CSGO Match Outcome Predictor

Justin Bell

TURTLE BEACH

How can we predict the outcome of a match using player and team statistics?

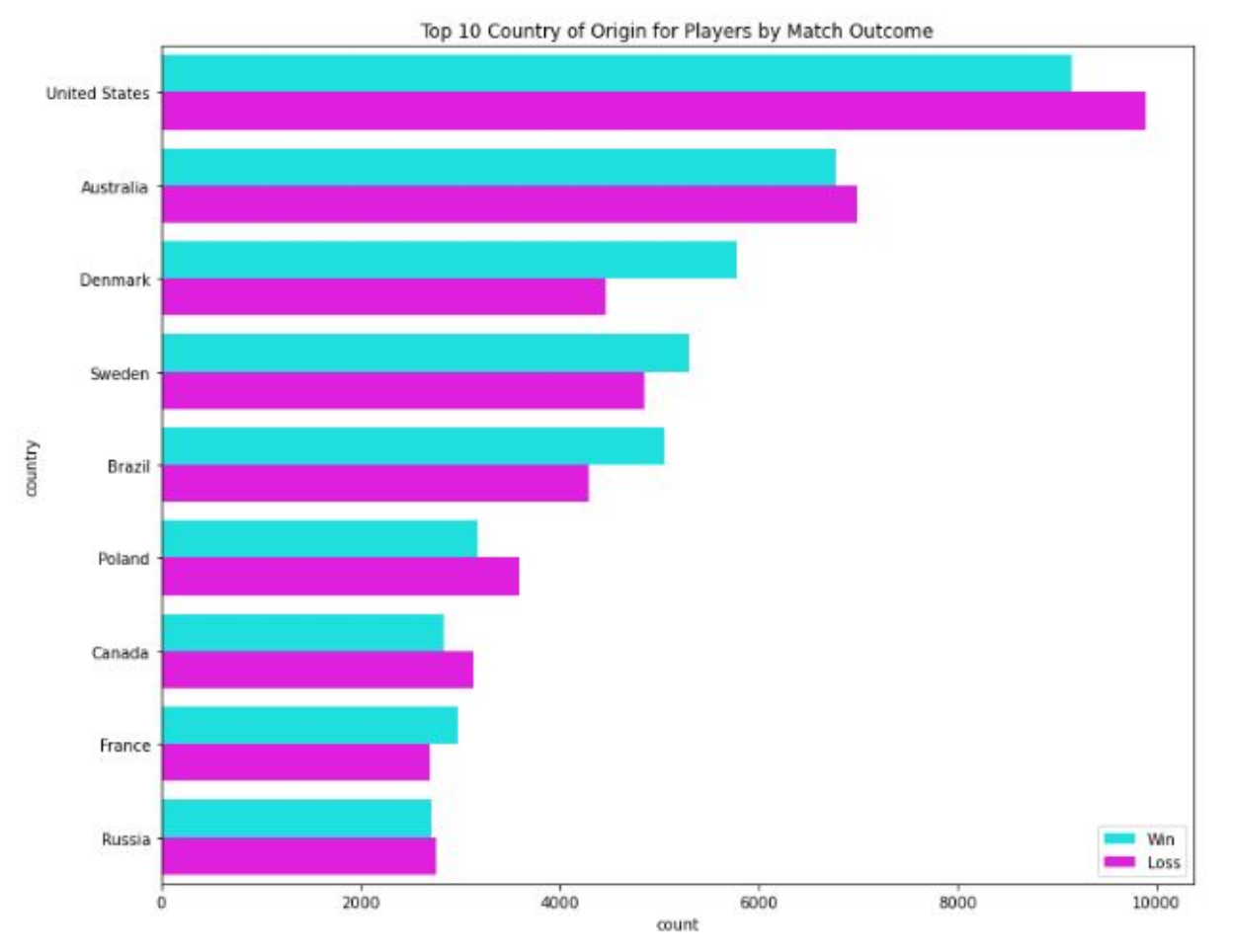
Data Wrangling

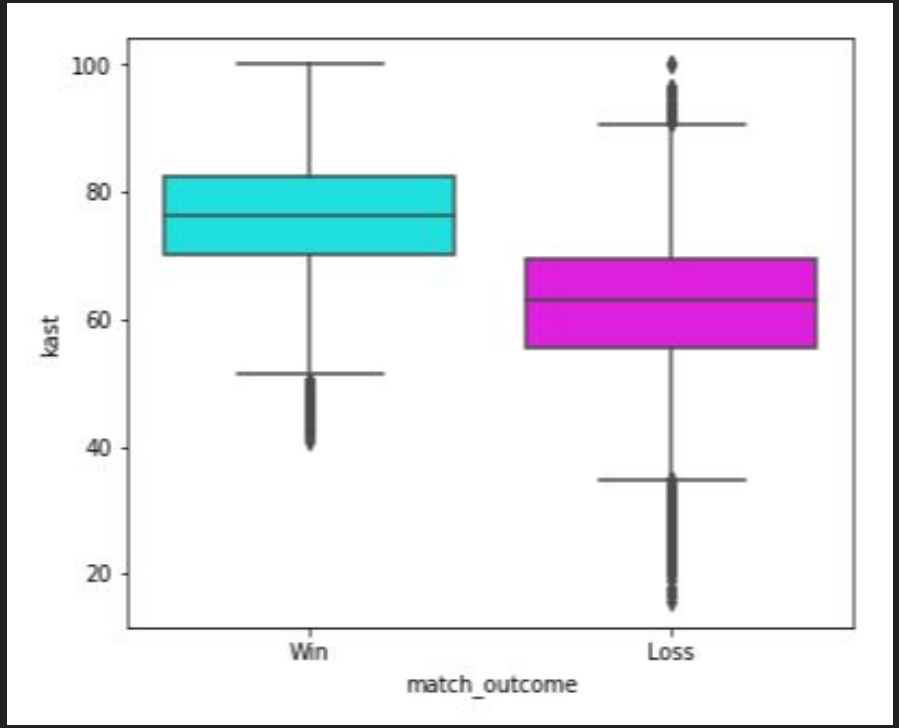
- Import Data
- Merge Dataframes
- Data Cleaning
  - Null values
  - Data types
  - Columns
- Create Team Dataframe

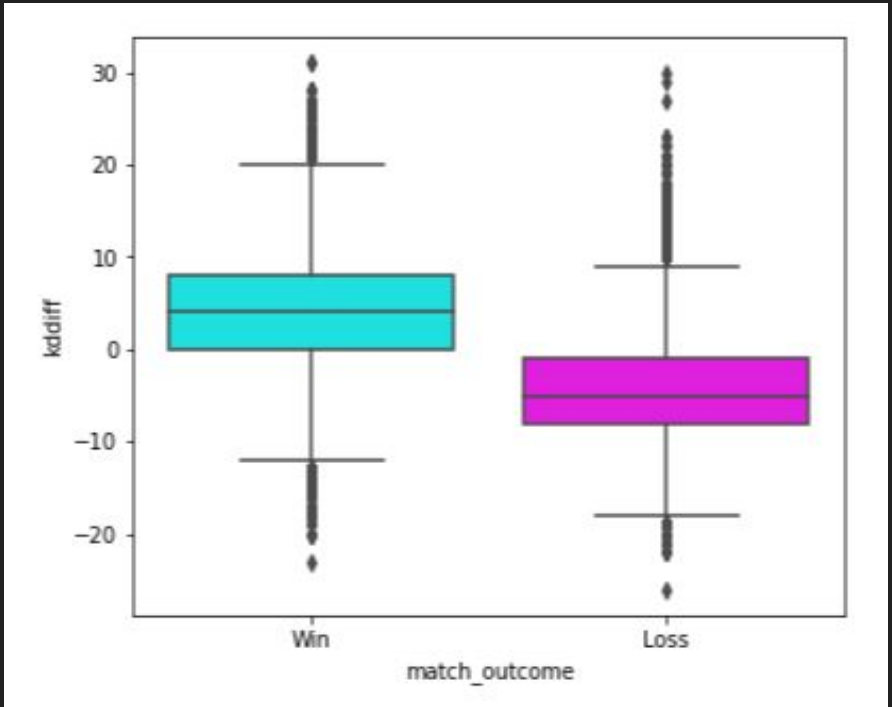
EDA

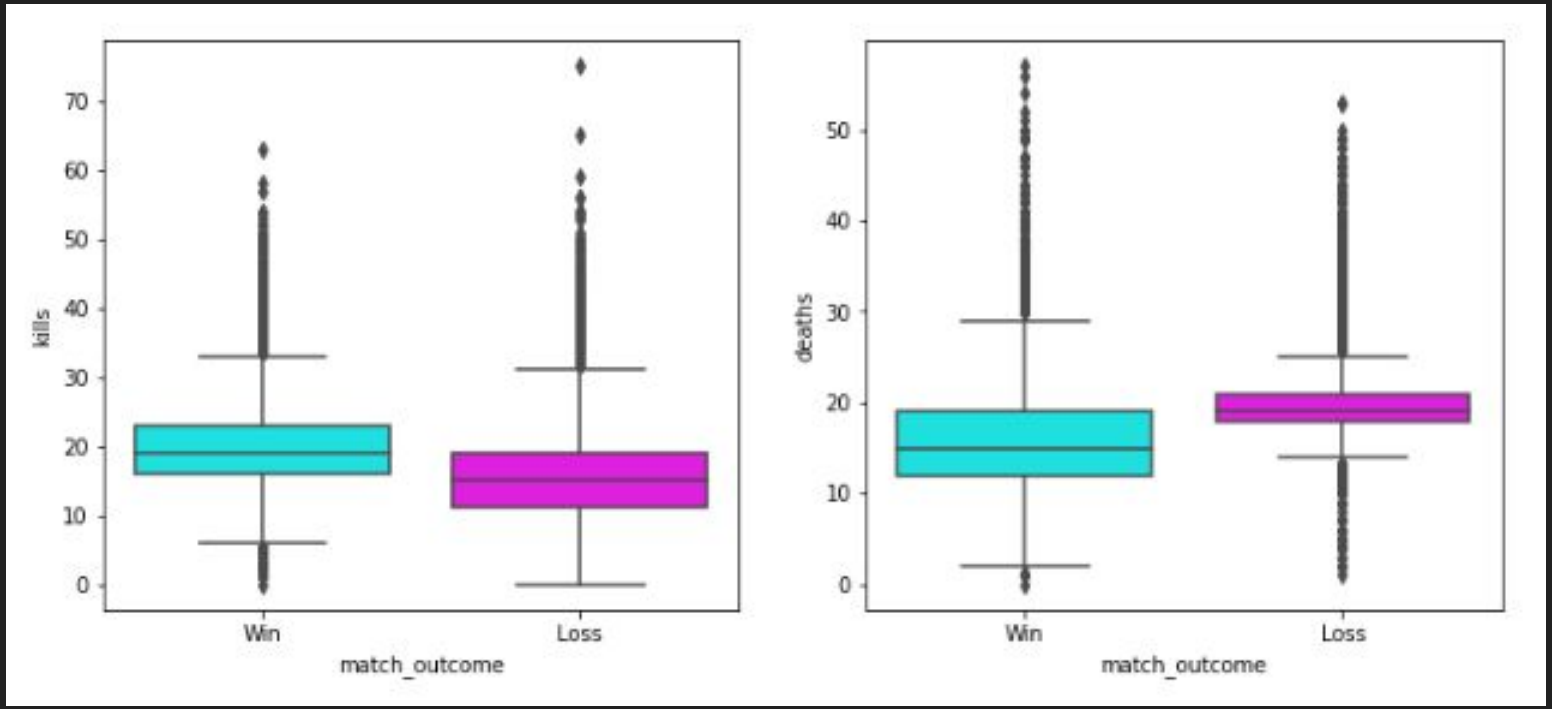
Individual Performance

- Split Data by Match Outcome
- Create Plots to Compare Distributions


Top 10 Country of Origin for Players by Match Outcome
United States
Australia
Denmark
Sweden
Brazil
Poland
Canada
France
Russia
country
0
2000
4000
6000
8000
10000
count
Win
Loss


100
80
60
40
20
kast
Win
Loss
match_outcome

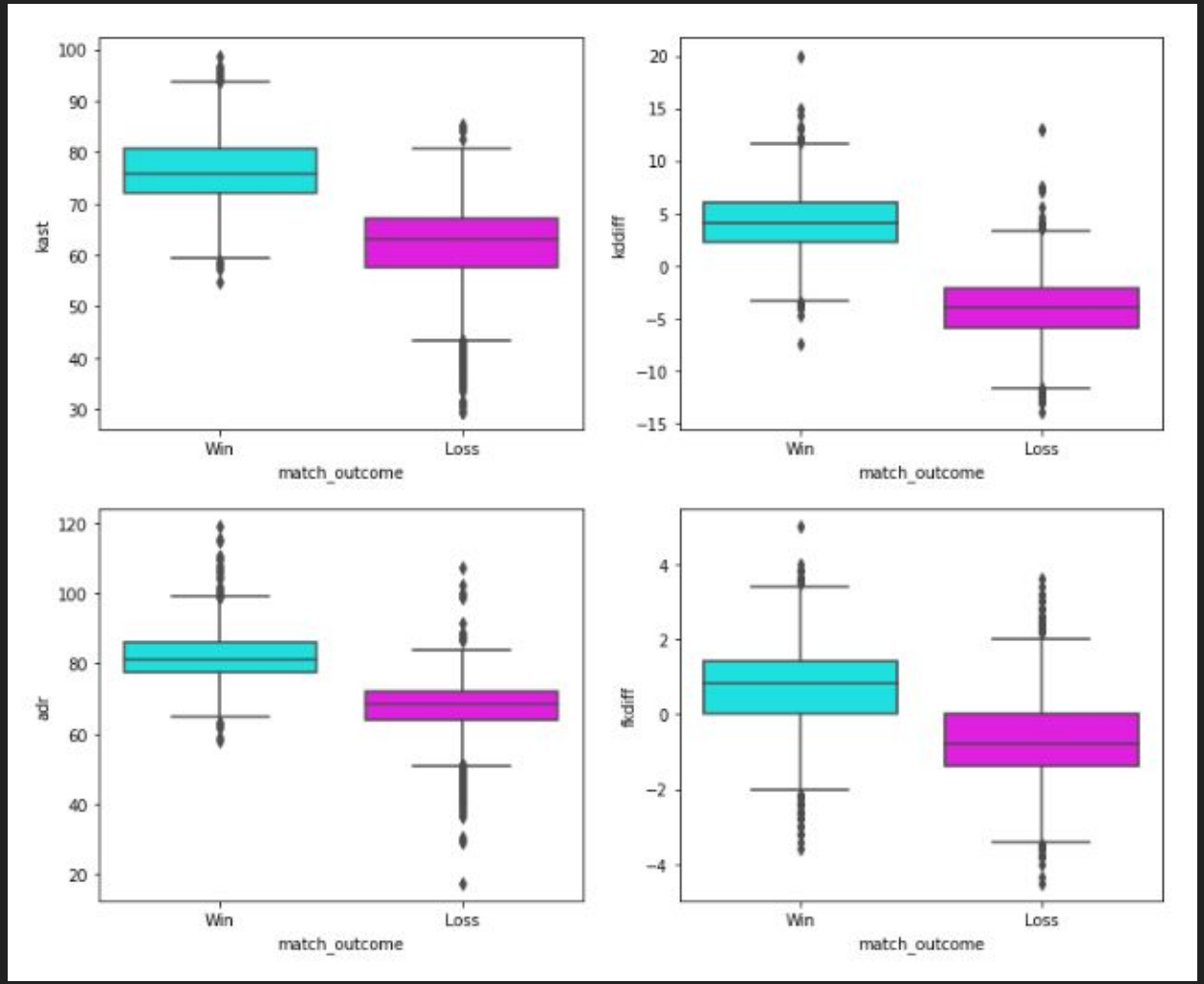

30
20
10
0
−10
−20
kddiff
Win
Loss
match_outcome


kills
70
60
50
40
30
20
10
0
Win
Loss
match_outcome
deaths
50
40
30
20
10
0
Win
Loss
match_outcome

EDA

Team Performance

- Split Data by Match Outcome
- Create Plots to Compare Distributions


kast
100
90
80
70
60
50
40
30
Win
Loss
match_outcome
kddiff
20
15
10
5
0
−5
−10
−15
Win
Loss
match_outcome
adr
120
100
80
60
40
20
Win
Loss
match_outcome
fkdiff
4
2
0
−2
−4
Win
Loss
match_outcome

Preprocessing

- Train/Test Split
- Split Numerical and Categorical Features
- Scale Numeric Features
- Encode Dummy Variables
- Merge Back Together

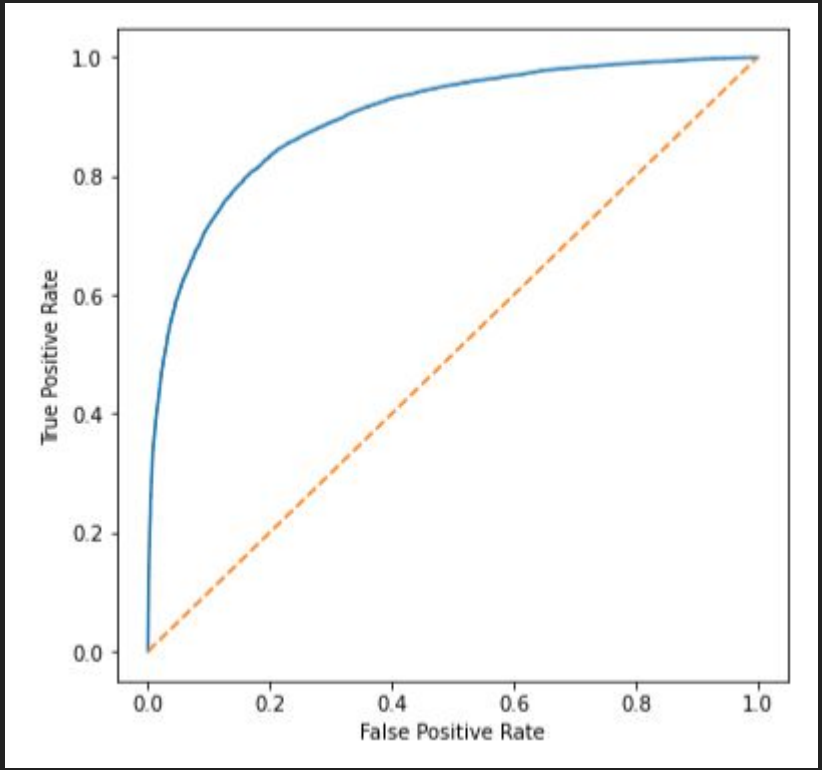
Modeling

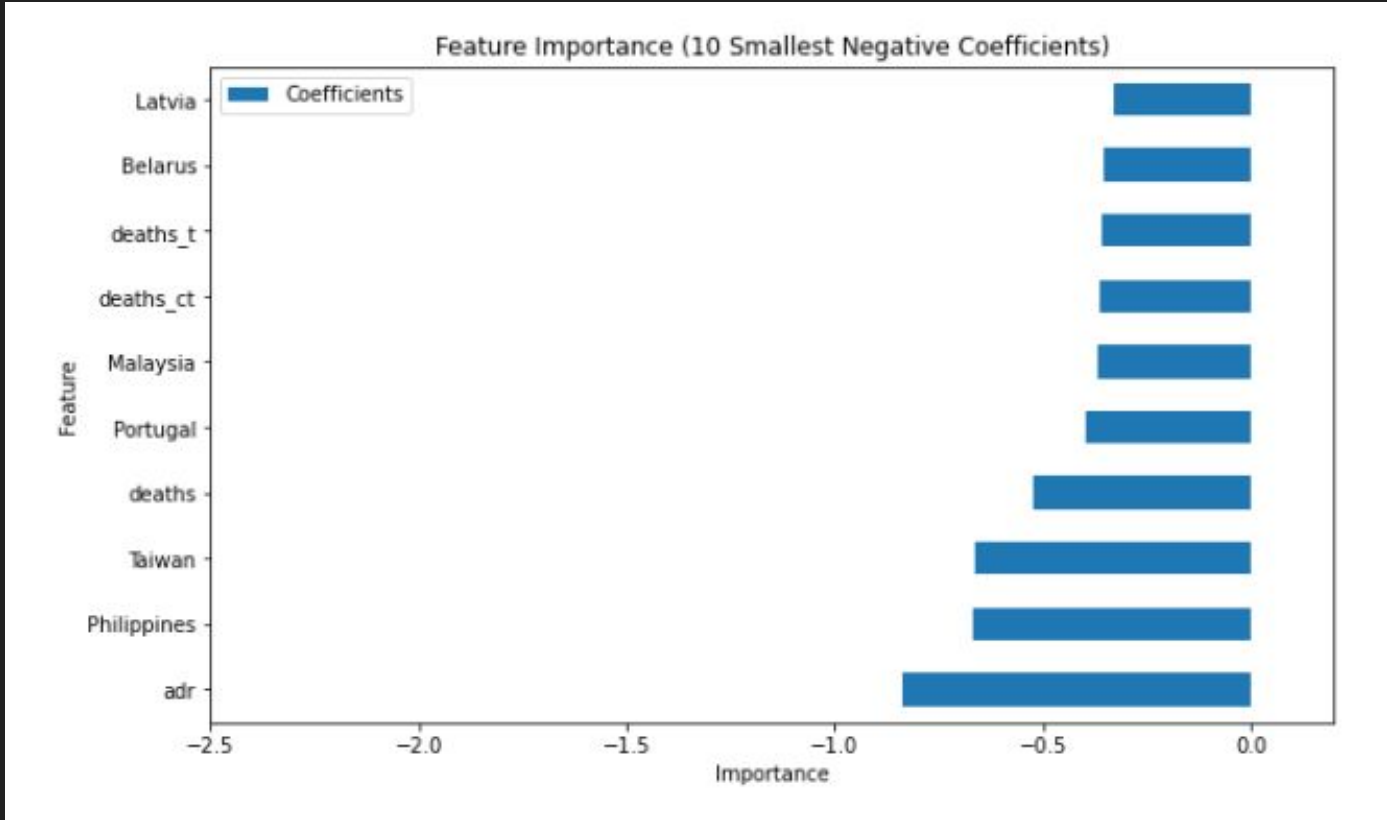
Individual Performance

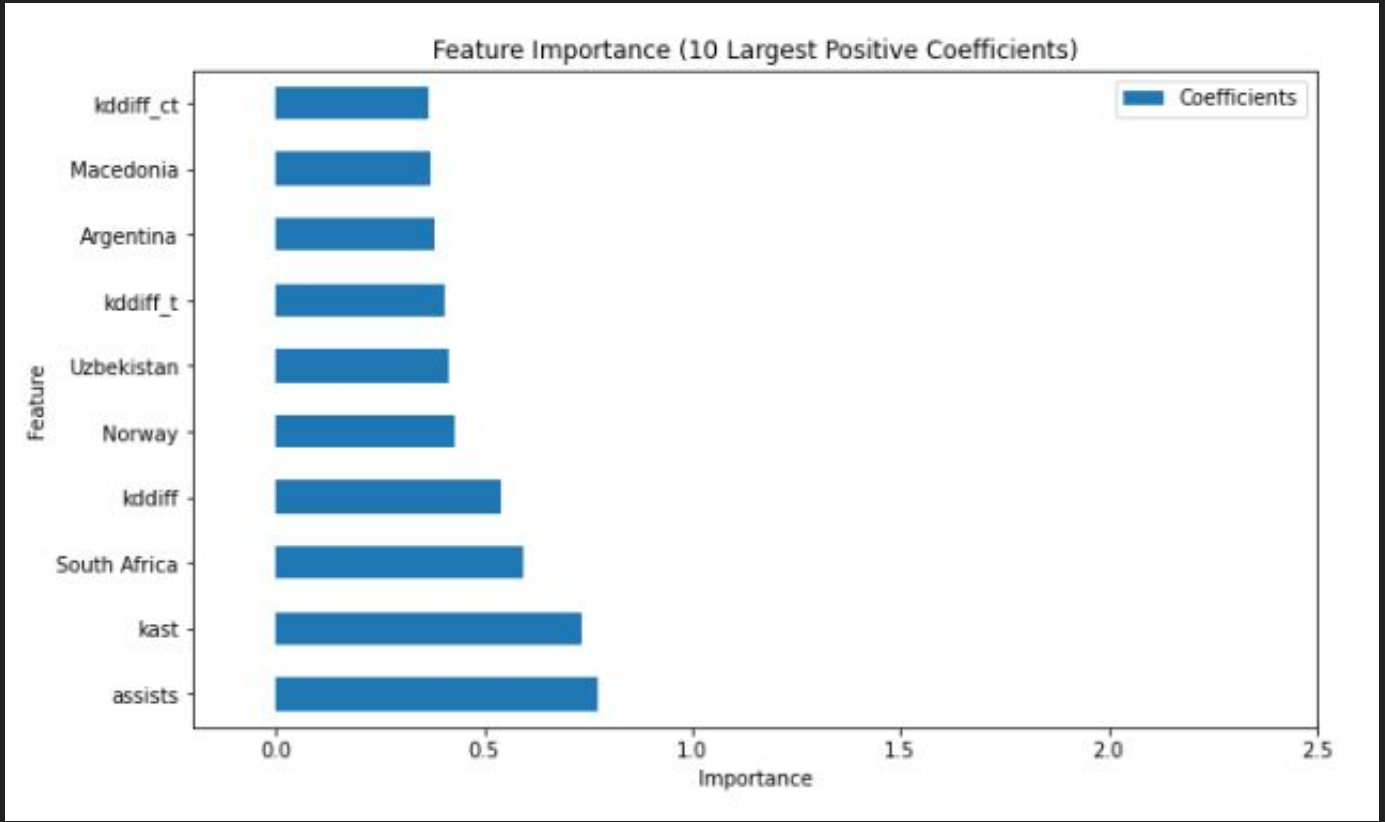
- Logistic Regression
  - F1-Score (Training): 0.817
- Random Forest
  - F1-Score (Training): 0.812
- Gradient Boosting
  - F1-Score (Training): 0.813

Best Model

- Logistic Regression
  - C: 0.1
  - Penalty: l2
  - Solver: liblinear
  - ROC-AUC Score: 0.898




Feature Importance (10 Smallest Negative Coefficients)
Coefficients
Latvia
Belarus
deaths_t
deaths_ct
Malaysia
Portugal
deaths
Taiwan
Philippines
adr
Feature
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
Importance


Feature Importance (10 Largest Positive Coefficients)
Coefficients
kddiff_ct
Macedonia
Argentina
kddiff_t
Uzbekistan
Norway
kddiff
South Africa
kast
assists
Feature
0.0
0.5
1.0
1.5
2.0
2.5
Importance

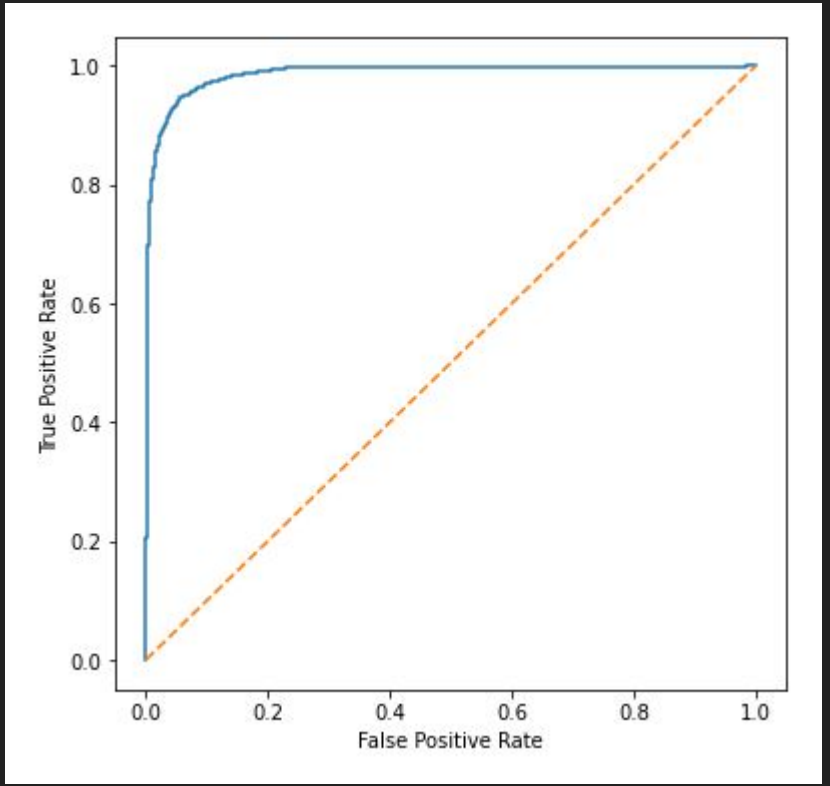
Modeling

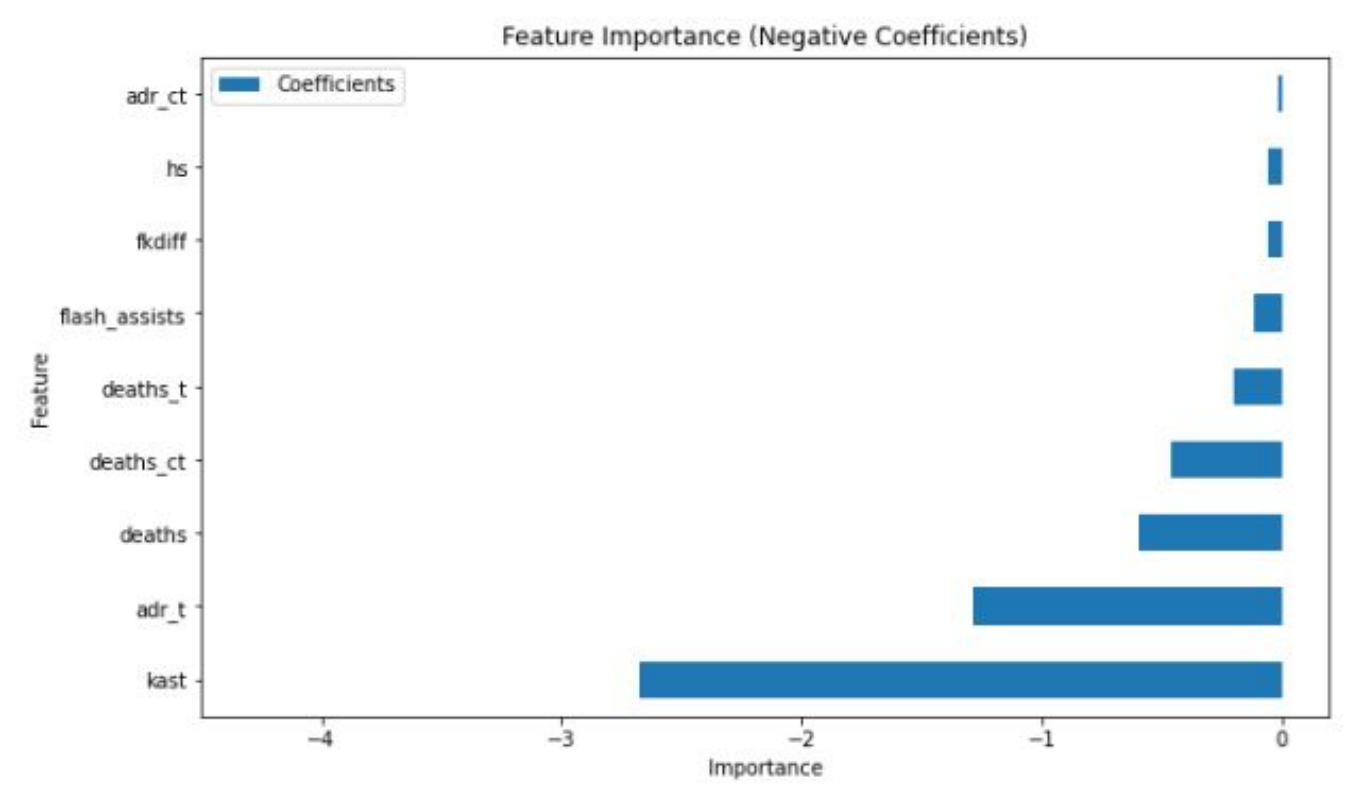
Team Performance

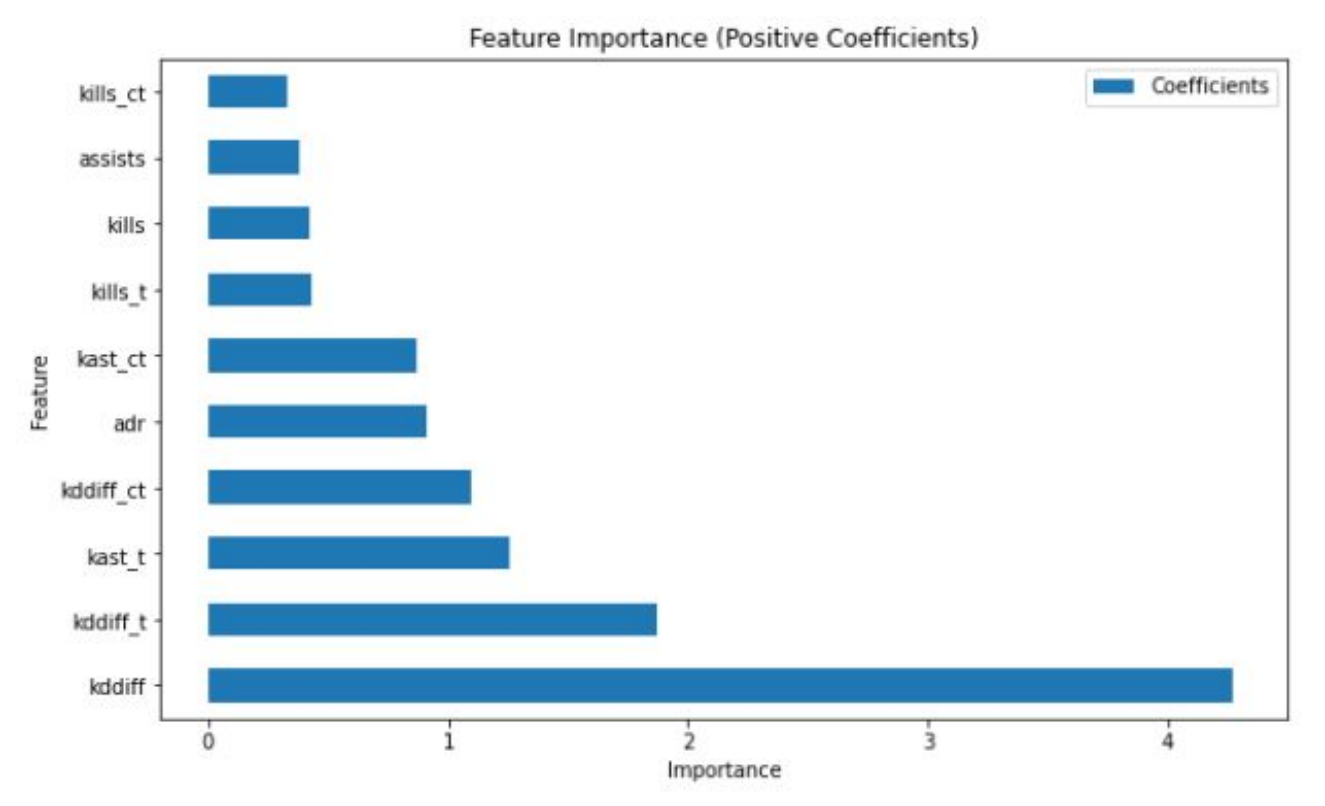
- Logistic Regression
  - F1-Score (Training): 0.945
- Random Forest
  - F1-Score (Training): 0.943
- SVM
  - F1-Score (Training): 0.945

Best Model

- Logistic Regression
  - C: 10.0
  - Penalty: l1
  - Solver: liblinear
  - ROC-AUC Score: 0.988




Feature Importance (Negative Coefficients)
Coefficients
adr_ct
hs
fkdiff
flash_assists
deaths_t
deaths_ct
deaths
adr_t
kast
Feature
−4
−3
−2
−1
0
Importance


Feature Importance (Positive Coefficients)
Coefficients
kills_ct
assists
kills
kills_t
kast_ct
adr
kddiff_ct
kast_t
kddiff_t
kddiff
Feature
0
1
2
3
4
Importance

Future Scope

- Explore more models
- Look at matches other than b01s
- Exploring T and CT differences further
- Map vetos
- Valorant comparisons