

Report: Data Storytelling

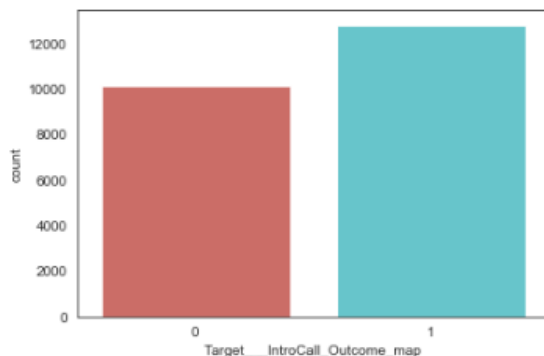
See Jupyter notebook here: [Link to Jupyter notebook in GitHub](#)

1. Ask the following questions and look for the answers using code and plots:

a. Can you count something interesting?

- i. The first question I had was how many records I had in my dataset and the number of qualified vs. disqualified demo calls.
 1. Of 22.9K records, 12.8K (56%) were qualified vs 10K (44%) disqualified. An interesting extension to this question is the volume of leads necessary to get to these qualification rates.

```
1    12822
0     10112
Name: Target__IntroCall_Outcome_map, dtype: int64
```

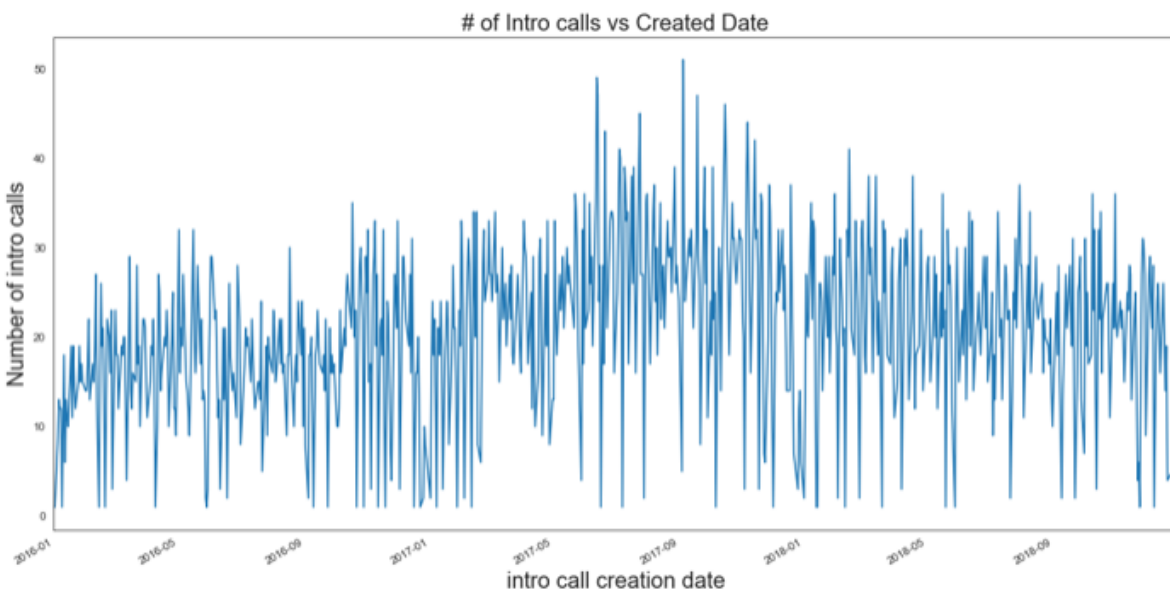


b. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?

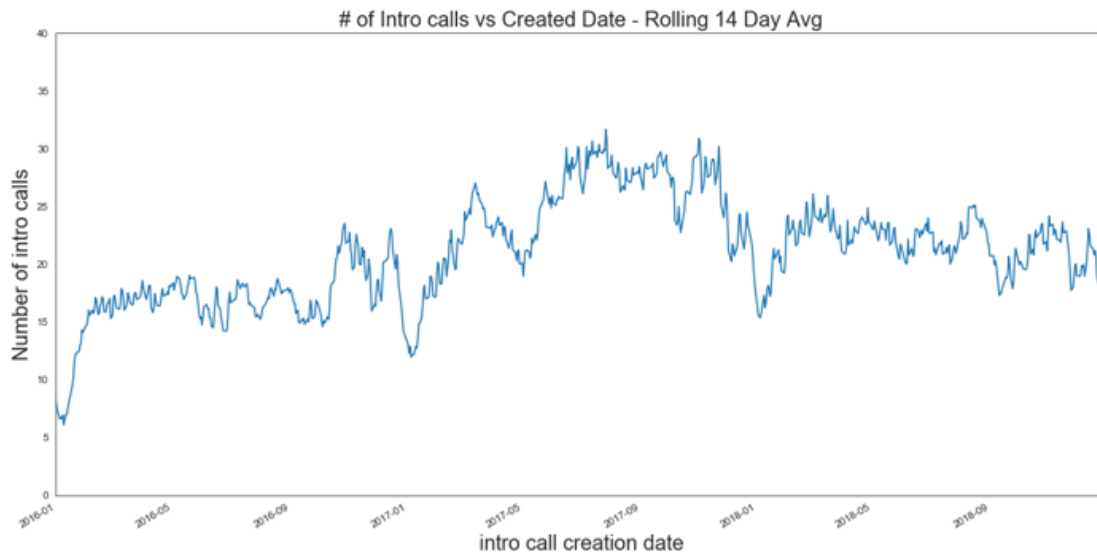
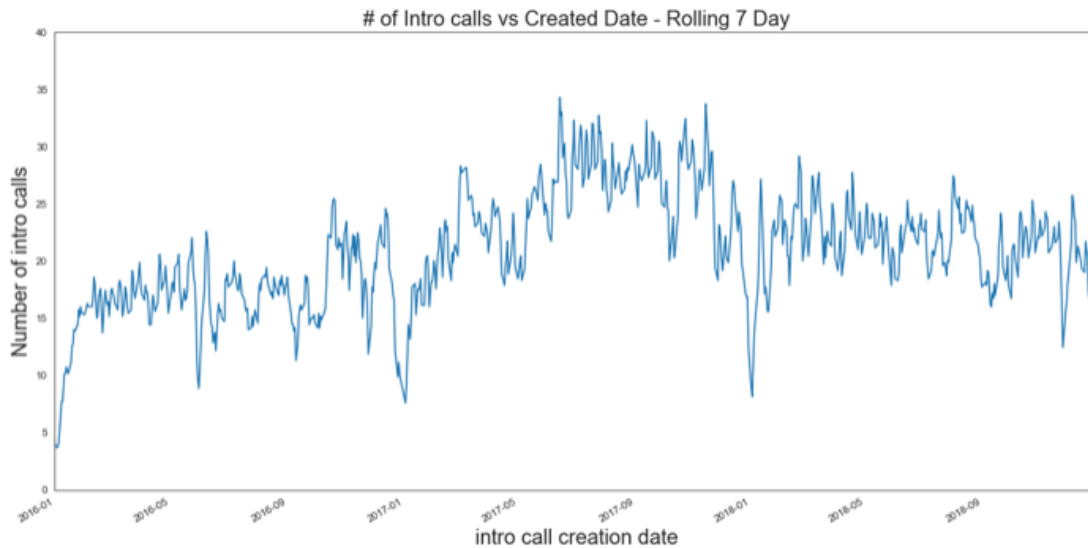
- i. Next I want to understand how the volume of intro calls has changed over time, both overall numbers and by qualification status (qualified vs disqualified).

The first graph below shows the volume of intro calls (min: ~ 0, high: ~50) versus their creation date (from Jan 2016 to Dec 2018).

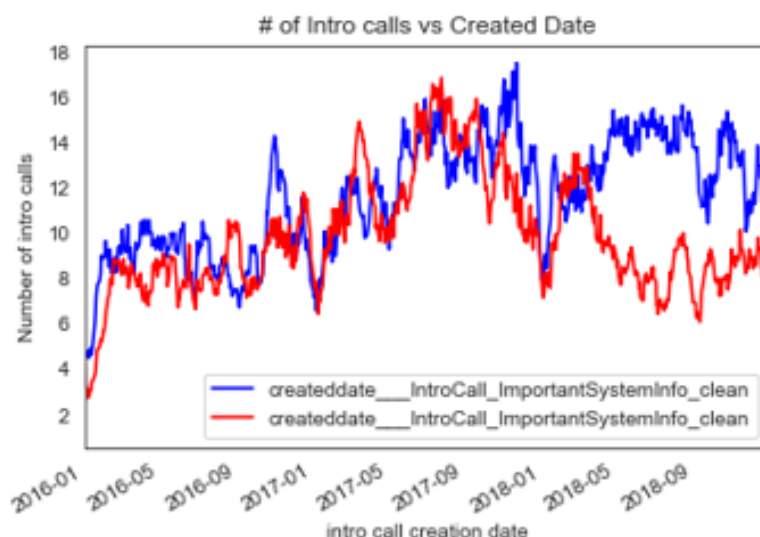
However the chart in its current state is difficult to discern for trends, so it might make sense to instead use a rolling mean.



The subsequent charts show the impact of taking a 7 day rolling average and then a 14 day rolling average. On inspection it seems like between May 2017 to Jan 2018 we had higher volumes of intro calls.



But what we're really interested in is understanding the drivers of qualified calls, so it makes sense to view the volume by time and qualification status.



Red line is “unqualified”, blue line is “qualified”. It looks like volume has remained high in the last year but disqualifieds are making up a smaller proportion (possibly reinforcing the marketing team’s assertion that they’re providing higher quality leads).

c. Can you make a bar plot or a histogram?

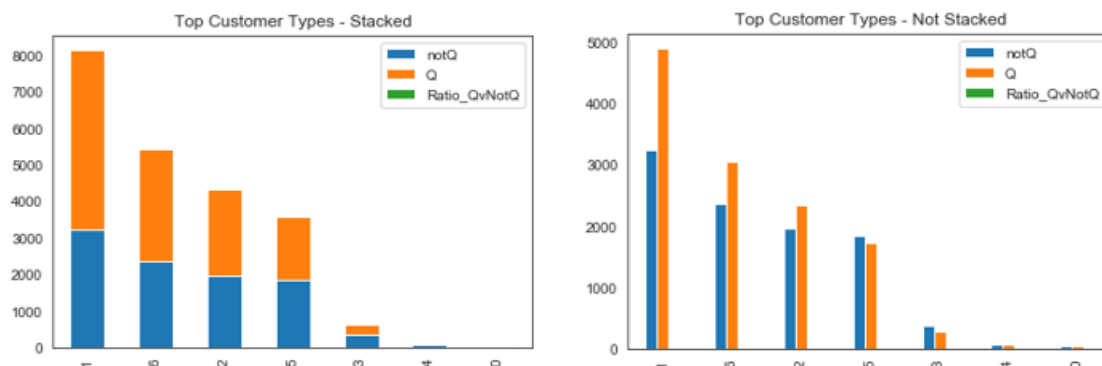
Next I want to understand the sources of the intro calls (like marketing channels, landing pages, business type, etc).



Originally I tried a stacked barchart but it really only shows the volume of intro calls by landing page. The charts 'Top 30 Lead Marketing Channels by Qualified' are sorted but the count of qualified

intro calls per traffic channel. What's interesting is that even though some marketing channels have produced a large volume of qualifieds, they aren't necessarily the same channels responsible for producing a high ratio of qualifieds to disqualifieds. For example, lead source 3 & 0 (corresponding to "Brand" and "Affiliate") have a ratio of 1.9 but are in sixth place and up, with additional intro call sources in between having a ratio of around 0.8~1.6. This is the first time I've seen the marketing funnel from the perspective of the intro calls (even within the company) so it's fascinating to see the different levels of quality.

The next question I was trying to answer was whether the customer type could be a driver of qualified intro calls. After generating the following charts, it seems that customer type could be a driver (as well as an indicator of the company's strategic focus on the enterprise space). 1 corresponds to 'Enterprise' (2 is 'Unknown', which doesn't exactly bode the best in terms of our data quality) and 4 corresponds to 'Nonprofits' (which makes sense, the company primarily markets to companies willing to invest significant resources in onboarding and digital adoption).



Countries is a little surprising as we have some EMEA and ANZ/APAC countries listed as the top producers of qualified intro calls. The company started in Israel and has major presence in AMER but it's interesting to see the UK (#111), Australia (#4), and Germany (#38) up in the top 8.

When we look at landing pages and try to create top 30 charts, we see some interesting trends where the top 30 best landing pages by qualifieds count aren't the same as the top 30 landing pages by ratio of qualified to disqualified intro calls.

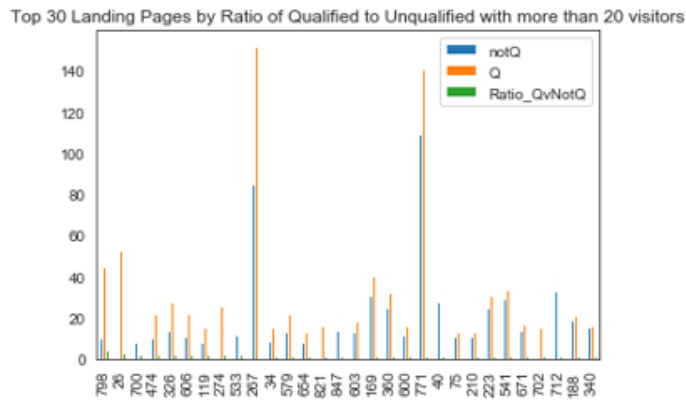
The first set of charts below show “Top 30 Landing Pages by Qualified Count” [L] and “Top 30 Landing Pages by Ratio (Qualified/Disqualified)” [R]. Landing pages had to have more than 20 visitors in order to weed out cases where only 5 people visited a landing page and all converted (or none converted). Notice how the top 5 landing pages are completely different depending on the particular cut.

Top 30 Landing Pages by Qualified Count

landingPage__Lead_MarketingInformation	notQ	Q	Ratio_QvNotQ
267 home	85.0	152.0	1.8
771 using-salesforce	109.0	141.0	1.3
245 employee-training-easier-st2-2	191.0	125.0	0.7
129 creating-online-tutorials	112.0	77.0	0.7
336 instantly-improve-ux	104.0	77.0	0.7
484 salesforce-never-easier	83.0	62.0	0.7

Top 30 Landing Pages by Ratio of Qualified to Unqualified with more than 20 visitors

landingPage__Lead_MarketingInformation	notQ	Q	Ratio_QvNotQ
798 walkme	10.0	45.0	4.5
26 Omri-SF2	18.0	53.0	2.9
700 training-walkmeblog-corporate-002-b	8.0	20.0	2.5
474 salesforce-corporatevideo-3-sf-never-easier	10.0	22.0	2.2
326 improve-ux-2-wm	14.0	28.0	2.0
606 software-corporate-4-simplify-enterprise-software	11.0	22.0	2.0
119 concur-user-onboarding-1	8.0	15.0	1.9



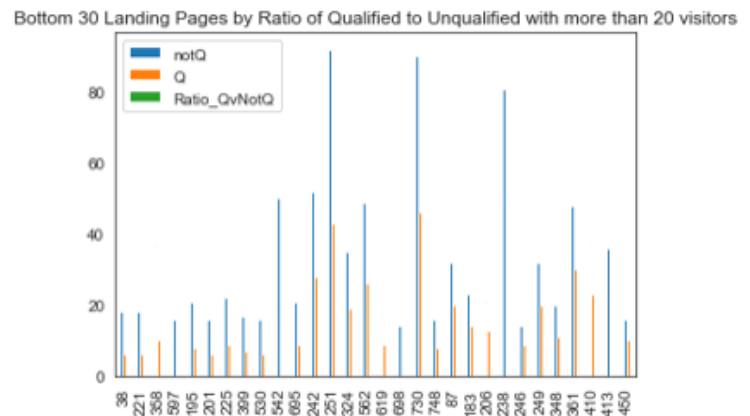
This next set of charts focuses on landing page rankings of disqualified intro calls.

Top 30 Landing Pages by Disqualified Count

landingPage__Lead_MarketingInformation	notQ	Q	Ratio_QvNotQ
245 employee-training-easier-st2-2	191.0	125.0	0.7
129 creating-online-tutorials	112.0	77.0	0.7
771 using-salesforce	109.0	141.0	1.3
336 instantly-improve-ux	104.0	77.0	0.7
251 employee-training-software	92.0	43.0	0.5
730 tutorial-creation	90.0	46.0	0.5

Bottom 30 Landing Pages by Ratio of Qualified to Unqualified with more than 20 visitors

landingPage__Lead_MarketingInformation	notQ	Q	Ratio_QvNotQ
38 SF-easier	18.0	6.0	0.3
221 elearning-solutions	18.0	6.0	0.3
358 make-tutorial-demo-tw	37.0	10.0	0.3
597 software-corporate-2-eliminate-frustration	16.0	5.0	0.3
195 dynamics-training-2-wm	21.0	8.0	0.4
201 effective	16.0	6.0	0.4

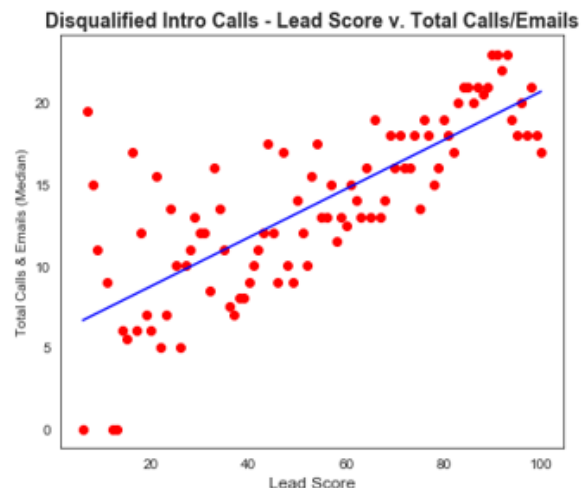
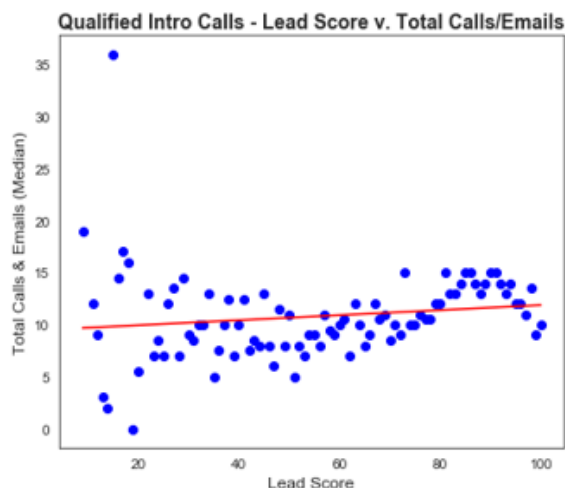


d. Can you compare two related quantities?

See previous charts

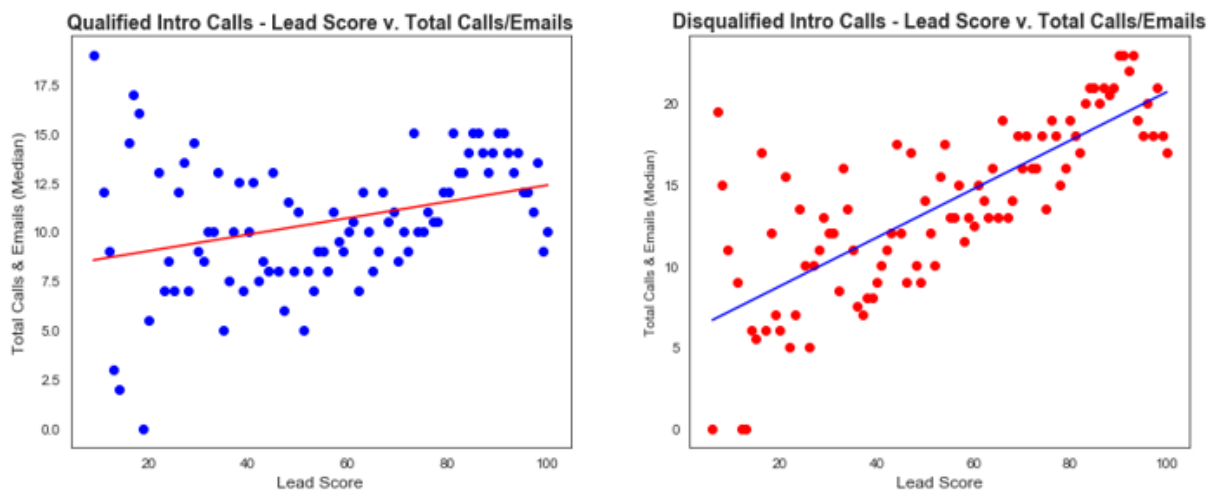
e. Can you make a scatterplot?

Next I want to understand the relationship between the marketing stamped lead score and the total calls & emails logged against that lead. Generating the charts below, it's surprising to see that the greater the lead score, the more engagement is needed for that lead. After separating out the qualified and disqualified intro calls we notice there are some outliers (especially in the Qualifieds chart), where a lead score (of around 20) had +35 emails and calls logged. We don't want to just address that one point, however, but a way to systematically address outliers.



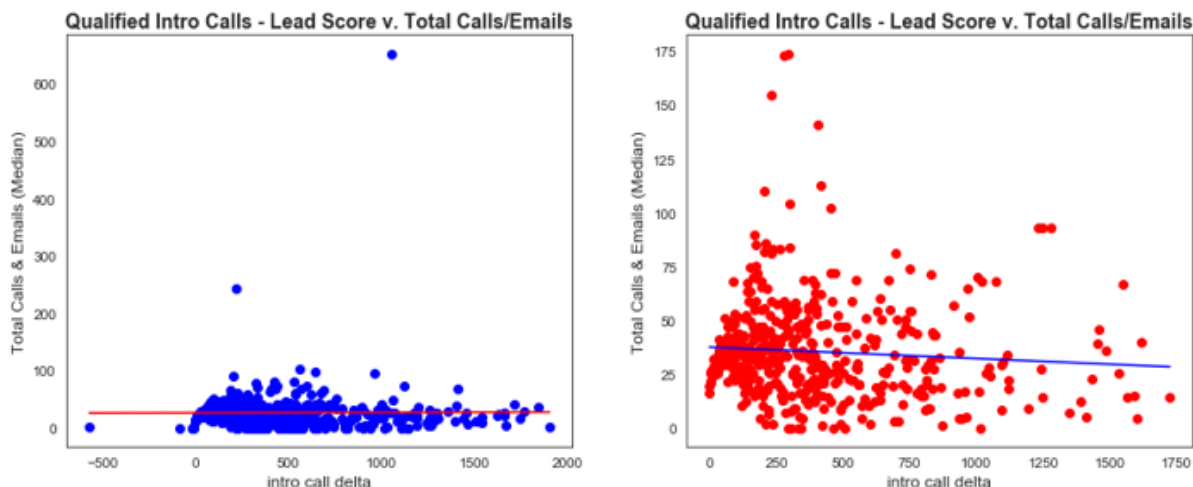
The following charts show outlier adjusted plots, where we can see a clearer relationship between qualified and disqualified intro calls. The trend is a bit more intuitive as we can imagine SDR's and their workflows.

For example, for qualified intro calls the relationship between total calls/emails and lead scores is less steep than for disqualified intro calls. In general the role of an SDR is to weed out the bad and keep hooked the good. For leads with higher lead scores, the positive slope could be showing that SDR's will engage more with prospects they view as positive bets. With disqualified vs qualified intro calls it's possible however that SDR's could also be putting additional effort into "pulling people across the line" or setting calls with individuals that might not be a great fit and therefore need extra coaxing.

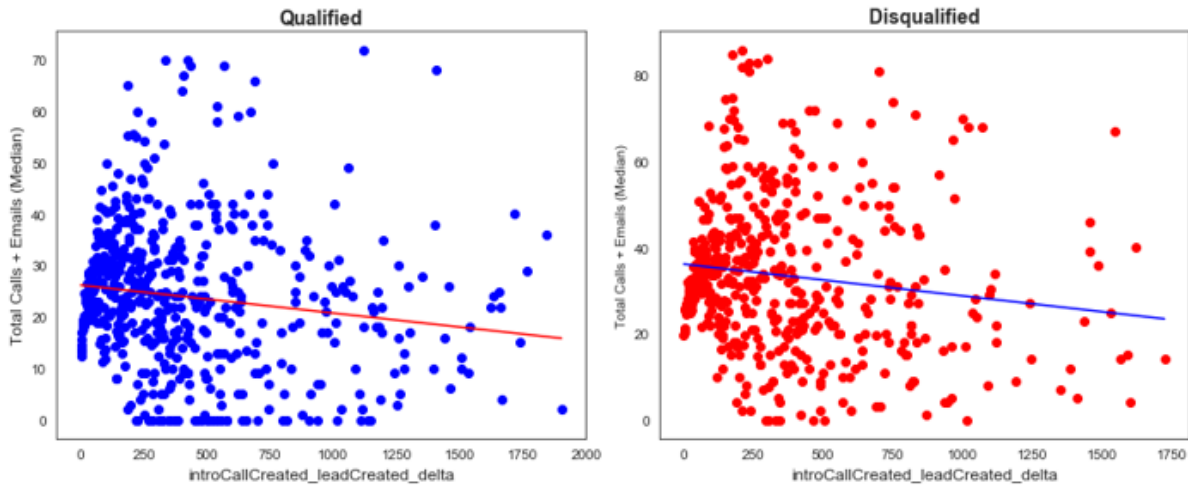


In trying to analyze the relationship between "introCallCreated_leadCreated_delta" (days elapsed between when the lead was created and when the intro call was created i.e. when a lead hit our system and someone talked to them) and total calls/emails there seems to be a negative relationship.

Top row of charts includes outliers, bottom row excludes outliers.



However given how similar the trend is between qualified and disqualified intro calls I'm not sure there is a meaningful relationship we can analyze.



b. Can you make a time-series plot?

Please see previous questions.

2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?

I'd like to further investigate the following variables as potential candidates for drivers of intro call status:

- Landing Page
- Lead/Marketing Channel
- Customer Type
- Creation Date

I'm ambivalent about the following variables given the mixed results I've seen in the previous charts:

- Lead Score
- Intro Call - Lead Creation Delta
- Region

I'm interested in seeing which of these variables end up being major contributors to the predictive models I'll be generating. Within the company we have a lot of untested beliefs about what drives a successful sales and being able to test out some of these assumptions will be incredibly helpful.

One belief we have internally that can't be tested directly (but could be indirectly) is that different sales teams have drastically different close rates. The difficulty in trying to test that question is the company has completely restructured the sales teams in the last 4-6 years and we don't have historical data on which individuals were on which teams handling specific deals. For the most part teams have been split across geographic territories so one way to determine if some teams (or regions) were more effective is analyzing the impact of geography on the qualification rates.

Another belief we have is that lead with higher lead scores should perform better and require less handholding.

3. **Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?**

Product details could be important in determining whether an intro call has been qualified. The company has marketed towards specific enterprise platforms and products (like SAP, Salesforce, etc) and we've had better luck in supporting some platforms as opposed to others. Product information is captured in a picklist (which then becomes a nested string in the data warehouse) so it'll take some additional cleaning to analyze the relationship between product of interest and qualification.