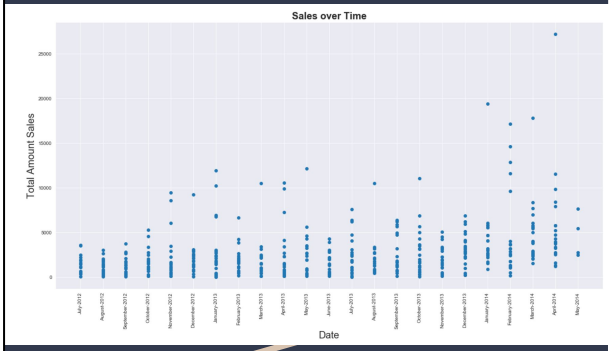
Strategies for Northwind

By: Isabella Lindgren

Part-Time Data Science Student

Hello, my name is Isabella Lindgren and today I will be presenting potential business strategies for Northwind Traders Food Company

Background



Increase Profitability

1. Discounts
2. Employees
3. Customer sales
4. Freight

The dataset we are analyzing for Northwind ranges from 2012 to 2014. As we can see, the amount of total revenue generated is steadily increasing over this time period. So how can we increase the profitability of this growing company? We will focus on four main areas of the company in order to evaluate potential strategies in increasing overall profit.

1. Discounts - Do discounts affect the quantity of items purchased and at what levels of discount?
2. Employee - Does employee location affect the amount of revenue brought in?
3. Customer sales- Do customers who place orders more frequently also spend more total money?
4. Freight- Is there a significant difference in freighting costs from region to region?

Methodology

- SQL Queries
- Hypothesis Testing

To investigate this data set, I utilized SQL queries to obtain relevant data and then used various methods of hypothesis testing to accurately answer our proposed questions

Hypothesis testing - is the statistical process of creating a clear, testable question, analyzing the relevant data, and forming a conclusion which answers our original question. For testing, we form two hypotheses: the alternative hypothesis and the null hypothesis. The null hypothesis is a statement reflecting no observed effect in our experiment, while the alternative hypothesis directly contradicts the null.

Discounts

Question: Do discounts affect the quantity of products ordered? If so, what levels of discount?

Findings/Conclusions:

- Significant difference between the quantity of discounted products vs non-discounted
- No difference between discount levels

Recommendation:

- Reduce discounts

The first question we are investigating is ‘Do discounts affect the quantity of products ordered?’

Why is this important?

- Goal: Optimize Profit
- We don’t want to discount items unnecessarily

Findings:

- We found that there is a small significant effect of discount on the quantity of products ordered
- But when looking at the levels of discount, there was no significant difference on the quantity of product ordered between the discount levels

Recommendation:

- We can give a lower discount (let’s say 5% instead of 25%) and still have the same effect on quantity ordered

Employee Location

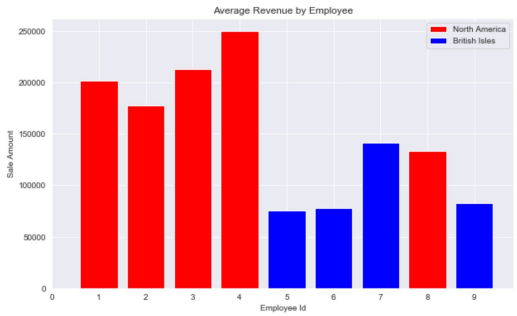
Question: Which employee region brings in more revenue?

Findings/Conclusions:

- North America brings in significantly more revenue than UK
- More employees? Less competition?

Recommendations:

- Investing in North American employee infrastructure



The next area we investigated was Employee location. There are only two locations: North America and the UK.

- We concluded that there is a large, significant difference between the amount of total revenue brought in by US employees compared to UK employees.
- This partly may be due to Northwind being more established in the US compared to the UK or there is more competition in the specialty food trading industry in the UK.

Recommendations:

- Investing in NA employee infrastructure: open more locations
- Training program in UK

Frequency of Sales

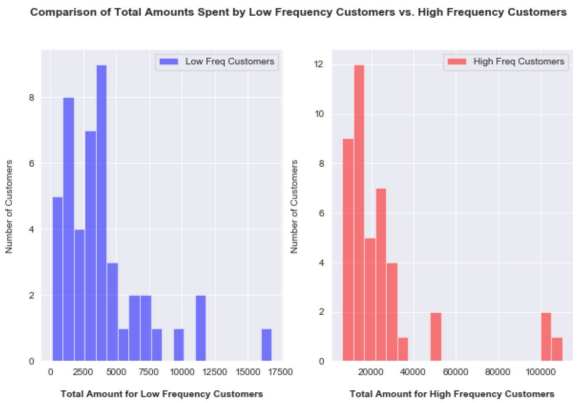
Question: Do customers who order more frequently also generate more sales?

Findings/Conclusions:

- There is a significant effect of order frequency on the total amount spent by customers

Recommendations:

- Membership Program
- Aggressive Marketing



Question - Do customers who order more frequently also generate more sales?

We found that there is a significant difference in the total amount spent by customers who place orders more frequently compared to customers who order less frequently.
Bias in our data: People who have been customers for a longer period of time would theoretically have more orders than people who became customers more recently. To reduce this bias, we will determine when customers placed their first order and sample from that month onwards.

Recommendations:

- We could create a 'membership' program that gives discounts or benefits to frequent shoppers
- Invest in aggressive marketing campaigns so people will purchase more frequently

Freight Costs

Question: Is there a significant difference in freight cost per region?

Findings/Conclusions:

- Significant difference between the freight costs
  - N America vs. S America
  - N America vs. S Europe
  - S Europe vs. W Europe
- Speedy Express significantly lower cost than United Package

Recommendations:

- New shipping warehouses in N America and W Europe
- Increase use of Speedy Express

The last area we investigated was Freight Costs. We looked at if there is a significant difference in freighting cost per region.

We looked at the total freight cost of each region and compared them with each other and determined that there was a significant difference between

- North America and South America
- North America and Southern Europe
- Southern Europe and Western Europe

We further investigated the difference in freight costs between shipping companies and found speedy express to have significantly lower cost than United Package.

Recommendations:

- May be beneficial to open shipping warehouses in North America and Western Europe in the future in order to decrease freight distance and cost for those high revenue regions.
- We may increase profit margins by freighting more of our products using Speedy Express

Recommendations

- Reduce discounts
- Invest in North America employee infrastructure
- Benefits for Frequent Customers
- Look into opening shipping warehouses in North America and Western Europe

Brief review of overall recommendations

Future Work

- Which products would benefit most from being discounted?
- Larger, more current dataset
- Rate of new customer growth, customer retention
- Determine if shipping company has an impact on the actual revenue

Thank you

Questions?

Appendix:

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
0.0	0.01	-19.7153	-80.3306	40.9001	False
0.0	0.02	-19.7153	-62.593	23.1625	False
0.0	0.03	-20.0486	-55.0714	14.9742	False
0.0	0.04	-20.7153	-81.3306	39.9001	False
0.0	0.05	6.2955	1.5381	11.053	True
0.0	0.06	-19.7153	-80.3306	40.9001	False
0.0	0.1	3.5217	-1.3783	8.4217	False
0.0	0.15	6.6669	1.551	11.7828	True
0.0	0.2	5.3096	0.2508	10.3684	True
0.0	0.25	6.525	1.3647	11.6852	True
0.01	0.02	0.0	-74.2101	74.2101	False
0.01	0.03	-0.3333	-70.2993	69.6326	False
0.01	0.04	-1.0	-86.6905	84.6905	False
0.01	0.05	26.0108	-34.745	86.7667	False
0.01	0.06	0.0	-85.6905	85.6905	False
0.01	0.1	23.237	-37.5302	84.0042	False
0.01	0.15	26.3822	-34.4028	87.1671	False
0.01	0.2	25.0248	-35.7554	85.805	False
0.01	0.25	26.2403	-34.5485	87.029	False
0.02	0.03	-0.3333	-55.6463	54.9796	False
0.02	0.04	-1.0	-75.2101	73.2101	False
0.02	0.05	26.0108	-17.0654	69.087	False
0.02	0.06	0.0	-74.2101	74.2101	False
0.02	0.1	23.237	-19.8552	66.3292	False
0.02	0.15	26.3822	-16.7351	69.4994	False
0.02	0.2	25.0248	-18.0857	68.1354	False
0.02	0.25	26.2403	-16.8823	69.3628	False
0.03	0.04	-0.6667	-70.6326	69.2993	False
0.03	0.05	26.3441	-8.9214	61.6096	False
0.03	0.06	0.3333	-69.6326	70.2993	False
0.03	0.1	23.5703	-11.7147	58.8553	False
0.03	0.15	26.7155	-8.6001	62.0311	False
0.03	0.2	25.3582	-9.9492	60.6656	False
0.03	0.25	26.5736	-8.7485	61.8957	False
0.04	0.05	27.0108	-33.745	87.7667	False
0.04	0.06	1.0	-84.6905	86.6905	False
0.04	0.1	24.237	-36.5302	85.0042	False
0.04	0.15	27.3822	-33.4028	88.1671	False
0.04	0.2	26.0248	-34.7554	86.805	False
0.04	0.25	27.2403	-33.5485	88.029	False
0.05	0.06	-26.0108	-86.7667	34.745	False
0.05	0.1	-2.7738	-9.1822	3.6346	False
0.05	0.15	0.3714	-6.2036	6.9463	False
0.05	0.2	-0.986	-7.5166	5.5447	False
0.05	0.25	0.2294	-6.3801	6.839	False
0.06	0.1	23.237	-37.5302	84.0042	False
0.06	0.15	26.3822	-34.4028	87.1671	False
0.06	0.2	25.0248	-35.7554	85.805	False
0.06	0.25	26.2403	-34.5485	87.029	False
0.1	0.15	3.1452	-3.5337	9.824	False
0.1	0.2	1.7879	-4.8474	8.4231	False
0.1	0.25	3.0033	-3.7096	9.7161	False
0.15	0.2	-1.3573	-8.1536	5.4389	False
0.15	0.25	-0.1419	-7.014	6.7302	False
0.2	0.25	1.2154	-5.6143	8.0451	False

Figure 1: Tukey Multi Comparison - Discount vs. No Discount

```
formula = 'Quantity ~ C(Discount)'  # for a linear model and generate table
lm = ols(formula, data1_pt2).fit()
table = sm.stats.anova_lm(lm, typ=1) # for one way analysis
print(table)
lm.summary()
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Discount)	4.0	1065.701253	266.425313	0.616376	0.650947
Residual	833.0	360060.198508	432.245136	NaN	NaN

Figure 2: One way ANOVA - difference between discount levels

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
0.05	0.1	-1.6853	-7.6359	4.2654	False
0.05	0.15	1.4599	-4.6486	7.5683	False
0.05	0.2	0.1026	-5.9639	6.169	False
0.05	0.25	1.318	-4.8232	7.4592	False
0.1	0.15	3.1452	-3.1197	9.41	False
0.1	0.2	1.7879	-4.4361	8.0118	False
0.1	0.25	3.0033	-3.2935	9.3	False
0.15	0.2	-1.3573	-7.7323	5.0176	False
0.15	0.25	-0.1419	-6.588	6.3042	False
0.2	0.25	1.2154	-5.1909	7.6217	False

Figure 3: Tukey Multi Comparison - Discount Levels

	EmployeeId	Region	Revenue
0	4	North America	249536.3550
1	3	North America	212652.6125
2	1	North America	201763.0405
3	2	North America	177431.4755
4	7	British Isles	140868.0110
5	8	North America	132962.2245
6	9	British Isles	82771.5430
7	6	British Isles	78006.7955
8	5	British Isles	75352.9100

Figure 4: Total Revenue brought in by each employee and their region

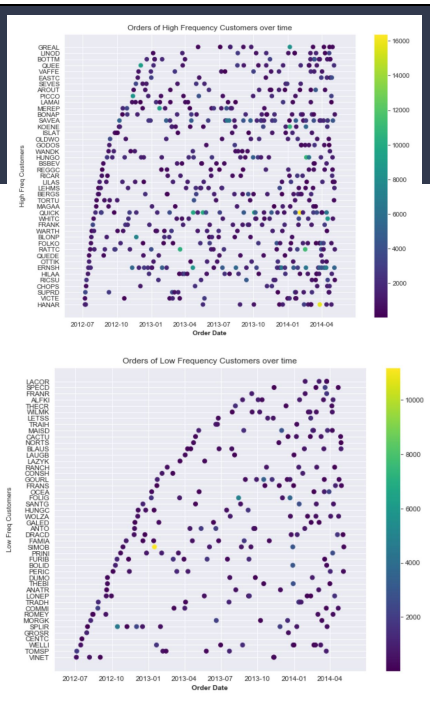


Figure 5: Date of first order for low frequency and high frequency customers

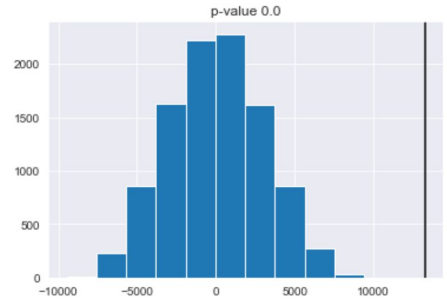


Figure 5: Monte Carlo simulation comparing means of samples of low frequency customers compared to mean of high frequency customers

Figure 6: Tukey comparison of freight costs per region

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
British Isles	Central America	-36.0275	-115.2593	43.2043	False
British Isles	Eastern Europe	-51.0211	-192.4108	90.3686	False
British Isles	North America	28.9349	-21.5488	79.4186	False
British Isles	Northern Europe	8.1239	-55.3872	71.6351	False
British Isles	Scandinavia	-33.7557	-112.9875	45.4761	False
British Isles	South America	-19.4789	-70.3636	31.4058	False
British Isles	Southern Europe	-39.0977	-99.9781	21.7826	False
British Isles	Western Europe	16.4866	-30.0998	63.0729	False
Central America	Eastern Europe	-14.9936	-166.1743	136.1872	False
Central America	North America	64.9624	-8.6118	138.5367	False
Central America	Northern Europe	44.1514	-38.9042	127.2071	False
Central America	Scandinavia	2.2718	-93.3433	97.8869	False
Central America	South America	16.5486	-57.3014	90.3986	False
Central America	Southern Europe	-3.0702	-84.1319	77.9914	False
Central America	Western Europe	52.5141	-18.4427	123.4708	False
Eastern Europe	North America	79.956	-58.3427	218.2547	False
Eastern Europe	Northern Europe	59.145	-84.4224	202.7124	False
Eastern Europe	Scandinavia	17.2654	-133.9154	168.4461	False
Eastern Europe	South America	31.5422	-106.9034	169.9878	False
Eastern Europe	Southern Europe	11.9233	-130.4998	154.3465	False
Eastern Europe	Western Europe	67.5077	-69.4165	204.4318	False
North America	Northern Europe	-20.811	-77.1063	35.4844	False
North America	Scandinavia	-62.6906	-136.2649	10.8836	False
North America	South America	-48.4138	-89.9439	-6.8837	True
North America	Southern Europe	-68.0326	-121.3423	-14.723	True
North America	Western Europe	-12.4483	-48.584	23.6873	False
Northern Europe	Scandinavia	-41.8797	-124.9353	41.176	False
Northern Europe	South America	-27.6028	-84.258	29.0524	False
Northern Europe	Southern Europe	-47.2217	-113.0015	18.5581	False
Northern Europe	Western Europe	8.3626	-44.4659	61.1912	False
Scandinavia	South America	14.2769	-59.5731	88.1269	False
Scandinavia	Southern Europe	-5.342	-86.4036	75.7196	False
Scandinavia	Western Europe	50.2423	-20.7144	121.199	False
South America	Southern Europe	-19.6189	-73.3084	34.0707	False
South America	Western Europe	35.9654	-0.7284	72.6593	False
Southern Europe	Western Europe	55.5843	5.9496	105.219	True

	Freight							
	count	mean	std	min	25%	50%	75%	max
CompanyName								
Federal Shipping	255.0	80.441216	119.362844	0.40	12.86	36.710	99.49	1007.64
Speedy Express	249.0	65.001325	75.393587	0.12	12.75	40.420	89.90	458.78
United Package	326.0	86.640644	138.219262	0.02	14.04	44.135	91.43	890.78

Figure 8: Basic statistics of freight costs of each shipping company.

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
Federal Shipping	Speedy Express	-15.4399	-39.825	8.9452	False
Federal Shipping	United Package	6.1994	-16.6823	29.0811	False
Speedy Express	United Package	21.6393	-1.3965	44.6752	False

Figure 8: Tukey comparison of Freight costs of Shipping companies