

# TEAM HDR: FINAL PRESENTATION



**Hasibul Jishan**

Passionately Exploring Business Analytics and the Analytics world.



**Daniel De las Heras**

MS Applied Mathematics and Data Science



**Ryan Pittman**

Gas Scheduling Contractor at  
National Grid, Statistics/Applied  
Mathematics and Data Science  
Graduate Student, Keyholder at  
Costello's Ace Hardware

BY HASIBUL, DANIEL, AND RYAN

# PROBLEM STATEMENT



**Netrality Data Centers** is looking to expand its business by targeting new companies to acquire as customers



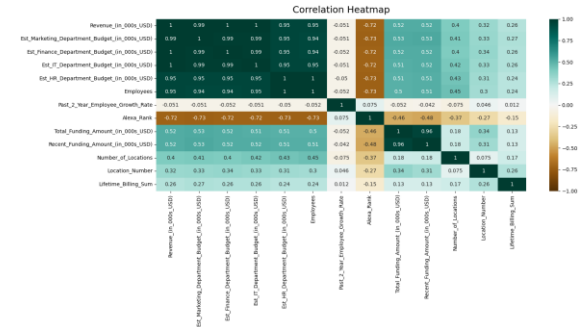
**Provided Three Datasets:**  
Current Customers, Current Billing, Prospect List  
(Not including Data Dictionary)



**Our goal:** To identify a consolidated list of companies that would be most advantageous for Netrality to gear their marketing campaigns toward and acquire as clients

# SUMMARY OF APPROACH - EDA

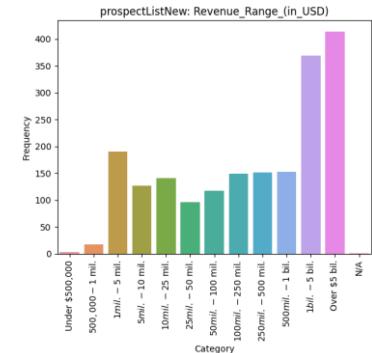
- CONDUCTED EXPLORATORY DATA ANALYSIS (EDA) ON THE THREE DATASETS
- PRINTED THE DESCRIPTIVE STATISTICS FOR THE QUANTITATIVE COLUMNS, CREATED VARIOUS GRAPHS FOR THE CATEGORICAL COLUMNS, AND CONSTRUCTED A CORRELATION HEAT MAP TO TEST FOR MULTICOLLINEARITY
- **KEY TAKEAWAY:** THE DATA WAS VERY SKEWED AND NONLINEAR



```

Founded_Year      8.081917
Revenue_(in_000s_USD)  86.217632
Est_Marketing_Department_Budget_(in_000s_USD)  130.992717
Est_Finance_Department_Budget_(in_000s_USD)    73.310792
Est_IT_Department_Budget_(in_000s_USD)         72.644307
Est_HR_Department_Budget_(in_000s_USD)         225.733661
Employees        201.276473
Past_2_Year_Employee_Growth_Rate  39.860218
Alexa_Rank        0.070918
Total_Funding_Amount_(in_000s_USD)  89.942488
Recent_Funding_Amount_(in_000s_USD)  147.241166
Number_of_Locations  50.052211
dtype: float64

```





# SUMMARY OF APPROACH - PREPROCESSING

- MERGED THE BILLING AND CURRENT CUSTOMER DATASETS
- LOG TRANSFORMED THE DATA TO ACCOUNT FOR SKEWNESS
- FILLED N/As WITH APPROPRIATE VALUES AND REMOVED OUTLIERS
- WE ALTERED SOME OF THE FEATURES IN OUR DATASET TO ENHANCE THEIR INFLUENCE IN OUR MACHINE LEARNING MODELS
  - EXAMPLES:
  - WE CHANGED THE “FOUNDED YEAR” COLUMN TO REFLECT THE AGE OF THE COMPANY IN YEARS BY SUBTRACTING EACH COMPANIES' FOUNDED YEAR FROM 2023
  - WE INCLUDED AN IT BUDGET RATIO COLUMN WHICH IS A PROPORTION OF HOW MUCH EACH COMPANY SPENDS ON ITS IT BUDGET OUT OF ITS ENTIRE REVENUE

	year	it_budget
CompanyID		
1475361	3.401197	0.148806
16400573	3.258097	0.154246
246280	3.555348	0.257051
43076614	3.218876	0.280992
126828694	3.713572	0.242453
...	...	...
372175531	3.583519	0.234457
347198798	2.772589	0.229900
32080939	3.135494	0.217065
39371338	3.091042	0.140490
441325878	2.397895	0.170075

# SUMMARY OF APPROACH – RESPONSE VARIABLES

- WE CREATED TWO BINARY COLUMNS TO ACT AS RESPONSE VARIABLES: “LABEL” AND “LABEL2”
  - SUMMED THE LIFETIME BILLING COEFFICIENTS FOR EACH COMPANY AND MADE A NEW COLUMN IN THE DATASET WITH THE TOTALS
  - THE “LABEL” COLUMN WHICH ASSIGNED A 1 TO COMPANIES WITH ABOVE AVERAGE TOTAL LIFETIME BILLING SUMS AND A 0 TO ALL OTHER COMPANIES THAT MADE AVERAGE OR BELOW AVERAGE BILLING PAYMENTS
  - THE “LABEL2” COLUMN TOOK COMPANY SIZE INTO CONSIDERATION BY DIVIDING THE LIFETIME BILLING COEFFICIENT SUMS BY THE NUMBER OF NETRALITY LOCATIONS EACH CURRENT COMPANY OCCUPIES
  - COMPANIES WITH AN AVERAGE LIFETIME BILLING SUM HIGHER THAN THE MEDIAN FOR ANY GIVEN LOCATION WERE ASSIGNED A 1 IN THE “LABEL2” COLUMN WHILE ALL OTHER COMPANIES RECEIVED A 0



# SUMMARY OF APPROACH – MODEL EXECUTION



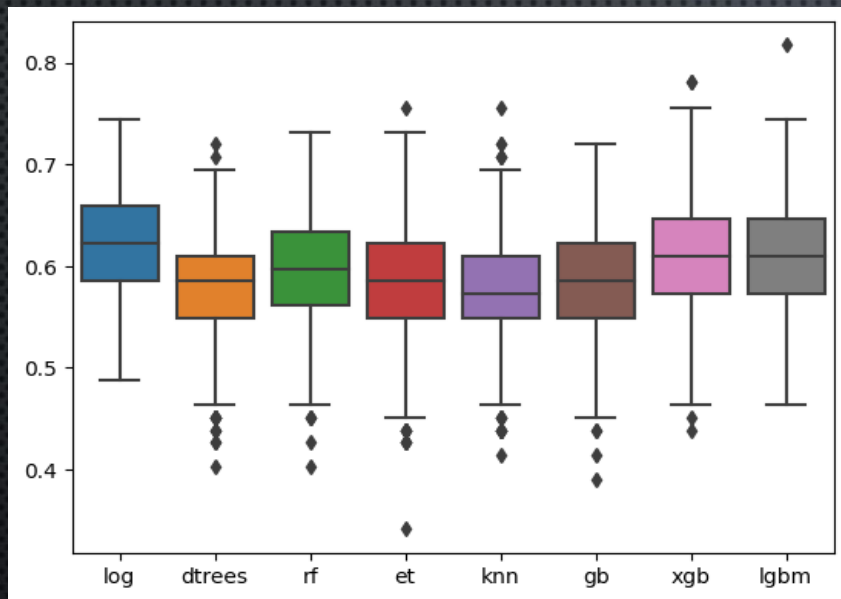
# FEATURE ENGINEERING

- LOGISTIC REGRESSION RESULTS
  - SHOWS MOST SIGNIFICANT AND INFLUENTIAL VARIABLES
- COEFFICIENTS: THE EXPECTED CHANGE IN LOG ODDS OF HAVING THE OUTCOME PER UNIT CHANGE

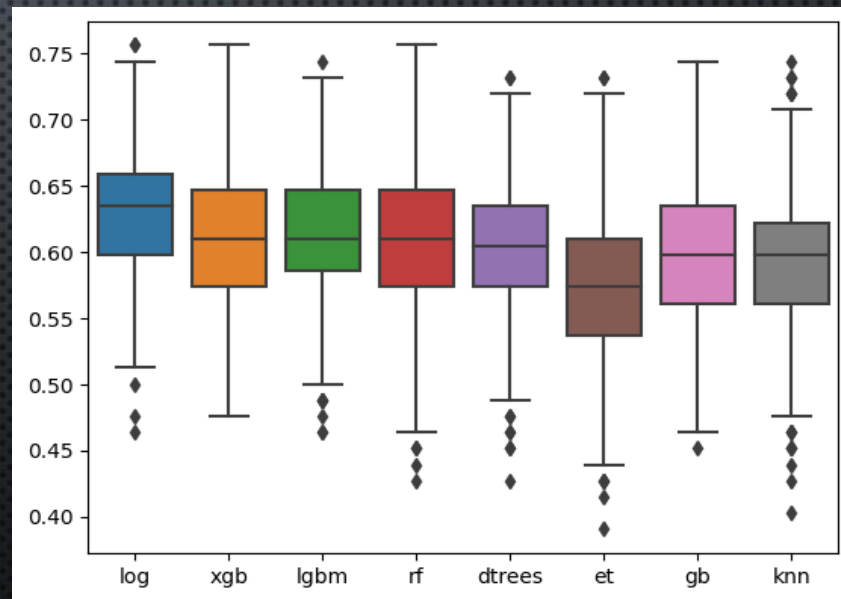
	Columns	Coefficient
4	ihc	-0.017244
5	alexa_rank	-0.022604
6	ownership	0.025065
2	employee_growth1	-0.028791
0	year	-0.071331
7	business_model	0.079181
11	recent_funding_date	-0.106475
3	employee_growth2	0.178993
14	it_budget	-0.179017
12	number_locations	0.180331
13	company_acquired	0.297038
8	certified_active_company	0.461595
10	recent_funding	-0.529807
1	revenue	0.590798
9	total_funding	0.699048

# MODEL ACCURACY PLOTS

“label”



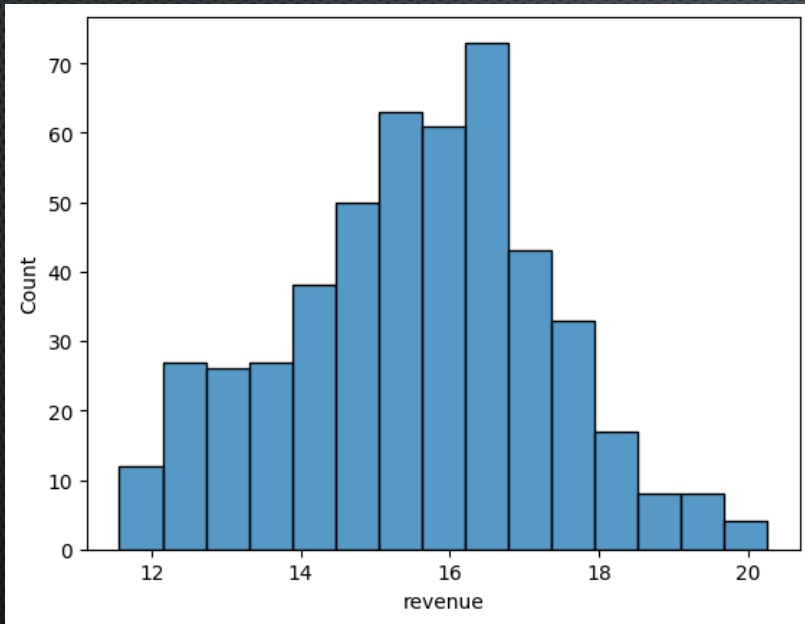
“label2”



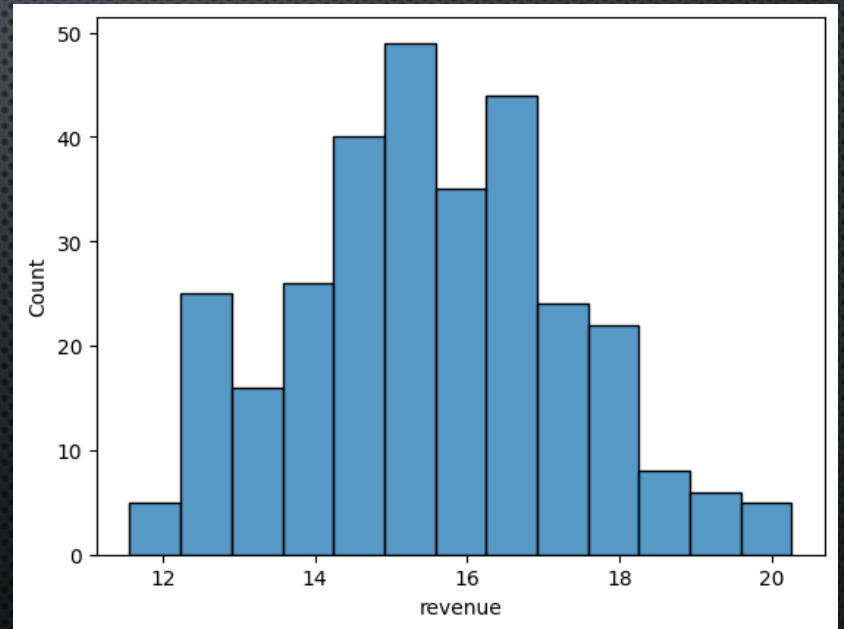


# REVENUE DISTRIBUTION OF RECOMMENDED LISTS

“label”

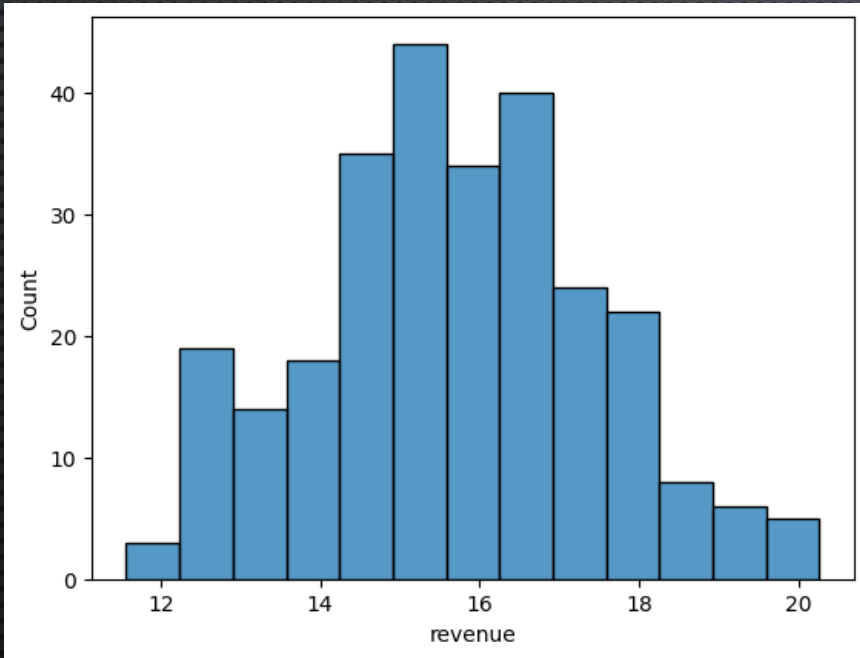


“label2”



# REVENUE DISTRIBUTION: FINAL LIST VS ACTUAL

Final List



Actual

