PREDICTING SALES SUCCESS FROM INTRO CALLS

Mikiko Bazeley

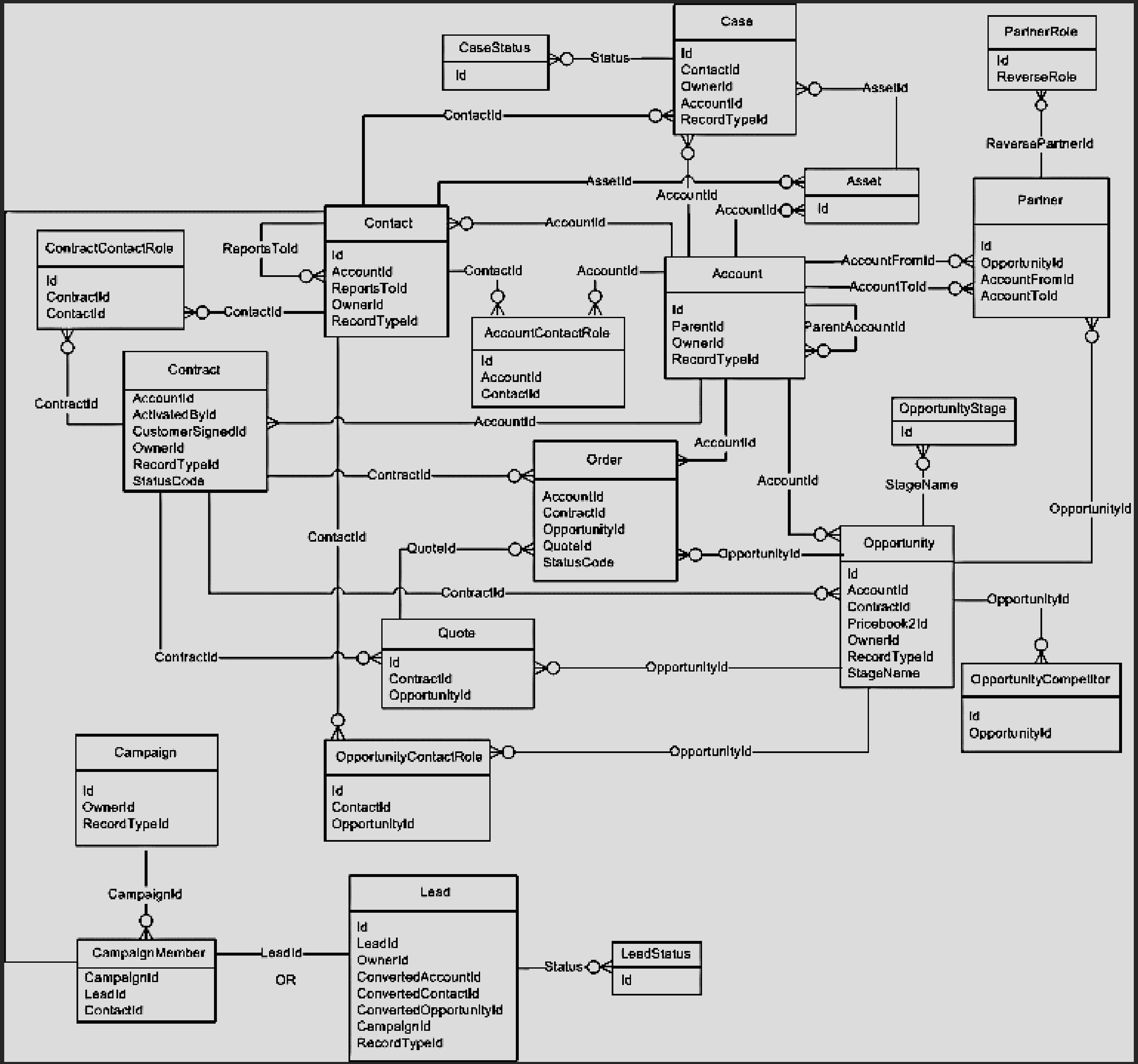
March 2019

Springboard Data Science Cohort

GH: https://github.com/MMBazel/springboard-program/tree/master/capstone1

CONTENTS

I. BACKGROUND
+
II. DATA SET DESCRIPTION



SALESFORCE STANDARD OBJECTS SCHEMA

I. Project Background

Context:

- This project is focused on understanding what are the drivers for demo call qualifications and whether we can create a robust predictive model.

- Two data sets are utilized (both internal and derived from the company's Redshift data warehouse) focused on leads and "intro calls" (demo calls). The data warehouse collects the data from Salesforce and processes it via ETL but still requires intense data cleansing and wrangling.

- Of the 107 variables in the initial lead dataset and 114 variables in the intro call data set, we'll be focusing on the core variables that relate to the demo calls. Demo calls (or intro calls, term are synonymous) occur after a lead has entered the system and before an opportunity can be created. As a result, predicting the outcome of demo calls (which are hosted by a live person and subjective) can be a tricky judgement call

- Being able to classify whether an intro calls will be qualified based on lead characteristics helps create consistency and transparency for both forecasting and training.

Goals:

- 1) Determine if correlations between key sales indicators (account level demographics, lead characteristics, etc) and qualification outcome.

- 2) Determine if there a statistically significant difference between qualified and disqualified intro calls in regards to key characteristics.

- 3) Create machine learning model that allows us to predict whether an intro call will be qualified and understand the different drivers of qualification across models.

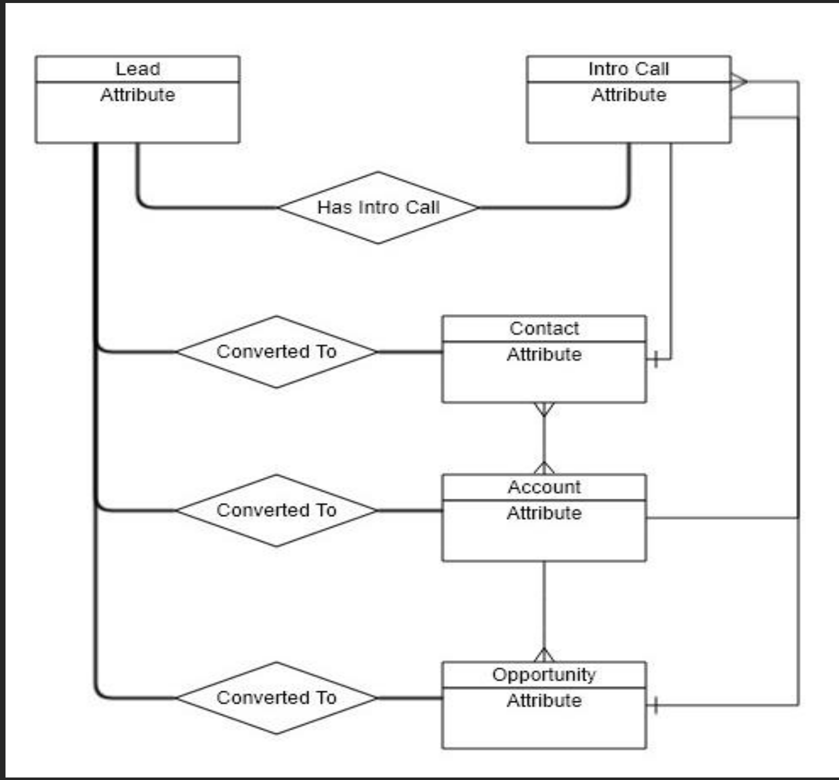
II. Detailed Data Set Description

Datasets

- Origin: AWS Redshift Database that collects data from all areas of the company (Salesforce, ADP, JIRA, Google Analytics, etc)

- Leads:
    - Volume: ~50K
    - Columns: ~107
    - Examples: id, name, title, converteddate

- Intro Calls:
    - Volume: ~23K
    - Columns: 114
    - Examples: lastactivitydate, name, systemmodstamp



III. DATA EXPLORATION + IV. ANALYSIS

Data Exploration:

- Volume Over Time
- Lead Marketing Channel
- Customer Types
- Lead Countries
- Landing Pages

Analysis:

- Lead Score vs. Intro Call Qualification
- Total Calls & Emails vs. Intro Call Qualification
- Lead Intro Call Delta vs. Intro Call Qualification

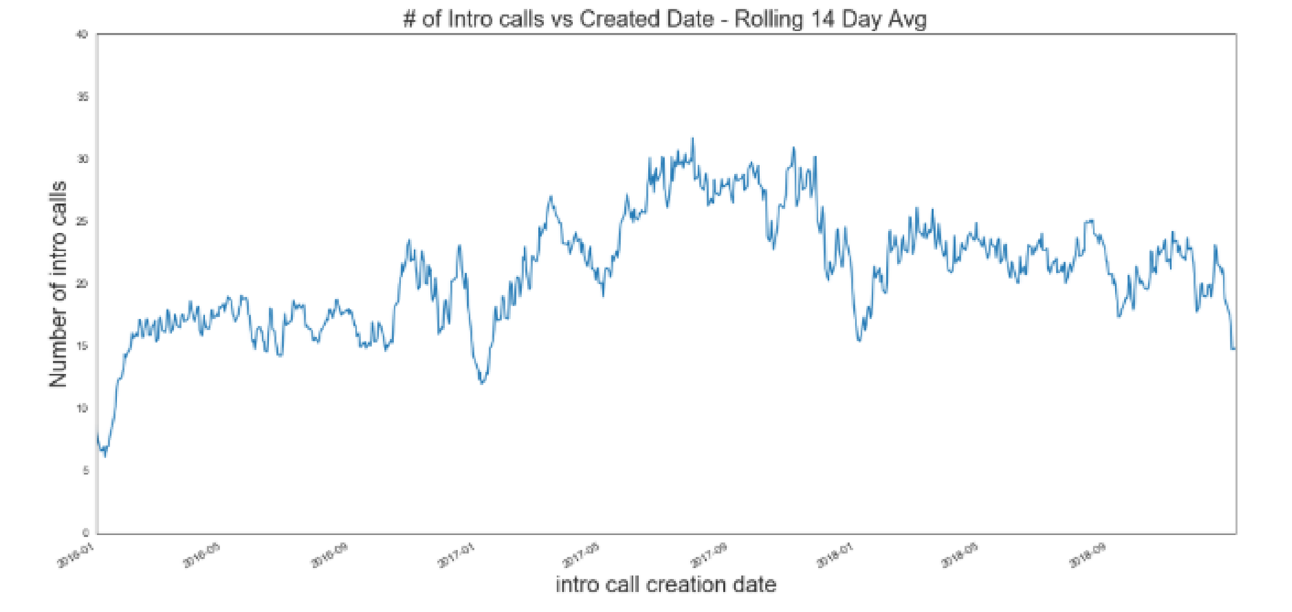
Data Exploration 1: Trends in Volume & Traffic Channels

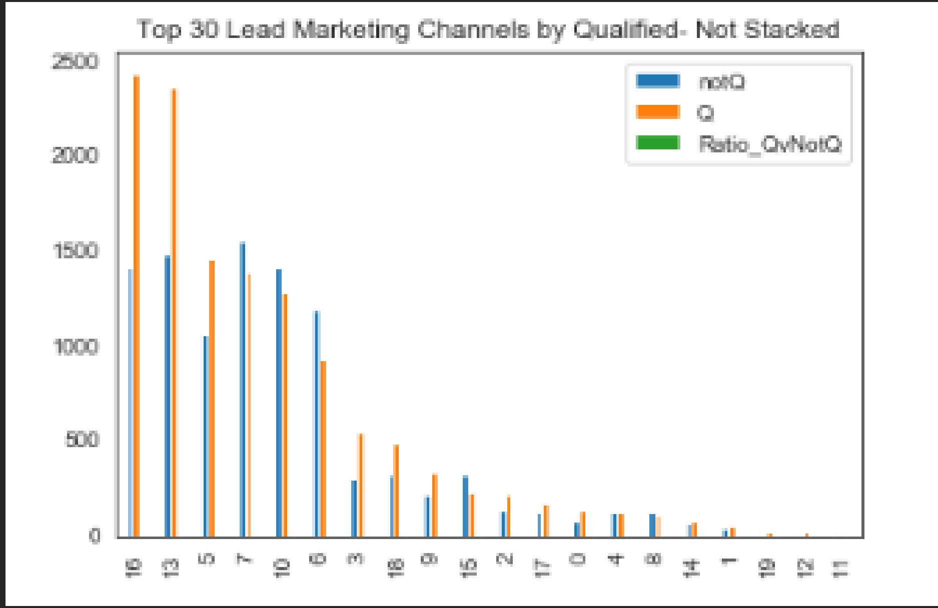
Intro Calls Volume Over Time

- Of 22.9K records, 12.8K (56%) were qualified vs 10K (44%) 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).
- One interesting trend to note is that there seemed to be a higher volume of closed intro calls between May 2017 and Jan 2018.

Traffic Channels (Qualified vs. Not Qualified)

- Note how some traffic channels have a much higher proportion of qualified intro calls.
- 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.

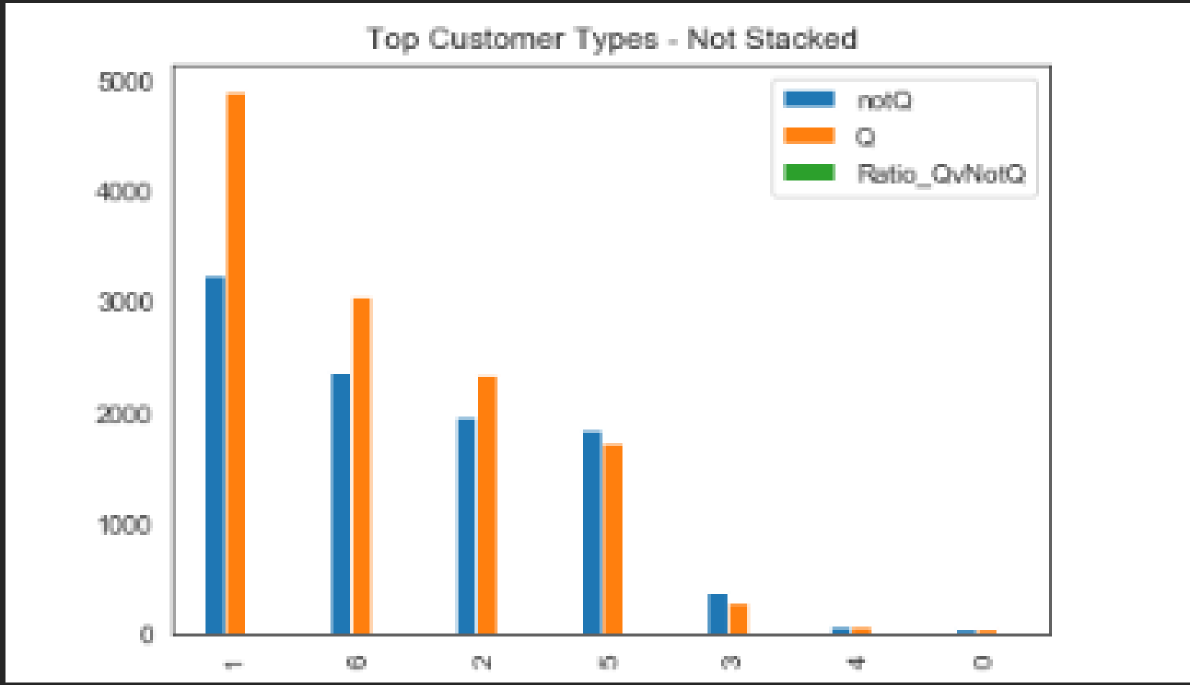




Data Exploration 2: Trends in Customer Types & Countries

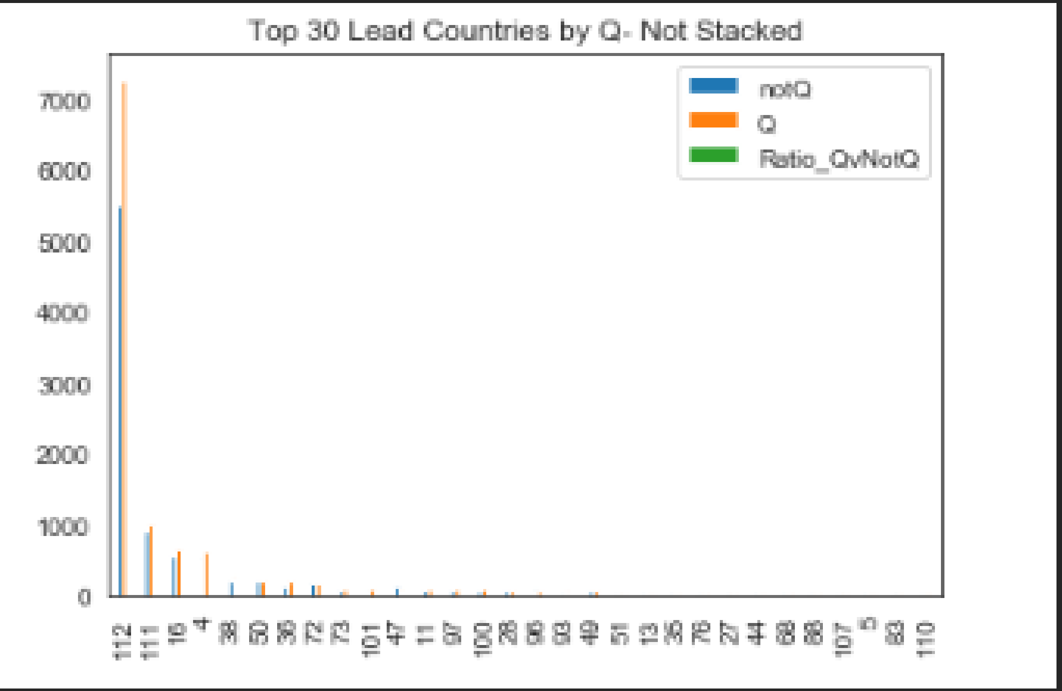
Customer Type (Qualified vs. Not Qualified)

- 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 (Qualified vs. Not Qualified)

- 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.

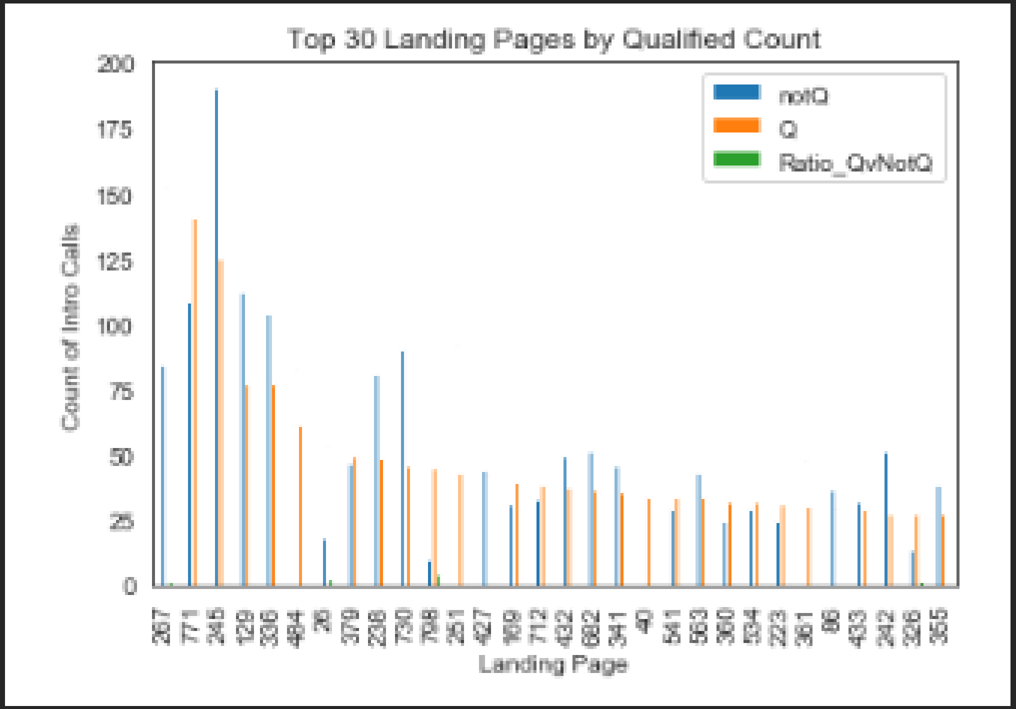


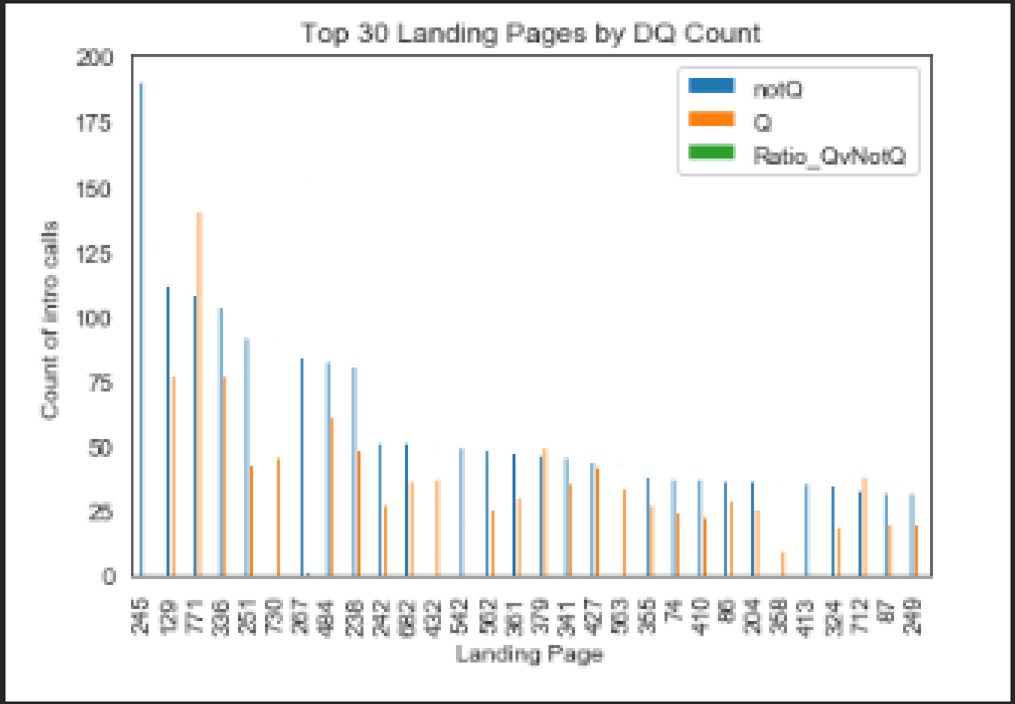
Data Exploration 3: Trends in Landing Pages (Qualified vs. Not Qualified)

Qualified

Not Qualified

- 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 qualified count aren't the same as the top 30 landing pages by ratio of qualified to disqualified intro calls





DATA ANALYSIS TAKEAWAYS

- Variables as potential candidates for drivers of intro call status:
  - Landing Page
  - Lead/Marketing Channel
  - Customer Type
  - Creation Date

- Potential Mixed Results:
  - Lead Score
  - Intro Call - Lead Creation Delta
  - Region

- Need to test for meaningful differences between qualified & disqualified intro calls:
  - Lead Score
  - Intro Call - Lead Creation Delta
  - Total Calls/Emails

Lead Score vs. Intro Call Qualification

Quality of lead can continue to impact downstream qualification & sales process beyond the Marketing to Sales Handoff , with lead scores differing between Qual & Not Qual

- We first want to understand the summary statistics of Qualified vs. Unqualified Intro Calls and whether the assertion that there is no difference (and lead scores should be 60+).

- From printing the summary statistics, we can already see that the assertion that the sales team doesn't interact with leads below 60 is false. Both samples of Qualified and Disqualified Intro Calls had a minimum below 60 (Qualified: 9, Disqualified: 6).

- However our Qualified sample is displaying an IQR of [82 (25%), 92 (75%)] and our Disqualified sample is displaying an IQR of [73 (25%), 91 (75%)], so it's possible the assertion that the majority of leads leading to demo calls should be around 70-90. We also observe a difference in means: Qualified (84), Unqualified (80).

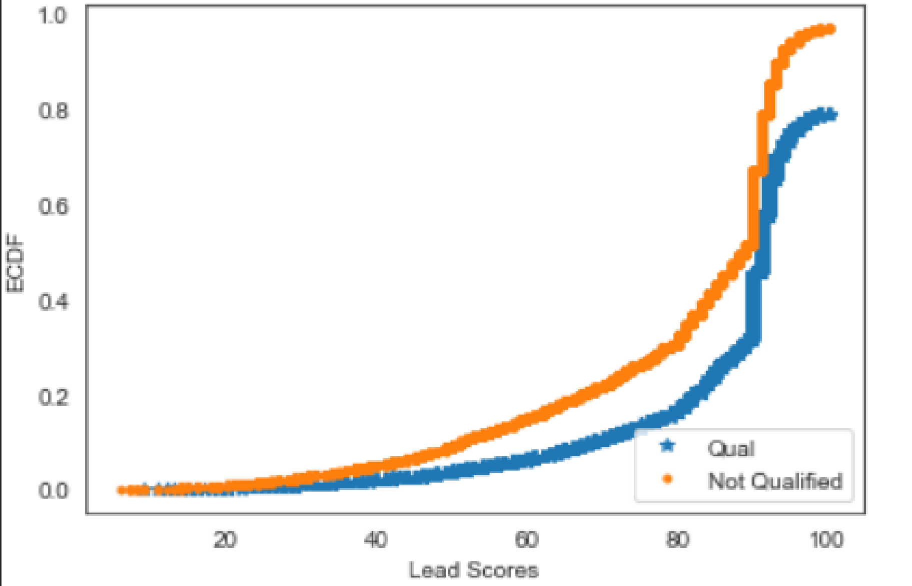
Summary of Qualified Intro Calls:

count	10198.000000
mean	84.204354
std	13.983466
min	9.000000
25%	82.000000
50%	90.000000
75%	92.000000
max	100.000000

Summary of Not Qualified Intro Calls:

count	9837.000000
mean	79.396869
std	17.703785
min	6.000000
25%	73.000000
50%	88.000000
75%	91.000000
max	100.000000

Fig 1A: ECDF of Lead Scores by Qualified vs. Unqualified Intro Calls

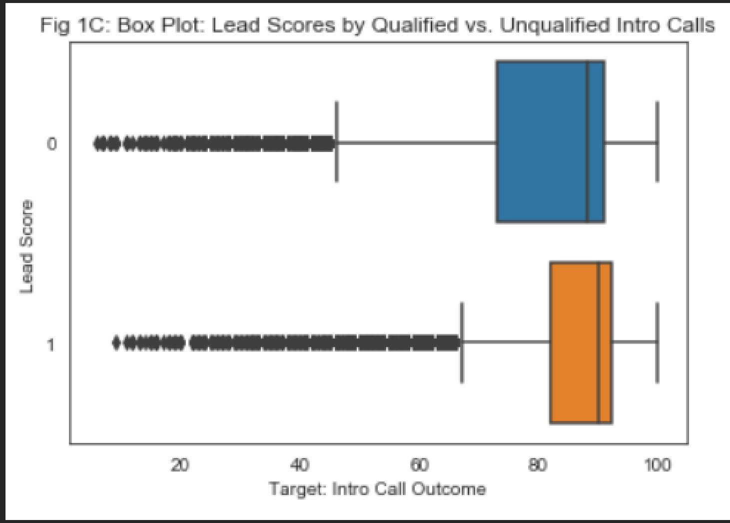


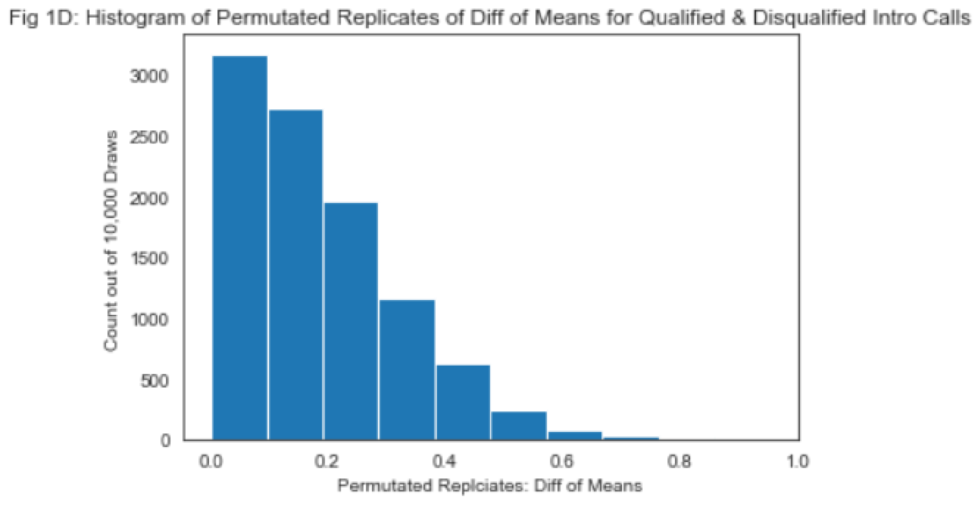
Lead Score vs. Intro Call Qualification

Quality of lead can continue to impact downstream qualification & sales process beyond the Marketing to Sales Handoff , with lead scores differing between Qual & Not Qual

Results of Statistical Tests:

- Permutation Test:
    - Empirical Diff of Mean: 4.8 (compare to FIG 1D)
    - Proportion of replicates with value as great or greater than empirical diff of means - p-value = 0.0000
    - Reject null hypothesis that two variables have identical distrb.
- Mann-Whitney:
    - MannwhitneyuResult(statistic=41177766.0, pvalue=6.510734883915887e-108)
- Welch's T-Test:
    - Ttest_indResult(statistic=array([21.28042586]), pvalue=array([2.58993265e-99]))





Total Calls & Emails vs. Intro Call Qualification

Level of engagement can be an important indicator of a qualified prospect & can show inefficiencies in engagements with not qualified prospects

Qualified:	totalCallsEmails
count	7029.000000
mean	28.165884
std	21.496974
min	0.000000
25%	12.000000
50%	24.000000
75%	40.000000
max	100.000000

Not Qualified:	totalCallsEmails
count	6905.000000
mean	37.026792
std	29.297481
min	0.000000
25%	14.000000
50%	30.000000
75%	52.000000
max	138.000000

- In an ideal sales world, most sales managers would like sales reps to engage in the minimum amount of correspondence needed to: (1) qualify a prospect and (2) ensure good prospects are pulled into the sales process.

- From printing the summary statistics, we can already see that Disqualified Intro Calls were associated with a higher mean of Total Calls & Emails compared to Qualified Intro Calls (36.9 vs. 28.0).

- We can also see a difference in the IQR of Disqualified vs Qualified Intro Calls, indicating that prospects of Disqualified Intro Calls could be taking up more sales rep time (Qualified: [12 (25%), 40 (75%)], Disqualified: [14 (25%), 52 (75%)]).]

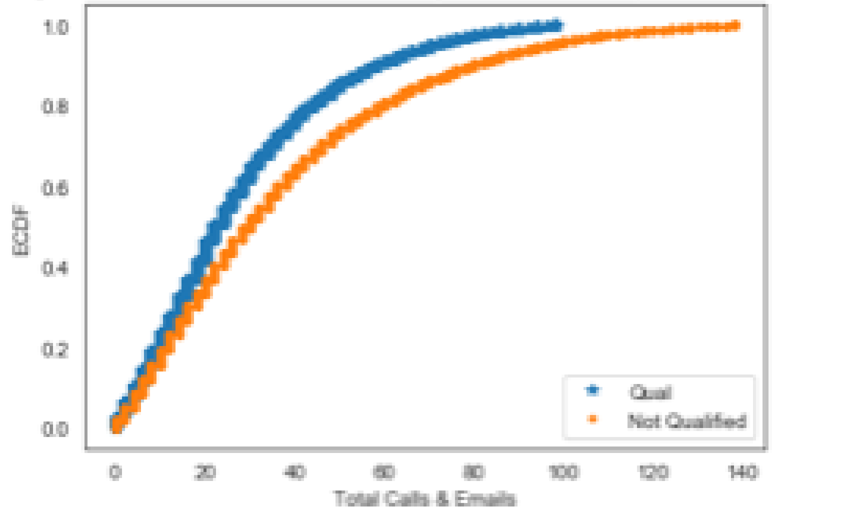


Fig 2A: ECDF of Total Calls & Emails by Qualified vs. Unqualified Intro Calls

Total Calls & Emails vs. Intro Call Qualification

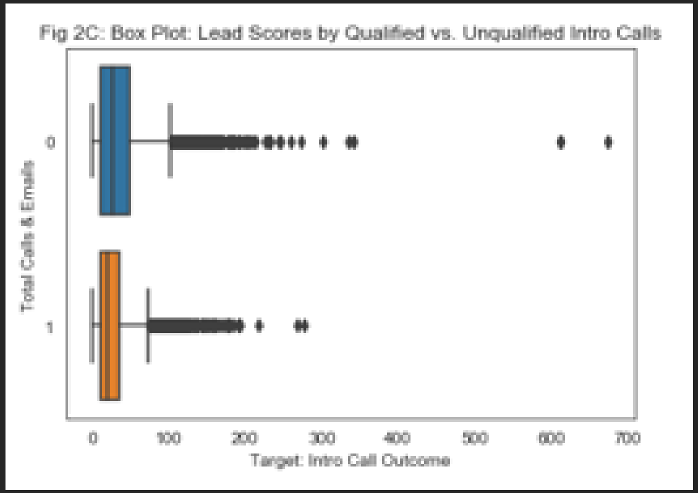
Level of engagement can be an important indicator of a qualified prospect & can show inefficiencies in engagements with not qualified prospects

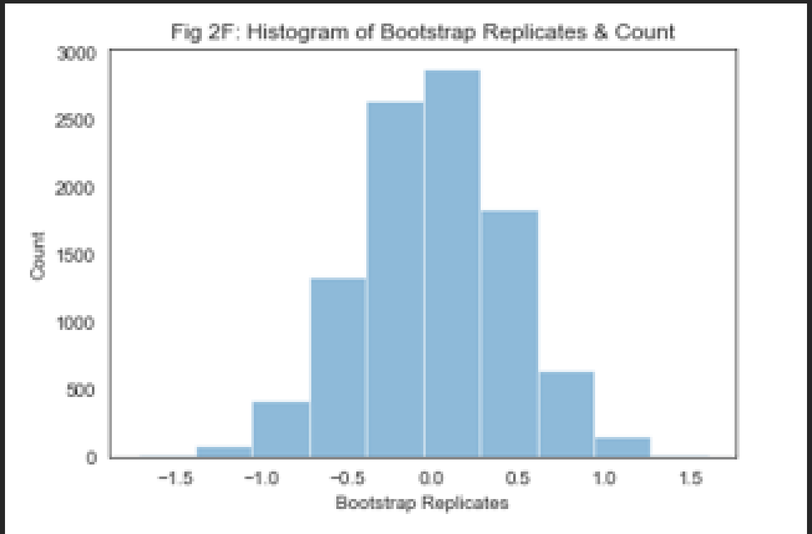
Results of Statistical Tests:

- Boostrap:
  - Empirical Diff of Mean: 8.860907985526804
  - Proportion of replicates with value as great or greater than empirical diff of means - p-value = 0.0000
  - Reject null hypothesis that two variables are same but come from difft groups.

- Mann-Whitney:
  - MannwhitneyuResult(statistic=20476861.0, pvalue=9.569972711739395e-58)

- Welch's T-Test:
  - Ttest_indResult(statistic=array([-20.32552895]), pvalue=array([2.1042651e-90]))





Lead Intro Call Delta vs. Intro Call Qualification

Freshness of Lead for Intro Call Doesn't Seem to Impact Outcome

Results of Statistical Tests:

- Permutation:
  - Empirical Diff of Mean: 1.1776803858285874
  - Proportion of replicates with value as great or greater than empirical diff of means - p-value = 0.2016
  - [Fig 3D] From the histogram of permuted replicates we can visually see that the empirical mean of 1.5 isn't an extreme value with about 12% of the permuted values having a value as great or greater than the empirical difference of means. The permutation test result doesn't seem to provide evidence to reject the null hypothesis that Qualified and Disqualified Intro Calls are significantly different with regards to the Time Delta

- Mann-Whitney:
  - MannwhitneyuResult(statistic=32711883.0, pvalue=1.2147849692726079e-05)

- Welch's T-Test:
  - Ttest_indResult(statistic=array([1.27657379]), pvalue=array([0.2017708]))

- We are seeing conflicted results from the Mann-Whitney test (which seems to reject the null hypothesis that the populations are similar) and Welch's T-Test (which doesn't result in a statistically significant p-value).

Fig 3C: Box Plot: Time between Lead and Intro Call Creation by Qualified vs. Unqualified Intro Calls

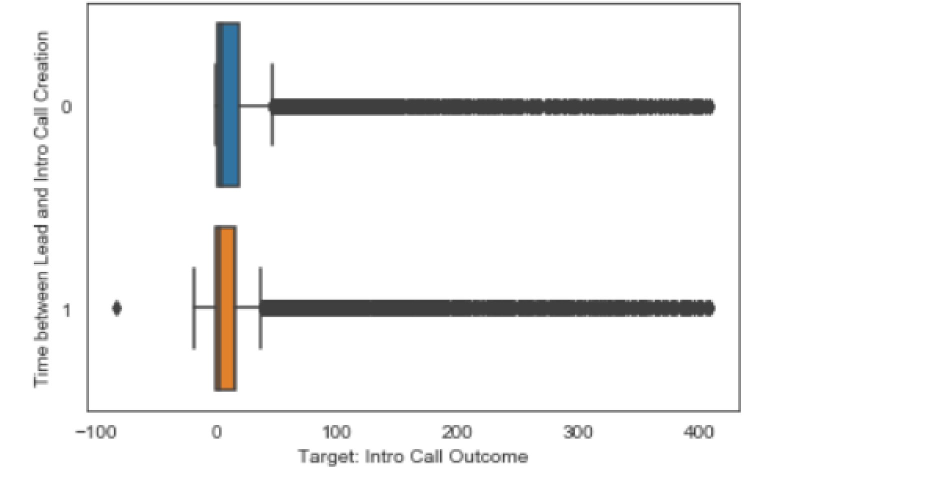
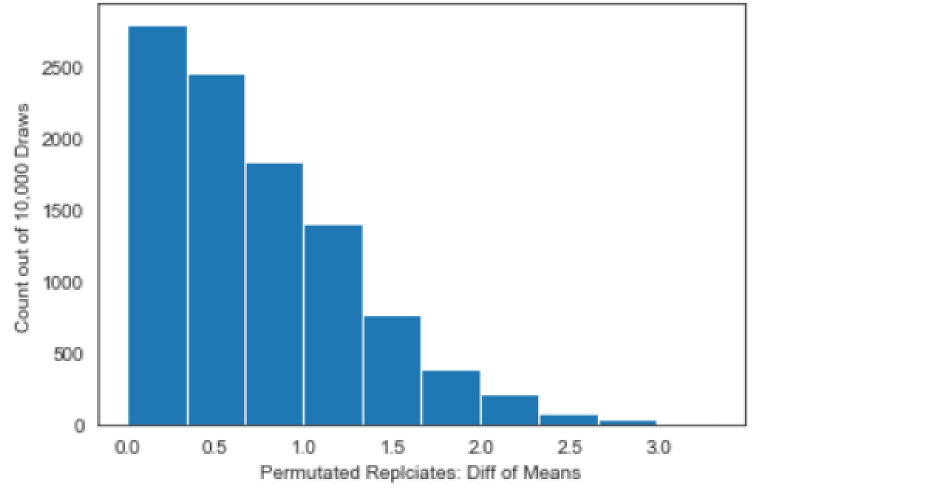


Fig 3D: Histogram of Permutated Replicates of Diff of Means for Qualified & Disqualified Intro Calls



V. PREDICTIVE MODELS

Prediction, hyperparameter tuning and model performance evaluation of:

- Logistic Regression
- Random Forest
- Gradient Boosted

Process:

1. Prepare master data set (1-Hot Encoding)
2. Scale data (StandardScaler)
3. Train-Test-Split data
4. Evaluate
5. Tune parameters using RandomizedSearchCV & GridSearchCV
6. Evaluate

PREDICTIVE MODELS OVERVIEW

My goal was to understand what features were important to predicting whether an intro call would be qualified. In order to classify whether intro calls would be classified, I initially built three models: a logistic regression model, a random forests model, and a gradient boosted model.

Leveraging different feature engineering techniques and hyperparameter tuning, I was able to attain 79.3 % accuracy in classifying Intro Call qualifications (table of results shown below) with the Gradient Boosted 1-Hot Encoded Model with Hyperparameter tuning.

The top 5 features across all three models in determining Intro Call Qualification Status included:

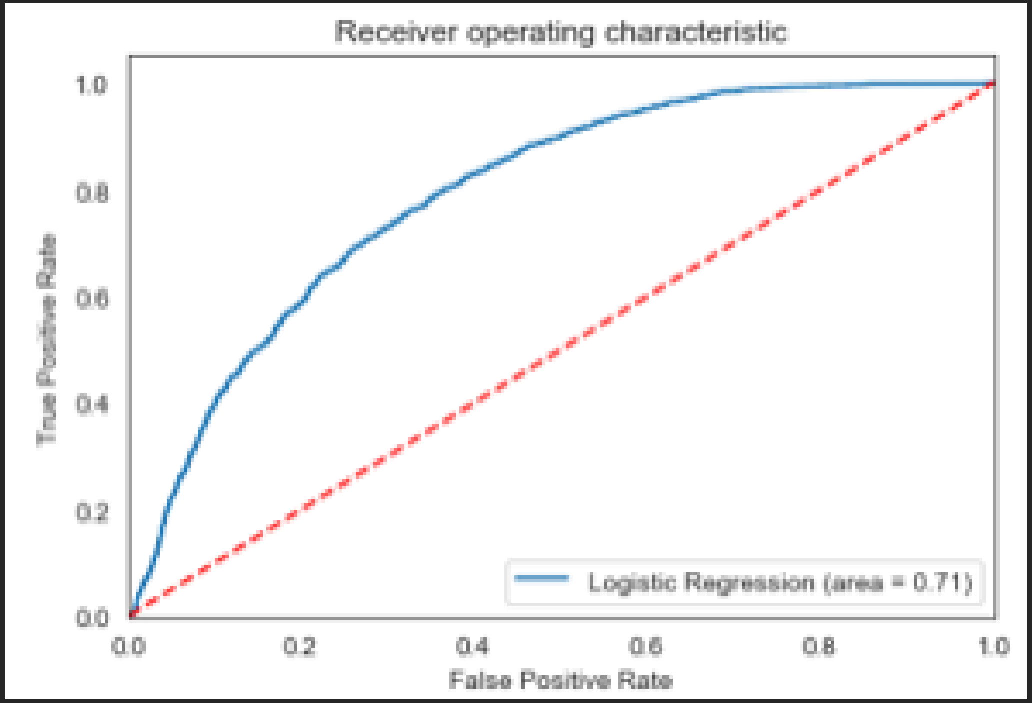
1. inferScore___Lead_AddedInfo
2. totalEMails___Lead_AddedInfo
3. totalCalls___Lead_AddedInfo
4. introCallCreated_leadCreated_delta
5. assignedToRole___IntroCall_OtherInfo_map

The features I assumed would be highly ranked but weren't included:

1. country___Lead_LeadCompanyInformation_map
2. trafficChannel___Lead_MarketingInformation_map_map
3. product2___IntroCall_MeetingDetails_WalkMe

Model 1: Logistic Regression

Model Performance & Features



Features:

Before tuning accuracy:

- Accuracy Score: 0.7307969707897584

With GridSearchCV:

- **Best parameter:** {'C': 10, 'max_iter': 100}
- **Best score:** 0.734551574897812
- **Test set accuracy:** 0.7311575910566174

Logistic model performed the worst of all three at classifying intro call qualifications. Additional parameter tuning provided no significant lift,

Model 2: Random Forest

Model Performance & Features

Before tuning:

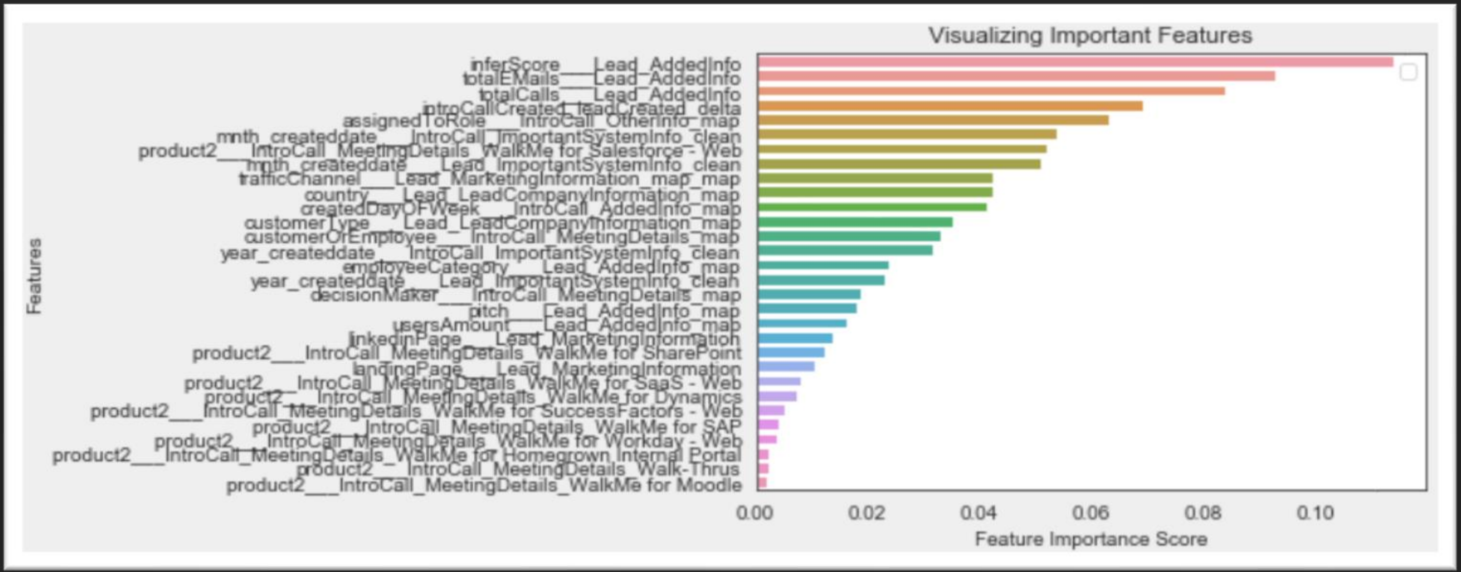
- Accuracy Score: 0.7370398196844478

- Confusion Matrix (Test Set):

[[2143 835]
[915 2762]]

- Confusion Matrix (Train Set):

[[6828 51]
[88 8560]]



Top Features	Importance
inferScore___Lead_AddedInfo	**0.109808**
totalEMails___Lead_AddedInfo	**0.101876**
totalCalls___Lead_AddedInfo	**0.079746**
introCallCreated_leadCreated_delta	**0.071043**
assignedToRole___IntroCall_OtherInfo_map	**0.060582**

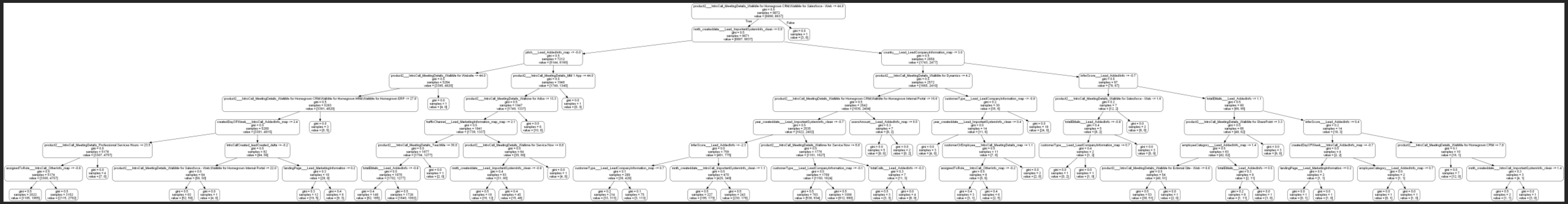
Random Forest might suffer from overfitting (see Confusion Matrix on Train Set) – Random Forest accuracy was #2 of all models

Model 2: Random Forest

Model Performance & Features

After tuning:

- **Best parameter:** {'bootstrap': False, 'max_depth': 50, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 13, 'n_estimators': 400}
- **Best score:** 0.791138017646668
- **Test set accuracy:** 0.7873779113448535



Model 3: Gradient Boosted

Model Performance & Features

Before tuning:

- Accuracy Score: 0.7767092411720511

- Confusion Matrix (Test Set):

[[1923 1050]

 [436 3246]]

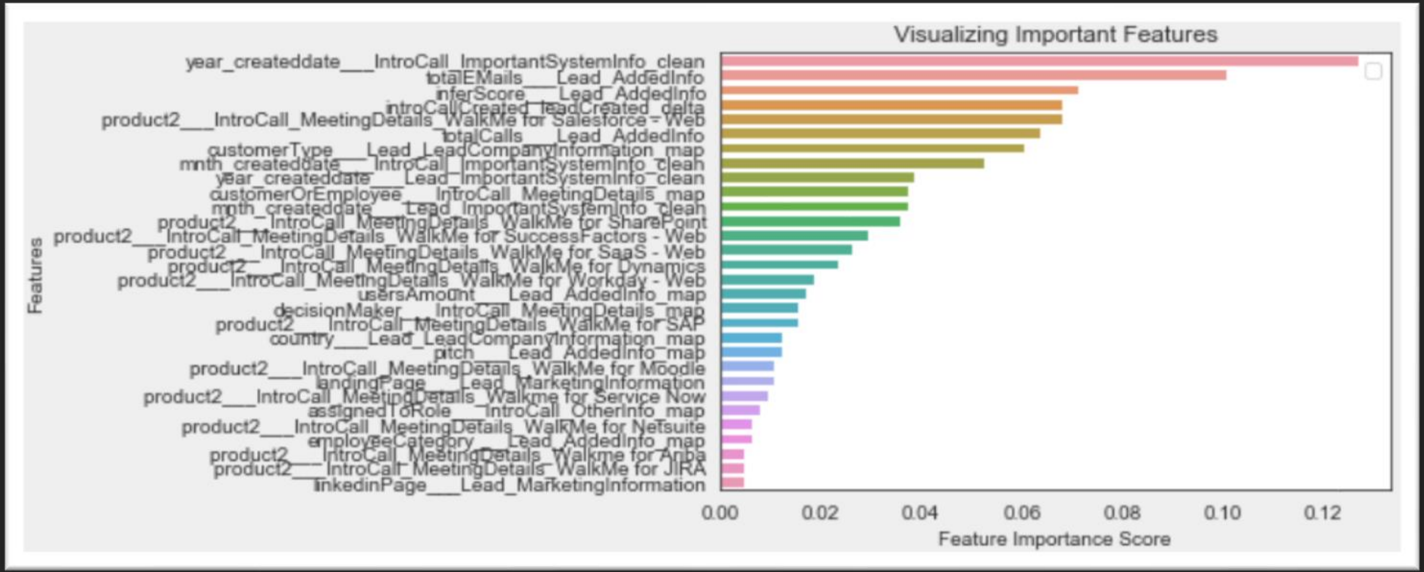
- Confusion Matrix (Train Set):

[[4499 2385]

 [894 7749]]

With GridSearchCV:

- Best parameters: {'colsample_bytree': 1.0, 'gamma': 2, 'max_depth': 6, 'min_child_weight': 10, 'subsample': 1.0}
- Best score: 0.7919108649449347
- Test set accuracy: 0.7912847483095417



Top Features	Importance
totalEMails___Lead_AddedInfo	**0.106646**
totalCalls___Lead_AddedInfo	**0.075734**
customerType___Lead_LeadCompanyInformation_map	**0.074189**
year_createddate___Lead_ImportantSystemInfo_clean	**0.071097**
introCallCreated_leadCreated_delt	**0.071097**

VI. RECOMMENDATIONS + FUTURE WORK

Project Based-Recommendations

Future Work

General Recommendations

VI. Project Based Recommendations + Future Work

Recommendations

- Given that disqualified intro calls were correlated with higher calls & emails, one possible suggestion could be to train the sales teams to front load discovery questions for earlier disqualification.

- Given also the difference in distributions of lead scores by qualified vs. disqualified, there could be down stream impact from marketing letting in poorer quality leads. Some of their assertions should also be evaluated as it seems their leads aren't as high quality as expected.

Future Work

- Using the model building process developed in this project, working with the sales org to ask similar questions:
    - Can we predict what deals will close?
    - Can we predict which existing customers are candidates for upsell/cross sell?
    - How can we utilize the insights from both the intro call qualification process to improve internal processes around data collection?
    - Once these labels are predicted, how can we then improve our forecasts?

VI. General Recommendations

For Similar Projects

- Never assume you have as much data as you think you do! Of 200+ fields "available" for analysis a majority ended up being duplicates or poorly populated. ~15 were used for model building across models.
- Don't assume that just because you have a "data warehouse", its contents are relevant and useful.
- Gantt charts are still around for a reason. Plan ahead as much as possible and under promise!

For the Organization

- Create tighter communication with upstream application teams
- Constant re-evaluation and assessment of data points being collected should be done
- Significant data cleaning and extraction time could have been saved by the presence of relevant documentation, especially around data owners.

A1. APPENDIX: ADDITIONAL RESOURCES

Project Page:

- List of resources (& links) used for the project: https://bit.ly/2T0kmxC

Project Repo:

- Jupyter notebook:
- Supporting Documentation:
- Presentation deck:

Springboard Program:

- Information about program: https://www.springboard.com/workshops/data-science-career-track/

A2. APPENDIX: ABOUT ME

Applied analytics and data science evangelist, Mikiko Bazeley is a seasoned analyst with 5+ years of working in high-impact roles for start-ups and enterprise tech companies.

A UCSD Economics & Anthropology graduate, Mikiko aims to strategically leverage data science to drive new insights for sales, marketing, finance & customer success organizations by with her experience in social research & modeling. Mikiko also earned certifications in GIS & Supply Chain Management.

Prior to joining WalkMe (where she leads the global sales analytics effort) Mikiko worked as a Data Scientist at Autodesk (focused on understanding product adoption & user health), as well as assisting with scaling analytics initiatives at Sunrun (the largest residential solar company in the US).

Please feel free to reach out:

- LinkedIn: https://www.linkedin.com/in/mikikobazeley/
- Email: mmbazel (at) gmail (dot) com

THANKS!