revenue** from these 250 customers in order to get the **expected profit**. To do so we've to predict the average sales amount from each of the 250 customers and then multiply that amount with the probability that a customer is likely to make a purchase inorder to get the expected revenue. With the average gross margin being 50% we multiply the expected revenue by 0.5 then subtract $6.50 cost inorder to get the expected profit from a respective customer. We will then aggregate/sum up the profit amounts to find out if the total exceeds the $10,000 mark.

✓ All following questions have been accurately answered:

- What decisions need to be made?
- What data is needed to inform those decisions?

Q1 ✅
Awesome: That's absolutely correct! The main decision here is that the company wants to determine if the expected profit from these customers satisfies the constraint of being minimum $10,000 and then decide to send the catalog out to these customers or not.

Q2 ⚠️
Comment: It is advised to list down the variables useful for the analysis, since we're asked "What data is needed?"

> For example- Demographic data (Gender, region), Sales purchase data etc.

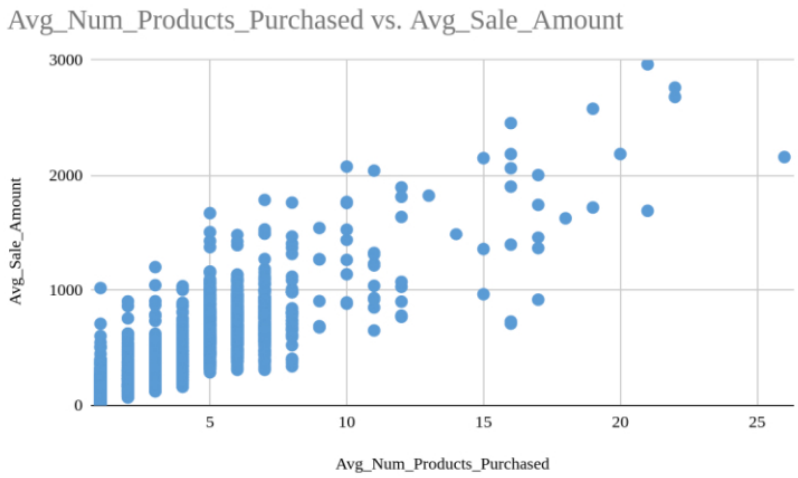
Analysis, Modeling, and Validation

✓ The section is written clearly and is concise. The section is written in less than 500 words.

✓ Each predictor variable is shown to have a linear relationship between the target variable whenever appropriate.

Each predictor variable should be significant (p-value <= 0.05)

✅
Awesome: You have used correct predictor and target variables. Amazing job showing the scatterplot for better understanding on the desired relationship between the target and the predictor variable! 👍🏽



COMMENT

Customer Segment is a categorical variable. As business analysts, we need to justify decisions based on the data. How to do that in cases with categorical variables?

As such, we cannot use a scatterplot to estimate its linear relationship with the target variable. In such cases, please feel free to use **Association Analysis tool** to find correlation between the predictor and target variable. For more information, please check Linear Regression Validation. This will be useful for both continuous and categorical predictor variables.

✓ p-values and R-squared values are used to justify how well the linear model works.

✅
Awesome: The justification regarding the p-values and R-Squared values being good for your model is presented briefly. Good job!

The following are the p-values of the predictor variables that I included in the model

Predictor Variable	p-value
Customer_SegmentLoyalty Club Only	2.2e-16
Customer_SegmentLoyalty Club and Credit Card	2.2e-16
Customer_SegmentStore Mailing List	2.2e-16
Avg_Num_Products_Purchased	2.2e-16

The following are the R-Squared values of the models:
Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366

Based on the p-values of the predictor variables and R-squared values we can conclude that this is a viable model that can be used to predict average sales amounts of the new customers.

These values indicate how significant are these variables when selected with other variables. Let me help you understand these in a better manner-

Adjusted R Squared
The higher the Adjusted R Squared value, the higher the explanatory power of the model. This represents the amount of variation in the target variable explained by the variation in the predictor variables. Any model with an adjusted r-square value above 0.70 is considered to be a strong model. Our present linear model has a value of 0.84, hence it is a good model.

p-value
The lower the p-value, the higher is the probability that your results are significant and have not been obtained randomly. It indicates that a statistically significant relationship exists between the predictor variable and target variable. When a predictor variable has a p-value below 0.05, we can conclude that the relationship between it and the target variable is considered to be statistically significant. On the other hand, if the p-value is high, it indicates that the relationship between the predictor variable and the target variable is random and not statistically significant.

**For more information, please feel free to read this excellent article on p- values*

✓ The regression equation given is correct. Each coefficient should have up to 2 digits after the decimal figures (ex: 1.28).

✅
Awesome: The linear equation is correct since it contains the correct predictor variables with correct coefficients up to 2 decimal places!

Equation:
Avg_Sales_Amount = 303.46 + Avg_Num_Product_Purchased x 66.98 + CustomerSegmentLoyalty Club Only x -149.36 + CustomerSegmentLoyalty Club and Credit Card x 281.84 + CustomerSegmentStore Mailing List x -245.42

Presentation/Visualization

✓ The section is written clearly and is concise. The section is written in less than 500 words.

1. What is your recommendation? Should the company send the catalog to these 250 customers?

I recommend that the company should send the catalog to the 250 customers because the predicted expected profit exceeds the $10,000 mark.

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)

I used past data of last year's customers to generate a linear regression model that can predict the average sales amount of the 250 new customers.
Using the Alteryx linear regression tool I found out that only the Avg_Num_Product_Purchased and CustomerSegment dummy variables had high correlation with the target variable(Avg_Sales_Amount). Using these variables to generate a model resulted in a model with an Adjusted R_Squared value of 0.8366 that indicates it to be a viable model.

Using the Alteryx score tool I applied the model to the data of the 250 new customers in order to get the expected Avg_Sales_Amount of each customer.
Using the Alteryx function I then multiplied the average sales amount from each of the 250 customers with the probability that a customer is likely to make a purchase inorder to get the expected revenue.
With the average gross margin being 50% I multiplied the expected revenue by 0.5 then subtracted $6.50 cost inorder to get the expected profit from a respective customer.
Using the Alteryx Summarize tool I then summed up the profit amounts to get the total which resulted in $21,987.44.

✓ All questions have been accurately answered and the recommendations are well justified.

- What is your recommendation?
- How did you come up with your recommendation?
- What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?

✅
Awesome: The recommendation is absolutely correct and the recommendation process has been justified well. Appreciate the process explained step-by-step! 👍🏽

✓ The profit calculation is correct.

✅
Awesome: The expected profit is absolutely correct. 👏🏽

3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?

The expected profit from the new catalog is $21,987.44

DOWNLOAD PROJECT

RETURN TO PATH