

Northwind Data Analysis

An Executive Summary of Four Statistical Hypothesis Tests

By: TJ Whipple
Flatiron School Project
January 2020

Project Focus

Using the Northwind Dataset from Microsoft, a statistical analysis was performed. The focus of the following hypothesis tests were to increase the Northwind business operating cost through speed and efficiency. We looked at the following features:

1. Effect of Discount on Order Quantity
2. Most profitable sales Regions
3. Unit Price versus Order Quantity
4. Shipping Company Order Speed

Test 1 - Order Discounts

Does discount have an effect on the quantity of an order amount?

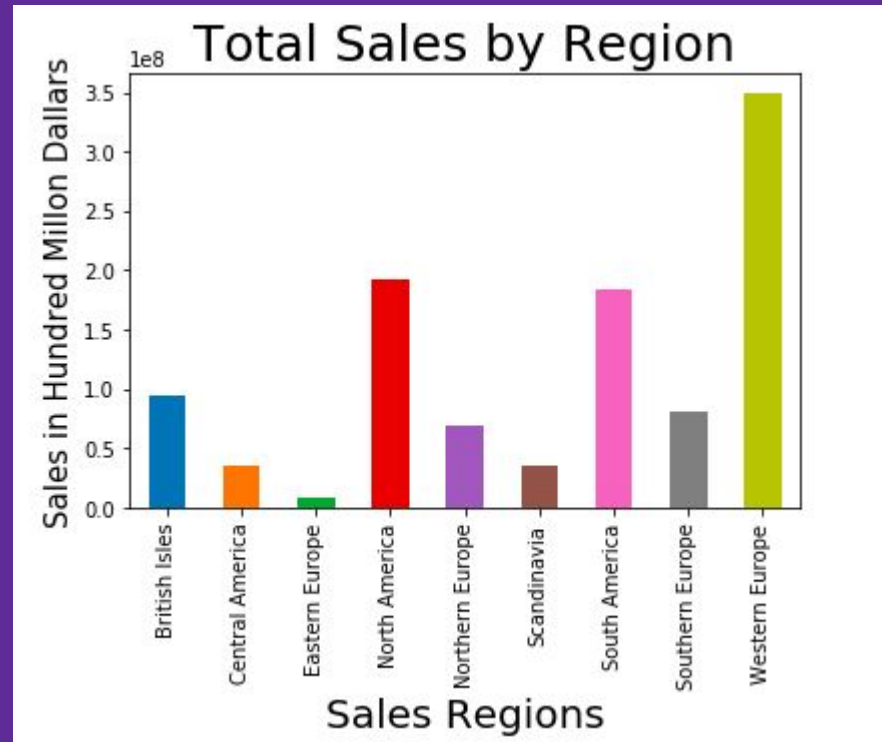


- As shown by the bar graph to the left, discounted orders had higher total product quantities than orders without a discount.
- 24 was the average sized order.
- Discounted orders averaged 27.
- 130 was the largest order size.

Test 2 - Order Regions

Which regions had the biggest orders and the most sales?

- The Region is an important aspect of sales.
- Western Europe had the most sales - around 595,000 orders for about \$349,000,000.
- Eastern Europe had the least - around 15,000 for \$8,860,000.



Test 3 - Unit Price

Does the unit price affect the quantity of order? Are more expensive items ordered as much as cheaper products?

UnitPrice	ProductName	Total_Quantity
18.0	Chai	2382
18.0	Chartreuse verte	2382
18.0	Lakkalikööri	2382
18.0	Steeleye Stout	2382
10.0	Aniseed Syrup	1843
10.0	Longlife Tofu	1843
10.0	Sir Rodney's Scones	1843
12.5	Gorgonzola Telino	1430
12.5	Scottish Longbreads	1430
14.0	Laughing Lumberjack Lager	1277

- Unit Price did not have an effect on the quantity of products ordered.
- The table on the right shows the top ten ordered products by quantity.
- The majority had a price of \$18.00
- The average price was around \$23.00
- The most expensive was \$263.50

Test 4 - Shipping Companies

Does the shipping company have an affect on the order processing speed? Which shipping company is the fastest?

- Federal Express was the fastest with the lowest average.
- United had the most overall orders.
- Nevertheless, the difference in speed between companies was negligible.



Summary

Discounted items had a significant impact on total quantity of orders. This was especially true with larger discounts more than smaller ones.

Unit Price was not an important feature on order quantity though it seems plausible that more expensive items might be less marketable.

Sales Region results were inconclusive but there is evidence that showed certain areas to be more profitable than others.

Shipping Company also proved to be of minor significance. Either the average time to ship out was similar enough between companies or the overall number of orders made the different options negligible.

Future Work

Discount - *I would like to see a control group of items where no discounts are offered and then applied, to validate sales of different discounts.*

Unit Price - *It would be helpful to have more information on the actual cost of items to be able to compare specific unit prices and markup values to overall net profit.*

Sales Region - *I recommend focusing on the countries with the least number of total orders. Again, use a control group that receives no additional sales support or marketing versus another country and gather data on emerging opportunities.*

Shipping - *The most obvious future work would be to include an arrival date for all orders. As well as confirmation that the order was successfully received as intended.*

Thank you for your time.

Questions or Comments?

Sources:

Microsoft Northwind Dataset and Flatiron School

Python Pandas, Jupyter Notebook, GitHub, Matplotlib

Presentation By: TJ Whipple