Abnormal Distribution

Yueru Zhang

Skyla Li

Oindrila Mandal

Brian Diller

Ben Botirov

Katherine Kouot

Marketing Analytics: Increasing Sales and Profit for GloBiz

Team 5 – Abnormal Distribution

Bekhruzbek Botirov, Brian Diller
Katherine Kouot, Skyla Li
Oindrila Mandal, Yueru Zhang

Overview and Data Summary

Data Source: Kaggle

Overview: The marketing team of a large company named GloBiz is exploring ways to grow their sales and profit margin. Management has contracted an analytics firm, Abnormal Distribution, to understand what factors they should focus on.

Outcome Variable

- Sales and Profit margin: indicate sales/profit margin by region/product

Data Summary

- Total # of Records: 51,290
- Market Information: Markets, Region, Country
- Product Information: Product Categories, Sub-Category, Product ID
- Purchase Information: Customer ID, Order ID, Order date, Order Priorities
- Sales Information: Profit, Sales, Quantity, Shipping Cost, Discount

Data Limitations

- Only has data from 2011-2014
- Ignores the influence of COVID-19
- Unsure how data was collected or calculated
- Lacking definitions of the variables and the variable interactions

Problem Statement

Problem Statement

GloBiz is unaware of factors that impact their firm's market performance.

Criteria for Success

Identify key variables influencing GloBiz sales and profit margin.

Identify most successful products and regions for increased marketing activity

Identify least successful products and regions for possible deprioritization

Executive Summary

Key Insights

1. **Certain countries and regions have very high profit margins**

2. **Profit margins were very uneven between product categories**

3. **Some product categories have high sales order values while other categories have smaller sales order values**

4. **Surprisingly, there was no correlation between sales and discount**

Our Approach

Initial Data Analysis:

- Analyzed the data set
- Checked for missing information
- Cleaned data set

Identifying Outcome Variables:

- Outcome variables that matter to marketers: Sales and Profit Margin

Analysis Plan:

- K-Means Cluster Analysis of the Marketing Mix – clustering based on regions and product lines
- Step-wise Multiple Regression on all numerical and categorical variables to identify significant ones

Identifying Tools:

- Tableau and Radiant (R-Studio).

Key Insights:

- Identify strong and weak products and regions
- Consider effect of previous promotional activity
- Develop initial conclusions

Overview of the Analysis

Need for this analysis

With goal to grow, need understand impacting factors
Company has ample data but needs meaningful, actionable conclusions
Who is their consumer and what are sales patterns?

How analytics helps solve this problem

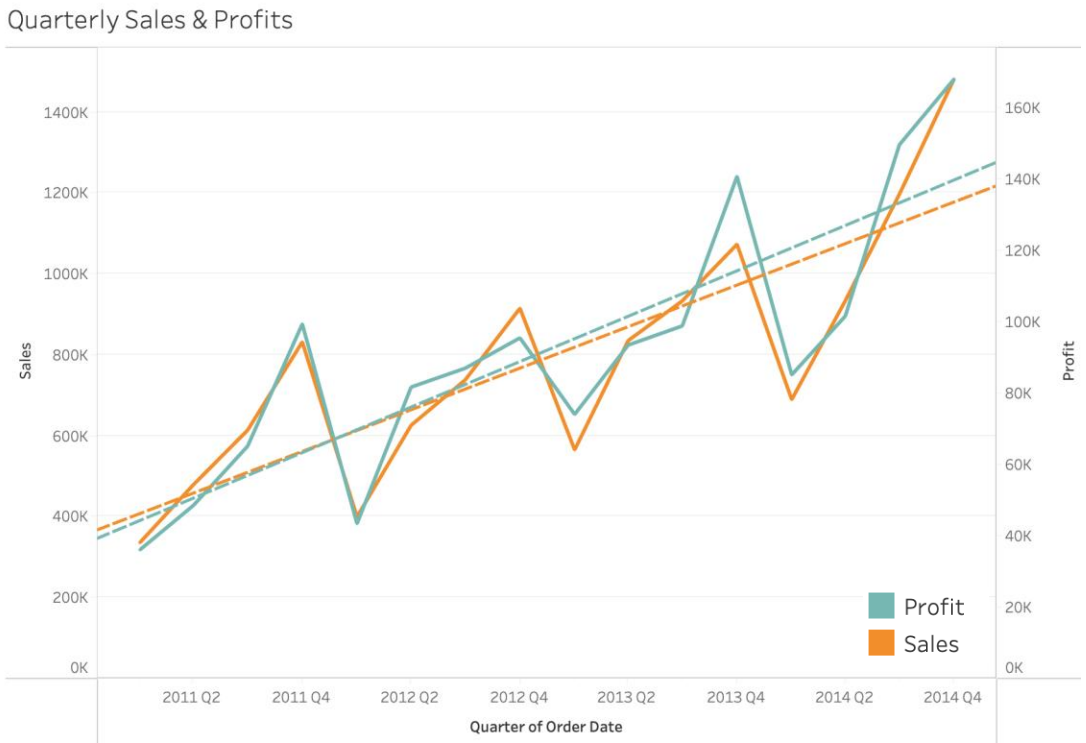
Identify key variables in as much detail as possible and understand their interactions
Determine profitability of specific regions and products
Understand characteristics of market segments to more efficiently target them
Evaluate past promotional activity
Synthesize findings into actionable changes to Globiz strategy

Profitability Analysis

Company Level Analysis

- **Upward trends**
- **Seasonality of sales and profits**

- **Don’t see correlation between discount and order quantity**



Profitability for Categories

Product Categories

- Furniture has the lowest margin
- Sales are evenly distributed across categories

Category	% of Sales	Profit Margin
Furniture	32.52%	7%
Office Supplies	29.96%	14%
Technology	37.53%	14%

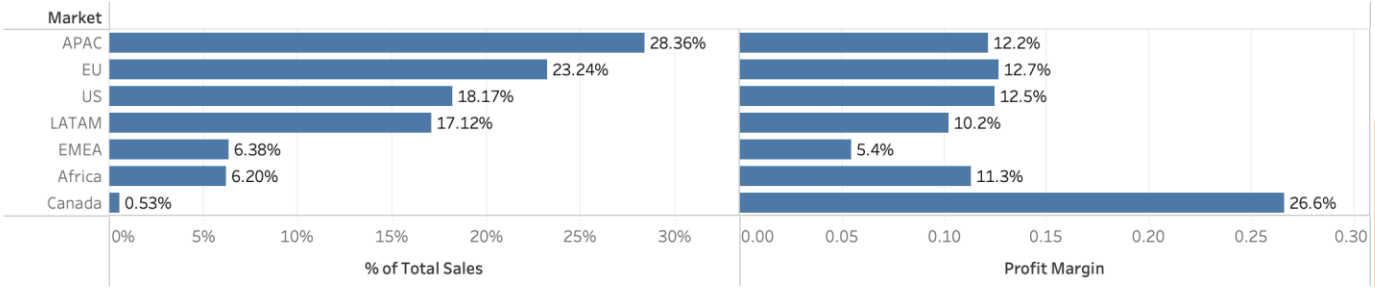
Product Sub-Categories

- Tables have the lowest margin
- Top 6 sales sub-categories: Bookcases, Chairs, Appliances, Storage, Copiers and Phones. Accounts for 66% of total sales
- Paper and Labels are highly profitable, but only account for less than 3% of total sales

Category	Sub-Catego..	% of Sales	Profit Margin
Furniture	Bookcases	11.60%	11%
	Chairs	11.88%	9%
	Furnishings	3.05%	12%
	Tables	5.99%	-8%
Office Supplies	Appliances	8.00%	14%
	Art	2.94%	16%
	Binders	3.65%	16%
	Envelopes	1.35%	17%
	Fasteners	0.66%	14%
	Labels	0.58%	20%
	Paper	1.93%	24%
	Storage	8.92%	10%
	Supplies	1.92%	9%
Technology	Accessories	5.93%	17%
	Copiers	11.94%	17%
	Machines	6.16%	8%
	Phones	13.50%	13%

Profitability for Markets & Product Categories

Percentage of Sales in Each Market & Net Profit Margin in Each Market



Product Categories

	Market													
	APAC		EU		US		LATAM		EMEA		Africa		Canada	
Category	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin
Furniture	37.48%	9.4%	26.52%	8.8%	32.30%	2.5%	37.51%	5.1%	28.36%	5.0%	24.83%	8.4%	15.83%	24.7%
Office Supplies	24.66%	11.9%	35.61%	15.3%	31.30%	17.0%	26.05%	14.1%	34.32%	5.4%	34.03%	10.7%	44.88%	26.5%
Technology	37.86%	15.1%	37.87%	13.0%	36.40%	17.4%	36.44%	12.8%	37.32%	5.8%	41.13%	13.7%	39.29%	27.6%

		Market													
		APAC		EU		US		LATAM		EMEA		Africa		Canada	
Category	Sub-Catego..	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin	% of Sales	Profit Margin
Furniture	Bookcases	14.08%	13.4%	12.37%	15.5%	5.00%	-3.0%	13.97%	8.2%	11.38%	8.7%	10.65%	8.6%	8.57%	23.4%
	Chairs	14.31%	12.1%	7.79%	8.6%	14.30%	8.1%	13.96%	9.5%	8.66%	-0.9%	7.15%	5.0%	4.78%	26.8%
	Furnishings	2.82%	16.0%	2.76%	16.8%	3.99%	14.2%	2.89%	0.4%	3.44%	5.2%	2.63%	11.2%	1.21%	14.1%
Lowest margin	Tables	6.28%	-8.9%	3.59%	-19.9%	9.01%	-8.6%	6.69%	-8.5%	4.88%	7.0%	4.41%	11.6%	1.27%	35.3%
Office Supplies	Appliances	8.58%	13.7%	9.37%	16.8%	4.68%	16.9%	8.41%	14.4%	8.47%	4.4%	7.94%	5.9%	11.86%	28.1%
	Art	1.76%	11.6%	5.45%	19.2%	1.18%	24.1%	1.90%	16.9%	4.73%	3.8%	4.90%	10.3%	6.16%	22.2%
	Binders	1.77%	17.0%	3.40%	19.8%	8.85%	14.9%	1.99%	12.3%	3.31%	10.9%	2.85%	11.9%	4.62%	25.4%
Low sales High margin	Envelopes	1.45%	10.0%	1.37%	21.5%	0.72%	42.3%	1.91%	15.2%	1.31%	7.7%	1.24%	15.7%	0.85%	30.2%
	Fasteners	0.78%	6.0%	0.69%	21.8%	0.13%	31.4%	0.88%	13.2%	0.79%	14.9%	0.74%	14.7%	0.71%	29.3%
	Labels	0.62%	11.1%	0.53%	21.0%	0.54%	44.3%	0.63%	17.6%	0.51%	9.5%	0.62%	16.1%	0.63%	30.6%
	Paper	1.67%	11.5%	1.45%	20.9%	3.42%	43.4%	1.76%	15.9%	1.27%	8.7%	1.72%	15.3%	2.27%	24.6%
	Storage	6.03%	11.5%	11.56%	8.2%	9.74%	9.5%	6.56%	11.4%	12.18%	3.5%	12.33%	12.3%	15.82%	27.5%
	Supplies	2.00%	5.9%	1.80%	18.2%	2.03%	-2.5%	2.00%	17.5%	1.76%	7.0%	1.69%	7.8%	1.96%	22.7%
Technology	Accessories	5.19%	8.5%	5.55%	20.5%	7.29%	25.1%	6.55%	19.1%	5.48%	8.1%	5.38%	15.4%	6.68%	29.0%
High sales High margin	Copiers	13.79%	16.3%	12.43%	15.4%	6.51%	37.2%	14.61%	13.0%	9.94%	10.2%	12.28%	14.6%	11.16%	35.7%
	Machines	5.31%	13.9%	7.60%	7.7%	8.24%	1.8%	1.89%	5.8%	7.69%	4.4%	8.82%	8.6%	6.39%	14.2%
	Phones	13.56%	16.7%	12.30%	10.3%	14.37%	13.5%	13.38%	10.4%	14.21%	2.6%	14.65%	15.4%	15.07%	26.6%

- The top 4 markets account for 87% of total sales; profit margins are 10%~13%
- EMEA has the lowest profit margin ~5%
- Canada has the highest profit margin ~27%, but only accounts for 0.5% of total sales

Product Categories

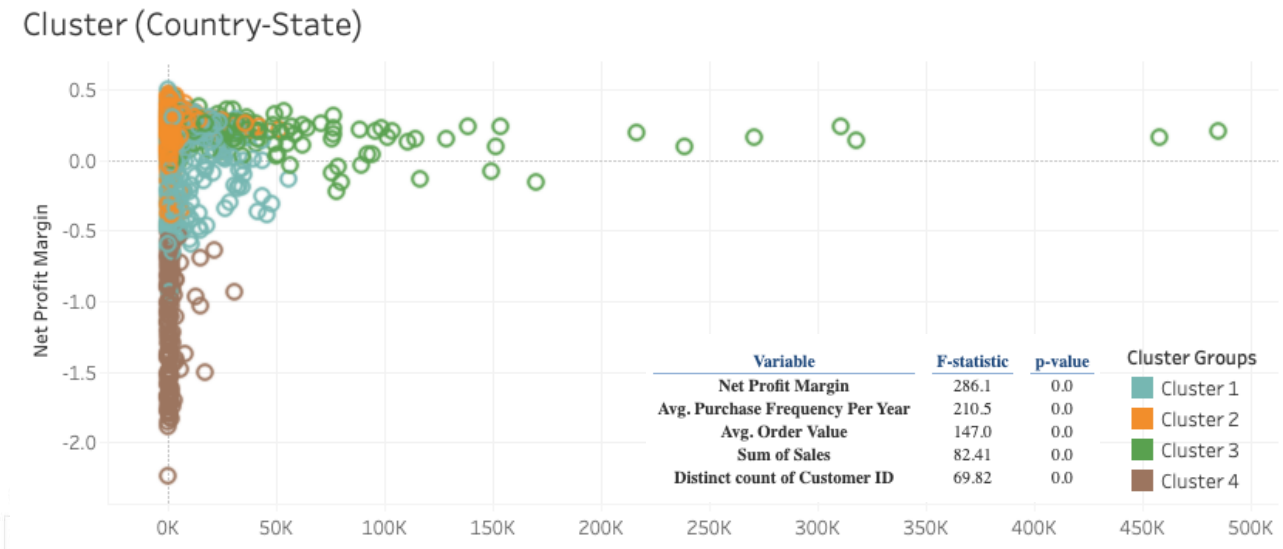
- % sales for each category is slightly different in each market; no category has a major advantage in sales
- Furniture has the lowest margin in each market. Especially in the EU, US and LATAM, the margin of furniture is significantly lower than that of other categories

Product Sub-Categories

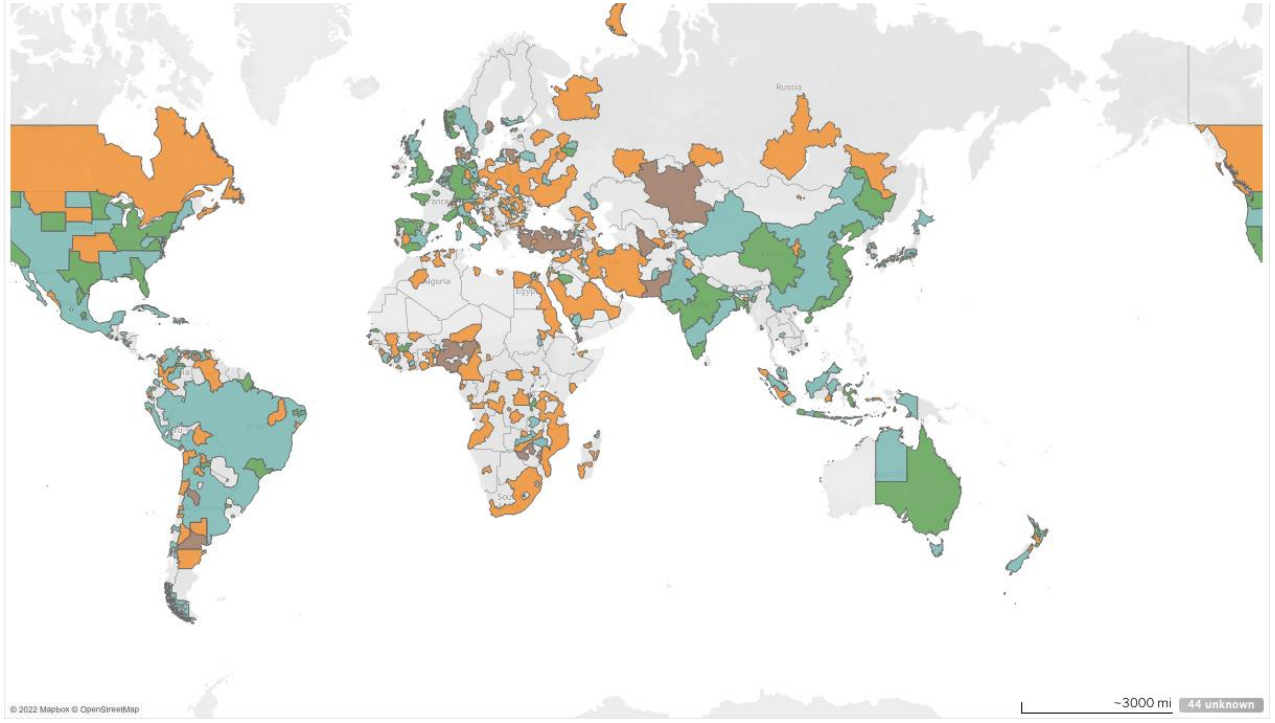
- Tables have the lowest margin in the top 4 m arkets
- Majority of office supplies have low sales but extremely high margins [area for growth
- Copiers and phones have both high sal and high margins

Cluster Analysis

K-Means Clustering Analysis by Country-State



Variable	F-statistic	p-value
Net Profit Margin	286.1	0.0
Avg. Purchase Frequency Per Year	210.5	0.0
Avg. Order Value	147.0	0.0
Sum of Sales	82.41	0.0
Distinct count of Customer ID	69.82	0.0



Variables: Sales, Profit margin, # of unique customers, Customer purchase frequency, Order value

	Centers					
Clusters	Number of Items	Sum of Sales	Net Profit Margin	Distinct count of Customer ID	Avg. Purchase Frequency Per Year	Avg. Order Value
Cluster 1	390	8531.7	0.05796	18.856	2.9348	429.91
Cluster 2	441	3695.0	0.20738	9.1769	1.7087	386.84
Cluster 3	152	48991.0	0.19549	69.625	2.639	1034.2
Cluster 4	136	1760.5	-1.1398	10.147	1.8862	174.76
Not Clustered	0					

Cluster Characteristics				
	Sales	Profit Margin	Customer Size	Purchase Frequency
Cluster 1	Medium	Medium	Medium	High
Cluster 2	Low	High	Small	Low
Cluster 3	High	High	High	High
Cluster 4	Low	Low	Small	Low

Strategy	
Cluster 1	Increase customer base & improve profitability
Cluster 2	Increase customer base & improve purchase frequency
Cluster 3	Top Performance Regions
Cluster 4	Consider to shut down

Multiple Regression Analysis

Multiple Regression Analysis Summary

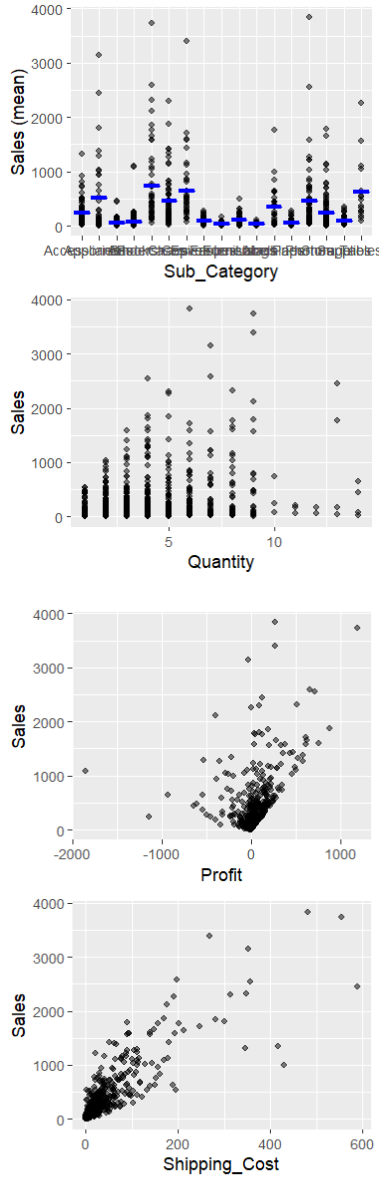
```
Linear regression (OLS)
Data     : Dataset
Response variable    : Sales
Explanatory variables: Sub_Category, Profit, Quantity, Shipping_Cost
Null hyp.: the effect of x on Sales is zero
Alt. hyp.: the effect of x on Sales is not zero
```

	coefficient	std.error	t.value	p.value	
(Intercept)	10.560	5.285	1.998	0.046	*
Sub_Category|Appliances	154.216	8.105	19.028	< .001	***
Sub_Category|Art	-53.460	6.219	-8.597	< .001	***
Sub_Category|Binders	-53.897	5.972	-9.025	< .001	***
Sub_Category|Bookcases	187.421	7.399	25.332	< .001	***
Sub_Category|Chairs	102.586	6.709	15.292	< .001	***
Sub_Category|Copiers	188.556	7.597	24.819	< .001	***
Sub_Category|Envelopes	-58.073	7.322	-7.931	< .001	***
Sub_Category|Fasteners	-70.991	7.345	-9.666	< .001	***
Sub_Category|Furnishings	-34.277	6.826	-5.021	< .001	***
Sub_Category|Labels	-77.632	7.200	-10.782	< .001	***
Sub_Category|Machines	176.716	8.543	20.685	< .001	***
Sub_Category|Paper	-66.594	6.659	-10.001	< .001	***
Sub_Category|Phones	126.999	6.761	18.783	< .001	***
Sub_Category|Storage	16.850	6.160	2.735	0.006	**
Sub_Category|Supplies	-39.200	7.327	-5.350	< .001	***
Sub_Category|Tables	442.295	10.638	41.577	< .001	***
Profit	0.994	0.017	57.457	< .001	***
Quantity	21.332	0.593	35.983	< .001	***
Shipping_Cost	3.036	0.056	54.542	< .001	***
Profit:Quantity	-0.032	0.003	-12.654	< .001	***
Profit:Shipping_Cost	-0.000	0.000	-10.024	< .001	***
Quantity:Shipping_Cost	0.297	0.008	36.024	< .001	***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-squared: 0.695,  Adjusted R-squared: 0.695
F-statistic: 5315.196 df(22,51267), p.value < .001
Nr obs: 51,290

Prediction error (RMSE):  269.175
Residual st.dev   (RSD):  269.235
```

Note: Screenshots are Radiant output using R



- **3 rounds** of multiple regression with Sales as the outcome variable
- **Significant Variables:** Sub-category, Profit, Quantity, Shipping Cost
- **Interaction Terms:** Model better explained by (higher R-squared) interaction terms between Profit, Quantity and Shipping cost
- **Scatter plots:** Shipping Cost and Profit may not have a very linear relationship with sales
- **Limitations:** Region was not statistically significant and did not improve adjusted R-square. So, Region was excluded from the model. We could not regress Country variable (as in our tableau analysis) as that field was character.

Multiple Regression Marketing Implications

Regression Equation

Sales =

Intercept	10
Sub-categories that deliver higher sales	+ 154 * Appliances + 187 * Bookcases + 102 * Chairs + 188* Copiers + 177 * Machines + 127 * Phones + 442 * Tables + 17 * Storage
Sub-categories that lead to lower sales	– 53 * Art – 54* Binders – 58 * Envelopes – 71 * Fasteners – 34 * Furnishings – 78 * Labels – 66 * Paper – 39 * Supplies
Marketing Mix	+ 0.994 * Profit + 21* Quantity + 3 * Shipping Cost
Interactions in Marketing Mix	– 0.032 * Profit * Quantity + 0.297 * Quantity * Shipping Cost

- Product lines (sub categories) that impact sales
  - Appliances, book-cases, chairs, copiers, phones, tables and storage drive up sales
  - Art, binders, envelopes, fasteners, furnishings, labels, paper and supplies drive down sales.
- Marketing Mix variables that impact sales
  - Profit – Increases with sales but reduces sales if quantity is large
  - Quantity – Increases with sales
  - Shipping Cost – Increases with sales and quantity

Limitations of Data and Analysis

- Data
  - How is "profit" variable calculated?
  - Is shipping cost deducted from "profit"?
- Discount variable
  - What kinds of promotions were included?
  - Applied over discrete time or regionally?
  - Difficult to analyze. Need more specific data.
- Complex interactions between variables
  - Especially profit, quantity, and shipping cost
- Other factors
  - Global and regional economic conditions
  - Competitive landscape

Summary of Analysis: Marketing Insights

- Regression analysis revealed sales in each order are impacted by the following parameters
  - Sub-category
  - Profit
  - Quantity
  - Shipping Cost
- Overall sales are impacted by below parameters
  - Country
  - Quarter (Seasonality)
  - Sales Mix/ Product Mix
- There is no statistically significant impact of below parameters on sales
  - Region
  - Sub-region
  - Discount
  - Order Priority
- The K-means clusters implied that the company would improve the overall sales by addressing the following factors in each cluster:
  - Profit Margin
  - Customer Size
  - Purchase Frequency

Areas for further research

- Lean into high-margin countries in all regions
  - Canada (highest profit margin and tiny market penetration)
  - EU, US (large markets with good margins, not saturated), but by country
  - Possibly not APAC—high margins but already saturated
- Cluster analysis by product and customer
- Within subcategories
  - Push office supplies because high margins but low sales. Area we can easily increase revenues. Copies and phones already selling well: maintain
- Consider effect of "discount" variable on sales
- Develop optimization strategy and recommendations

	Strategy
Cluster 1	Increase customer base & improve profitability
Cluster 2	Increase customer base & improve purchase frequency
Cluster 3	Top Performance Regions
Cluster 4	Consider to shut down

Thank you!
Questions?

Appendix

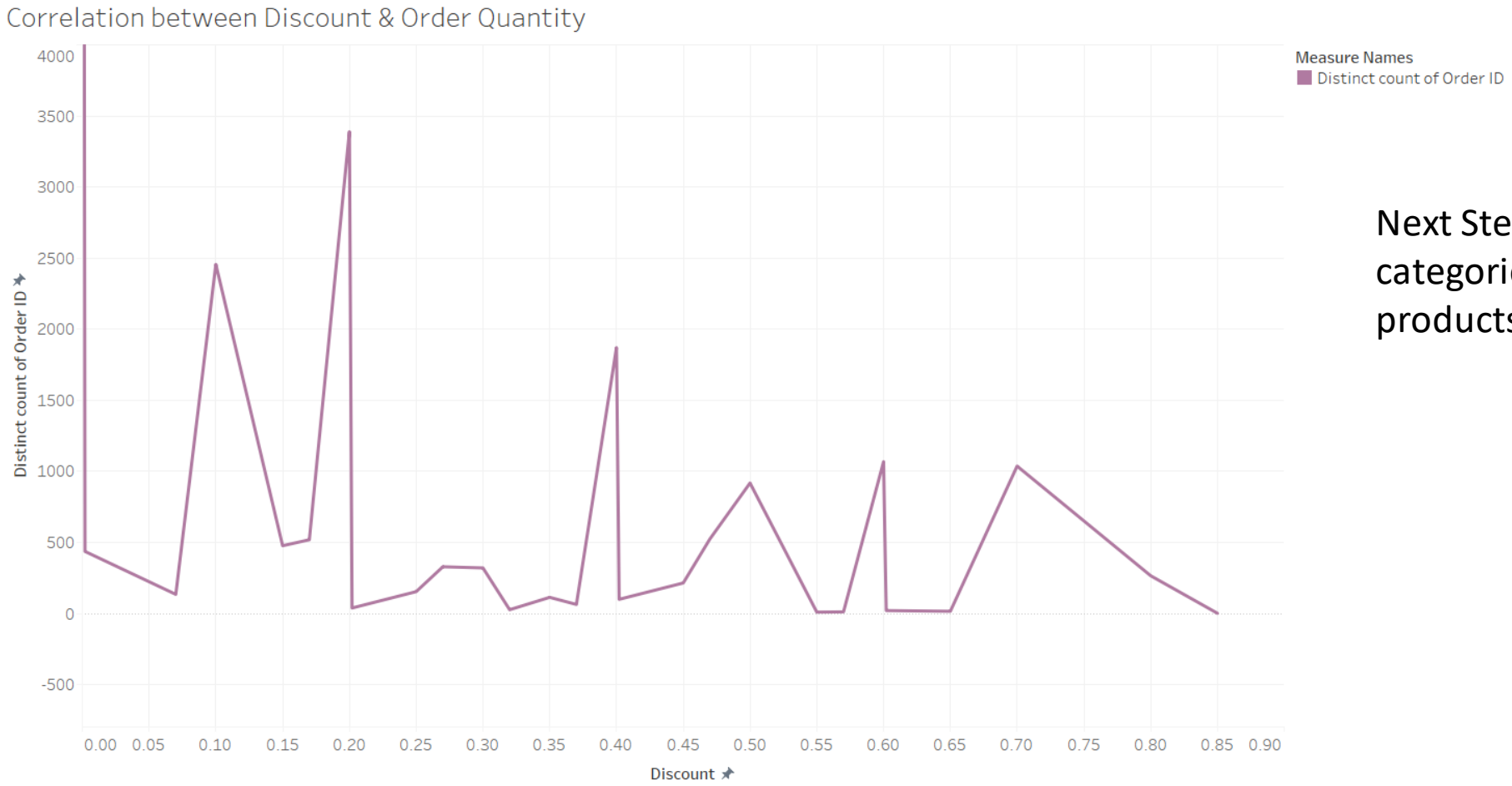
Country	Customer ID	Customer Name	Market	Order Date	Order ID	Order Priority	Product ID	Product Name	Region	Row ID	Segment	Ship Date	Ship Mode	State	Sub-Category	Discount	Number of Records	Profit	Quantity	Sales	Shipping Cost
Hungary	AT-7352	Annie Thurman	EMEA	1/1/2011	HU-2011-1220	High	OFF-TEN-10001585	Tenex Box, Single Width	EMEA	48883	Consumer	5/1/2011	Second Class	Budapest	Storage	0	1	29.64	4	66	8.17
Sweden	EM-141402	Eugene Moren	EU	1/1/2011	IT-2011-3647632	High	OFF-PA-10001492	Enermax Note Cards, Premium	North	11731	Home Office	5/1/2011	Second Class	Stockholm	Paper	0.5	1	-26.055	3	45	4.82
Algeria	TB-112801	Toby Braunhardt	Africa	1/1/2011	AG-2011-2040	Medium	OFF-TEN-10000025	Tenex Lockers, Blue	Africa	42433	Consumer	6/1/2011	Standard Class	Constantine	Storage	0	1	106.14	2	408	35.46
Australia	JH-159851	Joseph Holt	APAC	1/1/2011	IN-2011-47883	Medium	FUR-FU-10003447	Eldon Light Bulb, Duo Pack	Oceania	22255	Consumer	8/1/2011	Standard Class	New South Wales	Furnishings	0.1	1	37.77	5	114	4.7
Australia	JH-159851	Joseph Holt	APAC	1/1/2011	IN-2011-47883	Medium	OFF-PA-10001968	Eaton Computer Printout Paper, 8.5 x 11	Oceania	22254	Consumer	8/1/2011	Standard Class	New South Wales	Paper	0.1	1	15.342	2	55	1.8
Australia	JH-159851	Joseph Holt	APAC	1/1/2011	IN-2011-47883	Medium	OFF-SU-10000618	Acme Trimmer, High Speed	Oceania	22253	Consumer	8/1/2011	Standard Class	New South Wales	Supplies	0.1	1	36.036	3	120	9.72
Canada	MM-72602	Magdelene Morse	Canada	2/1/2011	CA-2011-1510	Medium	TEC-OKI-10002750	Okidata Inkjet, Wireless	Canada	49550	Consumer	6/1/2011	Standard Class	Ontario	Machines	0	1	3.12	1	314	24.1
Australia	KN-164501	Kean Nguyen	APAC	3/1/2011	IN-2011-79397	Critical	OFF-AP-10000304	Hoover Microwave, White	Oceania	20303	Corporate	3/1/2011	Same Day	New South Wales	Appliances	0.1	1	110.412	1	276	125.32
Iraq	LW-69902	Lindsay Williams	EMEA	3/1/2011	IZ-2011-4680	High	FUR-NOV-10002791	Novimex Swivel Stool, Set of Two	EMEA	46681	Corporate	7/1/2011	Standard Class	Ninawa	Chairs	0	1	253.32	4	667	81.26

Sample of the Data Set

- Data Source: Kaggle

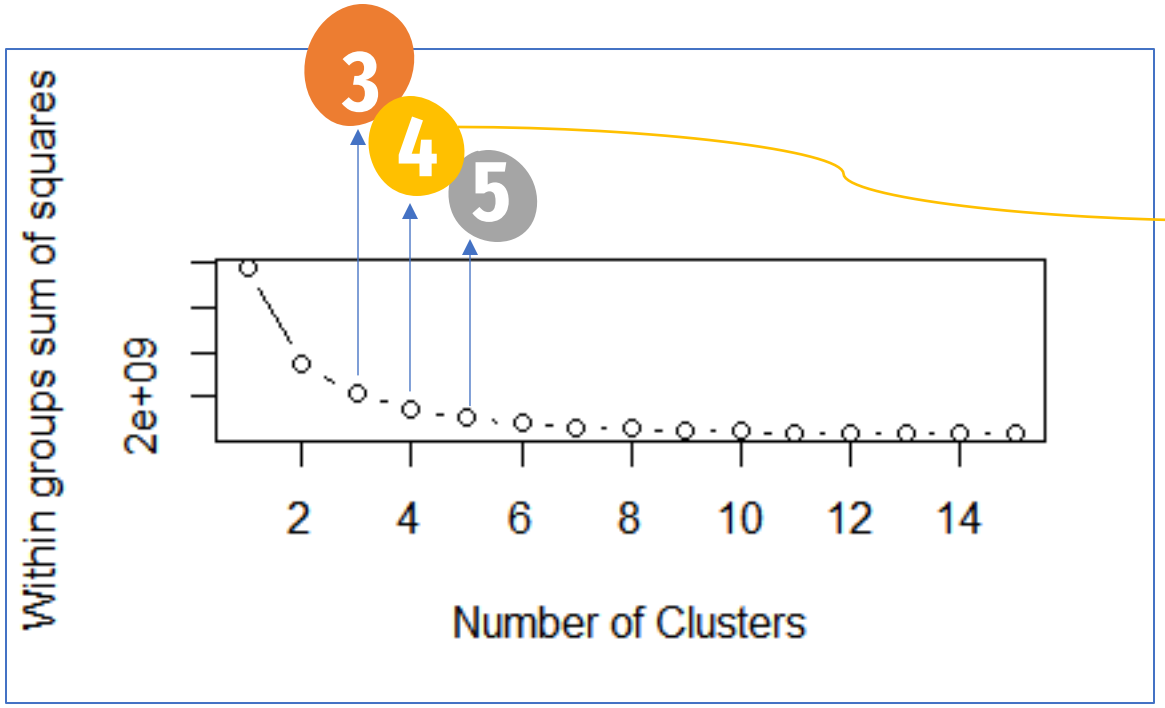
No correlation between Discount and Order Quantity

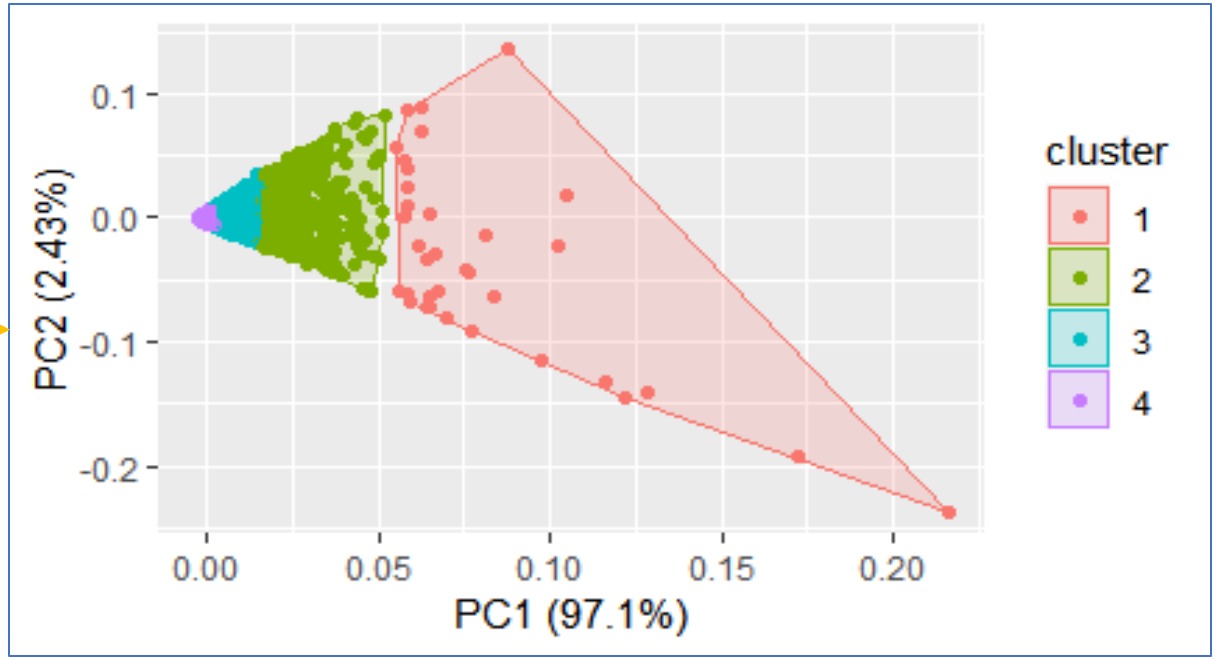
Appendix Slide

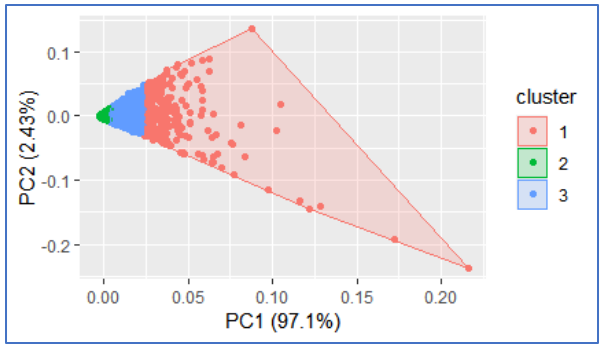


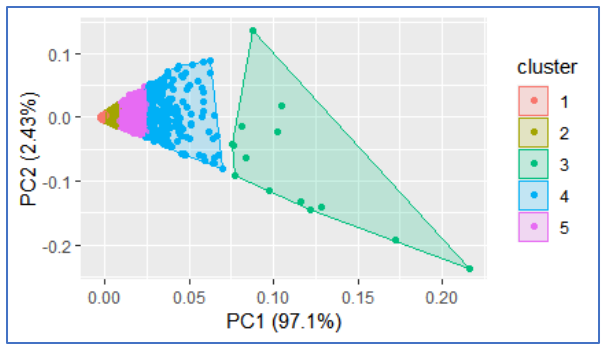
Next Step: Looking into product categories, sub-categories, products that offer high discounts

Alternate cluster analysis – Part 1 (Elbow method)









Alternate cluster analysis – Part 1 (Clusters)

2.1%

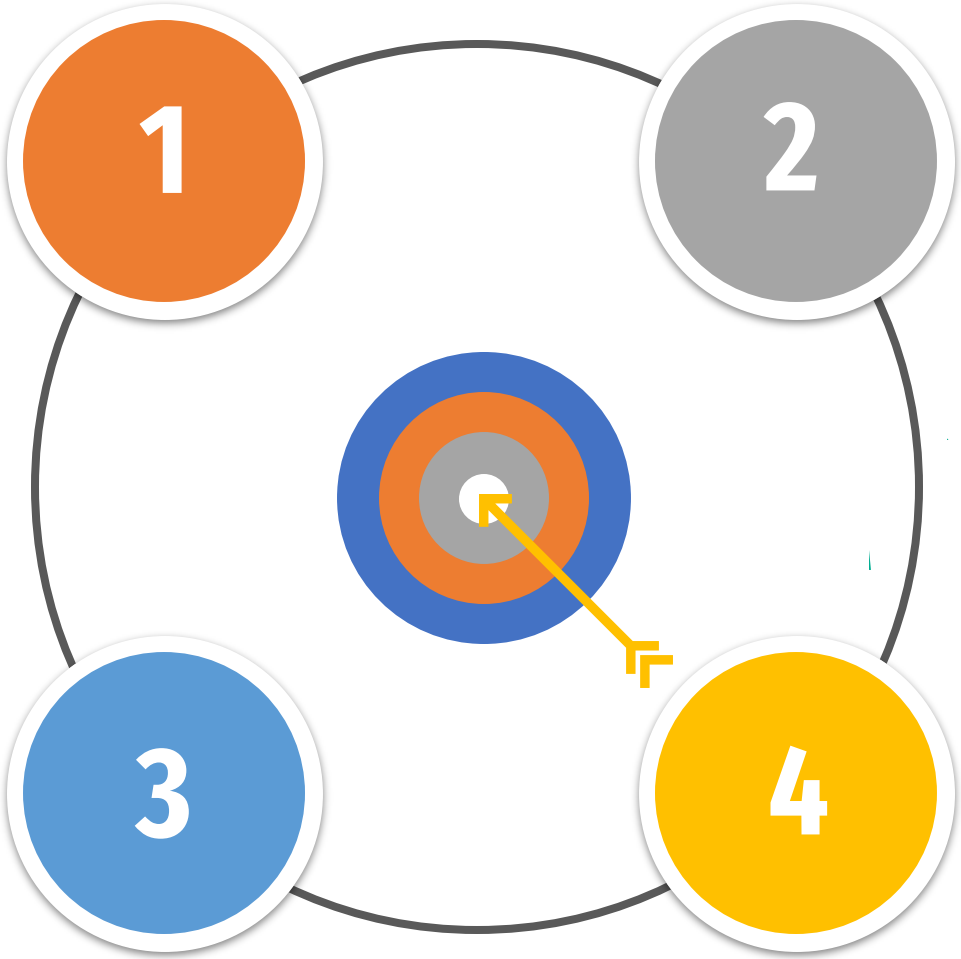
Profit	Quantity	Sales	Shipping cost
594.1	6.2	2313.5	243.9

Products: Technology, Furniture
Region: Central, Oceania, South
Segment: Consumer, Corporate
Market: APAC, EU, US

0.1%

Profit	Quantity	Sales	Shipping cost
2509.7	7.8	6880.7	404.8

Products: Technology, Office
Region: East, Central, South
Segment: Consumer, Corporate
Market: US, APAC, EU



86.0%

Profit	Quantity	Sales	Shipping cost
26.5	3.2	105.6	11.5

Products: Office supplies
Region: Central, South, EMEA
Segment: Consumer, Corporate
Market: LATAM, APAC, US

11.7%

Profit	Quantity	Sales	Shipping cost
192.9	4.6	794.6	87.4

Products: Technology, Furniture
Region: Central, South, North
Segment: Consumer, Corporate
Market: APAC, EU, LATAM

Multiple Regression – Part 1

```
Start:  AIC=572994.2
Sales ~ Category + Order_Priority + Region + Segment + Sub_Category +
    Discount + Profit + Quantity + Shipping_Cost


Step:  AIC=572994.2
Sales ~ Order_Priority + Region + Segment + Sub_Category + Discount +
    Profit + Quantity + Shipping_Cost

                Df  Sum of Sq        RSS    AIC
- Segment        2      74875 3640708576 572991
<none>                       3640633701 572994
- Region        12    5352674 3645986374 573046
- Discount       1    9685262 3650318962 573128
- Order_Priority 3  157469073 3798102773 575160
- Quantity       1  168008992 3808642692 575306
- Sub_Category  16  344024910 3984658610 577593
- Profit         1  614713464 4255347165 580994
- Shipping_Cost  1 2663929461 6304563162 601156

Step:  AIC=572991.2
Sales ~ Order_Priority + Region + Sub_Category + Discount + Profit +
    Quantity + Shipping_Cost

                Df  Sum of Sq        RSS    AIC
<none>                       3640708576 572991
- Region        12    5354523 3646063099 573043
- Discount       1    9684858 3650393434 573125
- Order_Priority 3  157450551 3798159126 575157
- Quantity       1  168034355 3808742930 575303
- Sub_Category  16  343971478 3984680053 577590
- Profit         1  614742905 4255451481 580992
- Shipping_Cost  1 2663925443 6304634018 601153


Backward stepwise selection of variables
------------------------------------------------------
Linear regression (OLS)
Data     : Dataset
Response variable    : Sales
Explanatory variables: Category, Order_Priority, Region, Segment, Sub_Category, Discount, Profit, Quantity, Shipping_Cost
Null hyp.: the effect of x on Sales is zero
Alt. hyp.: the effect of x on Sales is not zero
```

	coefficient	std.error	t.value	p.value	
(Intercept)	-187.949	7.579	-24.799	< .001	***
Order_Priority|High	132.628	4.811	27.569	< .001	***
Order_Priority|Low	153.424	6.937	22.116	< .001	***
Order_Priority|Medium	203.911	4.654	43.813	< .001	***
Region|Canada	10.943	14.190	0.771	0.441	
Region|Caribbean	-35.446	7.652	-4.633	< .001	***
Region|Central	-5.246	4.751	-1.104	0.270	
Region|Central Asia	22.418	7.166	3.128	0.002	**
Region|East	-11.192	6.448	-1.736	0.083	.
Region|EMEA	0.880	5.447	0.161	0.872	
Region|North	-9.453	5.584	-1.693	0.090	.
Region|North Asia	4.364	6.867	0.635	0.525	
Region|Oceania	9.973	6.057	1.646	0.100	.
Region|South	-7.826	5.194	-1.507	0.132	
Region|Southeast Asia	1.646	6.284	0.262	0.793	
Region|West	-18.990	6.251	-3.038	0.002	**
Sub_Category|Appliances	135.098	8.023	16.838	< .001	***
Sub_Category|Art	-47.335	6.169	-7.673	< .001	***
Sub_Category|Binders	-52.687	5.916	-8.905	< .001	***
Sub_Category|Bookcases	158.699	7.324	21.669	< .001	***
Sub_Category|Chairs	83.794	6.650	12.601	< .001	***
Sub_Category|Copiers	161.403	7.525	21.447	< .001	***
Sub_Category|Envelopes	-52.301	7.256	-7.208	< .001	***
Sub_Category|Fasteners	-66.047	7.276	-9.078	< .001	***
Sub_Category|Furnishings	-34.469	6.758	-5.100	< .001	***
Sub_Category|Labels	-70.477	7.128	-9.887	< .001	***
Sub_Category|Machines	149.388	8.474	17.628	< .001	***
Sub_Category|Paper	-56.530	6.598	-8.567	< .001	***
Sub_Category|Phones	106.057	6.694	15.844	< .001	***
Sub_Category|Storage	14.625	6.108	2.395	0.017	*
Sub_Category|Supplies	-37.354	7.263	-5.143	< .001	***
Sub_Category|Tables	374.648	10.541	35.543	< .001	***
Discount	70.647	6.050	11.677	< .001	***
Profit	0.715	0.008	93.029	< .001	***
Quantity	27.314	0.562	48.637	< .001	***
Shipping_Cost	5.030	0.026	193.656	< .001	***

- Ran multiple regression analysis with Sales as response variable and Order Priority, Region, Sub Category, Discount, Profit, Quantity and Shipping Cost as input variables.
- Apart from a few of the Region variables, all other factors were statistically significant
- Adjusted R-Square for the model was very high which confirmed this is a good model.

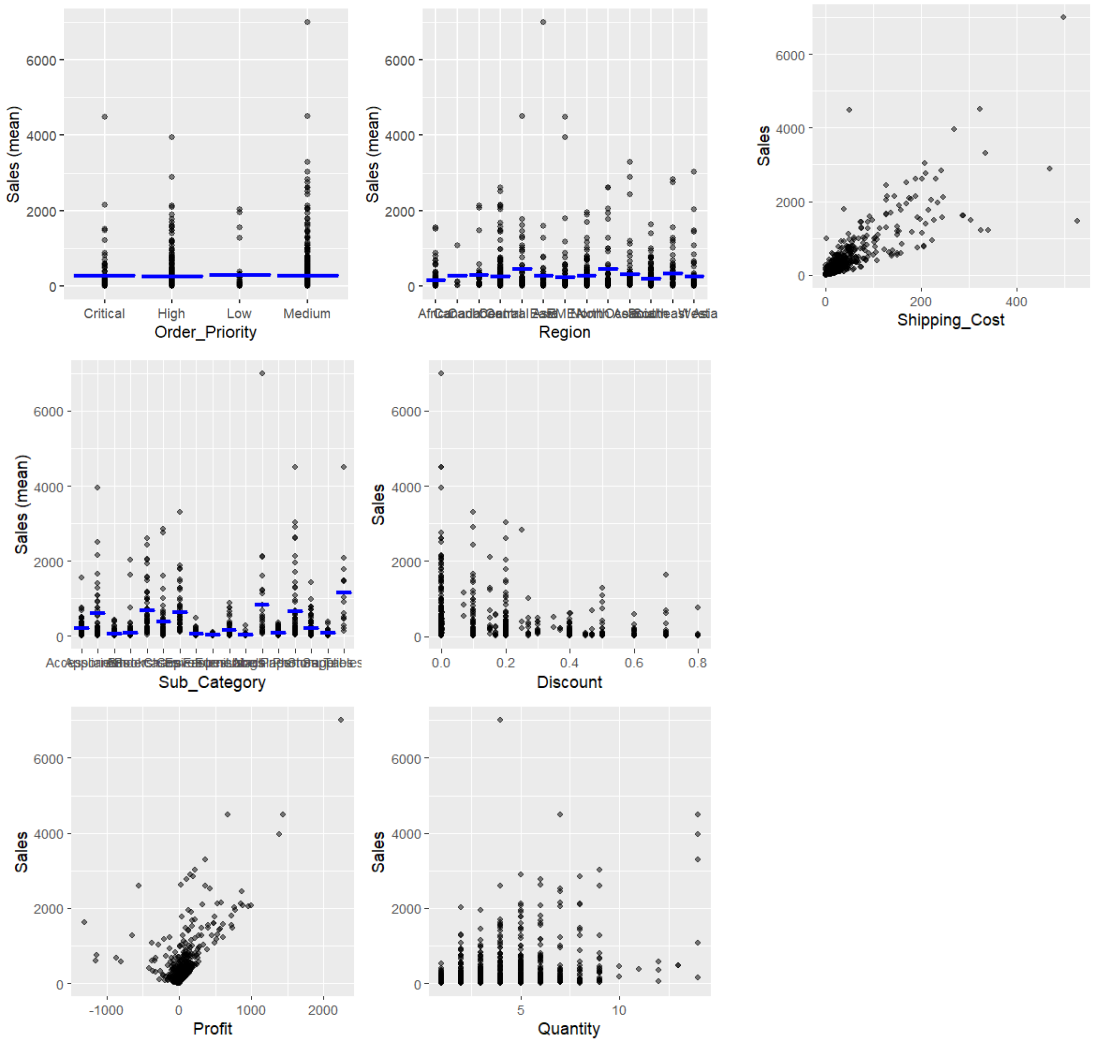
```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.701,  Adjusted R-squared: 0.701
F-statistic: 3439.786 df(35,51254), p.value < .001
Nr obs: 51,290

Prediction error (RMSE):  266.426
Residual st.dev   (RSD):  266.52
```

Note: Screenshots are Radiant output using R

Multiple Regression – Part 1 Scatter Plots



- Shipping cost and profits do not seem to have linear relationship with sales hence interaction terms may exist
- Order priority means are all similar hence this may not be an essential factor in regression

Note: Screenshots are Radiant output using R

Multiple Regression Radiant Output Using R – Part 2

```
Linear regression (OLS)
Data     : Dataset
Response variable    : Sales
Explanatory variables: Sub_Category, Discount, Profit, Quantity, Shipping_Cost
Null hyp.: the effect of x on Sales is zero
Alt. hyp.: the effect of x on Sales is not zero
```

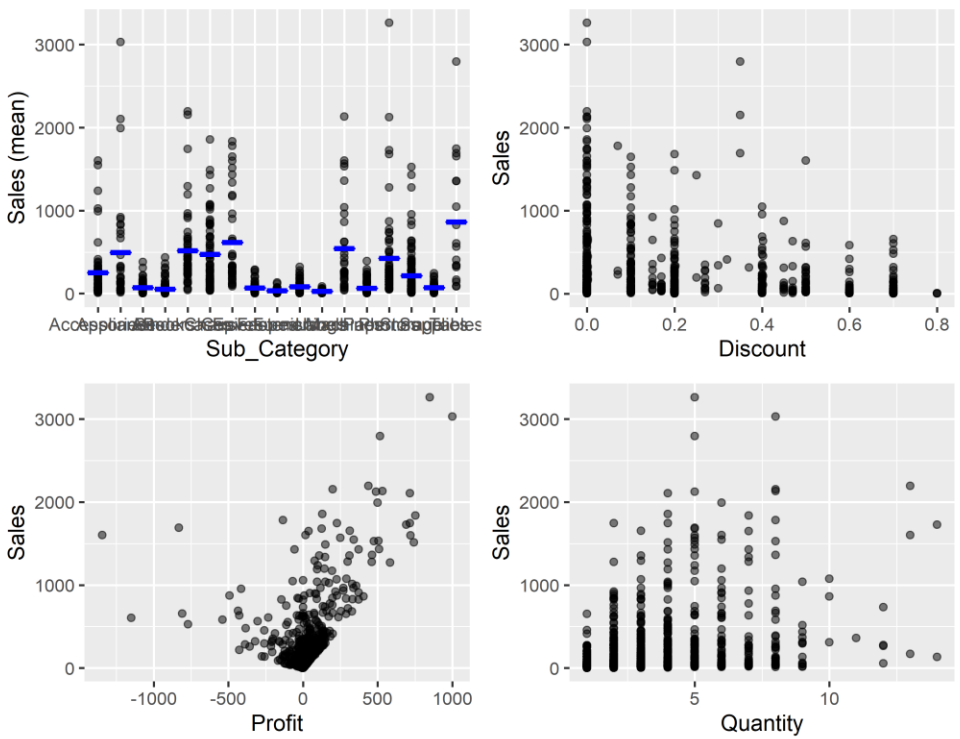
	coefficient	std.error	t.value	p.value	
(Intercept)	3.106	4.870	0.638	0.524	
Sub_Category|Appliances	117.268	7.176	16.342	< .001	***
Sub_Category|Art	-35.886	5.503	-6.521	< .001	***
Sub_Category|Binders	-44.216	5.286	-8.364	< .001	***
Sub_Category|Bookcases	156.873	6.553	23.940	< .001	***
Sub_Category|Chairs	99.473	5.942	16.742	< .001	***
Sub_Category|Copiers	147.290	6.731	21.884	< .001	***
Sub_Category|Envelopes	-40.648	6.477	-6.275	< .001	***
Sub_Category|Fasteners	-47.621	6.499	-7.328	< .001	***
Sub_Category|Furnishings	-20.650	6.039	-3.420	< .001	***
Sub_Category|Labels	-53.608	6.373	-8.412	< .001	***
Sub_Category|Machines	128.728	7.569	17.008	< .001	***
Sub_Category|Paper	-49.134	5.894	-8.336	< .001	***
Sub_Category|Phones	109.014	5.985	18.216	< .001	***
Sub_Category|Storage	26.923	5.448	4.942	< .001	***
Sub_Category|Supplies	-19.983	6.482	-3.083	0.002	**
Sub_Category|Tables	325.397	9.496	34.266	< .001	***
Discount	-14.101	8.976	-1.571	0.116	
Profit	1.834	0.017	105.412	< .001	***
Quantity	16.192	0.645	25.119	< .001	***
Shipping_Cost	2.134	0.052	41.171	< .001	***
Discount:Profit	-2.684	0.026	-104.713	< .001	***
Discount:Quantity	-1.412	2.375	-0.594	0.552	
Discount:Shipping_Cost	5.379	0.161	33.510	< .001	***
Profit:Quantity	-0.019	0.002	-8.173	< .001	***
Profit:Shipping_Cost	-0.001	0.000	-20.042	< .001	***
Quantity:Shipping_Cost	0.228	0.007	31.119	< .001	***

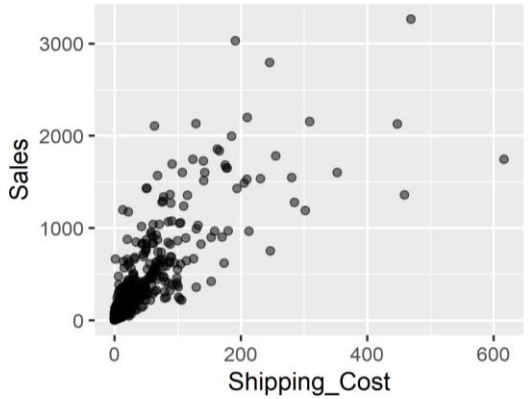
```
R-squared: 0.762,  Adjusted R-squared: 0.762
F-statistic: 6301.668 df(26,51263), p.value < .001
Nr obs: 51,290
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Ran multiple regression analysis with Sales as response variable and Sub Category, Discount, Profit, Quantity and Shipping Cost as input variables.
- The model includes interaction terms Discount : Profit, Discount : Quantity, Discount : Shipping Cost, Profit: Quantity, Profit : Shipping Cost and Quantity : Shipping Cost
- Apart from Discount and Discount : Quantity, all other factors were statistically significant
- Adjusted R-Square for the model was even higher 0.762 which confirmed this is a good model.

Note: Screenshots are Radiant output using R

Multiple Regression – Part 2 Scatter Plots





- Shipping cost and profits do not seem to have linear relationship with sales hence interaction terms may exist
- In the next part we plan to add interaction terms between financial terms

Note: Screenshots are Radiant output using R

Multiple Regression Radiant Output Using R – Part 3

```
Start:  AIC=574018.2
Sales ~ Sub_Category + Profit + Quantity + Shipping_Cost + Profit:Quantity +
    Profit:Shipping_Cost + Quantity:Shipping_Cost

                        Df Sum of Sq        RSS    AIC
<none>                                3716222095 574018
- Profit:Shipping_Cost   1   7283574 3723505669 574117
- Profit:Quantity        1  11606456 3727828551 574176
- Quantity:Shipping_Cost 1  94071306 3810293401 575298
- Sub_Category          16 456042522 4172264617 579923
-----------------------------------------------------
Backward stepwise selection of variables
-----------------------------------------------------
Linear regression (OLS)
Data     : Dataset
Response variable    : Sales
Explanatory variables: Sub_Category, Profit, Quantity, Shipping_Cost
Null hyp.: the effect of x on Sales is zero
Alt. hyp.: the effect of x on Sales is not zero


R-squared: 0.695,  Adjusted R-squared: 0.695
F-statistic: 5315.196 df(22,51267), p.value < .001
Nr obs: 51,290

Prediction error (RMSE):  269.175
Residual st.dev   (RSD):  269.235
```

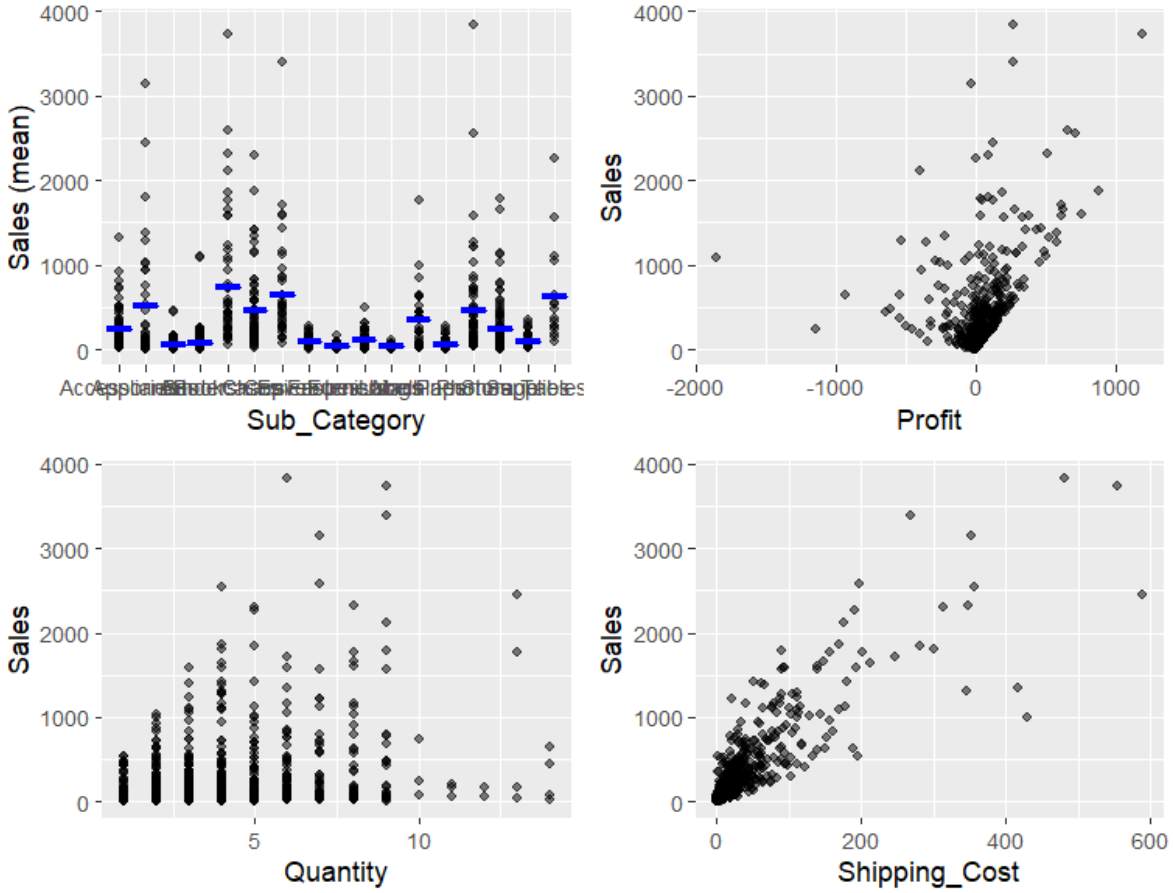
	coefficient	std.error	t.value	p.value	
(Intercept)	10.560	5.285	1.998	0.046	*
Sub_Category|Appliances	154.216	8.105	19.028	< .001	***
Sub_Category|Art	-53.460	6.219	-8.597	< .001	***
Sub_Category|Binders	-53.897	5.972	-9.025	< .001	***
Sub_Category|Bookcases	187.421	7.399	25.332	< .001	***
Sub_Category|Chairs	102.586	6.709	15.292	< .001	***
Sub_Category|Copiers	188.556	7.597	24.819	< .001	***
Sub_Category|Envelopes	-58.073	7.322	-7.931	< .001	***
Sub_Category|Fasteners	-70.991	7.345	-9.666	< .001	***
Sub_Category|Furnishings	-34.277	6.826	-5.021	< .001	***
Sub_Category|Labels	-77.632	7.200	-10.782	< .001	***
Sub_Category|Machines	176.716	8.543	20.685	< .001	***
Sub_Category|Paper	-66.594	6.659	-10.001	< .001	***
Sub_Category|Phones	126.999	6.761	18.783	< .001	***
Sub_Category|Storage	16.850	6.160	2.735	0.006	**
Sub_Category|Supplies	-39.200	7.327	-5.350	< .001	***
Sub_Category|Tables	442.295	10.638	41.577	< .001	***
Profit	0.994	0.017	57.457	< .001	***
Quantity	21.332	0.593	35.983	< .001	***
Shipping_Cost	3.036	0.056	54.542	< .001	***
Profit:Quantity	-0.032	0.003	-12.654	< .001	***
Profit:Shipping_Cost	-0.000	0.000	-10.024	< .001	***
Quantity:Shipping_Cost	0.297	0.008	36.024	< .001	***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Ran multiple regression analysis with Sales as response variable and Sub Category, Profit, Quantity and Shipping Cost as input variables.
- The model includes interaction terms Profit: Quantity, Profit : Shipping Cost and Quantity : Shipping Cost
- All factors were statistically significant
- Adjusted R-Square for the model 0.695 which confirmed this is a good model.

Note: Screenshots are Radiant output using R

Multiple Regression – Part 3 Scatter Plots



- Shipping cost and profits do not seem to have linear relationship with sales hence interaction terms may exist – regression results support this

Note: Screenshots are Radiant output using R