GW Business

Putting It All Together

Analytics to Provide Actionable Insights for SuperSuppliers

Final Project Deliverable by Jonathan Beck

The Stakeholder & Company

Industry

Physical and Electronic Retail.

B2C and B2B model.

Multiple distribution centers.

Company

SuperSupplier, a (fictional) retail-firm based in Washington D.C.

They have a physical presence; however, the focus recently has been on expanding their business by incorporating e- commerce.

Prides themselves on their large electronic catalog and high quality products.

Ned Hulseman

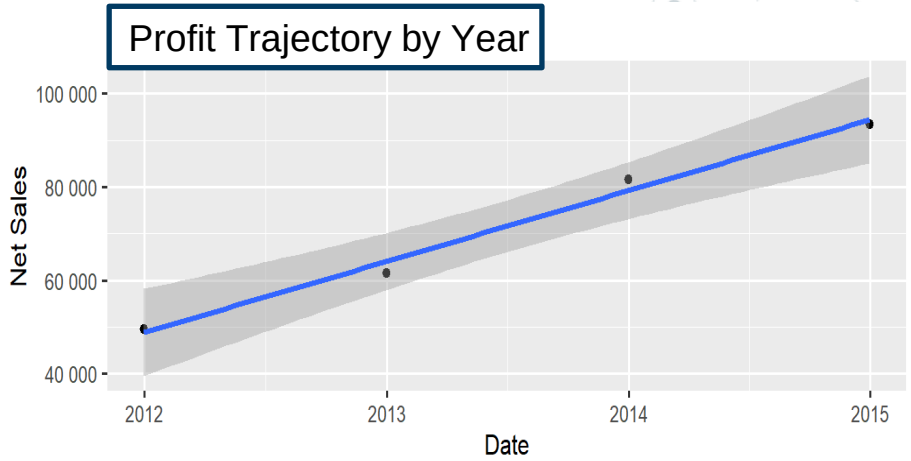
Newly appointed Vice President of Sales and Marketing at SuperSupplier.

Business Needs for SuperSupplier's

- Enhanced customer understanding.
  - *Who are SuperSupplier's most profitable clients? Least profitable clients? Which clients place the highest volume of orders?*
- Less variation in profit on a day-to-day basis
  - *Refined business model and strategy.*
- **GOAL:** Identify pain points in sales and provide solutions that can be incorporated promptly.

At a Glance: Year to Year

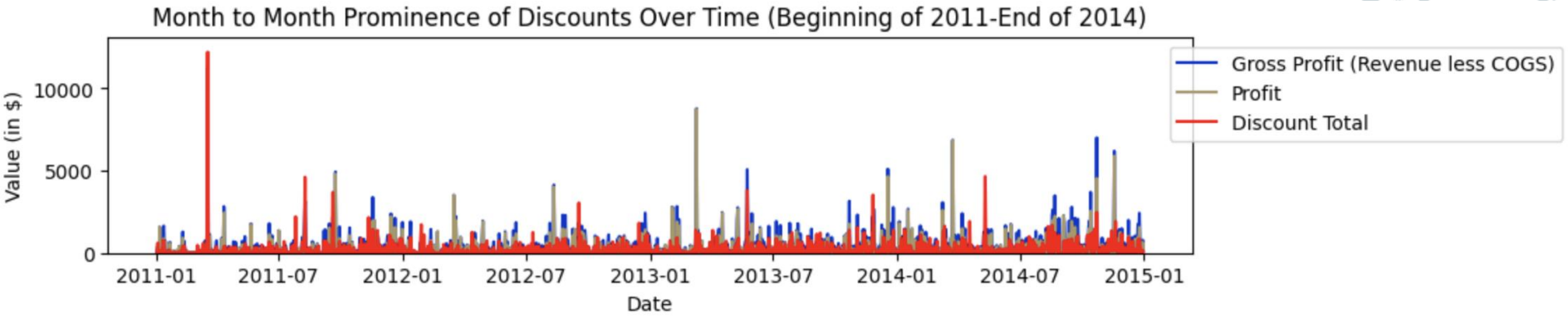
order_date	Cumulative Profit	Total Orders Placed	Total Goods Sold
2014-12-31	93507.5131	1692	12503
2013-12-31	81726.9308	1310	9810
2012-12-31	61618.6037	1038	7979
2011-12-31	49543.9741	969	7581



Cumulative Year to Year Bottom Line Analysis

- Upon initial observation of SuperSupplier's year to year trajectory, sales appear sturdy.
- Profit, total orders placed, and total goods sold have each increased year to year.
- Growth trend of profit is promising.

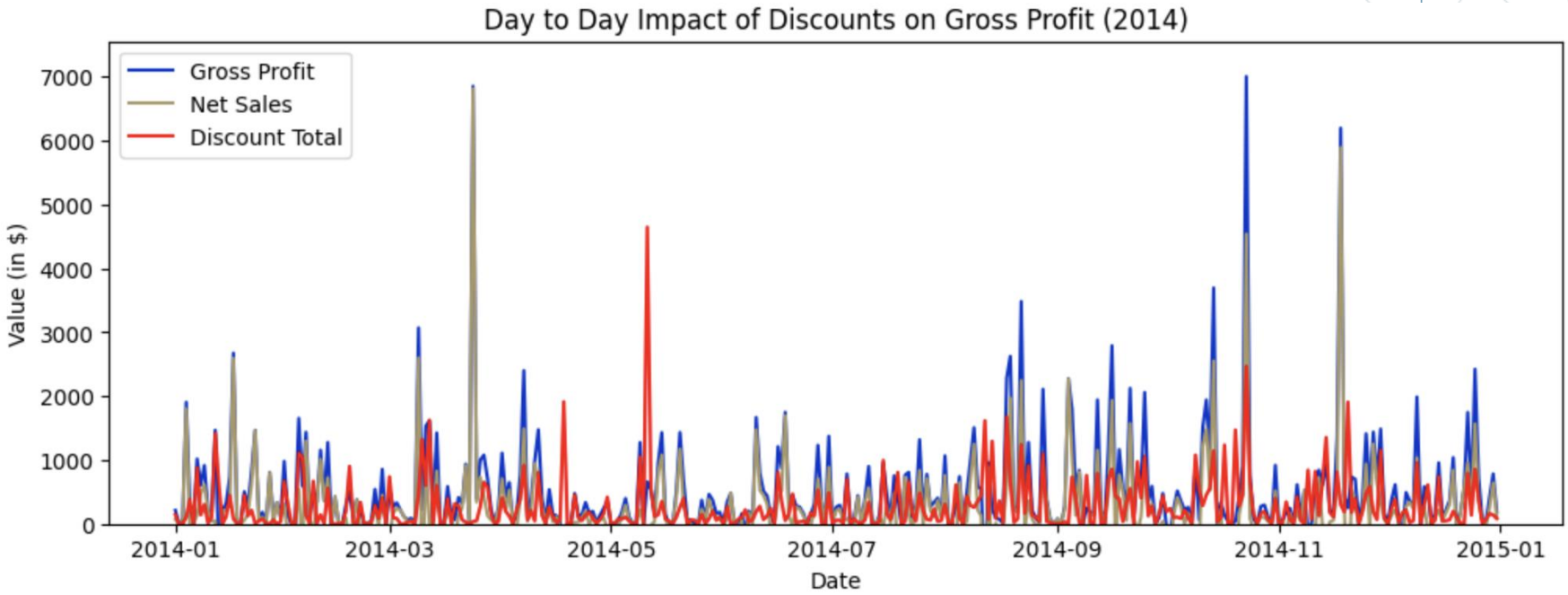
Prominence of Discounts in Sales



Business Issue for SuperSuppliers'

- Profitable sales are currently being hard capped by an over issuance of discounts being applied to each customer order.

2014 Discount Visual

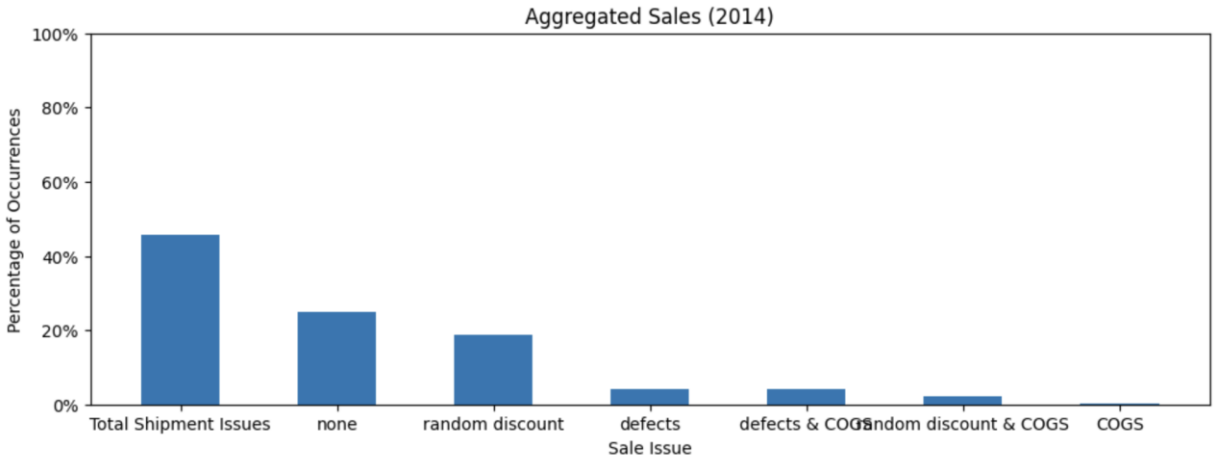


2014 Total Profit: $93,507

Total Monetary Value of Discounts in 2014: $101,398

Profit had discounts been more tightly controlled: ***$194,905***

The Discounts and Sale Issues



Sales Issues (2014)	Percentages
Total Shipment Issues	45.86%
none	25.19%
random discount	18.71%
defects & COGS	4.04%
defects	3.65%
random discount & COGS	2.44%
COGS	0.12%

Insight

- In the most recent sales period (2014), our analytics team have pinpointed a number of 'Sales Issues' as the primary cause for this negation in sales.
- **'Issues with Shipment' define almost half of SS' sales in 2014 and I believe this to be the primary issue to resolve going forward for SS to see an immediate spike in profit going into the next fiscal period.**

Standard vs Reality in Shipping

SS Shipping Standards

- ***Standard Class:*** 6 Days
- ***Second Class:*** 4 Days
- ***First Class:*** 3 Days
- ***Same Day:*** 0 Days

Prominent Issue:

- Most customers purchase using ***Standard Class,*** they expect these goods to be shipped within 6 days, and yet, on average, their goods are being shipped close to 50 days after being ordered!
- Customer retention is difficult, if not impossible, in the current state of business.

order_date	ship_mode	Average Days to Ship
2011	First Class	27.0
	Same Day	0.0
	Second Class	39.6
	Standard Class	53.2
2012	First Class	27.3
	Same Day	0.9
	Second Class	32.5
	Standard Class	53.1
2013	First Class	25.4
	Same Day	0.2
	Second Class	36.0
	Standard Class	48.5
2014	First Class	24.6
	Same Day	1.0
	Second Class	34.7
	Standard Class	56.0

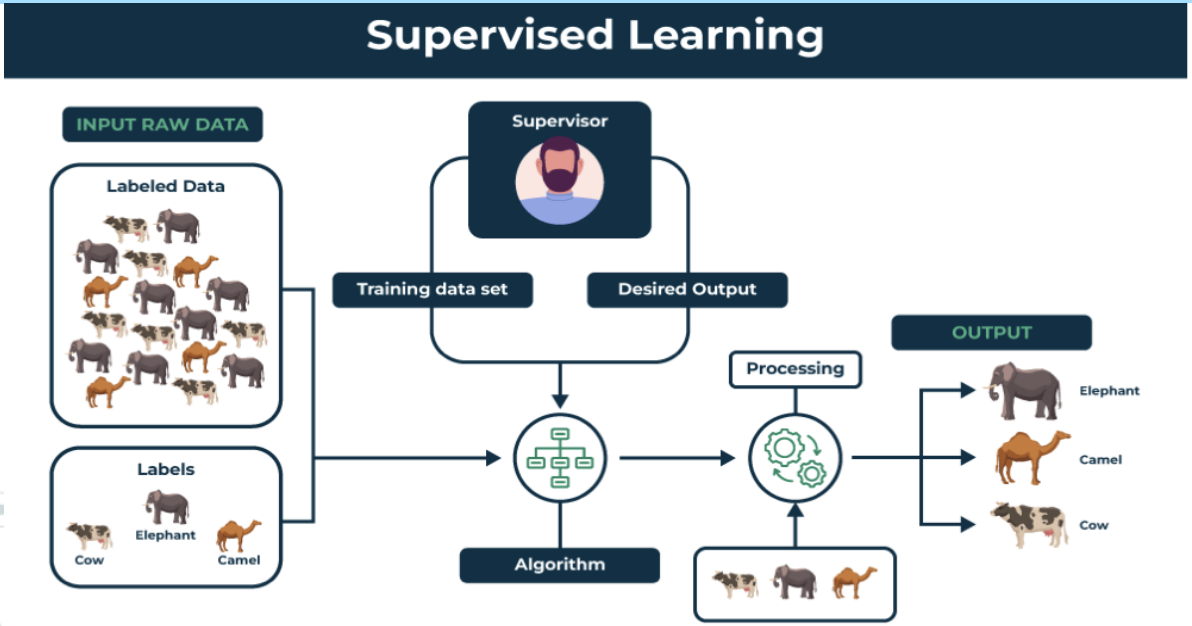
order_date	ship_mode	Total Goods Shipped
2011	First Class	932
	Same Day	356
	Second Class	1194
	Standard Class	3651
2012	First Class	962
	Same Day	383
	Second Class	1340
	Standard Class	3839
2013	First Class	1246
	Same Day	546
	Second Class	1618
	Standard Class	4506
2014	First Class	1842
	Same Day	642
	Second Class	2228
	Standard Class	5702

Solution: Classification Tree

GW Business

GWU Business Consulting LLP

- We have designed a Supervised Learning algorithm that is able to discern descriptive sales data attributes from each record of sale that is causing a **Shipment Issue** to arise.
- Moreover, the model is able to reasonably predict whether a sale will cause such an issue to occur.



The Model: Holistic Description

GW Business

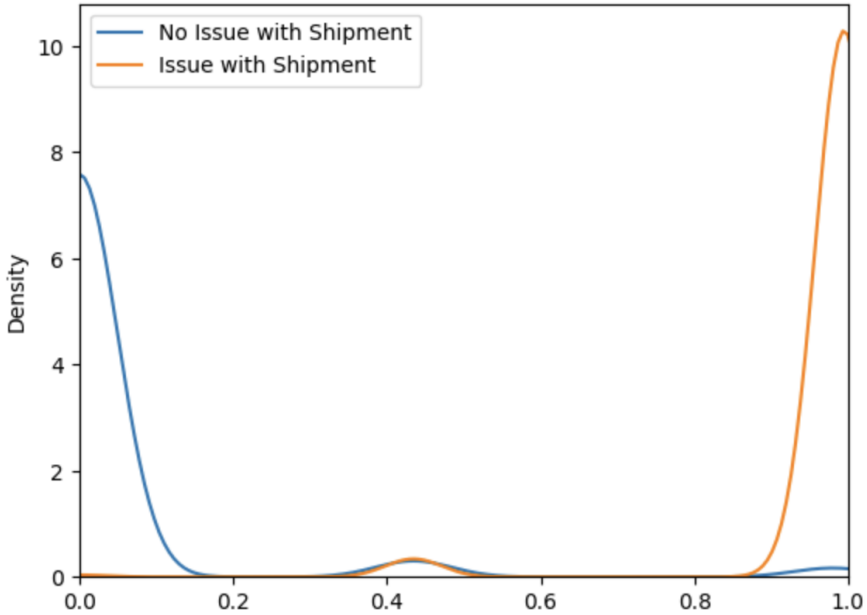
<u>Classification Tree</u> to Predict Probability that a Sale will lead to a *Shipment Issue*

- **Objective*:*** To predict whether a sales transaction will lead to an issue with shipment *without being aided by key historical variables, found in the descriptive sales data, such as 'Days (a past order took) to Ship'*. **Why?** So model can be applied to incoming, future orders.
- **Data Used in Training & Testing*:*** Each transaction that occurred from 2011-2014, as this provides the model with more data to learn from during training, to be able to make more accurate predictions on unseen (i.e. future) data.
- **Binary Classification:** Each observation was labeled with either a 0 (No Shipment Issue) or a 1 (Shipment Issue) for training. These output values were hidden from the model during testing.
- **Limiting Complexity of Model & Preventing Overfitting*:***
  - Manual Dimensionality Reduction of variables likely to be recognized as noise (e.g., Order #)
  - Model Depth capped at 4. Minimum sample size to split at 25.

Current State of Model: Even lacking key data points that would strongly aid its predictive capabilities, (data that, for whatever reason, is not yet stored with the sales transactions e.g., manufacturing details, supplier info, quantity of product held in inventory, etc.), and needing to discern 'Shipment Issues' from a number of other sales issues, its predictive capabilities are still commendable -> **roughly 63% predictive accuracy in classifying during testing.**

Going forward, we recommend obtaining descriptive data from vendors and supply side distributors to enhance the models predictive capabilities. This model will then be able to *confidently* flag orders that are likely to cause 'Shipping Issues' as they are received by the processing system, ensuring SS can take the necessary precautions to prevent the issue.

KDE Plot with Higher Quality Data



Visual Supports that Model is extremely confident in classification of data when given sufficient detail in shipment information

- Minimum overlap (uncertainty) when classifying data during testing.
- Model can confidently discern between a Shipment Issue and a Non-Shipment Issue with each transaction.

Business Value

- Model can effortlessly be fine-tuned to analyze alternative avenues worth pursuing, with similar predictive efficiency.
  - *For example, SS may not want model to split and classify using 'Customer Name' as a variable. A model with this variable removed was observed and its predictive qualities were practically identical to the presentation model (a few decimal points less accurate during testing). By simply removing this variable, model adjusted its splitting and classification methodology accordingly; moreover, the new model provided clear indication that specific zip codes and cities correlate with a higher frequency of issues with shipment.*
- Model is able to provide conclusive leads for SS management to pinpoint what is causing these shipment issues to arise and which common sales classes are causing these issues to occur most frequently.
- Going forward, as more data is collected with each sales transaction and other sales issues are resolved, model will aid SS in identifying likely 'Shipping Issues' before they occur.

For Best Results

- Obtain more supply side data from all suppliers and vendors of products being offered for immediate uptick in classification and predictive output of model.
  - Model can easily be retrained using these data points and will undoubtedly see greater predictive output from model.
- Take initiative (detailed in Appendix) to resolve the other issues occurring in sales.
- Closely monitor all transactions, in the next sales period, that have sales characteristics inline with those correctly classified by (current) model as having 'Shipment Issues', as soon as order is placed.

Business Outcome

- Reinstate trust within customer base.
- Retaining a larger portion of customers.
- Customers feel that their business is valued.
- Higher ratio of Customer Retention = More profitable day-to-day business operations
- Most frequently occurring issue causing loss in profit to occur will be understood and resolved.

BOTTOM LINE

Less shipment issues -> more loyal customers -> less discounts necessary to keep customers returning -> higher profits.

GW Business

GWU Business Consulting LLP we hope that these insights & recommendations prove valuable!

Any questions? Feel free to reach out!

If needed, attached is a link to a copy of your data of initial data:

Superstore Dataset (kaggle.com)

Table of Contents: **Core Appendix**

1. Additional Analysis
   I. General
   II. Product
   III. Client
2. Additional Sale Issues & Outcomes

1.1 Additional Analysis & Business Recommendations: General

GW Business

Month to Month

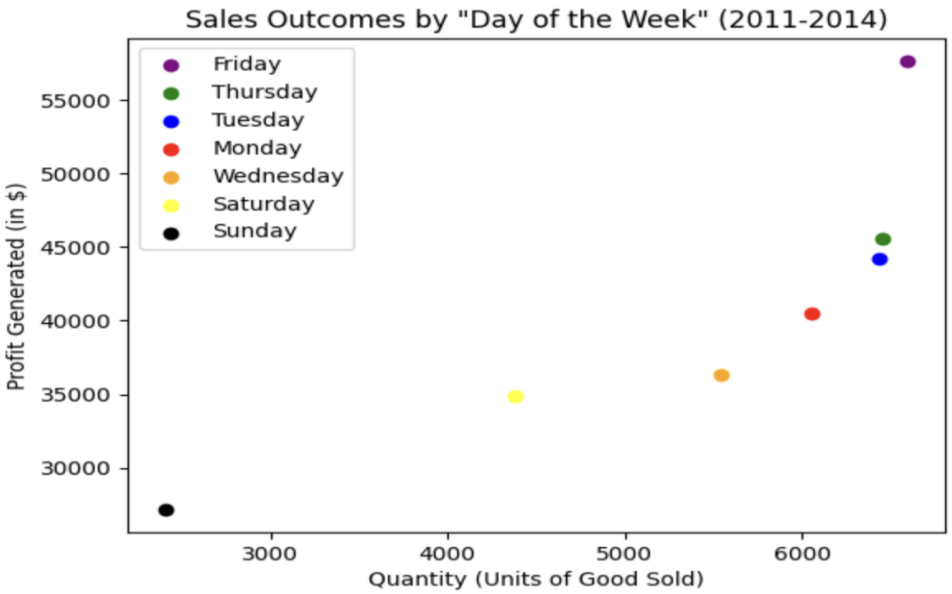
order_month	Cumulative Goods Sold	Cumulative Profit	Total Orders Placed
September	4302	36045.8966	565
March	3093	35523.0867	405
December	4132	34996.7550	559
August	3107	24914.4554	387
November	4788	24004.7050	628
October	3081	23268.4253	403
June	2770	21729.1049	385
May	3144	20981.6013	418
February	1931	20410.5561	254
January	2050	17745.6229	278
April	2703	17131.5296	375
July	2772	9645.2829	352

order_month	Goods Sold (2014)	Profit Realized (2014)	Total Orders Placed (2014)
March	1247	14047.2219	150
September	1293	12347.1450	168
January	837	9519.7930	121
November	1466	9287.5185	211
October	1184	9125.3986	156
December	1230	8608.8258	175
June	902	8399.6258	117
August	995	7260.4549	126
July	798	5701.4371	109
February	703	4764.5993	81
April	931	2930.5848	144
May	917	1514.9084	134

General Month to Month Trends (Cumulative vs 2014)

- March and September appear to be sales periods when SuperSupplier's should expect to receive the largest volume of sales.

Day of Week



Day of Week Trends

- Aggregated sales qualities, as per day of week (2011-2014) reveals a positive correlation with profit and quantity of goods ordered.
- SS can expect to receive highest volume of sales towards the end and beginning of the business week.

Goods Sold vs Net Sales (2014)

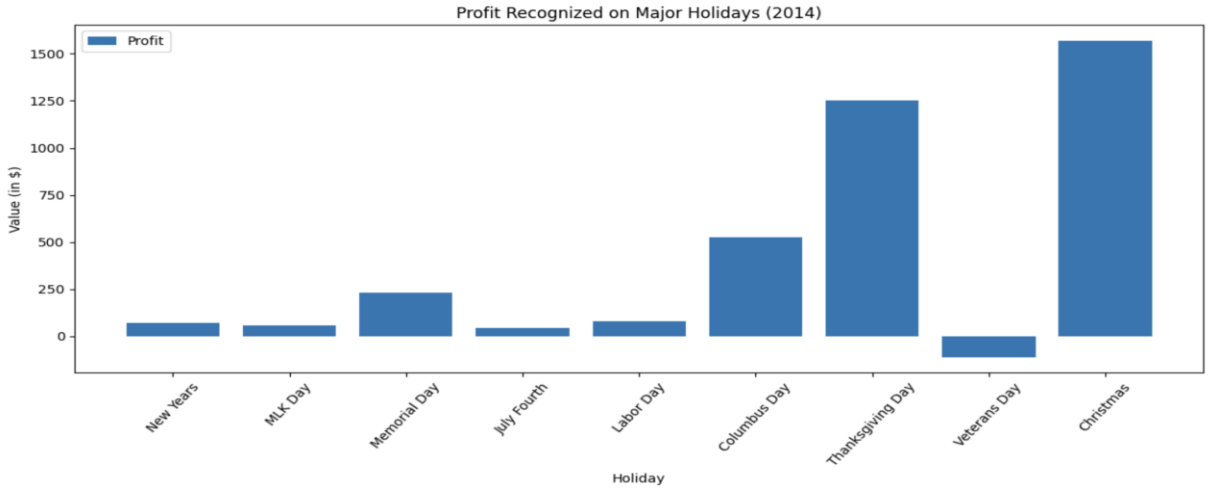
GW Business

Digging Deeper into Sales in 2014

- The chart to the right explicitly illustrates the large variation in number of goods sold and profit outcome, by day.
- Clearly see that, until discounts are removed from being a prominent part of SS' business model, a positive relationship between Net Sales and (Volume) of Goods Sold will not be achievable.
- ***Valuable Insight***
  - It is very likely that the prior VP in sales gradually realized that customers purchasing in volume often generated higher net sales; therefore, began distributing discounts to increase the likelihood that customers will add more to their carts.
  - We clearly see that, unfortunately, the opposite effect occurred, more often than not.

Date	Goods Sold	Profit	Sale Rank
2014-03-12	152	-7.3410	302
2014-02-12	149	718.8920	48
2014-03-09	126	2600.1853	4
2014-10-12	120	1487.1418	16
2014-03-10	116	-1095.5167	358
2014-11-20	114	-1659.0309	362
2014-09-12	114	1154.6045	24
2014-11-25	110	935.9258	33
2014-09-16	110	1934.9693	10
2014-05-09	108	224.4251	132
2014-03-11	105	929.1536	34
2014-11-14	103	-398.4248	350
2014-10-09	101	-152.8754	337
2014-12-23	100	957.1227	31
2014-12-09	97	1032.8847	27

Holiday Sales (2014)



Current Analysis

- It is clear that Christmas Day and Thanksgiving Day are two of SS's most profitable sales days; however, the major insight to be gained from this observation pertains to sales frequency and strange deviations from the norm.
  - ***Columbus Day:*** Customers purchased 78 items on this date and yet SS only generated $526 in profit for SS. Roughly $700 less than *Thanksgiving (in which 69 goods were sold)* and over $1000 less in profit than on *Christmas (in which only 54 goods were sold).*
  - ***Veterans Day:*** 64 goods were bough by SS clientele on this date and yet, this resulted in a $111 decrease in profit.

    - ***NOTE:*** The reasons for these deviations is due to discounts.

GW Business

1.2 Additional Analysis & Business Recommendations: Product

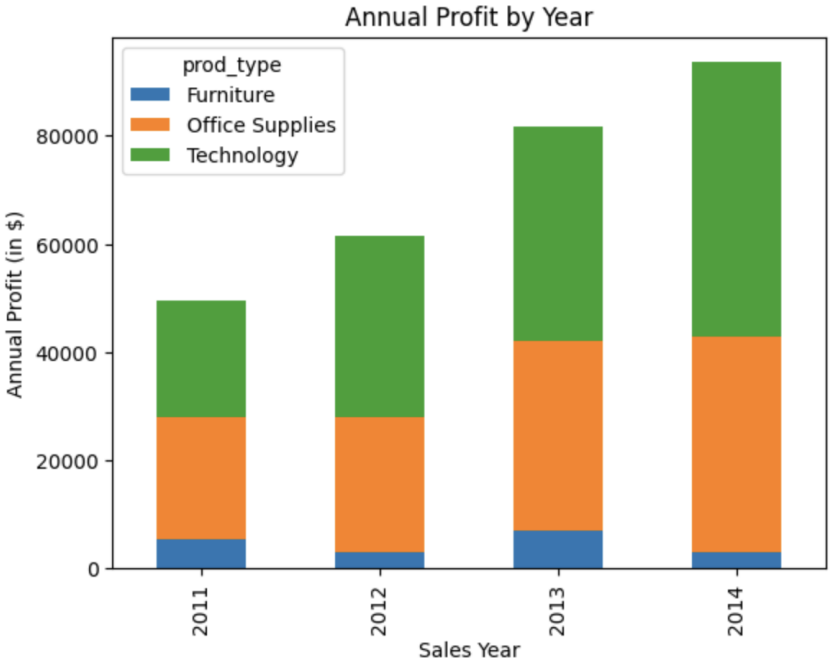
Profit and Product Base

Cumulative (2011-2014) by Product Type

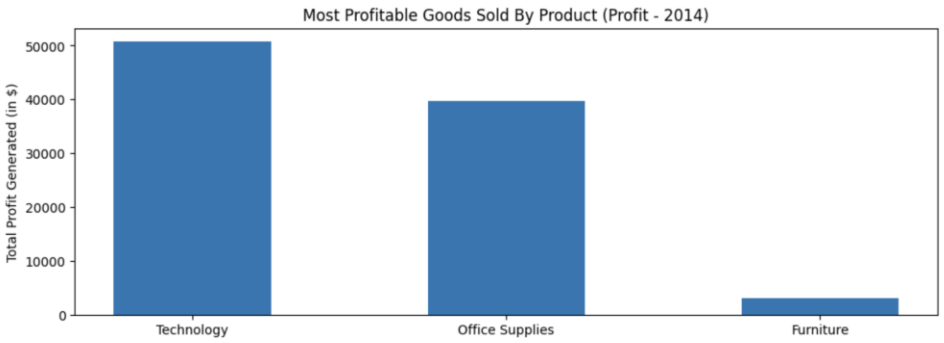
- ***Technology:*** $145,454 in profit
  - ***Total Goods Sold:*** 6,939
  - ***Discounts Applied***: 1,014
  - ***Average Price of Good:*** $132
  - ***Average Cost to Sell:*** $94

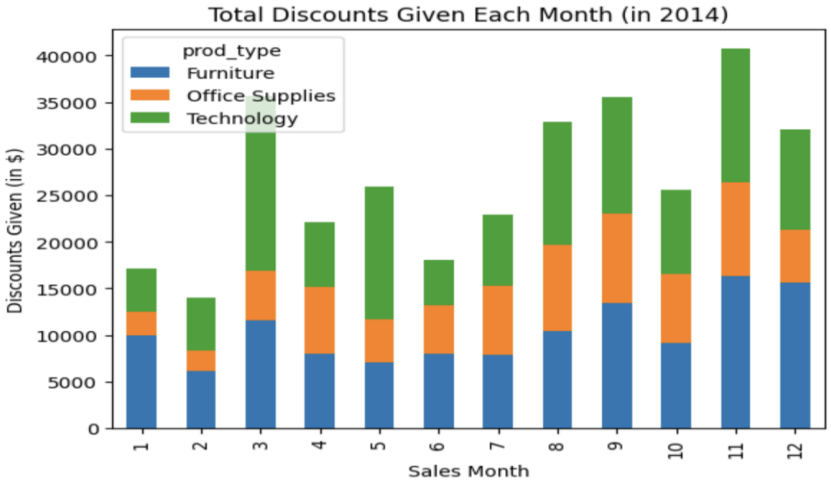
- ***Office Supplies:*** $122,490 in profit
  - ***Total Goods Sold:*** 22,906
  - ***Discounts Applied***: 2,897
  - ***Average Price of Good:*** $34
  - ***Average Cost to Sell:*** $25

- ***Furniture:*** $18,451 in profit
  - ***Total Goods Sold:*** 8,028
  - ***Discounts Applied***: 1,285
  - ***Average Price of Good:*** $102
  - ***Average Cost to Sell:*** $85



Profit vs Discounts on Products (2014)





- ***Technology***
  - **Sold:** 2,366
  - **Discounts Applied:** 342
  - **Average Price:** $131
  - **Average Cost to Sell:** $82
- ***Office Supplies***
  - **Sold:** 7,700
  - **Discounts Applied:** 974
  - **Average Price:** $34
  - **Average Cost to Sell:** $25
- ***Furniture***
  - **Sold:** 2,437
  - **Discounts Applied:** 412
  - **Average Price:** $97
  - **Average Cost to Sell:** $81

Overview of Product Sub-Types

GW Business

Cumulative Product Sub-Type Chart

- An insightful chart illustration that outlines how each product sub-type has performed over the years.
- **Phones:** Although it is the product that contributes to 14.9% of SS's total discounts (in terms of total $ lost) it still cumulatively generates the second most profit.
- **Tables, Bookcases, & General Supplies:** Three products that seem to be a trouble area in terms of sales.
  - ***Tables:*** Have lost over $17K in profit for SS since 2011!
  - ***Supplies:*** Average selling price of these goods is only $3 more than the cost to supply them.
- Most products within these sub-types should be removed from inventory.

prod_sub-type	Average Cost	Average Price	Sold	Total Profit	Discounts Given	Profit Rank	Product Type	Percentage of Total Discounts
Copiers	353.853605	648.004043	234	55617.8249	46	1	[Technology]	5.6
Phones	83.450611	111.661484	3289	44515.7306	578	2	[Technology]	14.9
Accessories	39.073685	56.499443	2976	41936.6357	304	3	[Technology]	3.0
Paper	8.175556	15.940055	5178	34053.5693	513	4	[Office Supplies]	1.6
Binders	28.095978	42.430974	5974	30221.7633	1186	5	[Office Supplies]	13.6
Chairs	122.916050	154.995135	2356	26590.1663	484	6	[Furniture]	15.4
Storage	61.803519	72.773210	3158	21278.8264	316	7	[Office Supplies]	4.1
Appliances	48.520925	62.668333	1729	18138.0054	195	8	[Office Supplies]	2.3
Furnishings	21.127006	27.173232	3563	13059.1436	386	9	[Furniture]	2.7
Envelopes	10.184564	19.470501	906	6964.1767	102	10	[Office Supplies]	0.4
Art	6.403543	9.119300	3000	6527.7870	298	11	[Office Supplies]	0.6
Labels	4.306747	8.439536	1400	5546.2540	125	12	[Office Supplies]	0.2
Machines	395.511863	505.939431	440	3384.7569	86	13	[Technology]	14.5
Fasteners	2.124538	3.421507	914	949.5182	89	14	[Office Supplies]	0.1
Supplies	68.874090	71.455806	647	-1189.0995	73	15	[Office Supplies]	0.9
Bookcases	128.121007	148.178208	868	-3472.5560	168	16	[Furniture]	6.4
Tables	174.244484	195.554528	1241	-17725.4811	247	17	[Furniture]	13.7

Product Sub-Types (2014)

Ranked by Frequency of Discounts

- This chart provides a deeper analysis as to how each product sub type has been performing recently.
  - **Copiers:** Not only is this SS's most profitable type of product to sell, but it causes the least amount of issues with discounts in terms of frequency of occurrences.
    - **Copiers** should be more heavily advertised to customers.
  - **Binders:** Most amount of discounts applied to these product types by a good amount and product that is responsible for 16.9% of the total discount value in 2014.
    - If impossible to avoid applying discounts, remove.

prod_sub-type	2014 Average Cost	2014 Average Price	2014 Goods Sold	2014 Total Profit	Freq of Discounts Applied (2014)	2014 Profit Rank	Product Type	Total Cost of Discounts (2014)	Percent of Discount Cost (2014)
Binders	30.410052	43.116529	2077	7737.0951	395	6	[Office Supplies]	17129.733900	16.9
Phones	81.997190	109.118558	1093	12872.0032	202	3	[Technology]	16007.350400	15.8
Paper	8.495039	16.578567	1707	12049.0614	183	4	[Office Supplies]	1830.577600	1.8
Chairs	120.792129	151.792291	674	7643.5493	148	7	[Furniture]	14277.170300	14.1
Furnishings	20.659047	26.782227	1097	4099.1628	126	9	[Furniture]	2721.673600	2.7
Art	5.760256	8.287217	1106	2227.1723	112	10	[Office Supplies]	576.692800	0.6
Storage	59.524726	70.615201	1026	7367.5855	110	8	[Office Supplies]	4055.592000	4.0
Accessories	38.405653	55.912409	1079	15672.3570	99	2	[Technology]	3106.774400	3.1
Tables	165.997980	184.767205	390	-8140.6947	82	17	[Furniture]	14873.403075	14.7
Appliances	46.650979	60.798189	654	7865.2683	65	5	[Office Supplies]	2541.832400	2.5
Bookcases	118.213902	136.602447	276	-583.6261	56	14	[Furniture]	4742.103159	4.7
Labels	4.050460	7.969133	448	1744.6093	36	11	[Office Supplies]	172.659200	0.2
Envelopes	8.272379	16.016076	241	1441.7590	31	12	[Office Supplies]	230.444800	0.2
Machines	399.643726	497.945250	121	-2869.2156	26	16	[Technology]	11499.239900	11.3
Fasteners	2.107440	3.518007	249	304.9489	21	13	[Office Supplies]	56.860800	0.1
Supplies	75.948854	78.116314	192	-955.3128	21	15	[Office Supplies]	1376.192000	1.4
Copiers	438.367505	832.181709	73	25031.7902	15	1	[Technology]	6199.912000	6.1

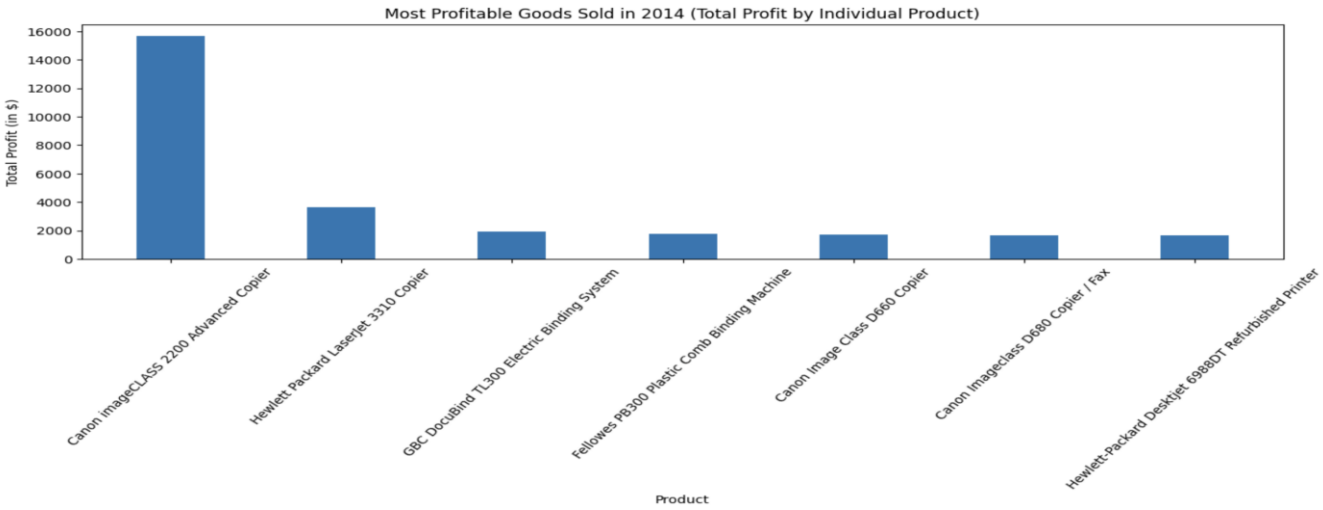
Product Sub-Types (2014)

Ranked by Number of Items Sold

- This chart illustrates which products are most popular, in terms of sales frequency, in 2014.
- **Copiers:** Unfortunately, Copiers is SS's least popular product in terms of times sold! The marketing team for SS certainly needs to prioritize making more sales of these products, whilst steering clear of applying discounts to do so.
- **(General) Supplies:** Clearly a product that should be strongly considered for discontinuation. It isn't popular and has led to an almost $1,000 decrease in net sales in 2014!

prod_sub-type	2014 Average Cost	2014 Average Price	2014 Goods Sold	2014 Total Profit	Freq of Discounts Applied (2014)	2014 Profit Rank	Product Type	Total Cost of Discounts (2014)	Percent of Discount Cost (2014)
Binders	30.410052	43.116529	2077	7737.0951	395	6	[Office Supplies]	17129.733900	16.9
Paper	8.495039	16.578567	1707	12049.0614	183	4	[Office Supplies]	1830.577600	1.8
Art	5.760256	8.287217	1106	2227.1723	112	10	[Office Supplies]	576.692800	0.6
Furnishings	20.659047	26.782227	1097	4099.1628	126	9	[Furniture]	2721.673600	2.7
Phones	81.997190	109.118558	1093	12872.0032	202	3	[Technology]	16007.350400	15.8
Accessories	38.405653	55.912409	1079	15672.3570	99	2	[Technology]	3106.774400	3.1
Storage	59.524726	70.615201	1026	7367.5855	110	8	[Office Supplies]	4055.592000	4.0
Chairs	120.792129	151.792291	674	7643.5493	148	7	[Furniture]	14277.170300	14.1
Appliances	46.650979	60.798189	654	7865.2683	65	5	[Office Supplies]	2541.832400	2.5
Labels	4.050460	7.969133	448	1744.6093	36	11	[Office Supplies]	172.659200	0.2
Tables	165.997980	184.767205	390	-8140.6947	82	17	[Furniture]	14873.403075	14.7
Bookcases	118.213902	136.602447	276	-583.6261	56	14	[Furniture]	4742.103159	4.7
Fasteners	2.107440	3.518007	249	304.9489	21	13	[Office Supplies]	56.860800	0.1
Envelopes	8.272379	16.016076	241	1441.7590	31	12	[Office Supplies]	230.444800	0.2
Supplies	75.948854	78.116314	192	-955.3128	21	15	[Office Supplies]	1376.192000	1.4
Machines	399.643726	497.945250	121	-2869.2156	26	16	[Technology]	11499.239900	11.3
Copiers	438.367505	832.181709	73	25031.7902	15	1	[Technology]	6199.912000	6.1

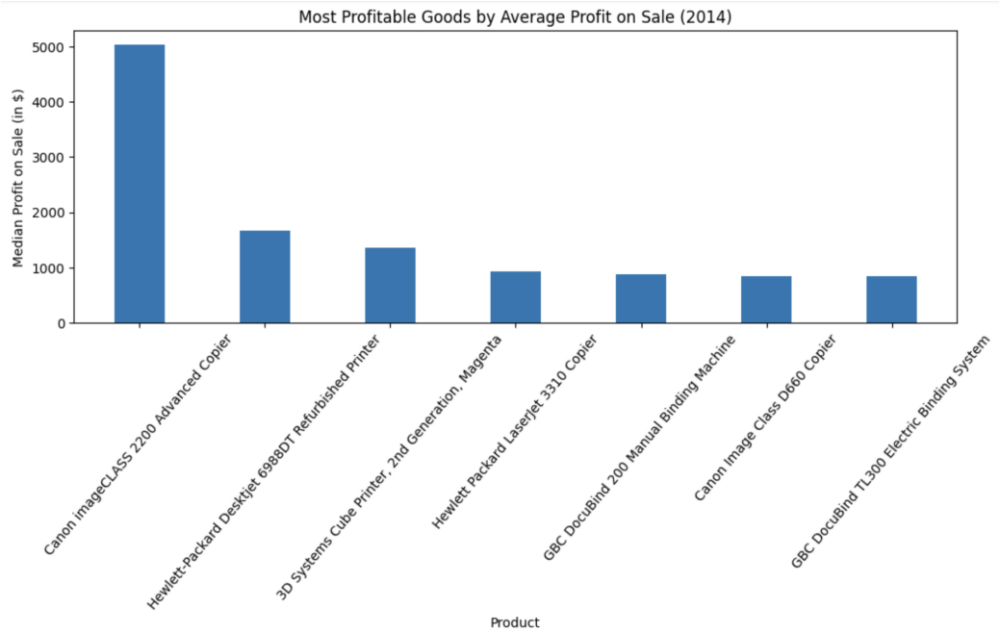
Individual Product Analysis (2014)



Goods that generated the most profit in 2014 (*Aggregated Total Net Sales*)

- ***Canon Copier:*** Only 11 of these products have been sold in 2014 and yet is still has generated a grand total of almost $16K in profit.
- ***The GBC DocuBind & the Plastic Comb Binding Machine:*** It is somewhat surprising to see that these products of the "Binder and Office Suppliers" class fall within the top 7 profit generating products. These two products could generate an even larger spike in profit if not for the fact that they, on average, are typically sold with 32% and 15% discounts applied to them, even though they tend to ship within only 5 days.

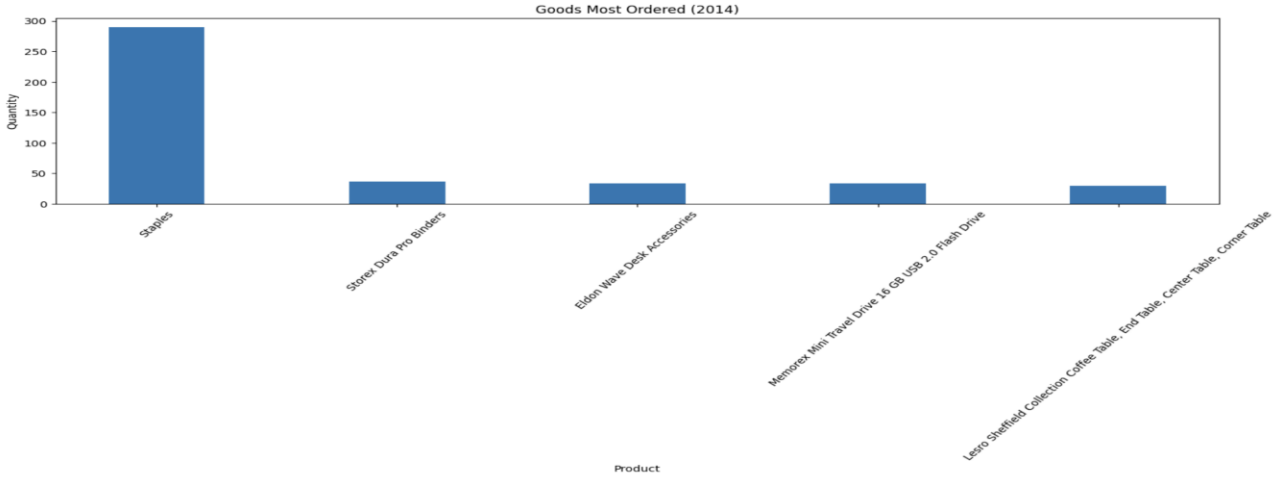
Individual Product Analysis (2014)



Goods that generated the most profit in 2014 (*Aggregated Median Net Sales*)

Canon Copier: Per order, on average, this product generates a little over $5K in profit!

Individual Product Analysis (2014)

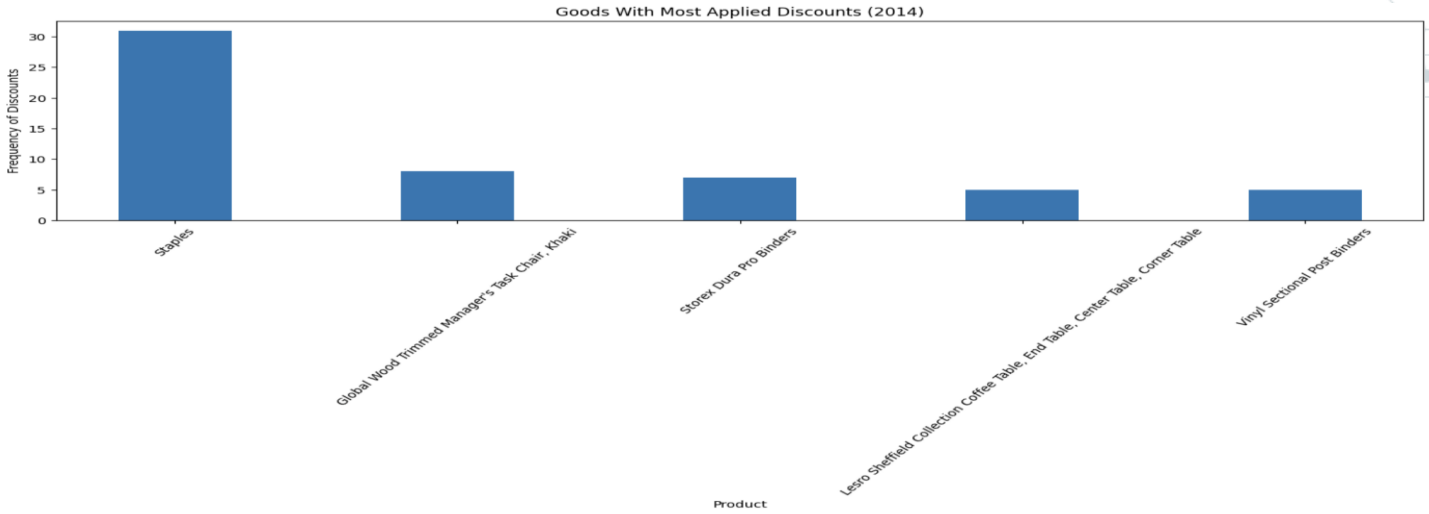


Most Ordered Goods in 2014

Noteworthy Background Detail on Some of Above Products

- ***Staples:*** This product has been sold 290 times; however, 31 of those occurrences have had discounts applied to them.
- ***Storex Dura Pro Binder:*** On (mean) average, this product costs $3.08 to sell with a selling price of only $3.59 (on average). Moreover, the median time delay in shipping is 40 days for this product and averages an applied 50% discount per order. For these reasons, it is clear why this frequently purchased product has only generated a total of $23.58 in sales in 2014, with a mean net sale of negative $2.

Individual Product Analysis (2014)

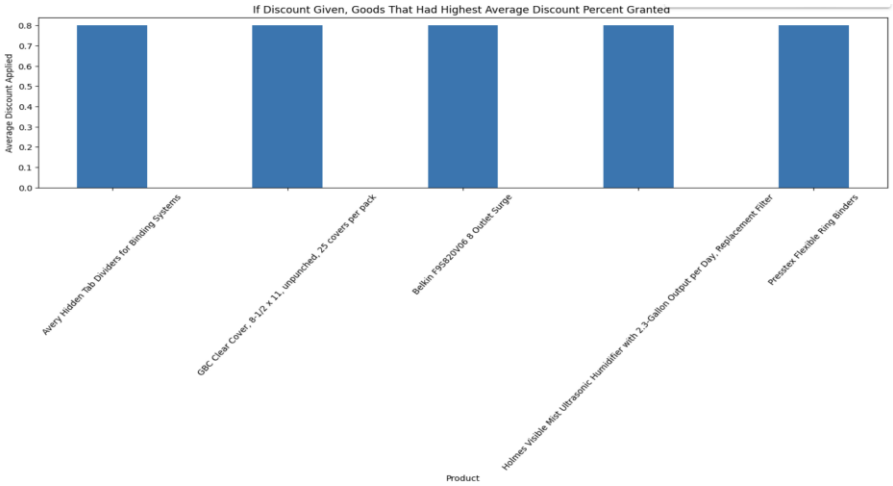


Products That Correlate with Highest Frequency of Discounts

Noteworthy Background Detail on Some of Above Products

Global Wood Trimmed Task Chair: On average, this product faces a $75 cost to sell with an average selling price of $81.42. Those are terrible margins for a product that should be PROFITABLE. Moreover, on average, this product takes 33 days to ship and has had 8 discounts applied to the only 27 sold chairs of this type. With each discount average 25%, this has all culminated in the product underperforming and generating a loss of $272 for SS in 2014.

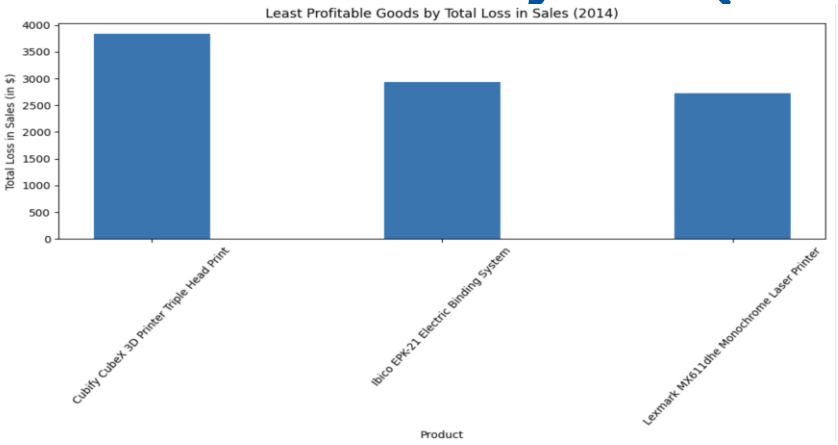
Individual Product Analysis (2014)



Goods With Highest, On Average, Applied Discount Percentage

- ***On average,*** each of these products being sold causes the sale to be negated with an 80% discount. Moreover, and perhaps unsurprisingly, each of these products (all of which are from the "Office Supplies" Product Base) have generated losses for SS in 2014.

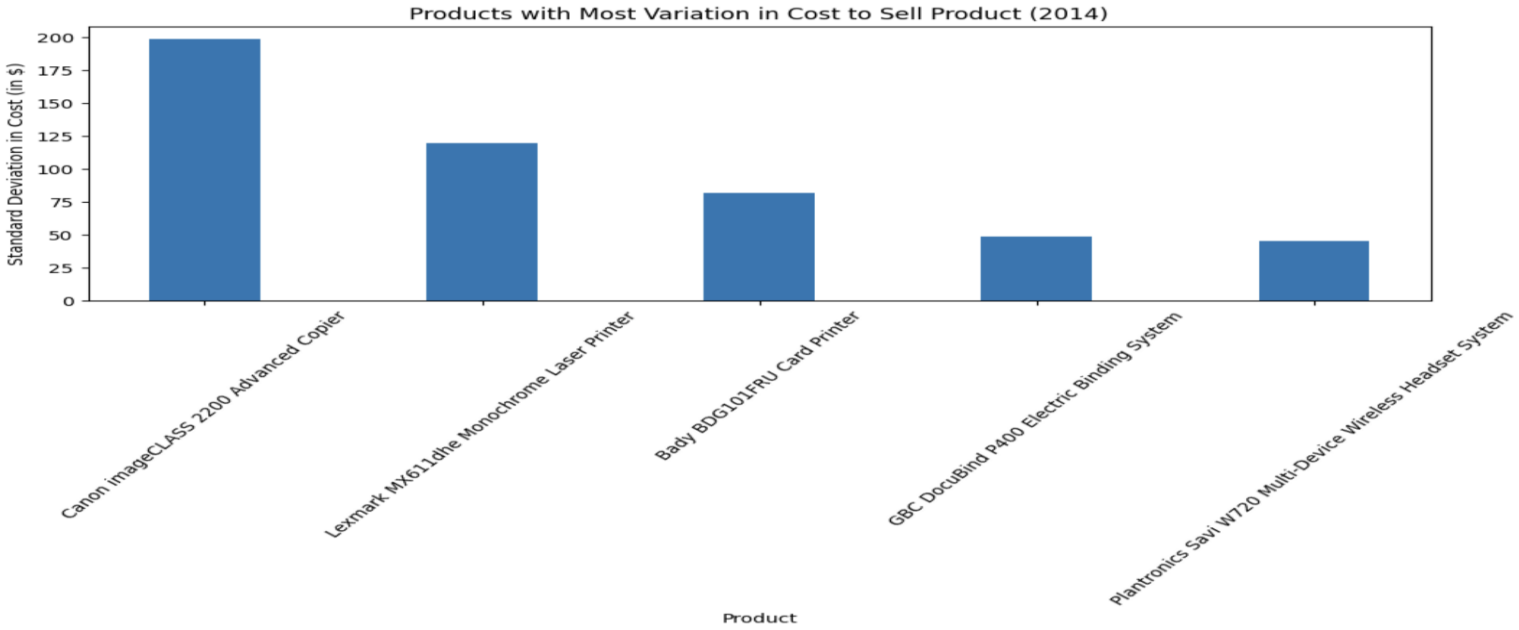
Individual Product Analysis (2014)



Least Profitable Goods 2014

- Each of these products totaled a loss of $2,000 in net sales each and each had a discount applied to their orders of more than 40% on average. The correlation of high discounts and loss in sales is clear.

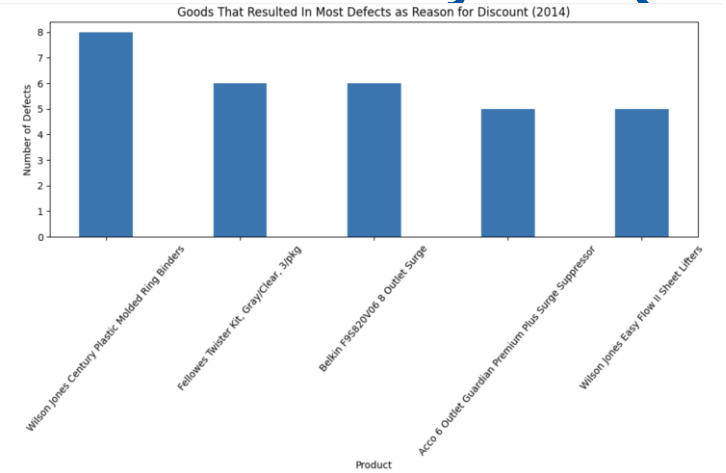
Individual Product Analysis (2014)



Goods Sold in 2014 with Largest Standard Deviations in Cost

- For SS to realize more consistency in sales, the root cause of what is causing these large deviations in cost for SS to sell these products must be discussed and resolved. Over a span of three years, even accounting for inflation, there is no reason for these large deviations in cost to occur.

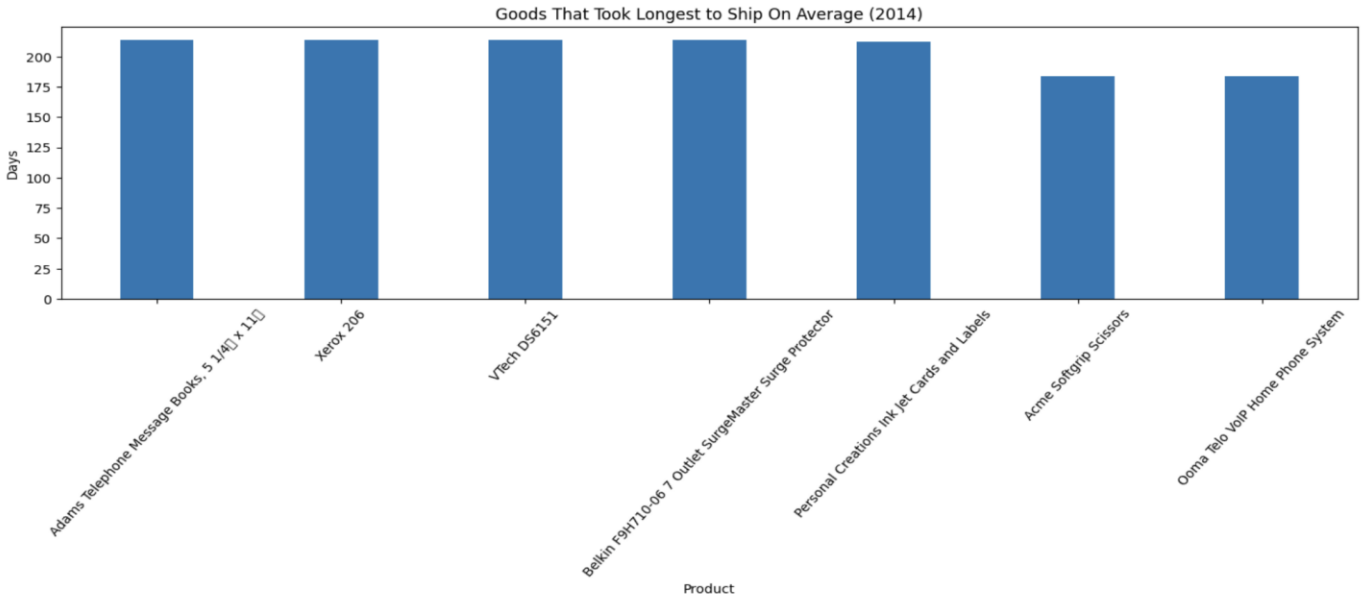
Individual Product Analysis (2014)



Goods that Resulted in Most Defects in 2014

- ***Background detailing of this analysis:*** After closely inspecting each sale, I was able to very closely estimate how many goods were defected in each order (in other words, the customer was likely unhappy with the order), by analyzing how long the good took to deliver and what the discount value was in accordance with the full-retail selling price of the good. Although this was not explicitly detailed in the data, it is fair estimation of the total frequency of occurrence that a customer ordered these products and were discontent with these sale.
  - ***In other words,*** say for example***,*** a product was purchased, and a discount was applied with no delay in shipment or any other valid reason *for* a discount to be applied, and the price of the product before being discounted was $1, with the given discount being less than or equal to the selling price, I recognized the likely reason for the discount being 'customer unsatisfaction', due to the quality of the product. Hence, the 'defect'.

Individual Product Analysis (2014)



Goods sold in 2014 that, on average, presented the largest delays in shipment.

- Urgent that these products are closely scrutinized.
- If shipping delays can't be improved for these products, they must be removed from inventory.

1.3 Additional Analysis: Client

Most Profitable Clients (2011-2014)

Most Profitable Customers (Total Net Sales Generated)

- Again, trying to differentiate customers currently based on "Total Net Sales" is not advised and does not paint a clear picture of the objectives to be realized by SS to reduce variation in sales on a maintainable basis.
  - For example, ***Tamara Chand*** is noted as being the #1 most profitable customer in terms of total net sales, yet $8400 of those $8981 recognized were due to one order. ***However,*** *Perhaps most importantly,* that was one of her last placed orders and she ordered using *Standard Shipping, her orders were shipped 214 days after being placed and yet* ***no discount was administered.***

cust_name	Total Profit Generated	Customer Sales Rank	Total Products Purchased	Total Orders Placed	Last Order Issue	Consequence of Issue	Date of Last Order	Days Since Order
Tamara Chand	8981.3239	1	42	5	date_of_shipment entry error	uncertainty & no profit	2013-11-27	399 days
Raymond Buch	6976.0959	2	71	6	defects	net sales decrease	2014-09-26	96 days
Sanjit Chand	5757.4119	3	87	9	none	none	2014-01-16	349 days
Hunter Lopez	5622.4292	4	50	6	none	none	2014-11-18	43 days
Adrian Barton	5444.8055	5	73	10	none	none	2014-11-20	41 days
Tom Ashbrook	4703.7883	6	36	4	none	none	2014-10-23	69 days
Christopher Martinez	3899.8904	7	34	4	random discount	net sales decrease	2014-11-24	37 days
Keith Dawkins	3038.6254	8	84	12	date_of_shipment entry error	uncertainty & no profit	2014-10-09	83 days
Andy Reiter	2884.6208	9	33	6	random discount	net sales decrease	2014-12-25	6 days
Daniel Raglin	2869.0760	10	57	8	date_of_shipment entry error	uncertainty & no profit	2014-12-08	23 days

Least Profitable Clients (2011-2014)

cust_name	Cumulative Profit Generated	Most Purchase Product Type	Total Orders Placed	Most Recent Order	Days Since Recent Order	Date Since First Order	Days Since First Order Placed
Cindy Stewart	-6626.3895	Machines	6	2014-03-11	295 days	2011-03-26	1376 days
Grant Thornton	-4108.6589	Binders	3	2014-05-11	234 days	2011-06-06	1304 days
Luke Foster	-3583.9770	Binders	7	2014-09-12	110 days	2011-06-12	1298 days
Sharelle Roach	-3333.9144	Binders	5	2014-12-23	8 days	2011-06-14	1296 days
Henry Goldwyn	-2797.9635	Binders	12	2014-08-12	141 days	2011-09-20	1198 days
Nathan Cano	-2204.8072	Binders	6	2014-11-20	41 days	2011-03-05	1397 days
Sean Braxton	-2082.7451	Machines	7	2014-11-14	47 days	2011-07-09	1271 days
Sean Miller	-1980.7393	Paper	5	2014-10-13	79 days	2011-03-18	1384 days
Christine Phan	-1850.3029	Binders	8	2014-07-24	160 days	2011-03-17	1385 days
Natalie Fritzler	-1695.9714	Chairs	7	2014-10-13	79 days	2011-12-02	1125 days

Least Profitable Clients

It's worth considering, should these customers really be avoided?

Professional Consultive Advice

Identifying customers as "least profitable" is premature until supply chain issues have been resolved. These negative net sales are due to SS's irresponsible past sales decisions.

Most Products by Client (2011-2014)

GW Business

cust_name	Total Profit Generated	Customer Sales Rank	Total Products Purchased	Total Orders Placed	Last Order Issue	Consequence of Issue	Date of Last Order	Days Since Order
Jonathan Doherty	1050.2668	95	150	11	date_of_shipment entry error	uncertainty & no profit	2013-10-29	428 days
William Brown	714.3311	150	146	11	date_of_shipment entry error	uncertainty & no profit	2014-12-11	20 days
John Lee	228.9070	395	143	11	date_of_shipment entry error	uncertainty & no profit	2014-11-25	36 days
Paul Prost	1495.0854	48	138	10	defects & COGS	net loss	2014-09-25	97 days
Steven Cartwright	1276.6513	65	133	11	date_of_shipment entry error	uncertainty & no profit	2014-06-27	187 days
Emily Phan	144.9578	483	124	17	none	none	2014-12-19	12 days
Cassandra Brandow	150.2089	480	122	10	none	none	2014-12-09	22 days
Chloris Kastensmidt	141.2831	489	122	13	random discount	net sales decrease	2014-11-22	39 days
Edward Hooks	1393.5154	52	120	12	random discount	net sales decrease	2014-08-18	135 days
Matt Abelman	1240.2266	67	117	11	random discount	net sales decrease	2014-10-28	64 days

Noteworthy Insight

Even considering the extravagant amount of discounts that have (most likely) been applied to these customers, it is clear that there is a relationship between customers ordering more products and those customers generating profitable sales for SS.

Customers ordering more products should be the clients that SS focuses on marketing to and treated as “most profitable”.

Recommended Strategy

Consider reaching out to these customers via email, with a survey, querying what they would like SS to enhance in the business model going forward. These customers will feel valued and will increase the probability they will continue choosing SS for their product needs.

- e.g., An easier to navigate electronic catalog? Increased inventory of some product?

Most Orders by Client (2011-2014)

GW Business

cust_name	Total Profit Generated	Customer Sales Rank	Total Products Purchased	Total Orders Placed	Last Order Issue	Consequence of Issue	Date of Last Order	Days Since Order
Emily Phan	144.9578	483	124	17	none	none	2014-12-19	12 days
Zuschuss Carroll	-1032.1490	777	105	13	date_of_shipment entry error & likely defects ...	net loss	2014-12-06	25 days
Joel Eaton	221.7967	403	90	13	major delayed shipment	lack of discount administered	2014-10-11	81 days
Sally Hughsby	558.4740	200	113	13	date_of_shipment entry error	uncertainty & no profit	2014-10-28	64 days
Chloris Kastensmidt	141.2831	489	122	13	random discount	net sales decrease	2014-11-22	39 days
Patrick Gardner	137.4613	495	83	13	date_of_shipment entry error	uncertainty & no profit	2014-12-11	20 days
Noel Staavos	-234.7714	721	113	13	defects & COGS	net loss	2014-10-07	85 days
Erin Ashbrook	-52.7406	672	87	13	major delayed shipment	net sales decrease	2014-01-08	357 days
Pete Kriz	2038.2676	25	107	12	none	none	2014-12-22	9 days
Lena Creighton	1288.3469	64	93	12	random discount	net loss	2014-12-15	16 days

Cumulative Graph Outlining The Most Individual Orders Placed By Customers Since Sales Have Been Recognized

Although these customers “bottom line” in terms of profit generated, may not look promising at first glance, my advice is to aim towards strengthening relationships with these customers going forward. They have all made purchases relatively recently and once the issues on the backend of sales are resolved they will likely be eager to place more frequent orders.

GW Business

2 Additional Sale Issues & Outcomes

‘Sale Issue’ Dictionary

- ***Random Discount:*** A discount that was applied to a sale for a reason that can’t be classified with the given data; thus, my impression is again that the previous VP in sales distributed these discounts to increase sales volume and customer retention would, without paying attention to profit margins of the goods being sold. <u>These discounts have been the primary reason for loss in sales in 2014.</u>
- ***Major Delayed Shipment:*** Discounts very likely issued because the order took well over 10 days past the standard average for the shipping class the good was sold under.
- ***Date of Shipment Entry Error:*** The order had a “date of shipment” recognized as being BEFORE the day ordered.
- ***COGS:*** The cost of the good was equal to or greater than the price of the good sold. *Note: The chart declares that orders marked with COGS resulted in no LOSS in sales, but all four of these issues in sales resulted in exactly no profit earned.*

Sales Issue	Occurrences (2014)	Occurrences (2011-2014)	Percentage of Occurrences (Ratio of 2014 over 2011-2014)	Profit Lost Due to Issue (2014)
random discount	621	1883	33.0%	26781.54
major delayed shipment	921	2647	34.8%	20134.02
defects	121	409	29.6%	17957.62
defects & COGS	134	399	33.6%	16010.63
date_of_shipment entry error	457	1461	31.3%	6250.51
date_of_shipment entry error & likely defects	48	163	29.4%	5933.53
major delayed shipment & COGS	56	158	35.4%	4182.95
random discount & COGS	81	234	34.6%	2377.65
date_of_shipment entry error & likely defects & COGS	17	76	22.4%	1406.20
date_of_shipment entry error & COGS	20	53	37.7%	342.05
delayed same day delivery	3	8	37.5%	21.51
none	836	2488	33.6%	0.00
COGS	4	15	26.7%	0.00

Outcomes of Sales Issues

Outcome of Sale Issue	Resulting Consequence (2014)	Frequency of Occurences (2011-2014)	Percentage of Occurrences (2014 freq over 2011-2014 freq)
net sales decrease	928	2715	34.2
none	836	2488	33.6
net loss	621	1871	33.2
lack of discount administered	477	1406	33.9
uncertainty & no profit	443	1464	30.3
no profit	14	50	28.0

Detail and Descriptions on "Outcomes" of Sale Issues

Initially, SS had a clientele base of 793 unique customers. As of 2014, this number has decreased to only 693customers.

As illustrated, a primary reason for this is likely due to extremely delayed shipping times; moreover, out of the 3319 recognized sales receipts in 2014 only roughly 25% of them could be confidently marked as having no sale issue.

- ***Lack of Discount Administered:*** These customers experienced EXTREME shipping delays (10 days or more past the expected shipping time) and yet were given no sales discount. Another likely reason some customers have not returned to making purchases from SS.
- ***Uncertainty & No Profit:*** These sales were recorded with shipping dates well before the actual order was placed, making the issue for the discount ultimately impossible to trace. A majority of these sales contained products being sold that cost more to sell than their respective sales revenues.

Table of Contents: **Model Appendix**

1. Presentation Model

   I. Why Shipment Issues?

   II. Why Classification Tree?

   III. Methodology Detail, Key Ratios, & Current Limitations

2. Envisioned Theoretical Model

GW Business

1.1 Presentation Model: Why Prioritize Shipment Issues?

Why Prioritize **Shipment Issues?**

GW Business

Reasoning

- Outside of **(Total) Shipment Issues**, the next most frequent *sales issue*, **random discount,** cumulatively arose in conjunction with 21.15% of SS' sales in 2014.
  - On aggregate, this amounts to less than half of the total percentage of **Total Shipment Issue** occurrences.
- Every other recognized ***sales issue*** makes up a small ratio of SS' total sales transactions.
- **Defects,** can be resolved by enhancing customer relationships; moreover, these issues arose in only roughly 7% of sales.
- **Random discounts** can be easily resolved by understanding that overzealous issuance of discounts to customer orders is not a good business strategy, especially considering profit margins.
- Obtaining the data to understand '**Shipment Issues**', by following the leads provided by the CART model, should naturally lead to understanding most of what is causing **COGS** issues to arise.

Percentages

Sales Issues (2014)	
Total Shipment Issues	45.86%
none	25.19%
random discount	18.71%
defects & COGS	4.04%
defects	3.65%
random discount & COGS	2.44%
COGS	0.12%

1.2 Presentation Model: Why Classification Tree?

Classification Model Selection

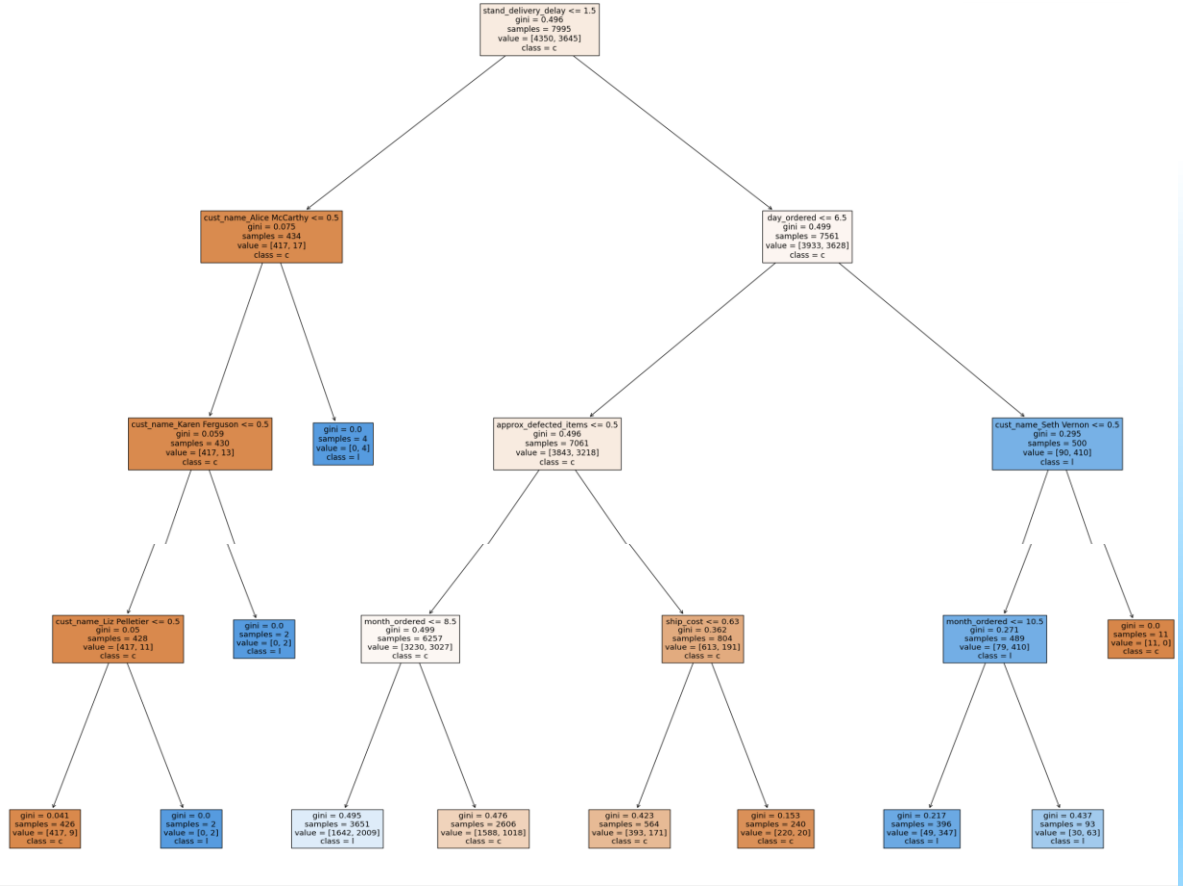
Why a Classification Tree was determined to be the appropriate Machine Learning algorithm to remedy the 'Shipment Issue' in sales

- **Scope of model is too complex for an alternative supervised learning algorithm, especially in the long haul.**
  - **More variables = increased likelihood that model will overfit and prove unreliable**
- **To pinpoint and characterize exactly what is causing 'Shipment Issues' to occur, a binary classification model is undoubtedly optimal.**
- **Most efficient model to implement with a Data Mart, without worry of overfitting causing model to lose usefulness.**
- **Can be built upon and fine-tuned as needed, without much coding effort.**
- **Easy to comprehend visually. Visual classification and splitting methodology crucial for understanding the issue at a deeper level.**
- **Predictive efficiency easily understood.**

GW Business

1.3 Presentation Model: Efficiency, Methodology Detail & Key Ratios

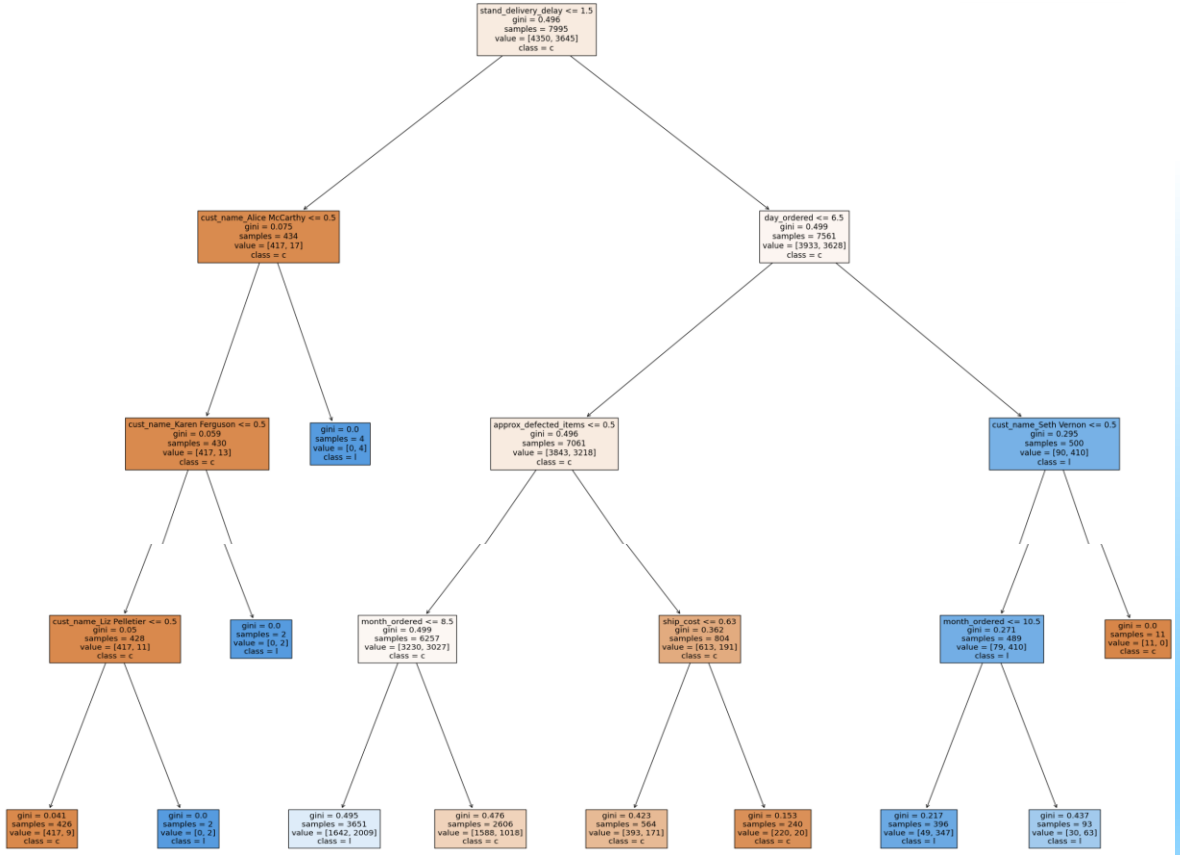
The Model: Classification Tree



Detail on Design & Efficiency

- Model trained on, cumulatively, 7995 transactions.
  - 4350 with no issue in shipment
  - 3645 with issue in shipment
- 80% of data used to train model -> accurately predicted 63.23% of the training data.
- During testing, model was able to predict appropriately for 63.28% of the data.
  - Predictive capabilities improve, definitively, as it learns
  - More transactions -> more accuracy
  - It's important to remember that a large portion of the 'no issue in shipment' observations had other issues in sales. The fact the model is able to discern & predict a good deal above *random guessing*, whilst differentiating these other sales issues *and* being trained with limited data, is applaudable and supports its usefulness going forward.

Analysis of Splitting Methodology



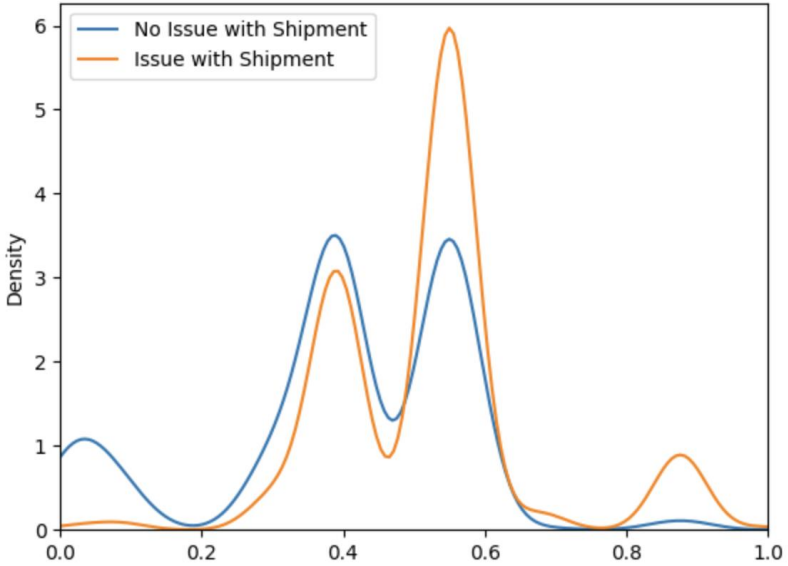
Some Insights Regarding Models Classification Methodology

- Model appears very confident in its classification of certain customers who ordered using 'Same Day' either having or not having experienced an issue with shipment. *E.g., Model classifies a customer named Alice McCarthy, who ordered with Same Day Delivery, as having had 4 orders containing shipping issues.*
- Additionally, model is mostly confident that orders placed on Friday, during the months of January -October, experience a higher-than-normal issue rate with shipment.

This information gives strong indication that other Customers ordering with similar customer demographics and characteristics are experiencing similar problems.

- **The distribution and production centers should be made aware of this issue.**
- **This information provides an avenue to pursue in negating this issue from future orders.**

KDE Plot with Given Limited Data



Visual of Models Current Confidence in Classifying Sales (Predictions)

- Model slightly more confident in inferring **Issues with Shipment**
- Ideally, plot does not display overlap when classifying.
  - Unless provided higher quality data to describe each record, models predictive confidence is not optimal

Cumulative Sales Issues	Aggregated Percentage of Occurrence
major delayed shipment	26.49%
none	24.89%
random discount	18.84%
date_of_shipment entry error	14.62%
defects	4.09%
defects & COGS	3.99%
random discount & COGS	2.34%
date_of_shipment entry error & likely defects	1.63%
major delayed shipment & COGS	1.58%
date_of_shipment entry error & likely defects ...	0.76%
date_of_shipment entry error & COGS	0.53%
COGS	0.15%
delayed same day delivery	0.08%

Reasons for Overlap

- **Once certain sales issues are circumvented (e.g., random discounts, COGS) and/or appropriate data is captured to fully detail an orders logistic information, this model will be able to foreshadow which orders are most likely to cause shipment issues without issue .**

Key Ratios

	Predicted 0	Predicted 1
True 0	653	425
True 1	309	612

```
Recall:   66.45%
Precision:  59.02%
```

Analysis of Confusion Matrix on Testing Performance

- **653 True Negatives:** The model correctly predicted that the transaction didn't have an issue with shipment.
- **612 True Positives:** The model correctly predicted these transactions did have issues with shipment.
- **425 False Positives:** The model incorrectly labeled these transactions as having issues with shipment, when they in fact, did not.
- **309 False Negatives:** The model incorrectly labeled these transactions as not having issues with shipment, when they in fact did.

It is promising that a False Negative occurred with the least frequency because this carries the greatest weight in purpose of model!

Analysis of Recall & Precision

- **Recall:** Model was able to successfully predict 66.45 percent of all the transactions that were defined as having had an 'Issues with Shipment', during testing.
- **Precision:** Metric detailing how accurate model was in its prediction of a transaction being labeled as having had an 'Issues with Shipment', when model labeled transaction as having an 'Issue with Shipment'.

Model Limitations

- **Model's current predictive capabilities and confidence in discerning 'Shipment Issues' is moderate.**
- **Model must make use of all transaction history to provide only a reasonable level of predictability.**
  - *Some of these transactions may not be relevant to SS' present business environment.*
- **Model not provided enough *quality data*:** data that efficiently provides model with *supply chain-related* attributes, to classify *supply chain-related* issues. This data needs to be obtained and that may take some time.
- **For optimum value, SS will likely have to interconnect the model with a data mart to properly train the model and build upon its current state**.

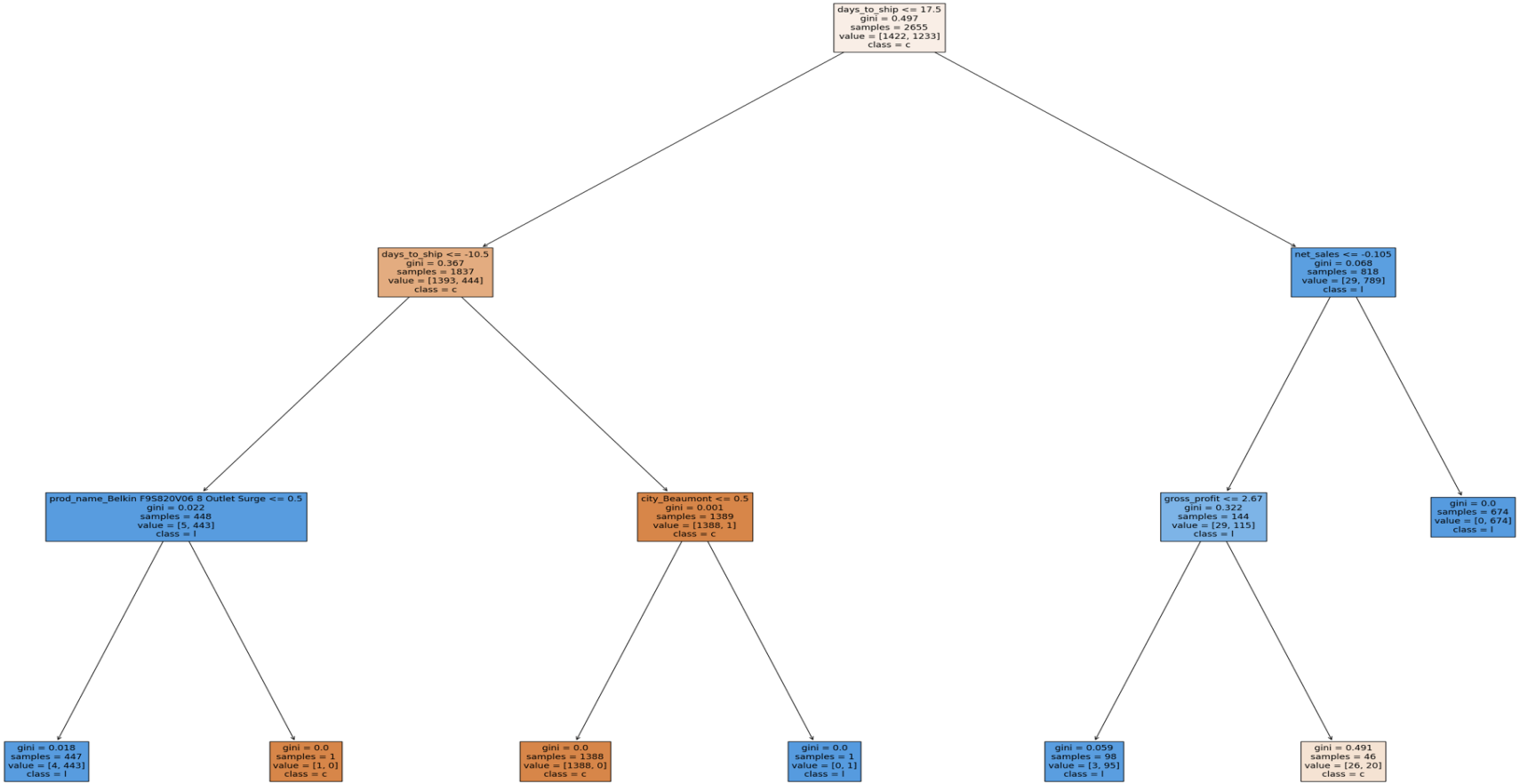
GW Business

2

Potential Model

(illustrates Presentation Model's potential, given additional data inputs)

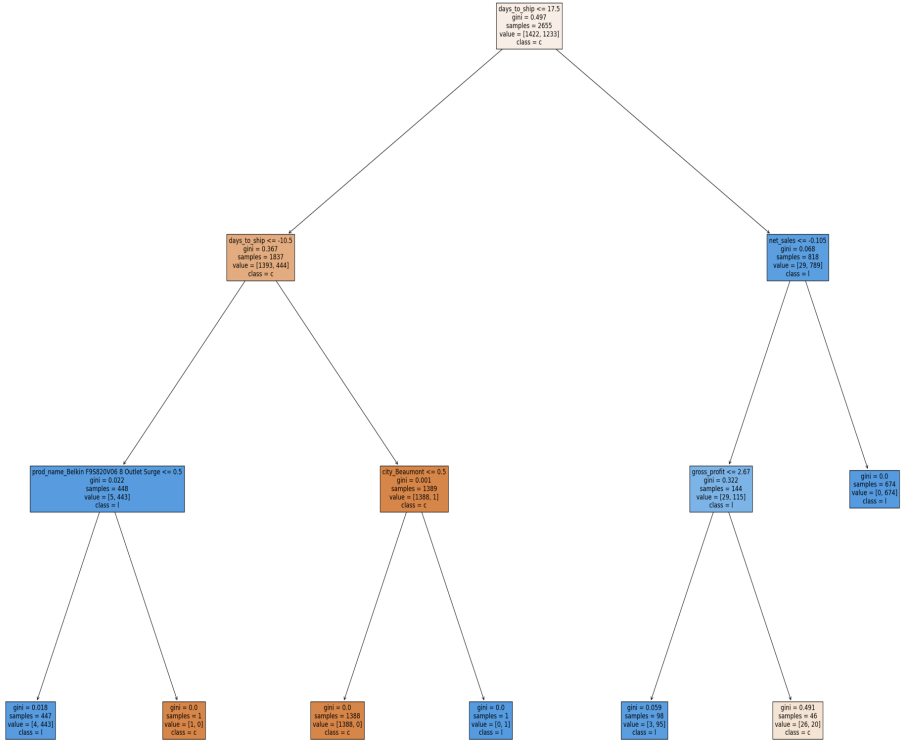
The Model: Classification Tree



Holistic Description of Model

<u>Classification Tree</u> to Predict Probability that a Sale will lead to a *Shipment Issue*

- **Objective:** To illustrate "primary model's" potential, as more 'Shipment data' is attributed to each transaction.
- **Data Used in Training & Testing:** Each transaction that occurred in the most recent sales period strictly (2014). ***Includes days took to ship for each transaction.***
- **Binary Classification:** Each observation was labeled with either a 0 (No Shipment Issue) or a 1 (Shipment Issue) for training. These output values were hidden from the model during testing.
- **Limiting Complexity of Model & Preventing Overfitting:**
  - Manual Dimensionality Reduction of variables likely to be recognized as noise (e.g., Order #)
  - Model Depth capped at 3. Minimum sample size to split at 20.



Key Ratios: Less input, higher quality variables

	Predicted 0	Predicted 1
True 0	367	8
True 1	10	279

Recall: 96.54%
Precision: 97.21%

Analysis of Confusion Matrix on Testing Performance

- **367 True Negatives:** The model correctly predicted that the transaction didn't have an issue with shipment.
- **279 True Positives:** The model correctly predicted these transactions did have issues with shipment.
- **8 False Positives:** The model incorrectly labeled these transactions as having issues with shipment, when they in fact, did not.
- **10 False Negatives:** The model incorrectly labeled these transactions as not having issues with shipment, when they in fact did.

Analysis of Recall & Precision

- **Recall:** Model was able to successfully predict 96.54 percent of all the transactions that were defined as having had an 'Issues with Shipment', during testing.
- **Precision:** Measures how accurate model was in its prediction of a transaction being labeled as having had an 'Issues with Shipment', when model label transaction as having an 'Issue with Shipment'.

Slides Carnival

Free templates for all your presentation needs

For PowerPoint and Google Slides

100% free for personal or commercial use

Ready to use, professional and customizable

Blow your audience away with attractive visuals