

Order Analysis - Northwind

John Kline

12/13/19

Framework & Data Sources

GOALS

Evaluate several questions using statistical hypothesis testing

- *Does discount size affect order size? If so, at what levels?*
- *Does discount size affect order revenue?*
- *Should we consolidate a distribution center to serve both German and American customers (i.e. if German and American customer purchase the same types of this will be an efficient consolidation)?*
- *Do certain employees drive larger order sizes than others? Or is order frequency the key driving factor behind total order volume?*

METHODS

Various statistical analyses in Python

DATA SOURCES

- *Northwind SQL database*

Data Background

1

Data Detail:

- ~2,150 products ordered to fictional company Northwind

2

Statistical tests

- One-Way ANOVA – test for whether any factor (e.g. any discount value) has a significant effect on the outcome variable (e.g. order size)
- Tukey's HSD – identifies which factors have significant effects
- Chi-Sq – evaluates whether category relative frequencies are significantly different from one another

3

Feature Construction

- Created values reflecting the difference in revenue and quantity of orders compared to the average revenue or quantity of orders for each product

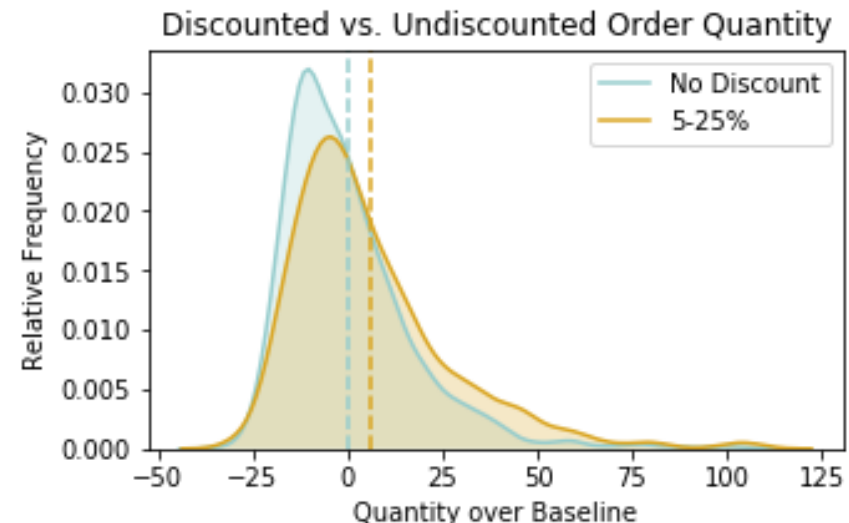
Does discount level affect order size?

Findings and Recommendations

- **Discount level is highly associated with larger order size – associated with up to a 25% increase in per-order quantities**
 - All order quantities except 10% had a statistically significant effect on order amount
 - 5%, 15%, 20%, and 25% discount levels had statistically indistinguishable impacts on order quantity
- **The optimal discount level is 5% - this should have the lowest impact on revenue and a high impact on order quantity**

Discount	Order Avg.	Discount Impact	Impact %	P-Value
0%	21.7	-	-	-
5%	28.0	+6.2	+23%	.001
10%	25.2	+4.0	+16%	.08
15%	28.4	+6.7	+24%	.001
20%	27.0	+5.8	+22%	.003
25%	28.2	+7.0	+25%	.001

P-values represent the results from a Tukey HSD test following positive one-way ANOVA



John Kline

Non-Technical Presentation

Flatiron School - Data Science Program

Do discounts affect order revenue?

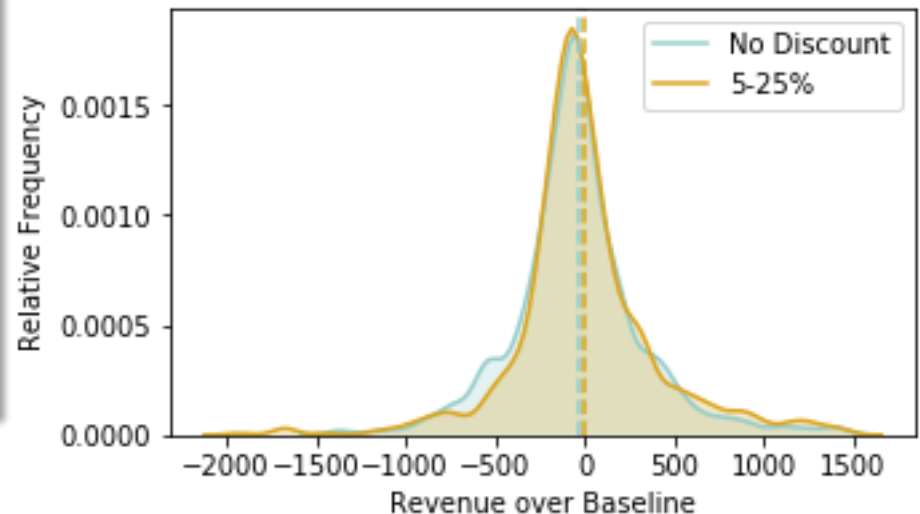
Findings and Recommendations

- **Discount level is NOT associated with a statistically significant increase in revenue per order**
 - The 5% discount level was the closest to having a significant difference in mean order revenue
 - Unit price varied significantly within each discount level driving uncaptured price variation
- **Discounts' potential impacts are drowned out by unit price variation not driven by discounts**

Discount	Discount Impact	Impact %	P-Value
0%	-	-	-
5%	+ \$124	+22%	.19
10%	- \$3	-0%	.9
15%	+ \$65	+11%	.85
20%	+ \$43	+8%	.9
25%	+ \$46	+8%	.9

P-values represent the results from a Tukey HSD test following negative one-way ANOVA

Discounted vs. Undiscounted Order Revenue



John Kline

Non-Technical Presentation

Flatiron School - Data Science Program

Should we combine US and GER distribution?

Findings and Recommendations

Key Question – Can Northwind combine distribution centers for scale (i.e. are US and GER customers purchasing patterns similar)?

- Product category
- Individual products
- **Potentially – if Category is the driving factor in dist. center design**
 - Half of categories are very similar
 - Half of categories differ materially
- **No – if Product is the driving factor in distribution center design**
 - Products within every category vary significantly in order pattern between the two countries

Category	USA vol.	Ger. vol.	Significance
Beverage	1620	1690	0.45
Condiment	870	820	0.45
Confection	1600	1600	0.45
Grains / Cereal	800	730	0.45
Dairy	1800	1560	.001
Produce	490	320	.001
Meat	560	1160	.001
Seafood	1400	1520	.001

P-values represent the results from two chi-sq. tests of order quantity tables

Category	Products	Significance
Beverage	All	<.001
Condiment	All	<.001
Confection	All	<.001
Grains / Cereal	All	<.001
Dairy	All	<.001
Produce	All	<.001
Meat	All	<.001
Seafood	All	<.001

P-values represent the results from two chi-sq. tests of order quantity tables across products

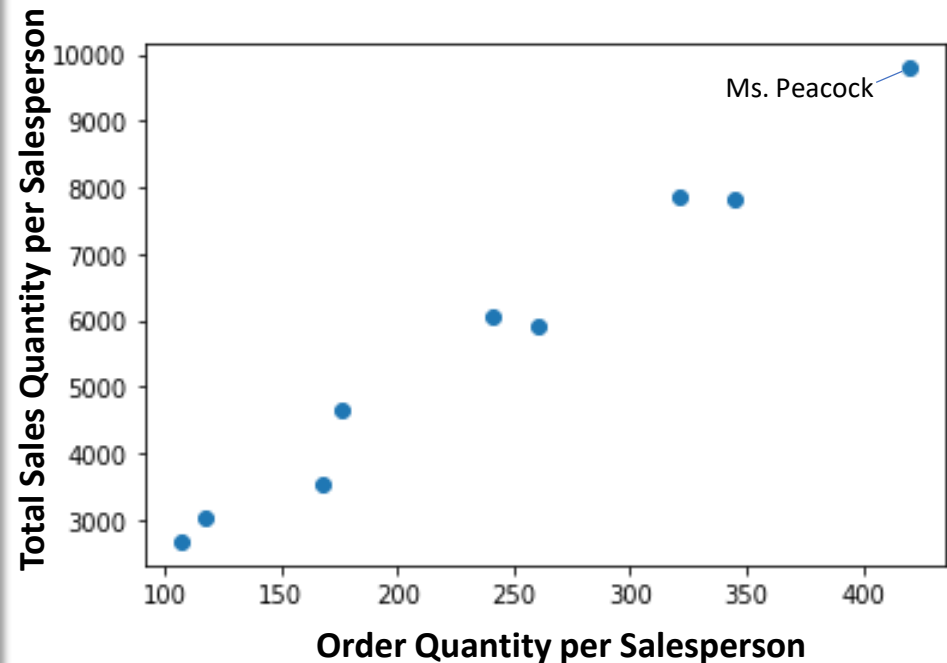
What metric should we tie salesperson comp to?

Findings and Recommendations

- **High difference in total sales volume between employees**
 - Range from 2,700 to 10,000 total order volume
 - Range form 250K to 80K total order revenue
- **No statistically significant difference in per-order quantity or per-order revenue between any pair of sales employees**
- **Total sales volume variation between employees is explained by difference in the number of orders**

Measure	Tukeys HSD
Revenue	Fail, all pairs
Quantity	Fail, all pairs

Number of Orders vs. Total Quantity Sold



Client Recommendation Summary

Recommendations

1. Discounts drive improved per- order volume – however, prices vary for unexplained reasons such that discounts do not drive improved per-order revenue
 - **Improve price discipline outside of discounts**
 - **Re-evaluate discount impact on top of stable pricing structure**
2. Is product category or product type a more important driver of distribution center design?
 - **If the former – half of product categories are similar in order patterns between the US and GER, and this could be a basis for consolidating**
 - **If the latter – do not consolidate, extensive variation in product patterns**
3. **Salespeople do not have different order sizes – the key factor is total order number – therefore sales comp. structure should incentivize total order frequency primarily**

Next Analytic Steps

1. Detailed review of pricing variation outside of discounts

- Is this driven by regional pricing differences?
- Are certain employees more associated with pricing variations (e.g. execs have more flexibility in base pricing?)
- How does price variability differ between products

Discussion & Questions

Thanks!

Appendix

Employee Sales Records

Employee	Total Quantity	Total Revenue
Peacock	10,000	250,000
Leverling	5,900	210,000
Davolio	7,800	200,000
Fuller	6,100	180,000
King	4,600	140,000
Callahan	5,900	130,000
Dodsworth	2,700	80,000
Suyama	3,500	80,000
Buchanan	3,000	80,000