Project Report

Superstore Statistical Analysis

(Microsoft Excel, 2019)

Index

S.No	TOPIC	PAGE No
1	INTRODUCTION	3
2	TASKS	4-16
3	SUMMARY	17
4	REFERENCES	17

Introduction

Superstores are very large supermarkets or shops selling household goods and other equipments. It's usually built outside city centers away from other shops.

The advantages of a Superstore:

- Saving labor cost due to self-service systems.
- Superstore has large turnover.
- Reasonable or low prices of goods.
- Low cost of operation.
- Freedom of selection.
- Shopping is very easy and quick.
- Due to adequate parking space, shopping becomes an easy and pleasing activity rather than boredom.
- High degree of efficiency due to elimination of service.
- High margin of profit to organizers.
- Advantages of large scale operations.

The disadvantages of a Superstore:

- Superstore requires huge financial resources.
- It is normally situated at a long distance from the residential localities.
- There is a lack of personal attention.
- It faces the problem of coordinating activities of various sections of the market.
- It requires large and extensive premises.
- Goods which require explanation by salesmen cannot be sold in such markets.

Most people prefer Superstore because of its special features, offers and discounts. Goods are found at relatively low prices when compared to other local stores. They take less profit from customers.

Now, Let's analyze the dataset of a Superstore in Excel using the Data Tool.

TASKS

Task 1: Descriptive Statistics

- Select the columns for Descriptive Statistical Analysis

- Select Data tab and Data Analysis

- Now select Descriptive statistics from the menu

- Select the input range and output range

- We can observe the Mean, Standard Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum, Count for Sales, Profit and Loss Columns
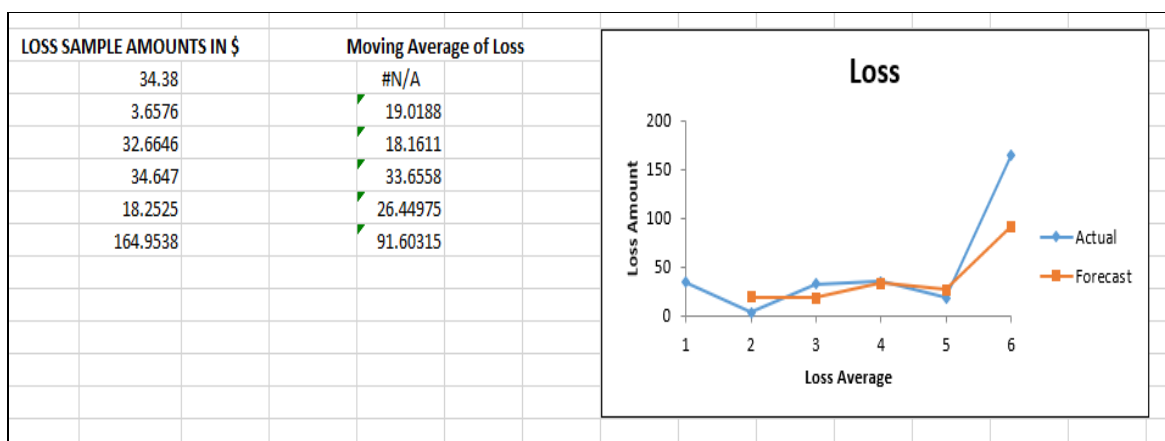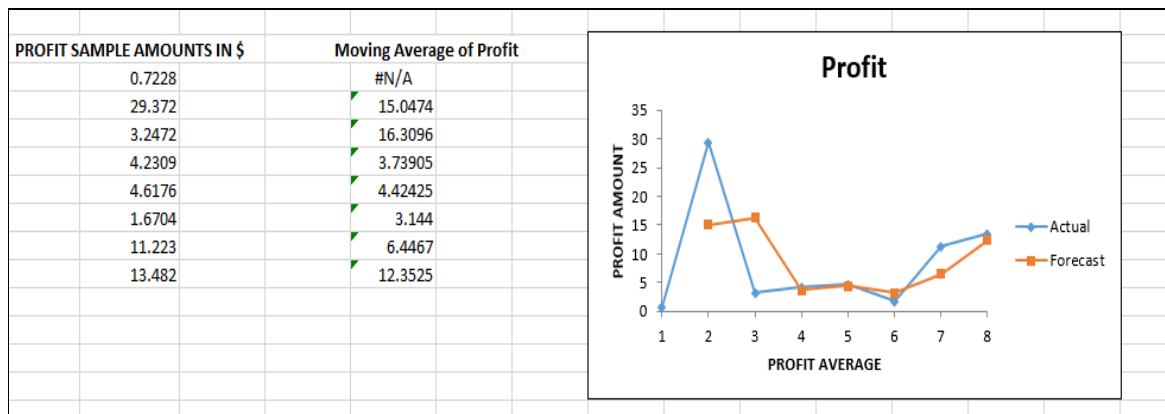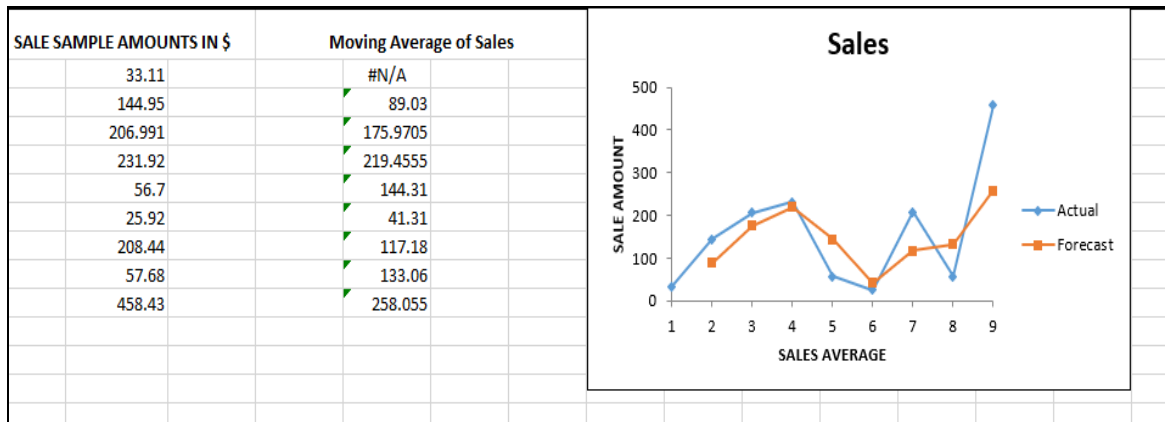
Sales in $			SALES OBSERVATION	
Mean	229.8580008		TOTAL AMOUNT EARNED (SALES)	$ 22,97,200.86
Standard Error	6.234321582		MAXIMUM SALE AMOUNT	$ 22,638.48
Median	54.49		MINIMUM SALE AMOUNT	$ 0.44
Mode	12.96		TOTAL SALES	9,994
Standard Deviation	623.2451005		AVERAGE SALE AMOUNT	$ 229.86
Sample Variance	388434.4553			
Kurtosis	305.3117532			
Skewness	12.97275234			
Range	22638.036			
Minimum	0.444			
Maximum	22638.48			
Sum	2297200.86			
Count	9994			

Profit in $	
Mean	54.91788377
Standard Error	2.38078019
Median	13.3176
Mode	6.2208
Standard Deviation	213.7139798
Sample Variance	45673.66515
Kurtosis	541.811814
Skewness	19.14304951
Range	8399.9132
Minimum	0.0628
Maximum	8399.976
Sum	442528.3074
Count	8058

PROFIT OBSERVATION	
TOTAL PROFIT	$ 4,42,528.31
MAXIMUM PROFIT	$ 8,399.98
MINIMUM PROFIT	$ 0.06
TOTAL SALES WHERE PROFIT EARNED	8,058
AVERAGE PROFIT	$ 54.92

Loss in $	
Mean	83.44804153
Standard Error	6.575495182
Median	18.0882
Mode	10.1736
Standard Deviation	284.4234223
Sample Variance	80896.68313
Kurtosis	200.8163361
Skewness	11.85555888
Range	6599.8885
Minimum	0.0895
Maximum	6599.978
Sum	156131.2857
Count	1871

LOSS OBSERVATION	
TOTAL LOSS	$ 1,56,131.29
MAXIMUM LOSS	$ 6,599.98
MINIMUM LOSS	$ 0.09
TOTAL SALES WHERE LOSS	1,871
AVERAGE LOSS	$ 83.45

Task 2: Sampling and Moving Average

- Sampling can be selected from the Data Analysis Tab
- Select the input range, number of random samples and output range
- Moving average can be selected from the Data Analysis Tab
- Moving average was calculated on sample sale, profit and loss amounts.

SALE SAMPLE AMOUNTS IN $	Moving Average of Sales
33.11	#N/A
144.95	89.03
206.991	175.9705
231.92	219.4555
56.7	144.31
25.92	41.31
208.44	117.18
57.68	133.06
458.43	258.055

PROFIT SAMPLE AMOUNTS IN $	Moving Average of Profit
0.7228	#N/A
29.372	15.0474
3.2472	16.3096
4.2309	3.73905
4.6176	4.42425
1.6704	3.144
11.223	6.4467
13.482	12.3525

LOSS SAMPLE AMOUNTS IN $	Moving Average of Loss
34.38	#N/A
3.6576	19.0188
32.6646	18.1611
34.647	33.6558
18.2525	26.44975
164.9538	91.60315

- Histogram was also created on the sales column based on price ranges.

HISTOGRAM		

SALES BIN		SALES BIN	Frequency
$ 100.00		$ 100.00	6229
$ 500.00		$ 500.00	2603
$ 1,000.00		$1,000.00	694
$ 2,000.00		$2,000.00	328
		More	140
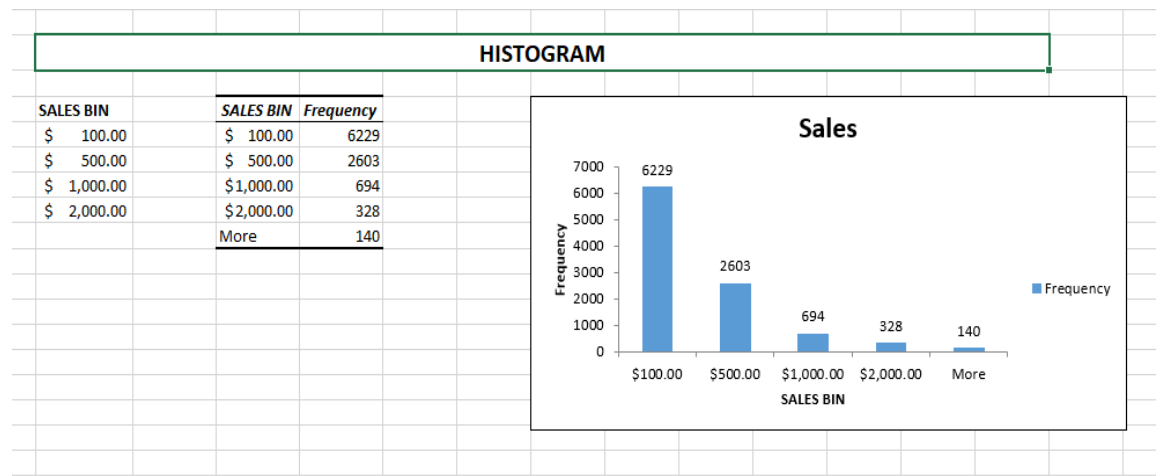
Task 3: Covariance and Correlation

- Covariance and correlation can be selected from the Data Analysis Tab

- If the covariance correlation has positive values then the data columns have a positive covariance and correlation.

- If the covariance correlation has negative values then the columns have a negative covariance and correlation.

- Covariance and correlation was calculated on Sales, Profit and Loss columns.

Covariance (Sales, Profit)		
	Sales	Profit
Sales	388395.5885	
Profit	90977.71906	45667.99703
Observation		
1. Sales and Profit data have a positive covariance		
2. Sales Increase then profit increases		

Covariance (Sales, Loss)		
	Sales in $	Loss in $
Sales in $	388395.5885	
Loss in $	-1695.999289	80853.44599

Observation
1. Sales and Loss data have a negative covariance
2. Sales Increase then loss decreases

Correlation (Sales, Profit)		
	Sales	Profit
Sales	1	
Profit	0.7559404	1

Observation
1. Correlation between sales and profit data is 0.75
2. Sales, Profit dataset has strong positive correlation

Correlation (Sales, Loss)		
	Sales in $	Loss in $
Sales in $	1	
Loss in $	-0.011128	1

| Observation |
|---|
| 1. Correlation between sales and loss data is 0 |
| 2. Sales, loss dataset has no correlation |

Task 4: Reports

- Data can be easily visualized in Excel through Charts, Bar graphs
- Select Insert Tab and Charts

BAR GRAPHS AND CHARTS
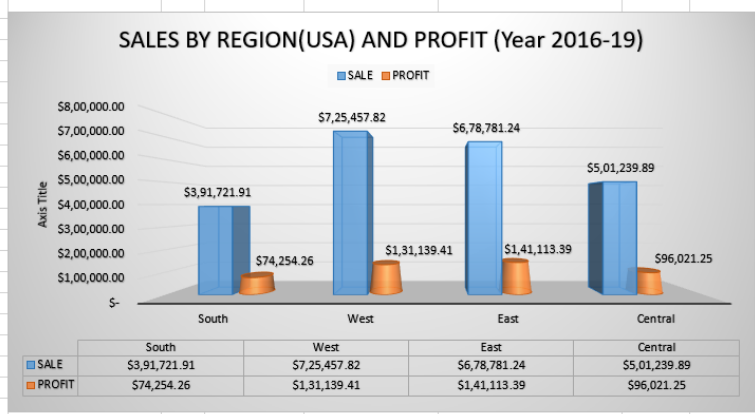
YEAR (2016-2019)

PRICE	SALES	TOTAL AMOUNTS
<=$500	8832	$ 8,22,216.73
>$500 and <= $1000	694	$ 4,84,815.05
>$1000 and <=$2000	328	$ 4,48,142.08
>$2000	140	$ 5,42,027.00

Sales by Price Range (Year 2016-19)
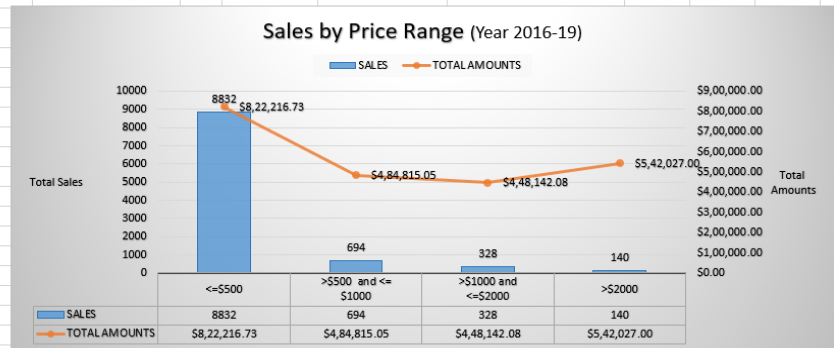
	YEAR	SALE	PROFIT
	2016	$4,84,247.50	$ 81,436.75
	2017	$4,70,532.51	$ 94,147.99
	2018	$6,09,205.60	$ 1,19,668.10
	2019	$7,33,215.26	$ 1,47,275.46

TOTAL SALE AMOUNT AND PROFIT BY YEAR
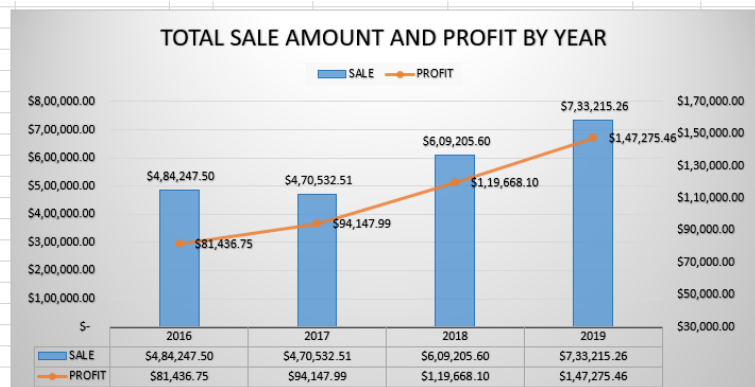
YEAR (2016-2019)

REGION	SALE	PROFIT
South	$3,91,721.91	$ 74,254.26
West	$7,25,457.82	$ 1,31,139.41
East	$6,78,781.24	$ 1,41,113.39
Central	$5,01,239.89	$ 96,021.25

SALES BY REGION(USA) AND PROFIT (Year 2016-19)

PERSON	SALES
Cassandra Brandow (South)	1,620
Anna Andreadi (West)	3,203
Chuck Magee (East)	2,848
Adam (Central)	2,323

SALES BY PERSON

SALES BY PERSON FROM DIFFERENT REGIONS OF USA

| Year	Sales	Returned
2016	1993	151
2017	2102	163
2018	2587	197
2019	3312	289

Total Sales and Returned
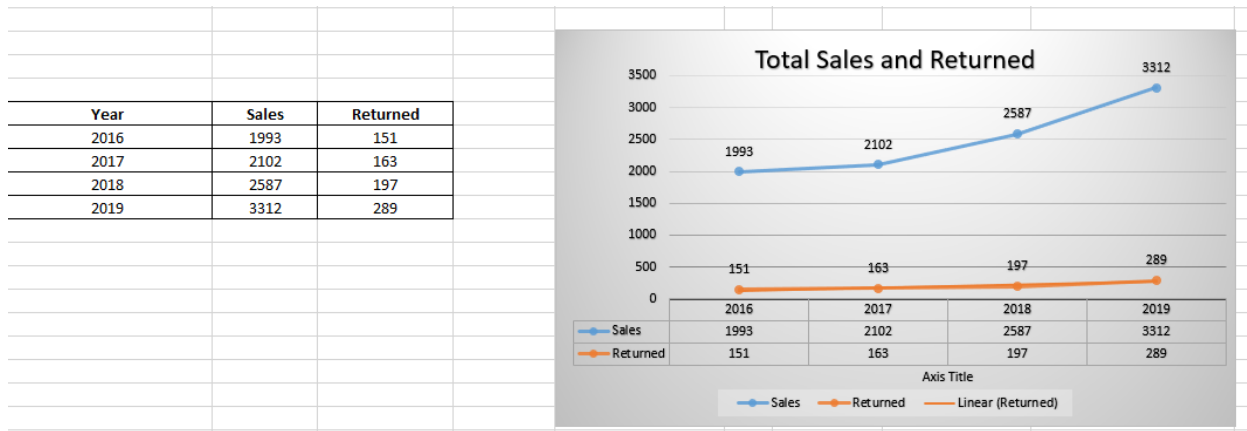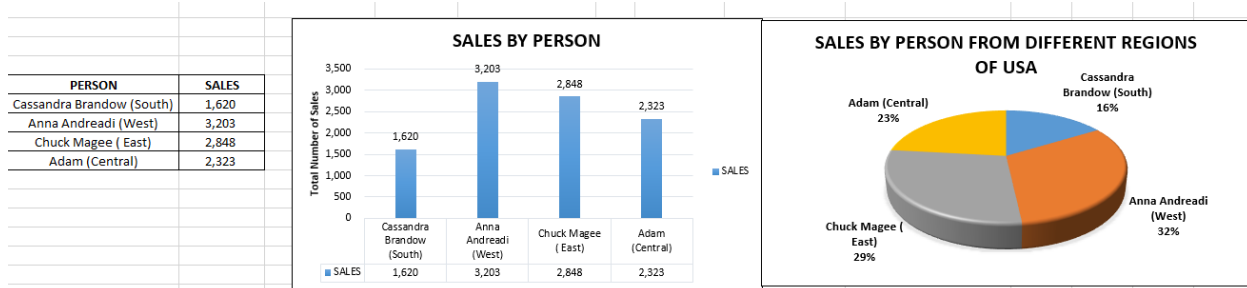
Task 5: Pivot Tables

- Pivot table is a powerful tool to calculate, summarize, and analyze data.
- Select Insert Tab and Pivot table.
- Select the required fields and drop them in Rows, Columns, Values and Filters

PIVOT TABLES

Region and Ship Mode (2016-2019)

Count of Produ Column Labels	First Class	Same Day	Second Class	Standard Class	Grand Total
Row Labels					
⊞ Central	299	120	465	1439	2323
⊞ East	490	155	530	1673	2848
⊞ South	234	83	329	974	1620
⊞ West	515	185	621	1882	3203
Grand Total	1538	543	1945	5968	9994
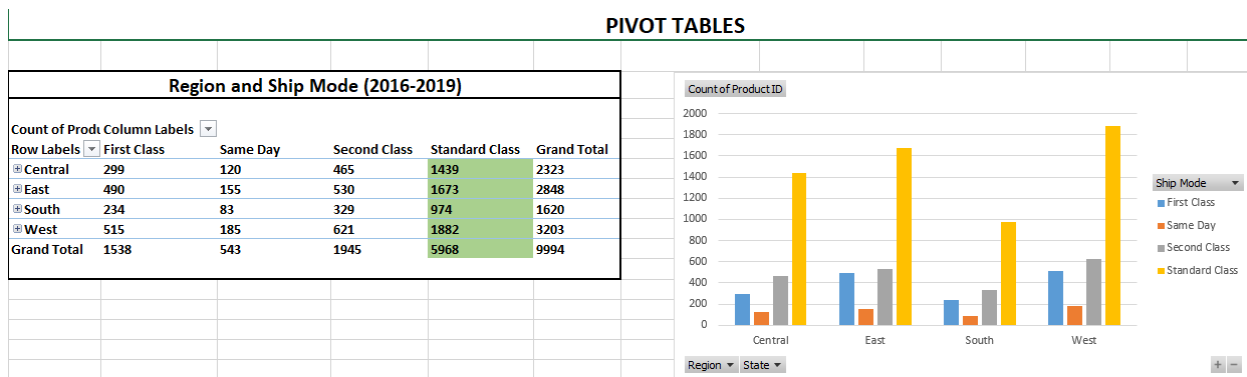
Region and Sales, Profit, Loss (2016-2019)

Row Labels	Sum of Sales	Sum of Profit2	Sum of Loss
Central	$ 5,01,239.89	$ 96,021.25	$ 56,314.89
East	$ 6,78,781.24	$ 1,41,113.39	$ 49,590.61
South	$ 3,91,721.91	$ 74,254.26	$ 27,504.83
West	$ 7,25,457.82	$ 1,31,139.41	$ 22,720.96
Grand Total	$ 22,97,200.86	$ 4,42,528.31	$ 1,56,131.29
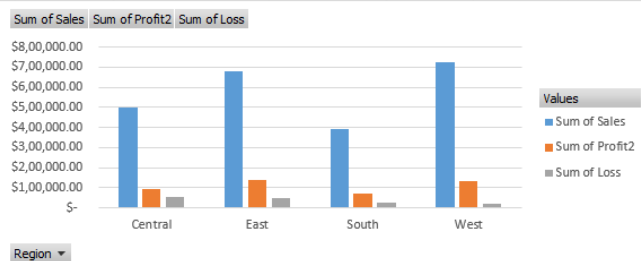
Year, Region and Sales, Profit, Loss

Row Labels	Sum of Sales	Sum of Profit2	Sum of Loss
2016	$ 4,84,247.50	$ 81,436.75	$ 31,892.77
Central	$ 1,03,838.16	$ 15,686.99	$ 15,147.44
East	$ 1,28,680.46	$ 25,457.59	$ 8,397.98
South	$ 1,03,845.84	$ 16,834.77	$ 4,955.65
West	$ 1,47,883.03	$ 23,457.39	$ 3,391.70
2017	$ 4,70,532.51	$ 94,147.99	$ 32,529.39
Central	$ 1,02,874.22	$ 20,728.47	$ 9,011.67
East	$ 1,56,332.06	$ 33,603.19	$ 12,512.18
South	$ 71,359.98	$ 15,438.90	$ 7,120.31
West	$ 1,39,966.25	$ 24,377.43	$ 3,885.23
2018	$ 6,09,205.60	$ 1,19,668.10	$ 37,872.93
Central	$ 1,47,429.38	$ 31,546.28	$ 11,647.12
East	$ 1,80,685.82	$ 36,925.26	$ 16,783.67
South	$ 93,610.22	$ 20,766.50	$ 3,063.69
West	$ 1,87,480.18	$ 30,430.06	$ 6,378.45
2019	$ 7,33,215.26	$ 1,47,275.46	$ 53,836.19
Central	$ 1,47,098.13	$ 28,059.50	$ 20,508.66
East	$ 2,13,082.90	$ 45,127.34	$ 11,896.78
South	$ 1,22,905.86	$ 21,214.09	$ 12,365.18
West	$ 2,50,128.37	$ 52,874.53	$ 9,065.57
Grand Total	$ 22,97,200.86	$ 4,42,528.31	$ 1,56,131.29
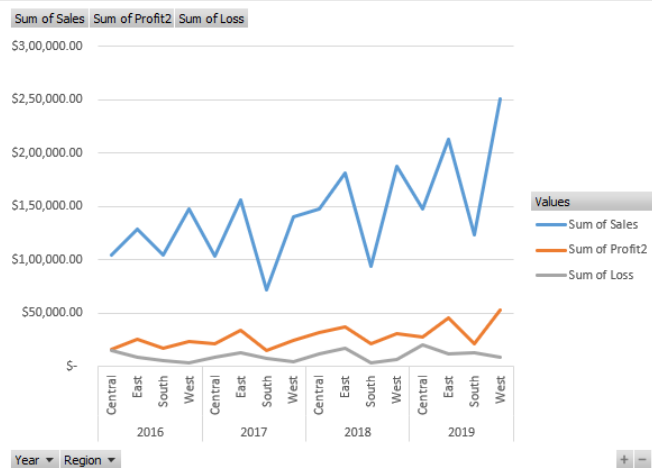
Region and Category Sales (2016-2019)

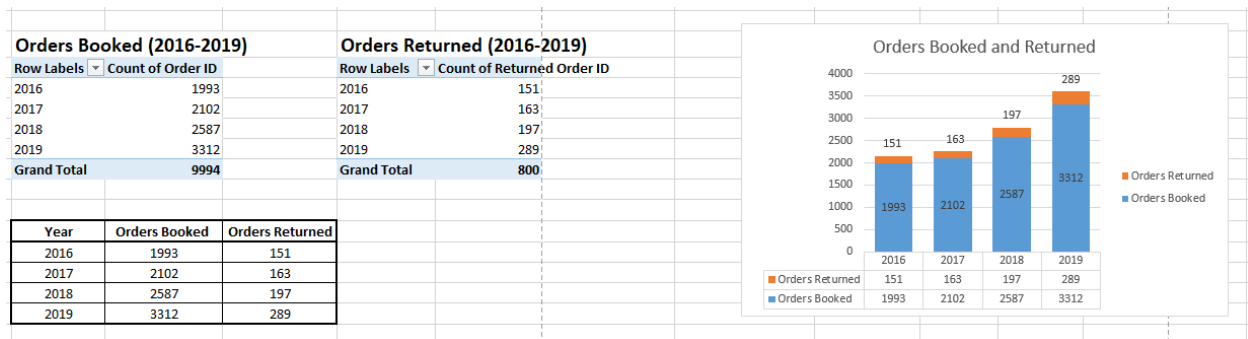| Sum of Sales Row Labels	Furniture	Office Supplies	Technology	Grand Total
Central	$ 1,63,797.16	$ 1,67,026.42	$ 1,70,416.31	$ 5,01,239.89
East	$ 2,08,291.20	$ 2,05,516.06	$ 2,64,973.98	$ 6,78,781.24
South	$ 1,17,298.68	$ 1,25,651.31	$ 1,48,771.91	$ 3,91,721.91
West	$ 2,52,612.74	$ 2,20,853.25	$ 2,51,991.83	$ 7,25,457.82
Grand Total	$ 7,41,999.80	$ 7,19,047.03	$ 8,36,154.03	$ 22,97,200.86
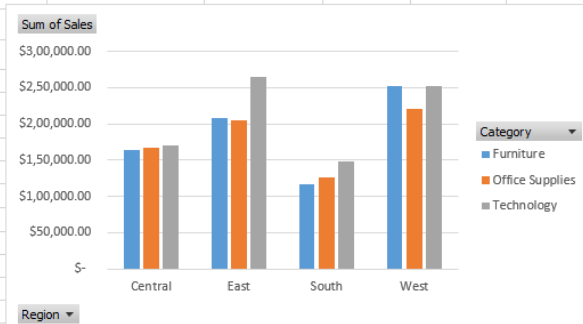
State Sales, Profit (2016-2019)

Row Labels	Sum of Sales	Sum of Profit2	Sum of Loss
Alabama	$ 19,510.64	$ 5,786.83	$ -
Arizona	$ 35,282.00	$ 3,228.84	$ 6,656.77
Arkansas	$ 11,678.13	$ 4,008.69	$ -
California	$ 4,57,687.63	$ 80,151.05	$ 3,769.67
Colorado	$ 32,108.12	$ 2,373.05	$ 8,900.90
Connecticut	$ 13,384.36	$ 3,531.11	$ 19.61
Delaware	$ 27,451.07	$ 10,063.27	$ 85.90
District of Colu	$ 2,865.02	$ 1,059.59	$ -
Florida	$ 89,473.71	$ 5,290.53	$ 8,689.83
Georgia	$ 49,095.84	$ 16,250.04	$ -
Idaho	$ 4,382.49	$ 826.72	$ -
Illinois	$ 80,166.10	$ 6,893.81	$ 19,501.70
Indiana	$ 53,555.36	$ 18,382.94	$ -
Iowa	$ 4,579.76	$ 1,183.81	$ -
Kansas	$ 2,914.31	$ 836.44	$ -
Kentucky	$ 36,591.75	$ 11,199.70	$ -
Louisiana	$ 9,217.03	$ 2,196.10	$ -
Maine	$ 1,270.53	$ 454.49	$ -
Maryland	$ 23,705.52	$ 7,102.30	$ 71.12
Massachusetts	$ 28,634.43	$ 7,351.76	$ 566.26
Michigan	$ 76,269.61	$ 24,463.19	$ -
Minnesota	$ 29,863.15	$ 10,823.19	$ -

State	Sales	Profit	Loss
Mississippi	$ 10,771.34	$ 3,172.98	$ -
Missouri	$ 22,205.15	$ 6,436.21	$ -
Montana	$ 5,589.35	$ 1,833.33	$ -
Nebraska	$ 7,464.93	$ 2,037.09	$ -
Nevada	$ 16,729.10	$ 3,426.35	$ 109.58
New Hampshir	$ 7,292.52	$ 1,811.82	$ 105.32
New Jersey	$ 35,764.31	$ 9,824.20	$ 51.29
New Mexico	$ 4,783.52	$ 1,162.81	$ 5.69
New York	$ 3,10,876.27	$ 79,069.69	$ 5,031.14
North Carolina	$ 55,603.16	$ 4,067.07	$ 11,557.99
North Dakota	$ 919.91	$ 230.15	$ -
Ohio	$ 78,258.14	$ 4,778.62	$ 21,750.00
Oklahoma	$ 19,683.39	$ 4,853.96	$ -
Oregon	$ 17,431.15	$ 1,700.01	$ 2,890.48
Pennsylvania	$ 1,16,511.91	$ 6,042.89	$ 21,602.85
Rhode Island	$ 22,627.96	$ 7,515.80	$ 230.17
South Carolina	$ 8,481.71	$ 1,769.06	$ -
South Dakota	$ 1,315.56	$ 394.83	$ -
Tennessee	$ 30,661.87	$ 1,915.32	$ 7,257.02
Texas	$ 1,70,188.05	$ 11,083.83	$ 36,813.19
Utah	$ 11,220.06	$ 2,546.53	$ -
Vermont	$ 8,929.37	$ 2,244.98	$ -
Virginia	$ 70,636.72	$ 18,597.95	$ -
Washington	$ 1,38,641.27	$ 33,790.52	$ 387.87
West Virginia	$ 1,209.82	$ 262.88	$ 76.95
Wisconsin	$ 32,114.61	$ 8,401.80	$ -
Wyoming	$ 1,603.14	$ 100.20	$ -
Grand Total	$ 22,97,200.86	$ 4,42,528.31	$ 1,56,131.29

V LOOKUP		
	Amount	**State**
Highest Sales	$ 4,57,687.63	California
Maximum Profit	$ 80,151.05	California
Maximum Loss	$ 36,813.19	Texas

Orders Booked (2016-2019)

Row Labels	Count of Order ID
2016	1993
2017	2102
2018	2587
2019	3312
Grand Total	9994

Orders Returned (2016-2019)

Row Labels	Count of Returned Order ID
2016	151
2017	163
2018	197
2019	289
Grand Total	800

Year	Orders Booked	Orders Returned
2016	1993	151
2017	2102	163
2018	2587	197
2019	3312	289

Orders Booked and Returned

Task 6: Tests

- For F-Test, T-Test and Anova Test, we require Null and Alternate Hypothesis

- Select Data Tab and Data Analysis

- Select the Test

- Select the input and output range

- The Tests were performed on Sales, Profit and Loss Columns of the Superstore Dataset

F-TEST (Sales and Profit dataset)				
NULL HYPOTHESIS: Sales and profit dataset have equal variances				
ALTERNATE HYPOTHESIS: Sales and profit dataset have unequal variances				

SALE SAMPLE	PROFIT SAMPLE		F-Test Two-Sample for Variances		
33.11	0.7228				
144.95	29.372			SALE SAMPLE	PROFIT SAMPLE
206.991	3.2472		Mean	120.713875	8.5707375
231.92	4.2309		Variance	7546.452934	90.83626215
56.7	4.6176		Observations	8	8
25.92	1.6704		df	7	7
208.44	11.223		F	83.07753705	
57.68	13.482		P(F<=f) one-tail	3.33922E-06	
			F Critical one-tail	3.78704354	

OBSERVATIONS
1. P=3.33922E-06= 0.00000333922
2.P<Alpha(0.05)
3.Reject Null Hypothesis and accept Alternate
Conclusion: Sales and Profit dataset have unequal variances

T-TEST (Sales, Profit)

NULL HYPOTHESIS: Sales and profit dataset have the same mean and unequal variances

ALTERNATE HYPOTHESIS: Sales and profit dataset have unequal mean and variances

SALE SAMPLE	PROFIT SAMPLE
33.11	0.7228
144.95	29.372
206.991	3.2472
231.92	4.2309
56.7	4.6176
25.92	1.6704
208.44	11.223
57.68	13.482

t-Test: Two-Sample Assuming Unequal Variances

	SALE SAMPLE	PROFIT SAMPLE
Mean	120.713875	8.5707375
Variance	7546.452934	90.83626215
Observations	8	8
Hypothesized Mean Difference	0	
df	7	
t Stat	3.629510674	
P(T<=t) one-tail	0.004201652	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.008403303	
t Critical two-tail	2.364624252	

OBSERVATIONS

1. P=0.008403303
2. P<Alpha(0.05)
3. Reject Null Hypothesis and accept Alternate
Conclusion: Sales and profit dataset have unequal mean and variances

ANNOVA TEST 1 (Sale, Profit, Loss)

NULL HYPOTHESIS: Sales, profit, loss dataset have Equal means

ALTERNATE HYPOTHESIS: Sales, profit, loss dataset have unequal means

SALE SAMPLE	PROFIT SAMPLE	LOSS SAMPLE
33.11	0.7228	20.889
144.95	29.372	0.3488
206.991	3.2472	34.38
231.92	4.2309	4.7976
56.7	4.6176	23.976
25.92	1.6704	3.6576
208.44	11.223	3.2175
57.68	13.482	5.1968

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
SALE SAMPLE	8	965.711	120.7139	7546.453
PROFIT SAMPLE	8	68.5659	8.570738	90.83626
LOSS SAMPLE	8	96.4633	12.05791	157.7019

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	65051.63164	2	32525.82	12.51797	0.000264	3.4668
Within Groups	54564.93734	21	2598.33			
Total	119616.569	23				

OBSERVATIONS

1. P=0.000264
2. P<Alpha(0.05)
3. Reject Null Hypothesis and accept Alternate
Conclusion: Sales, profit, loss dataset have unequal means

F-TEST (Profit, Loss)

NULL HYPOTHESIS: Profit and loss dataset have equal variances

ALTERNATE HYPOTHESIS: Profit and loss dataset have unequal variances

PROFIT SAMPLE	LOSS SAMPLE
0.7228	20.889
29.372	0.3488
3.2472	34.38
4.2309	4.7976
4.6176	23.976
1.6704	3.6576
11.223	3.2175
13.482	5.1968

F-Test Two-Sample for Variances		
	PROFIT SAMPLE	LOSS SAMPLE
Mean	8.5707375	12.0579125
Variance	90.83626215	157.7018525
Observations	8	8
df	7	7
F	0.575999969	
P(F<=f) one-tail	0.241957985	
F Critical one-tail	0.264058226	

OBSERVATIONS
1. P=0.241957985
2.P>Alpha(0.05)
3.Accept Null Hypothesis
Conclusion: Profit, loss dataset have equal variances

Summary

From the above Analysis, we can conclude the following.

- Standard class is the most common ship mode in all the regions.
- Year: (2016-2019)

    West Region has the highest Sales of **$7,25,457.82**

    South region has the lowest sales of **$3,91,721.91**

    East region has the highest profits of **$1,41,113.39**

    Central region has the highest loss of **$56,314.89**

    California State has made the highest sales of **$4,57,687.63** and a maximum profit of **$80,151.05**

    Texas has the highest loss of **$36,813.19**

- The Sales and profits have increased.
- From the Moving average graph, we can also predict an increase in sales and profit.
- Orders Returned are significantly smaller when compared to the orders booked.

References

I.   https://www.kaggle.com/

II. https://www.google.com/