When including all the variables together, I noticed that we could have different approaches for measuring conversion rates. So, I tested different measures. The variables and their descriptions are shown in Table 1. They are all considering a weekly period.

Table – Description of the Variables.

|  |  |
| --- | --- |
| total\_views\_mc2 | Total PDP page views per mc2; trigger event: detail |
| conversions\_purchase\_mc2 | How many PDP page views converted to purchase per mc2 |
| conv\_purc\_mc2\_rate | How many PDP page views converted to purchase by the total PDP page views per mc2 |
| total\_views | Total PDP page views for all mc2 combined; trigger event: detail |
| conversions\_purchase | How many PDP page views converted to purchase for all mc2 combined |
| conv\_purc\_rate | How many PDP page views converted to purchase by the total PDP page views for all mc2 combined |
| conv\_purc\_to\_all | How many PDP page views (per mc2) converted to purchase by the total PDP page views for all mc2 combined |
| total\_views\_perc | Total PDP page views per mc2 by the total PDP page views for all mc2 combined |
| oos\_rate | How many times items were oos by the items count per week |
| total\_price\* | Total price for units ordered, excluding canceled orders |
| ordered\_qty | Total units ordered, excluding canceled orders |
| autoship\_qty | Total units ordered that were flagged autoship, excluding canceled orders |
| conv\_order\_mc2\_rate | Total units ordered per mc2 by the total PDP page views per mc2 |
| total\_visits\_mc2 | Total visits (any event) per mc2 |
| total\_visits\_all | Total visits (any event) all mc2 combined |
| total\_visits\_exc | Total visits (any event) all mc2 combined, excluding the mc2 being studied |
| lagged\_oos\_rate | Lagged oos\_rate variable by one week |

\*target variable

First, we can see the correlations. For all variables, except Core Food, we have strong relationship between oos\_rate or lagged\_oos\_rate only for total\_views and conversions\_purchase. And this relationship is positive. This might happen due to people searching for replacements for the oos products.

Chart

Description automatically generated

For Core Food, we see some strong negative relationship between oos\_rate and lagged\_oos\_rate between and conv\_purc\_rate\_to\_all and total\_views\_perc. This might indicate that oos do have an impact on conversion and PDP page views per mc2 when compared to all mc2 combined. Some note is that these same conv\_purc\_rate\_to\_all and total\_views\_perc are negatively correlated to total\_price. This might indicate that when you have more visits to Core Food when compared to other mc2, customers spend less. They might be going to the website just for food.

Chart, treemap chart

Description automatically generated

When we check VIF for multicollinearity, we see that several of the variables do have multicollinearity. Which can make the modeling more complex.

['total\_views\_mc2' 'conversions\_purchase\_mc2' 'conv\_purc\_mc2\_rate'

'total\_views' 'conversions\_purchase' 'conv\_purc\_rate'

'conv\_purc\_rate\_to\_all' 'total\_views\_perc' 'oos\_rate' 'total\_price'

'ordered\_qty' 'autoship\_qty' 'conv\_order\_mc2\_rate' 'total\_visits\_mc2'

'total\_visits\_all' 'total\_visits\_exc' 'lagged\_oos\_rate']

feature VIF

0 total\_views\_mc2 5,778.43

1 conversions\_purchase\_mc2 2,112.93

2 conv\_purc\_mc2\_rate 2,649.75

3 total\_views 2,951.59

4 conversions\_purchase 1,789.68

5 conv\_purc\_rate 3,250.39

6 conv\_purc\_rate\_to\_all 1,822.38

7 total\_views\_perc 3,458.98

8 oos\_rate 7.75

9 total\_price 674.46

10 ordered\_qty 4,202.33

11 autoship\_qty 550.08

12 conv\_order\_mc2\_rate 1,913.79

13 total\_visits\_mc2 230.14

14 total\_visits\_all 3,675.41

15 total\_visits\_exc 3,157.08

16 lagged\_oos\_rate 7.96

We run a ElasticNet for identifying future importance, without considering ordered\_qty and autoship\_qty (these are too strongly correlated to total price and might depend on the same variables). Then, we get the following. It’s interesting to see that conv\_purc\_rate\_to\_all has the highest importance, while lagged\_oos\_rate is in the lowest above 0. This might indicate that the total\_price might be more related to the dynamics between mc2 than oos\_rate.

Chart, waterfall chart

Description automatically generated

If we take the scatter plot of the total\_price by these two variables, we see some trend but not clearly defined. With total\_price x oos\_rate with an upward trend and total\_price x conv\_purc\_rate\_to\_all with a downward trend.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

We perform the same analysis with Toys for comparison. We see some similarities between the two analyses. However, the correlations we mentioned before are weaker. The ElasticNet feature selection also shows a similar feature importance, with conv\_purc\_rate\_to\_all as the highest. I will run for other mc2s to see if anything changes drastically. The plots for Toys are presented below.

Chart

Description automatically generated

Chart, treemap chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart

Description automatically generated