

*Flatiron Module 2
Project: Hypothesis
Testing of the
Northwind Database*





Hello!

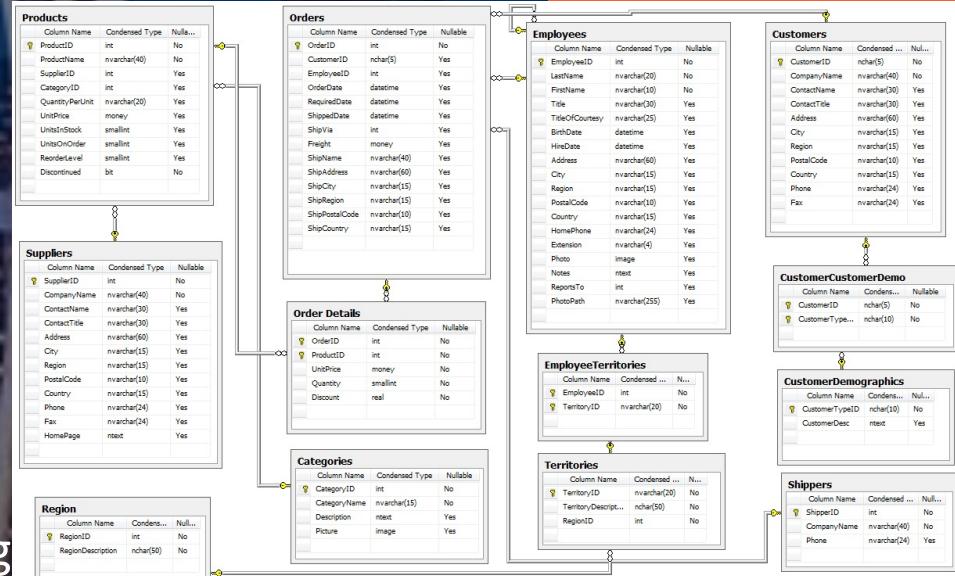
My name is Kiarash Ahmadi. I'm an aspiring Data Scientist with Flatiron Academy. I am also pursuing a Master's in Computational Science & Engineering at Georgia Tech. My field of work is in the building energy modeling industry.





Project Background

- Northwind is a fictional database made by Microsoft
- Formulate questions that will be answered via hypothesis testing
- Analysis of data is done beforehand to determine appropriate test by evaluating different parameters of the dataset





Questions to be answered

1. Does discount amount have a statistically significant effect on the quantity of a product in an order? If so, at what level(s) of discount??

2. Does customer region have a statistically significant effect on the quantity and/or price of a product in an order?

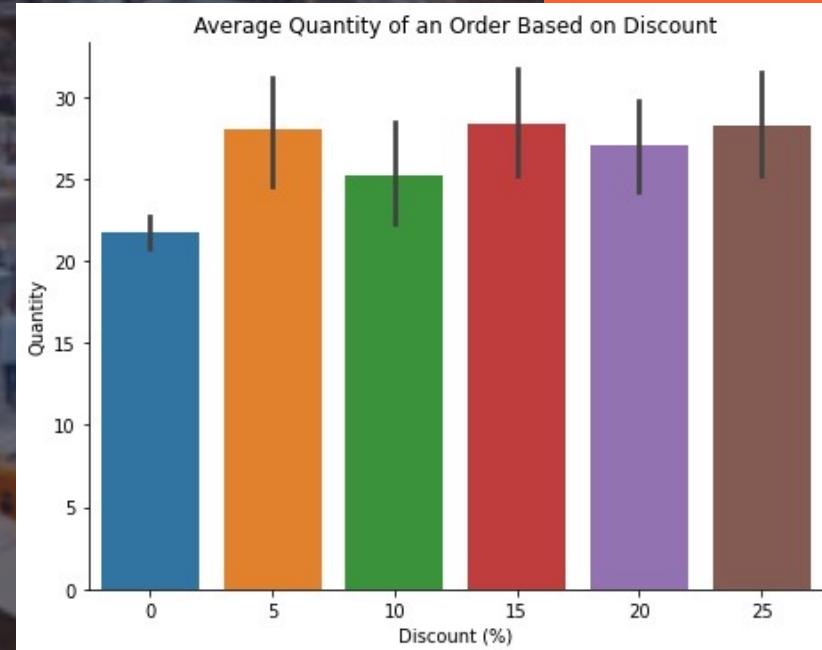
3. Does the shipping company have a statistically significant effect on the total price of an order?

4. Do categories have a statistically significant effect on the price of products in an order?



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

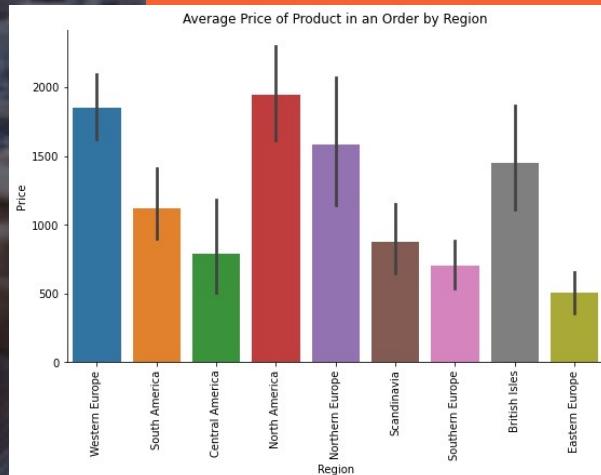
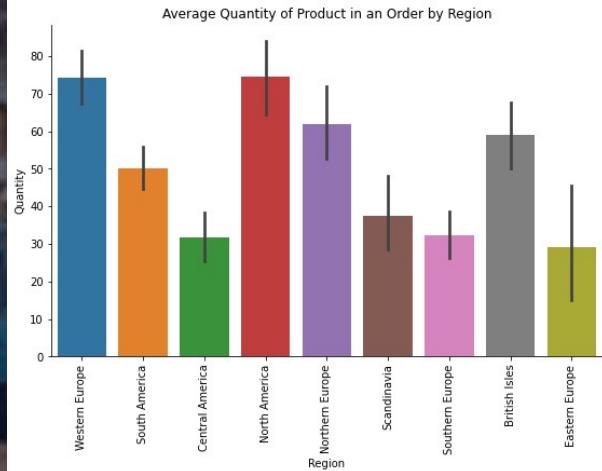
- It was found that discounts do lead to a significant difference in quantity of a product order
- More items should be discounted at 15% and 25% as it was found that these 2 discount groups have the largest effect
- Welch T-test and ANOVA was used





Does customer region have a statistically significant effect on the quantity and/or price of a product in an order?

- Use of Welch T-test and ANOVA led to conclusion that region does have a statistically significant effect on both quantity and price
- Thus, more focus can be put on regions where the average quantity and price in an order is lower:
 - Eastern Europe
 - Southern Europe
 - Scandinavia
 - Central America





Does the shipping company have a statistically significant effect on the total price of an order?

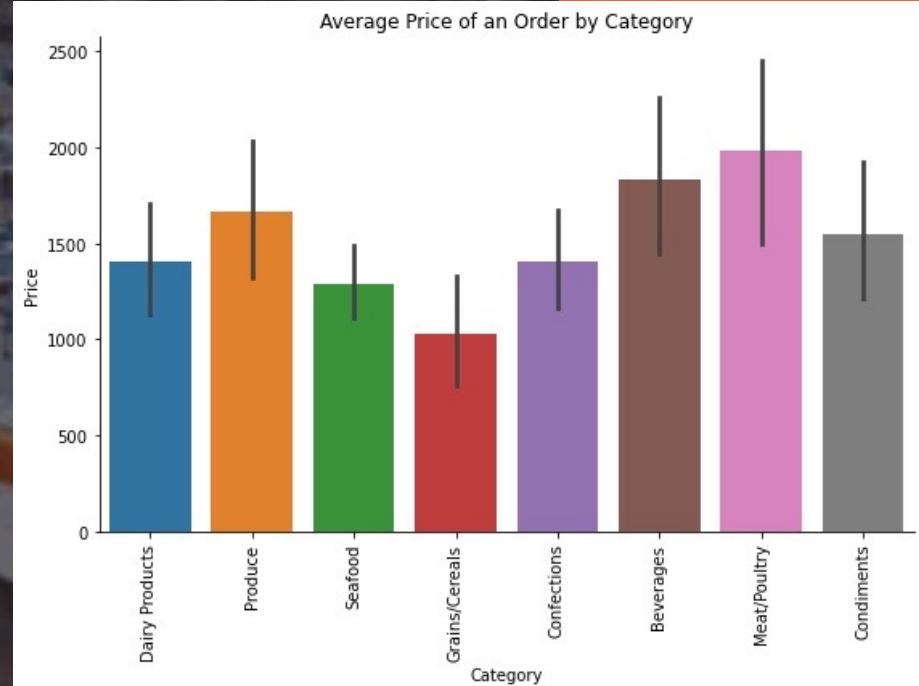
- It was found that shipping company used in an order does not have a significant effect on the average price of an order
- Thus, no focus should be put on using one shipping company over the other
- These results were gained using a Welch T-test and ANOVA





Do categories have a statistically significant effect on the price of products in an order?

- Categories of products within an order have a significant effect on the average price of an order
- Therefore, categories that are lower in price should have more focus via marketing and other methods in efforts to increase their price.
These categories are:
 - Seafood
 - Grains and Cereal





Recommendations

Discount

15% and 25% discount groups have the most impact on quantity ordered

Region

Focus on increasing average price in regions with lower prices and quantities ordered

Shipper

Shipping company has no statistically significant effect thus no focus should be put on one company over another

Category

Categories that are lower in average price should have more focus via marketing efforts to increase their total price

Future Work

Research Test Types

Research different tests outside of Flatiron curriculum that could be used to test hypotheses .

Resampling Methods

Perform resampling techniques on data to ensure same sample size among samples in question so that different hypothesis tests can be used.

Data Transformation

Transform the data in order to make it more normal. Would have to evaluate how this affects the interpretability of the results.



Thanks!

Any questions?

You can find me at:

- mkiah09@gmail.com

