



RESEARCH
PROBLEM

MULTIPLE
OUTCOMES

IDEAL
EXPERIMENT

DESCRIPTIVE
STATISTICS

SUPERSTORE DATASET ANALYSIS

Sophie Garrett, Karen Botros,
Grace Hsu, Seongjae Liew, Jesslyn
Noorjono, Chris Why, Vinni Guan





OVERVIEW OF DATASET

Information related to sales and profits for various products with varying discount amounts in different locations for a national Superstore company.

Variables analyzed:

- Sales
- Discount
- Product Category (Technology, Furniture, Office Supplies)
- Region (East, West, Central, South)



RESEARCH PROBLEM

Does offering discounts for different product categories affect sales within each region?



■ NULL VS ALTERNATIVE HYPOTHESIS

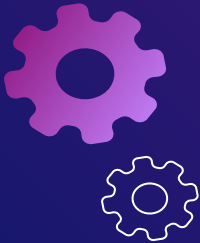
H_0

- No difference in sales of furniture across the regions based on the discount
- No difference in sales of office supplies across the regions based on the discount
- No difference in sales of technology across the regions based on the discount

VS

H_A

- A difference in sales of furniture across the regions based on the discount
- A difference in sales of office supplies across the regions based on the discount
- A difference in sales of technology across the regions based on the discount






■ MULTIPLE OUTCOMES



- All the regions might respond positively to discounts, or that all the regions respond negatively to discounts.
- A chance that some regions will respond positively to discounts and others won't.
- There might be an interaction effect between the region and category



IDEAL EXPERIMENT

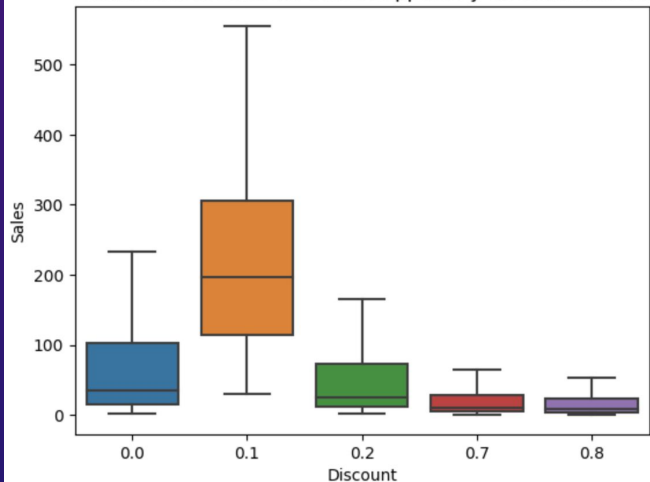
-  Randomly pick superstores within each region without any limitations
-  Implement discount strategies and put products on a discount and see if sales change
-  Compare superstores within each region that offer no discounts to those in the same region that do have discounts.





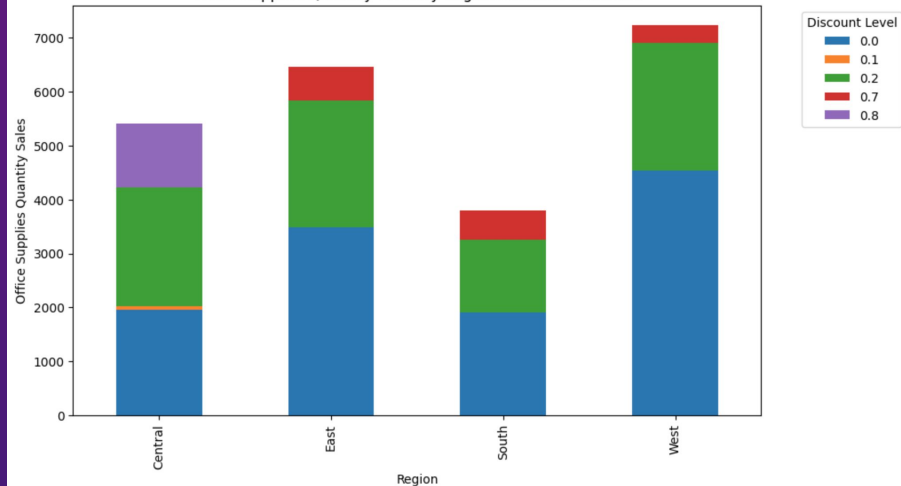
EDA - OFFICE SUPPLIES

Distribution of Sales for Office Supplies by discount amount



- H_0 : Office Supplies Sales remains constant across the different discount levels
- H_A : Office Supplies Sales is different across the different discount levels
- **ANOVA Statistic:** 8.2943
- **P-value:** <.01

Office Supplies Quantity Sales by Region and Discount Level

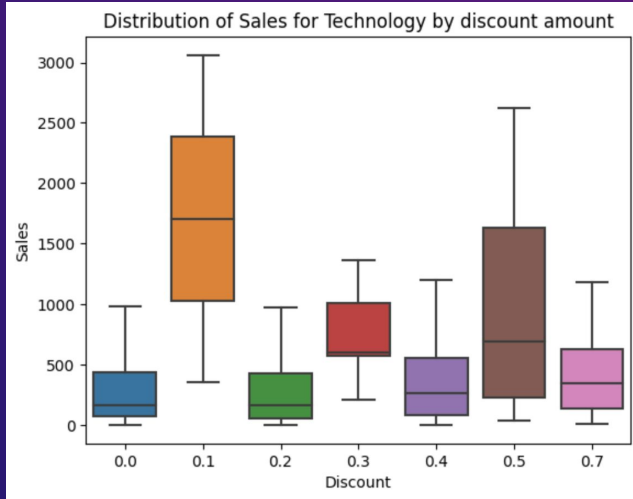


- H_0 : Office Supplies Sales remains constant across the different regions
- H_A : Office Supplies Sales is different across the different regions
- **ANOVA Statistic:** 0.1597
- **P-value:** .92

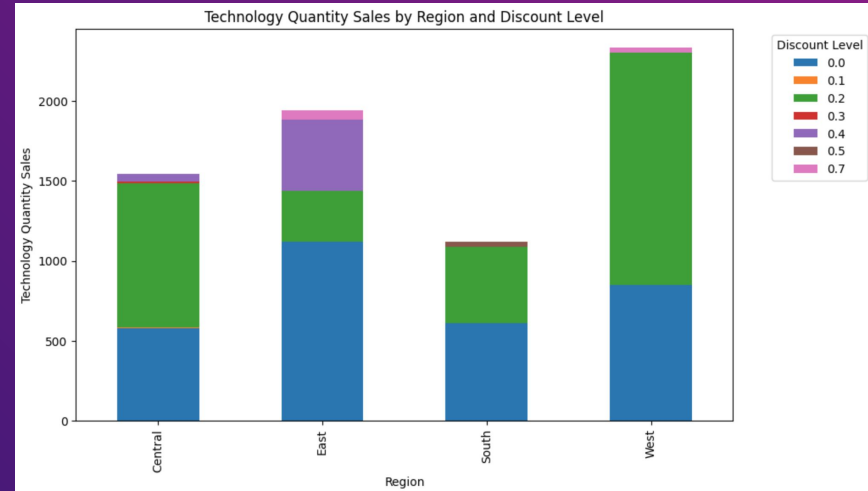




EDA - TECHNOLOGY



- H_0 : Technology Sales remains constant across the different discount levels
- H_A : Technology Sales is different across the different discount levels
- **ANOVA Statistic:** 14.0196
- **P-value:** <.01

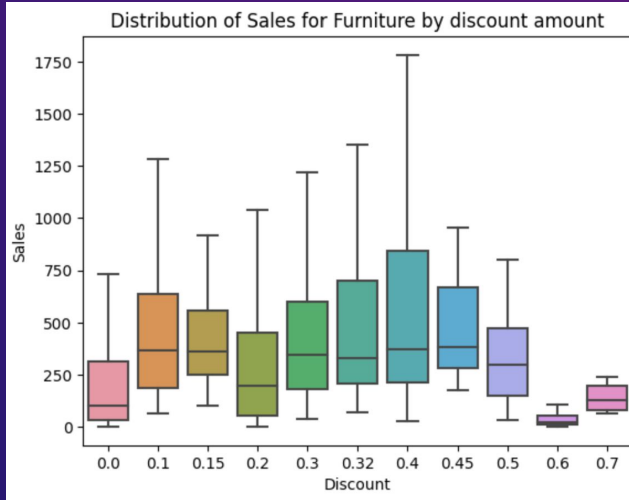


- H_0 : Technology Sales remains constant across the different regions
- H_A : Technology Sales is different across the different regions
- **ANOVA Statistic:** 0.9213
- **P-value:** .4297

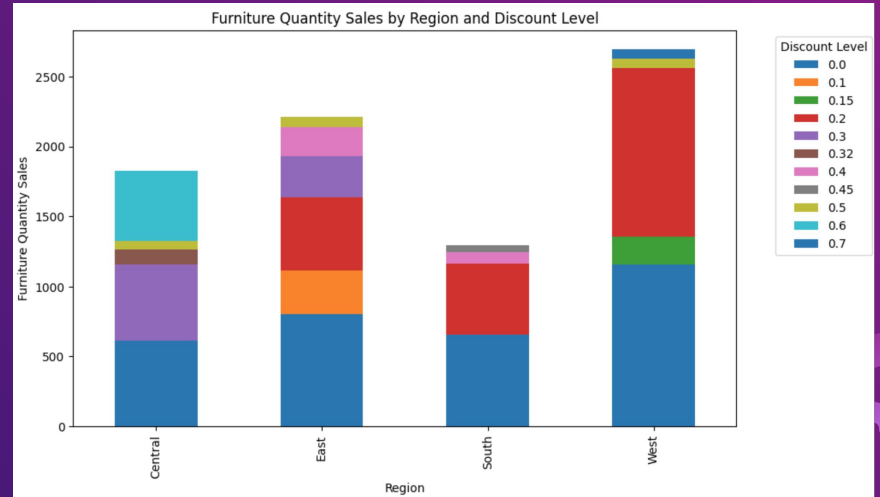




EDA - FURNITURE



- H_0 : Furniture Sales remains constant across the different discount levels
- H_A : Furniture Sales is different across the different discount levels
- **ANOVA Statistic:** 12.5183
- **P-value:** <.01

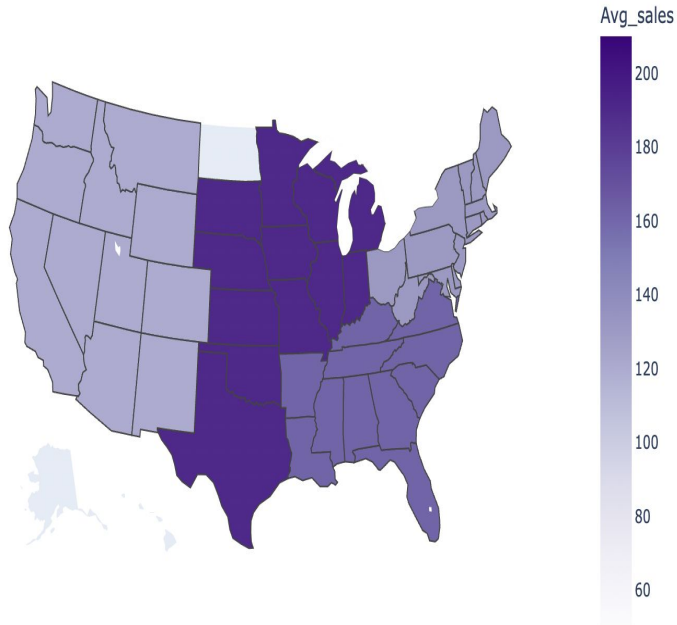


- H_0 : Furniture Sales remains constant across the different regions
- H_A : Furniture Sales is different across the different regions
- **ANOVA Statistic:** 0.1202
- **P-value:** .9482

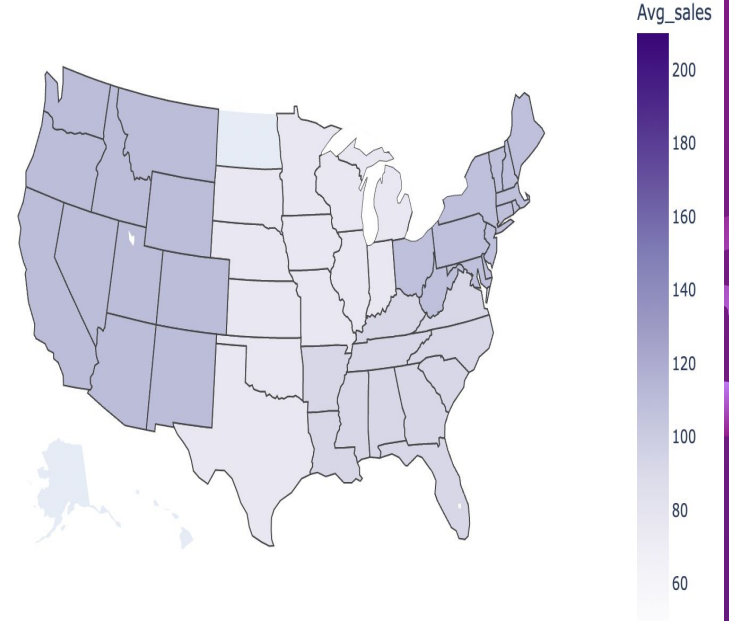


REGIONAL SALES HEAT MAP - OFFICE SUPPLY

Average Sales by Region in Office Supply Category without Discount

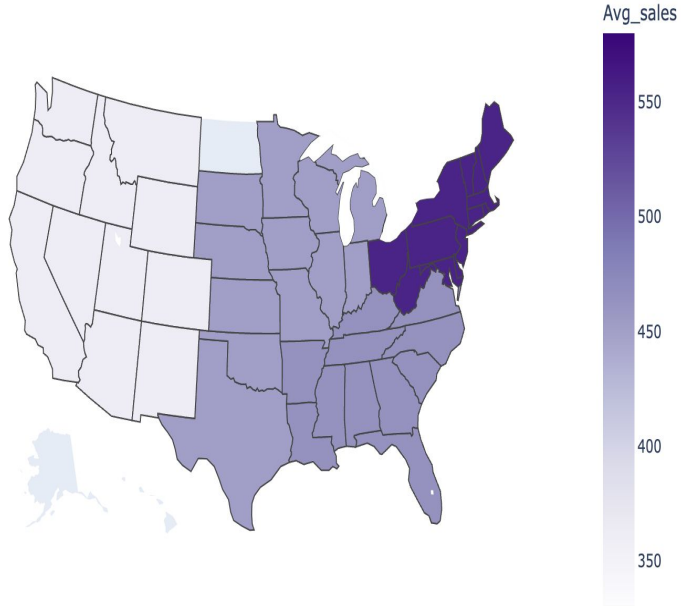


Average Sales by Region in Office Supply Category with Discount

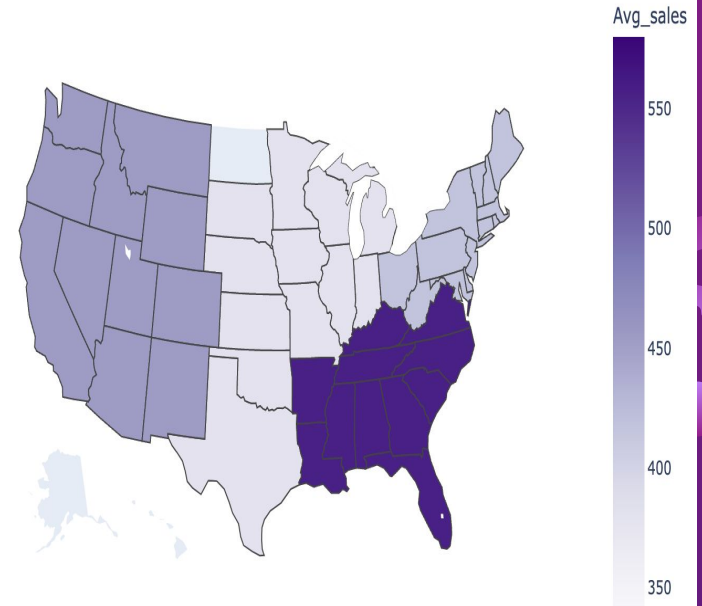


REGIONAL SALES HEAT MAP - TECHNOLOGY

Average Sales by Region in Technology Category without Discount

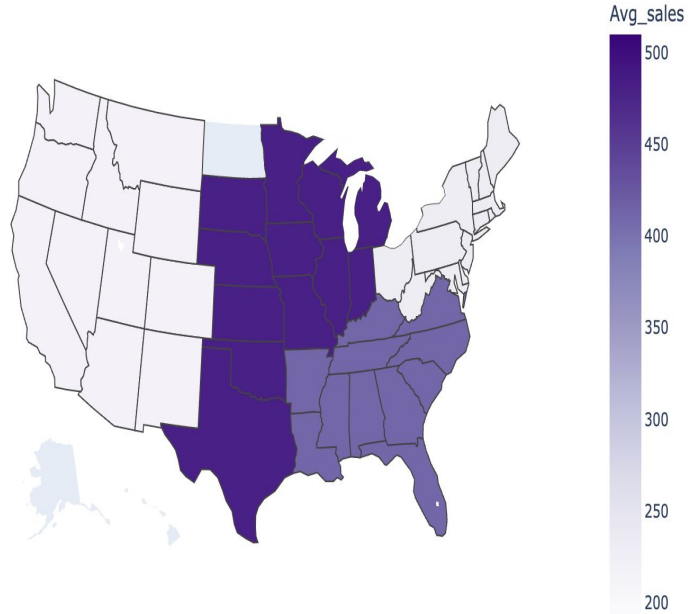


Average Sales by Region in Technology Category with Discount

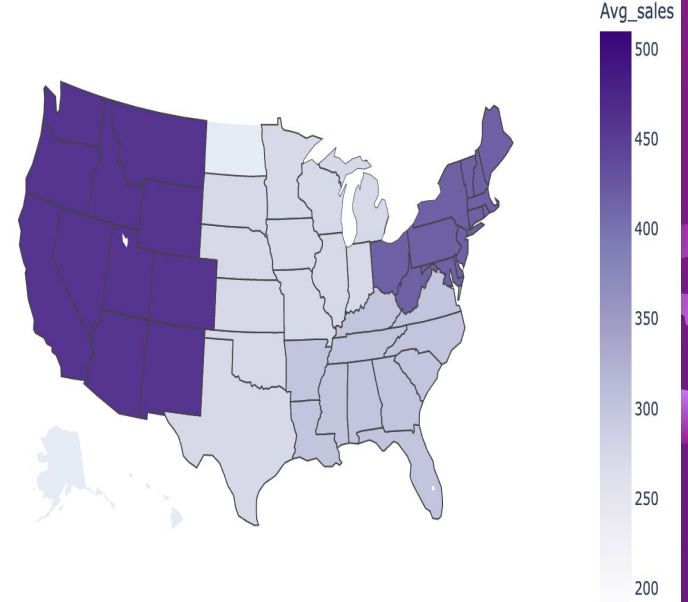


REGIONAL SALES HEAT MAP - FURNITURE

Average Sales by Region in Furniture Category without Discount



Average Sales by Region in Furniture Category with Discount



LOG-LINEAR REGRESSION ANALYSIS



Region: South, West, East, Central
Category: Furniture, Office Supplies, Technology



Baseline Region: South
Baseline Category: Furniture

CONSTANT

- Sales when discount is 0 for Furniture in the South
 $= \exp^{4.9333} = 138.84$

OLS Regression Results						
Dep. Variable:	Sales	R-squared:	0.263			
Model:	OLS	Adj. R-squared:	0.262			
Method:	Least Squares	F-statistic:	209.3			
Date:	Wed, 29 Nov 2023	Prob (F-statistic):	0.00			
Time:	04:07:26	Log-Likelihood:	-17289.			
No. Observations:	9994	AIC:	3.461e+04			
Df Residuals:	9976	BIC:	3.474e+04			
Df Model:	17					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	4.9333	0.058	85.236	0.000	4.820	5.047
Discount	0.2712	0.474	0.572	0.567	-0.658	1.201
West	-0.1351	0.053	-2.552	0.011	-0.239	-0.031
East	-0.0757	0.054	-1.400	0.161	-0.182	0.030
Central	0.1125	0.059	1.911	0.056	-0.003	0.228
Technology	0.1904	0.062	3.081	0.002	0.069	0.312
Office Supplies	-1.0873	0.048	-22.811	0.000	-1.181	-0.994
Discount_West	2.3832	0.535	4.451	0.000	1.334	3.433
Discount_East	1.4259	0.532	2.680	0.007	0.383	2.469
Discount_Central	-1.7946	0.491	-3.654	0.000	-2.757	-0.832
Discount_Technology	0.0559	0.680	0.082	0.935	-1.278	1.389
Discount_OfficeSupplies	-1.7744	0.482	-3.681	0.000	-2.719	-0.829
Discount_West_Technology	-1.0678	0.760	-1.405	0.160	-2.557	0.422
Discount_West_OfficeSupplies	-2.0818	0.554	-3.756	0.000	-3.168	-0.995
Discount_East_Technology	-1.4464	0.735	-1.968	0.049	-2.887	-0.006
Discount_East_OfficeSupplies	-1.4568	0.540	-2.698	0.007	-2.515	-0.398
Discount_Central_Technology	1.0559	0.774	1.365	0.172	-0.461	2.573
Discount_Central_OfficeSupplies	1.4038	0.496	2.831	0.005	0.432	2.376
Omnibus:	263.805	Durbin-Watson:	2.004			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	284.447			
Skew:	0.410	Prob(JB):	1.71e-62			
Kurtosis:	3.102	Cond. No.	162.			

LOG-LINEAR REGRESSION ANALYSIS

STATISTICALLY SIGNIFICANT

- West
- Technology
- Office Supplies
- Discount_West
- Discount_Central
- Discount_OfficeSupplies
- Discount_West_OfficeSupplies
- Discount_East_Technology
- Discount_East_OfficeSupplies
- Discount_Central_OfficeSupplies

OLS Regression Results						
=====						
Dep. Variable:	Sales	R-squared:	0.263			
Model:	OLS	Adj. R-squared:	0.262			
Method:	Least Squares	F-statistic:	209.3			
Date:	Wed, 29 Nov 2023	Prob (F-statistic):	0.00			
Time:	04:07:26	Log-Likelihood:	-17289.			
No. Observations:	9994	AIC:	3.461e+04			
Df Residuals:	9976	BIC:	3.474e+04			
Df Model:	17					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	4.9333	0.058	85.236	0.000	4.820	5.047
Discount	0.2712	0.474	0.572	0.567	-0.658	1.201
West	-0.1351	0.053	-2.552	0.011	-0.239	-0.031
East	-0.0757	0.054	-1.400	0.161	-0.182	0.030
Central	0.1125	0.059	1.911	0.056	-0.003	0.228
Technology	0.1904	0.062	3.081	0.002	0.069	0.312
Office Supplies	-1.0873	0.048	-22.811	0.000	-1.181	-0.994
Discount West	2.3832	0.535	4.451	0.000	1.334	3.433
Discount East	1.4259	0.532	2.680	0.007	0.383	2.469
Discount Central	-1.7946	0.491	-3.654	0.000	-2.757	-0.832
Discount Technology	0.0559	0.680	0.082	0.935	-1.278	1.389
Discount OfficeSupplies	-1.7744	0.482	-3.681	0.000	-2.719	-0.829
Discount West Technology	-1.0678	0.760	-1.405	0.160	-2.557	0.422
Discount West OfficeSupplies	-2.0818	0.554	-3.756	0.000	-3.168	-0.995
Discount East Technology	-1.4464	0.735	-1.968	0.049	-2.887	-0.006
Discount East OfficeSupplies	-1.4568	0.540	-2.698	0.007	-2.515	-0.398
Discount Central Technology	1.0559	0.774	1.365	0.172	-0.461	2.573
Discount Central OfficeSupplies	1.4038	0.496	2.831	0.005	0.432	2.376
=====						
Omnibus:	263.805	Durbin-Watson:	2.004			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	284.447			
Skew:	0.410	Prob(JB):	1.71e-62			
Kurtosis:	3.102	Cond. No.	162.			
=====						

LOG-LINEAR REGRESSION ANALYSIS

WEST

- Holding other variables constant, the expected sales in the West region are approximately 12.64% less than in the baseline 'South' region.

TECHNOLOGY & OFFICE SUPPLIES

- Technology: The coefficient 0.1904 suggests sales in the Technology category are approximately 20.97% ($\exp(0.1904)-1$) higher than in Furniture.
- Office Supplies: The coefficient -1.0873 suggests sales in Office Supplies are approximately 66.29% ($\exp(-1.0873)-1$) lower than in Furniture.

OLS Regression Results						
=====						
Dep. Variable:	Sales	R-squared:	0.263			
Model:	OLS	Adj. R-squared:	0.262			
Method:	Least Squares	F-statistic:	209.3			
Date:	Wed, 29 Nov 2023	Prob (F-statistic):	0.00			
Time:	04:07:26	Log-Likelihood:	-17289.			
No. Observations:	9994	AIC:	3.461e+04			
Df Residuals:	9976	BIC:	3.474e+04			
Df Model:	17					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	4.9333	0.058	85.236	0.000	4.820	5.047
Discount	0.2712	0.474	0.572	0.567	-0.658	1.201
West	-0.1351	0.053	-2.552	0.011	-0.239	-0.031
East	-0.0757	0.054	-1.400	0.161	-0.182	0.030
Central	0.1125	0.059	1.911	0.056	-0.003	0.228
Technology	0.1904	0.062	3.081	0.002	0.069	0.312
Office Supplies	-1.0873	0.048	-22.811	0.000	-1.181	-0.994
Discount West	2.3832	0.535	4.451	0.000	1.334	3.433
Discount East	1.4259	0.532	2.680	0.007	0.383	2.469
Discount Central	-1.7946	0.491	-3.654	0.000	-2.757	-0.832
Discount Technology	0.0559	0.680	0.082	0.935	-1.278	1.389
Discount OfficeSupplies	-1.7744	0.482	-3.681	0.000	-2.719	-0.829
Discount West Technology	-1.0678	0.760	-1.405	0.160	-2.557	0.422
Discount West OfficeSupplies	-2.0818	0.554	-3.756	0.000	-3.168	-0.995
Discount East Technology	-1.4464	0.735	-1.968	0.049	-2.887	-0.006
Discount East OfficeSupplies	-1.4568	0.540	-2.698	0.007	-2.515	-0.398
Discount Central Technology	1.0559	0.774	1.365	0.172	-0.461	2.573
Discount Central OfficeSupplies	1.4038	0.496	2.831	0.005	0.432	2.376
=====						
Omnibus:	263.805	Durbin-Watson:	2.004			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	284.447			
Skew:	0.410	Prob(JB):	1.71e-62			
Kurtosis:	3.102	Cond. No.	162.			
=====						

LOG-LINEAR REGRESSION ANALYSIS

INTERACTION TERM (COMPARE TO FURNITURE AND SOUTH)

- **Discount_West_Technology:** The effect of discounts on Technology sales in the West region is **not** different from the effect of discounts on the Furniture sales in the South region
- **Discount_West_OfficeSupplies:** ↓ sales by approximately 87.52% in the West region within the Office Supplies category.
- **Discount_East_Technology:** ↓ sales by approximately 76.46% in the East region within the Technology category.
- **Discount_East_OfficeSupplies:** ↓ sales by approximately 76.7% in the East region within the Office Supplies category.
- **Discount_Central_Technology:** The effect of discounts on Technology sales in the Central region is **not** different from the effect of discounts on the Furniture sales in the South region
- **Discount_Central_OfficeSupplies:** ↑ sales by approximately 307.06% in the Central region within the Office Supplies category.



KEY TAKEAWAYS

DISCOUNT IN THE WEST

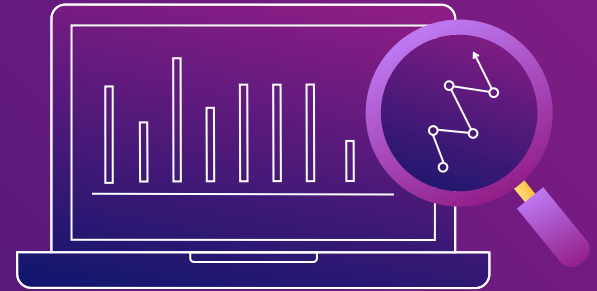
- Overall increase in sales for Technology & Furniture
- Discount in Office Supplies can reduce sales significantly

DISCOUNT IN THE EAST

- Discount on Technology & Office supplies reduce sales significantly

DISCOUNT IN THE CENTRAL

- Overall decrease in Sales
- But! Discounts are highly effective (increase sales) for Office Supplies and Technology





THANK YOU

