# PREDICTING SALES SUCCESS FROM INTRO CALLS

Mikiko Bazeley

March 2019

Springboard Data Science Cohort

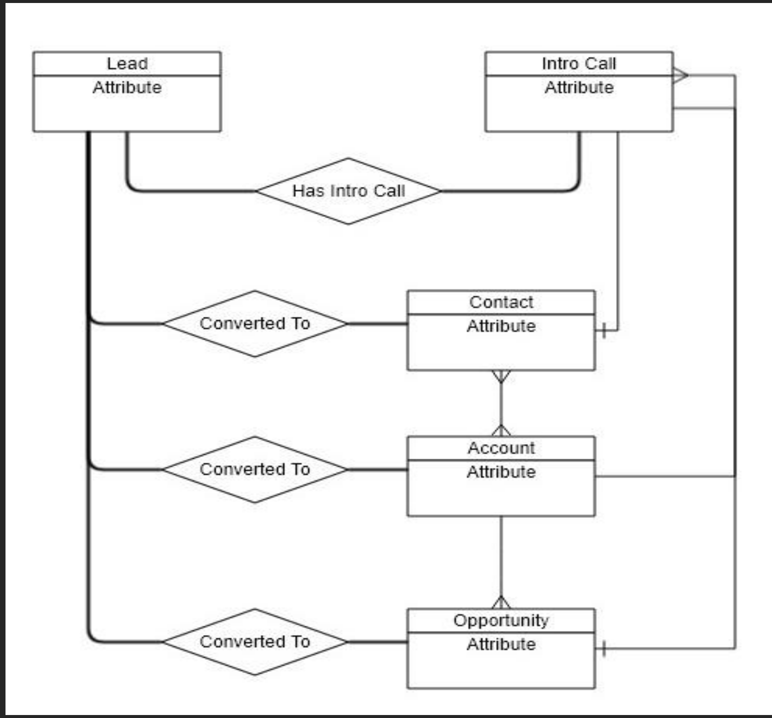
GH: https://github.com/MMBazel/springboard-program/tree/master/capstone1

# CONTENTS

# I. Project Overview

- **Goal:**
  - Predict qualification of sales demo calls using lead and demo call data.
- **Potential Interested Parties:**
  - Current company sales leadership
  - Current company marketing leadership
  - Other revenue operations professionals wanting to understand how to conduct similar analysis & projects
- Data Source:
  - Internal AWS Redshift data warehouse with tables mirroring Salesforce objects & fields

- **Outcome:**
  - Create classification model that predicts outcome of demo calls.



# II. Data Overview

amazon REDSHIFT

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

- ❖ Additional Details:
  - ❖Each row in the Intro Call table represents a single instance of a call but multiple rows in the Lead table could represent a single person, however, only a single Lead record will be attached to an Intro Call record.
  - ❖For this project, used data from Jan 1, 2016 to Jan 1, 2019
  - ❖Of 22.9K Intro Call records, 12.8K (56%) were qualified vs 10K (44%) disqualified indicating fairly well balanced classes

| **Leads Table/Object**<br><br>❖ Volume: ~50K<br>❖ Columns: ~107 | **Intro Call Table/Custom Object**<br><br>❖ Volume: ~23K<br>❖ Columns: 114 |
|---|---|
| •Id<br>•First Name<br>•Last Name<br>•Company<br>•Job Title<br>•Email<br>•Landing Page<br>•etc | •Id<br>•Lead Id<br>•Lead Name<br>•Company Name<br>•Assigned to<br>•Created Date<br>•Scheduled Call Date<br>•etc |

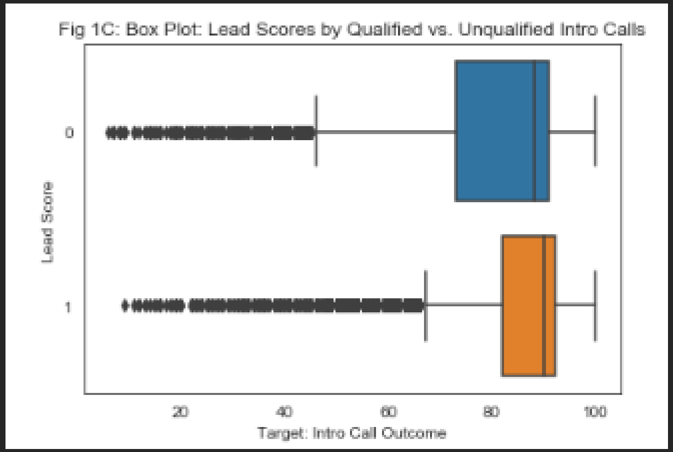
# III. Data Acquisition & Processing

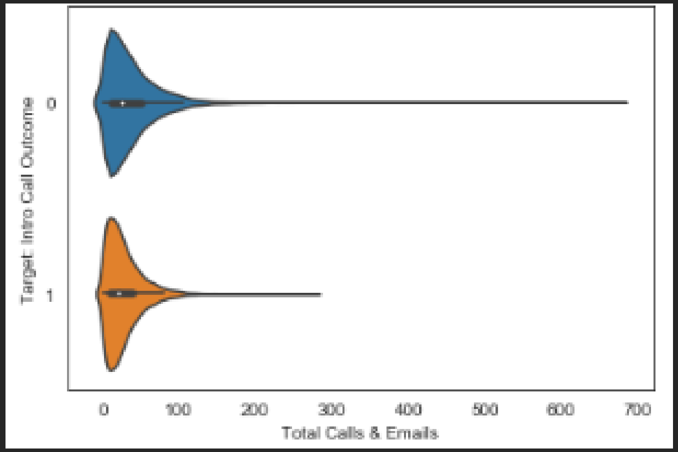
- **Important considerations needed to be handled given the data quality & source:**
  - Multiple intro calls for the same individual could have occurred throughout the time where the first call was scored as disqualified but the second call was then qualified.
  - A significant number of fields were duplicated across tables - due to the quasi relational structure of Salesforce, objects could be mimicked as tables but fields needed to be duplicated for the Salesforce user
  - Not all fields were used throughout time & the values didn't always reflect the field's stated purpose
  - Because of the tribal knowledge nature of the company, some fields could reflect data leakage without the business always knowing.

- **Process Used:**
  1. Query AWS Redshift db using sqlalchemy & psycopg2 (a postgreSQL driver) – Lead & Intro Calls tables loaded into pandas data frames
  2. Merge data frames on common identifier (left join Leads to Intro Calls)
  3. Clean & process fields, including:
     1. Standardizing datetime fields
     2. Remapping categorical fields
     3. Creating additional fields, especially datetime delta fields representing duration of time between stages of the sales cycle
  4. Perform exploratory analysis

# IV. Exploratory Analysis

- After data cleaning and processing I then analyzed a sample of potential key drivers of call qualifications. They included:
  - inferScore___Lead_AddedInfo ***
  - totalEMails___Lead_AddedInfo
  - totalCalls___Lead_AddedInfo ***
  - introCallCreated_leadCreated_delta ***
  - assignedToRole___IntroCall_OtherInfo_map
  - country___Lead_LeadCompanyInformation_map
  - trafficChannel___Lead_MarketingInformation_map_map
  - product2___IntroCall_MeetingDetails_WalkMe
- Features that seemed particularly promising ***



# IV. Feature Engineering & Selection

- **Feature Selection:**
  - Of the available 108 columns (Leads) & 115 columns (Intro Calls), 37 fields were used.
  - Fields of high interest included:
    - inferScore___Lead_AddedInfo
    - totalEMails___Lead_AddedInfo
    - totalCalls___Lead_AddedInfo
    - introCallCreated_leadCreated_delta
    - assignedToRole___IntroCall_OtherInfo_map
    - country___Lead_LeadCompanyInformation_map
    - trafficChannel___Lead_MarketingInformation_map_map
    - product2___IntroCall_MeetingDetails_WalkMe
- Feature Engineering performed:
  - Remapping of picklist values to cleaned dictionaries
  - Standardization of datetime fields

- Model 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

# V. Model Selection + Performance

- ❖ Decided to use some of the classic machine learning classifiers (Logistic Regression, Random Forest classifier, and a Gradient Boosted Trees)

- ❖ Best performing model: Random Forest classifier and Gradient Boosted tree models - resulted in an accuracy of 80% with additional hyperparameter tuning.

- ❖ The top 5 features across all three models in determining Intro Call Qualification Status included:
  - ❖ inferScore___Lead_AddedInfo
  - ❖ totalEMails___Lead_AddedInfo
  - ❖ totalCalls___Lead_AddedInfo
  - ❖ introCallCreated_leadCreated_delta
  - ❖ assignedToRole___IntroCall_OtherInfo_map

- ❖ The features I assumed would be highly ranked but weren't included:
  - ❖ country___Lead_LeadCompanyInformation_map
  - ❖ trafficChannel___Lead_MarketingInformation_map_map
  - ❖ product2___IntroCall_MeetingDetails_WalkMe

| Model | Performance | Performance with Param. Tuning | Optimal Params |
|---|---|---|---|
| Logistic Regression | 73% | 73% | {'C': 1, 'max_iter': 100} |
| Random Forest | 74% | 80% | {'bootstrap': False, 'max_depth': 60, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 400} |
| Gradient Boosted Tree | 78% | 80% | {'colsample_bytree': 1.0, 'gamma': 1, 'max_depth': 6, 'min_child_weight': 9, 'subsample': 1.0} |

# VI. Take-Aways

- Biggest takeaways from the project given the feature ranking & exploratory analysis:
  - Quality of lead can continue to impact downstream qualification & sales process beyond the Marketing to Sales Handoff, with lead scores differing between Qualified & Not Qualified Intro Calls (Lead Score).
  - Level of engagement can be an important indicator of a qualified prospect & can show inefficiencies in engagements with not qualified prospects (Total Calls + Emails).
  - Freshness of lead doesn't seem to impact outcome (Lead to Intro Call Creation Delta).

- Given the results of the models and tests, my project based recommendations are:
  - 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 downstream 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.

# 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 use her experience in social research & modeling to strategically leverage data science in order to drive new insights for sales, marketing, finance & customer success organizations. Mikiko is also GIS & Supply Chain Management certified.

Prior to joining WalkMe (where she focuses on scaling data science & global sales analytics) Mikiko worked as a Data Scientist at Autodesk (focused on understanding product adoption & user health), as well as assisting with scaling strategic finance 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!