**OOS vs Ordered Quantity Analysis**

**Note 1:** we continue the analysis with the exclusion of the following classes: Digital Gift Cards, Promotional Cards, Virtual Bundle, Perishable and Personalized.

**Note 2:** we use the snowflake product table to get the merch class 2 of each item throughout the studied period.

First, we combine the existing datasets to get the following variables:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| mc2 | Merch Class 2 |
| key\_0 | Week period |
| instock\_views\_perc | Percentage of in-stock PDP views for the mc2 from the total in-stock PDP views |
| oos\_views\_perc | Percentage of oos PDP views for the mc2 from the total oos PDP views |
| total\_views\_perc | Percentage of PDP views for the mc2 from the total PDP views |
| oos\_rate | Rate of oos occurrences weekly divided by the count of items |
| ordered\_qty | Quantity of ordered items in the week |
| autoship\_qty | Quantity of ordered items as autoship in the week |

We then look at the merch classes with the highest weekly percentages of PDP views from the total. Figure 1 shows that Core Food is by far the merch class with highest weekly percentages of PDP views from the total, with a median of almost 25% (1/4) PDP views from the total.



Figure – Distributions of Percentage of PDP Views per Merch Class 2 from the Total Number of PDP Views

We also plot a time series of 6 merch classes with high percentages of PDP views to understand their patterns (Figure 2). From the plot, we see that many of these classes have some seasonality, which explains the high variation show in their boxplot (Figure 1). We observe a drop in percentages of PDP views for each category in 2020, but this is due the distribution of these views into more categories. The total number of PDP views continues to increase (Figure 3).

Graphical user interface, chart, application

Description automatically generated

Figure 2 – Time Series of Percentage of PDP Views from Total for the 6 Merch Classes with High Weekly Traffic

Chart, line chart

Description automatically generated

Figure 3 - Total PDP Views per Week

A possible causal relationship between views and ordered quantity is shown in a correlation matrix of all merch classes (Figure 4), where the ordered quantity and percentage of PDP views per merch are strongly positive correlated. However, the oos rate does not show strong correlation to ordered quantity.



Figure 4 - Correlation Matrix for Data Including All Merch Classes Including Core Food

Then, we use the example of Core Food (the highest percentage PDP views) to compare to the behavior of all other merch classes combined and excluding Core Food. First, we see that the ordered quantity has an upward trend like the total PDP views for all merch classes (Figure 3) and oos rate (Figure 5). We check if the distribution of ordered quantity for Core Food comes from the same population of the other merch classes combined. Using KS two tail test, we find a p-value of 2.11e-15. We reject the null hypothesis that they come from the same distribution.

Chart, line chart, histogram

Description automatically generated

Figure - Time Series Ordered Quantity and OOS Rate

Then, we look at the correlations matrix of all merch classes excluding Core Food and of Core Food (Figure 6). We see that the overall correlations of all merch classes combined excluding Core Food does not change much. However, the correlation matrix for Core Food a lot has some changes. The high positive correlation between ordered quantity and oos rate might be explained by the fact that the more items ordered, higher is the probability that it might run out of stock. The not so interpretable correlation is the one between ordered quantity and percentage of PDP page views from total. One might suggest that the less someone spends on food, the more it buys other products, or that if they find the food they are looking for, they will spend less time looking at other products. However, more investigation should be done to understand what underlies this negative correlation.

Chart, bar chart, histogram

Description automatically generated

Figure 6 - Correlation Matrix for Core Food

This same pattern repeats in the results for OLS regression in each case.

OLS Regression Results – All Merch Classes Including Core Food

==============================================================================

Dep. Variable: ordered\_qty R-squared: 0.533

Model: OLS Adj. R-squared: 0.532

Method: Least Squares F-statistic: 2411.

Date: Mon, 26 Jul 2021 Prob (F-statistic): 0.00

Time: 16:43:24 Log-Likelihood: -59544.

No. Observations: 4235 AIC: 1.191e+05

Df Residuals: 4232 BIC: 1.191e+05

Df Model: 2

Covariance Type: nonrobust

====================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------

Intercept -9847.1481 7767.477 -1.268 0.205 -2.51e+04 5381.183

oos\_rate -4.254e+05 8.51e+04 -4.998 0.000 -5.92e+05 -2.59e+05

total\_views\_perc 6.267e+06 9.09e+04 68.971 0.000 6.09e+06 6.45e+06

==============================================================================

Omnibus: 2236.307 Durbin-Watson: 1.524

Prob(Omnibus): 0.000 Jarque-Bera (JB): 21538.834

Skew: 2.335 Prob(JB): 0.00

Kurtosis: 13.012 Cond. No. 19.3

==============================================================================

OLS Regression Results – All Merch Classes Excluding Core Food

==============================================================================

Dep. Variable: ordered\_qty R-squared: 0.336

Model: OLS Adj. R-squared: 0.335

Method: Least Squares F-statistic: 1022.

Date: Fri, 23 Jul 2021 Prob (F-statistic): 0.00

Time: 14:13:12 Log-Likelihood: -56047.

No. Observations: 4051 AIC: 1.121e+05

Df Residuals: 4048 BIC: 1.121e+05

Df Model: 2

Covariance Type: nonrobust

====================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------

Intercept 1.586e+04 6496.754 2.442 0.015 3125.095 2.86e+04

oos\_rate -3.693e+05 6.83e+04 -5.408 0.000 -5.03e+05 -2.35e+05

total\_views\_perc 4.873e+06 1.08e+05 45.019 0.000 4.66e+06 5.09e+06

==============================================================================

Omnibus: 2820.254 Durbin-Watson: 2.151

Prob(Omnibus): 0.000 Jarque-Bera (JB): 51352.981

Skew: 3.119 Prob(JB): 0.00

Kurtosis: 19.289 Cond. No. 28.0

==============================================================================

OLS Regression Results – Core Food

==============================================================================

Dep. Variable: ordered\_qty R-squared: 0.579

Model: OLS Adj. R-squared: 0.574

Method: Least Squares F-statistic: 123.8

Date: Mon, 26 Jul 2021 Prob (F-statistic): 1.53e-34

Time: 17:03:46 Log-Likelihood: -2594.9

No. Observations: 183 AIC: 5196.

Df Residuals: 180 BIC: 5205.

Df Model: 2

Covariance Type: nonrobust

====================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------

Intercept 2.884e+06 2.14e+05 13.509 0.000 2.46e+06 3.31e+06

oos\_rate 9.039e+06 2.59e+06 3.486 0.001 3.92e+06 1.42e+07

total\_views\_perc -6.643e+06 6.74e+05 -9.848 0.000 -7.97e+06 -5.31e+06

==============================================================================

Omnibus: 52.673 Durbin-Watson: 0.665

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1279.204

Skew: 0.219 Prob(JB): 1.68e-278

Kurtosis: 15.945 Cond. No. 104.

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

To try to understand Core Food more, we performed a new analysis considering the brands. We have several brands in the Core Food merch class, so we have done some resampling to consider only 20 brands. We first identified the quantiles for the weekly ordered quantity distribution given the brands and the sampled 5 of each of the quantiles. Figure 7 shows the ordered quantity of each of the 20 brands sampled.

Chart

Description automatically generated

Figure 7 - Ordered Quantity for Each of Sampled Brands

Something interesting happens when we perform the OLS considering the brands, the oos rate and percentage of PDP views from total now follow the same pattern as the combined merch classes (i.e. oos rate – negative coefficient, percentage views – positive coefficient), but neither are statistically significant. We also see those brands with high number of order are not always statistically significant (e.g. Hill’s Science Diet).

OLS Regression Results – Core Food with Brands

==============================================================================

Dep. Variable: ordered\_qty R-squared: 0.863

Model: OLS Adj. R-squared: 0.852

Method: Least Squares F-statistic: 82.16

Date: Mon, 26 Jul 2021 Prob (F-statistic): 3.98e-66

Time: 10:08:00 Log-Likelihood: -2505.6

No. Observations: 184 AIC: 5039.

Df Residuals: 170 BIC: 5084.

Df Model: 13

Covariance Type: nonrobust

=====================================================================================

coef std err t P>|t| [0.025 0.975]

-------------------------------------------------------------------------------------

Intercept 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

total\_views\_perc 1.13e+06 6.95e+05 1.626 0.106 -2.42e+05 2.5e+06

oos\_rate -1.246e+05 1.19e+06 -0.104 0.917 -2.48e+06 2.23e+06

BilJac 2.502e+05 1.08e+05 2.309 0.022 3.63e+04 4.64e+05

BlackGold 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

Blackwood 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

BlueSeal 1.741e+05 6.45e+04 2.701 0.008 4.69e+04 3.01e+05

Diamond 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

DogWhisperer 1.033e+05 1.13e+05 0.914 0.362 -1.2e+05 3.26e+05

Freshpet -1.669e+05 2.09e+05 -0.799 0.426 -5.8e+05 2.46e+05

GravyTrain 2.802e+05 5.11e+04 5.482 0.000 1.79e+05 3.81e+05

Halo 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

HillsIdealBalance 2.41e+05 8e+04 3.011 0.003 8.3e+04 3.99e+05

HillsScienceDiet 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

MaintainChunks 4.283e+05 8.21e+04 5.219 0.000 2.66e+05 5.9e+05

NaturalPlanet 0 0 nan nan 0 0

Olewo 2.993e+05 7.81e+04 3.833 0.000 1.45e+05 4.53e+05

OrgaNOMics -1.752e+05 1.14e+05 -1.542 0.125 -3.99e+05 4.91e+04

PetWellbeing 4.371e+05 9.39e+04 4.657 0.000 2.52e+05 6.22e+05

PurinaBeneful 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

Redbarn 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

TrueChews 1.309e+05 4.2e+04 3.119 0.002 4.81e+04 2.14e+05

WalkAbout 4.496e+04 2.47e+04 1.822 0.070 -3749.205 9.37e+04

==============================================================================

Omnibus: 84.865 Durbin-Watson: 1.045

Prob(Omnibus): 0.000 Jarque-Bera (JB): 1037.421

Skew: 1.344 Prob(JB): 5.33e-226

Kurtosis: 14.317 Cond. No. 6.36e+34

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The smallest eigenvalue is 5.57e-67. This might indicate that there are

strong multicollinearity problems or that the design matrix is singular.

It is important to notice that some considerations were not made here. We disregarded the time series quality of the data (e.g. trend and seasonality), which might affect the performance of linear models due to the autocorrelation. We have not done any transformation to the data as well. I am now working on testing with standardizing the oos rate and percentage of PDP views from total and either considering the lag ordered quantity in the regression or differencing it to be used in an ARIMA model.