

Identifying High-Potential Accounts for Sales Targeting

A Data-Driven Approach

Vitoria Lima

TAL (Target Account List): 30 High-Potential accounts yielding jointly potentially ~5.4 mio USD yearly*



*Based on an Assumption of 30\$ USD monthly subscription per vehicle. [Source](#)

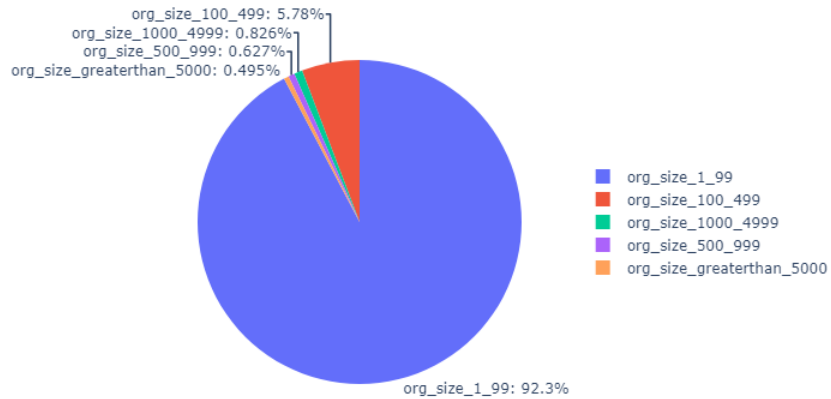
Agenda:

1. Data Overview
2. Feature Enhancing and Selection
3. Model: Additive Logistic Regression
4. Prediction Results
5. Insights and Business Impact
6. Limitations
7. Q&A

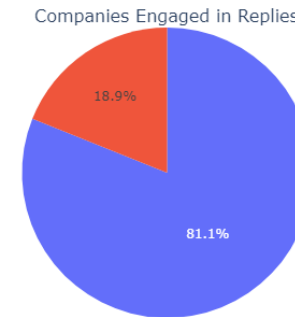
1: Data Overview

>3k companies across the 6 given datasets

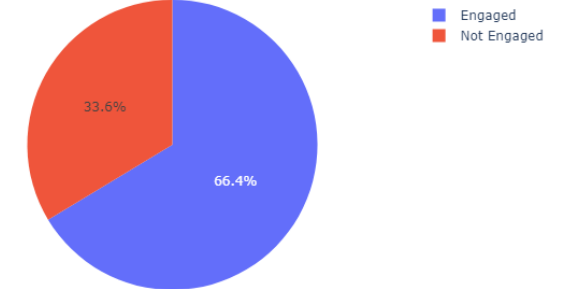
Distribution of Accounts by Organization Size



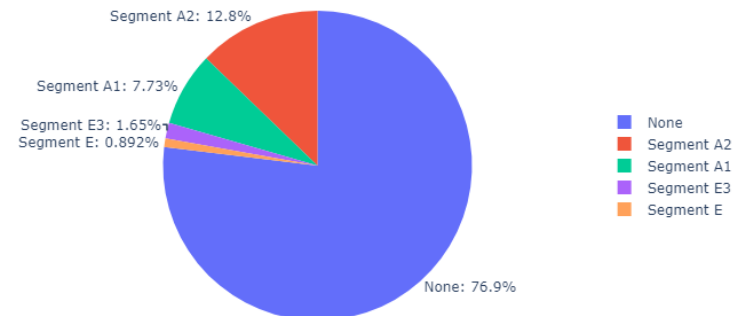
Distribution of Companies Engaged in Replies and Clicks



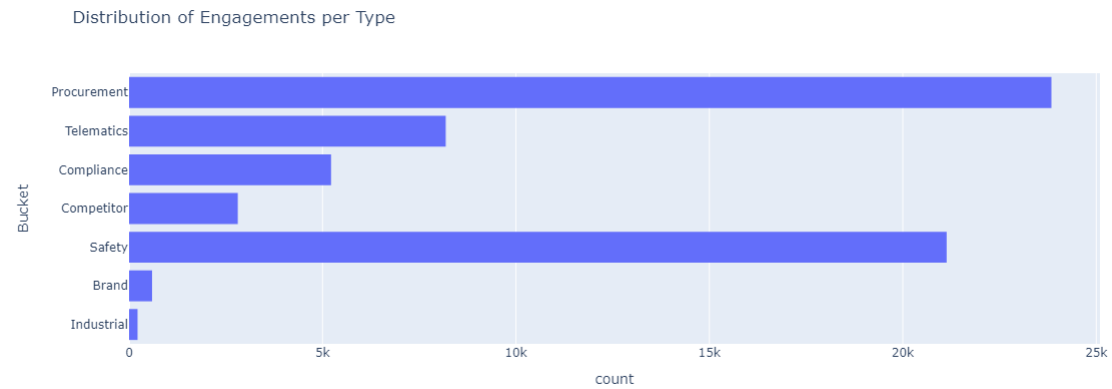
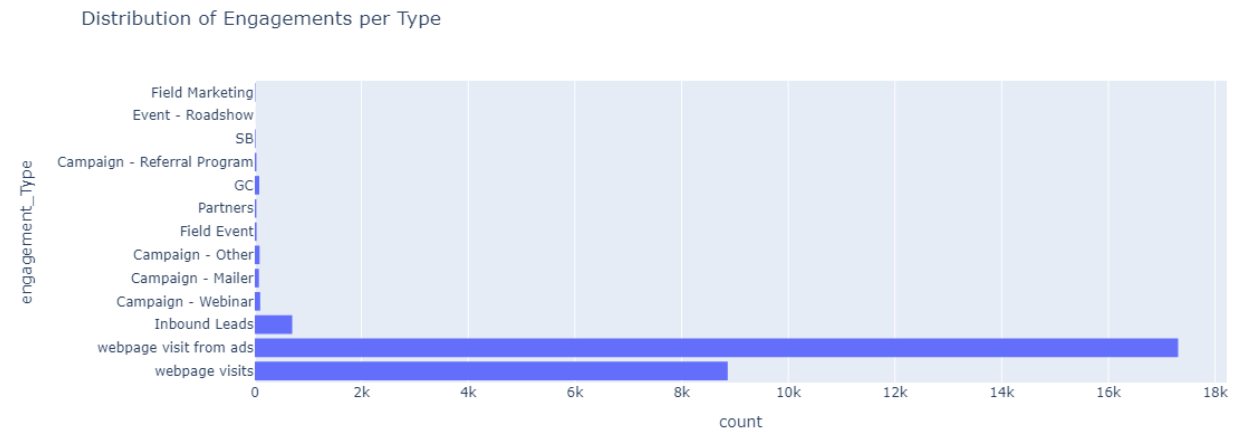
Companies Engaged in Clicks



Distribution of Companies by Segments

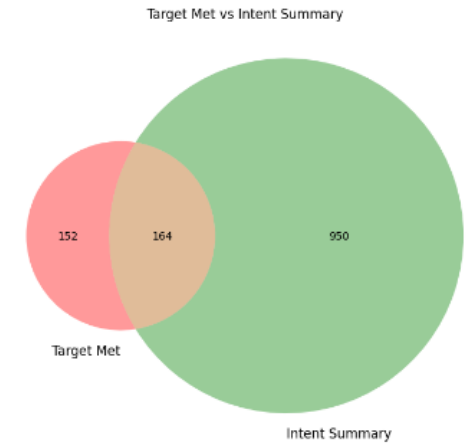
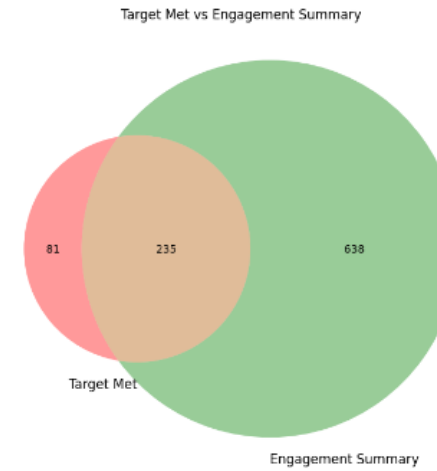
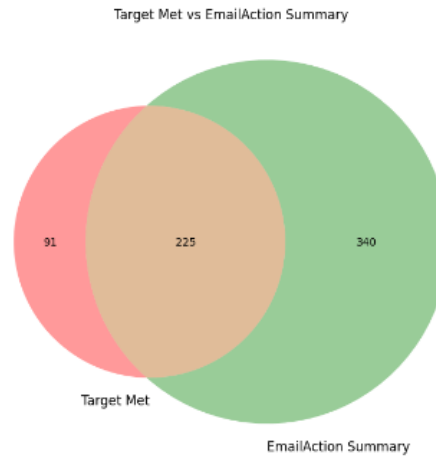
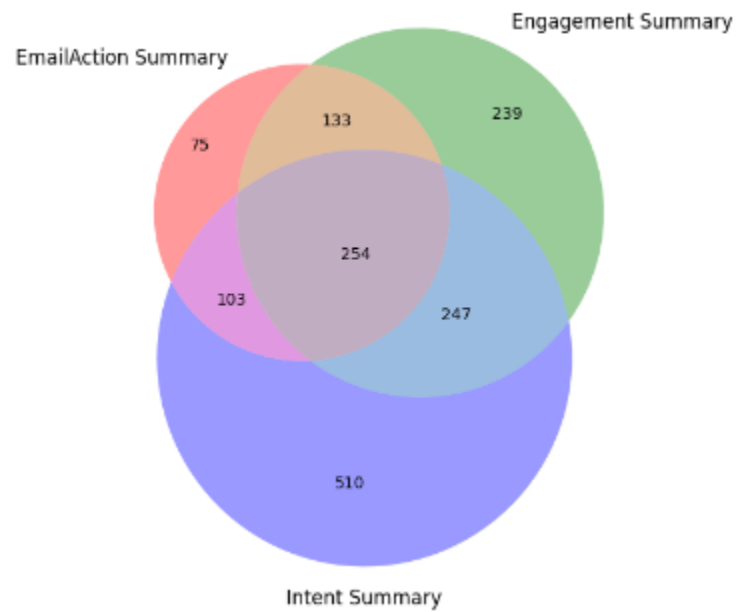


1.1: Data Overview



1.2: Data Sparsity

Venn Diagram of target_met Companies in Different Summaries



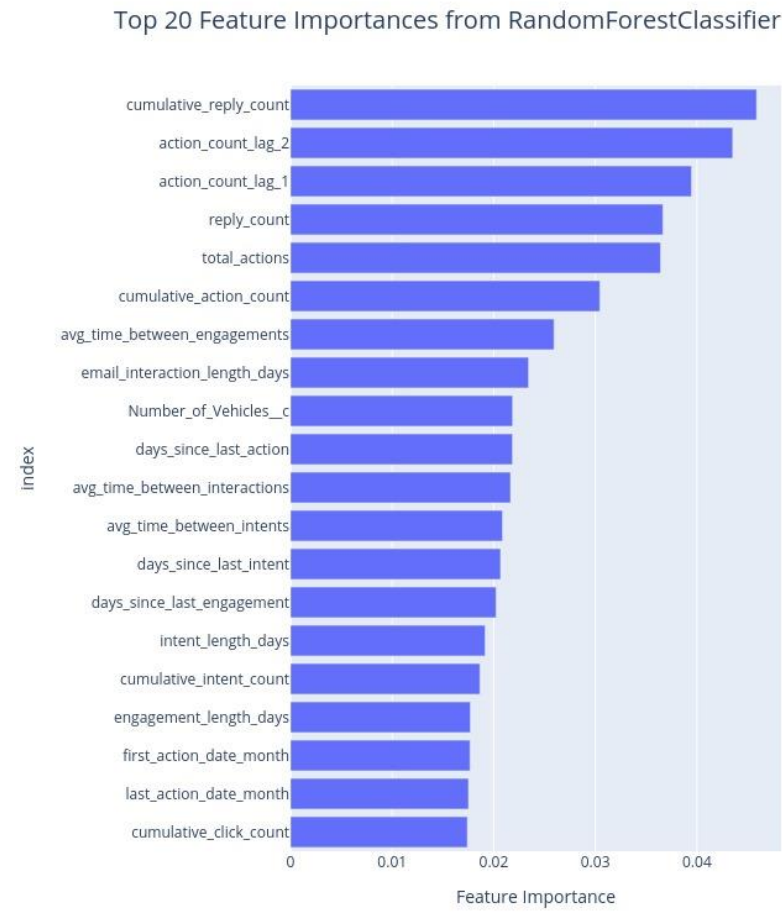
1.3: Training Data and Feature Engineering

- This has resulted in:
 - Train: 254 observations (out of which 122 sales)
 - Test : 132 observations
- Enhancement: Feature Engineering
 - New features: lag and rolling features, days between engagements, average amount of days between engagements, days since last engagement, average weekly engagement rate, etc.

...resulting in ~100 total features.

2: Feature Selection

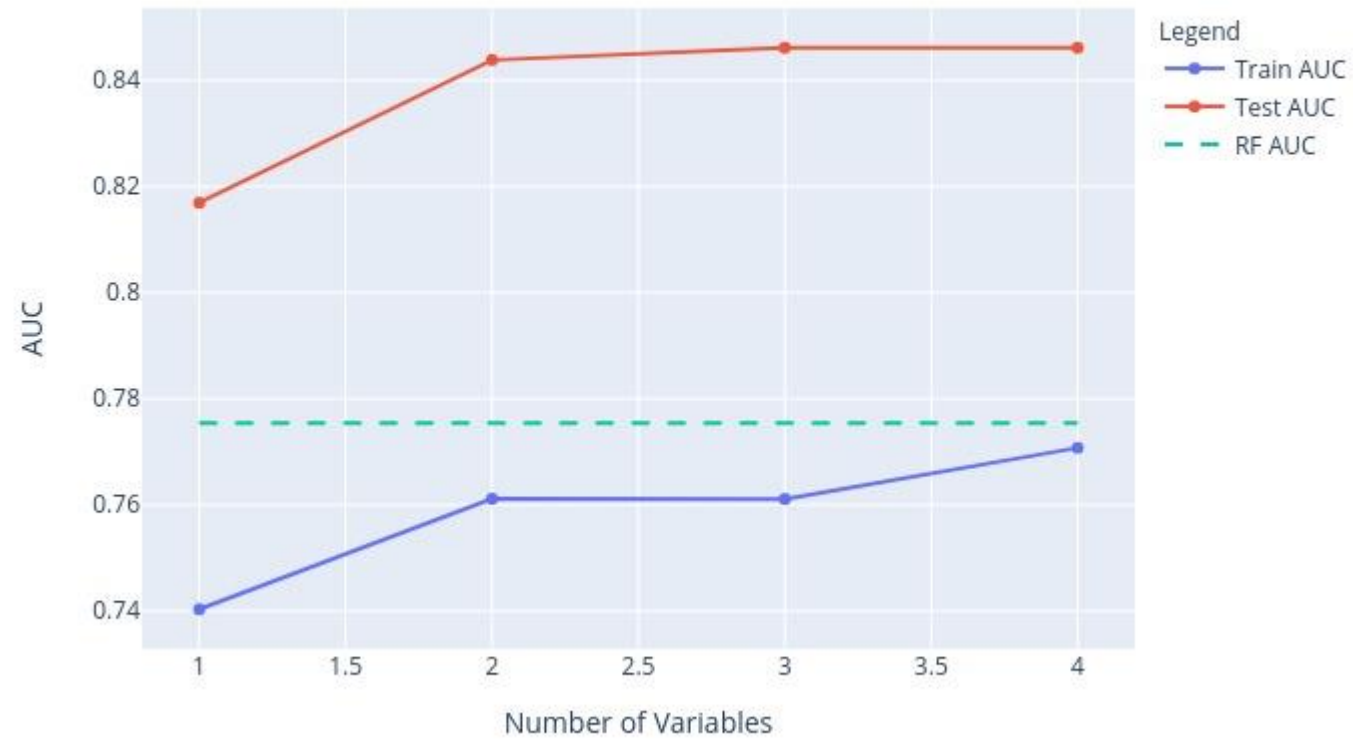
Random Forest



3: Modelling

Additive Logistic Regression

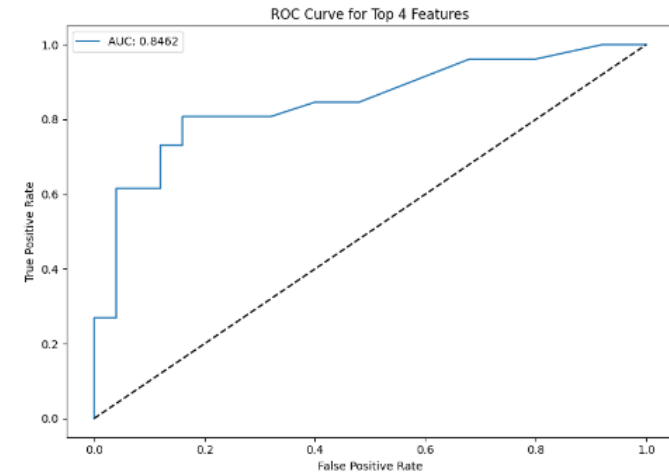
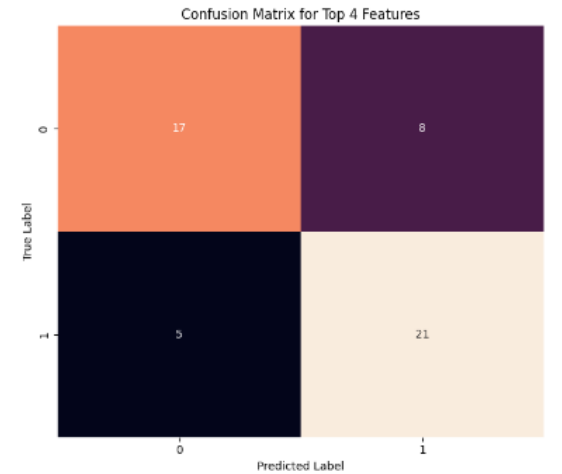
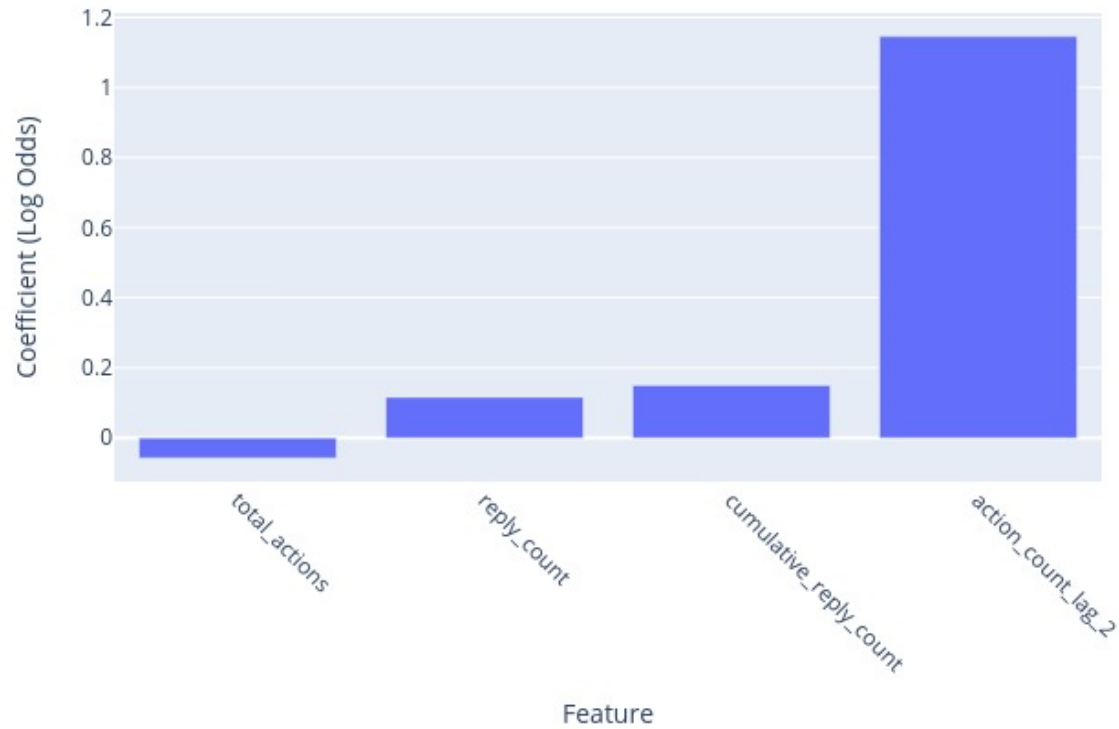
AUC Scores for Top 4 Features



3.1: Modelling

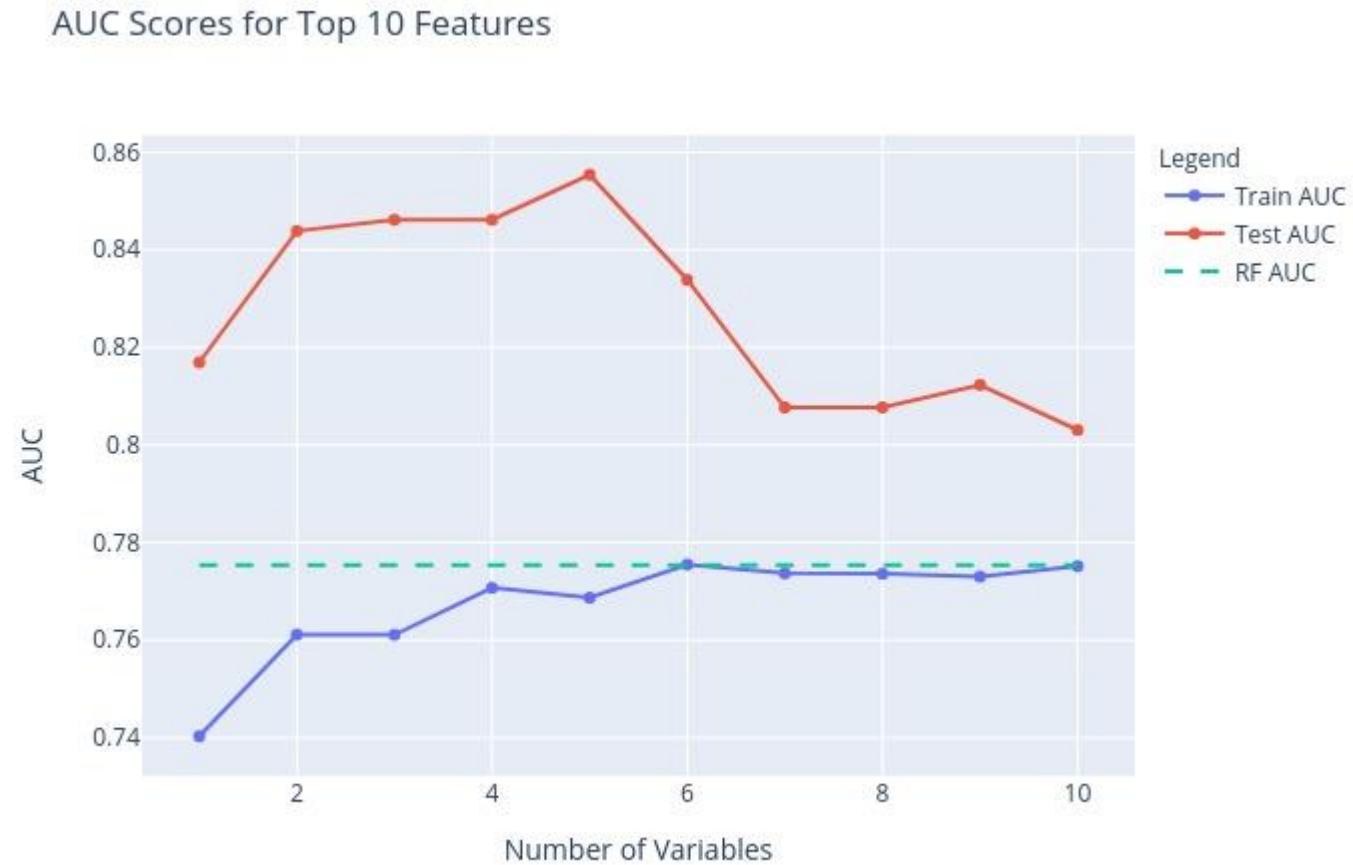
Additive Logistic Regression

LR Coefficients for Top 4 Features



3.2: Modelling

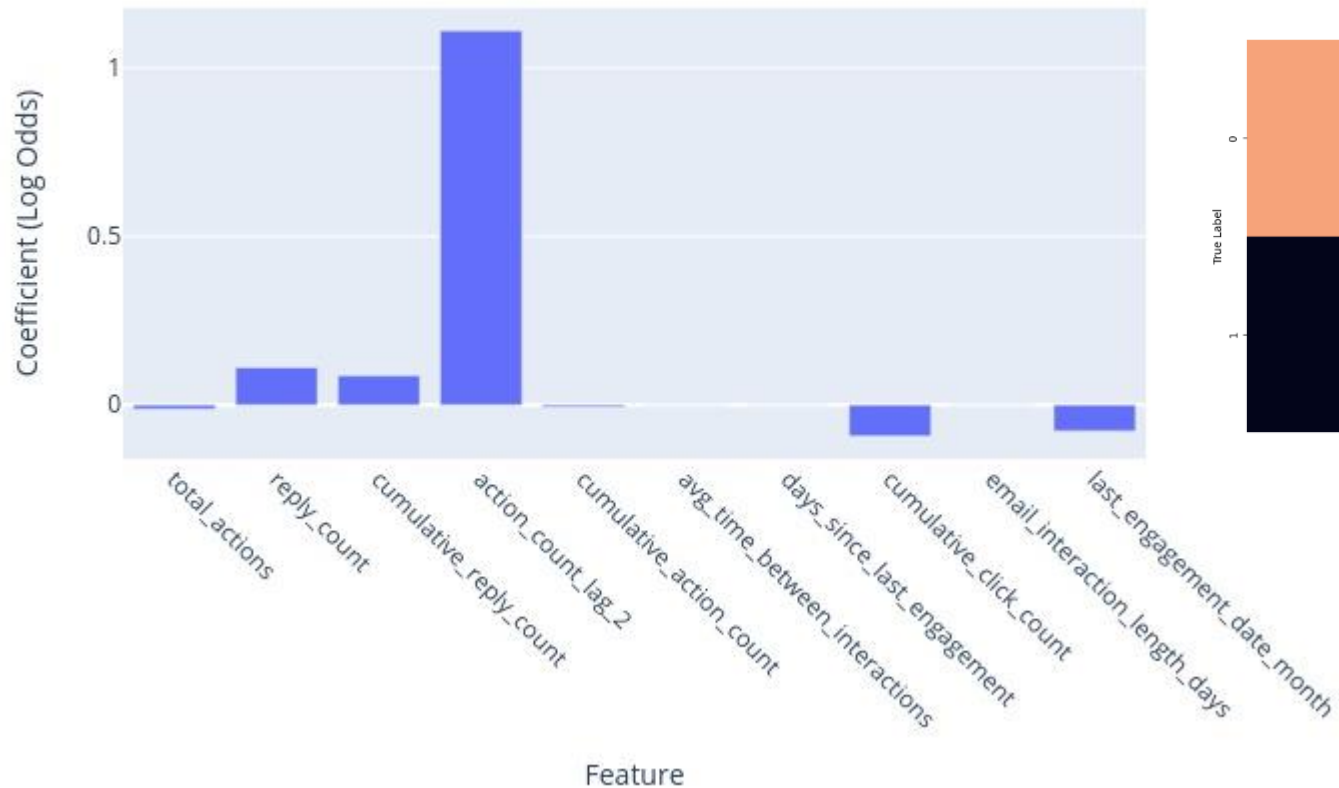
Additive Logistic Regression



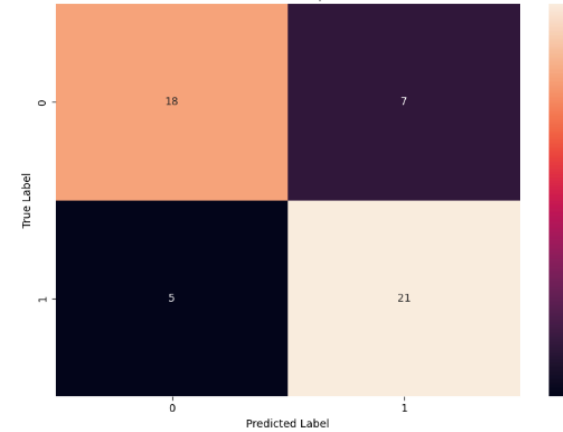
3.3: Modelling

Additive Logistic Regression

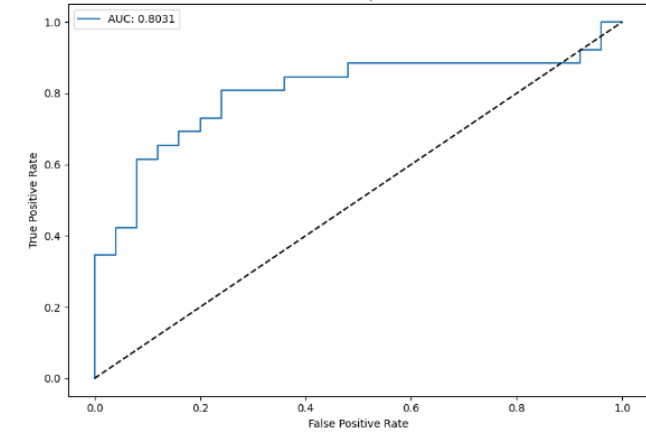
LR Coefficients for Top 10 Features



Confusion Matrix for Top 10 Features



ROC Curve for Top 10 Features



4: Prediction Results

- Model with 4 predictors:
 - 40 potential costumers
- Model with 10 predictors:
 - 35 potential costumers

...30 in common

4.1: Prediction and Lead Scoring

```
potential_accounts_high = potential_accounts[potential_accounts['Max_Proba']>0.7]
potential_accounts_high.head()
potential_accounts_low = potential_accounts[potential_accounts['Max_Proba']<0.7]
potential_accounts_low.head()
```

✓ 0.0s

	Annual_Potential_Yield	Predicted_Proba_10	Predicted_Proba_4	Max_Proba
id_number				
1158	46800.0	0.778099	0.864448	0.864448
1369	3600000.0	0.832385	0.765538	0.832385
2868	10080.0	0.828805	0.789253	0.828805
2944	43200.0	0.754701	0.810326	0.810326
2464	66600.0	0.720021	0.797857	0.797857
1361	25920.0	0.719157	0.784024	0.784024
506	32400.0	0.691803	0.759803	0.759803
1567	5400.0	0.656276	0.754359	0.754359
2822	27360.0	0.748654	0.728473	0.748654
2792	19800.0	0.636146	0.728473	0.728473

Total of ~3.9mio USD*

	Annual_Potential_Yield	Predicted_Proba_10	Predicted_Proba_4	Max_Proba
id_number				
2772	21600.0	0.677301	0.612369	0.677301
2796	56880.0	0.672899	0.607900	0.672899
1410	23400.0	0.594731	0.636298	0.636298
2347	54000.0	0.628731	0.561620	0.628731
1086	108000.0	0.570528	0.624376	0.624376
1042	7200.0	0.621915	0.547726	0.621915
767	7200.0	0.557432	0.612538	0.612538
775	7920.0	0.607234	0.477261	0.607234
945	9000.0	0.576478	0.528874	0.576478
911	18000.0	0.541531	0.561620	0.561620
2917	2160.0	0.506758	0.557370	0.557370
1670	14040.0	0.548986	0.499823	0.548986
1006	1800.0	0.536931	0.400959	0.536931
2797	108000.0	0.529697	0.452132	0.529697
2483	900000.0	0.510769	0.408200	0.510769
402	4320.0	0.507253	0.419324	0.507253
1882	16200.0	0.506069	0.419324	0.506069
2191	5400.0	0.468050	0.499823	0.499823
3005	23040.0	0.425038	0.442842	0.442842
1370	217440.0	0.442833	0.433591	0.442833

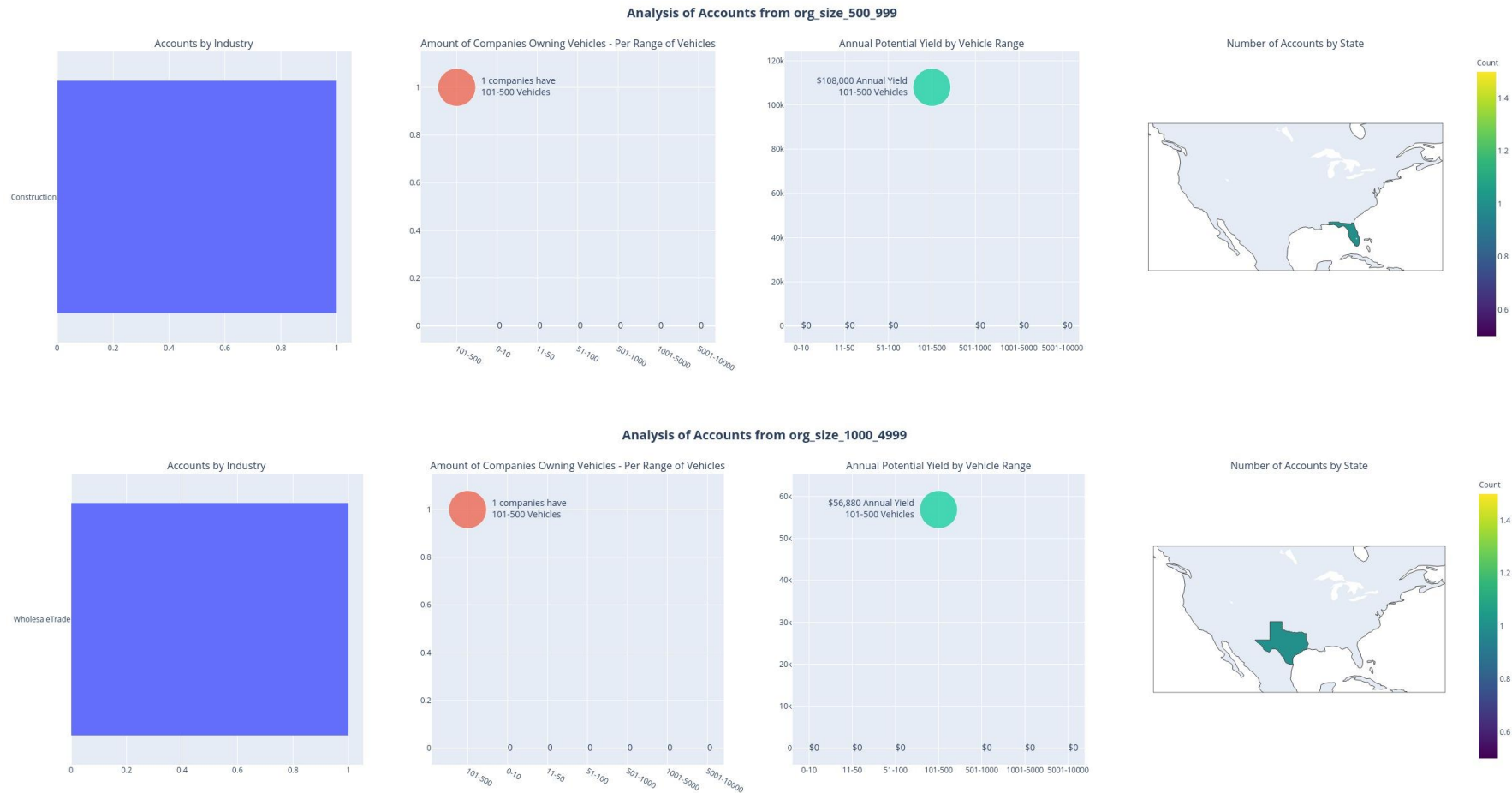
Total of ~1.6mio USD*

*Based on an Assumption of 30\$ USD monthly subscription per vehicle. [Source](#)

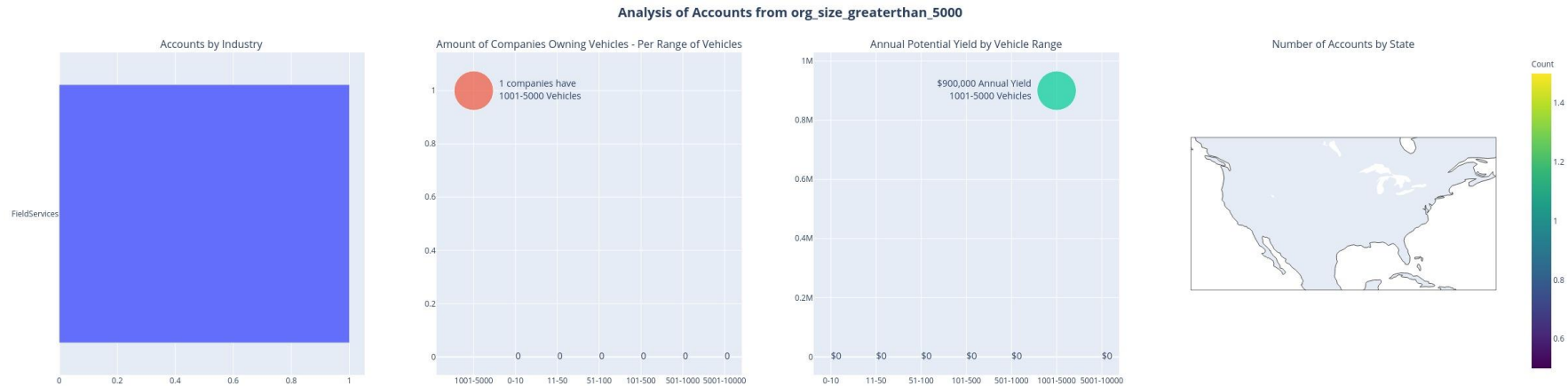
5: Insights and Business Impact



5.1: Insights and Business Impact

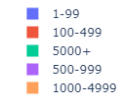
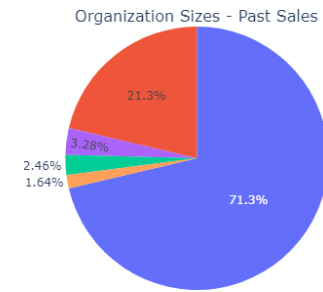
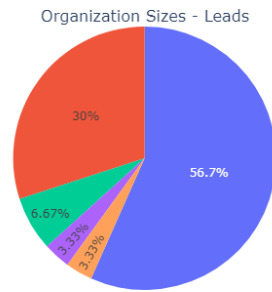


5.2: Insights and Business Impact

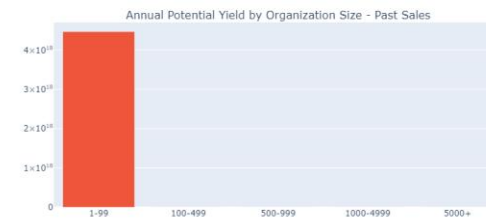
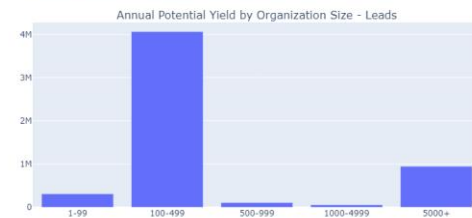


5.3: Sales Cockpit

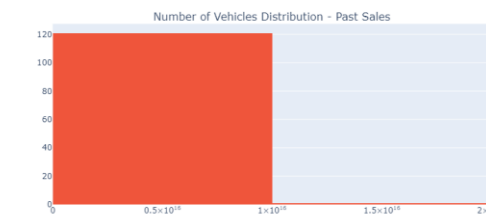
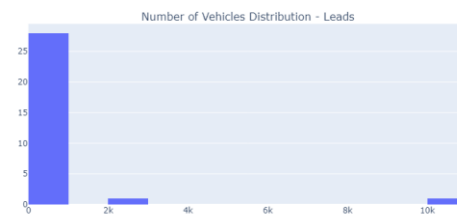
Organization Sizes



Annual Potential Yield by Organization Size



Number of Vehicles Distribution



6: Limitations

- Sparsity of data:
 - Trained only on 250~ observations
 - Predicted only on 100~ observations
 - ...the database has ~3k companies!
- Quantitative insights into probability of accounts might not align with Sales' qualitative goals to acquire a client
 - More alignment necessary between Sales strategic goals to ensure synergic and aligned data modelling

End of slides.