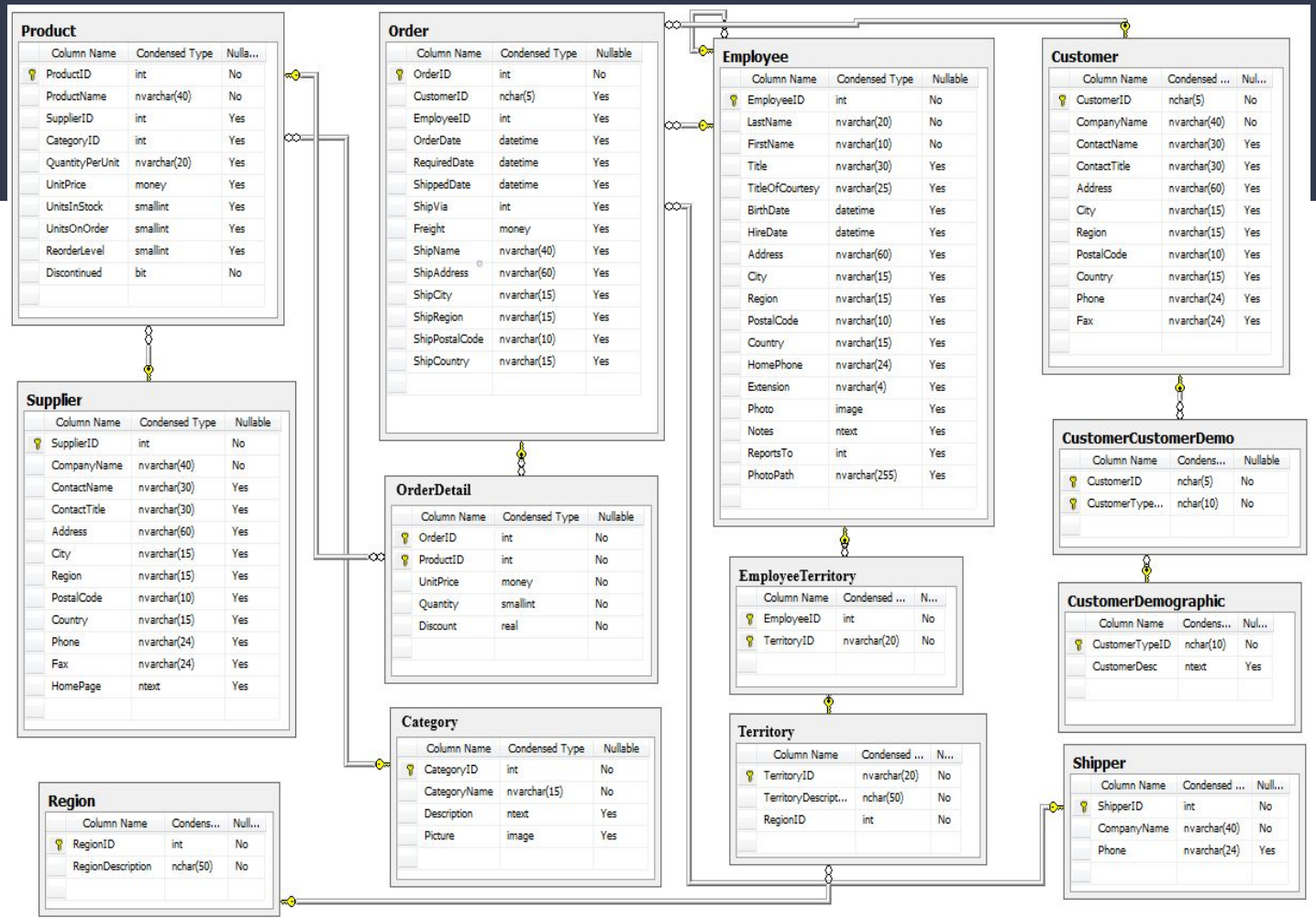
NorthWind Statistical Analysis

Susanna Han

STEPS

- Explored data
- Created 4 hypothesis’
- Connect tables for correlation
- Pulled information
- Ran tests through the grouped data.
- Reject/Accept hypothesis.



Tests applied on the data...

D'Agostino-Pearson's normality test - “is a goodness-of-fit measure of departure from **normality**, that is the test aims to establish whether or not the given sample comes from a **normally distributed population.**”

Levene's Test - “is a statistic used to assess the **equality of variances** for a variable calculated for two or more groups.”

Kruskal-Wallis - **“**used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.”

Statistical Power - is the probability that the test correctly rejects the null hypothesis.

Tukey's Pairwise - “used to test differences among sample means for significance.”

- Does *discount* have a statistically significant effect on the **number** of a product in an order?
- Does *discount* have a statistically significant effect on the **total amount** spent in an order?

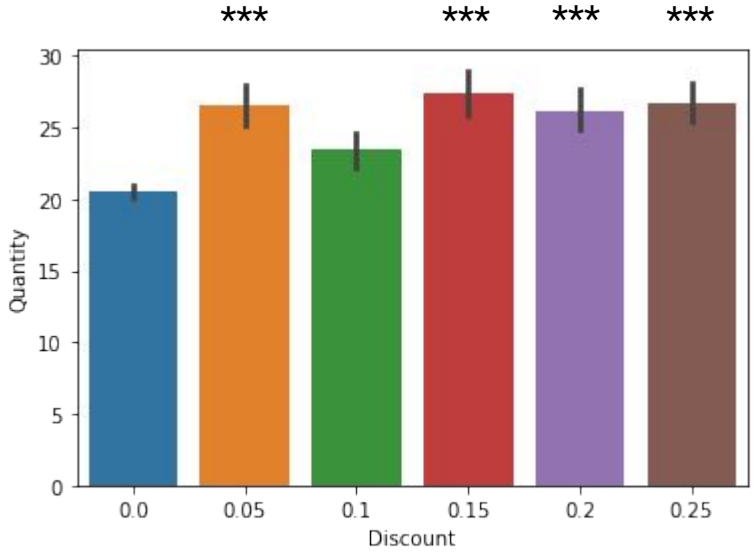
- Does the *ship region* have a statistically significant effect on the **product quantity sold** in an order?
- Does the *ship region* have a statistically significant effect on the **total amount spent** in an order?

- Does *discount* have a statistically significant effect on the quantity amount of a product sold in an order?

yes!

*** = p-value < 0.01

Discounting a product will **increase** the *quantity amount sold in an order*

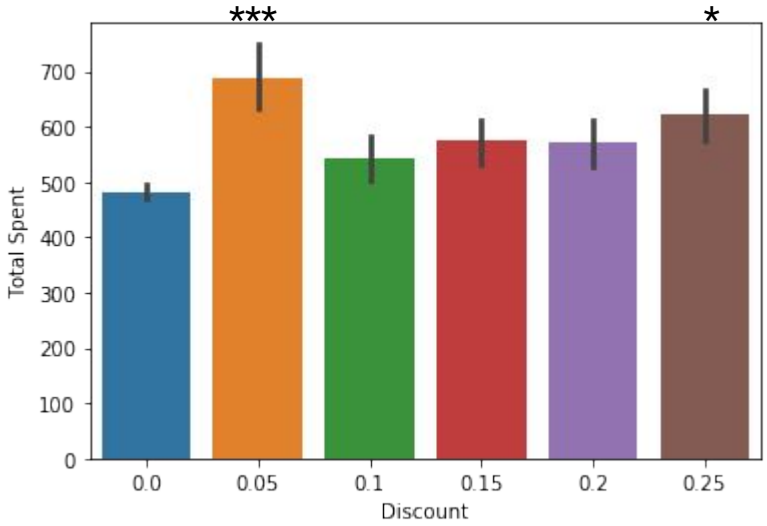


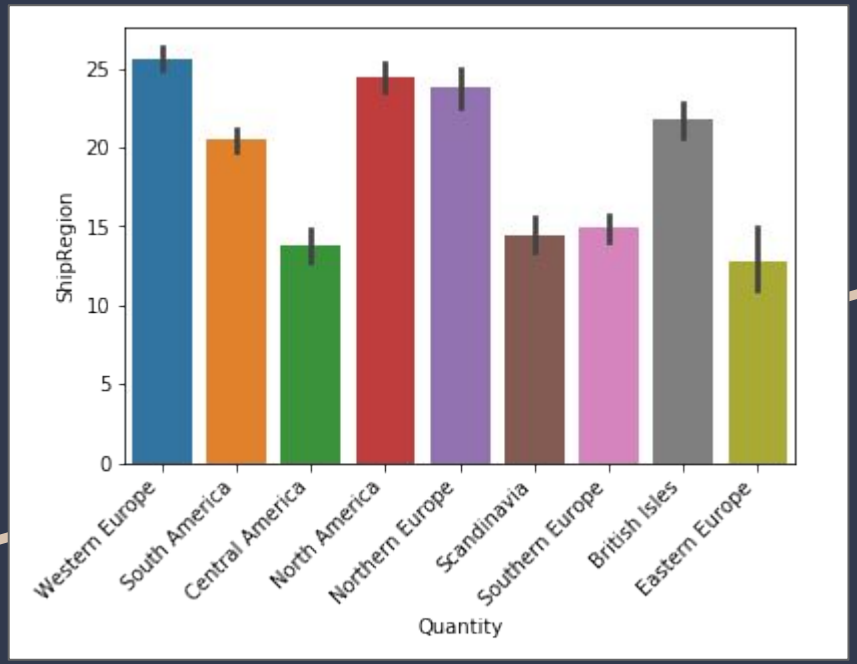
- Does *discount* have a statistically significant effect on the **total amount** spent in an order?

yes!

* = p-value < 0.05
*** = p-value <0.01

Discounting a product at 5% or 25% will **increase** the *total amount spent in an order.*





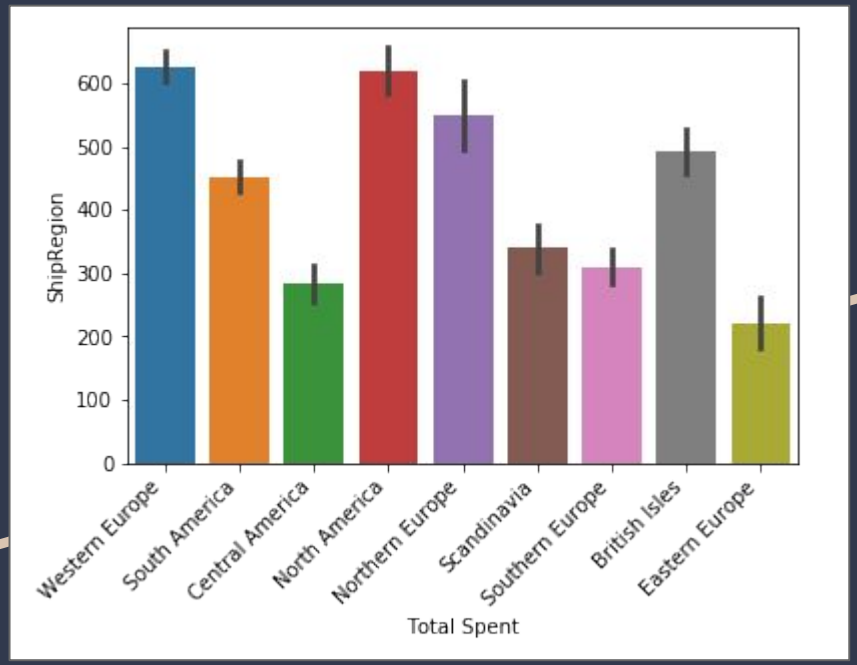
- Does the *ship region* have a statistically significant effect on the product **quantity sold** in an order?

yes!

(Except Eastern Europe)

TOP 3 SUPPLIER REGIONS

1. Western Europe
2. North America
3. Northern Europe



- Does the *ship region* have a statistically significant effect on the **total amount spent** in an order?

yes!

Except Eastern Europe

Increasing marketing efforts in Eastern Europe can help boost sales in that region.

Recommendations

- Discounting a product will **increase** the *quantity amount bought* and also increase the *total amount spent* in an order. A 5% discount is **NOT** significantly different than a 25% discount.
  - **Discount products at 5%** to *increase* sales.

- The TOP 3 ship regions with the most sales in total amount spent and quantity of products sold in an order are in *Western/Northern Europe* and *North America*.
  - **Increase product suppliers in those regions.**

- The only ship region that is **NOT** statistically significant in increasing the total amount spent and quantity of products sold is in *Eastern Europe*.
  - ***Consider increasing marketing efforts in that region.***

Future Work

- Compare different products using the ProductID and the quantity sold to see which products are doing well and which are not.
- Compare what products are correlated with where the customer regions are using the ShipRegion. To see which products are popular in which region.
- Compare different suppliers to see which supplier region is bringing in more money for the company.

Thank You!

Susanna Han