CASE STUDY - ALGORITHM DEVELOPMENT
LE HOANG NHAN

1. Combine the data sets and check which suppliers have brought the most revenue.

1.1 Combine the data sets

We have:

- Algorithm_Revenues_Posters_NL: 3278 records
- Algorithm_SalesPrice_Posters_NL: 3278 records
- Algorithm_Costs_Posters_NL: 3278 records
- Algorithm_CompetitorsPrices_Poster: 3650 (skus are duplicated)

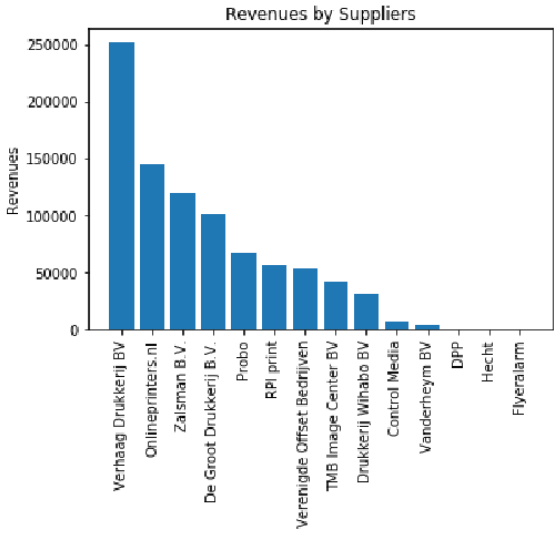
➔ I managed to maintain all 3278 records (not lost any)

1.2 Suppliers have brought the most revenue

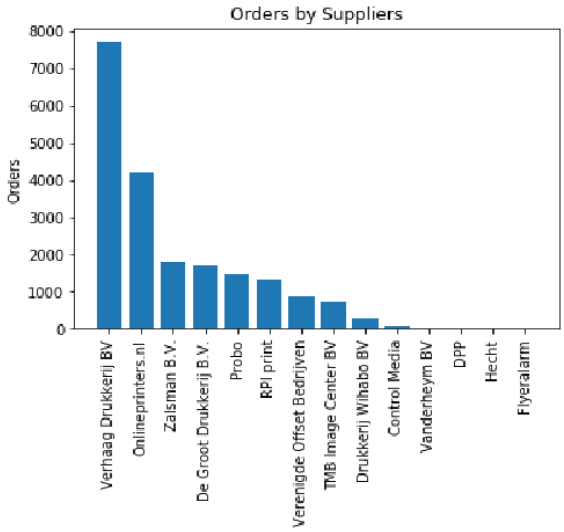
❖ **Combine data process:**

- Costs_Posters + Revenues = Costs_Revenues
- from Costs_Revenues --> Supplier_Revenues, which creates the below graphs:

Graph Revenues by Suppliers



Graph Orders by Suppliers



2. Create a competitiveness measure of our sales prices and an attractiveness measure of our cost prices versus each of the competitors

2.1 Competitiveness measure sales prices vs competitors'

❖ **Combine data process:**

- CompetitorsPrices + SalesPrice = SalesPrice_CompetitorsPrices, which has our price and competitors' prices.
- CompetitorsPrices pivot grouped by sku -> Price_pivot
- Add SalesPrice ---> Price_pivot

❖ **Calculate Price Index for competitors:**

- Price Index for particular sku for particular competitor:

PI of competitor i for sku j = [Price of competitor i / Our price] * 100

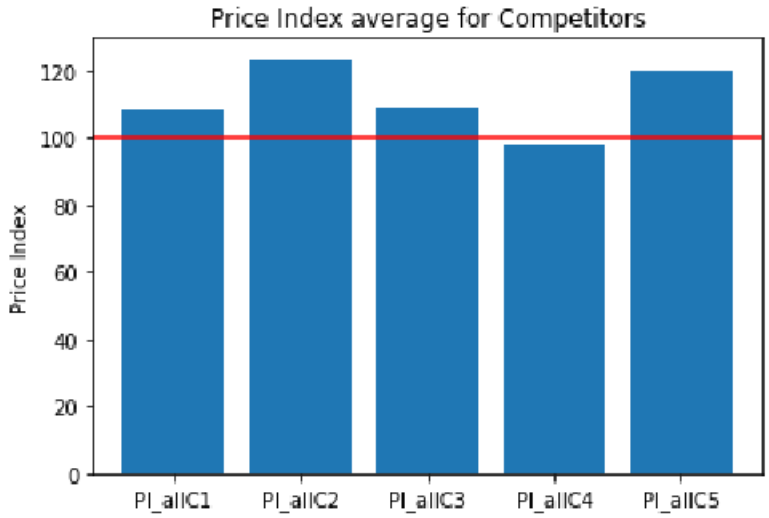
- Price Index average for particular sku (average all competitors):

PI of all competitors fro sku j = average of all PI of competitor i for sku j

- Price Index average for particular competitors (average all sku):

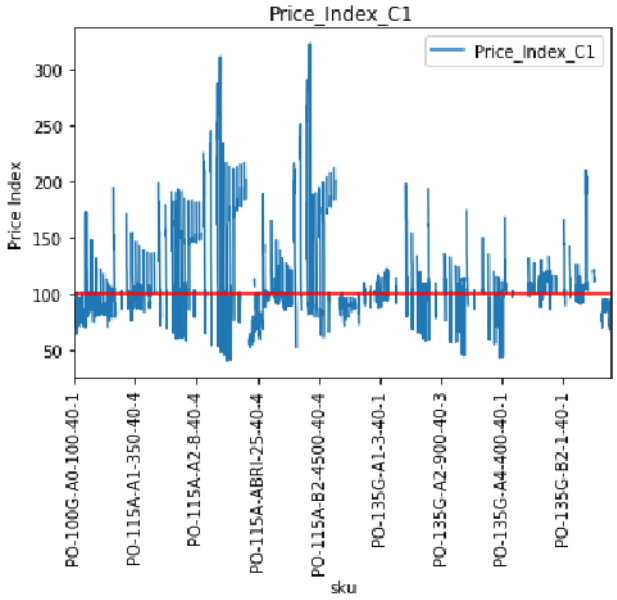
PI of competitor i for all skus = average of PI of competitor i for all skus

❖ **Graph Price Index average for Competitors**



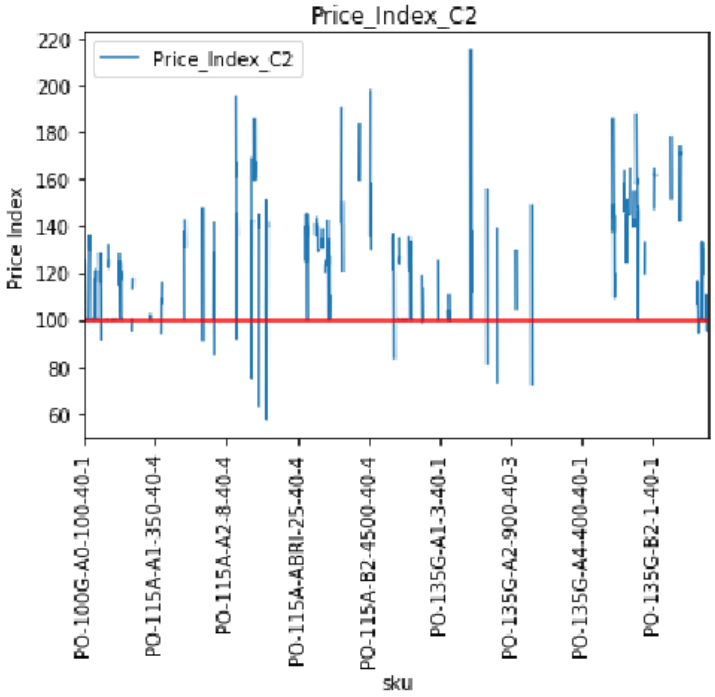
- For all skus, our average sale prices are 10% lower than Competitors 1 and 3, and even 20% lower than Competitors 2 and 5.
- However, average prices of Competitors 4 are lightly lower than our average prices.

❖ **Graph Price Index for particular sku for competitor 1**



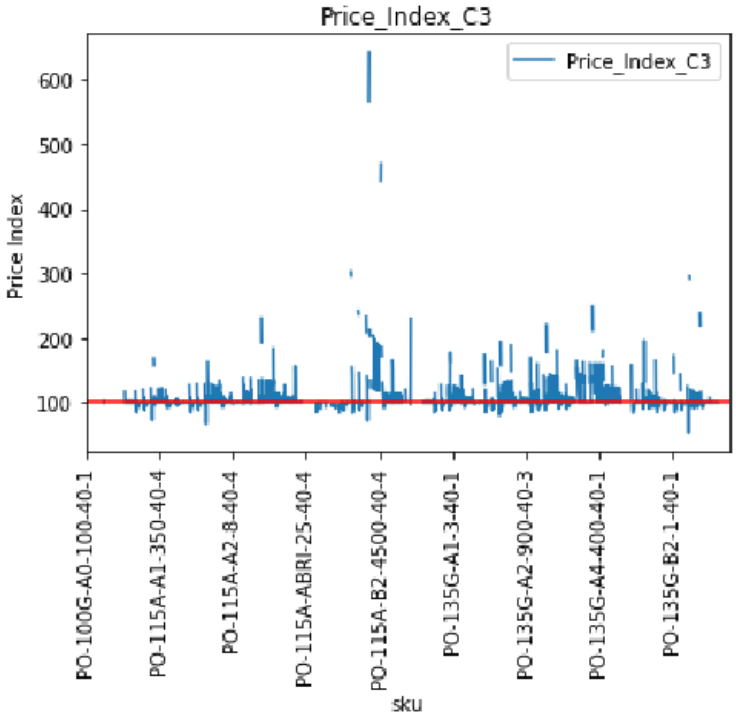
- We compete toughly with Competitors 1 for almost all our skus.
- Most our sale prices are lower (PI of Competitors 1 is above 100% line)
- Some Competitors prices is 50% lower than ours.

❖ **Graph Price Index for particular sku for competitor 2**



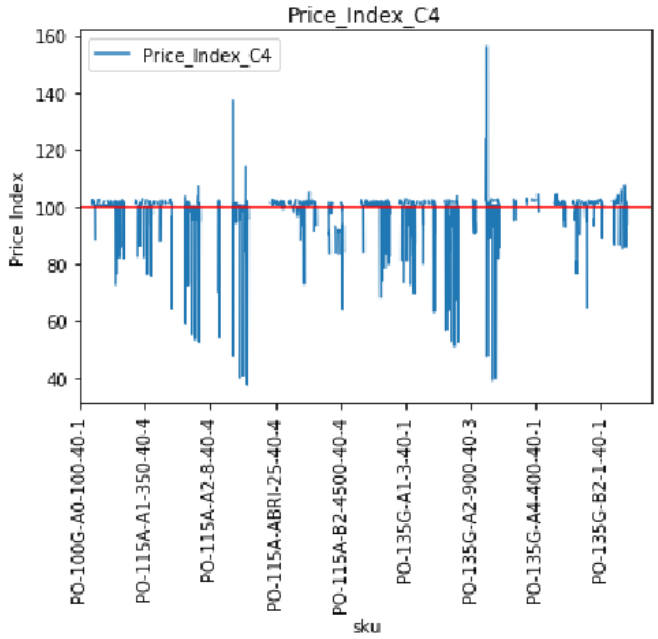
- We don’t compete much with Competitors 2, even only few our skus.
- We have many prices advantages

❖ **Graph Price Index for particular sku for competitor 3**



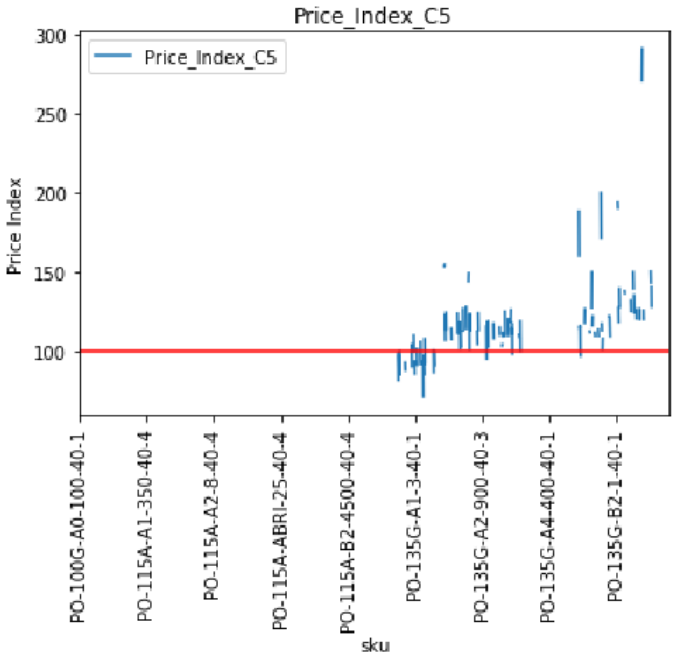
- We compete toughly with Competitors 2 for almost our skus.
- We don't have prices advantages

❖ **Graph Price Index for particular sku for competitor 4**



- We compete with Competitors 2 for a certain amount of our skus.
- We have many prices disadvantages. Its prices are much lower than ours.

❖ **Graph Price Index for particular sku for competitor 5**



- We don’t compete much with Competitors 2, even a few our skus
- We have many prices advantages

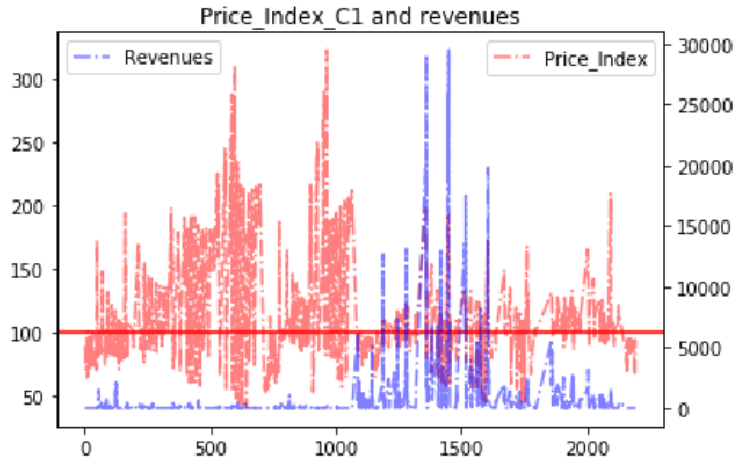
❖ **Rank competitors based on numbers of skus:**

competitor 1 > competitor 3 > competitor 4 > competitor 5 > competitor 2

❖ **Rank competitors based on price:**

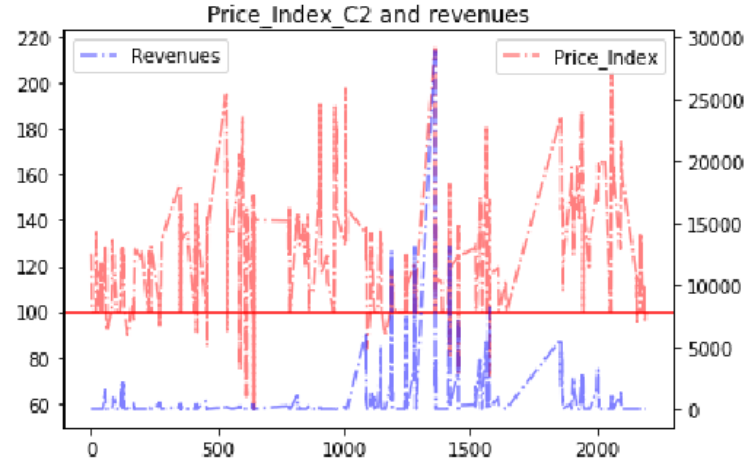
competitor 4 > competitor 3> competitor 1 > competitor 5 > competitors 2

❖ **Graph Price_Index_C1 and our revenues for each sku**



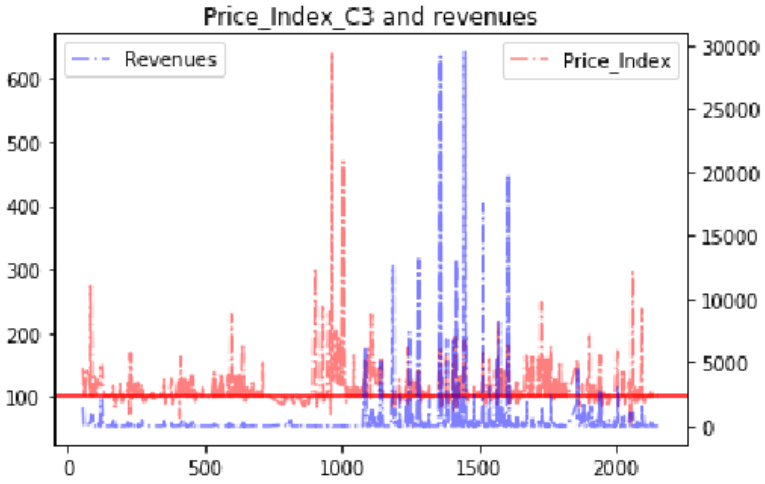
- Competitor 1 prices damage our revenues severely
- For some skus in where we have price advantages, our revenues are still low.

❖ **Graph Price_Index_C2 and our revenues for each sku**



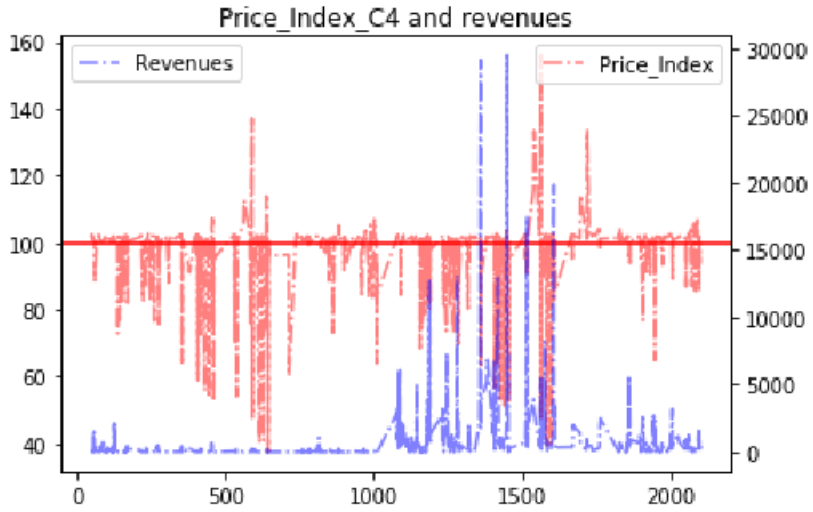
- For some skus in where we have price advantages, our revenues are still low.
- However, we deploy effectively our price advantages for other skus.

❖ **Graph Price_Index_C3 and our revenues for each sku**



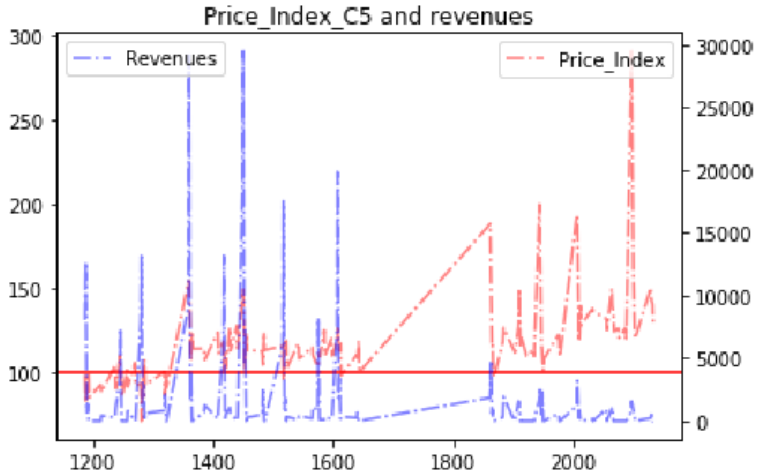
- For some skus in where we have price advantages, our revenues are still low.
- However, we deploy effectively our price advantages for other skus.

❖ **Graph Price_Index_C4 and our revenues for each sku**



- For some skus in where we have price disadvantages, Competitor 4 prices damage our revenues severely.
- However, for some skus in where we have price advantages, we deploy effectively our price advantages.

❖ **Graph Price_Index_C5 and our revenues for each sku**



We deploy effectively our price advantages for all our skus.

❖ **Rank competition based on numbers of skus:**

competitor 1 > competitor 3 > competitor 4 > competitor 5 > competitor 2

❖ **Rank competition based on price advantages:**

competitor 4 > competitor 3> competitor 1 > competitor 5 > Competitors 2

➔ **Overall, competitor 1, 3, 4 are main competitors**

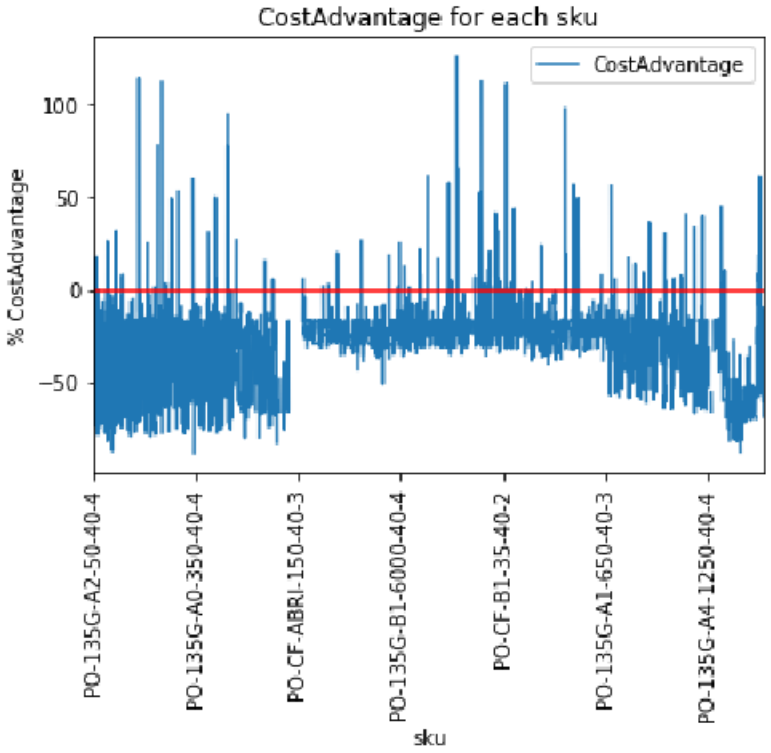
2.2 Attractiveness measure our cost prices vs competitors'

❖ **Calculate:**

- Total_Cost = supplier_price_salescountry + carrier_cost_salescountry
- CostAdvantage = (Total_Cost - Min among of Competitor Price Indexes) *100 / Min among of Competitor Price Indexes
- CostAdvantage: compare our production cost vs competitors' sale prices

--> the lower CostAdvantage is the better.

❖ **Graph CostAdvantage for each sku**



- Our production cost is much lower than competitors' sale prices

➔ **we still have room for reducing prices, if necessary.**

❖ **Correlation between CostAdvantage and number of Orders**

- 0.14

- CostAdvantage somehow affects the number of Orders on a certain extent.

3. Create a weighted importance score for each SKU

❖ **Use Machine Learning Linear Regression:**

- X: supplier_price_salescountry, carrier_cost_salescountry, carrier_days, extra_days, Price_Index_allcompetitors.

- supplier_price_salescountry, carrier_cost_salescountry: take production cost into account.
- carrier_days, extra_days: take delivery service into account.
- Price_Index_allcompetitors: take our sale prices and competitors' prices into account

- Y: Revenues

- X, Y are rescaled to MinMax [0,1] to improve the accuracy of model prediction and easier to compare the coefficients later.

- The model:

Revenues = intercept + a1 * supplier_price_salescountry + a2 * carrier_cost_salescountry + a3 * carrier_days + a4 * extra_days + a5 * Price_Index_allcompetitors

- Why don’t choose Xs:

- @Turnaround: correlated with carrier_days, extra_days
- @Material, @Size: cannot be considered as a sku’s design aspect.
- Sales_Price, CostAdvantage: correlated with Price_Index_allcompetitors

- Why don’t choose Lasso, Ridge regression:
  - This model is not complex and does not have many features. No need to reduce or choose features.
  - Chosen features are well reasonable.
- Split data into 2 parts:
  - Data in which skus have revenues: be used to train the model
  - Data in which skus don't have revenues: be used to predict revenues by the model.

❖ **Weighted importance score for each SKU**

- Weighted importance score for each SKU (Score): revenues are rescaled to MinMax [0,1]. Sku with maximum revenues has Score 1 and vice versa.
- Some SKUs is non-sold due to some features not seen in the data, not its features (supplier_price, carrier_cost, carrier_days, extra_days, Price_Index) included in this model.
- In this model, non-sold SKUs would have predicted revenues.
- Non-sold SKUs with high scores (predicted revenues) are worth for improving and could make real revenues.

❖ **5 skus have highest scores**

	sku	Revenues	Score
0	PO-135G-A2-50-40-4	29479.26729	1.000000
1	PO-135G-A2-100-40-4	29136.57392	0.988372
2	PO-135G-A2-50-40-1	25775.39229	0.874325
3	PO-135G-A2-100-40-1	19935.51579	0.676174
4	PO-135G-A3-50-40-4	19830.65735	0.672616

❖ **5 skus have lowest scores**

	sku	Revenues	Score
950	PO-115A-B1-1-40-1	10.75	0.000112
951	PO-135G-A4-7-40-4	9.57	0.000072
942	PO-115A-A3-4000-40-1	0.00	0.000054
952	PO-135G-A4-3-40-2	7.45	0.000000
820	PO-115A-A3-5000-40-1	0.00	-0.000596

4. Based on your weighted importance score, analyze what attributes are most important

❖ **In the model, coefficients are:**

- supplier_price_salescountry: -0.0063129
- carrier_cost_salescountry: -0.02693092
- carrier_days: -0.01000462
- extra_days: 0.00530029
- Price_Index_allcompetitors**: 0.11687582 (strongly significant)**

- Consider magnitude coefficient * std of the corresponding parameter:
[-0.0063129, -0.02693092, -0.01000462, 0.00530029, 0.11687582]
-> not much different from coefficients in the model.

❖ **Interpret the results:**

- Because we had rescaled Xs to MinMax [0,1], we can compare the magnitude coefficients with each other.
- supplier_price_salescountry, carrier_cost_salescountry, carrier_days have negative effects on revenues.
- extra_days has a positive effect on revenues because the company only require extra days to guarantee customers receiving the best delivery service.
- Price_Index_allcompetitors has a positive significant effect on revenues. Competitive price has strong advantages in competition. Customers have interested in low prices.

❖ **Suggested strategies:**

- Customers care about receiving products on time. Our company requiring extra days is not perceived as negative by customers.
- Customers have interested in low prices. Currently, our company has huge cost advantages (production cost is much lower than competitors' sale prices), so we still have room for reducing our sale price to boost revenues, even for non-sold skus.
- According to the data, competitive price is a good strategy to improve revenues for our company now.

❖ **Acknowledge drawbacks of the model:**

- few features due to the availability of the data. Features only capture cost production, delivery and sale prices.
- R square of the model is low. However, the ultimate goal is to get insights about these features' effects on revenues, not high R square. Furthermore, we capture significantly Price Index effect which affects strongly revenues.
- skus' scores are not predicted perfectly due to low R square of the model. However, in general, the scores can distinguish the most promising products for future improvement.