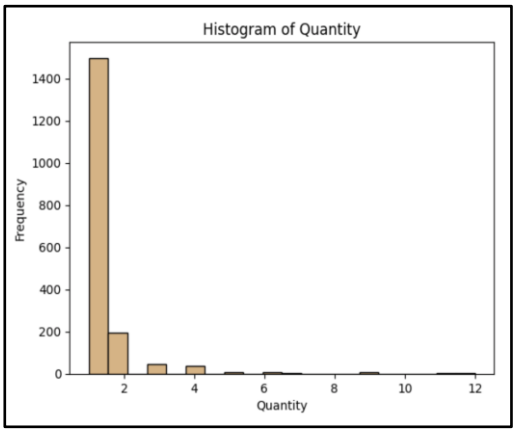
Jewlery Store Sales Analysis: Executive Summary

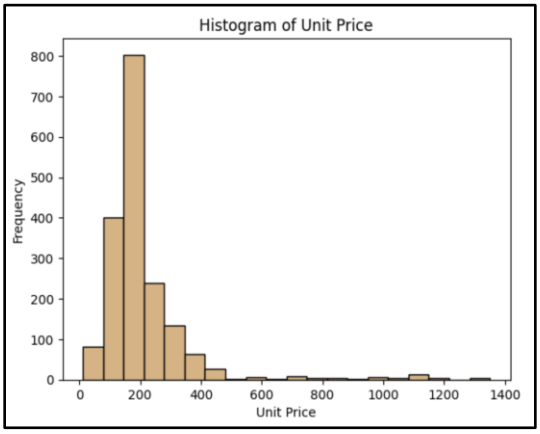
Rishi Sunkavalli

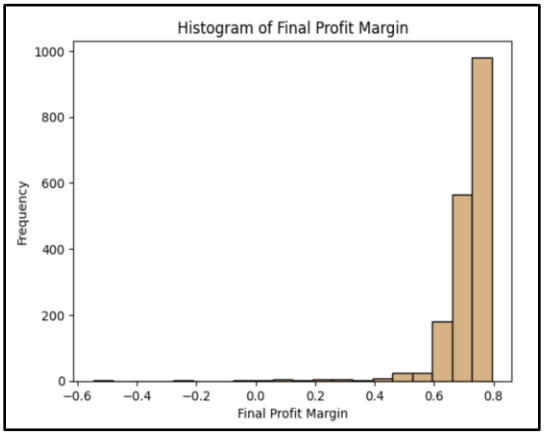
For this project, the dataset is derived from Signet, one of the largest jewelry retailers in the US, encompassing brands like Zales and Kay Jewelers. Signet primarily sells jewelry made from three types of materials: gold, silver, and 14k gold. The customers consist of members and non-members and are distributed throughout the United States. Overall, I was interested in investigating if there was a noticeable difference in purchases made between nonmembers and members. As a result, I used two different hypotheses to analyze if there was a difference in profit margin and in purchase quantity.

Prior to the regression analysis, I conducted comprehensive statistical tests to identify the variables within the main regression model with predictive power. The dependent variables include Quantity Purchased (continuous) and Final Profit Margin (calculated based on quantity, unit costs, and revenue) and the independent variables include Non-members vs members, location, price, and material type. The unit prices predominantly fall within the range of 0 to 500, with an average of around 200, showcasing a right-skewed distribution. The quantity distribution is also right-skewed, suggesting a focus on lower quantities. When it comes to profit margins, the distribution is skewed to the left. To ensure a more precise analysis, outliers with final profit margins less than 0.5 were excluded.

Histograms 1, 2, & 3: Quantity, Unit Price, and Final Profit Margin:







My first hypothesis focused on purchase quantity as the dependent variable, with customer type, region, and unit price as the independent variables. The null hypothesis (H0: $\beta i = 0$) stated that there was no statistically significant relationship between the quantity purchased and the customer type and location. Whereas, my alternative hypothesis (Ha: $\beta i \neq 0$), suggested a discernible relationship. My primary goal for the first regression was to explore how specific independent variables (customer type, unit price, and region) influence the quantity of items purchased.

Regression 1: Customer Type, Unit Price, & Region

OLS Regression Results

Dep. Variable:	Quantity	R-squared:	0.277
Model:	OLS	Adj. R-squared:	0.276
Method:	Least Squares	F-statistic:	228.8
Date:	Thu, 30 Nov 2023	Prob (F-statistic):	1.15e-125
Time:	23:39:27	Log-Likelihood:	-2192.1
No. Observations:	1795	AIC:	4392.
Df Residuals:	1791	BIC:	4414.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.5100	0.060	41.961	0.000	2.393	2.627
State_west	-0.0077	0.039	-0.198	0.843	-0.084	0.069
Customer Type_Non Member	-1.2197	0.047	-26.143	0.000	-1.311	-1.128
Unit Price	-0.0019	0.000	-12.895	0.000	-0.002	-0.002

Omnibus:	1885.996	Durbin-Watson:	0.989
Prob(Omnibus):	0.000	Jarque-Bera (JB):	134830.881
Skew:	5.112	Prob(JB):	0.00
Kurtosis:	44.209	Cond. No.	965.

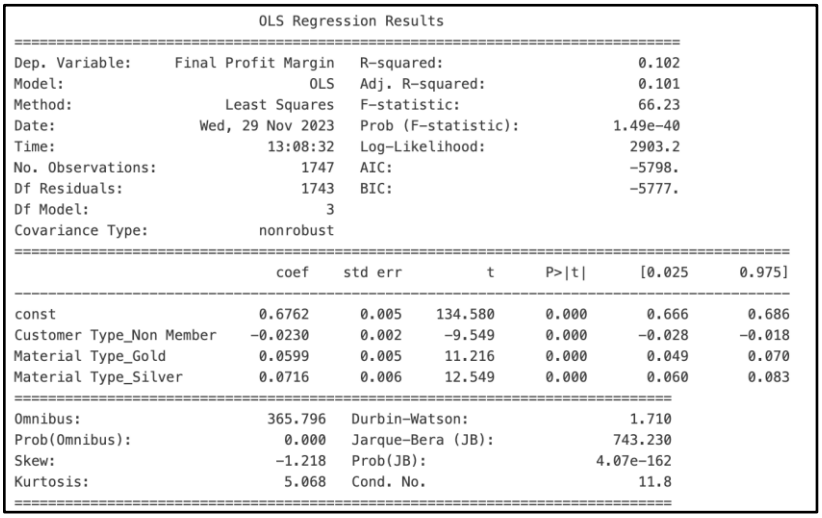
The regression revealed that both customer type and unit price were statistically significant. The p-values for both of these variables (~ 0.000) were below the predetermined alpha of 0.05, which led me to reject the null hypothesis. This rejection provided strong statistical evidence supporting the assertion that 'Customer Type' and 'Unit Price' are related to changes in the quantity of items purchased. However, the region variable did not have the same results. To analyze the 'Region' variable, I converted it into a dummy variable to assess its quantitative impact on the quantity purchased. The analysis yielded a p-value of 0.843, far greater than alpha of 0.05.

Consequently, I failed to reject the null hypothesis, indicating that this regression showed no significant relationship between the region customers live and the quantity they purchase. Furthermore, I introduced interaction terms to consider both the Unit Price and the customer's membership status. This adjustment was necessary as higher-priced items tended to be purchased in lower quantities, and members generally made more purchases compared to non-members.

The derived regression formula was Quantity = 2.51 - 1.2197 * Non_Member - 0.0019 * Unit Price. This formula suggests that by holding unit price constant, if the customer is a non-member, the quantity purchased would be decreased. Additionally, for each $1 increase in the unit price, there's an estimated decrease of around 2.5 units in the quantity purchased, maintaining the customer type constant.

The second regression's goal was to explore the relationship between the independent variables, material type and customer type with the dependent variable, final profit margin. The null hypothesis ($H0: \beta i = 0$) stated there was no statistically significant relationship between the final profit margin and the customer type and material type. Conversely, the alternative hypothesis ($Ha: \beta i \neq 0$) proposed a relationship between the variables.

Regression 2: Material Type & Customer Type



OLS Regression Results

Dep. Variable:	Final Profit Margin	R-squared:	0.102
Model:	OLS	Adj. R-squared:	0.101
Method:	Least Squares	F-statistic:	66.23
Date:	Wed, 29 Nov 2023	Prob (F-statistic):	1.49e-40
Time:	13:08:32	Log-Likelihood:	2903.2
No. Observations:	1747	AIC:	-5798.
Df Residuals:	1743	BIC:	-5777.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6762	0.005	134.580	0.000	0.666	0.686
Customer Type_Non Member	-0.0230	0.002	-9.549	0.000	-0.028	-0.018
Material Type_Gold	0.0599	0.005	11.216	0.000	0.049	0.070
Material Type_Silver	0.0716	0.006	12.549	0.000	0.060	0.083

Omnibus:	365.796	Durbin-Watson:	1.710
Prob(Omnibus):	0.000	Jarque-Bera (JB):	743.230
Skew:	-1.218	Prob(JB):	4.07e-162
Kurtosis:	5.068	Cond. No.	11.8

The regression model revealed that both variables are highly significant as the p-values fell below alpha of 0.05, and as a result, I rejected the null hypothesis. The regression formula derived from this regression was: Final Profit Margin = 0.6762 - 0.023*Non_Member* + 0.0599*Gold* + 0.0716*Silver*. This formula demonstrates that if everything else is held fixed, the non-member will have 0.023 less final profit margin than the member. From this regression formula, I wanted to investigate which type of material generated the highest profit margin. From Table 1 below, I concluded that silver for both members and nonmembers had the highest profit margin.

Table 1

	Member	Non-member
14K	0.6762	0.6532
Gold	0.7361	0.7131
Silver	0.7478	0.7248

Based on the regression model, I can tell that members indeed generate higher profit margins across regions and material types, indicating that the company should retain existing members through membership programs and regular engagements, meanwhile designing more marketing campaigns such as writing targeted emails to convert non-members into members. In addition, contrary to what I hypothesized at first, the silver material type is the one that generates more profit, demonstrating that a higher unit price does not necessarily lead to higher profit. Based on this insight, the jewelry company should invest more in the silver product line. Meanwhile, they should figure out a way to decrease costs for other product lines to maximize profits. Finally, the failure to reject the null hypothesis for the location variable assures the company to expand its customer base all around the U.S. Admittedly, only categorizing the West Coast and East Coast will not be detailed enough to assist the company in identifying the most profitable regions. Regarding future improvements, I should investigate the location in terms of different states to better understand this question.