Facebook Ads: Paid Sales VS Organic Sales

Exploring the relationship between Paid Sales, Organic Sales and the amount spent on an ad campaign.

Facebook Ads: Paid Sales VS Organic Sales

- During a Facebook Ad campaign sales are tracked for the product being advertised
- Using google analytics we are able to track which of these sales were the result of the campaign, we call these sales Paid Sales
- We are therefore also able to observe all other sales that occurred during that time, lets call those sales Untracked Sales
- Overtime we can track the daily Spend on an Ad campaign, as well as the Paid Sales and the Untracked Sales
- Here is a sample dataset taken from a recent campaign:

	Spend	Paid Sales	Untracked Sales
Date			
2014-01-01	4542.47	46	69
2014-01-02	5109.00	71	97
2014-01-03	12240.23	123	187
2014-01-04	7395.29	54	91
2014-01-05	5442.40	45	65

The data ranges from 1-Jan-2014 to 6-May-2014

Facebook Ads: Paid Sales VS Organic Sales

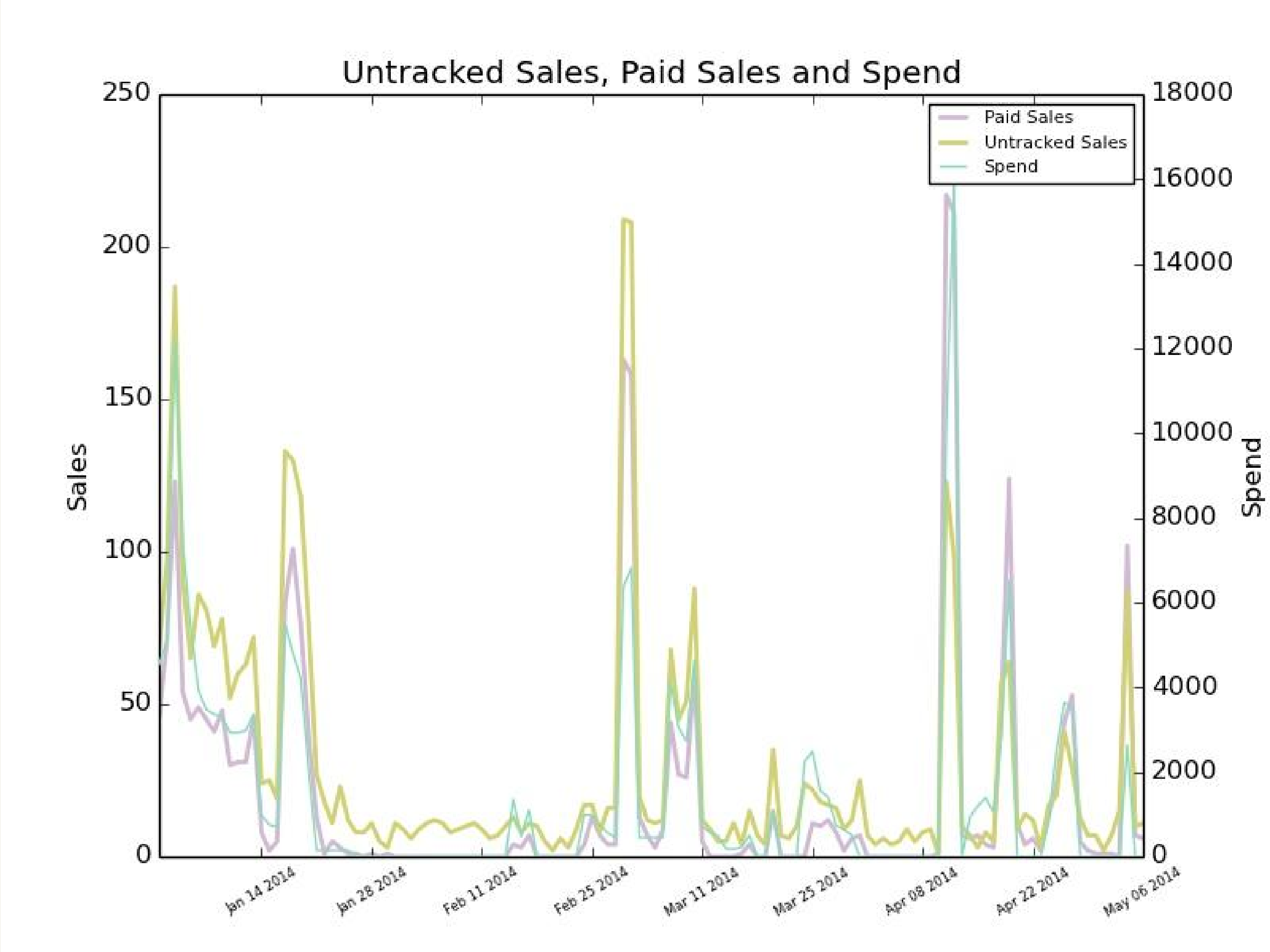
- The question is, were the Untracked Sales just Organic Sales ? i.e. Independent of campaign, or are they a direct result of the campaign but just not being picked up as Paid Sales?

GOAL: Separate Untracked Sales into Untracked Paid Sales and True Organic Sales

Untracked Sales = True Organic Sales + Untracked Paid Sales

- Let's start by observing how Untracked Sales move over time compared to Paid Sales and Spend

Facebook Ads: Paid Sales VS Organic Sales



Facebook Ads: Paid Sales VS Organic Sales

- There is a clear positive relationship between Paid Sales, Untracked Sales and Spend
- Untracked Sales appear to move much like Paid Sales, suggesting that many of the Untracked Sales are due to the campaign
- We could infer that:

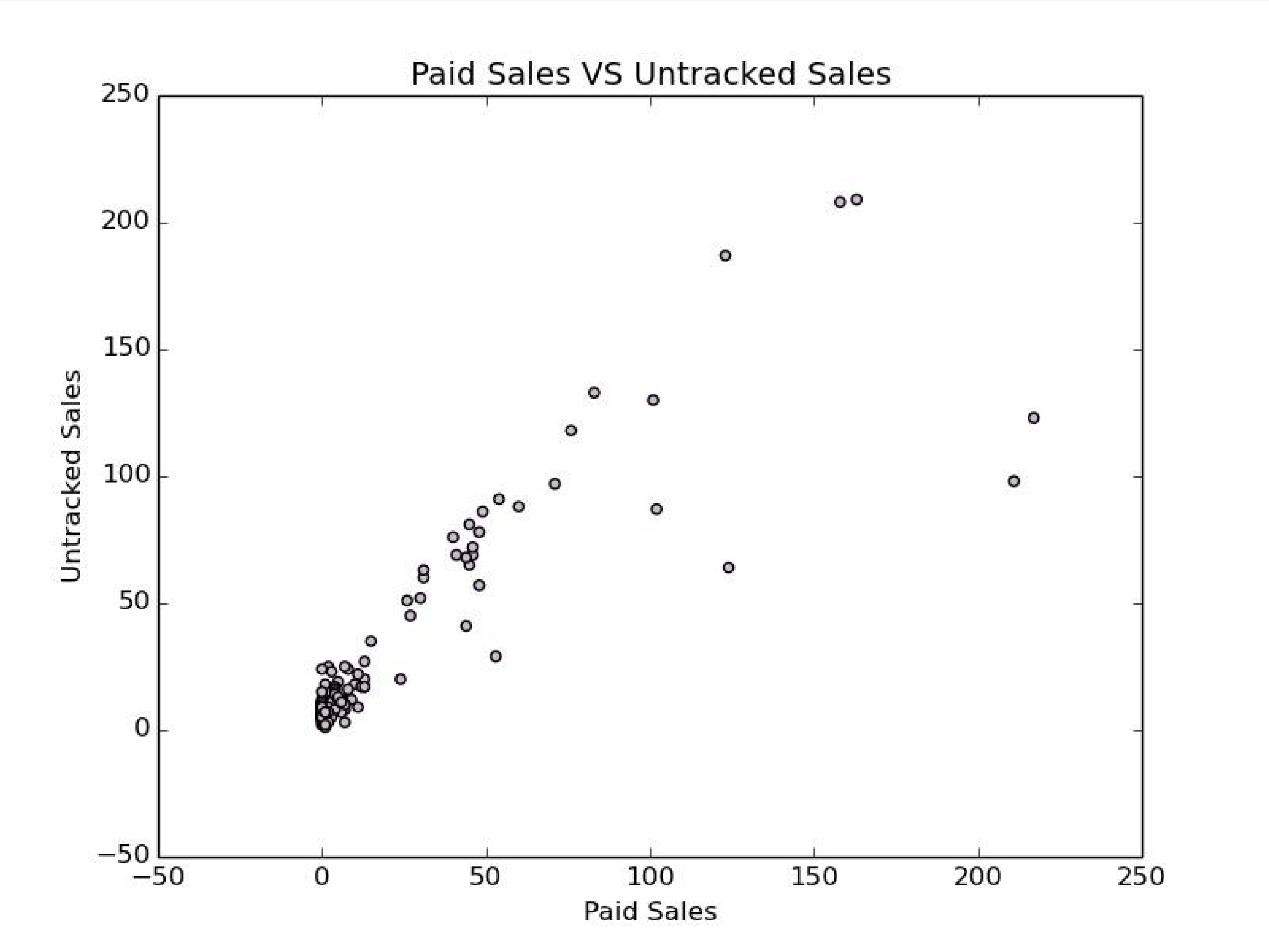
$$\text{Untracked Sales} = f(\text{Paid Sales})$$

- To gain an understanding of what "f" might be lets look at a scatter plot of Paid and Untracked Sales across the dataset. (If, for example, all the data appears to lie on a straight line then it would be reasonable to assume that

$$\text{Untracked Sales} = c_0 + c_1\text{Paid Sales}$$

where c_0 and c_1 are constant

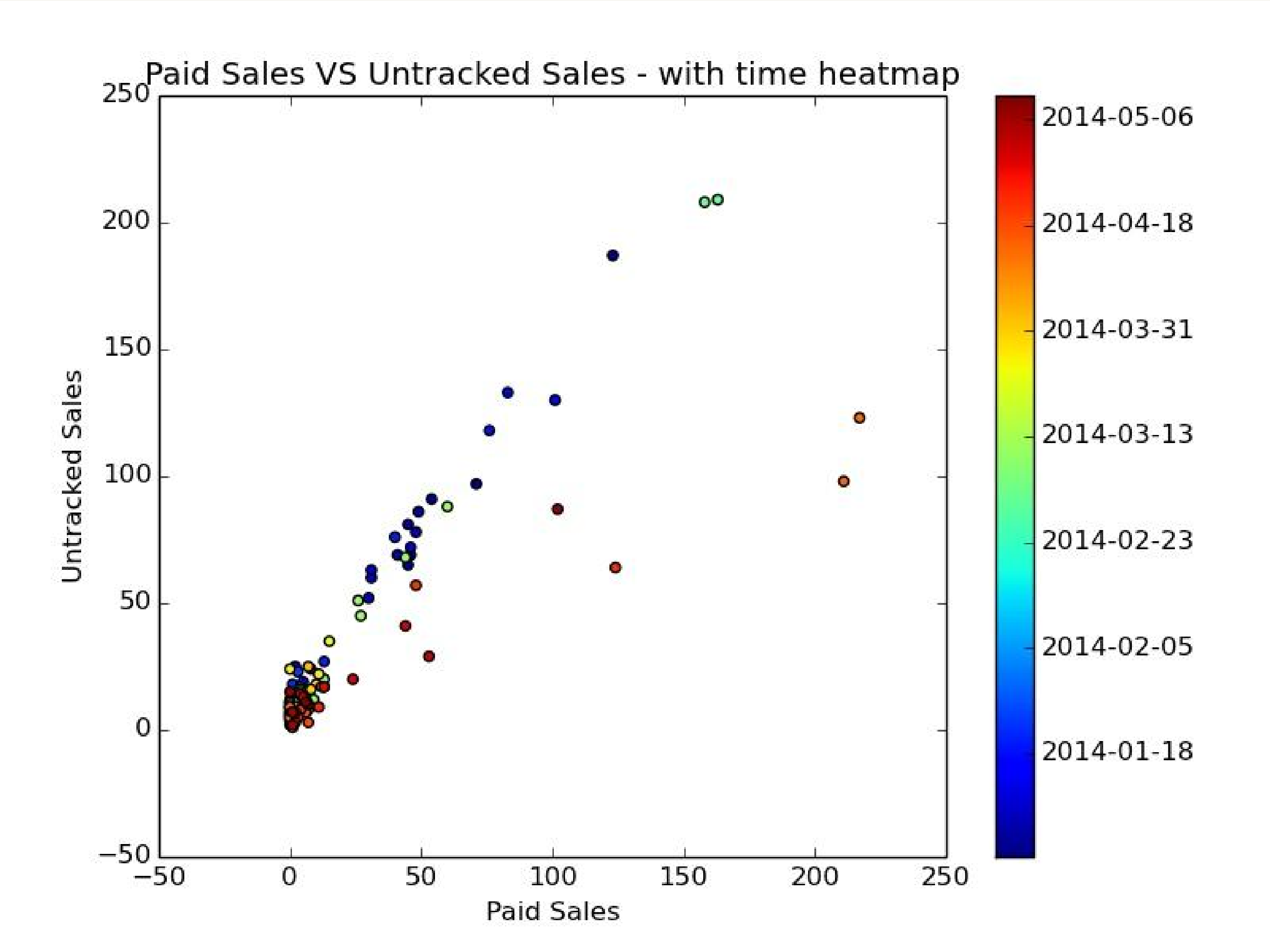
Facebook Ads: Paid Sales VS Organic Sales



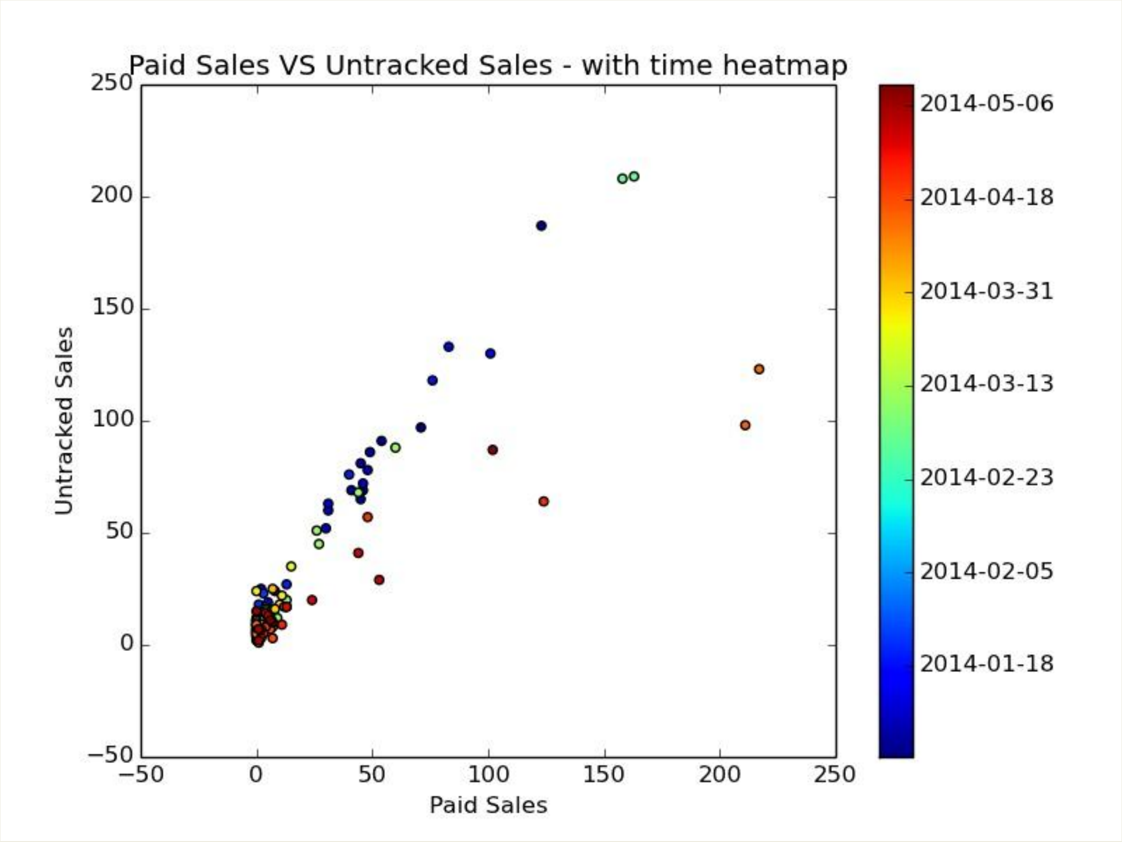
Facebook Ads: Paid Sales VS Organic Sales

- Some of the data appears to fall in a straight line but not all
- Now by adding a color map to the data we can observe how this relationship changes over time

Facebook Ads: Paid Sales VS Organic Sales



Facebook Ads: Paid Sales VS Organic Sales



- This view seems to suggest that there is a linear relationship that changes over time

i.e.

$$\text{Untracked Sales} = c_0(t) + c_1(t) \cdot \text{Paid Sales}$$

Facebook Ads: Paid Sales VS Organic Sales

Short Term model

- After observing the data it is reasonable to assume that the following relationship exists on a short term basis

$$\text{Untracked Sales} = c_0 + c_1 \cdot \text{Paid Sales}$$

- If we could find suitable values for c_0 and c_1 we could use c_0 as an estimate of True Organic Sales and $c_1 \cdot$ Paid Sales as an estimate of Untracked Paid Sales. Recall:

GOAL: Separate Untracked Sales into True Organic Sales and Untracked Paid Sales

$$\text{Untracked Sales} = \text{True Organic Sales} + \text{Untracked Paid Sales}$$
$$\text{Untracked Sales} = c_0 + c_1 \cdot \text{Paid Sales}$$

We assume that True Organic Sales are constant (short term)

We assume that Untracked Paid Sales are a multiple of Paid Sales (short term)

Facebook Ads: Paid Sales VS Organic Sales

Short Term model

- Suppose we used an OLS Linear regression model to predict Untracked Sales using Paid Sales
- Well, you might ask why? We already know how many Untracked Sales there were
- Running an OLS Linear regression model allows us to break down the obtained estimate of Untracked Sales into two components, in the exact format we would like:

$$\hat{y} \triangleq \hat{c}_0 + \hat{c}_1 \cdot \text{Paid Sales}$$

Where $\hat{y}$ represents the estimate of Untracked Sales obtained from running the linear regression model.

- At this point it is useful to go through an example with real numbers

Facebook Ads: Paid Sales VS Organic Sales

Short Term model

- Example : Suppose that we ran a regression model using Paid Sales to Predict Untracked Sales and that our model gave us values $\hat{c}_0 = 9$ and $\hat{c}_1 = 1.5$. Then suppose that, on a given day there were 20 Paid Sales. Then:

  $\hat{y} \triangleq 9 + 1.5 \cdot 20 = 39$ our model would give us an estimate of 39 Untracked Sales

  - 9 of which we can think of as the True Organic Sales of the estimate
  - And 30 of which we can think of as the Untracked Paid Sales of the estimate

  Further suppose that on the same day we knew that we actually had 35 Untracked Sales (not 39 as predicted by the model)

  Well we can still use the estimates of True Organic Sales and Untracked Paid Sales we obtained if we just rescale them by the actual number of Untracked Sales:

$$\text{Untracked Paid Sales} = 9\frac{35}{39} \approx 8, \qquad \text{True Organic Sales} = 30\frac{35}{39} \approx 27$$

i.e. 9 Untracked Paid Sales out of a possible 39 → 8 Untracked Sales out of a possible 35
i.e. 30 True Organic Sales out of a possible 39 → 27 True Organic Sales out of a possible 35

Facebook Ads: Paid Sales VS Organic Sales

<u>Short Term model</u>

- More generally then for

$$\text{Untracked Sales} = c_0 + c_1 \cdot \text{Paid Sales}$$

- Using an OLS Linear regression model we can use Paid Sales to Predict Untracked Sales:

$$\hat{y} \triangleq \hat{c}_0 + \hat{c}_1 \cdot \text{Paid Sales}$$

- And then we can find estimates of True Organic Sales and Untracked Paid Sales by rescaling our linear regression coefficients

$$\text{Untracked Paid Sales} = c_0 = \hat{c}_0 \frac{y}{\hat{y}}$$

$$\text{True Organic Sales} = c_1 \cdot \text{Paid Sales} = \hat{c}_1 \cdot \frac{y}{\hat{y}} \cdot \text{Paid Sales}$$

Facebook Ads: Paid Sales VS Organic Sales

- Now let's look at an example using our dataset:
- Take the 1st 10 rows from the dataset:

	Paid Sales	Untracked Sales
1-Jan-14	46	69
2-Jan-14	71	97
3-Jan-14	123	187
4-Jan-14	54	91
5-Jan-14	45	65
6-Jan-14	49	86
7-Jan-14	45	81
8-Jan-14	41	69
9-Jan-14	48	78
10-Jan-14	30	52

Using these data to train our regression model, we obtain the coefficients:

$c_0 = 9.325$ and $c_1 = 1.42$

Facebook Ads: Paid Sales VS Organic Sales

- Now we have an estimate of Untracked Sales on 10-Jan-14 of $9.325 + 1.42 \cdot 30 \cong 52$. Exactly matching the Actual Untracked Sales on that date, giving us an estimate of Untracked Paid Sales of 43 and True Organic Sales of 9

	Paid Sales	Untracked Sales	c0	c1	Untracked Paid Sales	True Organic Sales
1-Jan-14	46	69	N/A	N/A	N/A	N/A
2-Jan-14	71	97	N/A	N/A	N/A	N/A
3-Jan-14	123	187	N/A	N/A	N/A	N/A
4-Jan-14	54	91	N/A	N/A	N/A	N/A
5-Ja14	45	65	N/A	N/A	N/A	N/A
6-Jan-14	49	86	N/A	N/A	N/A	N/A
7-Jan-14	45	81	N/A	N/A	N/A	N/A
8-Jan-14	41	69	N/A	N/A	N/A	N/A
9-Jan-14	48	78	N/A	N/A	N/A	N/A
10-Jan-14	30	52	9.33	1.42	43	9

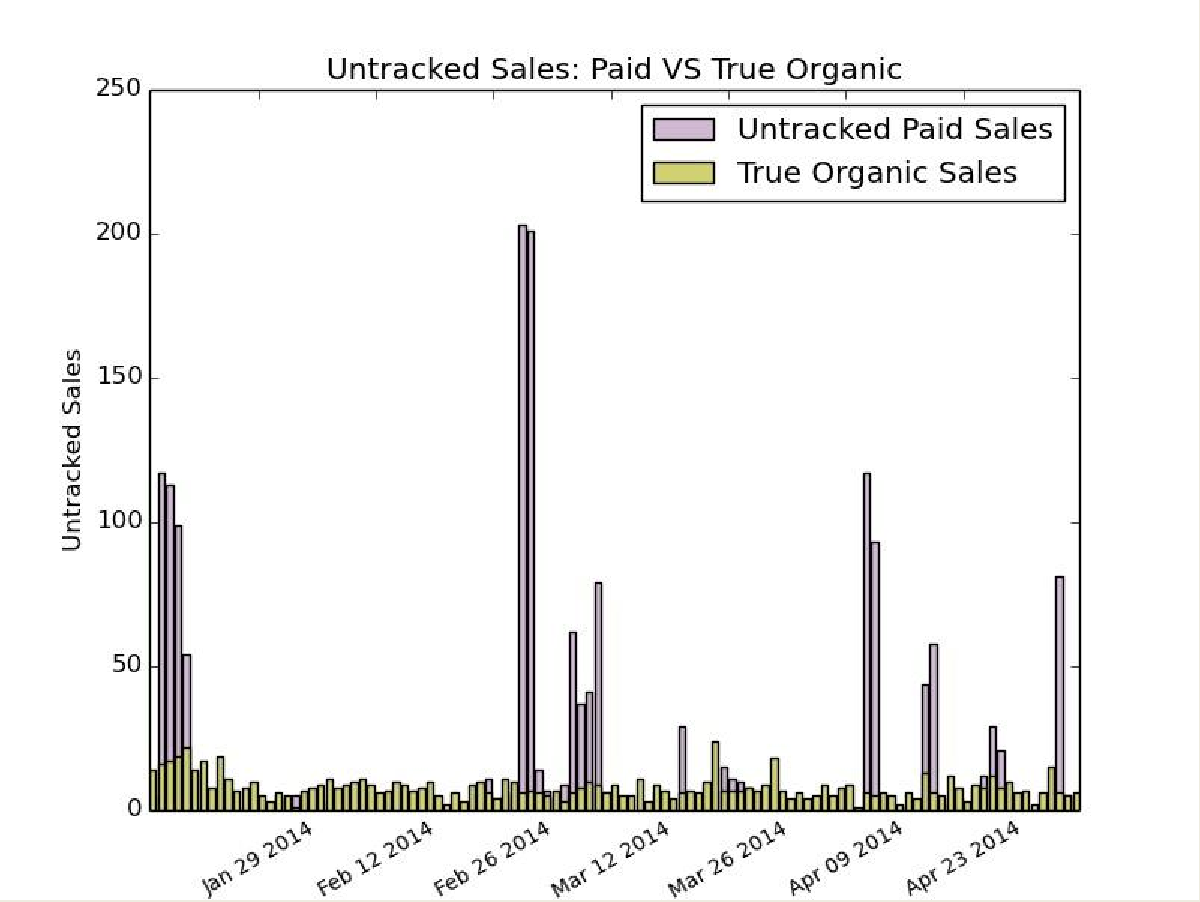
Facebook Ads: Paid Sales VS Organic Sales

- Now lets take this methodology and apply it on a 10 day rolling bases through the whole dataset 1 day at a time to obtain an estimate for Untracked Paid Sales and True Organic Sales for every day (except for the first 10 days)

	Paid Sales	Untracked Sales	c0	c1	Untracked Paid Sales	True Organic Sales
10-Jan-14	30	52	9.33	1.42	43	9
11-Jan-14	31	60	12.16	1.39	47	13
12-Jan-14	31	63	13.23	1.41	48	15
13-Jan-14	46	72	17.29	1.30	56	16
14-Jan-14	8	24	15.39	1.33	10	14
15-Jan-14	2	25	18.39	1.29	3	22
16-Jan-14	5	19	16.88	1.30	5	14
17-Jan-14	83	133	15.11	1.37	117	16
18-Jan-14	101	130	18.77	1.22	113	17
19-Jan-14	…	…	…	…	…	…

Facebook Ads: Paid Sales VS Organic Sales

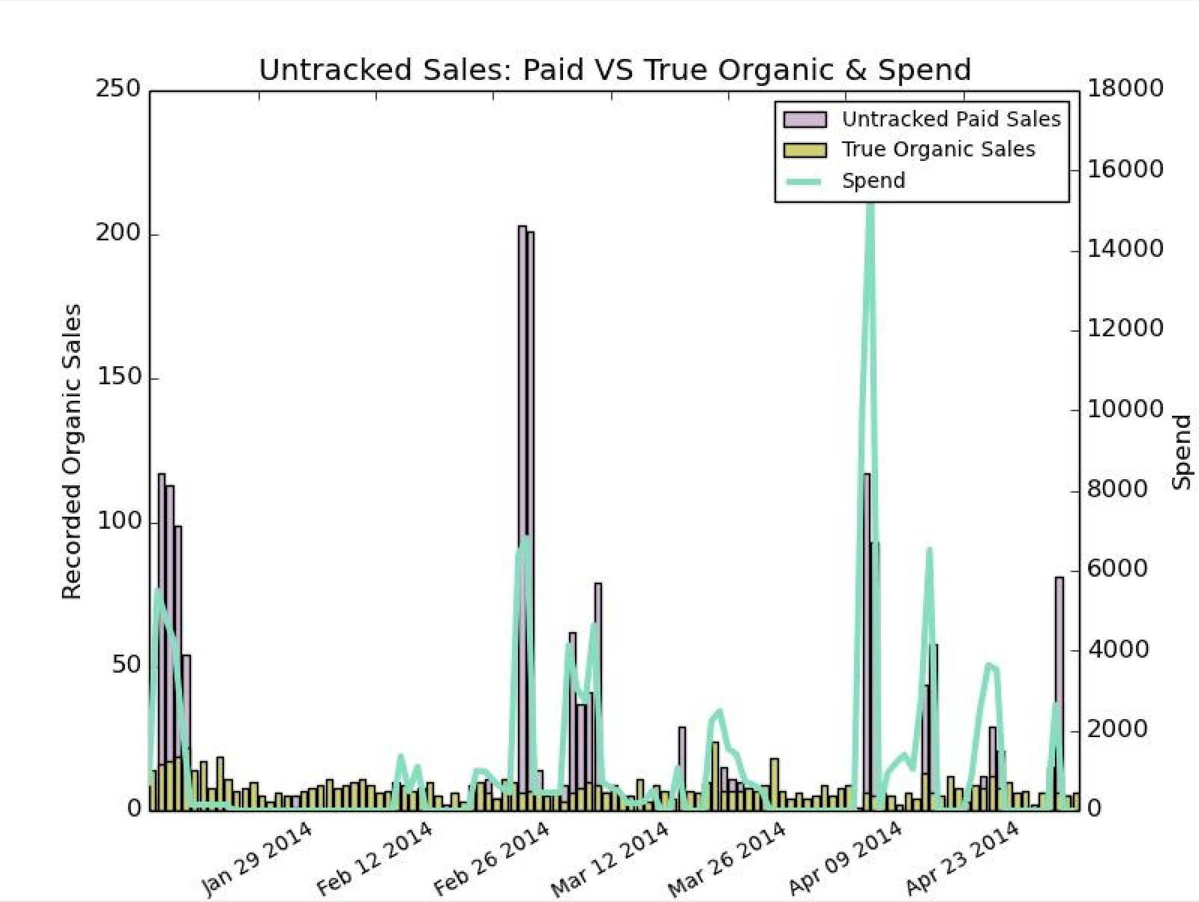
- Now lets take this methodology and apply it on a 10 day rolling bases through the whole dataset 1 day at a time to obtain an estimate for Untracked Paid Sales and True Organic Sales for every day (except for the first 10 days):



Facebook Ads: Paid Sales VS Organic Sales

- Now that we have a model in place that passes the gut check visually and gives results that make sense, the challenging next step is to find a quantitative method for validating the models accuracy given that we are unable to actually observe the actual breakdown of Untracked Sales
- However, what we do have is unused variable that is very meaningful, Spend

Facebook Ads: Paid Sales VS Organic Sales



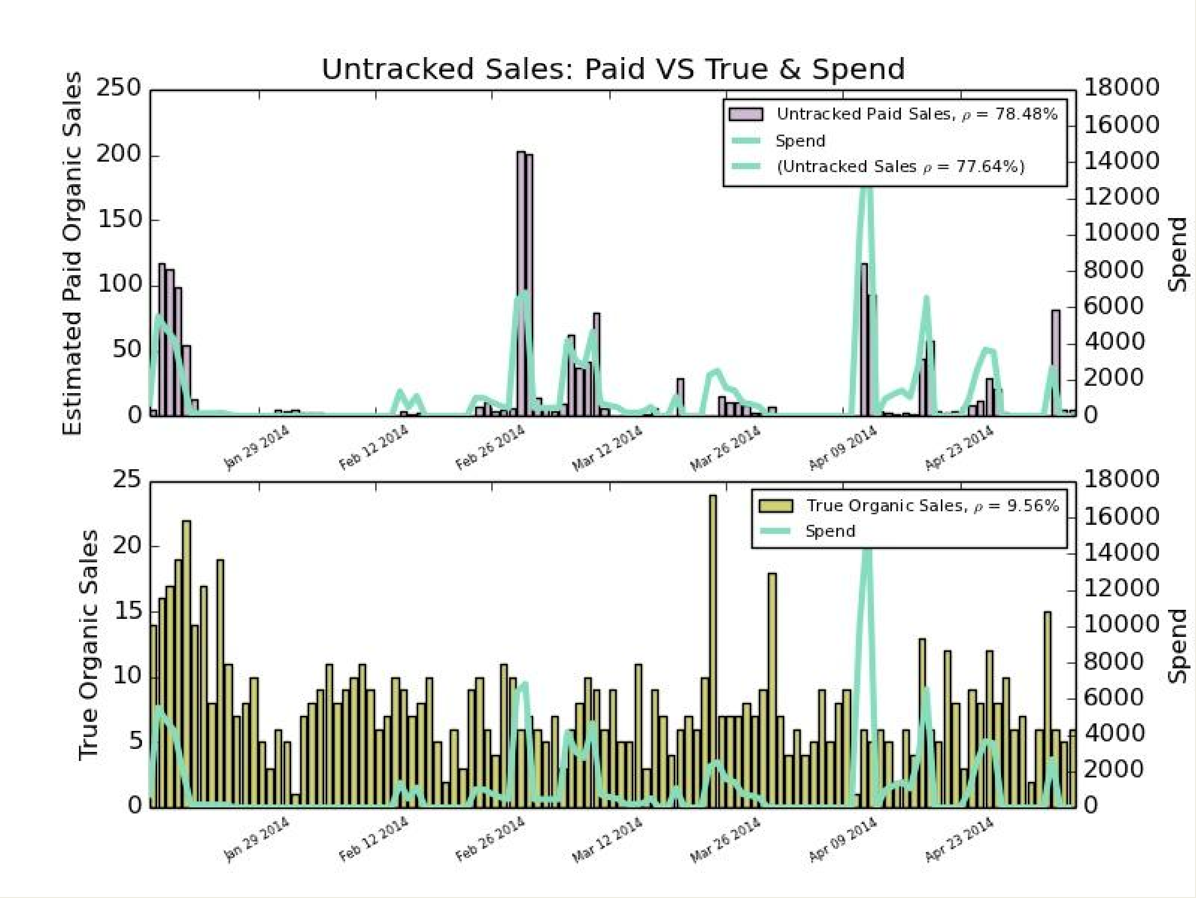
- We know overall Untracked Sales has a positive relationship with Spend (as seen above) and we would expect Untracked Paid Sales to have a stronger positive relationship with Spend and True Organic Sales to be independent of Spend

Facebook Ads: Paid Sales VS Organic Sales

- By observing the correlation between Spend and Untracked Sales, Spend and Untracked Paid Sales and, Spend and True Organic Sales we can quantify and compare these relationships

Facebook Ads: Paid Sales VS Organic Sales

- By observing the correlation between Spend and Untracked Sales, Spend and Untracked Paid Sales and, Spend and True Organic Sales we can quantify and compare these relationships



Facebook Ads: Paid Sales VS Organic Sales

- We observe a lift in correlation of 0.84% moving from Untracked Sales to Untracked Paid Sales. Moreover our estimate of True Organic Sales has a very low correlation with Spend, just 9.56%.
- Not only does correlation with Spend give us a mechanism for validating the model. We could also use it to help parameterize the model. i.e. we could use correlation with Spend to find the most suitable window length for the rolling regression model
- Here are the correlation results across different window lengths for the rolling regression:

Window Length	Paid Untracked	True Organic
5 days	78.29%	8.92%
6 days	78.40%	7.09%
7 days	78.48%	8.28%
8 days	78.51%	8.79%
9 days	78.44%	11.98%
10 days	78.48%	9.56%
11 days	78.50%	8.72%
12 days	78.45%	9.85%
13 days	78.40%	10.23%
14 days	78.26%	11.02%

Facebook Ads: Paid Sales VS Organic Sales

- Next steps:
  - Refine the model: At each given time step, verify whether the model value makes sense, could we use an alternative when we think it doesn’t? i.e. observe the most recent average Untracked Sales value when there was zero spend. Would this result in improved accuracy?
  - Testing this methodology on a new dataset to observe whether we see similar results