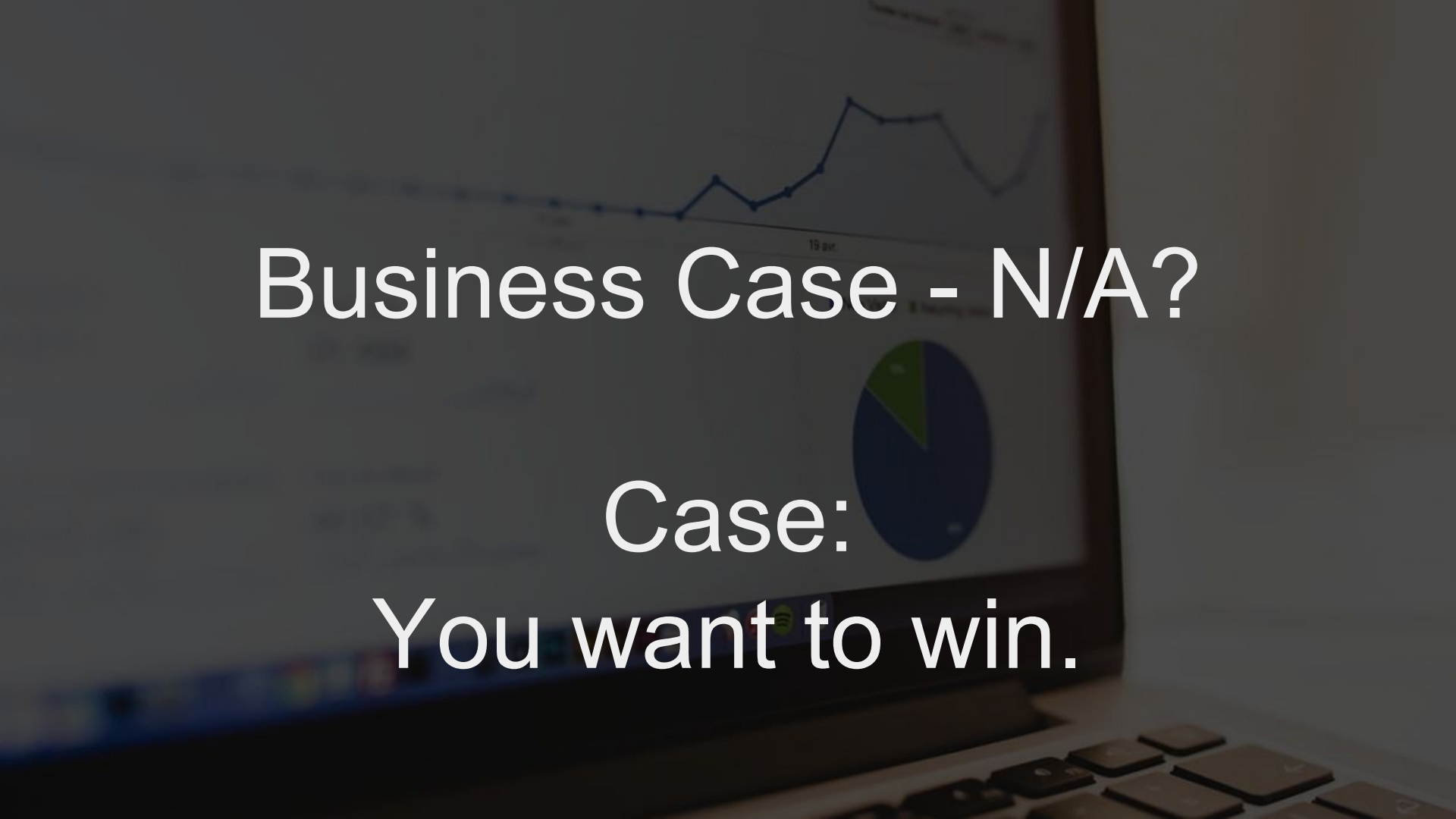
Analisis on Test Data



Business Case - N/A?

Case:
You want to win.

Tools

pandas

numpy

matplotlib

scipy-stats

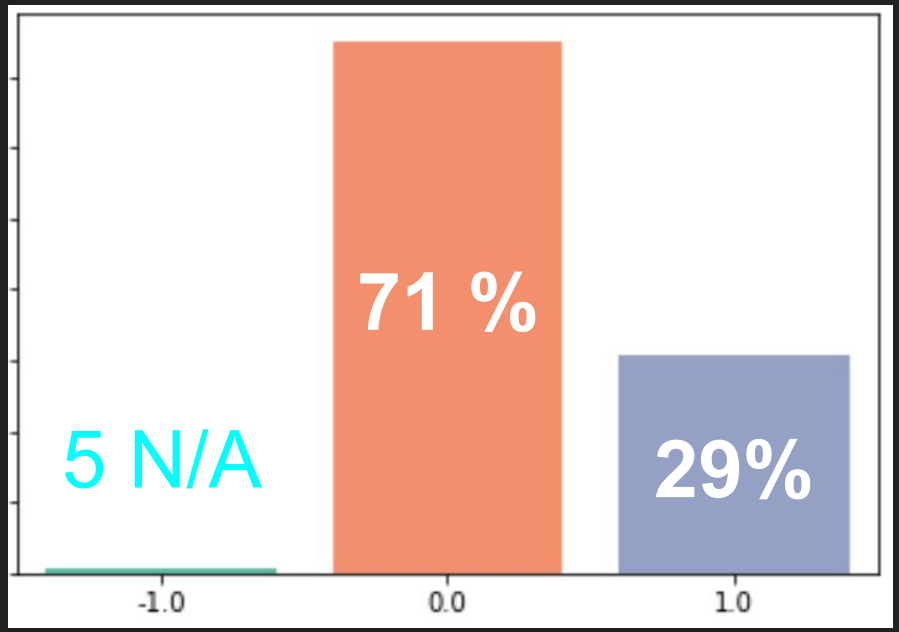
seaborn

Data

Predictors

Account Name
Contact: Title
Lead Source
Industry
Location
Amount
Close Date & Created Date

Target: Won



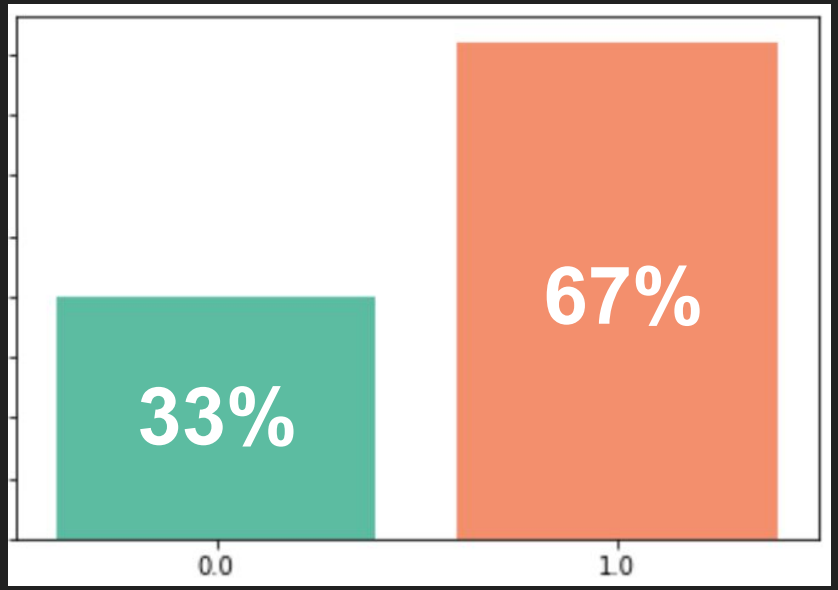
(1061 , 11) - csv

(1056 , 11)

Methodology

Generate new features
Data Processing
Correlations
Subsets of data

Won



Duplicate Ratio: 12 %

Account Name

Some Accounts Repeat

Up to 8 Times

TITLE

Fragmented Feature With 411 Unique Values

(Ceo = CEO = ceo = Chief executive officer)

Solution → Cluster

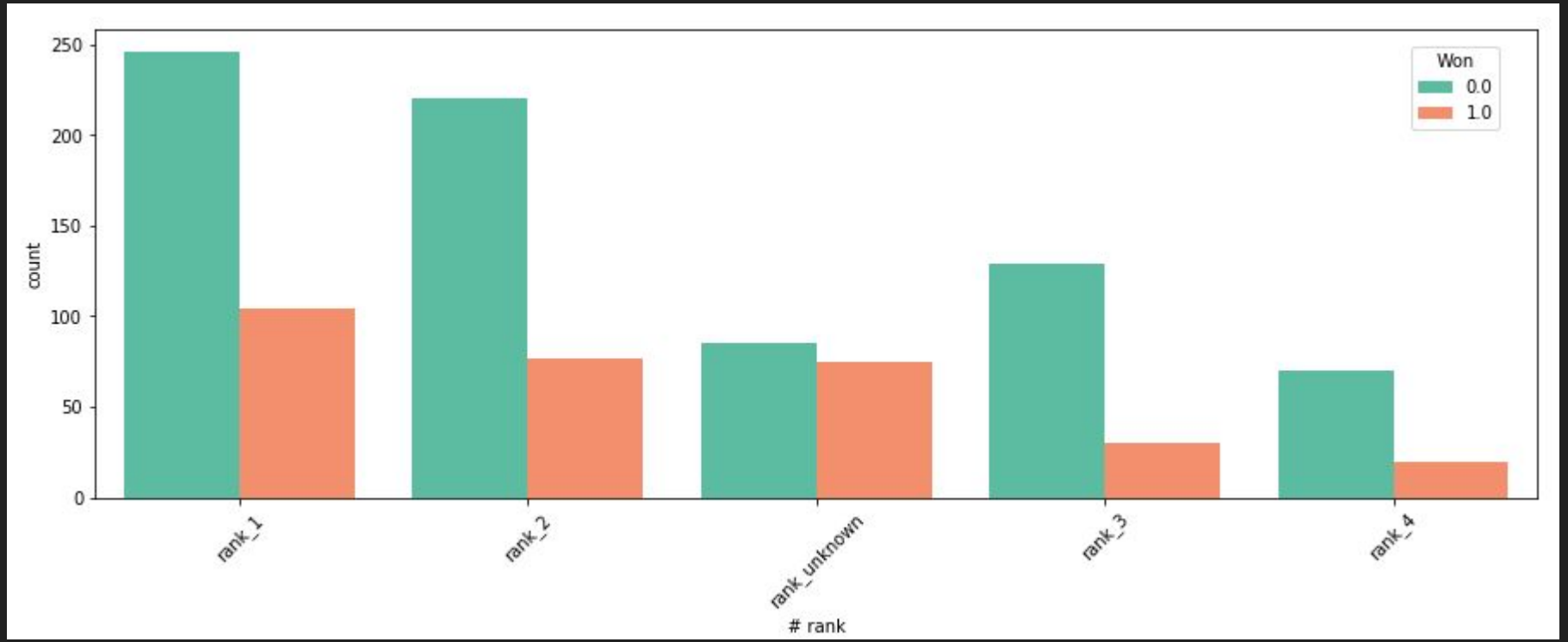
Rank:
Leadership
Hierarchy

(5)

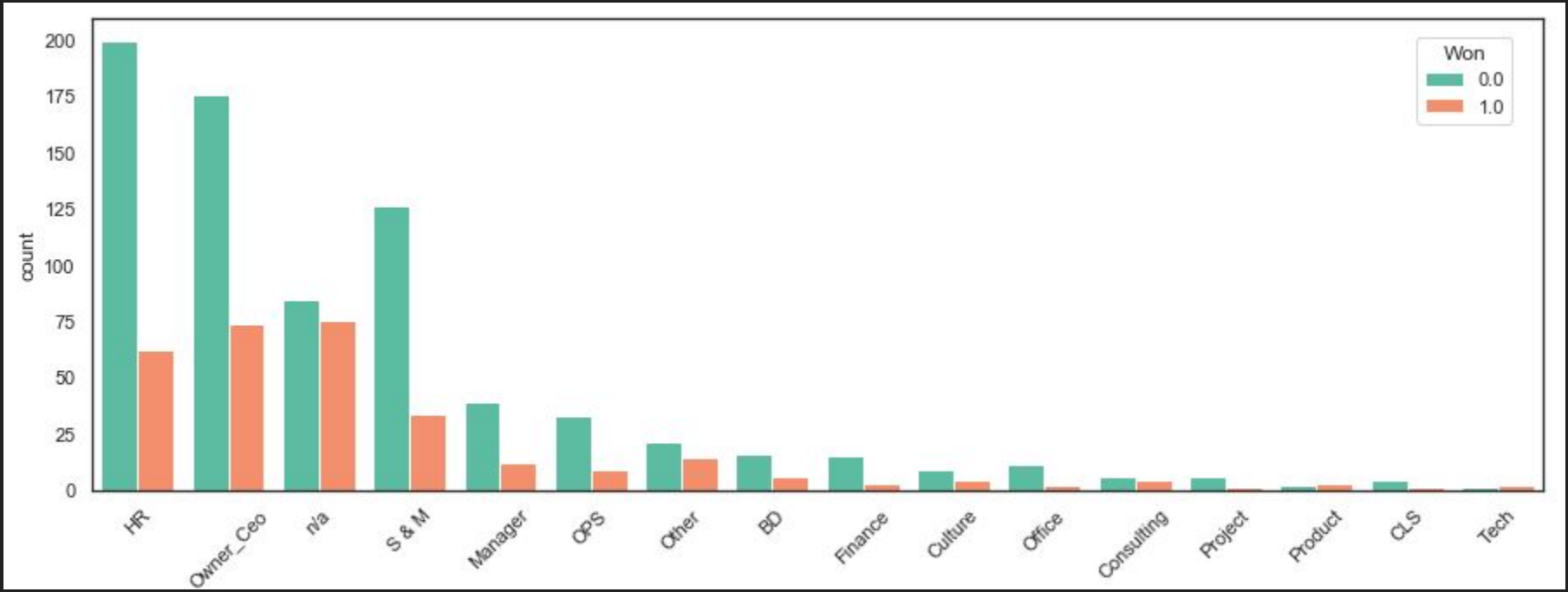
Job Type:
Professional
Occupation

(16)

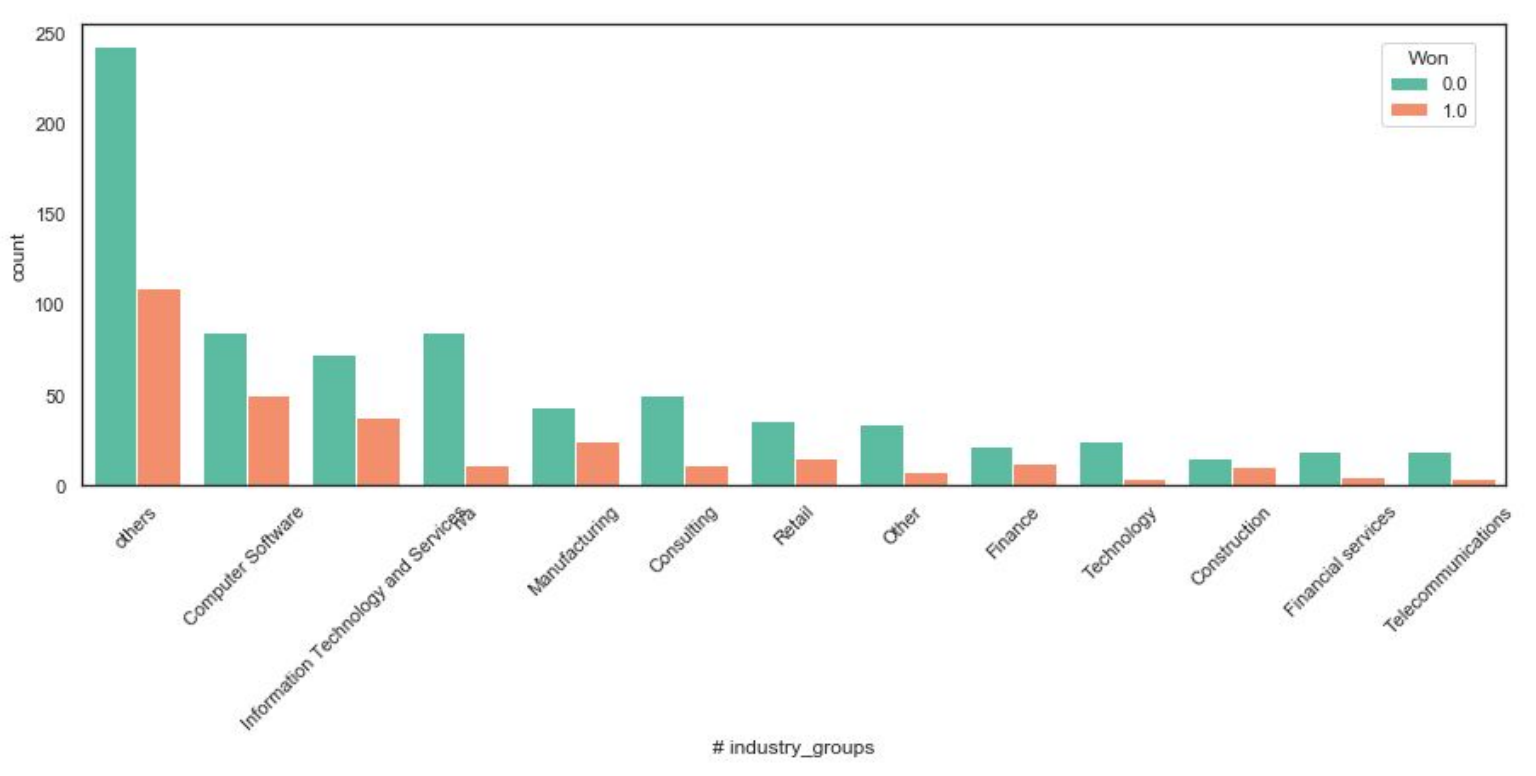
TITLE: Leadership Rank



TITLE: Job Category



Industry

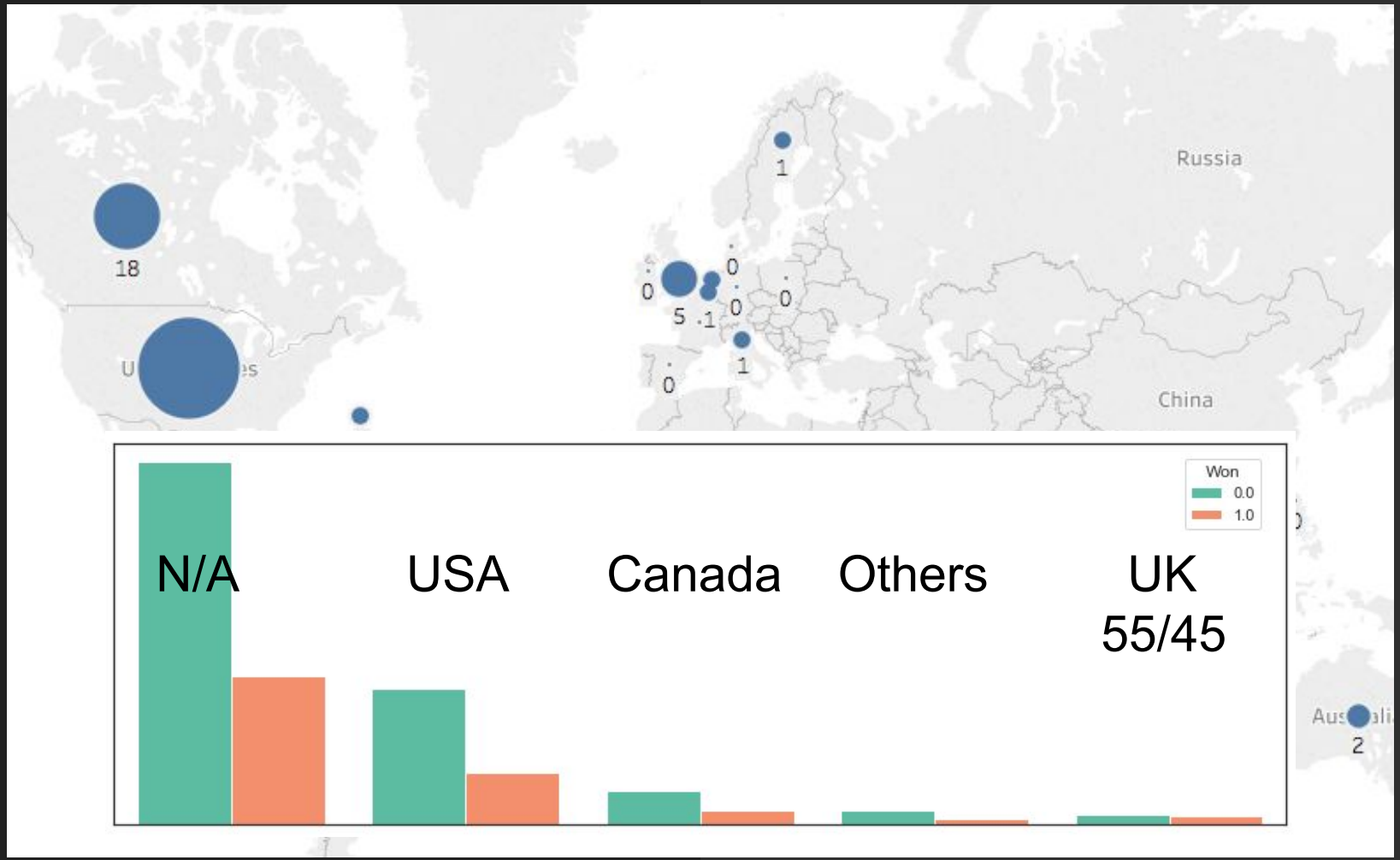

Won
0.0
1.0
count
250
200
150
100
50
0
others
Computer Software
Information Technology and Services
Manufacturing
Consulting
Retail
Other
Finance
Technology
Construction
Financial services
Telecommunications
industry_groups

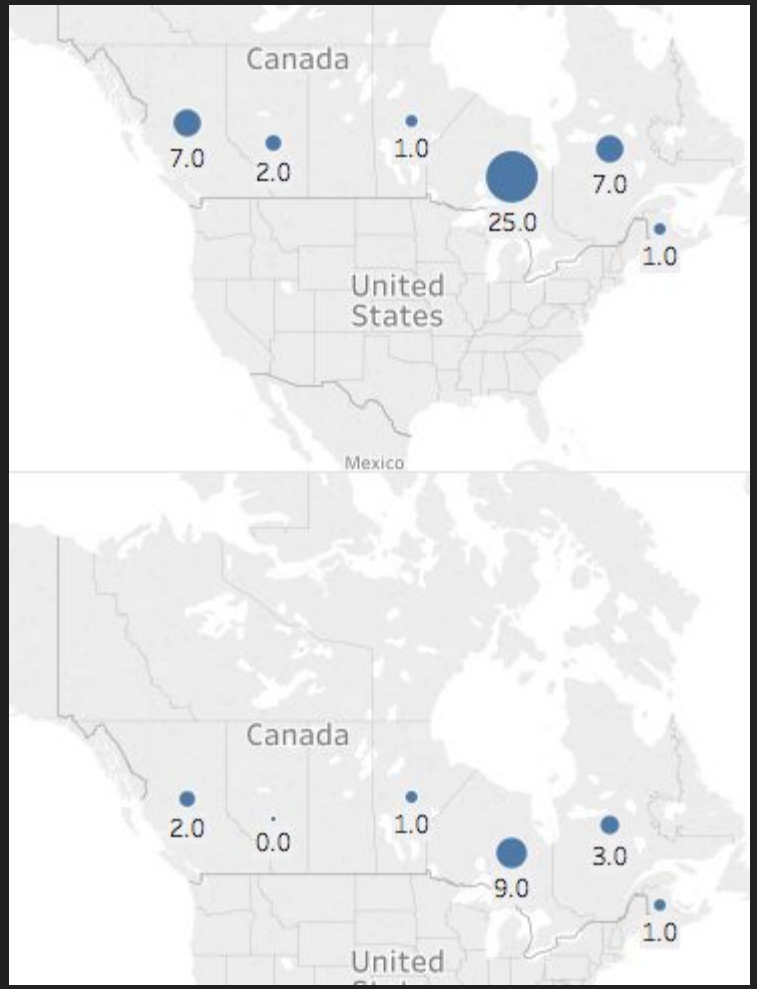
Location

City - State/Province - Country

~ 60 % missing data

No Alignment Among Null values

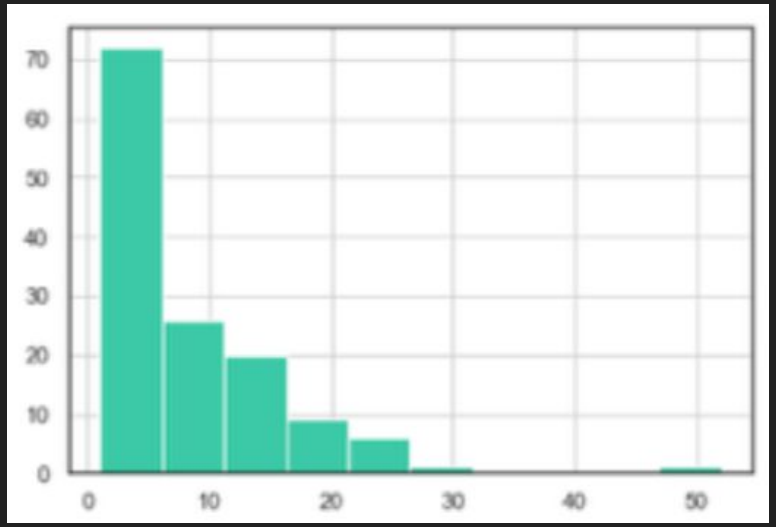

Russia
China
18
1
0
0
5
1
0
0
0
1
0
Won
0.0
1.0
N/A
USA
Canada
Others
UK
55/45
2


Canada
7.0
2.0
1.0
25.0
7.0
1.0
United
States
Mexico
Canada
2.0
0.0
1.0
9.0
3.0
1.0
United

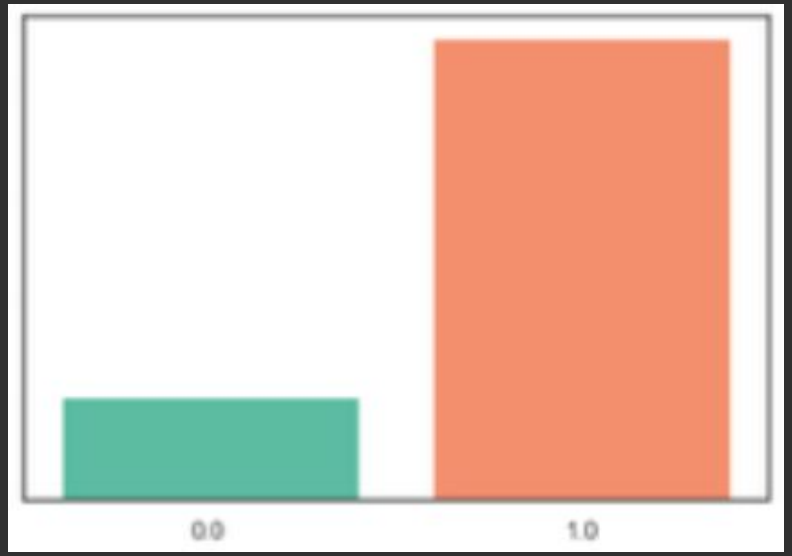
PROVINCE

TIME
Create Date & Close Date

Number of Registered Days



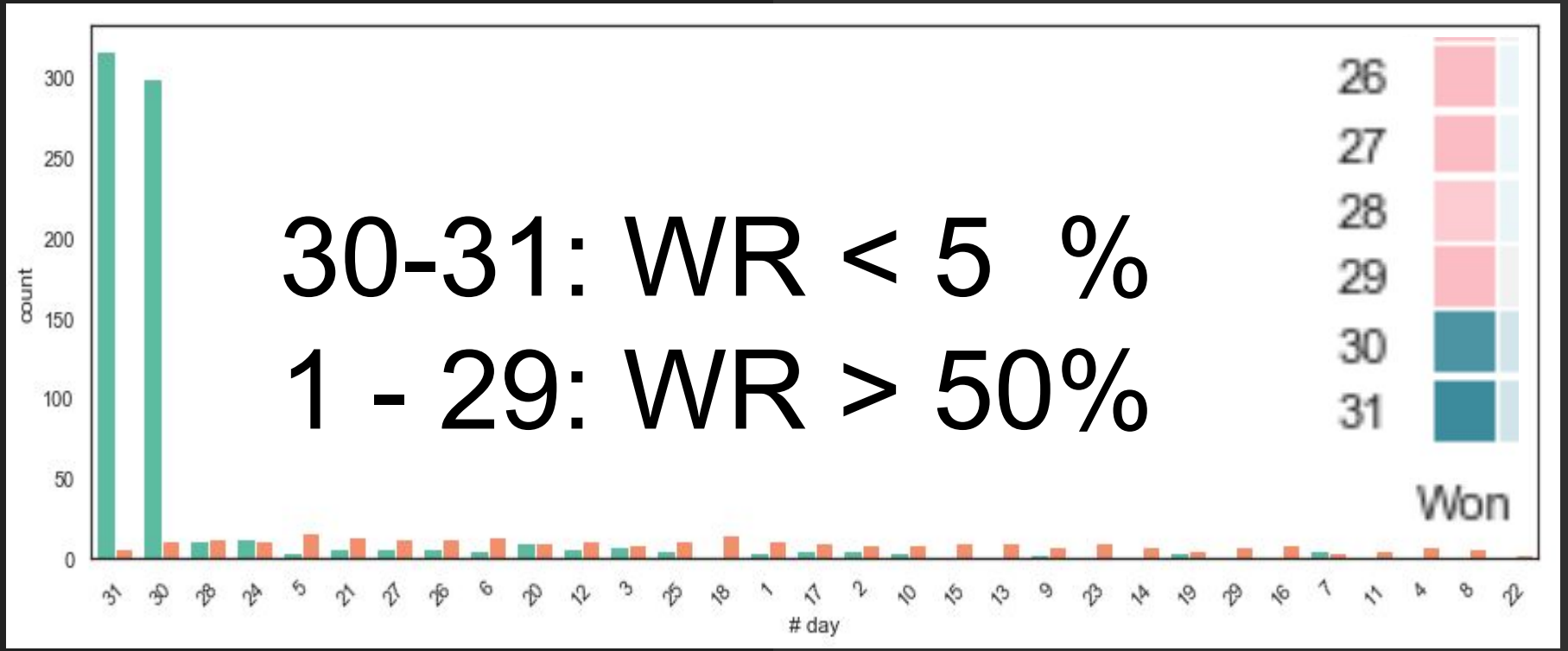
Created_date - Close_date



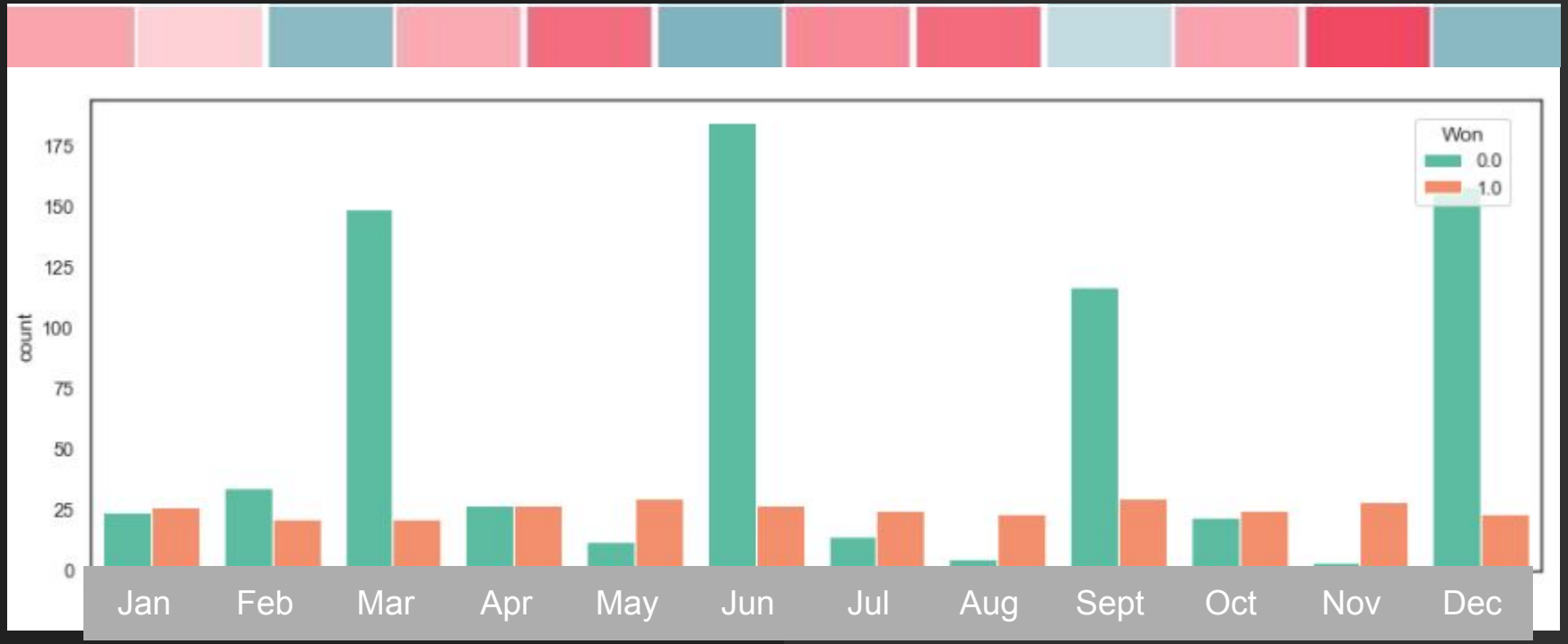
Abnormal Registered Days

Count: 11 - WR: 85%

Time - Day



Win: Nov, Aug, May



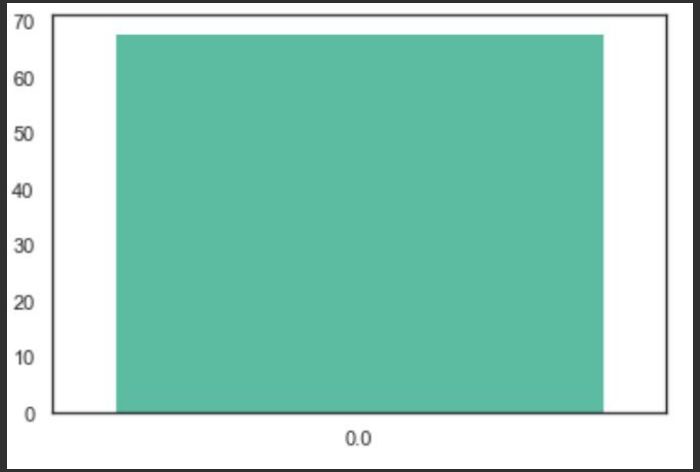
Lose: Mar, Jun, Dec

Time - Month

2017	Year	2018
4.149.785	Won Amount	4.446.500
9.746.500	NON_Won_Amont	11.712.730
13.896.285	Total	16.159.230
14%	Win Rate	15%

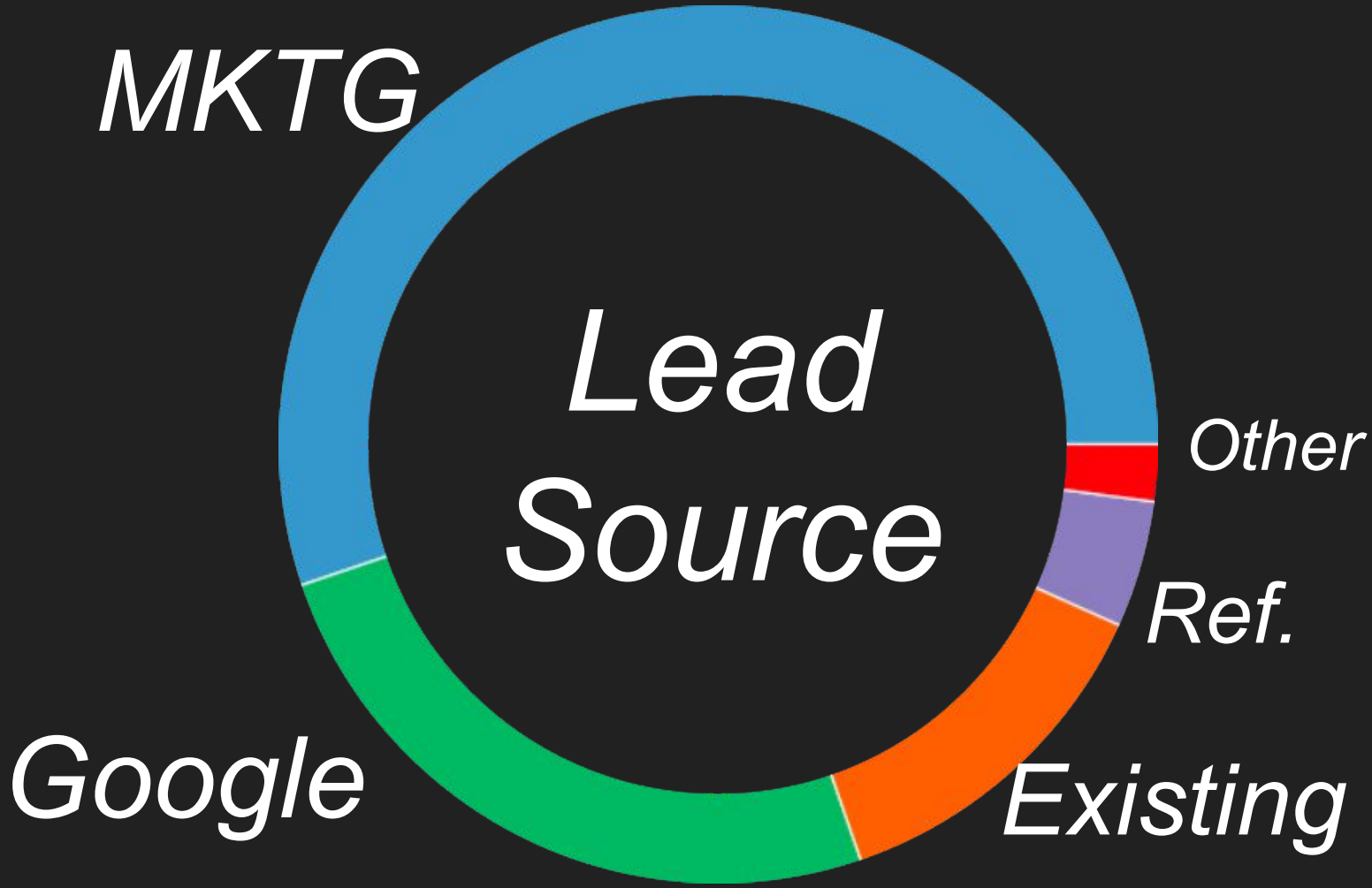
AMOUNT

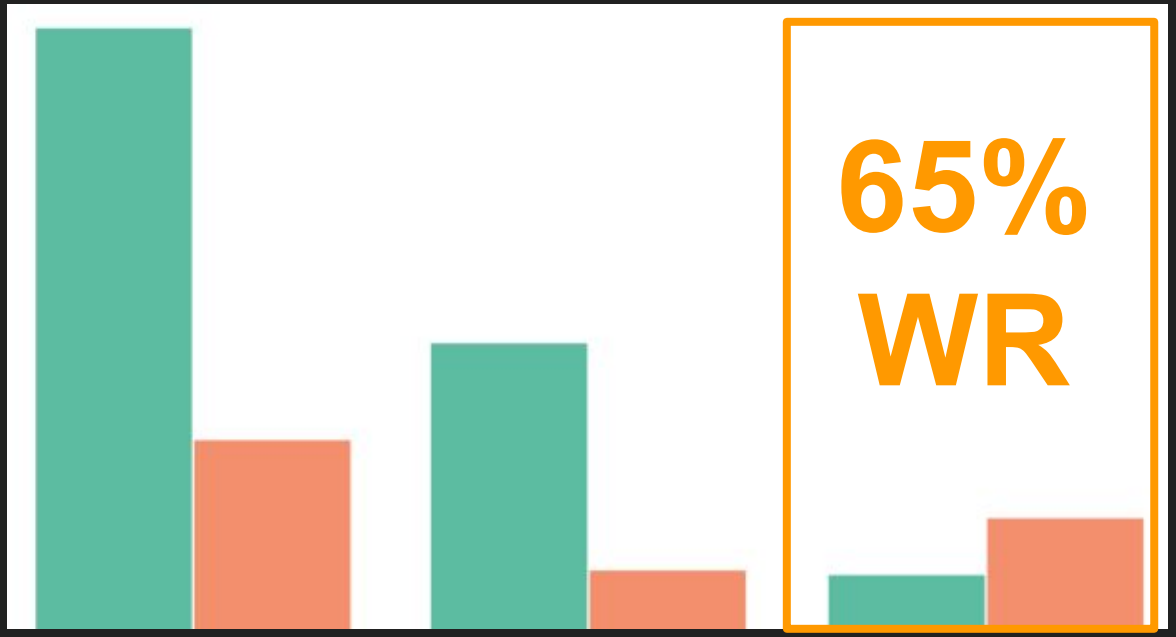
Subcategory: Missing Values



68 Data Points

0 % win rate


MKTG
Lead Source
Other
Ref.
Existing
Google



Existing Customers

Still want to WiN?

Create multiple account names

Do not declare your title

Enroll from UK

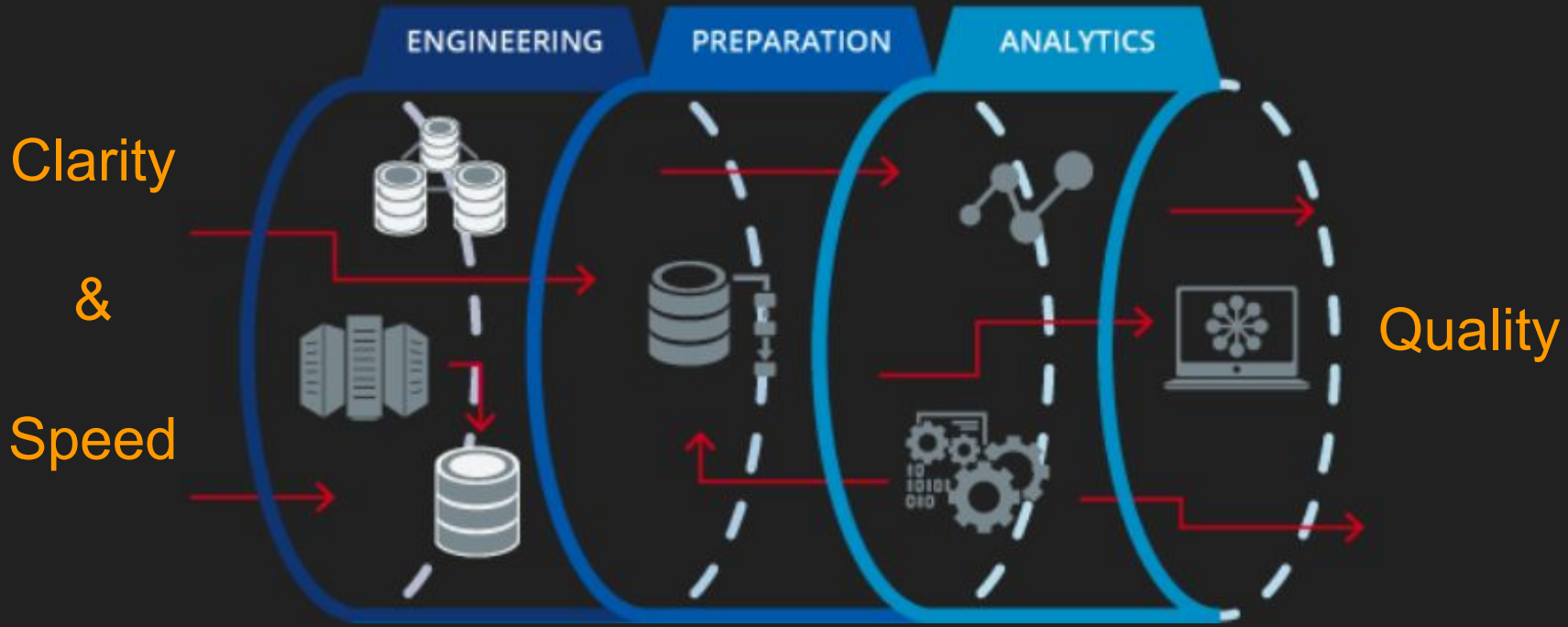
Twitch those registered days

Close your account at the middle of the month

Preferably Nov.

Invest some amount

Major Recommendation: Data Processing Pipeline

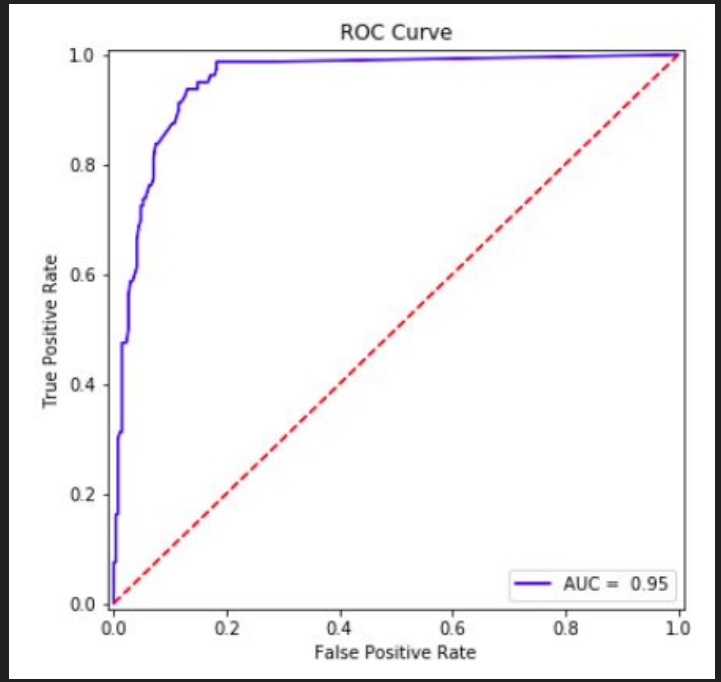


Minor Recommendations

Check issues about:

- Duplicate Account names
- Abnormal registered days
- Undocumented Amounts

-Bonus-
ML: GradientBoostClassifier



	precision	recall	f1-score	support
0.0	0.95	0.93	0.94	269
1.0	0.77	0.84	0.80	80

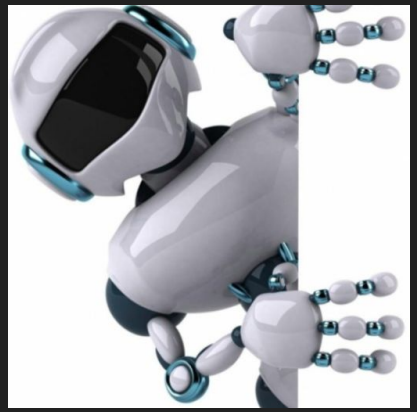
Questions & Comments

Jan

Education

Critical Thinking
Problem Solving Skills
Analytical Mindset

Mechatronics Engineering



Emphasis on $
Business Metrics
Growth Process
Business acumen

MBA

Change-Management

Probability
Statistics
ML algorithms
Databases
Data visualization

Projects
Network

Projects

Fraud Detection

Churn Prediction

A/B testing(multi variete)

Equipment Price Prediction

Customer Behaviour Prediction

Cash/Card prediction

Customer Behaviour is my passion!

Vision and Where I see myself in future?

TECH IS COOL

Effectively

Integrating Business and Data Science