Win Predictions in Super Smash Brothers Melee

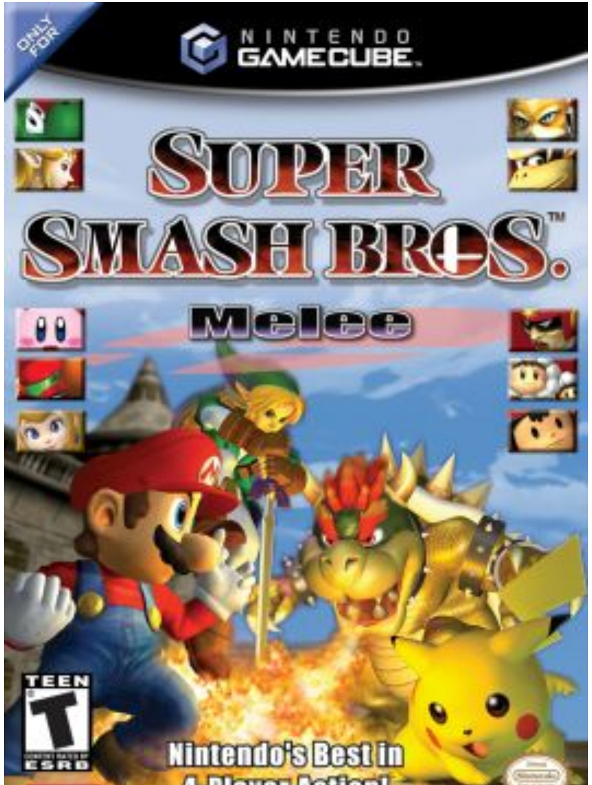
A “Slippi” classification problem

Dakota Nelson

Theo Zinos

Dan Ruppin

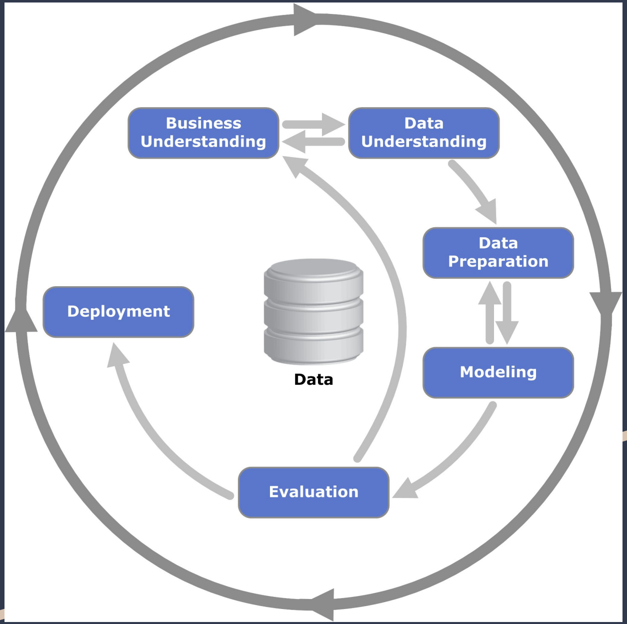
What is Super Smash Bros. Melee?



Super Smash Brothers Melee is a 2001 crossover fighting video game developed by HAL Laboratory and published by Nintendo for the GameCube. It is the second installment in the Super Smash Bros. series

Melee became a highly technical game over the 23 years it has been out. Peaking in tournament play in 2016 and still contains one of the largest audiences for fighting games.

Research question



Can we create a data-driven model to classify matchup outcomes of win or loss with match statistics?

- We will attempt to maximize accuracy

Purpose?

- Learn what matters to get better at the game
- Human error and reaction times.

Melee Game Rules

- Melee is a 2D game in which fighters attempt to knock the other players off the screen

- Percentage increases causes the player to die easier, resets back to 0% per stock

- The game end when one players' life pool (stocks or **lives**) reaches 0.


FPS: 60
0%
0%

Key Assumptions

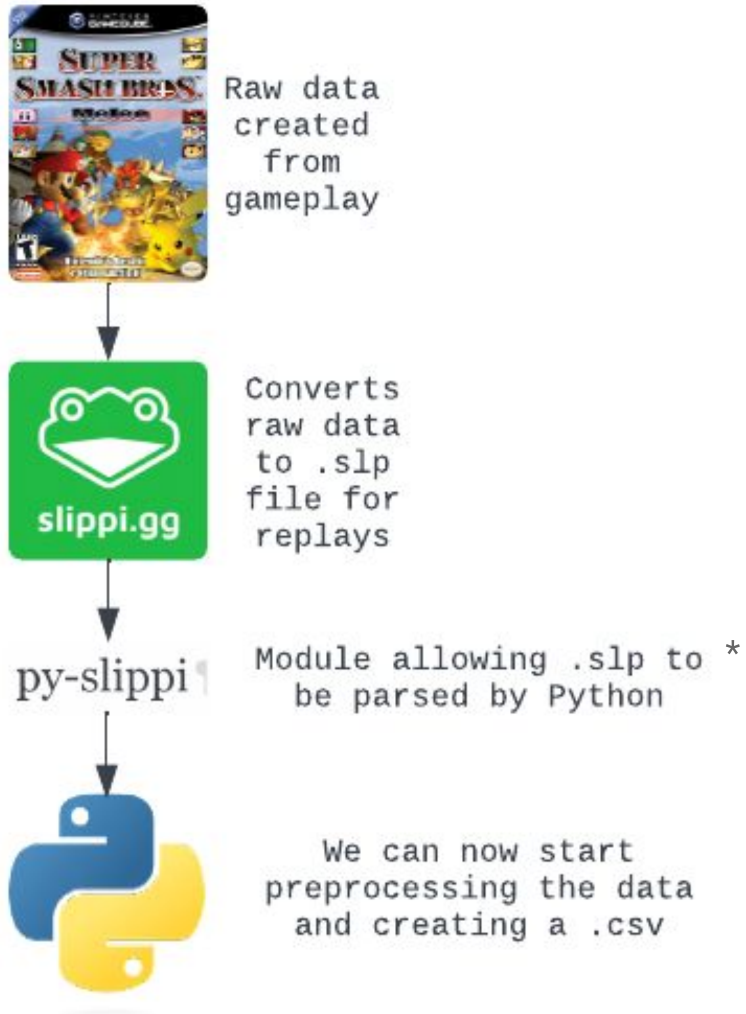
Strong imbalance exists between characters. This is due to characters’ frame data allowing certain characters to combat at much higher speeds than others.

Only 5 maps used in competitive melee. We will stick with this as the data available follows this rule.

Due to certain characters being stronger than others, there will be a much higher frequency of high tier characters than lower tier.

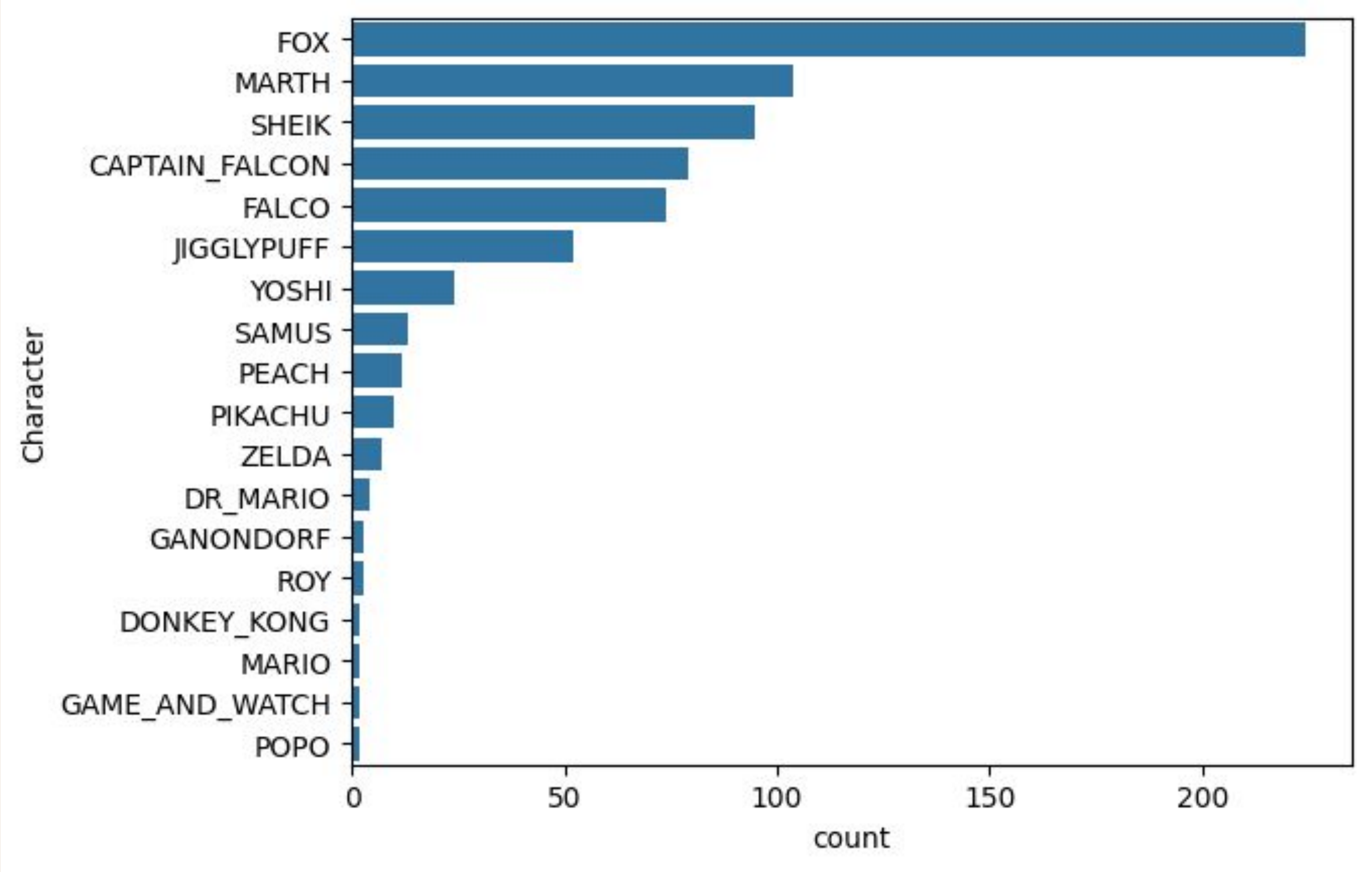
Items are not used in matches

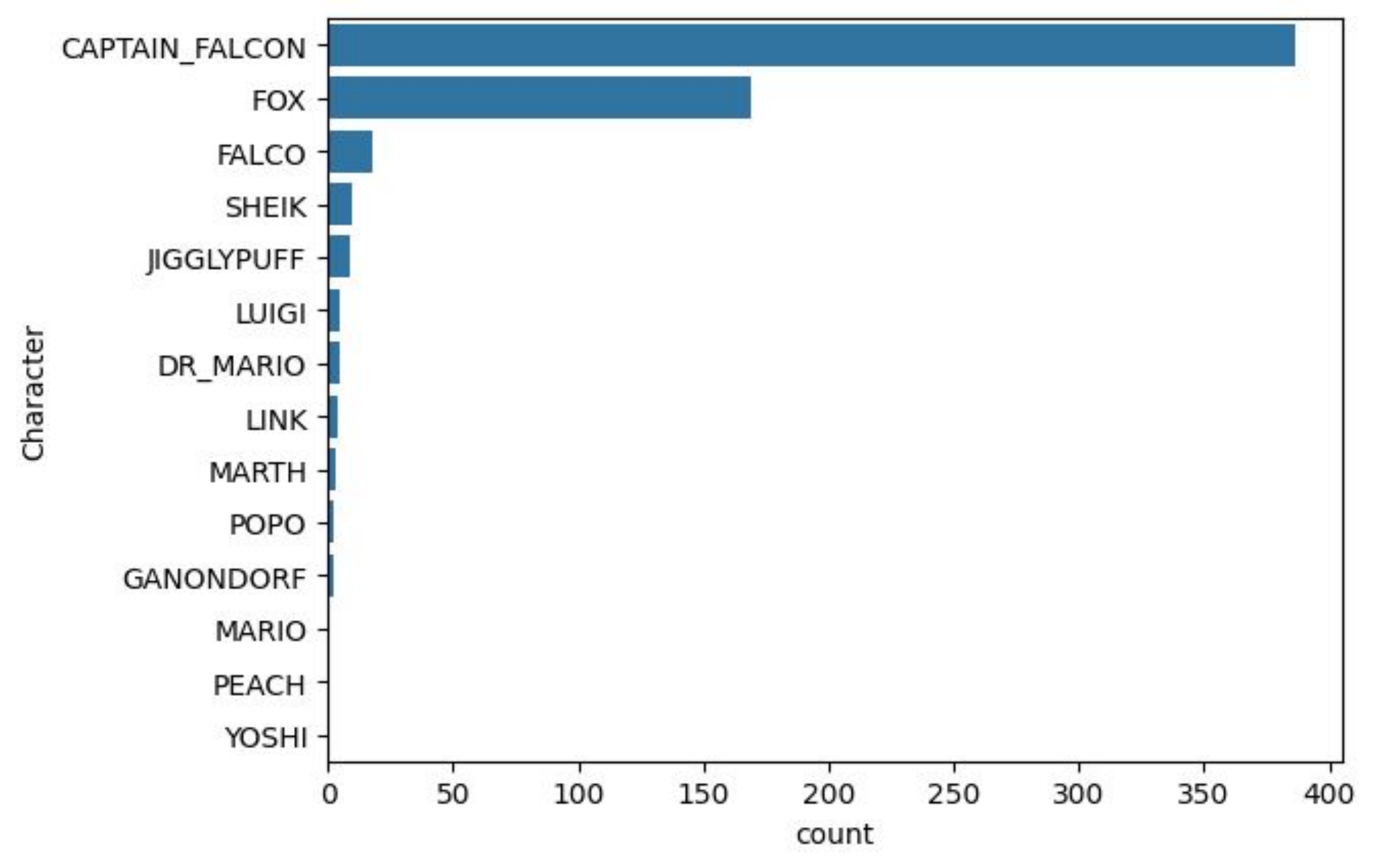
Getting the data



Tournament Match Data

Casual Gamer Data


FOX
MARTH
SHEIK
CAPTAIN_FALCON
FALCO
JIGGLYPUFF
YOSHI
SAMUS
PEACH
PIKACHU
ZELDA
DR_MARIO
GANONDORF
ROY
DONKEY_KONG
MARIO
GAME_AND_WATCH
POPO
Character
0
50
100
150
200
count


CAPTAIN_FALCON
FOX
FALCO
SHEIK
JIGGLYPUFF
LUIGI
DR_MARIO
LINK
MARTH
POPO
GANONDORF
MARIO
PEACH
YOSHI
Character
0
50
100
150
200
250
300
350
400
count

Data Structure

ENTRY	5
ENTRY_START	29
ENTRY_END	30
FALL	154
LANDING	557
TURN	242
DASH	1469
KNEE_BEND	656
JUMP_F	1422
WAIT_4	236
WAIT_ITEM	552
SQUAT_WAIT_1	33
ATTACK_AIR_N	544
LANDING_AIR_N	207

- JSON Format
  - Information about the state of each player in every frame.
  - General information: Date, Stage, End Condition, Mode
- Action States
  - How many matter?

Data Problems

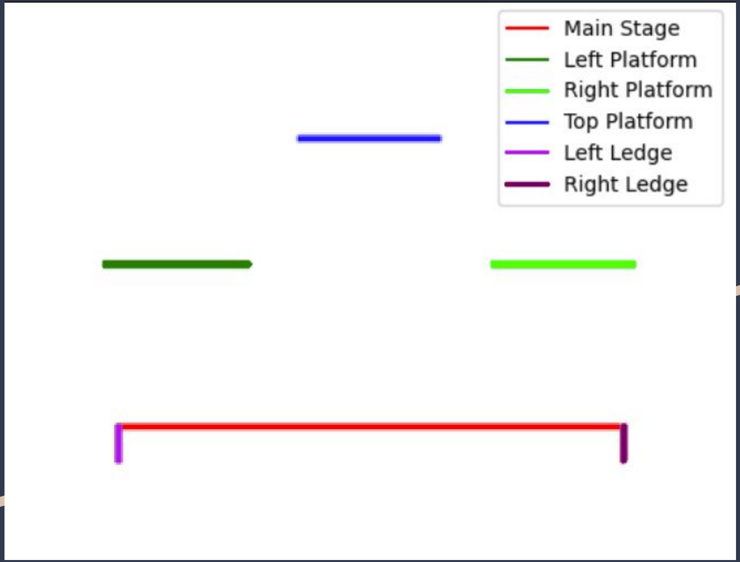
Offense		
Kills	4	2
Damage Done	543.7	308.1
Opening Conversion Rate	68.4% (13 / 19)	31.3% (5 / 16)
Openings / Kill	4.8	8.0
Damage / Opening	28.6	19.3
Defense		
Actions (Roll / Air Dodge / Spot Dodge)	4 / 2 / 5	9 / 1 / 1
Neutral		
Neutral Wins	10 (83%)	2 (17%)
Counter Hits	8 (38%)	13 (62%)
Beneficial Trades	1 (100%)	0 (0%)

- Lacking key information otherwise shown directly in Slippi
- Incredibly difficult to turn into something useable
- “Illegal” games in the datasets
- Character Generalization
- Action Relation

Problem Solutions

- All characters have the same inputs, but the actions performed are different
- Restricting to available general data
- Removing illegal games
    - CPU, Multiplayer, Early End
- Remove direct correlation

Variables we chose manually

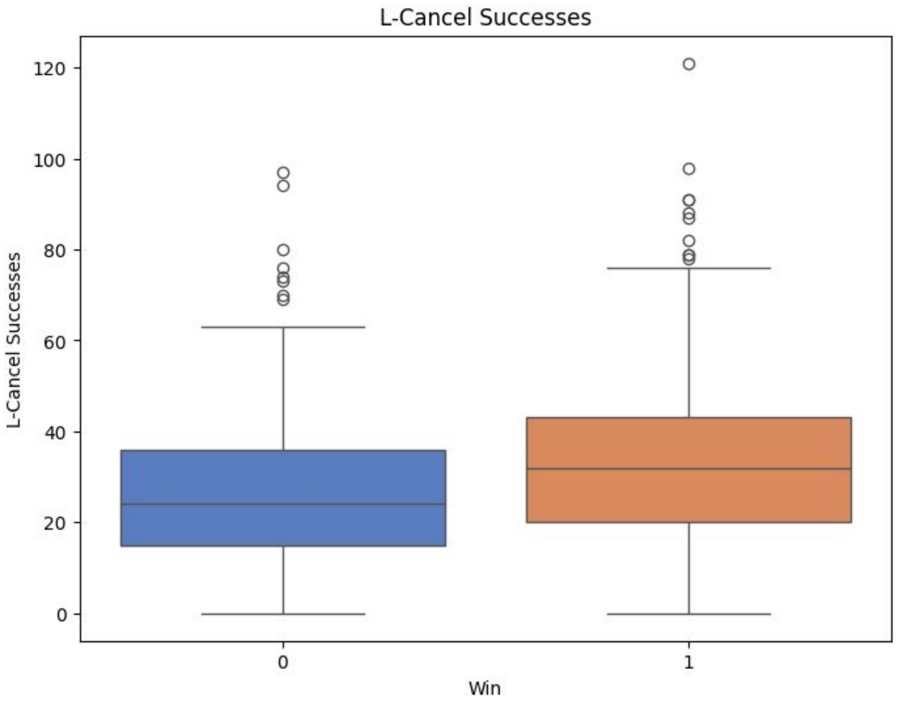


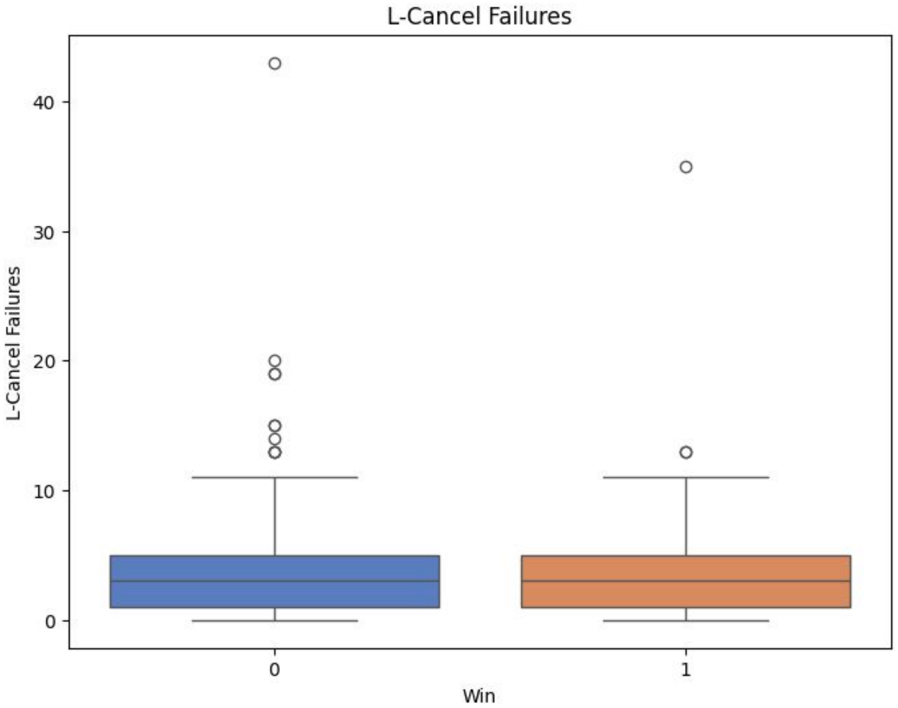
- Win/Loss Dependent Variable
- Character
- Stage
- Max Combo
- Mean X-Distance from center
- Most frequent platform
- L-Cancel Successes*
- L-Cancel Failures*

L-Cancel???

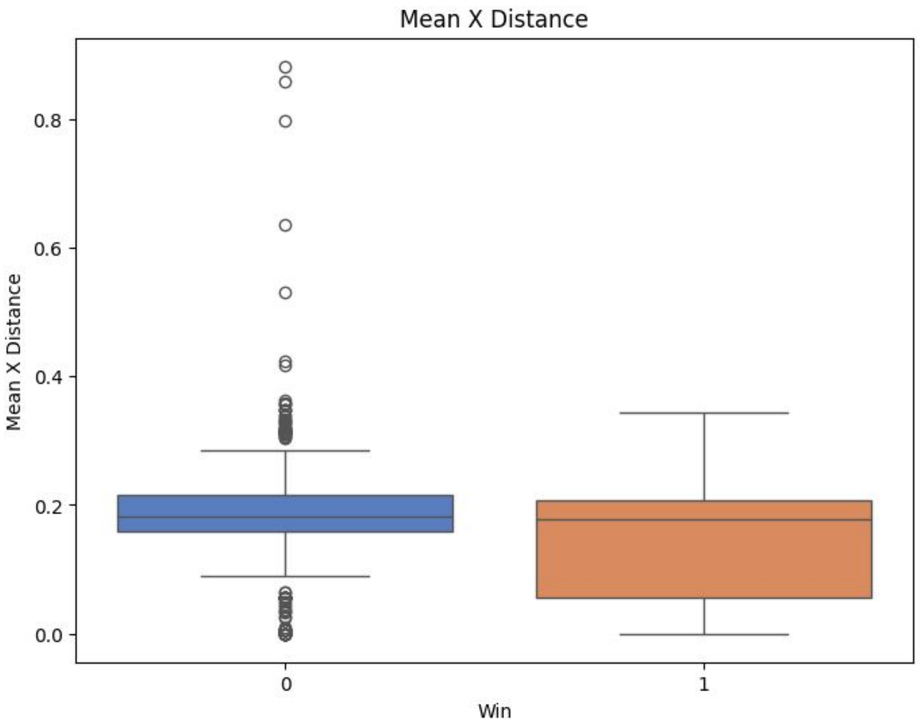
- Short for Land Cancelling
- Uses precise inputs to cancel the animation
- Lets skilled players play faster
- No “consequence” of failure
    - Extremely punishable

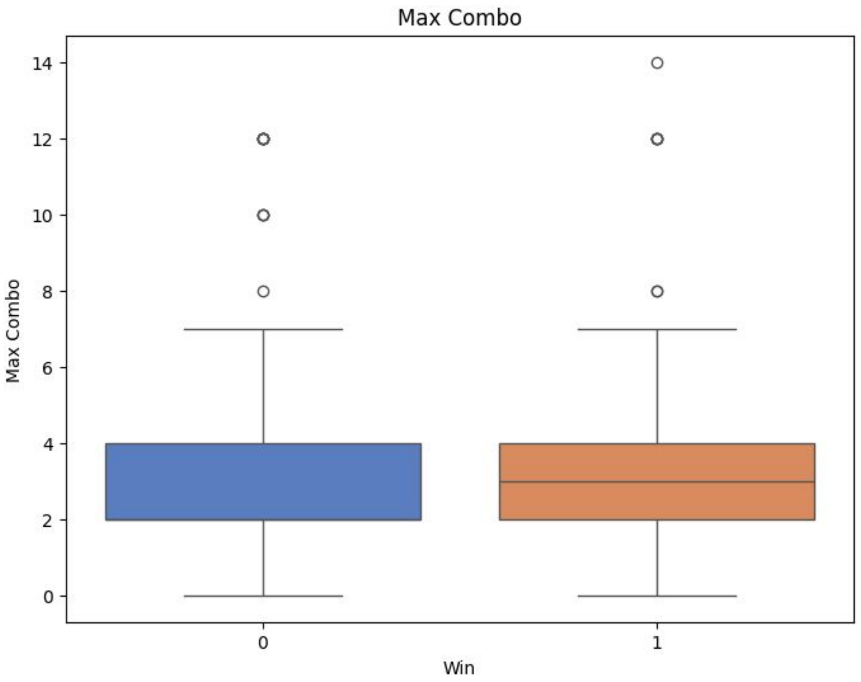
Data Visualizations (Professional)



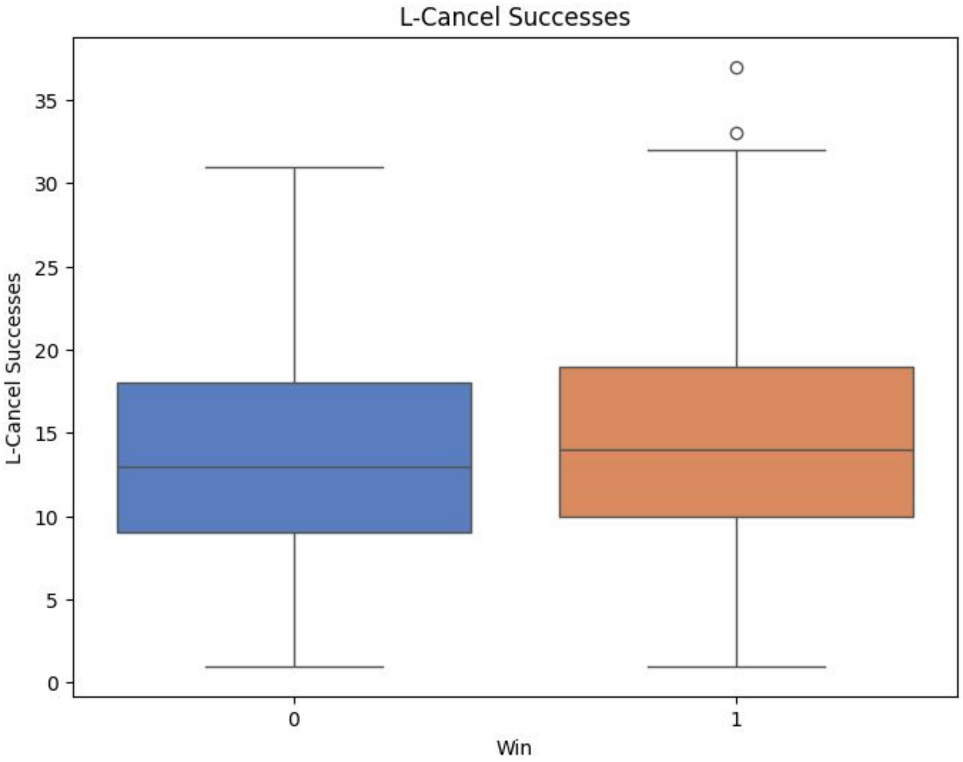


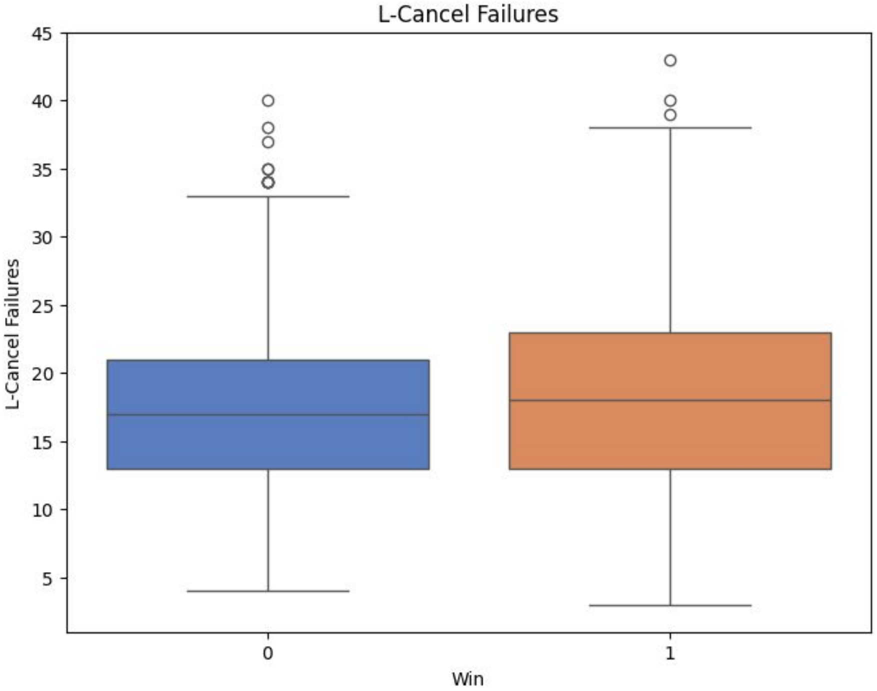
Data Visualizations (Professional)



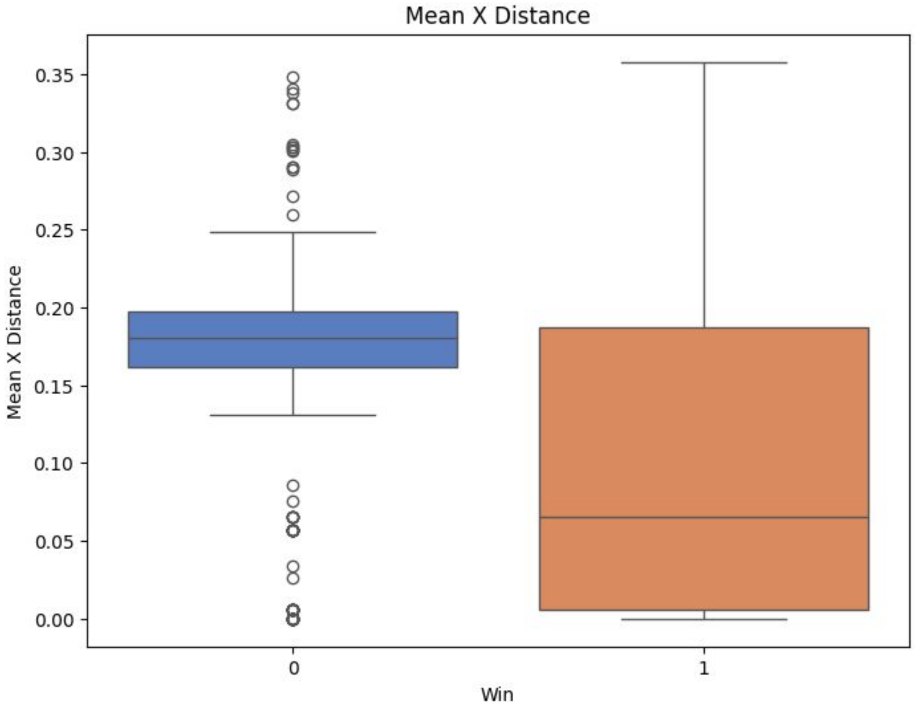


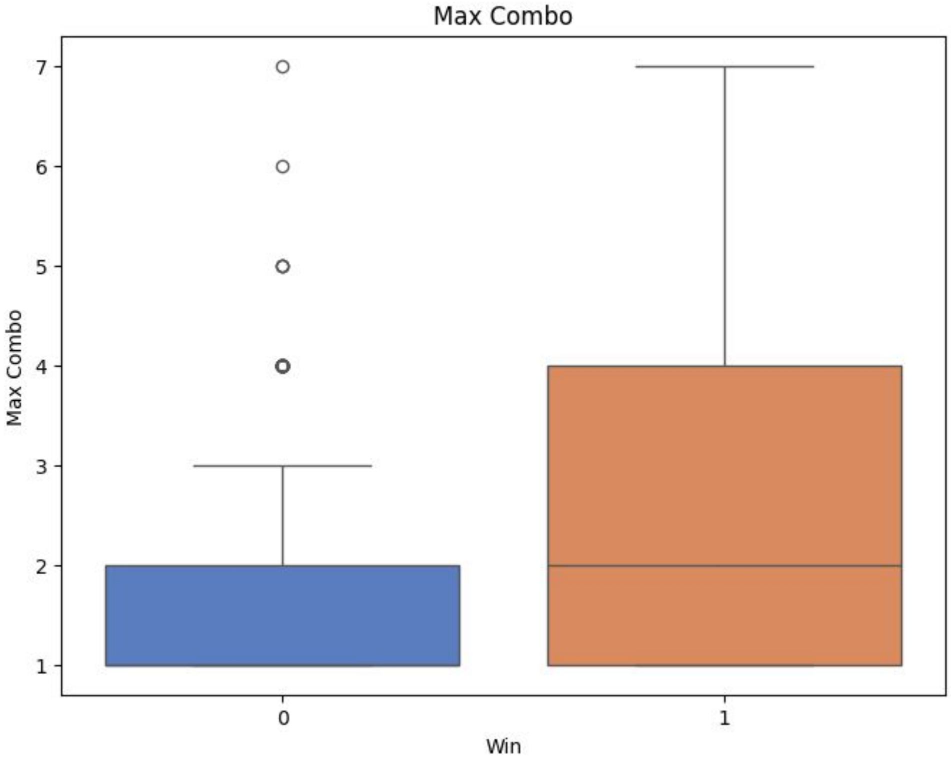
Data Visualizations (Casual)



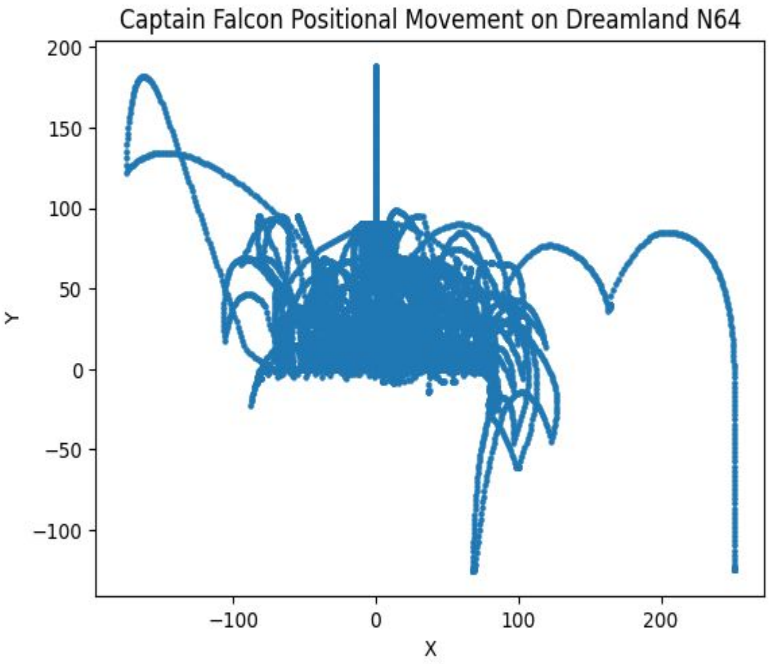


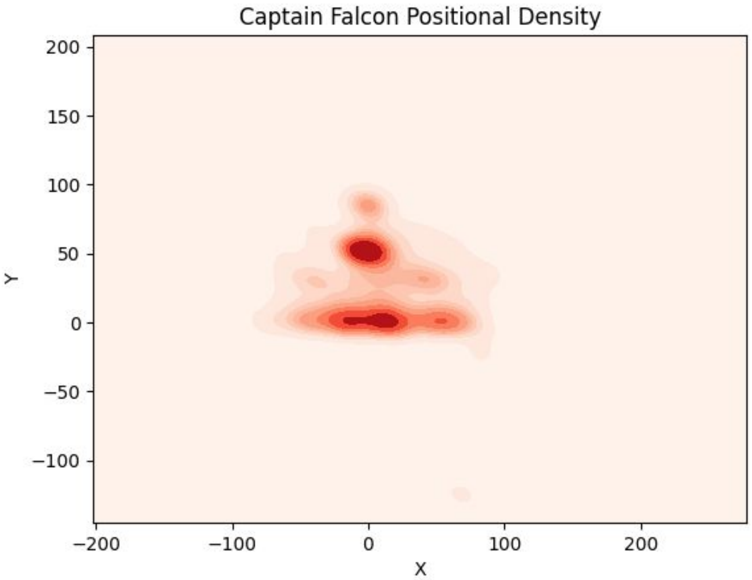
Data Visualizations (Casual)





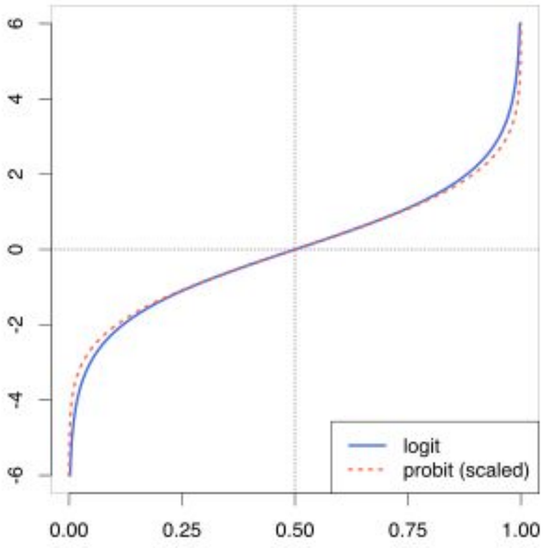
The players position



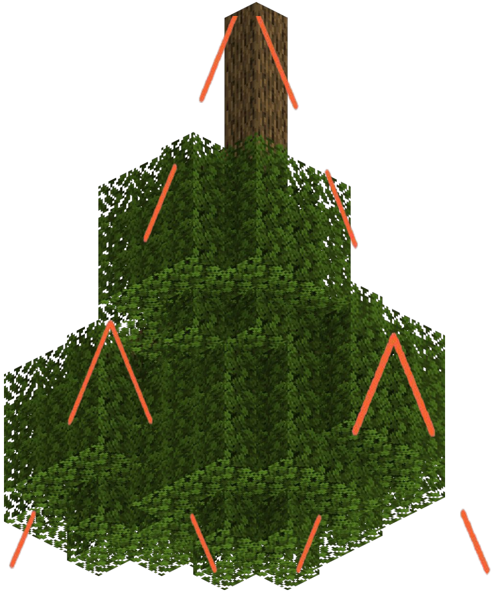


The models we used

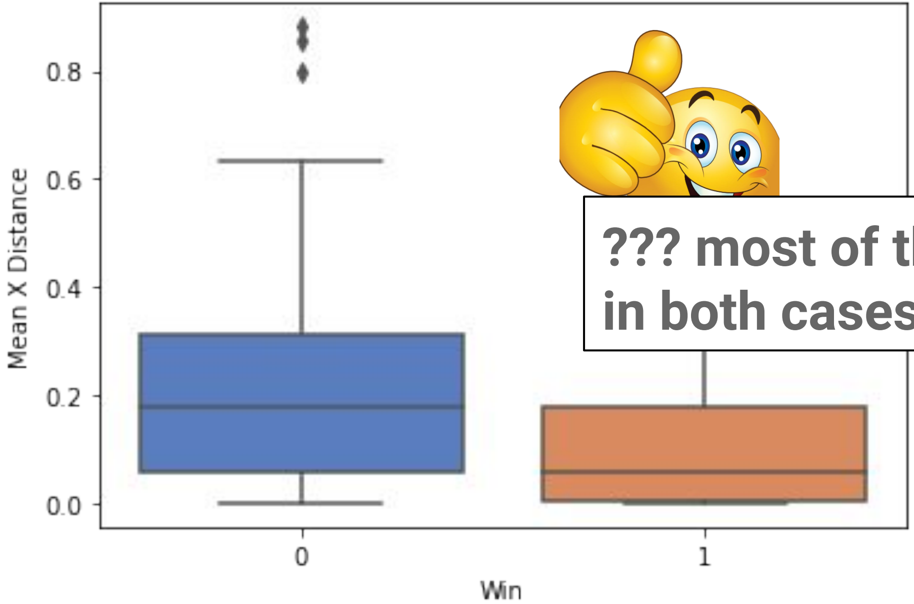
Logit

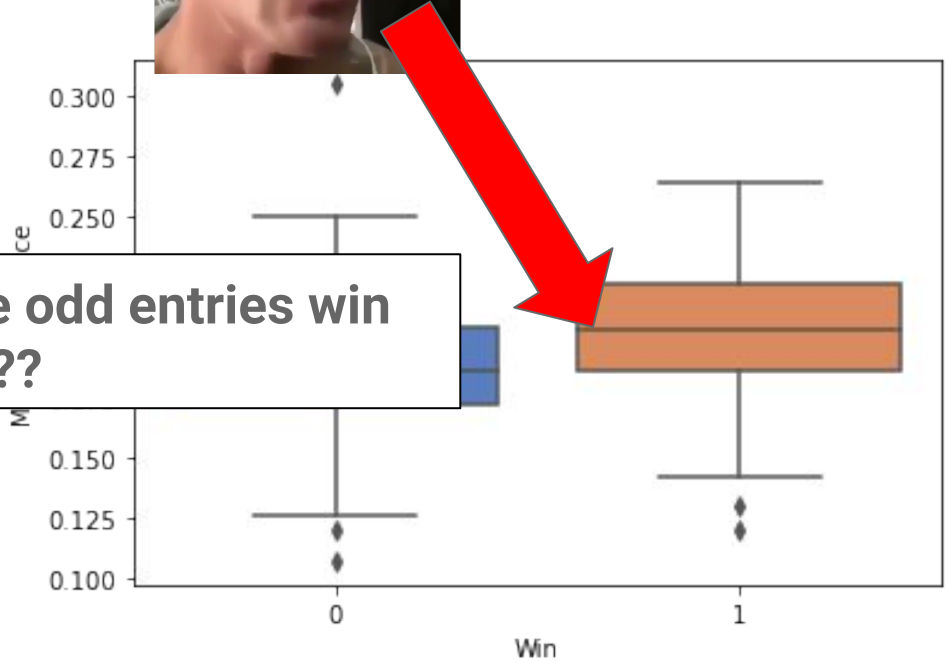


Random Forest



Logit – Initial Data Considerations

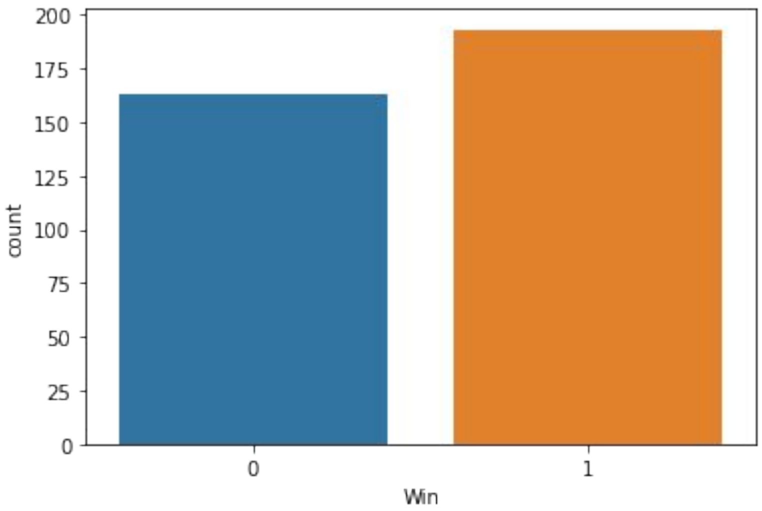




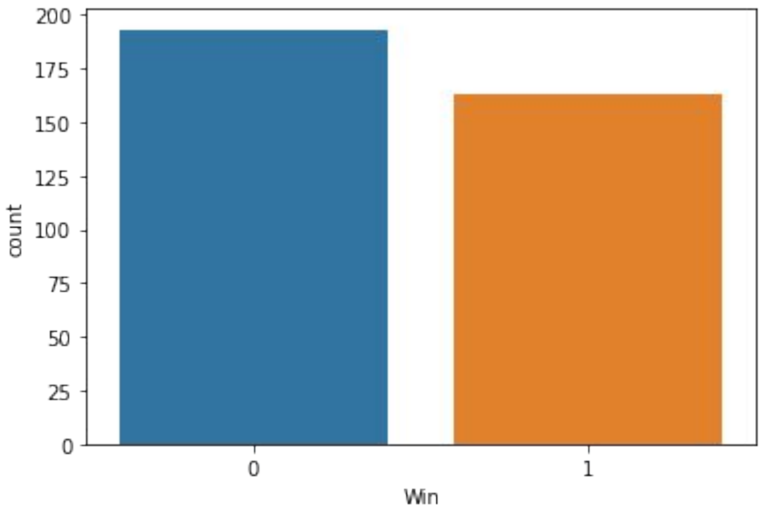
??? most of the odd entries win in both cases???

Logit – Even VS Odd data split

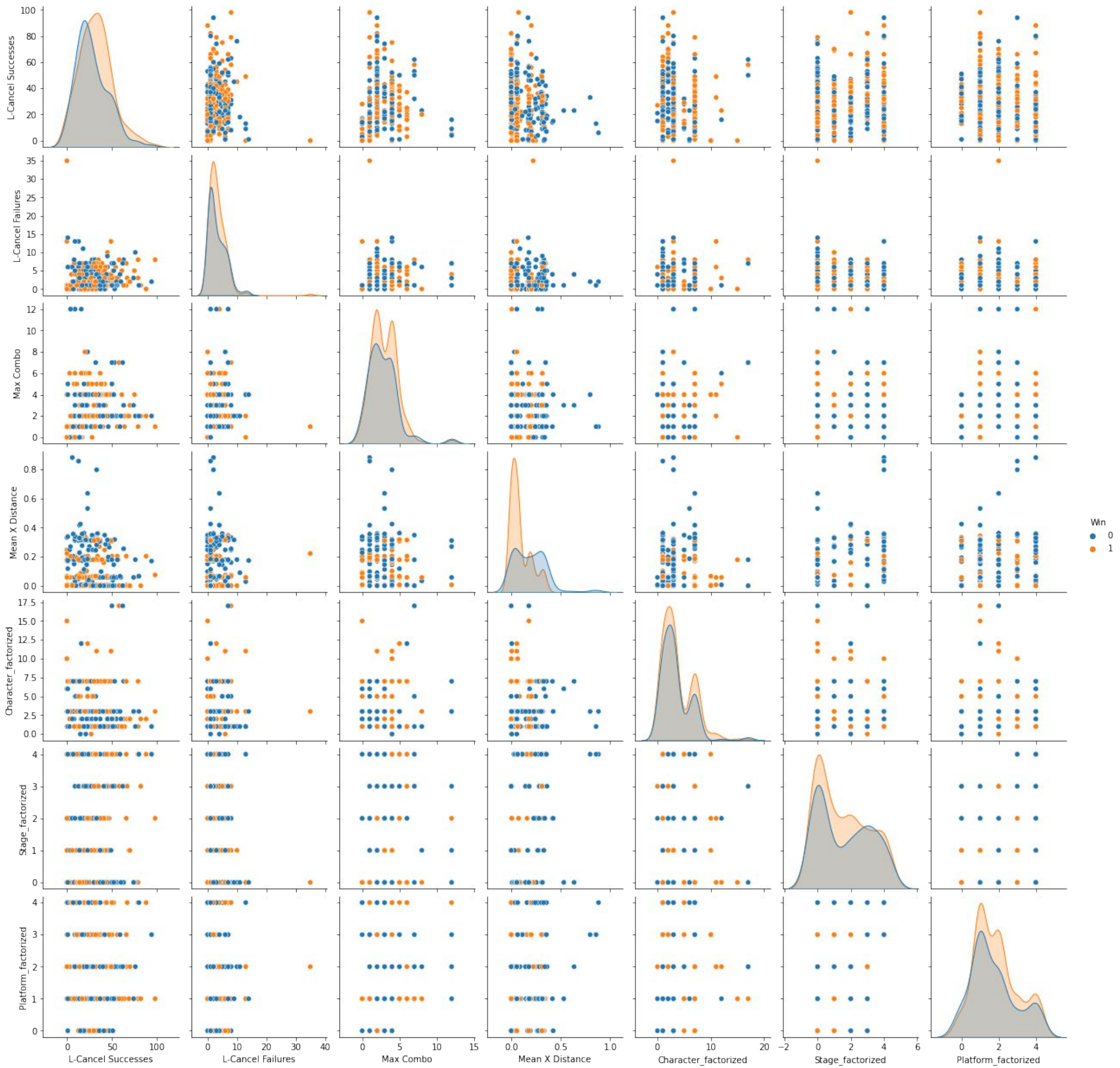
Odd



Even



Considering the data



Hypothesis Tests with linregress

Max Combo and Mean X Distance were significant in the casual set

	p-value	test_statistic	statistically_significant
L-Cancel Successes	0.036851	0.115138	True
L-Cancel Failures	0.942368	0.004001	False
Max Combo	0.461867	0.040703	False
Mean X Distance	0.0	-0.358753	True
Character_factorized	0.327417	0.054159	False
Stage_factorized	0.579231	-0.030681	False
Platform_factorized	0.370227	0.049559	False

Outcome of Logit models trained only on significant features

Tournament set

- Average Accuracy: 0.6717

Casual set

- Average Accuracy: 0.6819

Outcome of Logit models using greedy feature selection

Tournament Set

['constant', 'Platform', 'Character', 'Mean X Distance', 'L-Cancel Successes', 'Max Combo']. (no L-Cancel Failures)

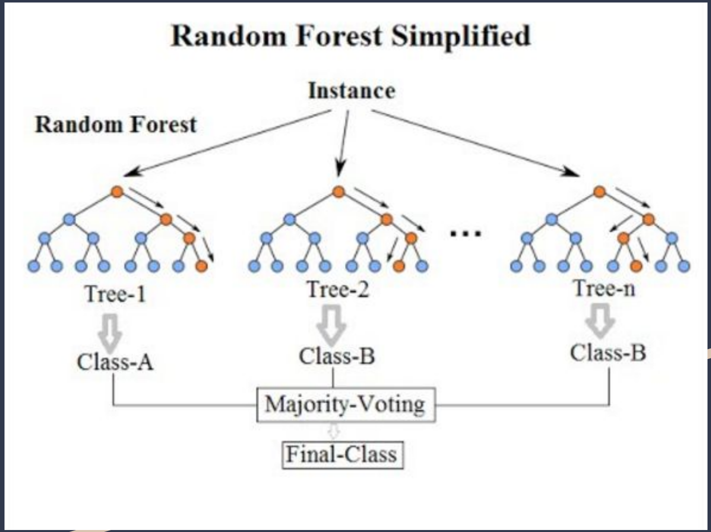
Average Accuracy: 0.6748

Casual set

['constant', 'Platform', 'Stage', 'Character', 'Max Combo']

Average Accuracy: 0.6983

Random Forest

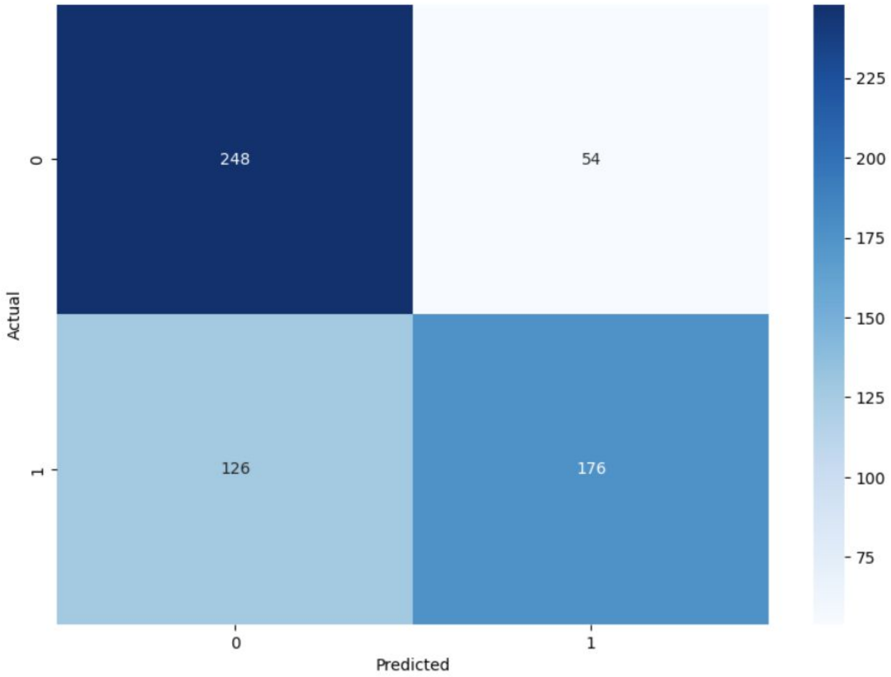


First we one-hot encoded the three categorical features [Character, Stage] as these were nominal data (no ordered significance)

Second we used RandomizedSearch with a 5-stratifiedKfold Hyperparameter tuning, giving the following parameters for the model:

```
RandomForestClassifier
RandomForestClassifier(criterion='entropy', max_depth=17, max_features='log2',
                       max_leaf_nodes=47, min_samples_leaf=6,
                       min_samples_split=20, n_estimators=30)
```

Outcome of Random Forest model



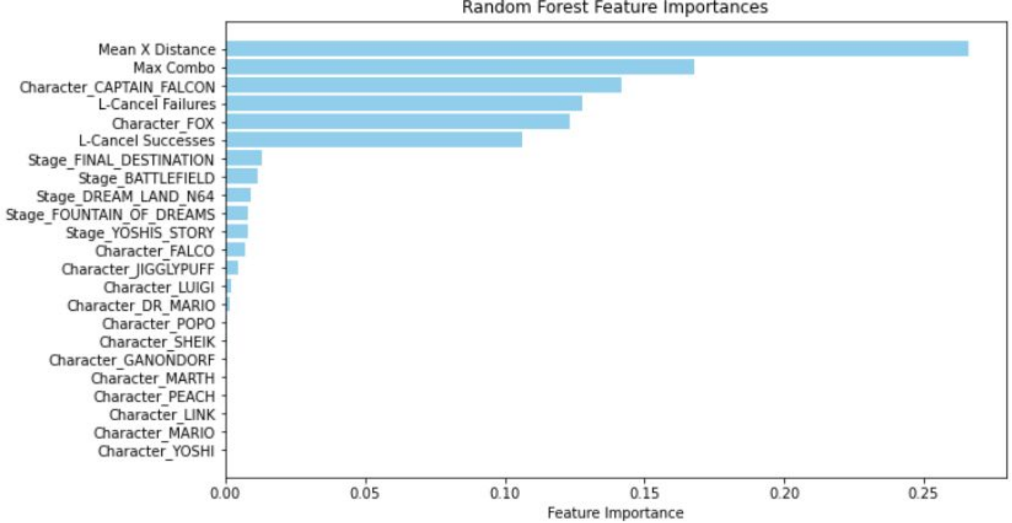
Tournament Set:

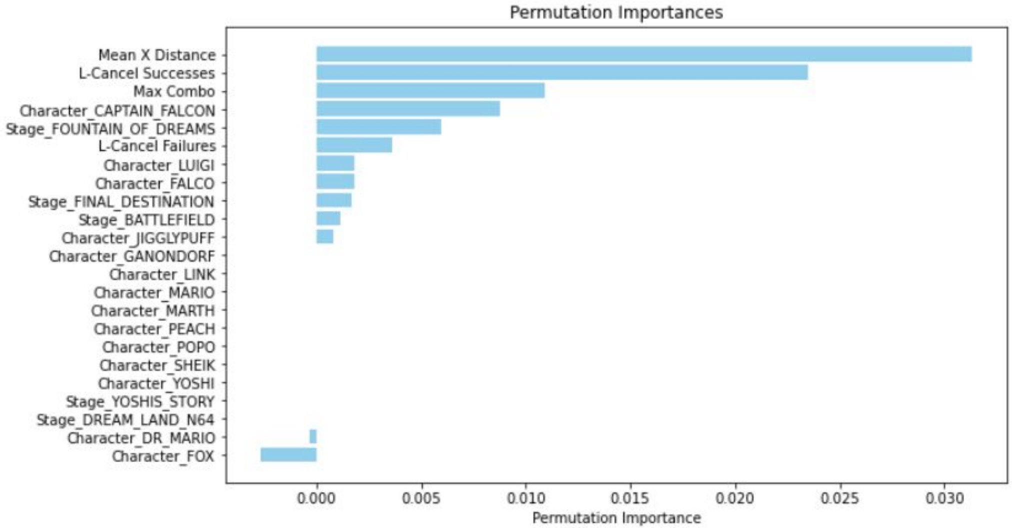
Average Accuracy: 0.6733

Casual Set:

Average Accuracy: 0.7003

Casual Data Set





Going forward

- Pulling data directly from Slippi would allow for more valuable data and features
- Slippi live feed predictions
- Use of a DNN could result in more complex relationship creation.

Biggest Takeaways

Logit model

-Simple

-Good accuracy

-Greedy model performed better than the Random Forest in the tournament set

Random Forest

-Good accuracy

-Good insight to features

-Room for improvement