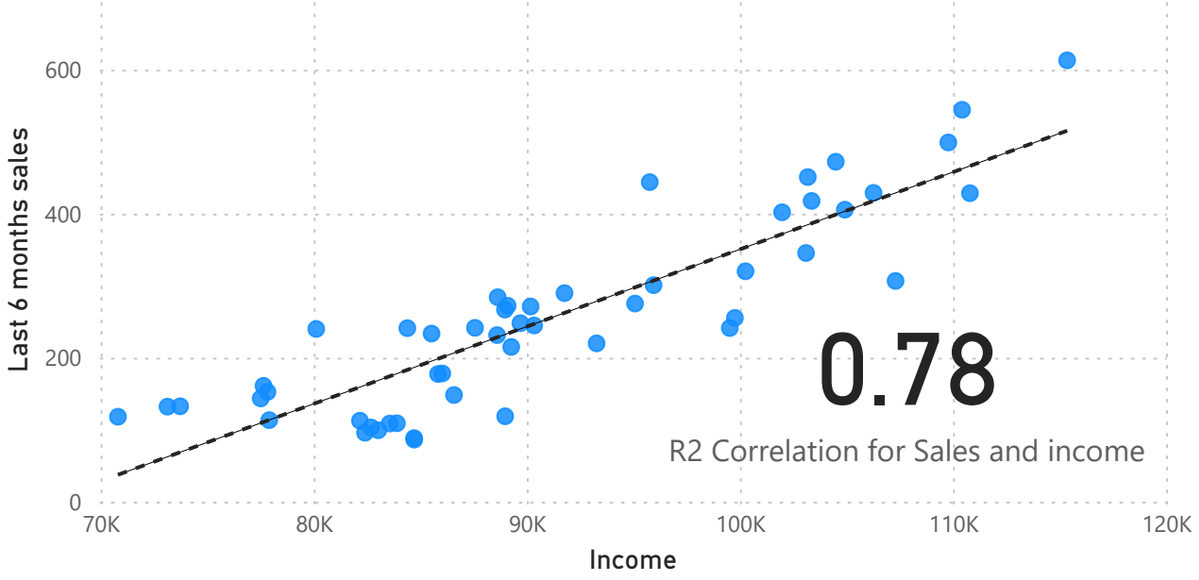
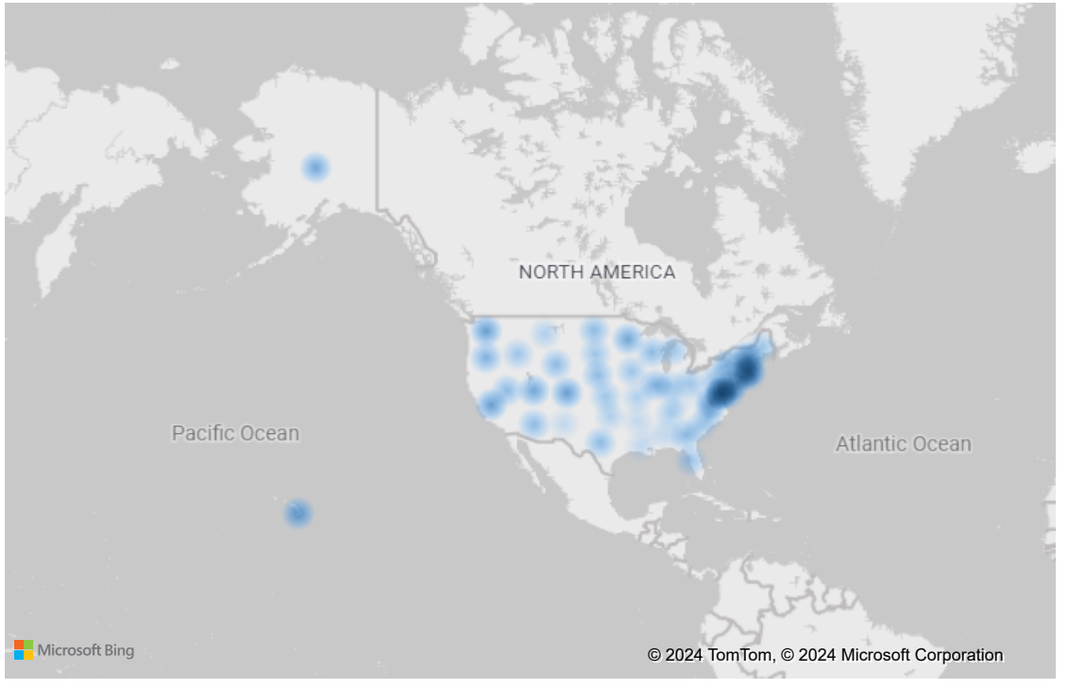
Average income and 6 months sales correlation



Sum of Average Income by State



0.69

R2 Correlation of Customer rating and Return rate

34,805.63

Standard deviation of Predicted Customer Income

$y = 0.01X + -722.14$

Y Formula

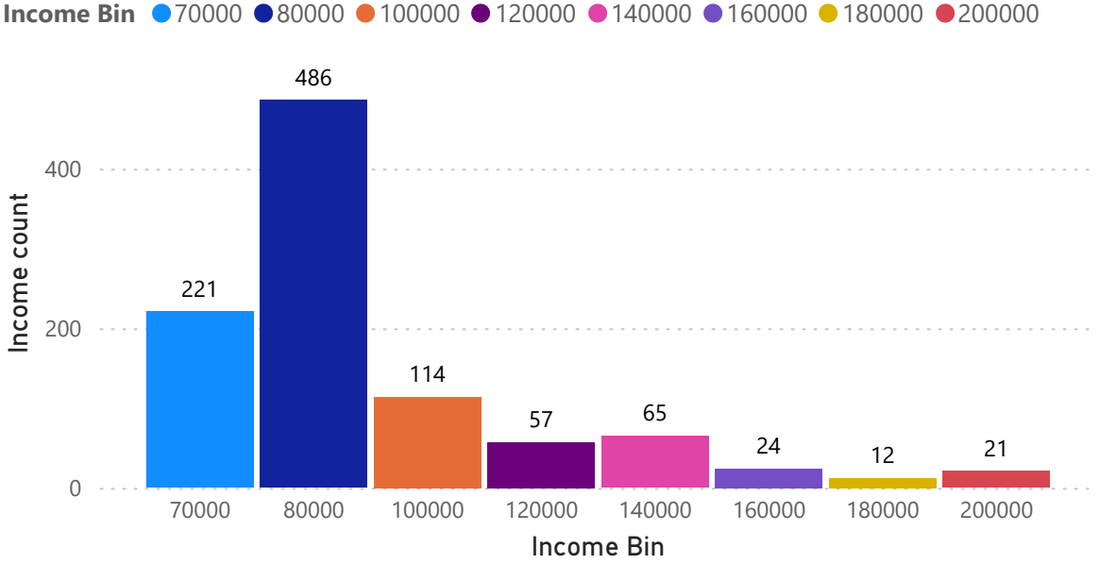
0.011 m

-722.14 b

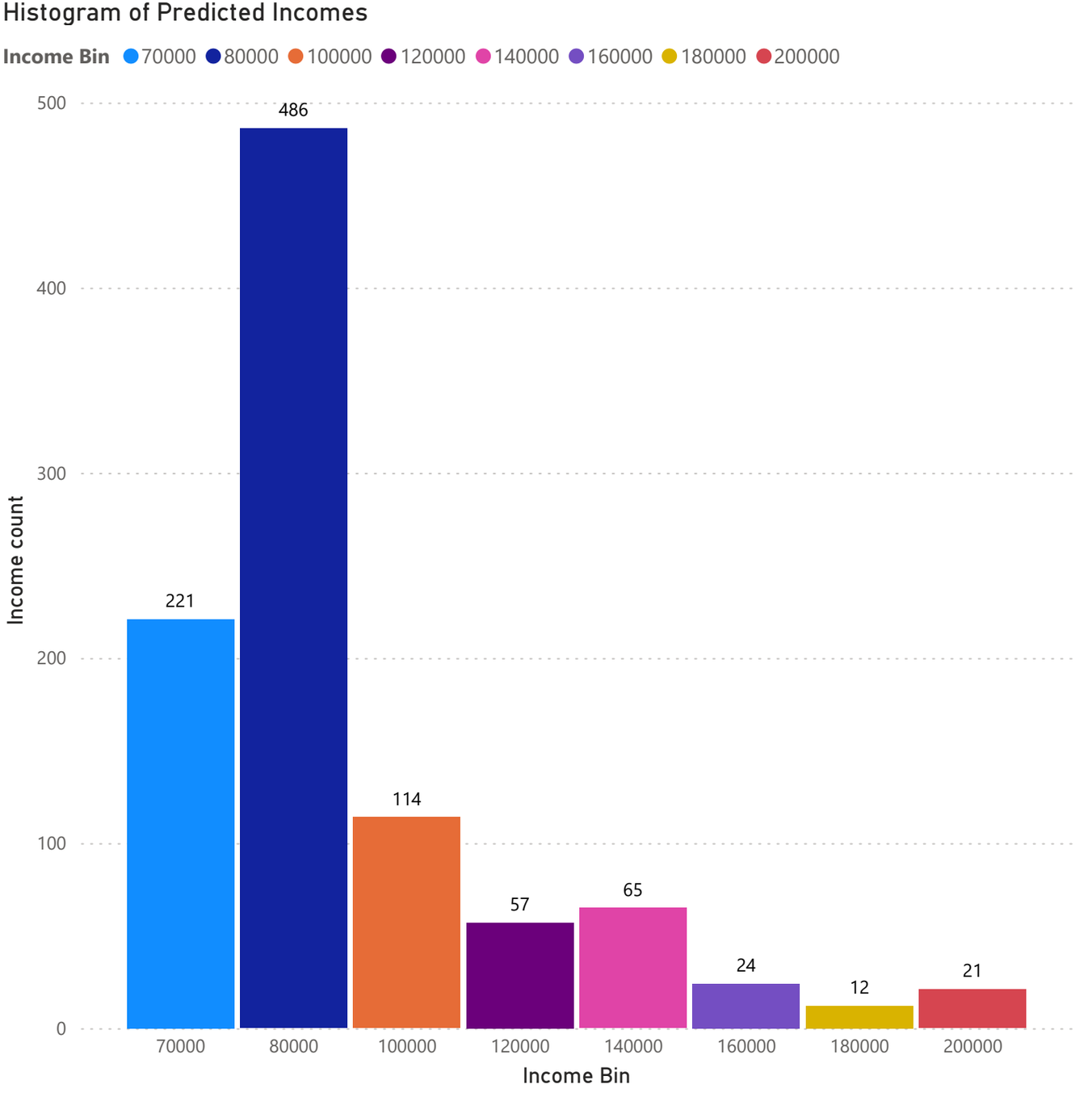
$x = -722.14 - Y / -0.01$

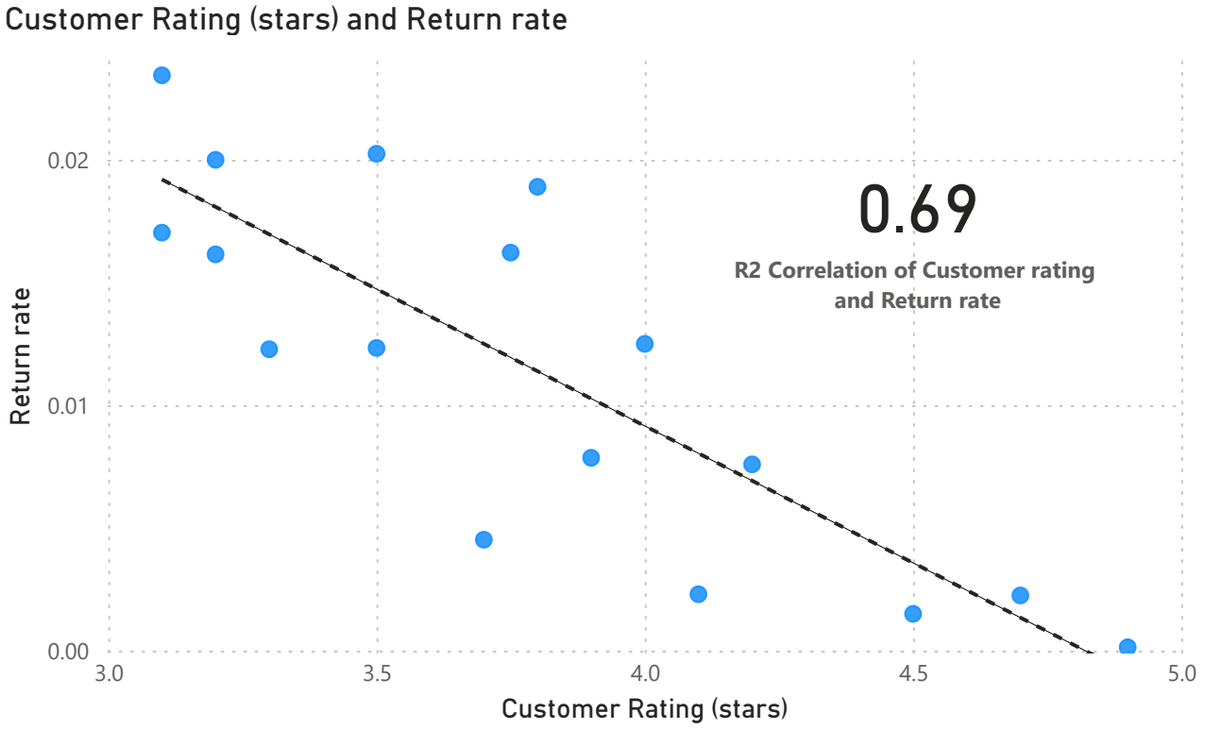
X Formula

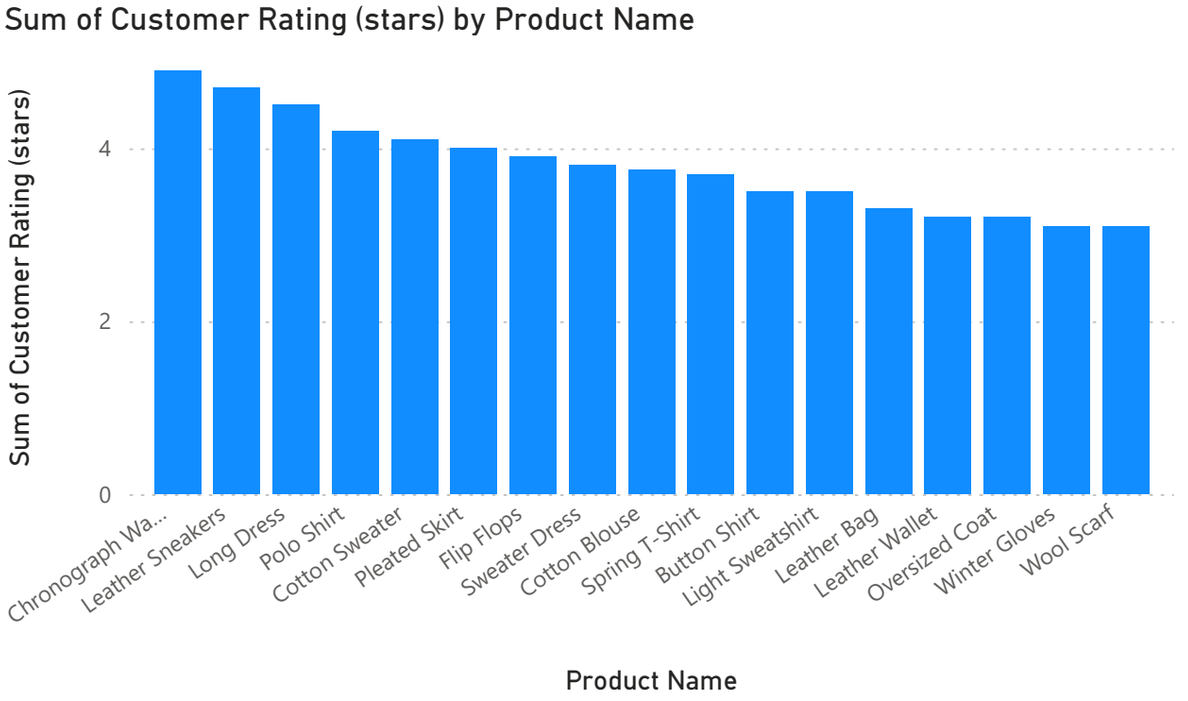
Histogram of Predicted Incomes

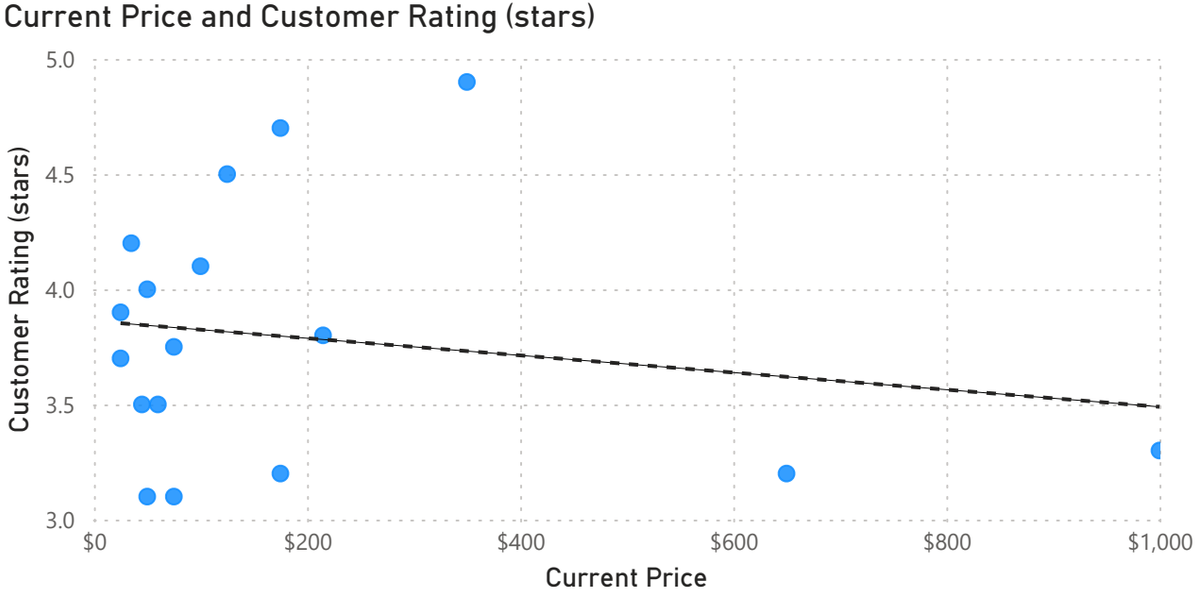


Full Name	Sum of Predicted Customer Income	Sum of Last 6 Months Purchases	Sum of Income Bin
Jon Little	597,214.00	5,250.00	200000
Jody Foster	322,214.00	2,500.00	200000
Tonya Rodriguez	307,214.00	2,350.00	200000
Edna Barnett	249,566.00	1,773.52	200000
Larry Cain	245,814.00	1,736.00	200000
Floyd Quinn	220,614.00	1,484.00	200000
Margarita Walters	218,878.00	1,466.64	200000
Sue Burgess	217,310.00	1,450.96	200000
Mercedes Stanley	214,510.00	1,422.96	200000
Nichole Schultz	210,454.00	1,382.40	200000
Rickey Green	209,950.00	1,377.36	200000
Juana Cohen	207,790.00	1,355.76	200000
Raymond Santos	207,646.00	1,354.32	200000
Damon Sullivan	206,854.00	1,346.40	200000
Charlene Sparks	206,614.00	1,344.00	200000
Craig Tran	206,494.00	1,342.80	200000
Lorena Strickland	206,422.00	1,342.08	200000
Todd Bush	205,702.00	1,334.88	200000
Lola Sanders	205,486.00	1,332.72	200000
Ed Diaz	205,398.00	1,331.84	200000
Gerard Owens	205,270.00	1,330.56	200000
Darlene Henry	197,214.00	1,250.00	180000
Erik Watson	197,214.00	1,250.00	180000
Estelle Harrington	197,214.00	1,250.00	180000
Frederick Rivera	197,214.00	1,250.00	180000
Genevieve Matthews	197,214.00	1,250.00	180000
Gladys Ward	197,214.00	1,250.00	180000
Harvey Chapman	197,214.00	1,250.00	180000
Lionel Greer	197,214.00	1,250.00	180000
Lisa Herrera	197,214.00	1,250.00	180000
Nora Hicks	197,214.00	1,250.00	180000
Seth Lambert	197,214.00	1,250.00	180000
Total	**100,815,580.00**	**286,015.80**	**91890000**




Customer Rating (stars) and Return rate
0.02
0.01
0.00
Return rate
3.0
3.5
4.0
4.5
5.0
Customer Rating (stars)
0.69
R2 Correlation of Customer rating and Return rate


Sum of Customer Rating (stars) by Product Name
Sum of Customer Rating (stars)
4
2
0
Chronograph Wa...
Leather Sneakers
Long Dress
Polo Shirt
Cotton Sweater
Pleated Skirt
Flip Flops
Sweater Dress
Cotton Blouse
Spring T-Shirt
Button Shirt
Light Sweatshirt
Leather Bag
Leather Wallet
Oversized Coat
Winter Gloves
Wool Scarf
Product Name


Current Price and Customer Rating (stars)
5.0
4.5
4.0
3.5
3.0
Customer Rating (stars)
$0
$200
$400
$600
$800
$1,000
Current Price

Summary report:

Using the linear regression method on average incomes and average sales, the scatter-plot showed that increases in customer income result in increases of sales, and vice versa. Income being the independent variable(X) and sales as dependent (Y). To address the questions, valuable data were uncovered:

- **1-2.**After thorough analysis, The resulting correlation R2 value between sales and income is **0.78**. As for the correlation R2 value between customer rating and return rate, the result is **0.69**.

- **3.** The linear regression formula used to predict customer incomes, is by the use of $x = b\text{-}y/\text{-}m$ formula , retrieving the intercept(b = 0.011) from the regression table minus the average sales divided by the negative value of the slope(m=-722.14) which is also negative resulting in a positive value. The linear regression formula used to predict customer sales, is by the use of $y= m(x) + b$, multiplying the slope value with the available Average income data plus the intercept value.

- **4**. To predict which customer has the highest income, A matrix table was used followed by sorting by descending order to identify the customer with the highest predicted income. The customer with highest predicted income is "*Jon Little*".

- **5**. After using the linear regression method on customer ratings and return rates, it was found that an increase in customer ratings(X) result in a decrease of return rates(Y), and vice versa. Therefore by determining which product has the highest customer rating, will be advertised the most, which is the "Chronograph watch".

Additional variables were added for further analysis, which are the current price of products as the independent variable and customer ratings as dependent to help determine market conditions. Although it does have a slight positive relationship, It was not enough to have a correlation since there are not enough high priced products to compete with lower cost products.

After observation, some key elements from the available data were believed to be missing . The "purchase list" did not contain a unique identifier for which products were purchased by individual customers. Therefore, It is recommended to record which products were sold to which customer, that might help with better and more accurate analysis in the future.