Social Media Analytics

Kum&Go

Jamie Moore, Purna Chandra Kuntla, Jeff Gioimo, Jerry Jacob
MS Business Analytics Capstone Project

Table of Contents

Synopsis

Kum & Go is a convenience store chain primarily located in the Midwestern United States. The company, based in Des Moines, Iowa, operates 400 stores in 11 states—primarily in its home state of Iowa. Other states include: Missouri, Kansas, Arkansas, Oklahoma, Nebraska, North Dakota, South Dakota, Minnesota, Montana, Colorado, and Wyoming. Kum & Go was ranked as the 24th-largest convenience store chain in the United States by Convenience Store News' 2019.

The company adopts several marketing techniques. Loyalty rewards is one of the marketing strategies. Once the user joins the rewards program, their purchases will be tracked and they will be awarded freebies and discounts. They will also be asked to answer a survey once every 30 days and rate the store they visited.

A couple of years ago, Kum & Go actively started their marketing campaign on social media sites like Facebook, Twitter and Instagram. They also created a new position and hired a Director to oversee their social media marketing strategies. Now two years into their social media campaign, they want to know whether the social media strategies made any improvements related to their sales, loyalty reward registration or customer ratings.

Project Management

We reached out to different companies like Wells Fargo, Non-Profits, and Kum & Go to get real time projects and datasets. Finally we received a project from Kum & Go. We communicated with Kum & Go via email to set up a meeting. We went to the new Kum & Go main office in downtown Des Moines to discuss and understand the project and datasets. After the initial meeting we followed up with additional questions using email communication.

Most of the project planning was done during our weekly meetings. This meeting more or less served as our standup meeting and weekly sprint planning. Work will be assigned for the coming week and clear milestones were set. A GitHub project was created for source code management. Collaboration features like bug tracking, feature requests and task management were done using this utility. We also used Zoom meetings extensively for online meetings and virtual collaboration.

Datasets

Table Name	Table Description
Facebook_Activity	Daily activity for facebook with the number of posts and the number of different actions users made that day.
Twitter_Activity	Daily activity for twitter with the number of posts and the number of different actions users made that day.
Instagram_Activity	Daily activity for instagram with the number of posts and the number of different actions users made that day.
Sale_Transactions	Daily transaction count per day for each store.
Loyalty_Signups	Monthly sign ups for each state.
NPS_Rating	Monthly rating per month for each store.
Stores_Master	The store's dataset. Includes information about store location.

Challenges

After receiving the datasets, we thought it was going to be straight forward. After further review, we realized the data did not have the same date format (some were daily, weekly, monthly). We had to decide how to format the 3 so we could use them all together in our evaluations.

Another challenge we had was the data had fields we did not know how to interpret. For example, Facebook Likes has 3 categories; Page LIkes, Organic Page LIkes, Paid Page LIkes. Due to this, we had to do extra research to understand how they differentiated to be able to appropriately score them for use in correlations. Another unique issue with social media dataset was to consider the time lag. The influence of the social media activity may not be seen on the same day, but on the subsequent days.

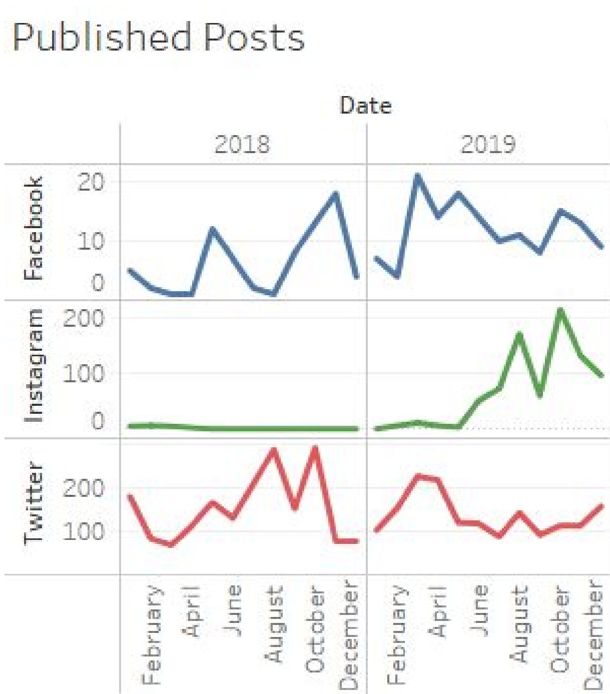
We spent a good amount of time reading the documentation from Sprout Social to try to understand these metrics. Sprout Social is the social media management and reporting tool used by Kum & Go. Data for some of these metrics: [Facebook](), [Twitter](), [Instagram]().

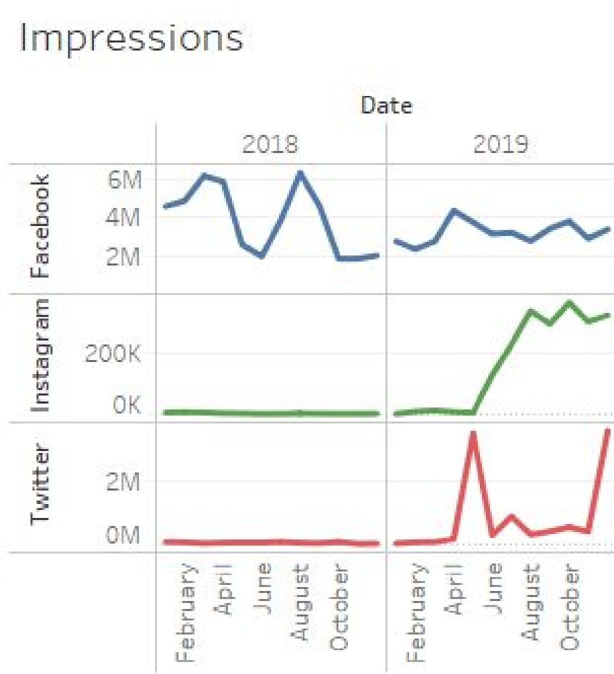
A big challenge for us was the different target datasets were collected at different aggregation levels. This meant that we had daily data for sales transactions… but only weekly data for loyalty sign ups… and finally monthly data for survey results. Further, some data was at a store level… while other data was at a state level. This made it difficult to work with some of these since we would have to aggregate our independent variables in order to compare apples to apples.

Exploratory Analysis

We noticed that there was some missing social media activity data. We followed up on this with the client and it seemed to be unintended but also something that they couldn't recover unfortunately. We were able to impute the missing data by using a linear interpolation method with the data we did have and filling the values for the missing dates.

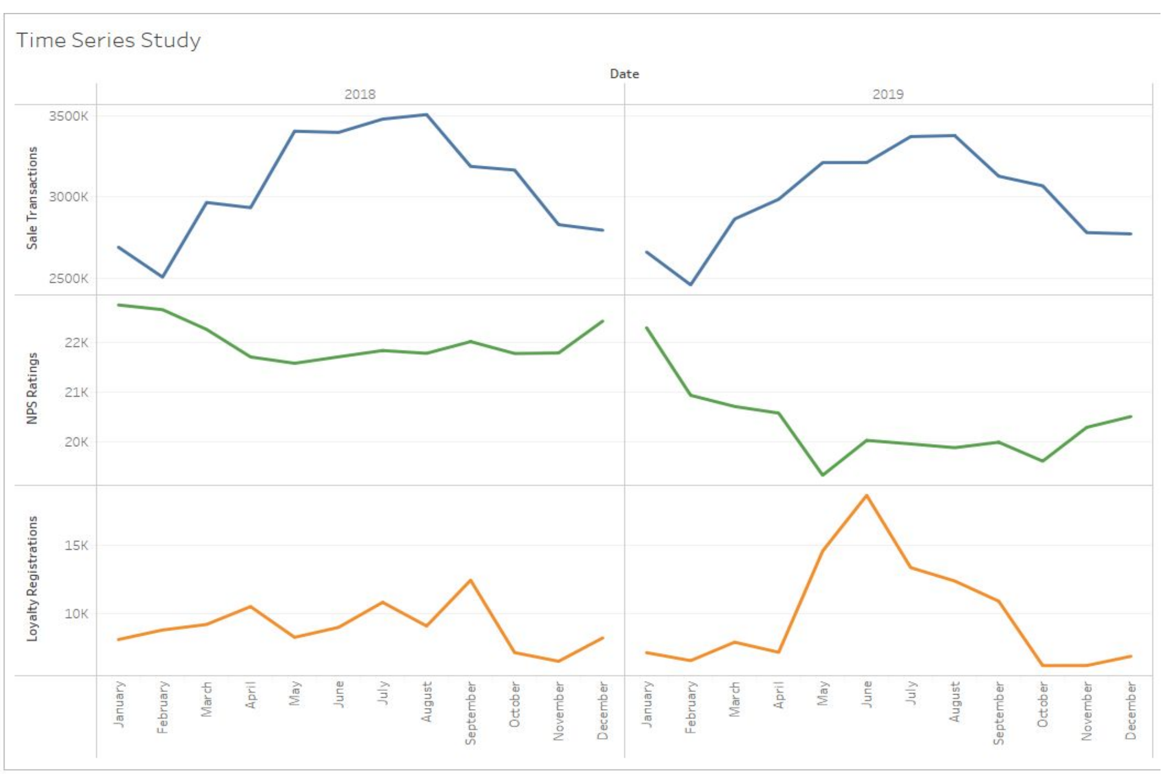
The below graph suggests that Kum & Go was only publishing posts on Twitter and Facebook in the 2018, even though user engagement was found only with facebook. In 2019, after their hire of a new Director to manage social media interactions, the user engagement was found to be increasing for Instagram and Twitter



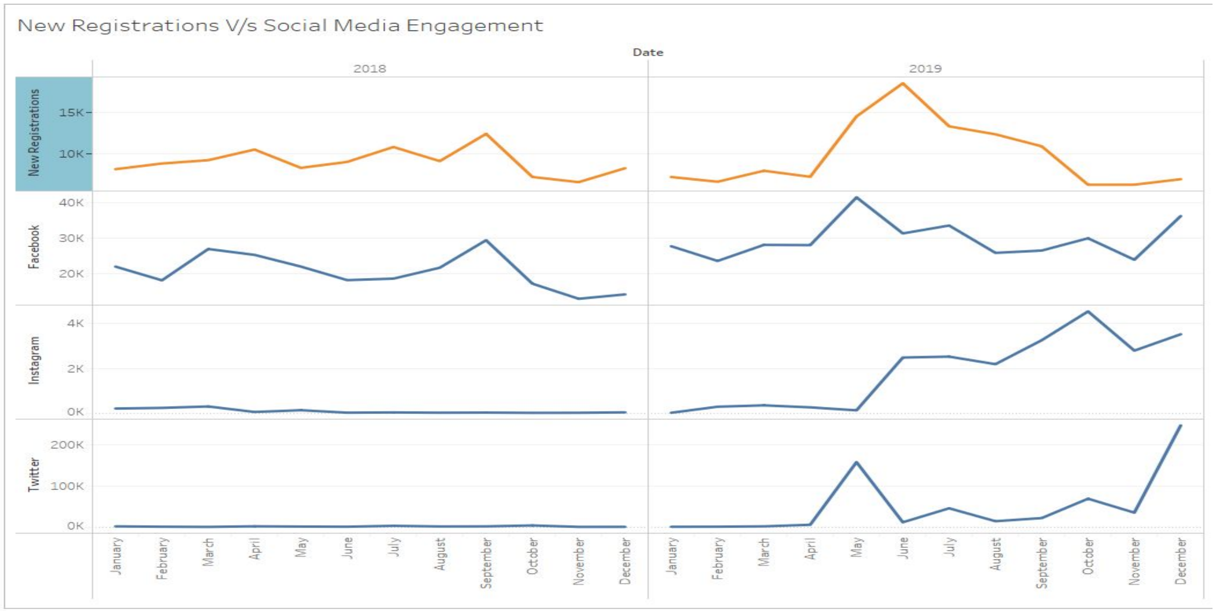


The below Twitter post was found to be an outlier compared to the other posts which got a very high reach in the month of May. The above graph shows a spike for Twitter impressions in the month of May 2019, which was solely due to this re-tweet.

The below graph shows a time series study on Sales, NPS Rating and Loyalty registrants. Compared with the social media activity, we suspect some relation with the heightened social media activity started in May 2019 with the number of loyalty program registrants during the same period.



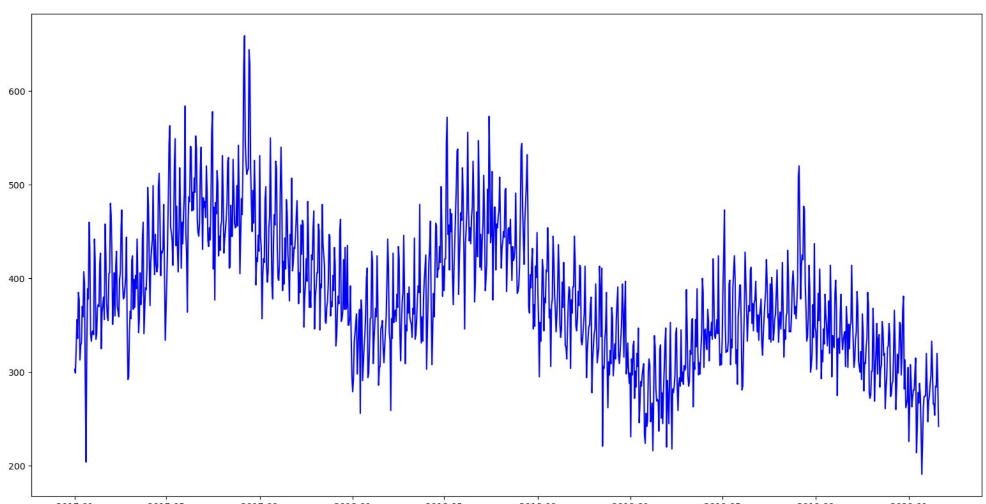
A closer look between new registrations and social media, which do show some relationship.



Monthly and weekly seasonality and trend with Sales transactions

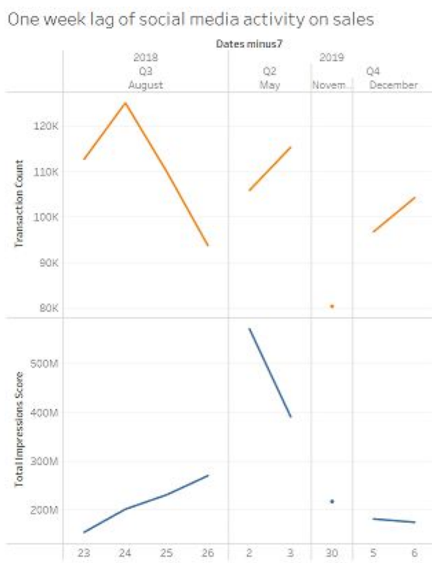
We found that the sales transactions have a very clear seasonality where the highest sales were noticed during the summer months and lowest were in the winter months. The sales also showed a very clear trend where the transaction count gradually increased from winter months to summer months and gradually decreased towards the winter months.

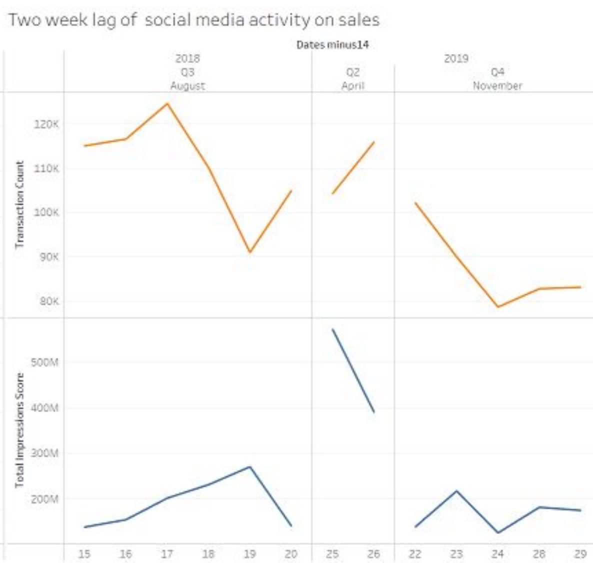
There was also a weekly trend for most stores. Most stores did the majority of their business on Friday evening. Some stores had positive overall trends, some negative, and some neutral. Here's an example for a store in Pleasant Hill IA:



You can see the monthly, weekly and overall downward trend all in one graph!

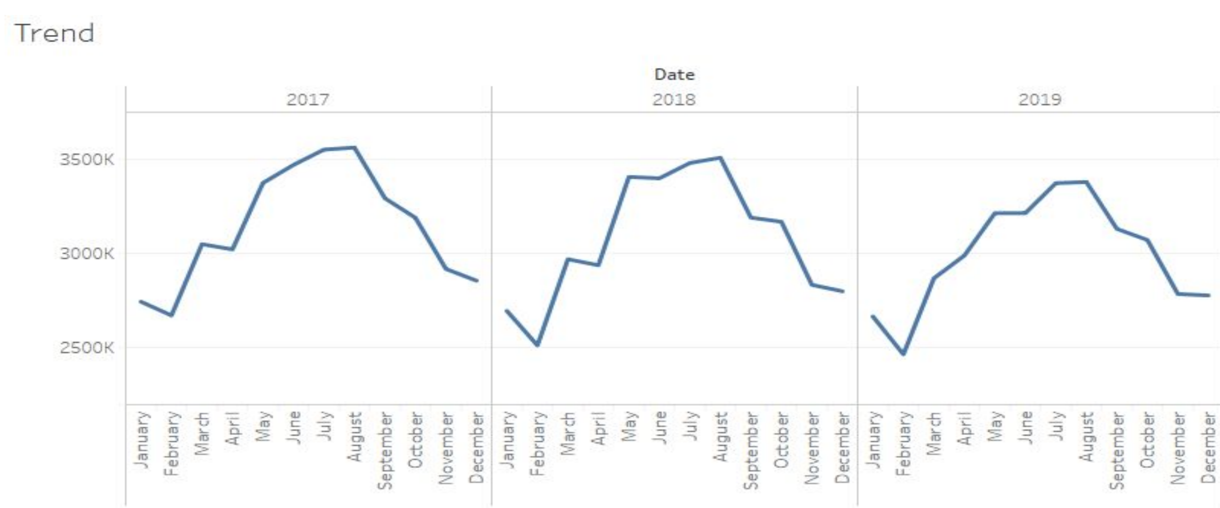
The below graphs are related to our exploratory analysis to see whether there is any correlation between the social media activity and sale transactions with one week and two week offsets. We couldn’t find any positive correlation during this phase of study.





Time Series Decomposition

With the sales transactions, we were able to see a clear seasonality where the monthly sales every year were found to be similar. Summer months have the highest sales and winter months have lower sales. We also found a trend where the sale was increasing from winter months to summer months and decreasing from summer to winter months.

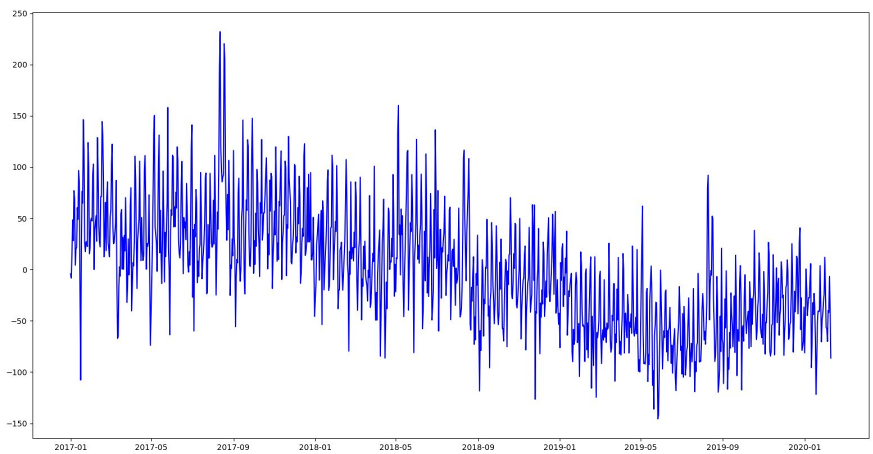


Since our eventual idea is to see whether the social media activity has any effects on the outcome, we want to remove this seasonality and trend. Time series decomposition would remove this seasonality and trend we found with sales transactions.

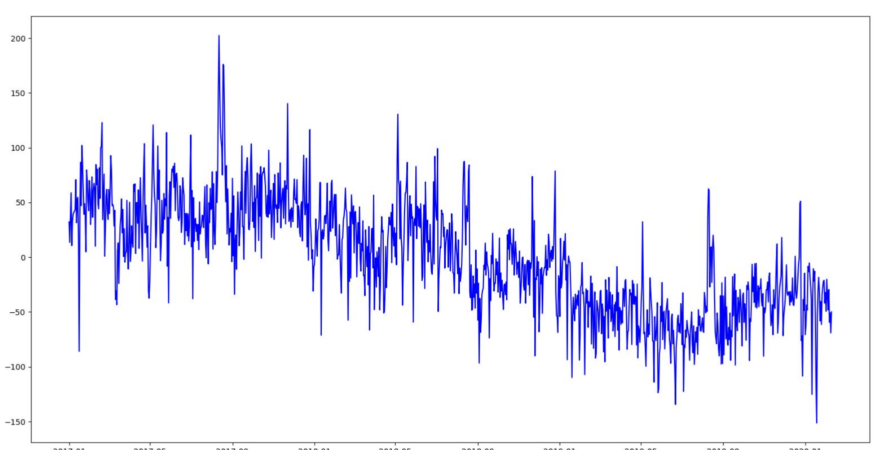
Seasonality

	Date											
Year of Date	January	February	March	April	May	June	July	August	September	October	November	December
2017	2,740,052	2,666,969	3,046,142	3,018,312	3,371,843	3,467,959	3,549,404	3,560,192	3,290,255	3,186,736	2,914,245	2,852,128
2018	2,690,809	2,507,638	2,966,032	2,934,255	3,404,041	3,396,717	3,478,370	3,506,068	3,187,698	3,165,301	2,829,756	2,795,674
2019	2,661,525	2,460,014	2,864,699	2,985,710	3,211,190	3,211,471	3,371,052	3,377,022	3,127,596	3,068,549	2,781,139	2,773,281

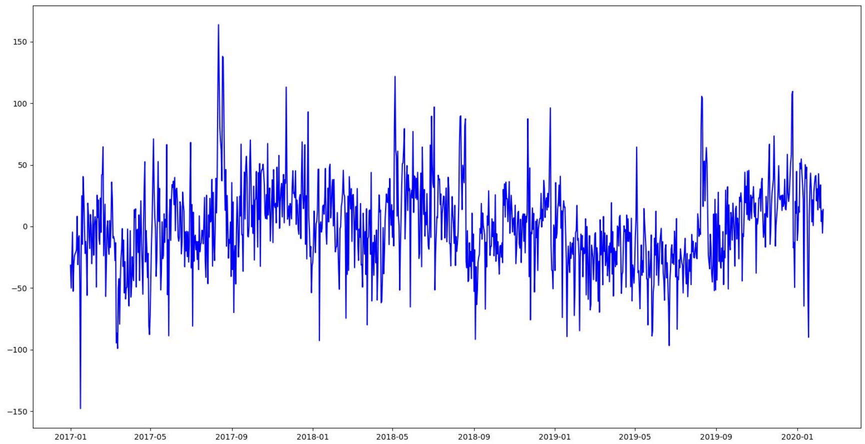
Using just the Pleasantville store again as an example we can see how each type of smoothing was removed one at a time. First we fit a polynomial curve to the data to account for 12 months. The below graph shows Pleasantville with monthly seasonality removed:



You can see that even with the monthly seasonality there are still a lot of swings up and down in the data. Turns out if you zoom in and count the peaks, they are every seven days… on Fridays. So we fit another polynomial curve to account for seven days. Then used that to do another smoothing.



This one doesn't seem much different. But if you examine the peaks and count their cadence you'll find them not set to every Friday now. Peaks for the week were much more varied. You might also notice that you can still see the downward trend. So the next step was to fit a line to the trend and then use that for the last level of smoothing.



Now finally for one store we had data that had no seasonality or trends! Of course this was only one store.   In order to do this for all 303 stores we wrote a Python script to iterate over each store, do the smoothing and then rewrite the dataset for use in further analysis.

Social Media Scoring

Kum & Go is active on three social media platforms; Facebook, Twitter and Instagram. It was unable to link a specific social media platform activity with the outcome of sales or loyalty or ratings. So for some of our analysis, we needed a way to combine the social media activity on the assumption that the outcome was based on a combined effect.

Our thought was that if we could boil some or all of the social media metrics for any given day down into a social media score, then we could use that score in our analysis. For instance on days where there was a high social media score we hoped to see an uptick in sales, survey results or loyalty signups.
Our first attempt at a social media score was to just take the number of raw posts for each of the three main platforms and add them all together. This was about as simple as we could get. Thinking a bit more about it we tried 'weighting' the post counts by the number of followers each media platform had. So while there might be a lot less facebook posts… they compensated by having a lot more followers that had liked their page.

For other social media scores we considered some metrics that were supposed to be important according to the sprout documentation. The two we picked were awareness and engagement. Awareness is essentially made up of:

- Impressions are how many times a post shows up in someone's timeline
- Reach is the potential unique viewers a post could have (usually your follower count plus accounts that shared the post's follower counts).

Engagement was primarily made up of:

- Likes, comments, shares and clicks

Multiple Linear Regression Analysis

Sales Analysis:

We combined all the social media data such as Twitter, Facebook, and Instagram into one data frame and added the smoothed transaction counts to the same dataframe. Then tried to fit the data using multiple linear regression using all the features of the social media data with target as smoothed transaction counts for each store. Determined the significant features of resulting models whose p-values are less than 0.05. Took the significant features and created the model with multiple linear regression. Then calculated the adjusted R-squared value of the final model for each store to see whether there is any influence of social media on transaction counts for corresponding stores . We repeated the process for each day offsets starting from day1 to day 30. This is to determine the effect of social activity on the smoothed transaction counts after 1 day after the activity, 2 days after the activity, etc upto 30 days after the activity.

Below is the heatmap of the adjusted R-squared values of the stores that has some significant values.

Stores	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	0.46282	0.46349	0.45429	0.50325	0.50199	0.55086	0.59016	0.60853	0.62754	0.65307	0.63625	0.65263	0.60278	0.58582	0.61212
424	0.24101	0.21009	0.28266	0.24091	0.23059	0.24141	0.06192	0.2133	0.22433	0.19487	0.18082	0.17217	0.11616	0.17294	0.22
601	0.18937	0.21651	0.21577	0.2237	0.23716	0.28565	0.26911	0.35054	0.30405	0.29737	0.30998	0.29123	0.23105	0.2787	0.21961
685	0.3161	0.3009	0.32517	0.28303	0.30655	0.26	0.27695	0.17774	0.19935	0.22685	0.25764	0.31357	0.26113	0.2443	0.33092
1108	0.30484	0.28022	0.31359	0.28308	0.25616	0.23449	0.24647	0.1512	0.21865	0.21636	0.24301	0.23579	0.18674	0.1935	0.03856
1111	0.29662	0.27475	0.27127	0.29238	0.23707	0.18307	0.30444	0.24979	0.21049	0.28466	0.19174	0.29573	0.28011	0.25173	0.33756
1399	0.3188	0.30833	0.3285	0.3226	0.30109	0.29972	0.31461	0.29537	0.29103	0.26776	0.20191	0.17511	0.00016	0.17783	0.15438
1423	0.35396	0.33074	0.34537	0.30777	0.31393	0.31377	0.286	0.19475	0.2531	0.22014	0.29294	0.23274	0.24788	0.30968	0.38969
1702	0.26385	0.25227	0.27378	0.2244	0.29498	0.18944	0.27473	0.05402	0.19944	0.22418	0.14287	0.04268	0.17854	0.16653	0.21764
1834	0.44176	0.42881	0.431	0.42291	0.43011	0.41086	0.41118	0.34143	0.38237	0.41655	0.28582	0.34685	0.25703	0.31441	0.34556
1849	0.2711	0.25776	0.26855	0.29872	0.27858	0.31007	0.31382	0.27539	0.1851	0.28101	0.10918	0.0443	0.0628	0.20948	0.3223
2413	0.38935	0.39302	0.40048	0.42178	0.46008	0.50414	0.3946	0.33277	0.25939	0.19372	0.15599	0.13487	0.12353	0.15036	0.31329
2590	0.28768	0.28693	0.27413	0.29316	0.33337	0.3245	0.33917	0.33779	0.29567	0.33325	0.40767	0.18911	0.07278	0.02291	0.27028
2659	0.2556	0.21748	0.15778	0.11465	0.2026	0.22049	0.14642	0.05481	0.18246	0.22357	0.17573	0.30874	0.27398	0.31606	0.35088
2890	0.36796	0.40674	0.45193	0.40454	0.46449	0.29294	0.39115	0.35088	0.39907	0.4021	0.5055	0.55055	0.55411	0.44832	0.60109
2920	0.21722	0.22676	0.19432	0.23709	0.24929	0.30337	0.2636	0.25113	0.29244	0.3103	0.31798	0.35568	0.33265	0.31038	0.28841
3169	0.24301	0.24319	0.30747	0.27891	0.30994	0.30606	0.31703	0.30524	0.30824	0.28659	0.26069	0.29875	0.25792	0.29559	0.36421
5512	0.3523	0.34068	0.32047	0.288	0.2847	0.26784	0.27695	0.20194	0.24335	0.2729	0.26167	0.29717	-0.0011	0.2111	0.16478
8437	0.43868	0.431	0.42709	0.40979	0.40926	0.42085	0.41975	0.39999	0.41406	0.43031	0.40317	0.4041	0.36939	0.36297	0.40986

Stores	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
4	0.6422	0.67352	0.81244	0.75265	0.68995	0.61336	0.62029	0.58004	0.38433	0.29393	0.29393	0.50974	0.41647	0.21953	0
424	0.26423	0.20756	0.13648	0.15483	0.22733	0.25876	0.13657	0.25487	0.0288	0.11816	0.24553	0.2276	0.28187	-0.003	0.27435
601	0.20546	0.27836	0.25078	0.27529	0.28121	0.24237	0.24329	0.11125	0.13815	0.08068	0.27597	0.27205	0.25244	0.00243	0.24176
685	0.42787	0.3143	0.39654	0.36816	0.33921	0.30552	0.4373	0.4525	0.01335	0.48799	0.52669	0.44177	0.47876	0.47995	0.53087
1108	0.24241	0.24838	0.30808	0.34929	0.30273	0.3524	0.38509	0.42474	0.29127	0.35361	0.13653	0.43381	0.43445	0.45897	0.48421
1111	0.32984	0.3192	0.32361	0.27143	0.27327	0.18104	0.37753	0.08722	0.35287	0.43451	0.0939	0.29417	0.34883	0.29605	0.23723
1399	0.13452	0.00612	0.02029	0.00627	0.00232	0.01914	0.06224	0.04047	0.00084	0.02164	0.05006	0.02219	0.10635	0.06613	0.05023
1423	0.44781	0.4538	0.44627	0.39067	0.25536	0.39297	0.51038	0.50002	0.46863	0.5502	0.49591	0.28049	0.4973	0.44746	0.47144
1702	0.11722	0.15046	0.1057	0.17574	0.09463	0.11329	0.07304	0.0299	0.01218	0.14947	0.21173	0.16503	0.12121	-0.0008	0.1259
1834	0.35807	0.39531	0.47929	0.44686	0.38808	0.39498	0.34825	0.28545	0.30878	0.37612	0.37011	0.38185	0.39838	0.36483	0.45074
1849	0.42204	0.46747	0.41136	0.4596	0.35633	0.3649	0.38651	0.52222	0.56357	0.52437	0.58602	0.49053	0.46454	0.5937	0.58326
2413	0.40018	0.53428	0.53216	0.46953	0.45141	0.52011	0.69827	0.74368	0.51297	0.60282	0.62805	0.38564	0.66646	0.72971	0.81787
2590	0.23887	0.35534	0.38756	0.54874	0.50821	0.50053	0.36922	0.32062	0.504	0.73394	0.54695	0.58878	0.53613	0.40766	0.54337
2659	0.38835	0.43818	0.43765	0.39292	0.43534	0.45655	0.49113	0.52969	0.49914	0.53875	0.51717	0.49042	0	0.37166	0.27333
2890	0.71388	0.7119	0.76362	0.80711	0.70568	0.71244	0.73753	0.79648	0.74792	0.72457	0.83552	0.71256	0.61033	0.58195	0.61744
2920	0.28432	0.26418	0.34017	0.3268	0.3057	0.36361	0.30908	0.33283	0.33154	0.26405	0.20225	0.14497	######	0.00574	-0.0041
3169	0.35202	0.34054	0.31199	0.32389	0.20667	0.27036	0.13245	0.25639	0.03348	0.26623	0.26235	0.21942	0	0.01386	0.17538
5512	0.2296	0.24211	0.27162	0.31648	0.23408	0.15771	0.1718	0.17144	0	0.19642	0.02699	0.00649	0	0.01818	0.22255
8437	0.39078	0.36997	0.37782	0.3958	0.36343	0.39964	0.37124	0.35147	0.34354	0.01152	0.32674	0.26073	0.16826	0.01588	0.03909

Among all the stores, Stores 4 (Hampton,IA) and 2890 (Kemmerer,WY) have good fit of social media activity for sales transaction counts. For Hampton store the best fit was found for sales transactions after 18 days of social media activity and for store Kemmerer the best fit was found for the sales transactions after 26 days of social media activity.

Alternately joined the social media data 'ActivityScore', 'WeightedActivityScore', 'TotalImpressions' with smoothed transaction counts and created a dataframe. Then tried to fit the data using multiple linear regression with target as smoothed transaction counts with different levels store, city, and region. Below is the heatmap of the adjusted R-squared values of the stores, city, state and region that has some significant values. Unfortunately, we could not find a strong adjusted R-squared value for any of the models at different levels of aggregations for sales transaction counts.

Store	R Squared_adj	RMSE (1367.0 - 22852.0)
All	0.067512895	131367.5359
4	0.259354111	747.8628524
13	0.114636269	539.6613092
88	0.083376907	618.7064036
97	0.148998089	559.1455689
130	0.131182171	579.2231907
139	0.053343872	372.4390483
151	0.117246487	469.7779149
154	0.083035185	644.973477
433	0.070886008	515.0593543
643	0.100939308	206.3580105
649	0.134726243	1156.408908
667	0.043551986	488.257464
907	0.106265013	759.6323722
916	-0.020984438	969.5525901
940	0.070016197	564.3806611
949	0.018425739	463.8951923
1045	0.105883528	809.030306
1102	0.078093294	477.2683748
1105	0.154505307	440.6147307

State	R Squared_adj	RMSE (2203.0 - 812628.0)
All	0.067512895	131367.5359
AR	0.077240169	20901.45681
CO	0.033459765	21459.16659
IA	0.078630881	38151.11766
MN	0.09899082	782.1970749
MO	0.068616988	19654.17032
MT	0.120855563	579.8693123
ND	0.066003483	3234.079354
NE	0.071759301	8545.173902
OK	0.074971614	16263.21939
SD	0.088231597	3089.99902
WY	0.054787727	5323.202547

Region	R Squared_adj	RMSE (204474.0 - 1381534.0)
All	0.049959494	305344.8417
CENTRAL	0.078406238	50881.84405
SOUTH	0.073756088	54725.66241
WEST	0.040703815	29591.59071

City	R Squared_adj	RMSE (1367.0 - 242036.0)
All	0.067512895	131367.5359
Ames	0.127692736	2695.775724
Ankeny	0.072851182	2479.942459
Evans	-0.020984438	969.5525901
Fairfield	0.11529883	483.3426229
Gillette	0.086739755	1107.083785
Hampton	0.259354111	747.8628524
Harrison	0.0862715	429.2132152
Lowell	0.058107695	1437.621194
Marion	0.025845529	1164.619695
Norwalk	0.064299866	849.4096775
Okoboji	-0.006724381	557.5646451
Plentywood	0.120855563	579.8693123
Silt	0.046279872	437.310844
Silverthorne	0.0164586	755.7866006
Tiffin	0.069860742	537.7373861
Tioga	0.136167412	700.4757065
Tipton	0.093406417	591.8159437
Windsor Height	0.051325783	496.9934671
Winterset	0.100107933	649.3943909

Loyalty Analysis:

Joined the combined social media data with loyalty numbers and created a dataframe.Then tried to fit the data using multiple linear regression using all the features of the social media data with target as loyalty for each state. Determined the significant features of resulting models whose p-values are less than 0.05. Took the significant features and created the model with multiple linear regression. Then calculated the adjusted R-squared value of the final model for each state to see whether there is any influence of social media on loyalty for the corresponding state.

Below is the heatmap of the adjusted R-squared and RMSE values, and significant features of the model for the states that have some significant values.

State	rsquared_adj	RMSE (0 - 882)	Significant Features
CO	0.733250314	79.58993316	['Twitter_Following', 'Twitter_Engagement Rate (per Impression)', 'Facebook_Total Fans', 'Facebook_Page Actions', 'Instagram_Followers', 'Instagram_Followers Gained', 'Instagram_Followers Lost', 'Instagram_Following', 'Instagram_Engagements', 'Instagram_Likes', 'Instagram_Saves']
IA	0.760926659	92.16274037	['Twitter_Following', 'Facebook_Total Fans', 'Instagram_Followers', 'Instagram_Saves']
MO	0.533591935	68.78355311	['Twitter_Following', 'Facebook_Total Fans', 'Facebook_Page Actions', 'Instagram_Engagements', 'Instagram_Likes', 'Instagram_Comments']
ND	0.601611962	21.63298547	['Twitter_Following', 'Instagram_Followers']
OK	0.75026349	42.53217287	['Twitter_Following', 'Facebook_Total Fans', 'Facebook_Page Actions', 'Instagram_Followers']
SD	0.724326488	13.41028382	['Twitter_Following', 'Twitter_Video Views', 'Twitter_Engagement Rate (per Impression)', 'Facebook_Total Fans', 'Facebook_Shares', 'Instagram_Followers']

Among all the states, Iowa and Oklahoma states have good fit for social media activity for loyalty.

Rating Analysis:

Joined the social media data 'ActivityScore', 'WeightedActivityScore', 'TotalImpressions' with nps score (rating) numbers and created a dataframe.Then tried to fit the data using multiple linear regression with target as nps score for each store. Determined the significant features of resulting models whose p-values are less than 0.05. Took the significant features and created the model with multiple linear regression. Then calculated the adjusted R-squared value of the final model for each store to see whether there is any influence of social media scores on nps rating for the corresponding store.

This is the heatmap of the adjusted R-squared and RMSE values, and significant features of the model for the stores that have some significant values. Among all the stores, store1462 (Webb City, MO) has a good fit of social media activity for rating score.

Store	rsquared_adj	RMSE (50.0 - 100.0)	Significant Features
154	0.297384423	19.92476139	['ActivityScore']
277	0.430540966	9.132757183	['ActivityScore']
364	0.245788803	8.388170671	['ActivityScore']
466	0.328073159	12.29431818	['ActivityScore']
478	0.25861597	4.498202413	['ActivityScore']
511	0.34870198	7.303355357	['ActivityScore']
604	0.271780399	8.308289071	['ActivityScore']
667	0.244861629	8.732741542	['ActivityScore']
751	0.249342978	6.539252253	['ActivityScore']
1198	0.301639649	8.777319941	['ActivityScore']
1219	0.235689847	12.25743478	['ActivityScore']
1408	0.386611457	22.68529791	['ActivityScore', 'WeightedActivityScore']
1459	0.40363327	9.877173523	['ActivityScore']
1462	0.568508742	11.80084176	['ActivityScore']
1477	0.328902045	12.94549729	['ActivityScore']
1618	0.324438606	9.362618014	['ActivityScore']
2335	0.230302181	22.45710651	['WeightedActivityScore']
2425	0.292882747	34.35923854	['ActivityScore', 'WeightedActivityScore']
2548	0.212425281	8.89760519	['ActivityScore']
2806	0.224328068	11.01711155	['ActivityScore']

Correlation Analysis

For correlation analysis we took our previous media scores and tried to correlate it to higher smoothed transactions, loyalty scores, and survey results. The first issue with this was that we anticipated that there would be some lag… but how much lag. After a good day of social media activity, how many days should pass before you would expect there to be an impact? The fact of the matter was that we didn't have any idea… and it could be different for different areas… independent variables, etc. We ultimately decided that we would have to do it programmatically and over a range of offsets.

So for sales transactions for instance we calculated correlations between different types of social media activity scores and transactions for the current day, the next day, the day after that… and so on all the way to 30 days. We did similar work for loyalty sign ups and survey results.

The nice thing about this approach was that the correlations scores could be visually represented as a matrix. The matrix output lent itself well to a heatmap!

In addition to the date offsets. We also checked for correlations at different aggregation levels. For instance for transactions (which were tracked by store) we had the ability to roll those up to city, state and region levels for instance. Checking for correlations at these different aggregations allowed us to check for correlations that might exist at higher aggregations that were getting lost when trying to look into the relationships at too fine a resolution.

The resulting matrix heatmap was 31 by 494. There was one of these for each variation of the social media score we came up with (so 4). Good thing correlations are fast to calculate!

Sales Analysis:
Unfortunately, none of the social media activity scores for any offset within 30 days had any significant negative or positive correlations. Here's a small section of the matrix. The coloring is scaled for bright green for positive correlations and bright red for negative correlations. As you can see there aren't any bright colors.

Level	ID	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Region	CENTRAL	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
Region	SOUTH	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.050	0.050	0.050
Region	WEST	0.038	0.037	0.037	0.036	0.036	0.035	0.034	0.034	0.033	0.033	0.032	0.032	0.031	0.031	0.030	0.029	0.029
Market	NE CENTR	0.104	0.113	0.121	0.119	0.117	0.115	0.083	0.051	0.019	0.053	0.086	0.120	0.120	0.120	0.120	0.111	0.103
Market	CENTRAL I	0.035	0.036	0.038	0.039	0.041	0.042	0.044	0.045	0.046	0.048	0.049	0.051	0.052	0.054	0.055	0.056	0.058
Market	DES MOIN	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067
Market	SOUTH IN	-0.023	-0.016	-0.010	-0.003	0.003	0.010	0.016	0.018	0.019	0.021	0.022	0.024	0.025	0.027	0.013	0.000	-0.014
Market	IOWA CIT	0.021	0.024	0.028	0.032	0.035	0.039	0.042	0.039	0.036	0.033	0.031	0.028	0.025	0.022	0.009	-0.004	-0.017
Market	AMES	0.024	0.022	0.020	0.019	0.017	0.019	0.020	0.021	0.023	0.024	0.005	-0.015	-0.035	-0.055	-0.074	-0.040	-0.006
Market	NW CENT	0.043	0.050	0.058	0.065	0.054	0.044	0.034	0.023	0.002	-0.019	-0.040	-0.062	-0.038	-0.014	0.011	0.035	0.039
Market	SE CENTR	0.056	0.046	0.061	0.076	0.038	0.001	0.052	0.103	0.098	0.094	0.073	0.052	0.042	0.032	0.044	0.055	0.049
Market	LITTLE RO	0.071	0.068	0.064	0.061	0.057	0.054	0.050	0.047	0.043	0.039	0.041	0.043	0.045	0.047	0.049	0.050	0.052
Market	SIOUX CIT	0.051	0.059	0.068	0.062	0.056	0.050	0.016	-0.017	-0.051	0.003	0.056	0.110	0.094	0.079	0.064	0.043	0.022
Market	FRONT RA	0.014	0.008	0.003	-0.003	-0.009	-0.015	-0.021	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.027	-0.040
Market	DENVER	0.025	0.026	0.026	0.027	0.019	0.011	0.003	-0.005	-0.028	-0.051	-0.074	-0.096	-0.053	-0.010	0.034	0.077	0.059
Market	OMAHA	0.096	0.093	0.091	0.088	0.086	0.083	0.081	0.078	0.076	0.073	0.071	0.068	0.065	0.063	0.060	0.062	0.063
Market	JONESBO	0.049	0.047	0.046	0.044	0.043	0.042	0.040	0.039	0.037	0.036	0.034	0.033	0.033	0.034	0.034	0.034	0.035
Market	SOUTH TO	0.033	0.037	0.041	0.045	0.049	0.053	0.057	0.061	0.065	0.069	0.073	0.078	0.074	0.070	0.066	0.062	0.058
Market	NW ARKA	0.054	0.053	0.052	0.051	0.050	0.049	0.048	0.046	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038	0.037
Market	SPRINGFIE	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
Market	JOPLIN	0.027	0.031	0.035	0.039	0.043	0.047	0.051	0.047	0.042	0.038	0.034	0.030	0.026	0.022	0.006	-0.010	-0.025
Market	CEDAR RA	0.071	0.070	0.068	0.067	0.065	0.064	0.066	0.068	0.070	0.072	0.074	0.076	0.049	0.023	-0.004	-0.030	-0.057
Market	SIOUX FAL	0.116	0.110	0.104	0.098	0.092	0.086	0.080	0.083	0.085	0.088	0.090	0.092	0.095	0.097	0.071	0.044	0.017
Market	COLORAD	0.037	0.035	0.033	0.031	0.029	0.027	0.025	0.023	0.021	0.019	0.017	0.015	0.013	0.010	0.011	0.011	0.012
Market	SW CENTR	0.039	0.061	-0.053	0.057	0.066	-0.011	0.041	0.034	-0.016	-0.039	-0.020	-0.025	-0.062	-0.016	0.027	0.039	-0.023
Market	WILLISTO	0.075	0.072	0.069	0.066	0.062	0.059	0.056	0.053	0.056	0.058	0.061	0.064	0.067	0.069	0.072	0.075	0.065

The colors are very subdued because they are all close to 0. So basically it found no correlations.

Loyalty and Survey Analysis:

The next two targets were more interesting… The correlations tended to be a lot more likely to be positive or negative which was great. But they were all over the place with seemingly little rhyme or reason. Also, awareness and engagement seemed to correlate much better than just generic posts (even the weighted ones) when trying to use a social media activity score.

Here's engagement vs loyalty. We can show the entire thing because it had to be aggregated by state and weeks. Since we were looking at weeks the offsets only take us out about 4 periods.

State	0	1	2	3	4
AR	0.100212	0.037519	0.056383	0.057926	0.044019
CO	0.032778	-0.07983	-0.03459	0.025228	0.045951
IA	0.006597	-0.05531	0.00264	0.044076	-0.02212
MN	-0.03252	0.013645	0.087865	0.096969	0.066273
MO	0.094836	0.014784	0.054477	0.095082	0.07094
ND	0.14076	0.042816	0.151555	0.173825	0.139689
NE	0.129447	-0.0193	0.017342	0.025523	-0.00081
OK	0.127682	0.019539	0.083012	0.08262	0.059343
SD	0.116485	0.081625	0.053181	0.102935	0.075684
WY	0.023497	-0.06909	-0.02397	0.059316	0.044002
MT	0.032093	0.070184	-0.0346	0.222078	0.268192

No super strong correlations here. Certainly better than just straight weighted posts from above. But you can also see that it seems to be kind of all over the place.

Here's engagement vs survey results:

City	Casper	0.288586	-0.02294	0.019619	0.089147
City	Gillette	0.130063	0.190217	0.226333	0.147047
City	Kemmerer	0.272287	0.315959	-0.03963	0.122911
City	Laramie	-0.17339	-0.10067	-0.2025	-0.45397
City	Rawlins	0.277012	0.198632	0.163492	0.038186
City	Rock Springs	-0.25948	-0.14983	-0.16153	-0.10151
City	Saratoga	0.045775	-0.09478	-0.29478	-0.27307
City	Dacono	0.026457	0.19881	-0.1708	-0.13176
City	Idaho Springs	0.021867	-0.06635	-0.24058	0.029123
City	Bevington	-0.49494	-0.64432	-0.45283	-0.41529
City	Eldora	0.197258	0.372804	0.283038	0.332894
City	Hettinger	0.093897	-0.4028	-0.05513	0.069467
City	Tioga	0.071047	0.079432	0.101863	0.148674
City	Glenwood Springs	-0.11175	-0.15208	-0.30887	-0.20699
City	Windsor Heights	0.004445	-0.5058	-0.07314	-0.12606
City	Sloan	-0.06211	0.102286	-0.34961	-0.00913
Store	4	0.00026	-0.34558	0.01143	-0.10188
Store	13	-0.02363	-0.04155	-0.05827	-0.18007
Store	25	0.003201	0.117057	0.080667	-0.00716
Store	73	-0.36032	-0.45511	-0.2812	-0.23766
Store	79	0.032016	0.042213	0.082001	0.13469
Store	85	-0.25062	-0.3668	0.034992	-0.43834
Store	88	-0.224	-0.63633	-0.25939	-0.42368
Store	97	-0.30763	-0.3534	-0.13765	-0.10178
Store	130	0.078739	-0.07629	-0.26898	-0.05666
Store	139	-0.14428	0.09906	-0.07202	0.063885
Store	151	-0.30293	-0.24888	-0.45391	-0.34663
Store	154	0.035526	0.275746	0.075452	0.263859
Store	160	-0.01708	-0.30139	0.058423	-0.18386
Store	178	-0.27781	-0.22777	-0.19717	-0.18145
Store	187	0.106353	-0.10467	0.367644	-0.01969
Store	199	-0.4001	-0.2759	-0.12682	-0.18682
Store	223	-0.3426	-0.34229	-0.23555	-0.12371
Store	226	-0.22835	-0.15444	-0.1801	-0.12304
Store	229	0.291946	-0.02327	0.168103	-0.07609
Store	241	0.466108	0.266555	0.475909	0.00935

Interesting right? There's some correlations that are way better (strongly negative or positive). But what's up with all of the seemingly sporadic results. Like why is Bevington negatively correlated while Eldora is positively correlated. It doesn't seem to be noise or random. If that were the case the correlations should be close to zero. But when you consider the original question about whether social media activity can drive some target variable like survey results…

the answer is not clear here at all. If correlations here suggest causation… then it also suggests that it's just as likely to drive it down as drive it up!

Findings and Recommendations

With the challenges we had with the dataset, it was a little difficult to accurately study the influence of social media with the target variables of sales, loyalty and nps. But with the available set, we couldn't find any relationship with sales or nps rating. We do think that the new loyalty registrations does have some positive influence with social media activity.

Recommendations for future efforts would include trying even more ways of deriving a social media activity score. Kum & Go can also collect the social media data with review comments and analyze the data using text mining and modeling to find any relation between social media data and sales transactions, loyalty, and rating.

As a team we learned on how to maximize the analysis of the datasets when we don't have significant influence of the data on the target variables.

Appendix

Given below are some of the resource repositories related to this project.

GitHub:
https://github.com/miner-league/social-media-effect-on-sales-analysis

Tableau Public:
https://public.tableau.com/profile/jerry.jacob#!/vizhome/SocialMediaExproratory/Dashboard2

Presentation Slide PDF:
https://iowa-my.sharepoint.com/:p:/g/personal/jgioimo_uiowa_edu/EewDHFQOqMZKvbldla9ZyVEBLc2gEvSalTWOh3jKMICOKw?e=g5iOo8