Adidas US Sales

Evolution: Key Metrics and Strategic Insights

Using SAS

Project By-Nandini Desai
Supervised By-Mr. Sun Makosso

AGENDA

01 OBJECTIVE

02 EDA

03 KEY FINDINGS

04 RECOMMENDATIONS

OBJECTIVE

Primary-
To analyze and evaluate the sales performance of Adidas in the US market, identify key trends, challenges, and growth opportunities and Provide actionable insights that can guide strategic decisions to enhance market share, customer engagement, and profitability."

Secondary-
Predicting Sales and Profit Margin

DATA METHODOLOGY

DATA INSIGHTS

- Number of Rows and Columns
- Data Types f Each Variable

UNIVARIATE ANALYSIS

- Univariate Analysis of Categorcal and Numerical Variables
- Checking Missing or duplicate values
- Checking outliers

BIVARIATE ANALYSIS

- T Test
- Anova

SEGMENTATION

- Region Wise
- Product Wise
- Retailer Wise

SALES STUDY

- Understanding the Trend
- Understanding the KPI
- Predictive Analysis

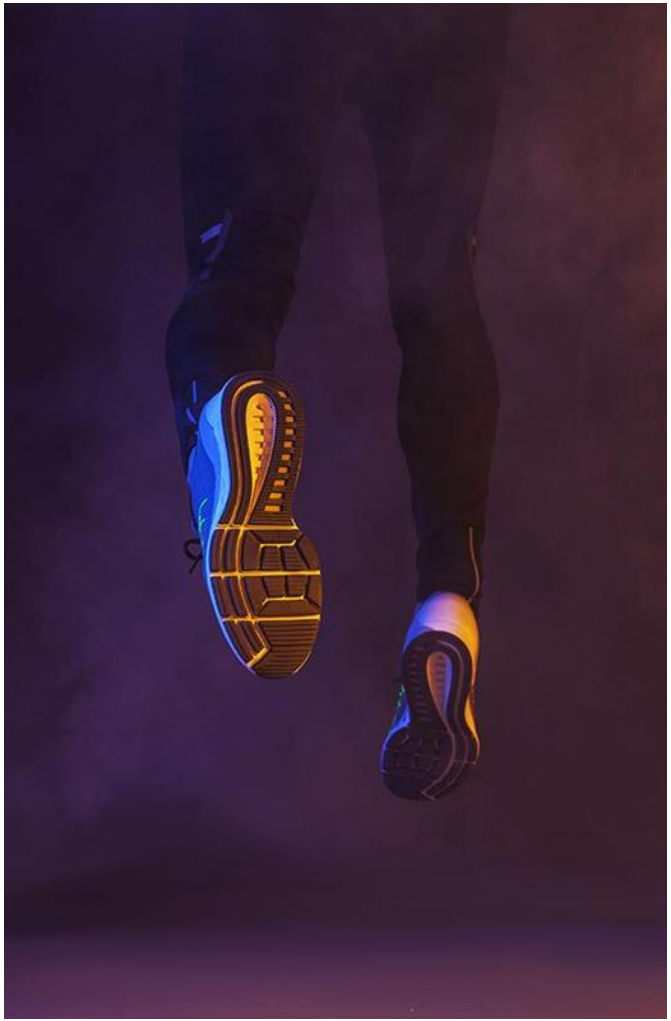
CONCLUSION

- Profit areas
- Non profit areas
- Steps to improve

01

DATA INSIGHTS

Dataset Overview and Datatypes



DATA SOURCE

Obs	Retailer	Retailer_ID	Invoice_Date	Region	State	City	Product	Price_per_Unit	Units_Sold	Total_Sales	Operating_Profit	Operating_Margin	Sales_Method
1	Foot Locker	1185732	01/01/2020	Northeast	New York	New York	Men's Street Footwear	$50.00	1,200	$600,000	$300,000	50%	In-store
2	Foot Locker	1185732	01/02/2020	Northeast	New York	New York	Men's Athletic Footwear	$50.00	1,000	$500,000	$150,000	30%	In-store
3	Foot Locker	1185732	01/03/2020	Northeast	New York	New York	Women's Street Footwear	$40.00	1,000	$400,000	$140,000	35%	In-store
4	Foot Locker	1185732	01/04/2020	Northeast	New York	New York	Women's Athletic Footwear	$45.00	850	$382,500	$133,875	35%	In-store
5	Foot Locker	1185732	01/05/2020	Northeast	New York	New York	Men's Apparel	$60.00	900	$540,000	$162,000	30%	In-store

Obs	Retailer	Retailer_ID	Invoice_Date	Region	State	City	Product	Price_per_Unit	Units_Sold	Total_Sales	Operating_Profit	Operating_Margin	Sales_Metho
9644	Foot Locker	1185732	01/24/2021	Northeast	New Hampshire	Manchester	Men's Apparel	$50.00	64	$3,200	$896	28%	Outlet
9645	Foot Locker	1185732	01/24/2021	Northeast	New Hampshire	Manchester	Women's Apparel	$41.00	105	$4,305	$1,378	32%	Outlet
9646	Foot Locker	1185732	02/22/2021	Northeast	New Hampshire	Manchester	Men's Street Footwear	$41.00	184	$7,544	$2,791	37%	Outlet
9647	Foot Locker	1185732	02/22/2021	Northeast	New Hampshire	Manchester	Men's Athletic Footwear	$42.00	70	$2,940	$1,235	42%	Outlet
9648	Foot Locker	1185732	02/22/2021	Northeast	New Hampshire	Manchester	Women's Street Footwear	$29.00	83	$2,407	$650	27%	Outlet

DATA TYPES-DUPLICATES-MISSING VALUES

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
6	City	Char	14	$14.	$14.	City
3	Invoice_Date	Num	8	MMDDYY10.		Invoice Date
12	Operating_Margin	Num	8	PERCENT12.		Operating Margin
11	Operating_Profit	Num	8	NLMNY15.		Operating Profit
8	Price_per_Unit	Num	8	NLMNY15.2		Price per Unit
7	Product	Char	25	$25.	$25.	Product
4	Region	Char	9	$9.	$9.	Region
1	Retailer	Char	13	$13.	$13.	Retailer
2	Retailer_ID	Num	8	BEST.		Retailer ID
13	Sales_Method	Char	8	$8.	$8.	Sales Method
5	State	Char	14	$14.	$14.	State
10	Total_Sales	Num	8	NLMNY15.		Total Sales
9	Units_Sold	Num	8	COMMA15.		Units Sold

Observations	9648
Variables	13
Indexes	0
Observation Length	144
Deleted Observations	0
Compressed	NO
Sorted	NO

The MEANS Procedure

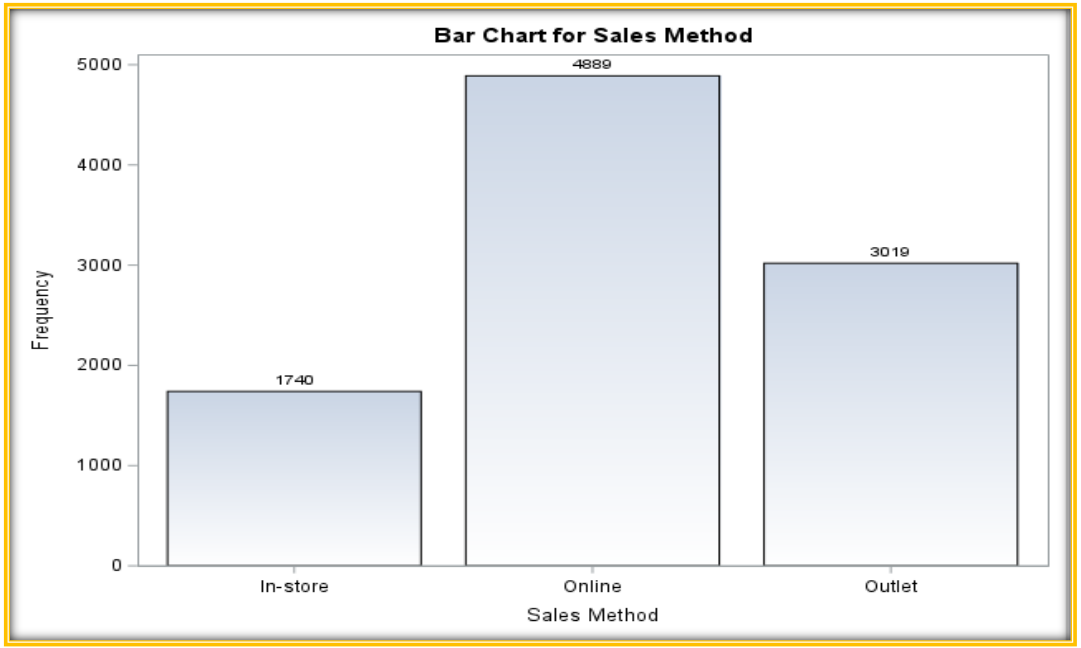
Variable	Label	N	N Miss
Retailer_ID	Retailer ID	9648	0
Invoice_Date	Invoice Date	9648	0
Price_per_Unit	Price per Unit	9648	0
Units_Sold	Units Sold	9648	0
Total_Sales	Total Sales	9648	0
Operating_Profit	Operating Profit	9648	0
Operating_Margin	Operating Margin	9648	0

The SAS System

city_n	product_n	Region_n	Retailer_n	Retailer_ID_n	salesm_n	State_n
9648	9648	9648	9648	9648	9648	9648

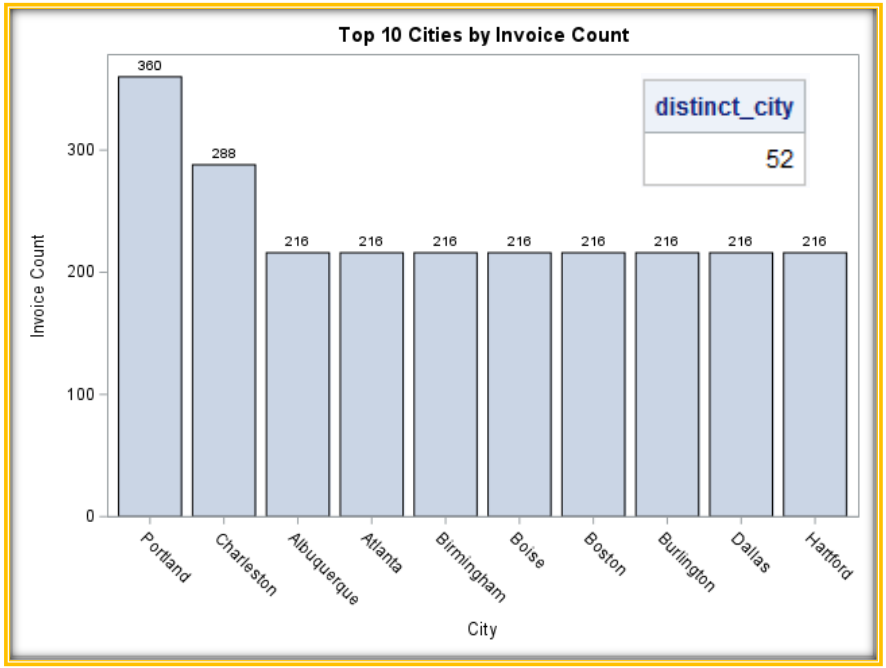
UNIVARIATE ANALYSIS -
CATEGORICAL VARIABLES

Sales Method

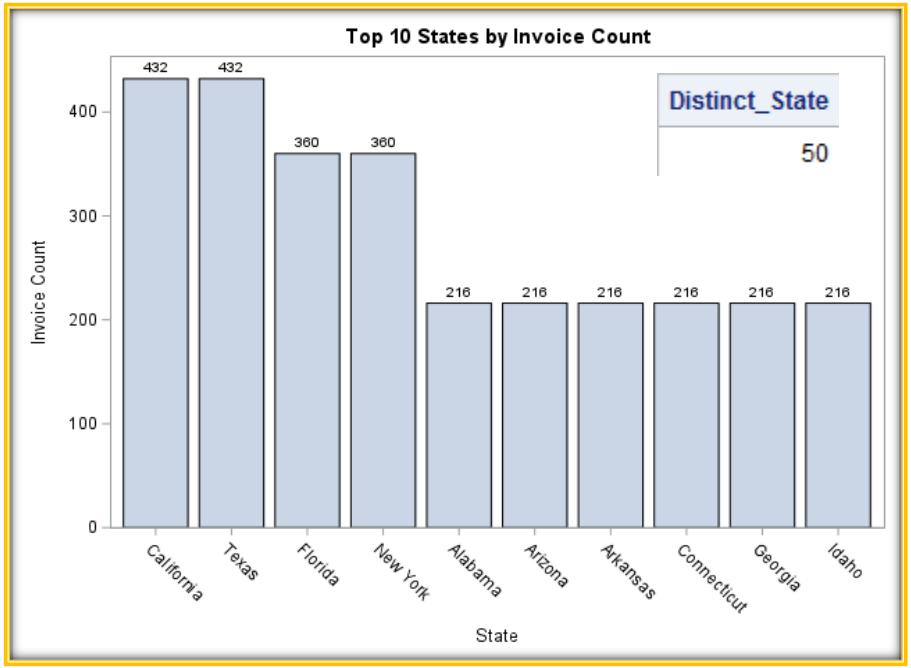


UNIVARIATE ANALYSIS -
CATEGORICAL VARIABLES

City

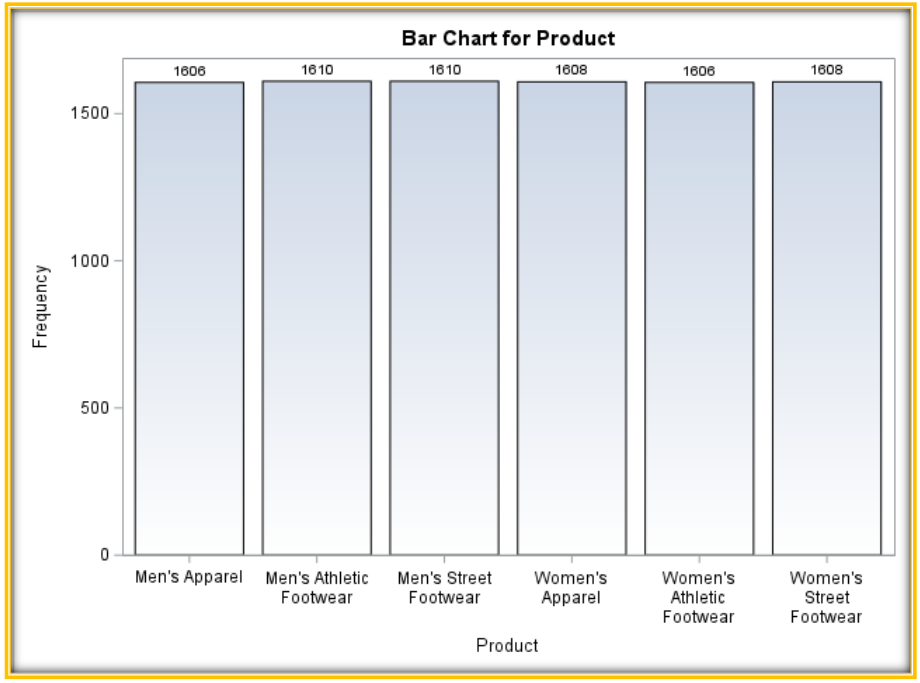


State

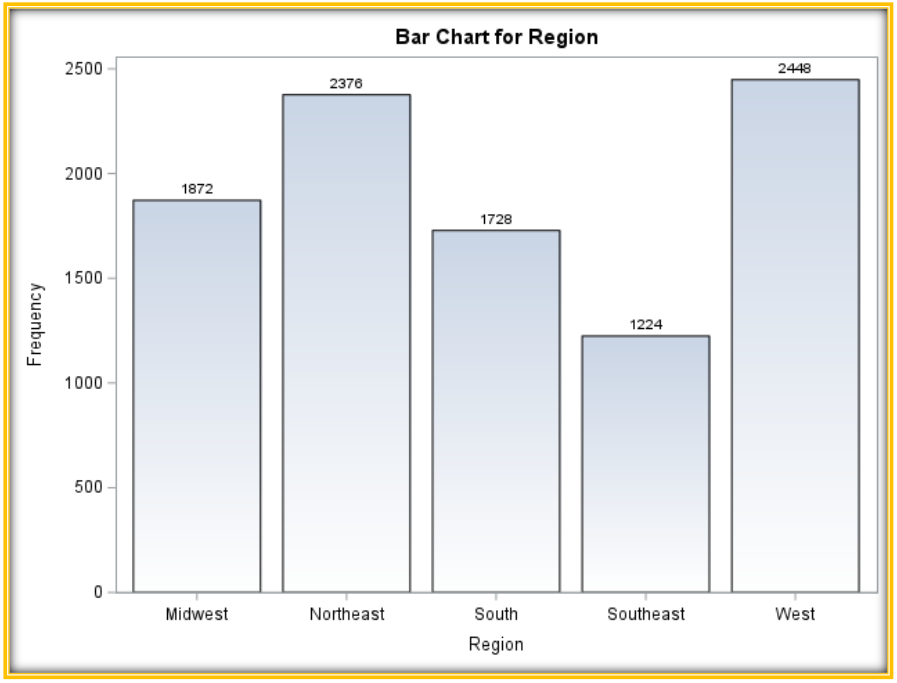


UNIVARIATE ANALYSIS -
CATEGORICAL VARIABLES

Product

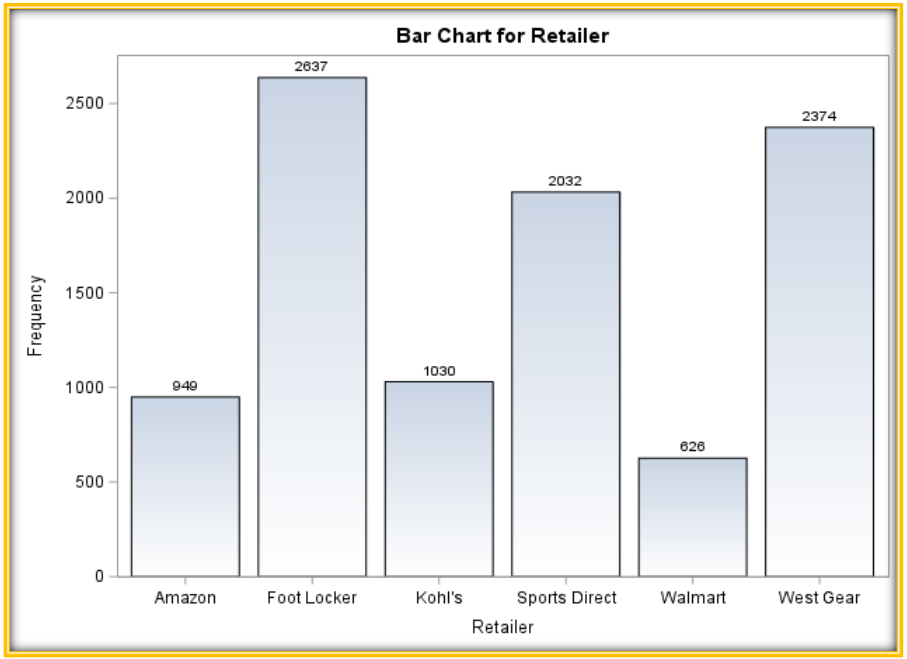


Region

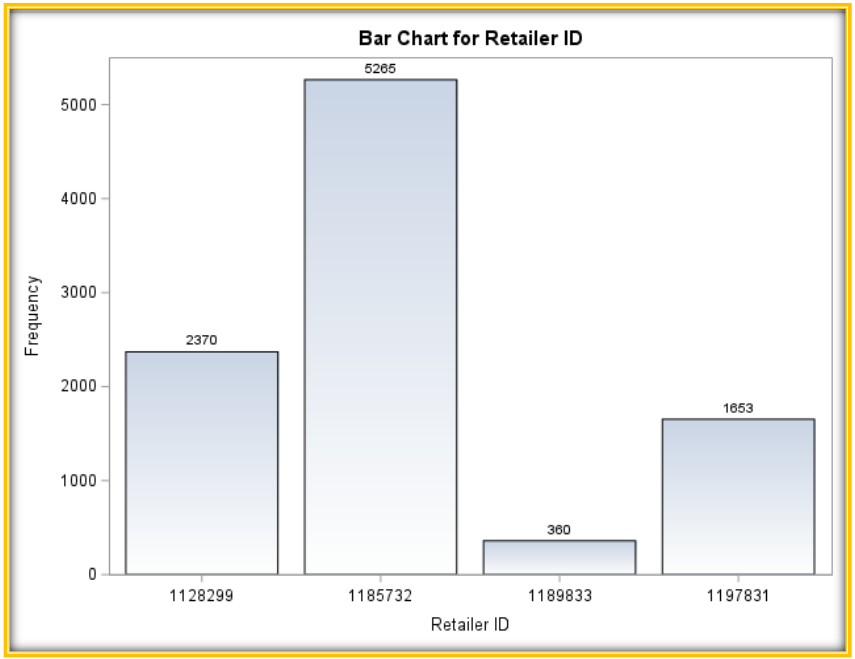


UNIVARIATE ANALYSIS -
NUMERICAL VARIABLES

Retailer



Retailer_ID



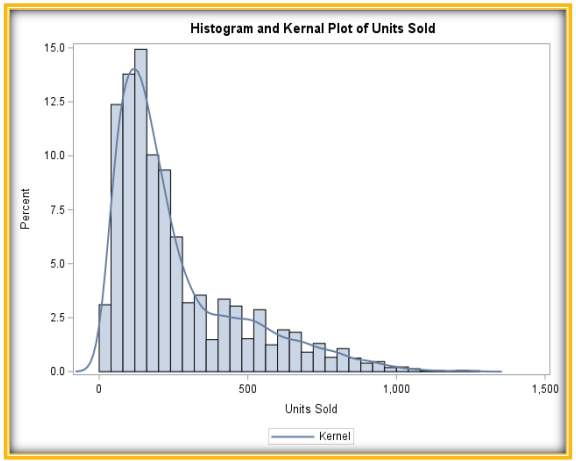
UNIVARIATE ANALYSIS -
NUMERICAL VARIABLES

Variable	Label	N	N Miss	Variance	Std Dev	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Mean	Sum	Minimum	Maximum
Retailer_ID	Retailer ID	9648	0	694869491.89	26360.38	2.25	1173323.66	1174375.78	1173849.72	11325302133	1128299.00	1197831.00
Invoice_Date	Invoice Date	9648	0	27599.40	166.13	0.74	22407.32	22413.95	22410.64	216217849.00	21915.00	22645.00
Price_per_Unit	Price per Unit	9648	0	216.25	14.71	32.52	44.92	45.51	45.22	436250.00	7.00	110.00
Units_Sold	Units Sold	9648	0	45903.93	214.25	83.39	252.65	261.21	256.93	2478861.00	0.00	1275.00
Total_Sales	Total Sales	9648	0	20140155804	141916.02	152.15	90441.29	96105.58	93273.44	899902125.00	0.00	825000.00
Operating_Profit	Operating Profit	9648	0	2936893573.9	54193.11	157.42	33343.74	35506.75	34425.24	332134761.45	0.00	390000.00
Operating_Margin	Operating Margin	9648	0	0.01	0.10	22.98	0.42	0.42	0.42	4081.02	0.10	0.80

UNIVARIATE ANALYSIS -
NUMERICAL VARIABLES

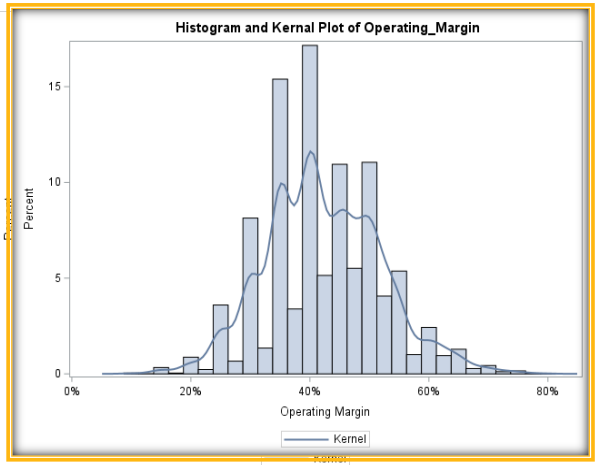
Units Sold

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.180815	Pr > D	<0.0100
Cramer-von Mises	W-Sq	95.50453	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	528.2038	Pr > A-Sq	<0.0050



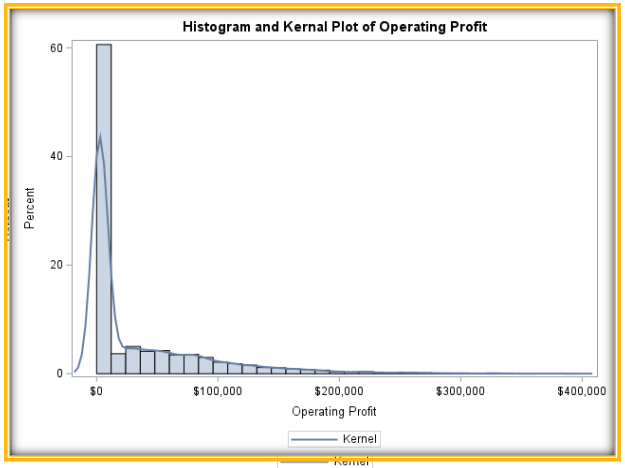
Operating Margin

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.083958	Pr > D	<0.0100
Cramer-von Mises	W-Sq	7.251366	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	40.72366	Pr > A-Sq	<0.0050



Operating Profit

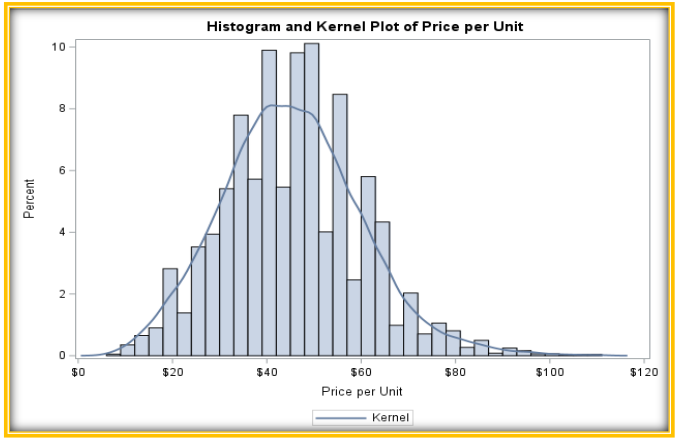
Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.297937	Pr > D	<0.0100
Cramer-von Mises	W-Sq	217.9787	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1142.555	Pr > A-Sq	<0.0050



UNIVARIATE ANALYSIS -
NUMERICAL VARIABLES

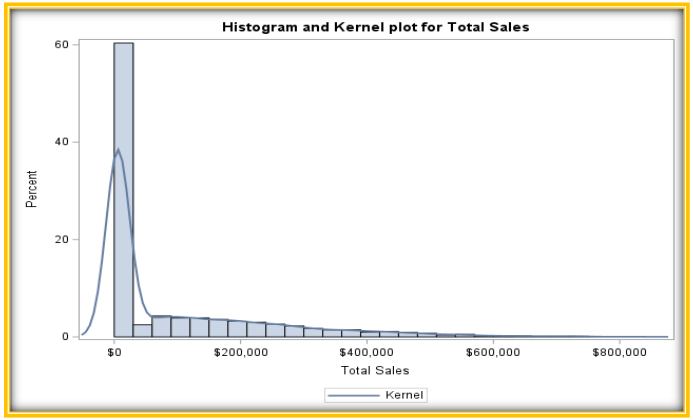
Price Per Unit

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.050967	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.590678	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	16.29257	Pr > A-Sq	<0.0050



Total Sales

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.297937	Pr > D	<0.0100
Cramer-von Mises	W-Sq	217.9787	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1142.555	Pr > A-Sq	<0.0050



UNIVARIATE ANALYSIS -
NUMERICAL VARIABLES

Highest

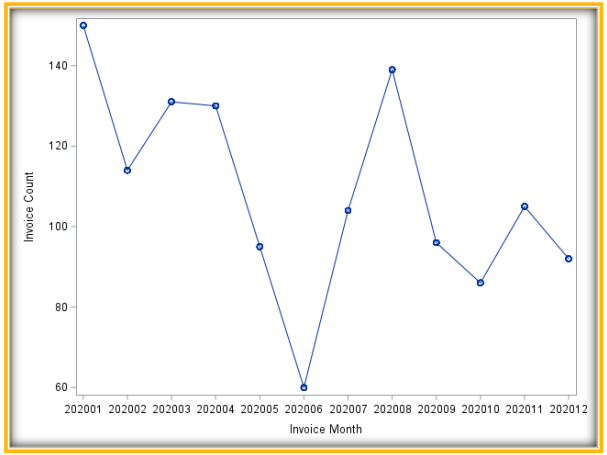
Invoice Date	invoice_count
01/17/2021	77
08/17/2021	72
03/16/2021	72
04/17/2021	72
11/17/2021	71
07/16/2021	70
12/16/2021	69
01/10/2021	67
01/24/2021	67
10/17/2021	65
02/17/2021	65
06/17/2021	64
01/23/2021	59
05/18/2021	54
09/18/2021	54
05/19/2021	54
09/19/2021	54
04/11/2021	54

Lowest

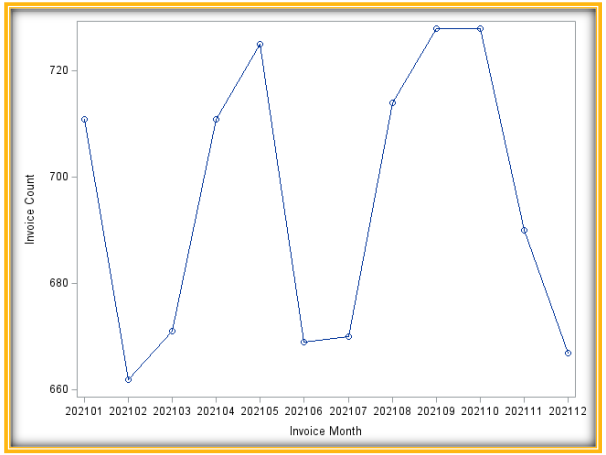
09/26/2020	2
05/14/2020	2
09/25/2020	2
09/24/2020	2
05/13/2020	2
09/23/2020	2
09/22/2020	2
05/12/2020	2
09/21/2020	2
09/20/2020	2
05/11/2020	2
09/19/2020	2
09/18/2020	2
05/10/2020	2
09/17/2020	2
05/09/2020	2
05/08/2020	2

Invoice Trend

Year 2020

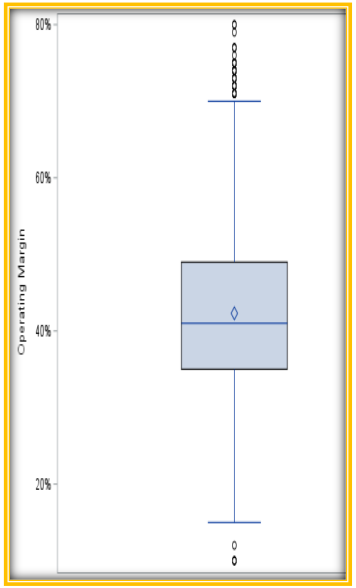


Year 2021

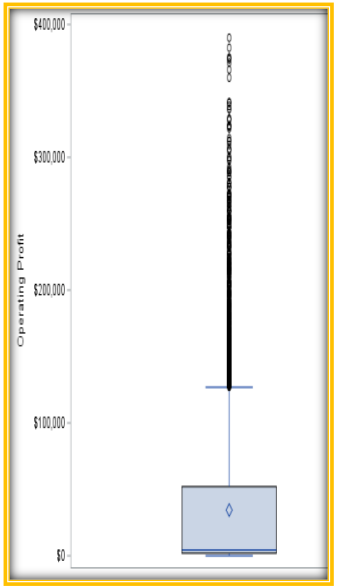


OUTLIERS DETECTION

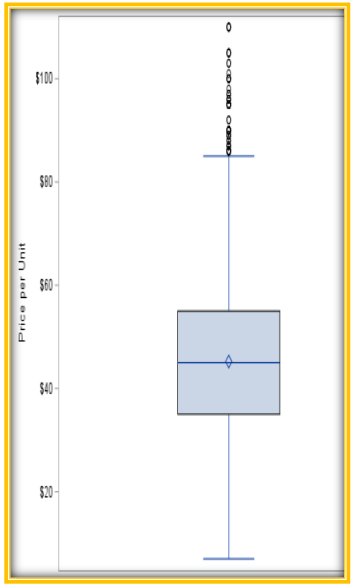
Operating Margin



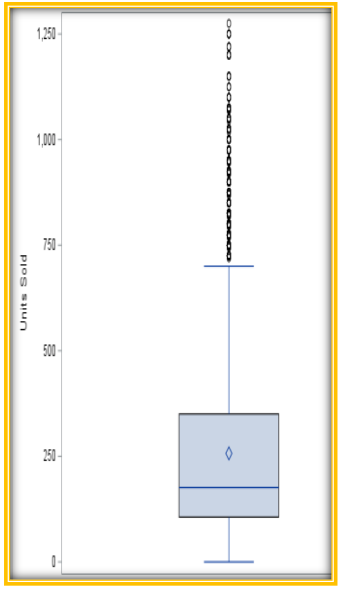
Operating Profit



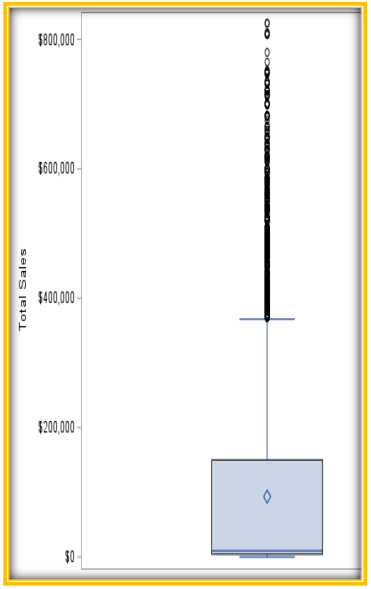
Price Per Unit



Units Sold

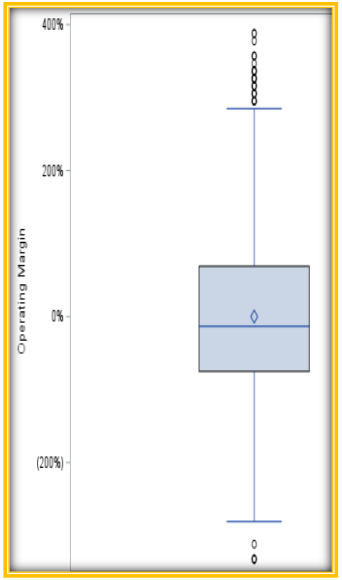


Total Sales

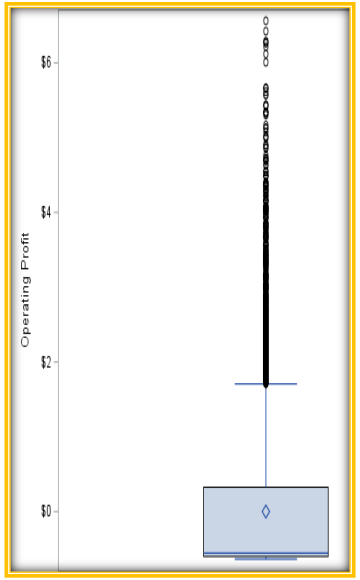


OUTLIERS DETECTION-
POST TRANSFORMATION

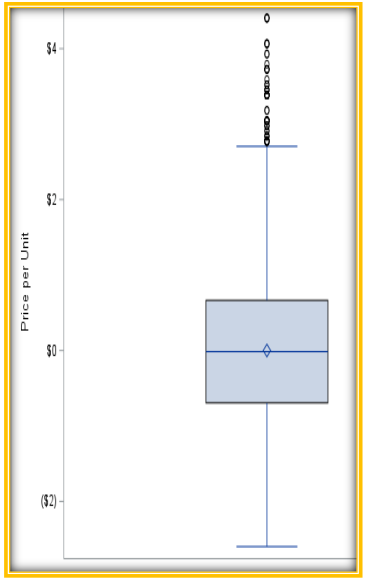
Operating Margin



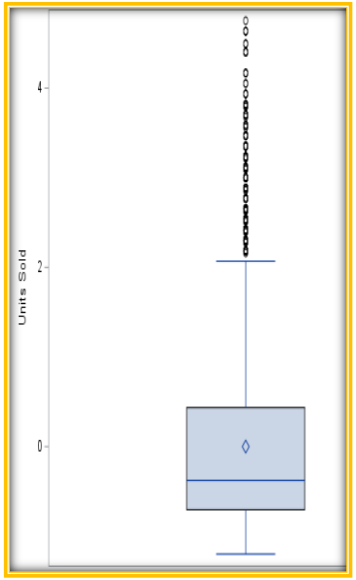
Operating Profit



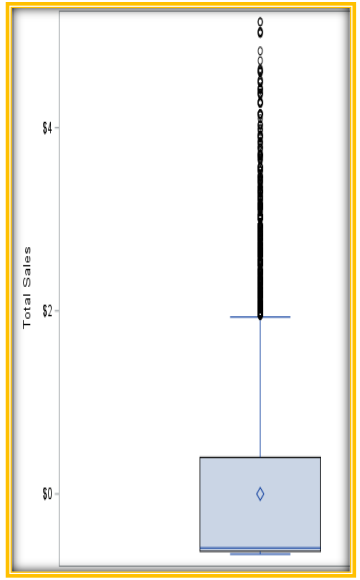
Price Per Unit



Units Sold



Total Sales



BIVARIATE ANALYSIS-
CATEGORICAL VS CATEGORICAL

Region and Product

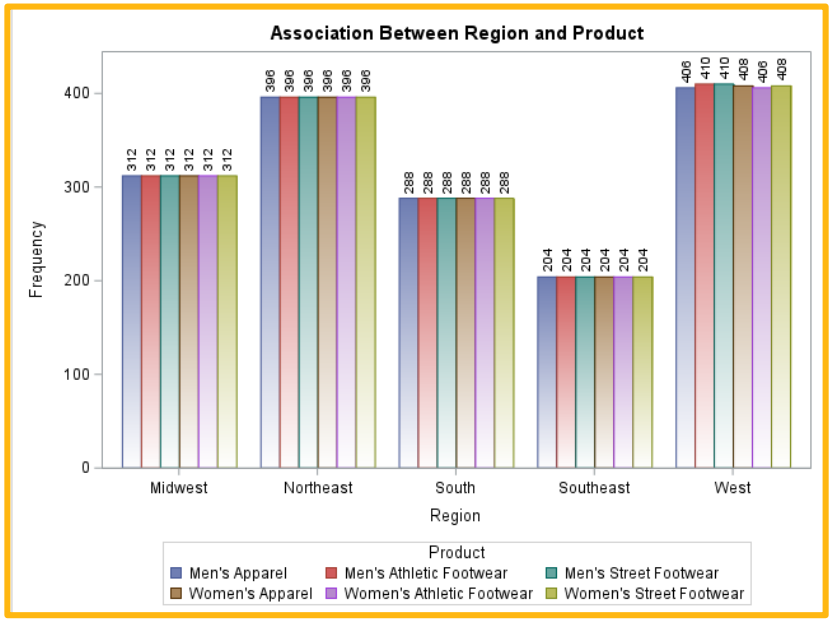


Table of Region by Product

Region(Region)	Product(Product) Men's Apparel	Men's Athletic Footwear	Men's Street Footwear	Women's Apparel	Women's Athletic Footwear	Women's Street Footwear	Total
Midwest	312 3.23	312 3.23	312 3.23	312 3.23	312 3.23	312 3.23	1872 19.40
Northeast	396 4.10	396 4.10	396 4.10	396 4.10	396 4.10	396 4.10	2376 24.63
South	288 2.99	288 2.99	288 2.99	288 2.99	288 2.99	288 2.99	1728 17.91
Southeast	204 2.11	204 2.11	204 2.11	204 2.11	204 2.11	204 2.11	1224 12.69
West	406 4.21	410 4.25	410 4.25	408 4.23	406 4.21	408 4.23	2448 25.37
Total	1606 16.65	1610 16.69	1610 16.69	1608 16.67	1606 16.65	1608 16.67	9648 100.00

Statistics for Table of Region by Product

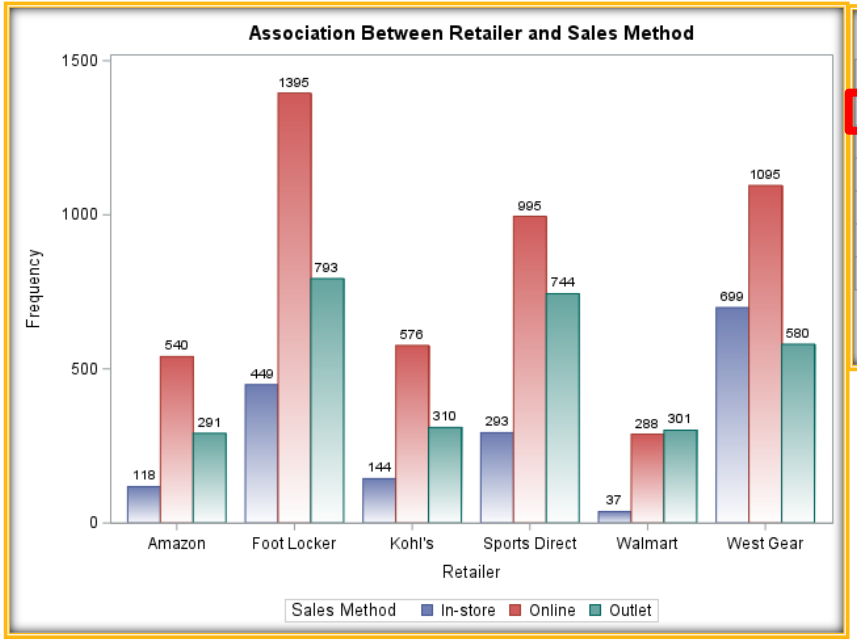
Statistic	DF	Value	Prob
Chi-Square	20	0.0293	1.0000
Likelihood Ratio Chi-Square	20	0.0293	1.0000
Mantel-Haenszel Chi-Square	1	0.0003	0.9871
Phi Coefficient		0.0017	
Contingency Coefficient		0.0017	
Cramer's V		0.0009	

Sample Size = 9648

- P value of >0.05- We fail to reject null hypothesis of Independency
- Region and Products are not Statistically significantly associated with each other.
- Region and Product are Independent of each other

BIVARIATE ANALYSIS-
CATEGORICAL VS CATEGORICAL

Retailer and Sales Method



Statistics for Table of Retailer by Sales_Method

Statistic	DF	Value	Prob
Chi-Square	10	405.8288	<.0001
Likelihood Ratio Chi-Square	10	394.1055	<.0001
Mantel-Haenszel Chi-Square	1	52.3423	<.0001
Phi Coefficient		0.2051	
Contingency Coefficient		0.2009	
Cramer's V		0.1450	

Sample Size = 9648

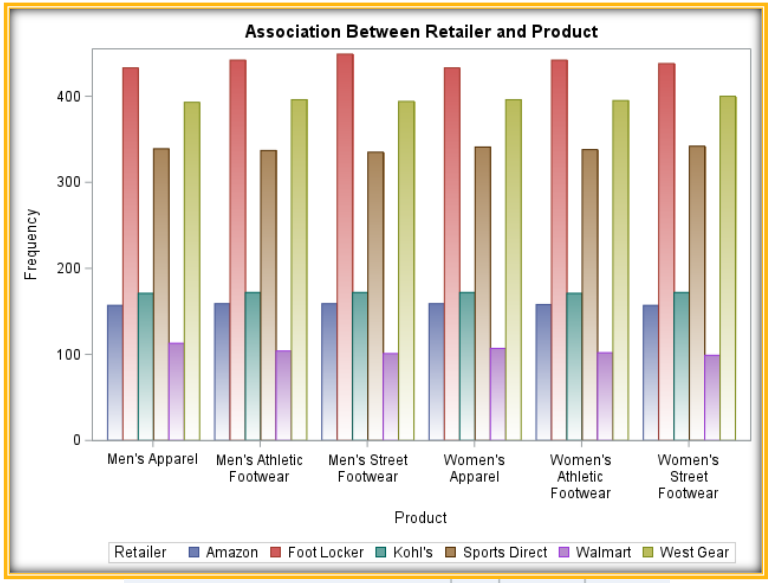
P value of <0.05- We reject null hypothesis of Independency Retailer and Sales Method are Statistically significantly associated with each other. Footlocker has highest share in Online Sales where as Walmart has Lowest share in In store Sales

Table of Retailer by Sales_Method

Retailer(Retailer)	Sales_Method(Sales Method) In-store	Online	Outlet	Total
Amazon	118 1.22	540 5.60	291 3.02	949 9.84
Foot Locker	449 4.65	1395 14.46	793 8.22	2637 27.33
Kohl's	144 1.49	576 5.97	310 3.21	1030 10.68
Sports Direct	293 3.04	995 10.31	744 7.71	2032 21.06
Walmart	37 0.38	288 2.99	301 3.12	626 6.49
West Gear	699 7.25	1095 11.35	580 6.01	2374 24.61
Total	1740 18.03	4889 50.67	3019 31.29	9648 100.00

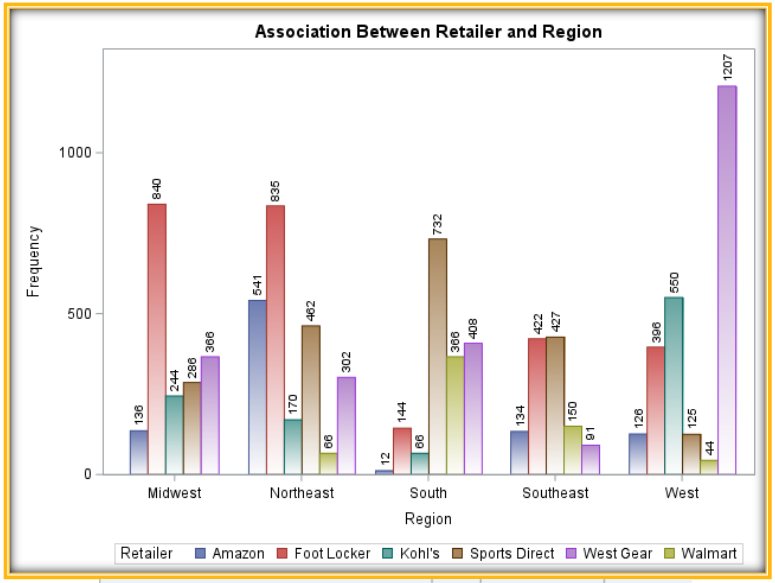
BIVARIATE ANALYSIS-
CATEGORICAL VS CATEGORICAL

Retailer and Product



Statistic	DF	Value	Prob
Chi-Square	25	1.8532	1.0000

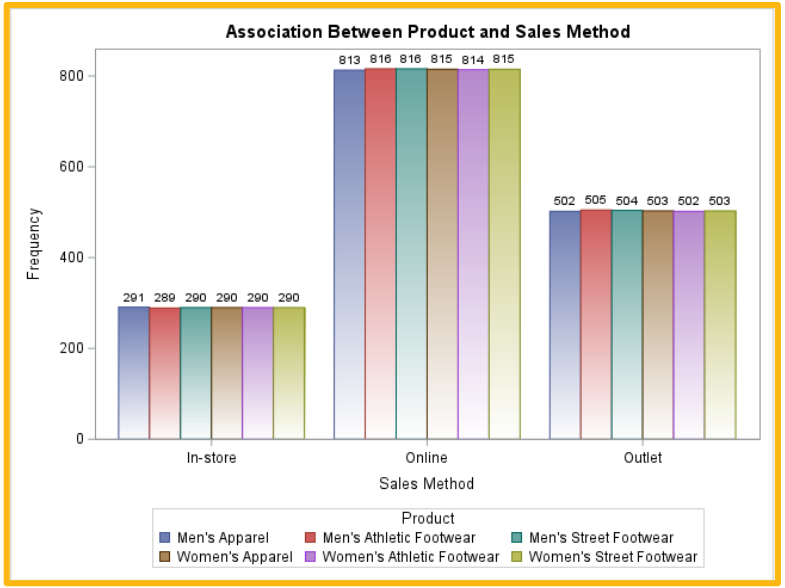
Region and Retailer



Statistic	DF	Value	Prob
Chi-Square	20	4423.6640	<.0001

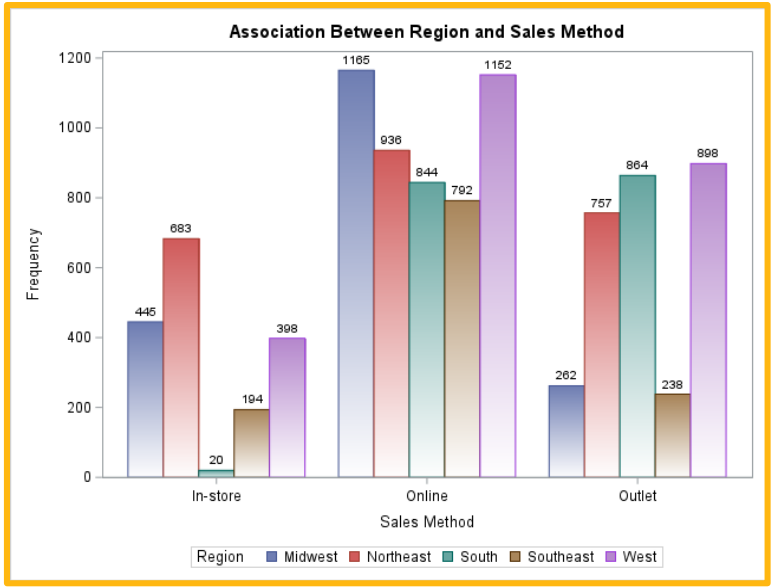
BIVARIATE ANALYSIS-
CATEGORICAL VS CATEGORICAL

Product and Sales Method



Statistic	DF	Value	Prob
Chi-Square	10	0.0189	1.0000

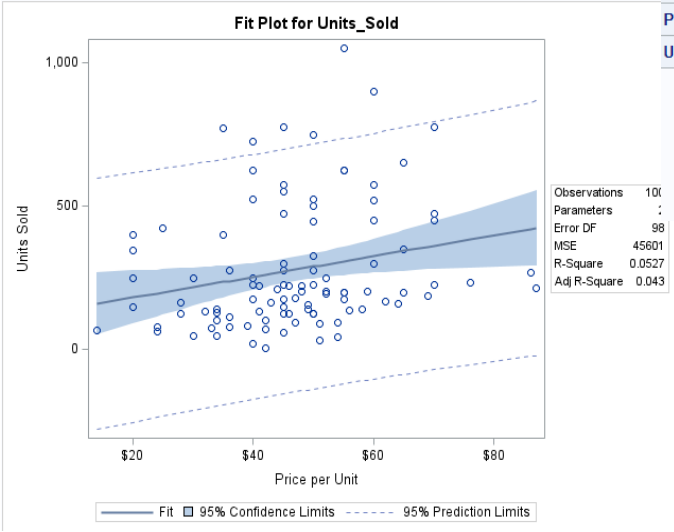
Region and Sales Method



Statistic	DF	Value	Prob
Chi-Square	8	1079.9202	<.0001

BIVARIATE ANALYSIS-
NUMERICAL VS NUMERICAL

Units sold Vs Price Per Unit



The CORR Procedure

1 With Variables:	Price_per_Unit
1 Variables:	Units_Sold

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Price_per_Unit	100	46.89000	13.90363	4689	14.00000	87.00000	Price per Unit
Units_Sold	100	278.82000	218.28909	27882	6.00000	1050	Units Sold

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	Units_Sold
Price_per_Unit Price per Unit	0.22951 0.0216

The REG Procedure
Model: MODEL1
Dependent Variable: Units_Sold Units Sold

Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	248487	248487	5.45	0.0216
Error	98	4468876	45601		
Corrected Total	99	4717363			

Root MSE	213.54338	R-Square	0.0527
Dependent Mean	278.82000	Adj R-Sq	0.0430
Coeff Var	76.58826		

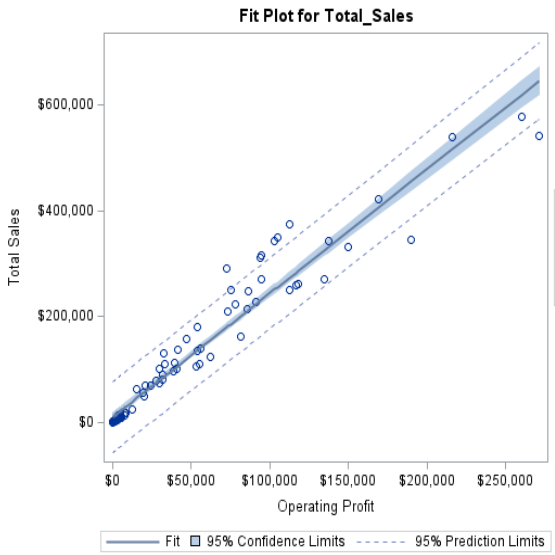
R square is very low and RMSE is Very High
Very week correlation found between Units Sold and Price Per Unit
Price per unit explains only 5% variation in Units sold

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	109.85925	75.46468	1.46	0.1487
Price_per_Unit	Price per Unit	1	3.60334	1.54362	2.33	0.0216

BIVARIATE ANALYSIS-
NUMERICAL VS NUMERICAL

Operating Profit Vs Total Sales



Observation	
Parameters	
Error DF	
MSE	
R-Square	0.9426
Adj R-Square	0.942

The CORR Procedure

1 With Variables:	Operating_Profit
1 Variables:	Total_Sales

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Operating_Profit	100	41448	57726	4144825	120.96000	271250	Operating Profit
Total_Sales	100	106637	139544	10663672	252.00000	577500	Total Sales

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	Total_Sales
Operating_Profit Operating Profit	0.97088 <.0001

- Very Strong correlation found between Operating Profit and Total Sales
- Operating Profit explains only 94% variation in Total Sales
- Correlation has relatively high error 31.5%

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales Total Sales

Number of Observations Read	100
Number of Observations Used	100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.817123E12	1.817123E12	1609.36	<.0001
Error	98	1.106516E11	1129097868		
Corrected Total	99	1.927775E12			

Root MSE	33602	R-Square	0.9426
Dependent Mean	106637	Adj R-Sq	0.9420
Coeff Var	31.51077		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	9360.11046	4143.76628	2.26	0.0261
Operating_Profit	Operating Profit	1	2.34694	0.05850	40.12	<.0001

BIVARIATE ANALYSIS-
NUMERICAL VS NUMERICAL-CORRELATION

Pearson Correlation Coefficients, N = 9648 Prob > |r| under H0: Rho=0					
	Total_Sales	Price_per_Unit	Units_Sold	Operating_Profit	Operating_Margin
Total_Sales Total Sales	1.00000	0.43581 <.0001	0.91343 <.0001	0.95631 <.0001	-0.36459 <.0001
Price_per_Unit Price per Unit	0.43581 <.0001	1.00000	0.26587 <.0001	0.39455 <.0001	-0.13749 <.0001
Units_Sold Units Sold	0.91343 <.0001	0.26587 <.0001	1.00000	0.89238 <.0001	-0.30548 <.0001
Operating_Profit Operating Profit	0.95631 <.0001	0.39455 <.0001	0.89238 <.0001	1.00000	-0.21192 <.0001
Operating_Margin Operating Margin	-0.36459 <.0001	-0.13749 <.0001	-0.30548 <.0001	-0.21192 <.0001	1.00000

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL- ANOVA

Operating Profit Vs Sales Method

The UNIVARIATE Procedure
Variable: Operating_Profit (Operating Profit)
Sales_Method = In-store

Moments			
N	1740	Sum Weights	1740
Mean	73328.3261	Sum Observations	127591288
Std Deviation	56537.0185	Variance	3196434463
Skewness	1.66358447	Kurtosis	3.28715061
Uncorrected SS	1.49147E13	Corrected SS	5.5586E12
Coeff Variation	77.1011988	Std Error Mean	1355.37128

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.85262	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.427067	Pr > D	<0.0100
Cramer-von Mises	W-Sq	11.20894	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	67.41568	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Operating_Profit (Operating Profit)
Sales_Method = Online

Moments			
N	4889	Sum Weights	4889
Mean	19749.4736	Sum Observations	96555176.5
Std Deviation	43353.5166	Variance	1879527398
Skewness	3.37554984	Kurtosis	13.459984
Uncorrected SS	1.1094E13	Corrected SS	9.18713E12
Coeff Variation	219.517327	Std Error Mean	620.032297

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.385156	Pr > D	<0.0100
Cramer-von Mises	W-Sq	207.5168	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	995.5268	Pr > A-Sq	<0.0050

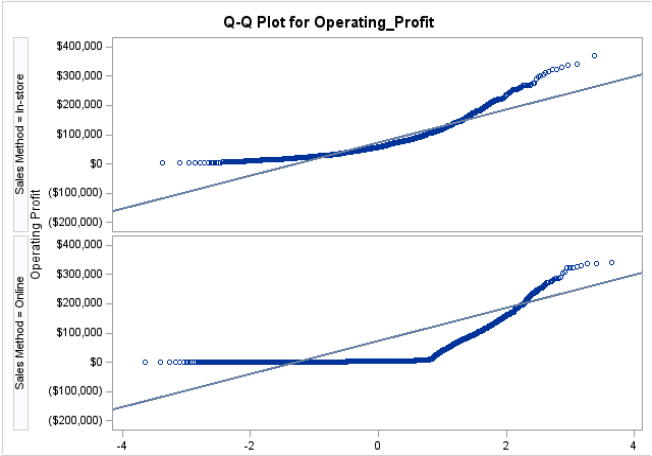
The UNIVARIATE Procedure
Variable: Operating_Profit (Operating Profit)
Sales_Method = Outlet

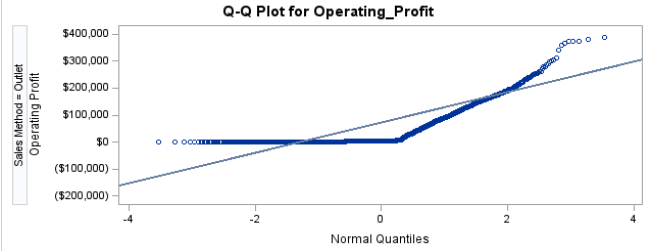
Moments			
N	3019	Sum Weights	3019
Mean	35769.5586	Sum Observations	107988297
Std Deviation	57258.555	Variance	3278542121
Skewness	2.1044602	Kurtosis	5.11384683
Uncorrected SS	1.37573E13	Corrected SS	9.89464E12
Coeff Variation	160.076213	Std Error Mean	1042.09864

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.307708	Pr > D	<0.0100
Cramer-von Mises	W-Sq	75.81803	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	388.1119	Pr > A-Sq	<0.0050

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Operating Profit Vs Sales Method





The GLM Procedure

Levene's Test for Homogeneity of Operating_Profit Variance
ANOVA of Absolute Deviations from Group Means

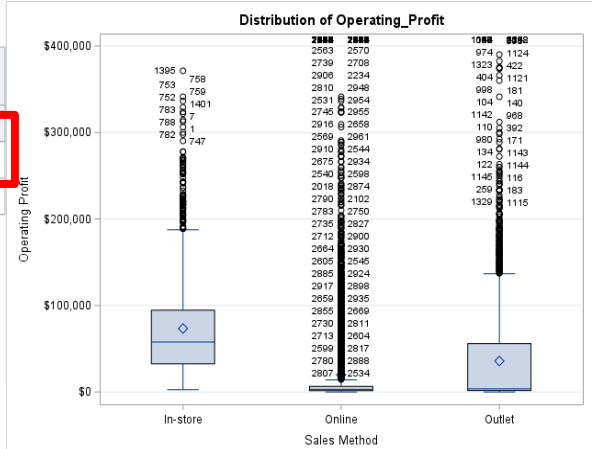
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Sales_Method	2	6E11	3E11	234.36	<.0001
Error	9645	1.235E13	1.2801E9		

Welch's ANOVA for Operating_Profit

Source	DF	F Value	Pr > F
Sales_Method	2.0000	662.23	<.0001
Error	4063.7		

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
2106.0121	2	<.0001



- All three methods have outliers.
- Operating Profit is not normally Distributed in any category of Sales Method
- But since the quantity of each group being more than 30 as per
- CLT we Can skip the normality assumption
- P value for Homoscedascity is <0.05, so we reject null hypothesis
- Operating Profit does not have equal variance in any sales method

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Operating Profit Vs Sales Method

The ANOVA Procedure

Dependent Variable: Operating_Profit Operating Profit

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.6918427E12	1.8459214E12	722.55	<.0001
Error	9645	2.464037E13	2554729867.7		
Corrected Total	9647	2.8332212E13			

R-Square	Coeff Var	Root MSE	Operating_Profit Mean
0.130305	146.8235	50544.34	34425.24

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Sales_Method	2	3.6918427E12	1.8459214E12	722.55	<.0001

Comparisons significant at the 0.05 level are indicated by ***.

Sales_Method Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
In-store - Outlet	37559	33834	41283	***
In-store - Online	53579	50125	57033	***
Outlet - In-store	-37559	-41283	-33834	***
Outlet - Online	16020	13156	18884	***
Online - In-store	-53579	-57033	-50125	***
Online - Outlet	-16020	-18884	-13156	***

- P Values is < 0.05. We reject Null Hypothesis
- Means of Operating profit are not equal in all categories of Sales Method
- Operating Profit is not equal in all Sales Methods

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL- ANOVA

Total Sales Vs Retailer

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = Amazon

Moments

N	949	Sum Weights	949
Mean	81874.5121	Sum Observations	77698912
Std Deviation	113390.215	Variance	1.28573E10
Skewness	1.6031961	Kurtosis	2.35189627
Uncorrected SS	1.85503E13	Corrected SS	1.21888E13
Coeff Variation	138.49269	Std Error Mean	3680.80212

Tests for Normality

Test		Statistic		p Value
Shapiro-Wilk	W	0.741218	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.298564	Pr > D	<0.0100
Cramer-von Mises	W-Sq	17.22966	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	93.72886	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = Foot Locker

Moments

N	2637	Sum Weights	2637
Mean	83464.0576	Sum Observations	220094720
Std Deviation	134053.6	Variance	1.79704E10
Skewness	2.20666938	Kurtosis	5.06041026
Uncorrected SS	6.57399E13	Corrected SS	4.73699E13
Coeff Variation	160.612369	Std Error Mean	2610.49832

Tests for Normality

Test		Statistic		p Value
Kolmogorov-Smirnov	D	0.271146	Pr > D	<0.0100
Cramer-von Mises	W-Sq	60.45497	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	318.6931	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = Kohl's

Moments

N	1030	Sum Weights	1030
Mean	99140.5369	Sum Observations	102114753
Std Deviation	130908.219	Variance	1.7137E10
Skewness	1.13152979	Kurtosis	0.01877409
Uncorrected SS	2.77576E13	Corrected SS	1.76339E13
Coeff Variation	132.04308	Std Error Mean	4078.94924

Tests for Normality

Test		Statistic		p Value
Shapiro-Wilk	W	0.748712	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.326125	Pr > D	<0.0100
Cramer-von Mises	W-Sq	21.14662	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	113.0676	Pr > A-Sq	<0.0050

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL- ANOVA

Total Sales Vs Retailer

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = Sports Direct

Moments

N	2032	Sum Weights	2032
Mean	89798.719	Sum Observations	182470997
Std Deviation	133219.357	Variance	1.77474E10
Skewness	1.62152321	Kurtosis	2.01431089
Uncorrected SS	5.24306E13	Corrected SS	3.6045E13
Coeff Variation	148.353294	Std Error Mean	2955.3266

Tests for Normality

Test		Statistic		p Value
Kolmogorov-Smirnov	D	0.328439	Pr > D	<0.0100
Cramer-von Mises	W-Sq	47.00504	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	246.151	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = Walmart

Moments

N	626	Sum Weights	626
Mean	119102.891	Sum Observations	74558410
Std Deviation	185747.169	Variance	3.4502E10
Skewness	1.64419916	Kurtosis	1.85037324
Uncorrected SS	3.04439E13	Corrected SS	2.15638E13
Coeff Variation	155.955214	Std Error Mean	7423.94998

Tests for Normality

Test		Statistic		p Value
Shapiro-Wilk	W	0.678251	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.354561	Pr > D	<0.0100
Cramer-von Mises	W-Sq	17.18567	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	87.48597	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Total_Sales (Total Sales)
Retailer = West Gear

Moments

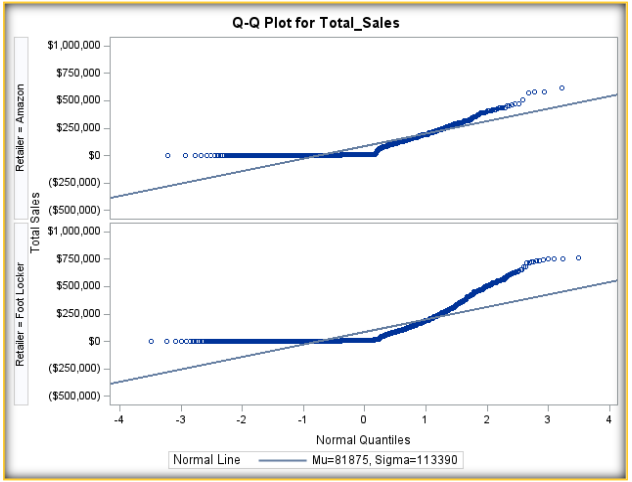
N	2374	Sum Weights	2374
Mean	102343.864	Sum Observations	242964333
Std Deviation	156931.144	Variance	2.46274E10
Skewness	1.76837241	Kurtosis	2.51529809
Uncorrected SS	8.33067E13	Corrected SS	5.84408E13
Coeff Variation	153.33713	Std Error Mean	3220.83727

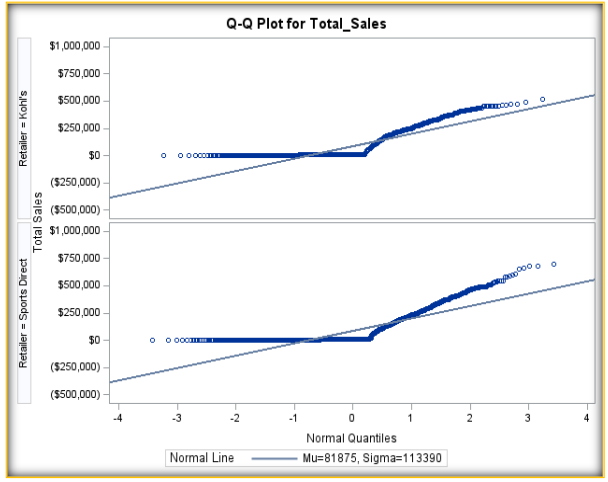
Tests for Normality

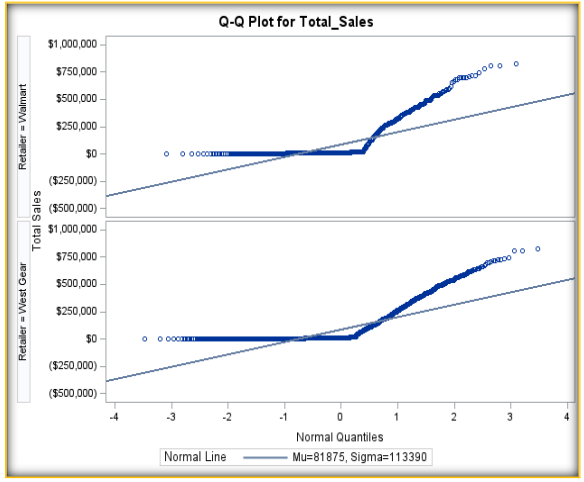
Test		Statistic		p Value
Kolmogorov-Smirnov	D	0.30144	Pr > D	<0.0100
Cramer-von Mises	W-Sq	56.17112	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	293.8409	Pr > A-Sq	<0.0050

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Total Sales Vs Retailer







- Total Sales is not normally Distributed in any Retailer Category
- But since the quantity of each group being more than 30 as per CLT we can skip the normality assumption

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Total Sales Vs Retailer

Levene's Test for Homogeneity of Total_Sales Variance
ANOVA of Absolute Deviations from Group Means

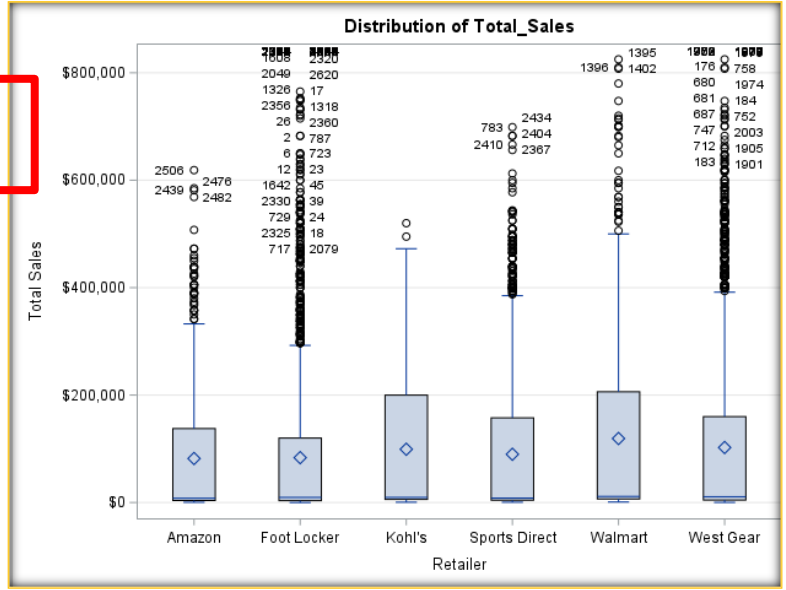
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Retailer	5	2.066E12	4.132E11	52.78	<.0001
Error	9642	7.549E13	7.8289E9		

Welch's ANOVA for Total_Sales

Source	DF	F Value	Pr > F
Retailer	5.0000	9.05	<.0001
Error	3085.5		

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
87.6044	5	<.0001



- P value for Homoscedascity is <0.05, so we reject null hypothesis
- Total Sales does not have equal variance in any Retailer Category
- All 6 groups have outliers.
- .

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Total Sales Vs Retailer

The ANOVA Procedure

Dependent Variable: Total_Sales Total Sales

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.0499984E12	209999672960	10.48	<.0001
Error	9642	1.9324208E14	20041701377		
Corrected Total	9647	1.9429208E14			

R-Square	Coeff Var	Root MSE	Total_Sales Mean
0.005404	151.7782	141568.7	93273.44

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Retailer	5	1.0499984E12	209999672960	10.48	<.0001

- P Values is < 0.05. We reject Null Hypothesis
- Means of Total Sales are not equal in all categories of Retailer
- Total Sales is not equal in all Retailers

Comparisons significant at the 0.05 level are indicated by ***.

Retailer Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Walmart - West Gear	16759	-4409	37927	
Walmart - Kohl's	19962	-3914	43839	
Walmart - Sports Direct	29304	7768	50840	***
Walmart - Foot Locker	35639	14692	56585	***
Walmart - Amazon	37228	12970	61487	***
West Gear - Walmart	-16759	-37927	4409	
West Gear - Kohl's	3203	-14375	20782	
West Gear - Sports Direct	12545	-1693	26784	
West Gear - Foot Locker	18880	5550	32209	***
West Gear - Amazon	20469	2375	38563	***
Kohl's - Walmart	-19962	-43839	3914	
Kohl's - West Gear	-3203	-20782	14375	
Kohl's - Sports Direct	9342	-8679	27362	
Kohl's - Foot Locker	15676	-1635	32988	
Kohl's - Amazon	17266	-3933	38465	

Sports Direct - Walmar	-29304	-50840	-7768	***
Sports Direct - West Gea	-12545	-26784	1693	
Sports Direct - Kohl's	-9342	-27362	8679	
Sports Direct - Foot Locke	6335	-7572	20242	
Sports Direct - Amazon	7924	-10599	26448	
Foot Locker - Walmart	-35639	-56585	-14692	***
Foot Locker - West Gea	-18880	-32209	-5550	***
Foot Locker - Kohl's	-15676	-32988	1635	
Foot Locker - Sports Direct	-6335	-20242	7572	
Foot Locker - Amazon	1590	-16245	19424	
Amazon - Walmart	-37228	-61487	-12970	***
Amazon - West Gear	-20469	-38563	-2375	***
Amazon - Kohl's	-17266	-38465	3933	
Amazon - Sports Direc	-7924	-26448	10599	
Amazon - Foot Locker	-1590	-19424	16245	

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-ANOVA

Anova Test Results among All categorical Vs. Numerical Variables

Categorical Feature	Numerical Feature
Retailer	Total Sales
	Units Sold
	Operating P
	Operating Margin
	Price Per Units
Sales Method	Operating P
	Total Sales
	Units Sold
	Operating Margin
	Price Per Units
Region	Operating P
	Total Sales
	Units Sold
	Operating Margin
	Price Per Units
Product	Operating P
	Total Sales
	Units Sold
	Operating Margin
	Price Per Units

For all combinations :

- P Values is < 0.05. We reject Null Hypothesis
- Means of Numerical Features are not equal in All respective categories of Categorical Features
- Numerical Features are not equal in all Categories of Categorical Features of this dataset

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL- TTEST

Units Sold Vs Profit Class

The UNIVARIATE Procedure
Variable: Units_Sold (Units Sold)
Profit_Class = High Profit

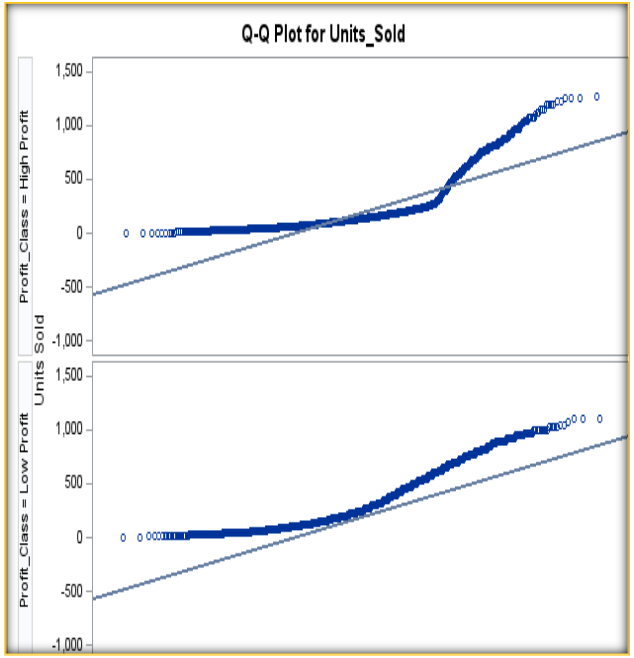
Moments			
N	4412	Sum Weights	4412
Mean	187.735494	Sum Observations	828289
Std Deviation	183.761498	Variance	33768.288
Skewness	2.66257745	Kurtosis	7.81463748
Uncorrected SS	304451163	Corrected SS	148951918
Coeff Variation	97.8831938	Std Error Mean	2.76653883

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.228128	Pr > D	<0.0100
Cramer-von Mises	W-Sq	81.10379	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	442.8506	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: Units_Sold (Units Sold)
Profit_Class = Low Profit

Moments			
N	5236	Sum Weights	5236
Mean	315.235294	Sum Observations	1650572
Std Deviation	220.687201	Variance	48702.8409
Skewness	0.93536091	Kurtosis	0.17603089
Uncorrected SS	775277922	Corrected SS	254959372
Coeff Variation	70.0071361	Std Error Mean	3.04984191

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.1254	Pr > D	<0.0100
Cramer-von Mises	W-Sq	23.05603	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	137.5915	Pr > A-Sq	<0.0050



BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-TTEST

Units Sold Vs Profit Class

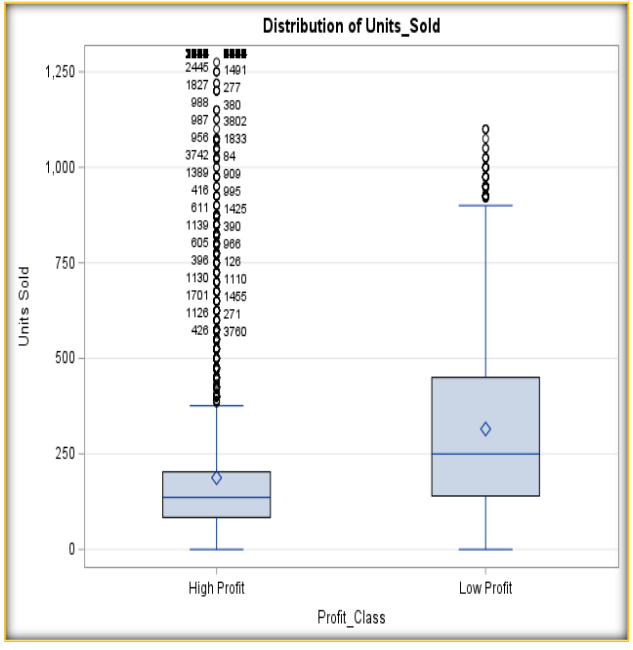
The GLM Procedure

Levene's Test for Homogeneity of Units_Sold Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Profit_Class	1	10095679	10095679	556.97	<.0001
Error	9646	1.7484E8	18126.0		

Welch's ANOVA for Units_Sold

Source	DF	F Value	Pr > F
Profit_Class	1.0000	958.77	<.0001
Error	9644.7		



- Units Sold is not normally Distributed in any Profit Class
- But since the quantity of each group being more than 30 as per CLT we can skip the normality assumption
- P value for Homoscedascity is <0.05, so we reject null hypothesis
- Units Sold does not have equal variance in any Profit Class
- Both groups of Profit Class have outliers with Majority in High Profit
-

BIVARIATE ANALYSIS-
CATEGORICAL VS NUMERICAL-TTEST

Units Sold Vs Profit Class

The TTEST Procedure

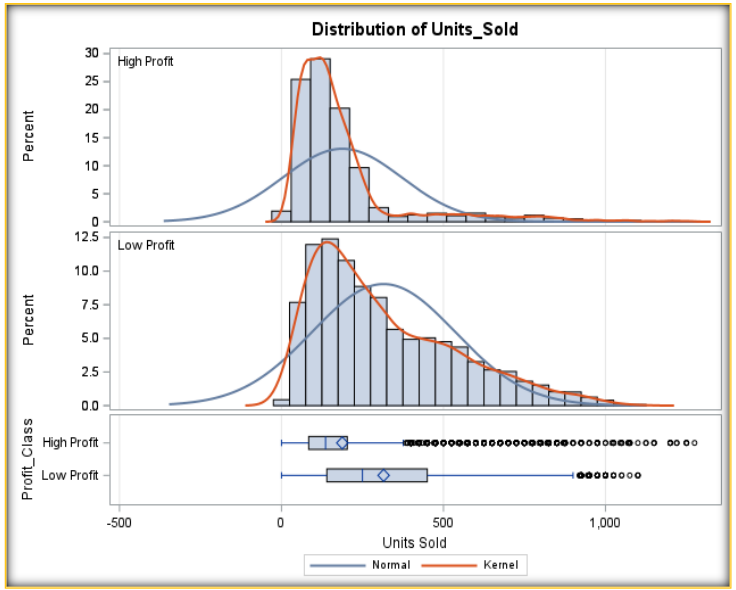
Variable: Units_Sold (Units Sold)

Profit_Class	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
High Profit		4412	187.7	183.8	2.7665	0	1275.0
Low Profit		5236	315.2	220.7	3.0498	0	1100.0
Diff (1-2)	Pooled		-127.5	204.6	4.1819		
Diff (1-2)	Satterthwaite		-127.5		4.1177		

Profit_Class	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
High Profit		187.7	182.3	193.2	183.8	180.0	187.7
Low Profit		315.2	309.3	321.2	220.7	216.5	225.0
Diff (1-2)	Pooled	-127.5	-135.7	-119.3	204.6	201.8	207.6
Diff (1-2)	Satterthwaite	-127.5	-135.6	-119.4			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	9646	-30.49	<.0001
Satterthwaite	Unequal	9644.7	-30.96	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	5235	4411	1.44	<.0001



- P value for folded F 0.0001 is <0.05- We Reject H0 for Homoscedascity
- Sathherwaite P value 0.0001<0.05- we reject H0
- Means of Units Sold is not equal for both Profit Class
- Units Sold is not equal in both profit Class

CHECKING MULTICOLLINEARITY

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	34380	2056.17575	16.72	<.0001	0
Operating_Profit	Operating Profit	1	1.74392	0.01342	129.96	<.0001	5.82818
Operating_Margin	Operating Margin	1	-194986	3312.24414	-58.87	<.0001	1.14229
Price_per_Unit	Price per Unit	1	838.00993	22.96108	36.50	<.0001	1.25649
Units_Sold	Units Sold	1	169.08846	3.32074	50.92	<.0001	5.57882

Collinearity Diagnostics (intercept adjusted)

Number	Eigenvalue	Condition Index	Proportion of Variation			
			Operating_Profit	Operating_Margin	Price_per_Unit	Units_Sold
1	2.21741	1.00000	0.03019	0.03675	0.04949	0.03063
2	0.88927	1.57909	0.00719	0.80414	0.13013	0.00009830
3	0.80228	1.66249	0.01491	0.09049	0.69698	0.03235
4	0.09104	4.93515	0.94771	0.06862	0.12340	0.93693

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales Total Sales

Number of Observations Read	9648
Number of Observations Used	9648

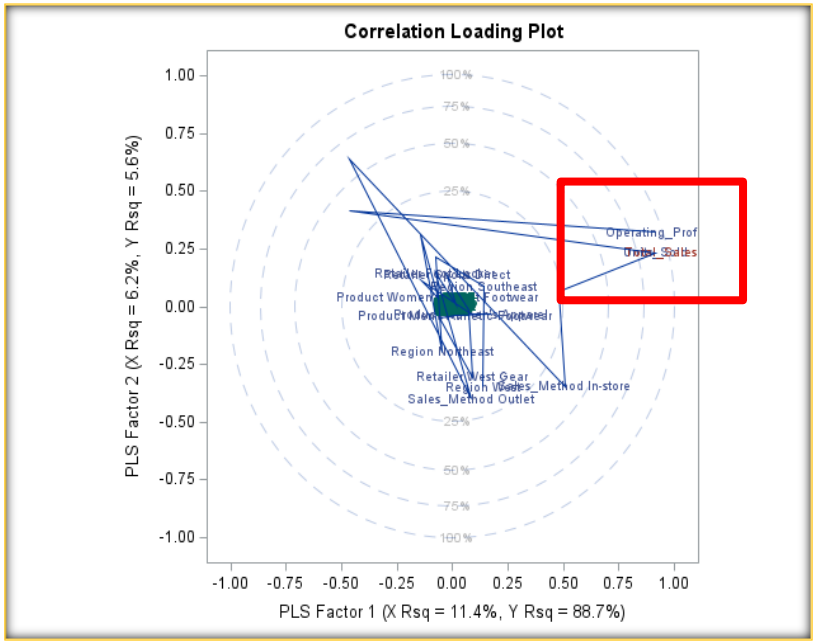
Analysis of Variance

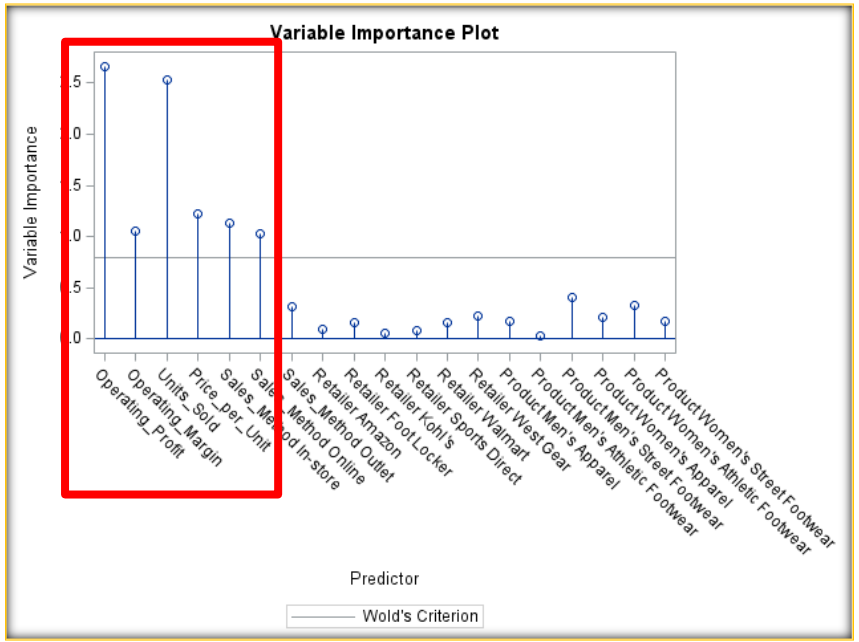
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1.858513E14	4.646283E13	53080.6	<.0001
Error	9643	8.440767E12	875325846		
Corrected Total	9647	1.942921E14			

Root MSE	29586	R-Square	0.9566
Dependent Mean	93273	Adj R-Sq	0.9565
Coeff Var	31.71954		

- VIF of Operating Profit and Units_sold is >5
- R squared is excellent using all 4 numerical predictors

CHECKING MULTICOLLINEARITY





CHECKING MULTICOLLINEARITY

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	89385	1970.66475	45.36	<.0001	0
Operating_Profit	Operating Profit	1	2.35141	0.00692	339.89	<.0001	1.22101
Operating_Margin	Operating Margin	1	-242716	3578.33741	-67.83	<.0001	1.05080
Price_per_Unit	Price per Unit	1	566.32354	25.15500	22.51	<.0001	1.18864

Collinearity Diagnostics (intercept adjusted)

Number	Eigenvalue	Condition Index	Proportion of Variation		
			Operating_Profit	Operating_Margin	Price_per_Unit
1	1.51401	1.00000	0.23152	0.11904	0.21261
2	0.88902	1.30500	0.03324	0.84068	0.16898
3	0.59697	1.59253	0.73524	0.04028	0.61841

The REG Procedure
Model: MODEL1
Dependent Variable: Total_Sales Total Sales

Number of Observations Read	9648
Number of Observations Used	9648

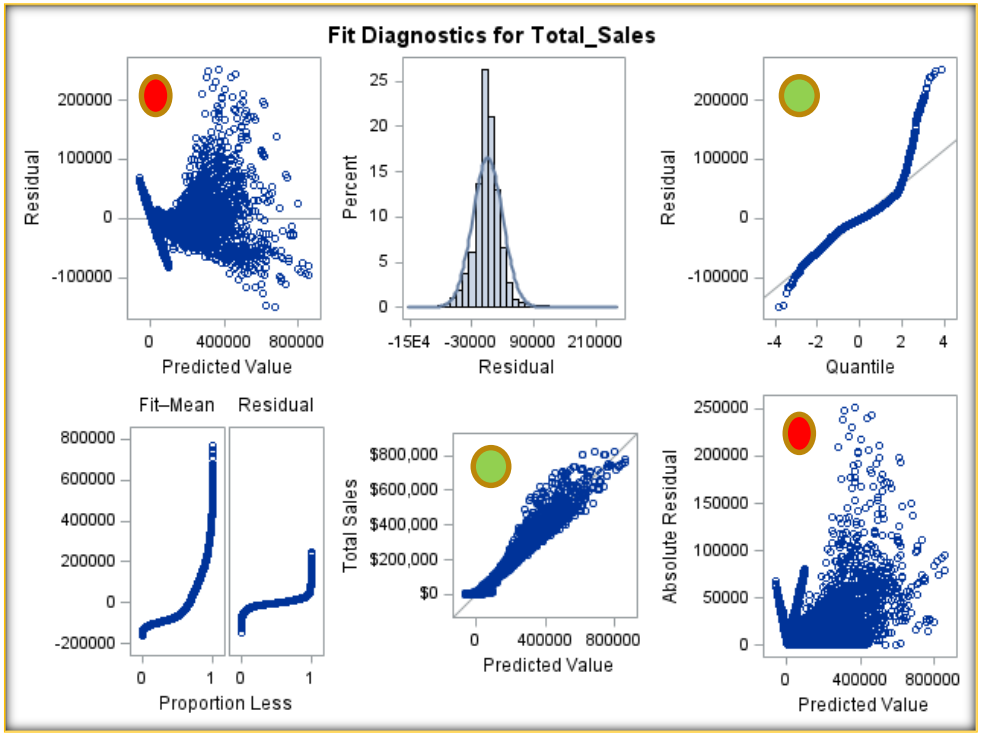
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.835818E14	6.119394E13	55101.8	<.0001
Error	9644	1.071025E13	1110561392		
Corrected Total	9647	1.942921E14			

Root MSE	33325	R-Square	0.9449
Dependent Mean	93273	Adj R-Sq	0.9449
Coeff Var	35.72838		

- After removing Units sold VIF of rest of numerical predictors become less than 5.
- R squared is still excellent using 3 numerical predictors

REGRESSION ASSUMPTIONS



- **Linearity** appears mostly satisfied, but potential deviations at high values.
- **Residuals are not perfectly normal**, especially in the tails.
- **Homoscedasticity is violated** (variance of residuals increases with predicted values).
- **Possible correlation in residuals**, indicating systematic error patterns.

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

The GLM Procedure

Dependent Variable: Total_Sales Total Sales

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	1.8617777E14	9.3088885E12	11044.3	<.0001
Error	9627	8.1143131E12	842870379		
Corrected Total	9647	1.9429208E14			

R-Square	Coeff Var	Root MSE	Total_Sales Mean
0.958237	31.12594	29032.23	93273.44

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Retailer	5	1.0499984E12	209999672960	249.15	<.0001
Product	5	4.5648856E12	912977112984	1083.18	<.0001
Sales_Method	2	3.0862513E13	1.5431256E13	18308.0	<.0001
Region	4	6.9229438E12	1.7307359E12	2053.38	<.0001
Price_per_Unit	1	3.3607559E13	3.3607559E13	39872.7	<.0001
Units_Sold	1	9.7027394E13	9.7027394E13	115115	<.0001
Operating_Profit	1	1.0171019E13	1.0171019E13	12067.1	<.0001
Operating_Margin	1	1.9714568E12	1.9714568E12	2338.98	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Retailer	5	28879921694	5775984338.8	6.85	<.0001
Product	5	142155800876	28431160175	33.73	<.0001
Sales_Method	2	73814903247	36907451624	43.79	<.0001
Region	4	49767666618	12441916654	14.76	<.0001
Price_per_Unit	1	918393406768	918393406768	1089.60	<.0001
Units_Sold	1	2.2909135E12	2.2909135E12	2717.99	<.0001
Operating_Profit	1	1.1822054E13	1.1822054E13	14025.9	<.0001
Operating_Margin	1	1.9714568E12	1.9714568E12	2338.98	<.0001

Strongest Predictor

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	24819.4008	B	2402.533856	10.33	<.0001
Retailer Amazon	-976.1232	B	1216.520328	-0.80	0.4223
Retailer Foot Locker	-1070.1557	B	904.454724	-1.18	0.2368
Retailer Kohl's	-3409.4309	B	1107.152920	-3.08	0.0021
Retailer Sports Direct	-4854.6580	B	971.388289	-5.00	<.0001
Retailer Walmart	-3896.0679	B	1426.441859	-2.73	0.0063
Retailer West Gear	0.0000	B	.	.	.
Product Men's Apparel	-109.7870	B	1065.503572	-0.10	0.9179
Product Men's Athletic Footwear	-637.3037	B	1028.309175	-0.62	0.5354
Product Men's Street Footwear	-9817.0449	B	1069.781954	-9.18	<.0001
Product Women's Apparel	-1707.2314	B	1065.681217	-1.60	0.1092
Product Women's Athletic Footwear	3942.3277	B	1031.389281	3.82	0.0001
Product Women's Street Footwear	0.0000	B	.	.	.
Sales_Method In-store	7584.5429	B	978.560788	7.75	<.0001
Sales_Method Online	5898.2495	B	763.084196	7.73	<.0001
Sales_Method Outlet	0.0000	B	.	.	.
Region Midwest	4417.0206	B	1005.930577	4.39	<.0001
Region Northeast	332.0714	B	956.706084	0.35	0.7285
Region South	565.3635	B	1100.189163	0.51	0.6073
Region Southeast	-4245.0194	B	1142.902436	-3.71	0.0002
Region West	0.0000	B	.	.	.
Price_per_Unit	853.0026		25.841418	33.01	<.0001
Units_Sold	191.9430		3.681700	52.13	<.0001
Operating_Profit	1.6800		0.014185	118.43	<.0001
Operating_Margin	-186069.3509		3847.349225	-48.36	<.0001

- Insignificant Features based on p values
- Retailer Amazon
- Retailer Footlocker
- Product Men’s Apparel
- Product Men’s Athletic Footwear
- Product Women’s Apparel
- Region Northeast
- Region South

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

The GLM Procedure
Least Squares Means

Sales_Method	Total_Sales LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
In-store	95876.5866	853.6306	<.0001	1
Online	94190.2932	479.7057	<.0001	2
Outlet	88292.0437	596.2392	<.0001	3

Least Squares Means for effect Sales_Method
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Total_Sales

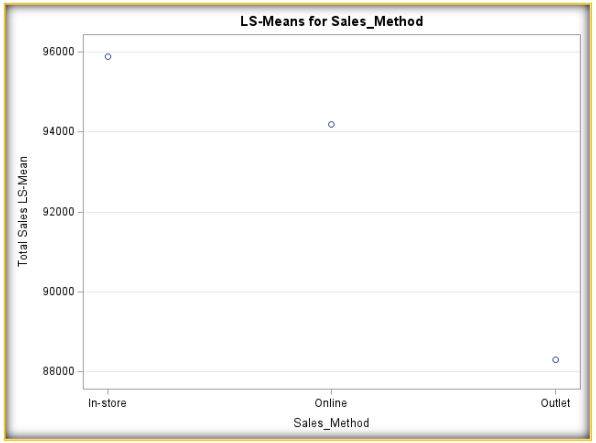
i/j	1	2	3
1		0.0902	<.0001
2	0.0902		<.0001
3	<.0001	<.0001	

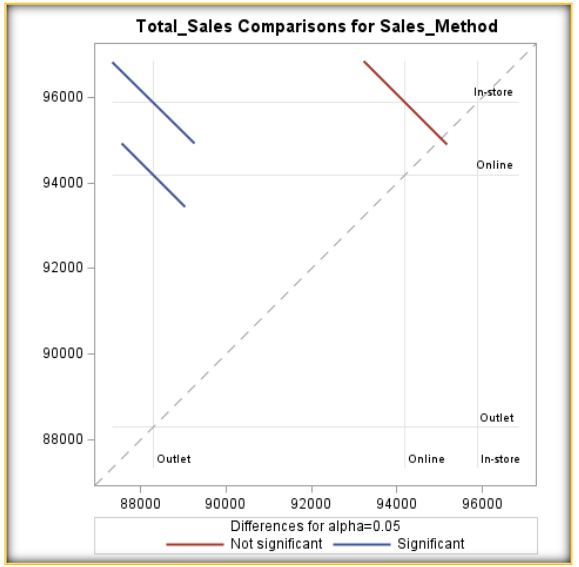
Sales_Method	Total_Sales LSMEAN	95% Confidence Limits	
In-store	95877	94203	97550
Online	94190	93250	95131
Outlet	88292	87123	89461

Least Squares Means for Effect Sales_Method

i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	1686.293376	-264.119483	3636.706234
1	3	7584.542874	5666.357808	9502.727940
2	3	5898.249498	4402.443896	7394.055100

Least Square Means –Sales Method





- In store Outperform where has Outlet method has lower sales
- Outlet sales differ significantly from both Online and In-store sales (blue lines).
- The difference between In-store and Online sales is not statistically significant (red line), meaning their total sales are similar.

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

Least Square Means –Retailer

Retailer	Total_Sales LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
Amazon	94177.9240	1026.2795	<.0001	1
Foot Locker	94083.8914	626.8616	<.0001	2
Kohl's	91744.6163	989.6119	<.0001	3
Sports Direct	90299.3891	687.9684	<.0001	4
Walmart	91257.9792	1247.7928	<.0001	5
West Gear	95154.0471	661.9643	<.0001	6

Least Squares Means for effect Retailer
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Total_Sales

i/j	1	2	3	4	5	6
1		0.9332	0.0785	0.0014	0.0708	0.4223
2	0.9332		0.0370	<.0001	0.0443	0.2368
3	0.0785	0.0370		0.2299	0.7622	0.0021
4	0.0014	<.0001	0.2299		0.4829	<.0001
5	0.0708	0.0443	0.7622	0.4829		0.0063
6	0.4223	0.2368	0.0021	<.0001	0.0063	

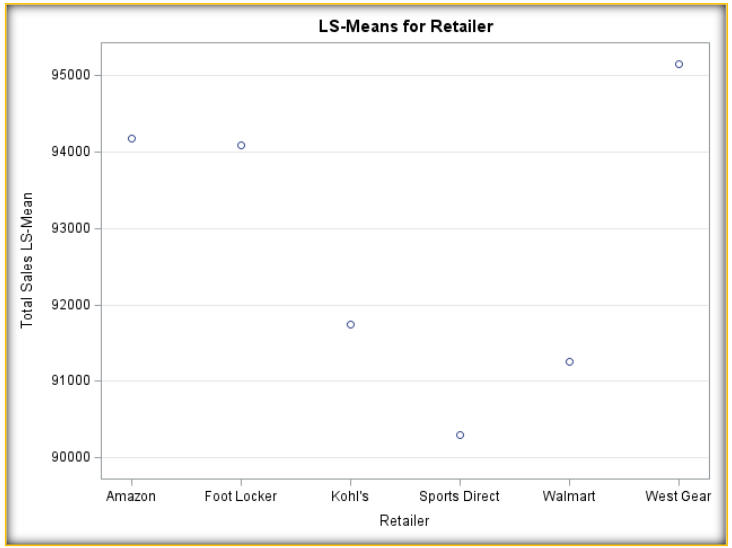
Retailer	Total_Sales LSMEAN	95% Confidence Limits	
Amazon	94178	92166	96190
Foot Locker	94084	92855	95313
Kohl's	91745	89805	93684
Sports Direct	90299	88951	91648
Walmart	91258	88812	93704
West Gear	95154	93856	96452

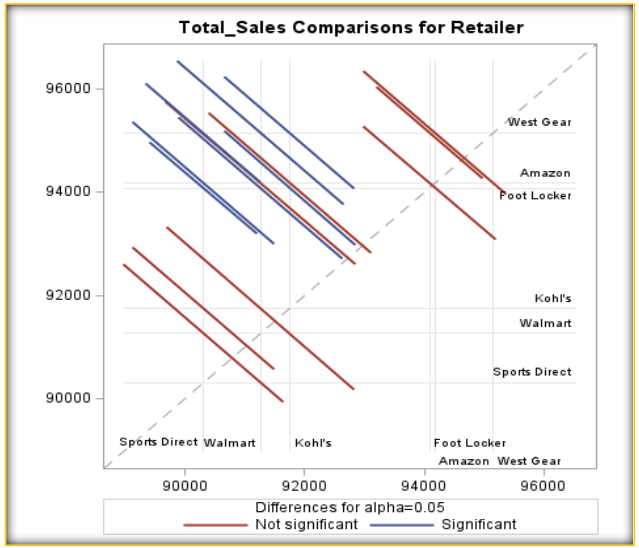
Least Squares Means for Effect Retailer

i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	94.032537	-2104.969047	2293.034120
1	3	2433.307709	-277.079148	5143.694567
1	4	3878.534861	1504.530282	6252.539440
1	5	2919.944790	-248.116075	6088.005655
1	6	-976.123159	-3360.758999	1408.512681
2	3	2339.275173	140.759530	4537.790816
2	4	3784.502325	1998.965147	5570.039503
2	5	2825.912253	72.368209	5579.456297
2	6	-1070.155695	-2843.077282	702.765891
3	4	1445.227152	-914.141523	3804.595827
3	5	486.637080	-2665.146249	3638.420409
3	6	-3409.430868	-5579.683574	-1239.178163
4	5	-958.590072	-3636.267875	1719.087731
4	6	-4854.658020	-6758.783480	-2950.532561
5	6	-3896.067948	-6692.194163	-1099.941734

LINEAR REGRESSION
PRODUCT PLOT

Least Square Means –Retailer





- West Gear outperform followed by Amazon and Footlocker
- Sports Direct Has lower Total Sales as compared to Others
- Comparisons involving Sports Direct and Walmart seem to have significant differences (blue lines).
- Comparisons involving West Gear, Amazon, and Foot Locker are more likely to be statistically not significant (red lines)
- There is a clear separation in Total Sales between some retailer groups.

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

Least Square Means –Region

Region	Total_Sales LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
Midwest	96989.4412	766.0564	<.0001	1
Northeast	92904.4920	652.4550	<.0001	2
South	93137.7841	849.3724	<.0001	3
Southeast	88327.4013	922.4045	<.0001	4
West	92572.4207	737.3478	<.0001	5

Least Squares Means for effect Region
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Total_Sales

i/j	1	2	3	4	5
1		<.0001	0.0006	<.0001	<.0001
2	<.0001		0.8326	<.0001	0.7285
3	0.0006	0.8326		<.0001	0.6073
4	<.0001	<.0001	<.0001		0.0002
5	<.0001	0.7285	0.6073	0.0002	

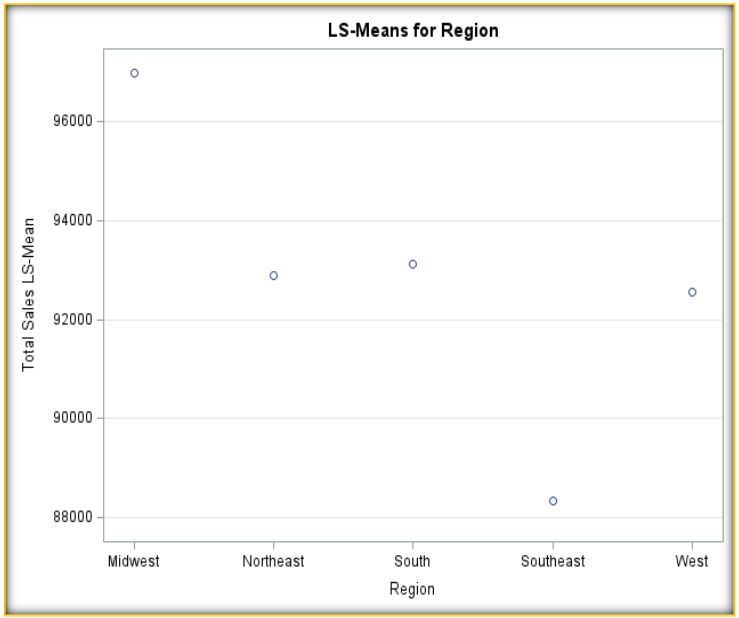
Region	Total_Sales LSMEAN	95% Confidence Limits	
Midwest	96989	95488	98491
Northeast	92904	91626	94183
South	93138	91473	94803
Southeast	88327	86519	90136
West	92572	91127	94018

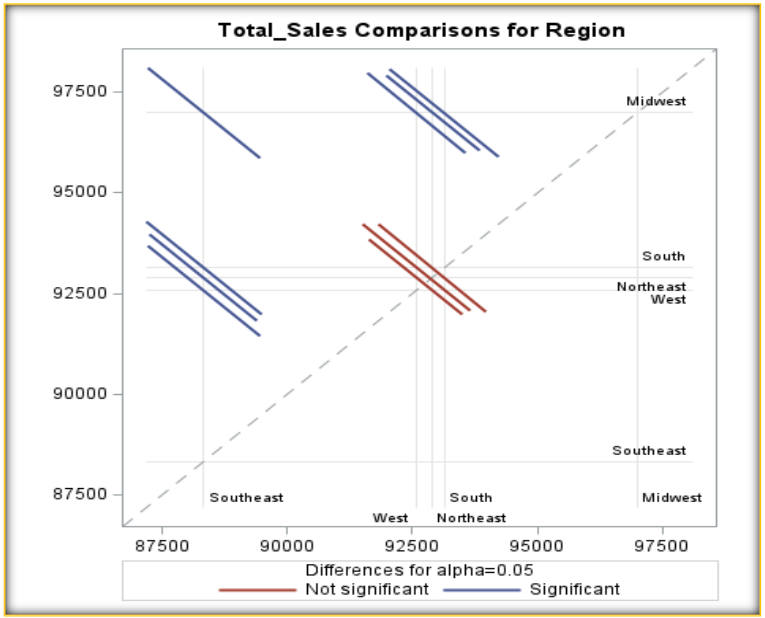
Least Squares Means for Effect Region

i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	4084.949194	2231.949545	5937.948844
1	3	3851.657097	1651.187723	6052.126471
1	4	8662.039963	6415.525958	10909
1	5	4417.020564	2445.184951	6388.856177
2	3	-233.292097	-2396.672777	1930.088582
2	4	4577.090768	2431.731634	6722.449903
2	5	332.071370	-1543.273878	2207.416618
3	4	4810.382866	2498.902729	7121.863002
3	5	565.363467	-1591.238810	2721.965744
4	5	-4245.019398	-6485.348679	-2004.690118

LINEAR REGRESSION
PRODUCT PLOT

Least Square Means –Region





- Midwest Outperform whereas South east has Lower Sales
- Comparisons involving the Midwest and Southeast regions seem to be statistically significant (blue lines), suggesting that their total sales differ from other regions.
- Comparisons involving the South, Northeast, and West regions are mostly not significant (red lines), indicating that their total sales may be similar.

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

Least Square Means –Product

Product	Total_Sales LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
Men's Apparel	94064.6940	789.9123	<.0001	1
Men's Athletic Footwear	93537.1774	783.9860	<.0001	2
Men's Street Footwear	84357.4362	828.3616	<.0001	3
Women's Apparel	92467.2497	782.5679	<.0001	4
Women's Athletic Footwear	98116.8088	777.7258	<.0001	5
Women's Street Footwear	94174.4811	785.6180	<.0001	6

Least Squares Means for effect Product
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Total_Sales

i/j	1	2	3	4	5	6
1		0.6184	<.0001	0.1262	0.0001	0.9179
2	0.6184		<.0001	0.3094	<.0001	0.5354
3	<.0001	<.0001		<.0001	<.0001	<.0001
4	0.1262	0.3094	<.0001		<.0001	0.1092
5	0.0001	<.0001	<.0001	<.0001		0.0001
6	0.9179	0.5354	<.0001	0.1092	0.0001	

Product	Total_Sales LSMEAN	95% Confidence Limits	
Men's Apparel	94065	92516	95613
Men's Athletic Footwear	93537	92000	95074
Men's Street Footwear	84357	82734	85981
Women's Apparel	92467	90933	94001
Women's Athletic Footwear	98117	96592	99641
Women's Street Footwear	94174	92635	95714

Least Squares Means for Effect Product

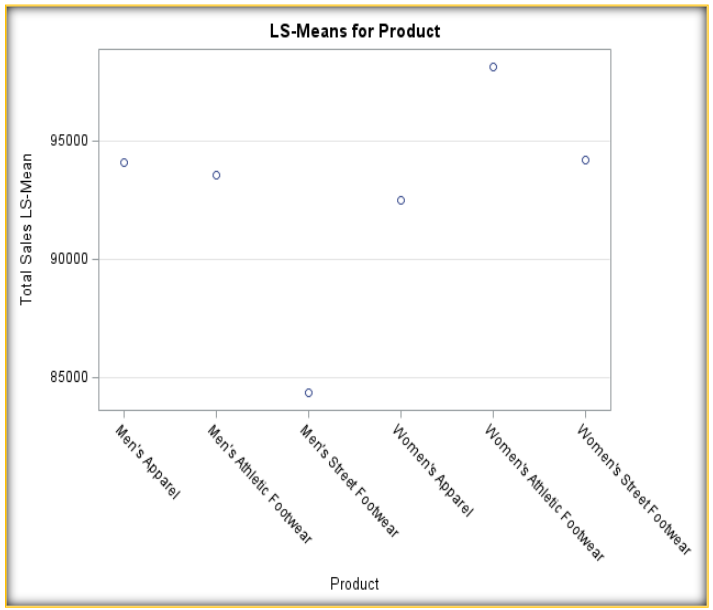
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	527.516639	-1548.475780	2603.509058
1	3	9707.257881	7496.232565	11918
1	4	1597.444368	-450.109982	3644.998718
1	5	-4052.114761	-6109.575950	-1994.653572
1	6	-109.787019	-2198.398237	1978.824199
2	3	9179.741242	7103.081429	11256
2	4	1069.927729	-993.059335	3132.914794
2	5	-4579.631400	-6618.447236	-2540.815564
2	6	-637.303658	-2653.006032	1378.398716
3	4	-8109.813513	-10209	-6010.527137
3	5	-13759	-15903	-11616
3	6	-9817.044900	-11914	-7720.047152
4	5	-5649.559129	-7729.126140	-3569.992118
4	6	-1707.231387	-3796.190827	381.728053
5	6	3942.327742	1920.587712	5964.067772

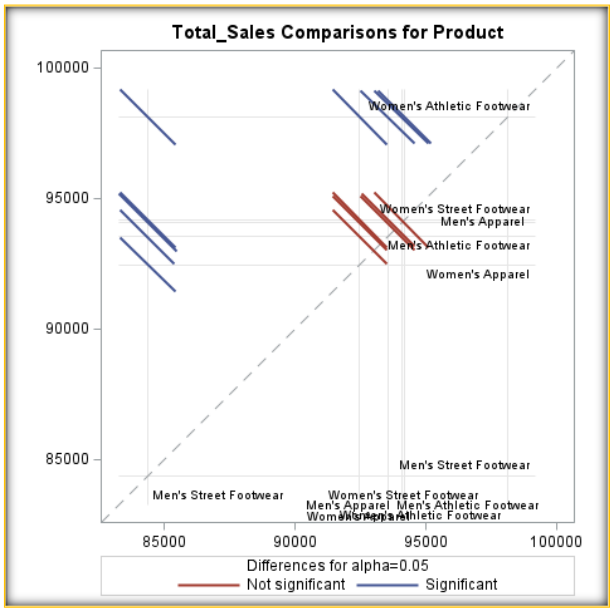
3	4	-8109.813513	-10209	-6010.527137

3	5	-13759	-15903	-11616
3	6	-9817.044900	-11914	-7720.047152
4	5	-5649.559129	-7729.126140	-3569.992118
4	6	-1707.231387	-3796.190827	381.728053
5	6	3942.327742	1920.587712	5964.067772

LINEAR REGRESSION
PRODUCT PLOT

Least Square Means –Product



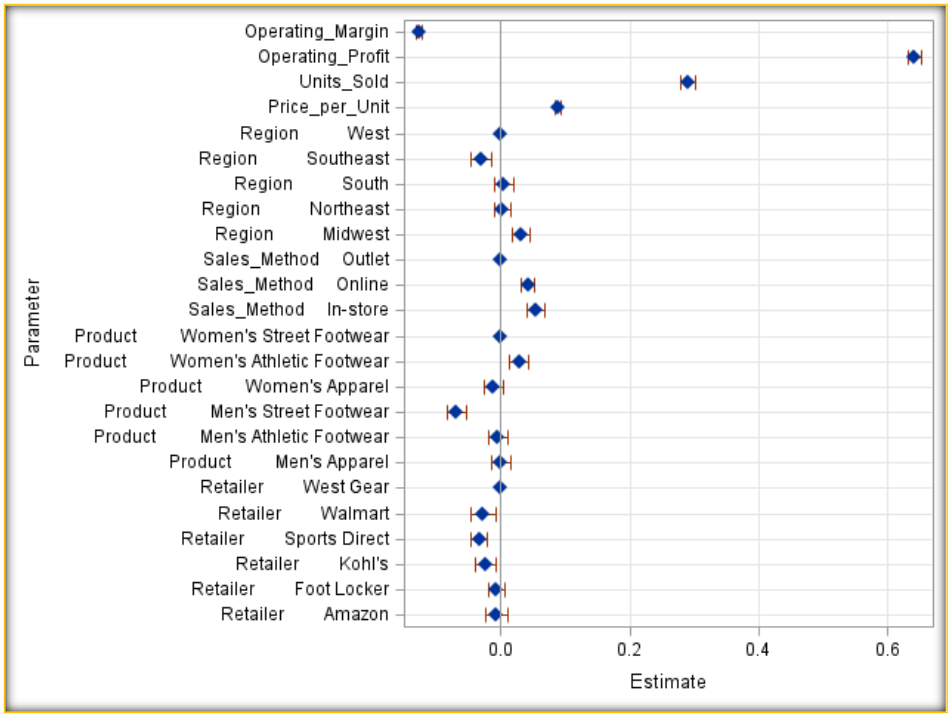


- Women's Athletic Footwear is top performers, while Men's Street Footwear have lower total sales.

- Some product groups (Women's Street Footwear vs. Men's Apparel) have similar sales, meaning there is no significant difference.

LINEAR REGRESSION
USING PROC GLM FOR MIXED VARIABLES

Visualisation of Coefficient



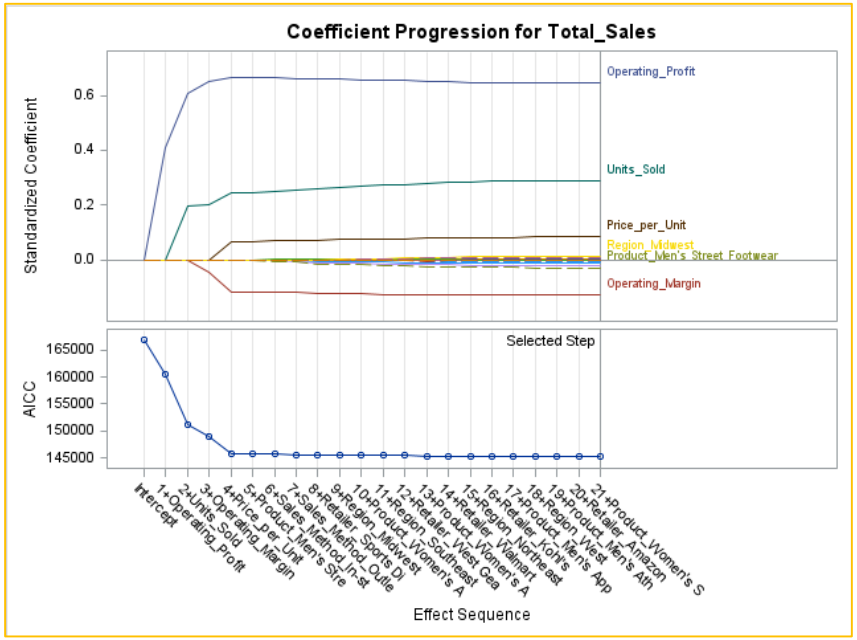
Top Predictors

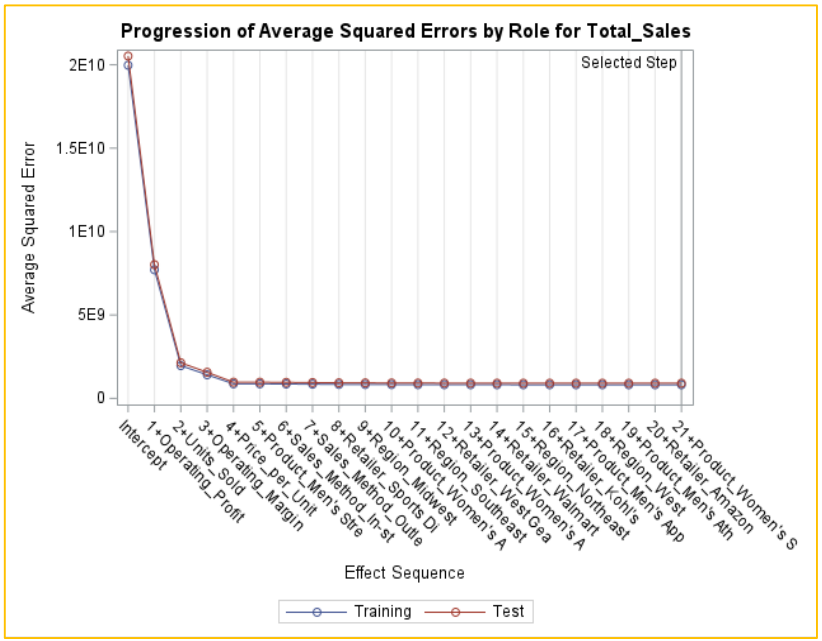
- Operating Profit
- Units Sold
- Price per Unit
- Operating Margin

Not Significant

- Region South
- Region Northeast
- Product Woman’s Apparel
- Product Men’s Apparel
- Retailer Footlocker
- Retailer Amazon

LINEAR REGRESSION
PREDICTIVE MODELLING





LINEAR REGRESSION
PREDICTIVE MODELLING

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	20	1.293918E14	6.46959E12	7912.40
Error	6733	5.505254E12	817652482	
Corrected Total	6753	1.348971E14		

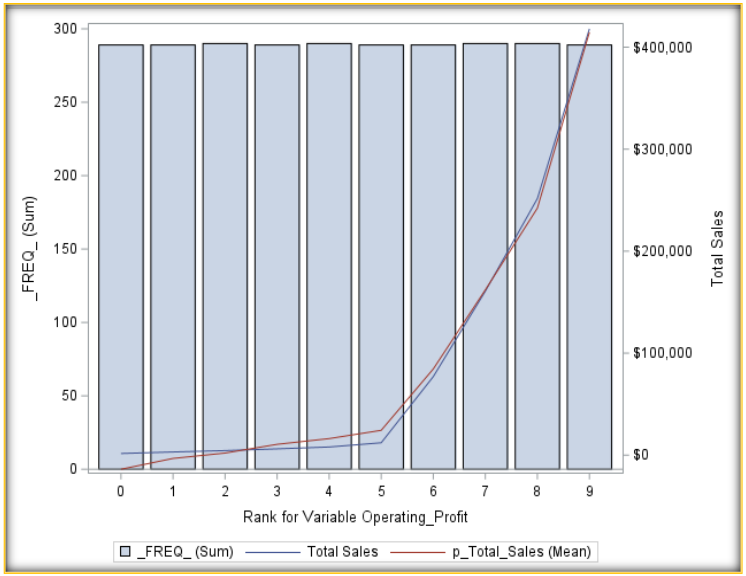
Root MSE	28595
Dependent Mean	92840
R-Square	0.9592
Adj R-Sq	0.9591
AIC	145382
AICC	145382
SBC	138769
ASE (Train)	815110181
ASE (Test)	905935666

LINEAR REGRESSION OUTPUT

Parameter	DF	Estimate
Intercept	1	31748
Operating_Profit	1	1.691713
Operating_Margin	1	-185758
Units_Sold	1	191.019394
Price_per_Unit	1	803.655850
Retailer_Amazon	1	2.624555
Retailer_Kohl's	1	-443.702567
Retailer_Sports Direct	1	-4093.914224
Retailer_Walmart	1	-1746.090033
Retailer_West Gear	1	1129.286334
Sales_Method_In-store	1	229.570145
Sales_Method_Outlet	1	-6026.116080
Region_Midwest	1	4238.647288
Region_Northeast	1	786.571435
Region_Southeast	1	-3438.072148
Region_West	1	-314.637370
Product_Men's Apparel	1	213.730994
Product_Men's Athletic Footwear	1	-21.774194
Product_Men's Street Footwear	1	-10502
Product_Women's Apparel	1	-2356.999050
Product_Women's Athletic Footwear	1	3301.132350
Product_Women's Street Footwear	0	0

LINEAR REGRESSION
PREDICTION VISUALIZATION

Performance by Operating Profit

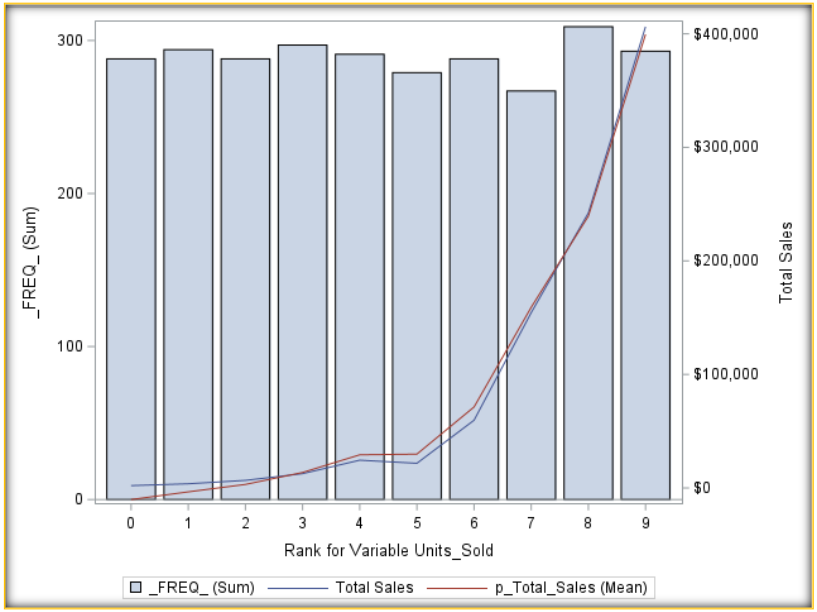


Rank for Variable Operating_Profit	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
0	1	Total_Sales	Total Sales	1	1594.94	.	1594.94	1594.94
		p_Total_Sales		1	-13836.62	.	-13836.62	-13836.62
1	1	Total_Sales	Total Sales	1	3092.88	.	3092.88	3092.88
		p_Total_Sales		1	-3236.02	.	-3236.02	-3236.02
2	1	Total_Sales	Total Sales	1	4471.85	.	4471.85	4471.85
		p_Total_Sales		1	2038.37	.	2038.37	2038.37
3	1	Total_Sales	Total Sales	1	6153.36	.	6153.36	6153.36
		p_Total_Sales		1	10609.71	.	10609.71	10609.71
4	1	Total_Sales	Total Sales	1	7982.23	.	7982.23	7982.23
		p_Total_Sales		1	16214.27	.	16214.27	16214.27
5	1	Total_Sales	Total Sales	1	12126.90	.	12126.90	12126.90
		p_Total_Sales		1	24341.09	.	24341.09	24341.09
6	1	Total_Sales	Total Sales	1	76965.55	.	76965.55	76965.55
		p_Total_Sales		1	84484.31	.	84484.31	84484.31
7	1	Total_Sales	Total Sales	1	160737.07	.	160737.07	160737.07
		p_Total_Sales		1	162015.18	.	162015.18	162015.18
8	1	Total_Sales	Total Sales	1	251547.41	.	251547.41	251547.41
		p_Total_Sales		1	242165.40	.	242165.40	242165.40
9	1	Total_Sales	Total Sales	1	418006.92	.	418006.92	418006.92
		p_Total_Sales		1	414645.20	.	414645.20	414645.20

Higher Operating Profit correlates with significantly higher Total Sales

LINEAR REGRESSION
PREDICTION VISUALIZATION

Performance by Units Sold

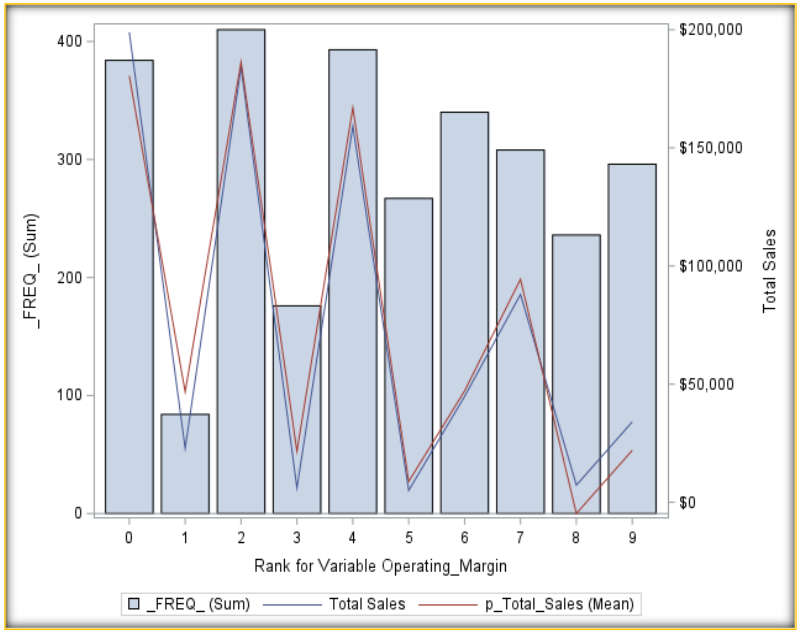


Rank for Variable Units_Sold	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
0	1	Total_Sales	Total Sales	1	2079.94	.	2079.94	2079.94
		p_Total_Sales		1	-10186.08	.	-10186.08	-10186.08
1	1	Total_Sales	Total Sales	1	3704.11	.	3704.11	3704.11
		p_Total_Sales		1	-3417.63	.	-3417.63	-3417.63
2	1	Total_Sales	Total Sales	1	6704.94	.	6704.94	6704.94
		p_Total_Sales		1	3127.60	.	3127.60	3127.60
3	1	Total_Sales	Total Sales	1	12608.47	.	12608.47	12608.47
		p_Total_Sales		1	13663.89	.	13663.89	13663.89
4	1	Total_Sales	Total Sales	1	24471.53	.	24471.53	24471.53
		p_Total_Sales		1	29205.42	.	29205.42	29205.42
5	1	Total_Sales	Total Sales	1	21747.27	.	21747.27	21747.27
		p_Total_Sales		1	29821.01	.	29821.01	29821.01
6	1	Total_Sales	Total Sales	1	59670.19	.	59670.19	59670.19
		p_Total_Sales		1	71336.62	.	71336.62	71336.62
7	1	Total_Sales	Total Sales	1	154422.35	.	154422.35	154422.35
		p_Total_Sales		1	159169.39	.	159169.39	159169.39
8	1	Total_Sales	Total Sales	1	242275.89	.	242275.89	242275.89
		p_Total_Sales		1	239428.04	.	239428.04	239428.04
9	1	Total_Sales	Total Sales	1	406237.20	.	406237.20	406237.20
		p_Total_Sales		1	399505.18	.	399505.18	399505.18

Higher Rank of Units Sold correlates with significantly higher Total Sales

LINEAR REGRESSION
PREDICTION VISUALIZATION

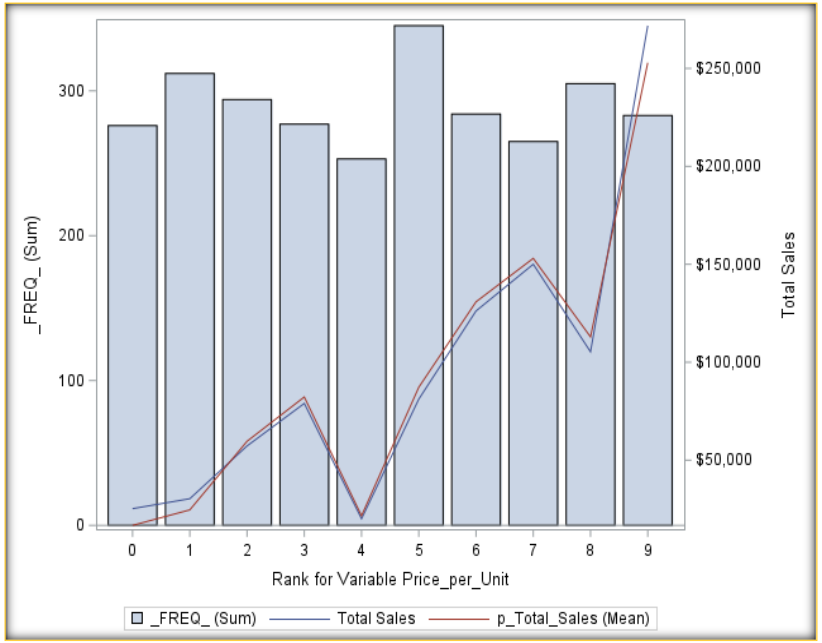
Performance by Operating_Margin



Rank for Variable Operating_Margin	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
0	1	Total_Sales	Total Sales	1	198935.38	.	198935.38	198935.38
		p_Total_Sales		1	180435.87	.	180435.87	180435.87
1	1	Total_Sales	Total Sales	1	22815.05	.	22815.05	22815.05
		p_Total_Sales		1	46741.46	.	46741.46	46741.46
2	1	Total_Sales	Total Sales	1	184012.30	.	184012.30	184012.30
		p_Total_Sales		1	186381.60	.	186381.60	186381.60
3	1	Total_Sales	Total Sales	1	6096.86	.	6096.86	6096.86
		p_Total_Sales		1	21592.26	.	21592.26	21592.26
4	1	Total_Sales	Total Sales	1	159172.87	.	159172.87	159172.87
		p_Total_Sales		1	167004.30	.	167004.30	167004.30
5	1	Total_Sales	Total Sales	1	5052.69	.	5052.69	5052.69
		p_Total_Sales		1	8844.89	.	8844.89	8844.89
6	1	Total_Sales	Total Sales	1	44808.18	.	44808.18	44808.18
		p_Total_Sales		1	47145.29	.	47145.29	47145.29
7	1	Total_Sales	Total Sales	1	87980.42	.	87980.42	87980.42
		p_Total_Sales		1	94392.63	.	94392.63	94392.63
8	1	Total_Sales	Total Sales	1	7264.79	.	7264.79	7264.79
		p_Total_Sales		1	-4729.23	.	-4729.23	-4729.23
9	1	Total_Sales	Total Sales	1	34060.87	.	34060.87	34060.87
		p_Total_Sales		1	21993.21	.	21993.21	21993.21

LINEAR REGRESSION
PREDICTION VISUALIZATION

Performance by Price_per_unit



Rank for Variable Price_per_Unit	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
0	1	Total_Sales	Total Sales	1	25257.87	.	25257.87	25257.87
		p_Total_Sales		1	16718.16	.	16718.16	16718.16
1	1	Total_Sales	Total Sales	1	30334.06	.	30334.06	30334.06
		p_Total_Sales		1	24625.61	.	24625.61	24625.61
2	1	Total_Sales	Total Sales	1	57259.88	.	57259.88	57259.88
		p_Total_Sales		1	59653.10	.	59653.10	59653.10
3	1	Total_Sales	Total Sales	1	78976.61	.	78976.61	78976.61
		p_Total_Sales		1	82262.02	.	82262.02	82262.02
4	1	Total_Sales	Total Sales	1	20085.55	.	20085.55	20085.55
		p_Total_Sales		1	21627.51	.	21627.51	21627.51
5	1	Total_Sales	Total Sales	1	81140.45	.	81140.45	81140.45
		p_Total_Sales		1	87220.42	.	87220.42	87220.42
6	1	Total_Sales	Total Sales	1	126174.08	.	126174.08	126174.08
		p_Total_Sales		1	130848.15	.	130848.15	130848.15
7	1	Total_Sales	Total Sales	1	150028.80	.	150028.80	150028.80
		p_Total_Sales		1	153037.68	.	153037.68	153037.68
8	1	Total_Sales	Total Sales	1	105320.28	.	105320.28	105320.28
		p_Total_Sales		1	112937.09	.	112937.09	112937.09
9	1	Total_Sales	Total Sales	1	271815.64	.	271815.64	271815.64
		p_Total_Sales		1	252912.99	.	252912.99	252912.99

SUMMARY

- Using Linear Regression Model 95.82% of variance in Total sales can be explained using independent variables
- Operating Profit, Units Sold, Price per unit and Operating margin are Top 4 important features for prediction
- No Clear Correlation between Operating Margin and Total Sales., suggesting that higher margins do not necessarily lead to higher revenue
- Strong positive correlation between Units sold and Total Sales clearly indicates that total revenue is directly influenced by sales volume
- High Priced Products tend to generate more revenue but mid ranged products showed inconstancy
- Apart from top 4 features following are important contributors in total sales
    - Sales Method- Instore
    - Retailer- West Gear,Amazon,Foot locker
    - Region- Midwest
    - Product- Woman’s Athletic Footwear


THANK YOU
KEEP SHOPPING WITH ADIDAS