



DOTA 2 HERO RECOMMENDER

BY: ZACKARIAS CHIA



WHAT IS DOTA 2?



Hero Selection Process, AKA Drafting

Teams take turns to draft or ban heroes in a precise predetermined order depending on the game mode. If a hero is banned by one team, this hero cannot be picked by either team

A Multiplayer Online Battle Arena (MOBA) game

Each player controls a single character with a set of distinctive abilities that improves over the course of the game. Team wins by being the first to destroy the other team's 'ancient'.

TEAM LIQUID PICKS AND BANS											
	2		4		6		8		9		12
✗ BAN		✗ BAN		✗ BAN		✓ PICK		✓ PICK		✗ BAN	
	14		15		17		19		22		
✓ PICK		✓ PICK		✓ PICK		✗ BAN		✓ PICK		✗ BAN	
OG PICKS AND BANS											
	1		3		5		7		10		11
✗ BAN		✗ BAN		✗ BAN		✓ PICK		✓ PICK		✗ BAN	
	13		16		18		20		21		
✓ PICK		✓ PICK		✓ PICK		✗ BAN		✓ PICK		✗ BAN	



OBJECTIVE

Create a predictor model to determine match wins based on hero selection only in the MOBA game Dota 2.

PROBLEM STATEMENT



Esports is increasingly being recognised as a legitimate sport

- In 2017, the IOC recognised the growing popularity of esports.
- Esports are played in other international sporting tournaments such as SEA games



Dota 2 is one of the most competitive esports

- The International (TI) is the largest world championship tournament for Dota 2, with prize pools of US\$40M in 2021

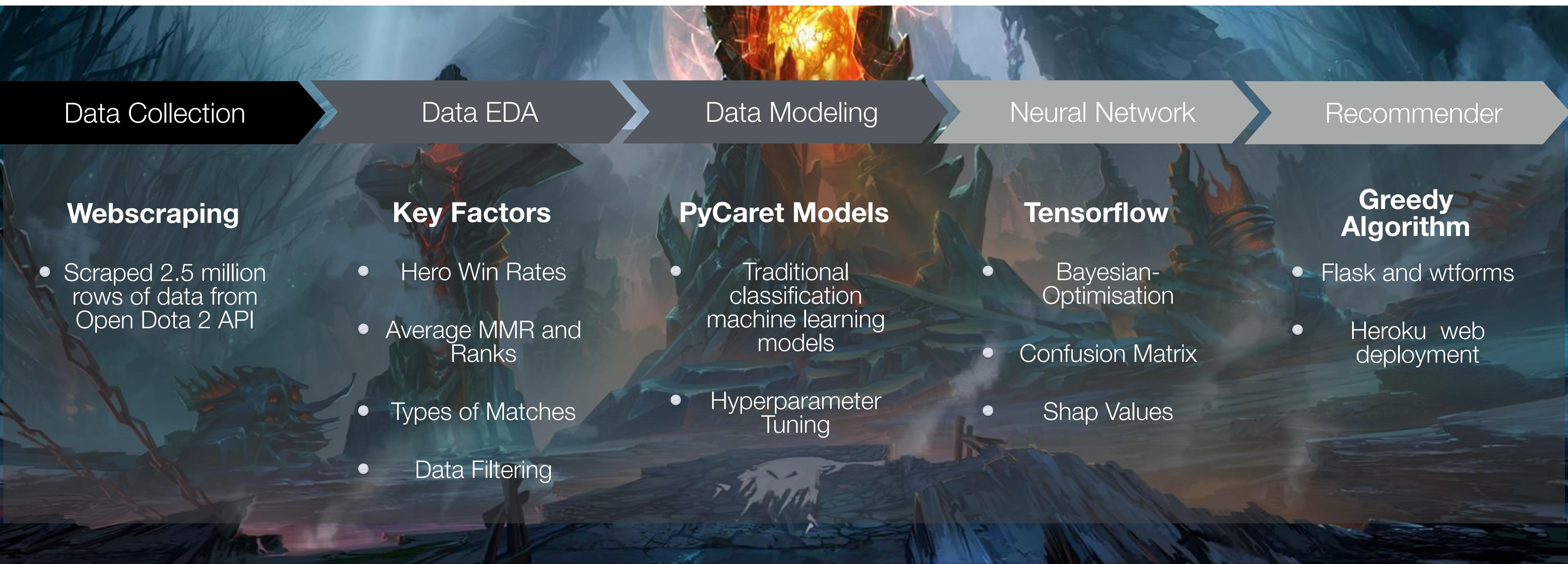


Tool to improve win rates

- Given the competitive nature, a tool to help predict optimal hero composition will be useful for teams as it can give a team an implicit advantage before even the match begins



METHODOLOGY



Data Collection

Data EDA

Data Modeling

Neural Network

Recommender

Webscraping

- Scraped 2.5 million rows of data from Open Dota 2 API

Key Factors

- Hero Win Rates
- Average MMR and Ranks
- Types of Matches
- Data Filtering

PyCaret Models

- Traditional classification machine learning models
- Hyperparameter Tuning

Tensorflow

- Bayesian-Optimisation
- Confusion Matrix
- Shap Values

Greedy Algorithm

- Flask and wtforms
- Heroku web deployment



DATA COLLECTION



WEBSRAPING

Pull ~2.5m matches from Open Dota 2 API. Parse match data from Patch 7.3.1d





DATA EDA



MATCH DURATION



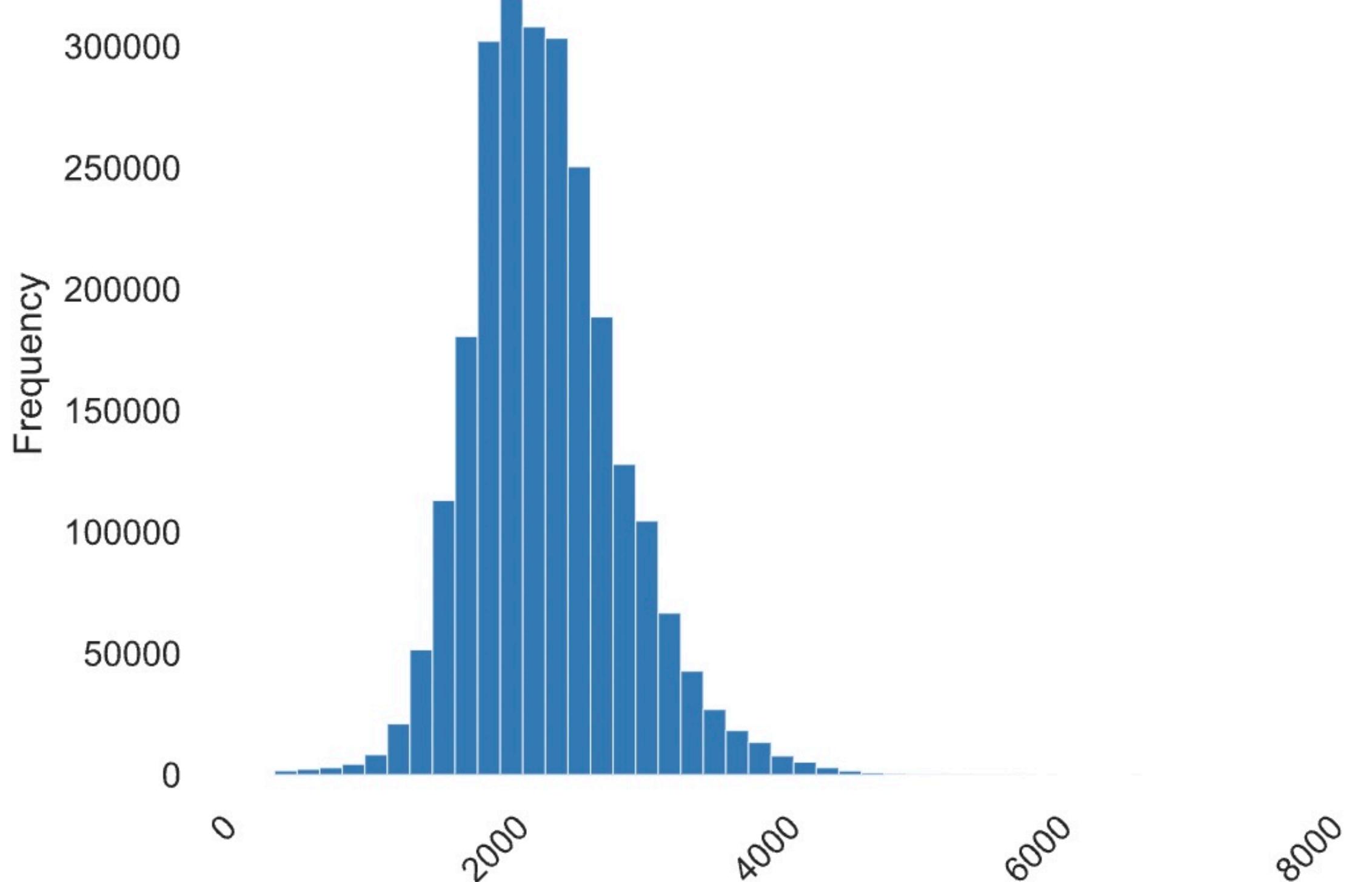
Median Duration -
2311 seconds
(38.5mins)

However, there are some games that are very short (minimum 361 seconds).

Unlikely that these are full matches

Quantile statistics

Minimum	361
5-th percentile	1585
Q1	1987
median	2311
Q3	2681
95-th percentile	3345
Maximum	8677
Range	8316
Interquartile range (IQR)	694



Histogram with fixed size bins (bins=50)

AVERAGE MMR

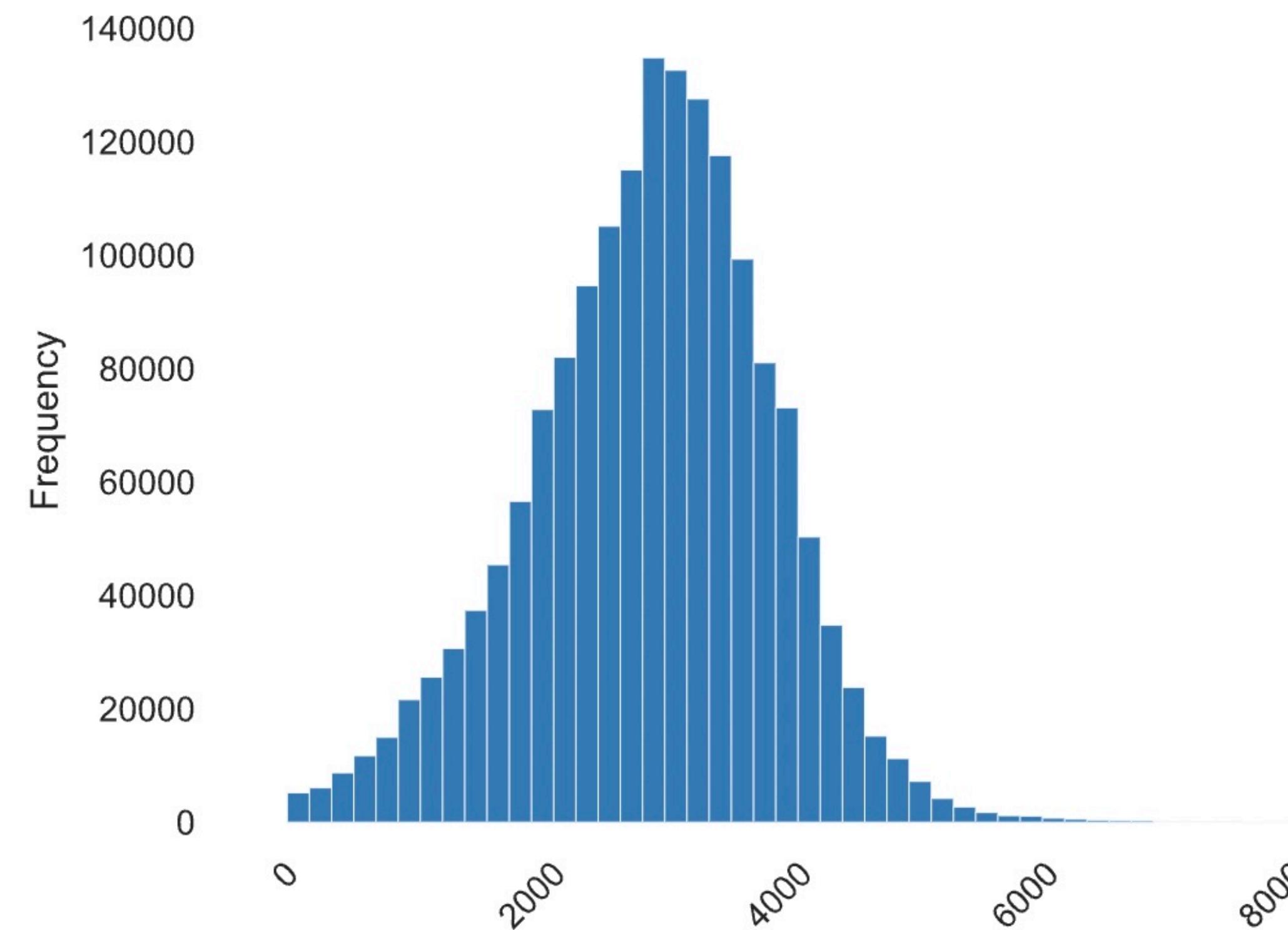
✓ Median MMR of 3012

Most people are not that good at the game.
3000 MMR is considered casual. For context - pro players are above 7k MMR

However, 33% of values missing - mainly as displaying public MMR is a personal choice

Quantile statistics

Minimum	1
5-th percentile	1188
Q1	2333
median	3012
Q3	3595
95-th percentile	4435
Maximum	9005
Range	9004
Interquartile range (IQR)	1262



Histogram with fixed size bins (bins=50)



AVERAGE RANK TIER

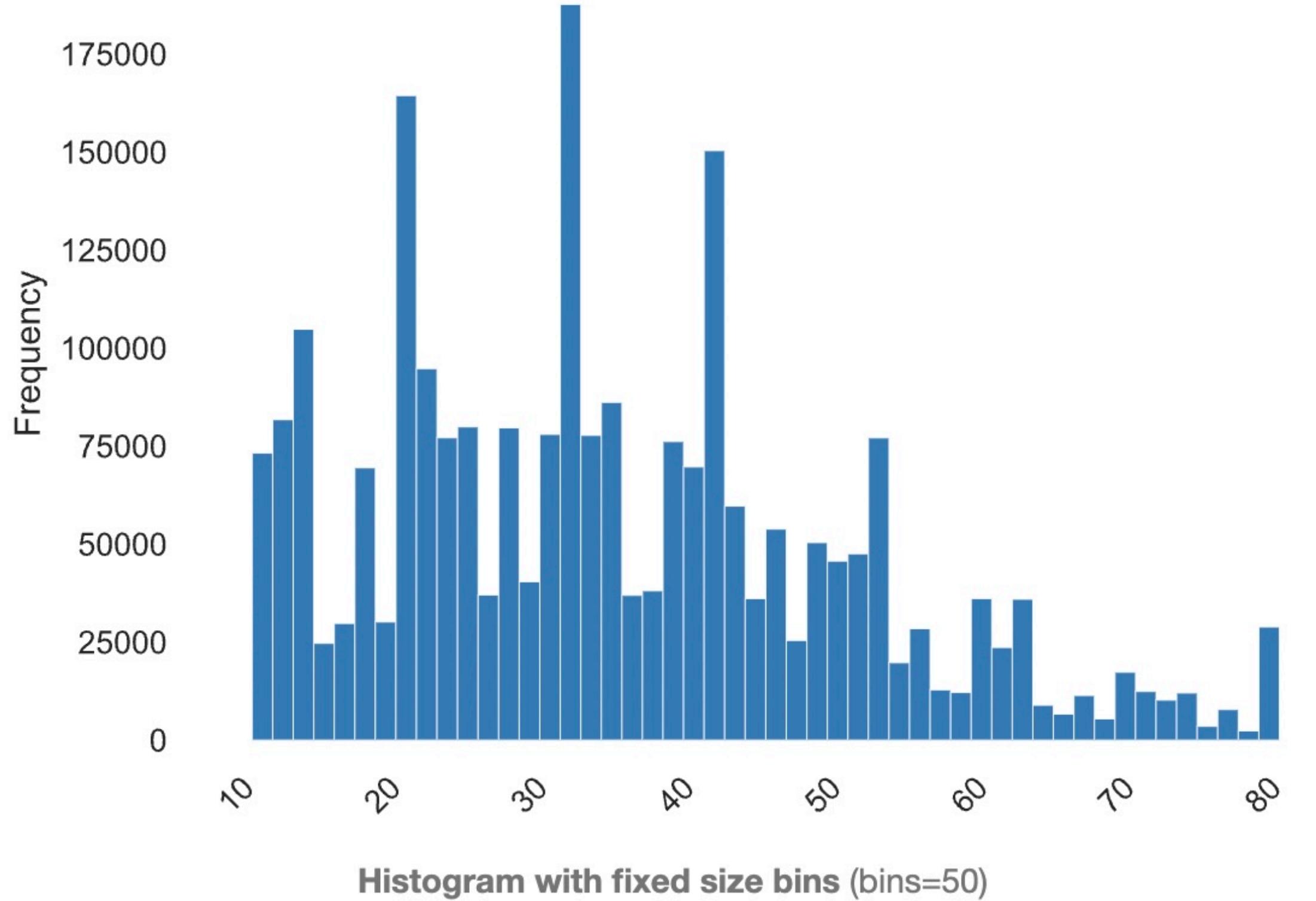


Median value of 33, similar to MMR

81 is immortal rank - which is the highest rank you can climb

Quantile statistics

Minimum	11
5-th percentile	13
Q1	23
median	33
Q3	45
95-th percentile	65
Maximum	81
Range	70
Interquartile range (IQR)	22

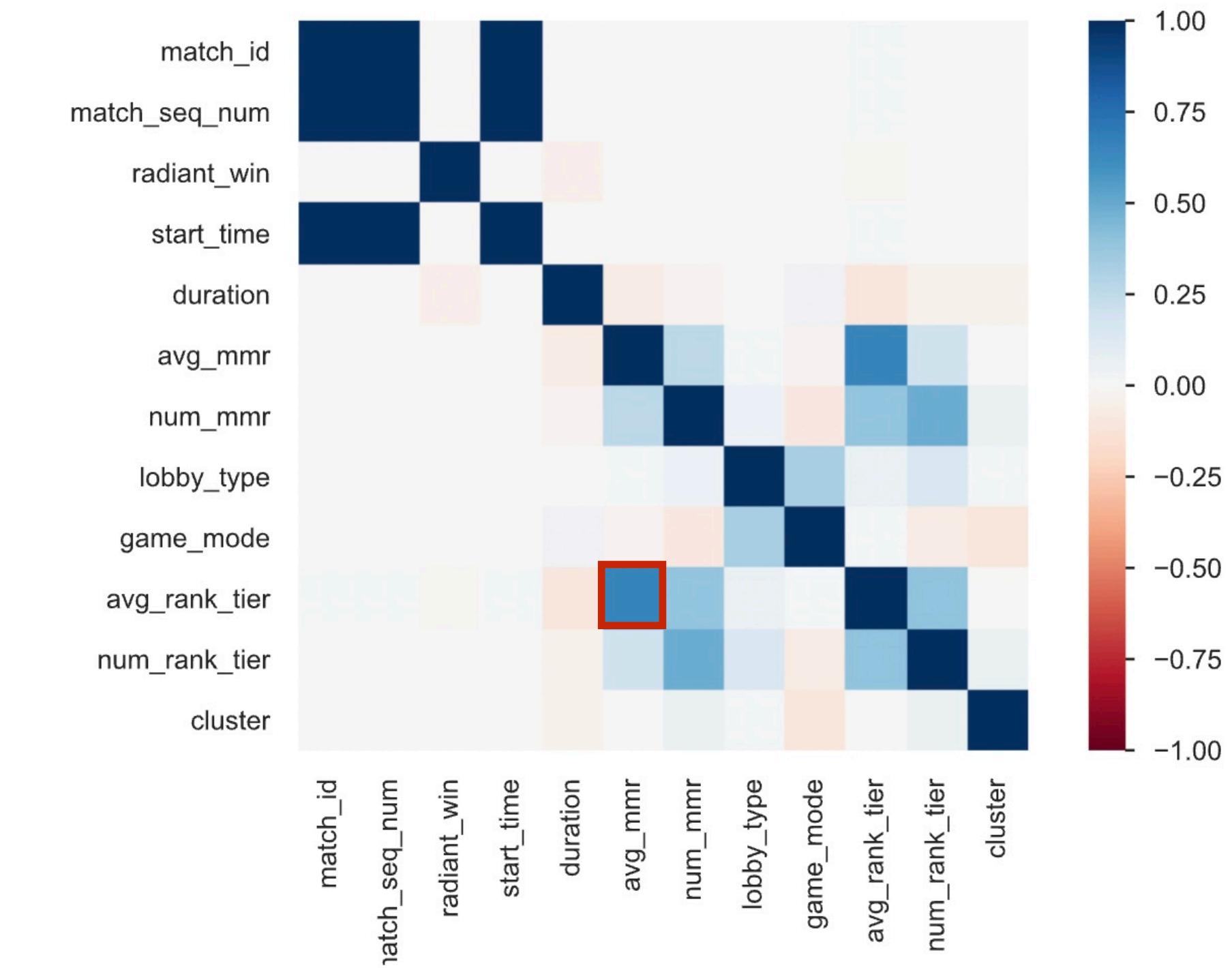
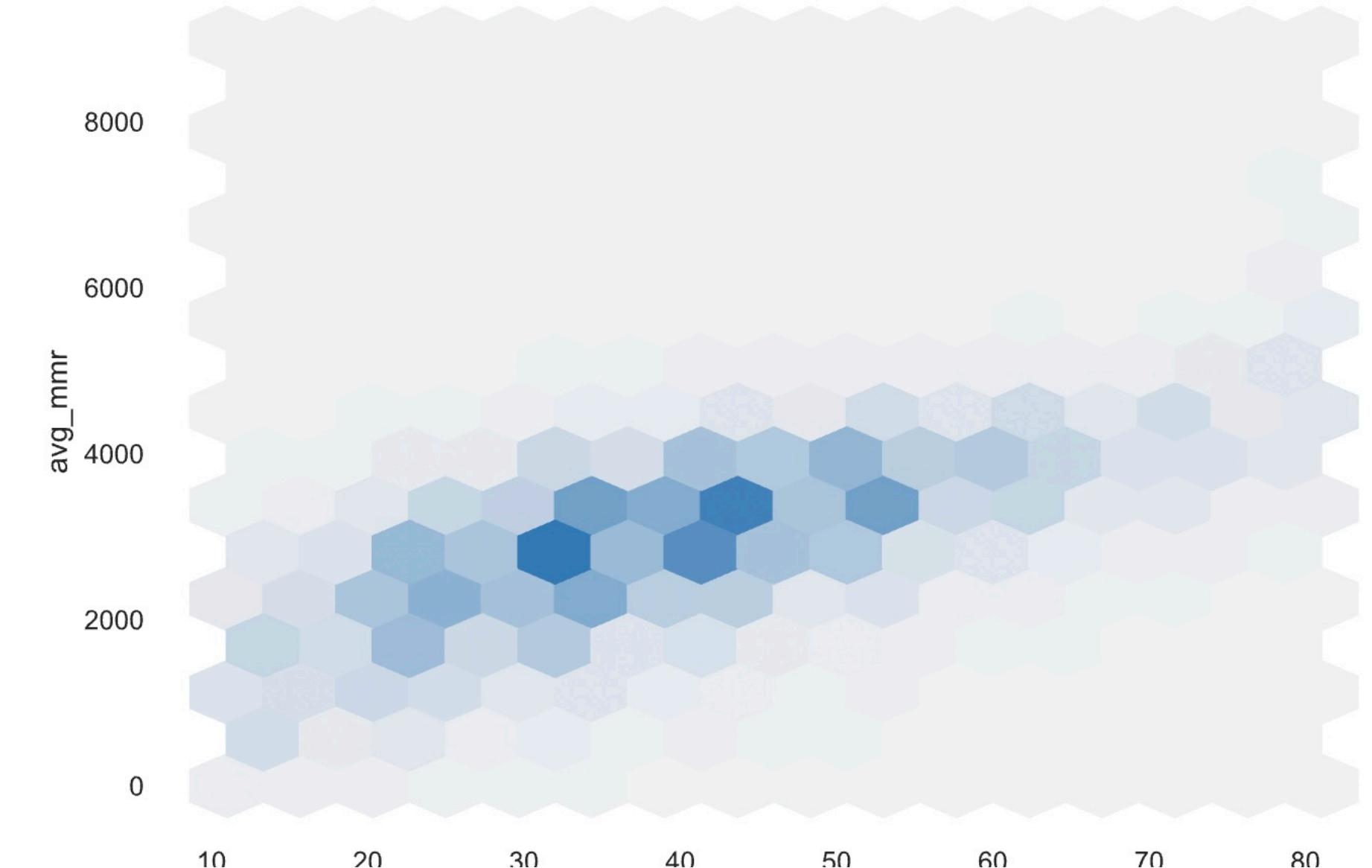




AVG MMR VS AVG RANK TIER



High Correlation
between Average MMR
and Average Rank Tier





LOBBY TYPE



Most matches are played in Ranked Mode

Lobby Type 7 refers to Ranked Matches.

Lobby Type 0 refers to non-ranked matches

Common Values

Value	Count	Frequency (%)
7	2032326	82.0%
0	442834	17.9%
1	1930	0.1%
9	1810	0.1%

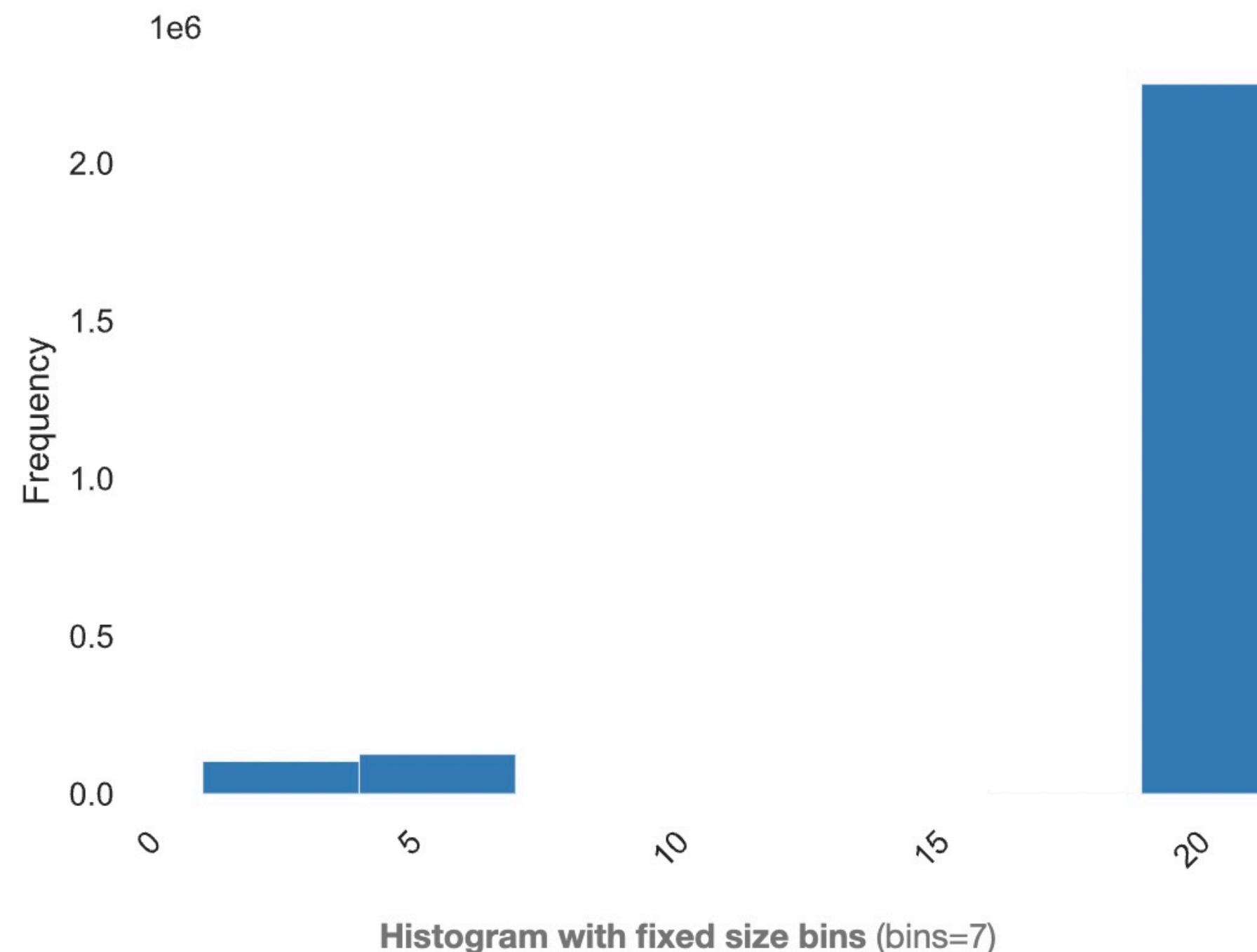


GAME MODE



Most are Game Mode
22

Game Mode 22
corresponds to the “All
Pick” Mode

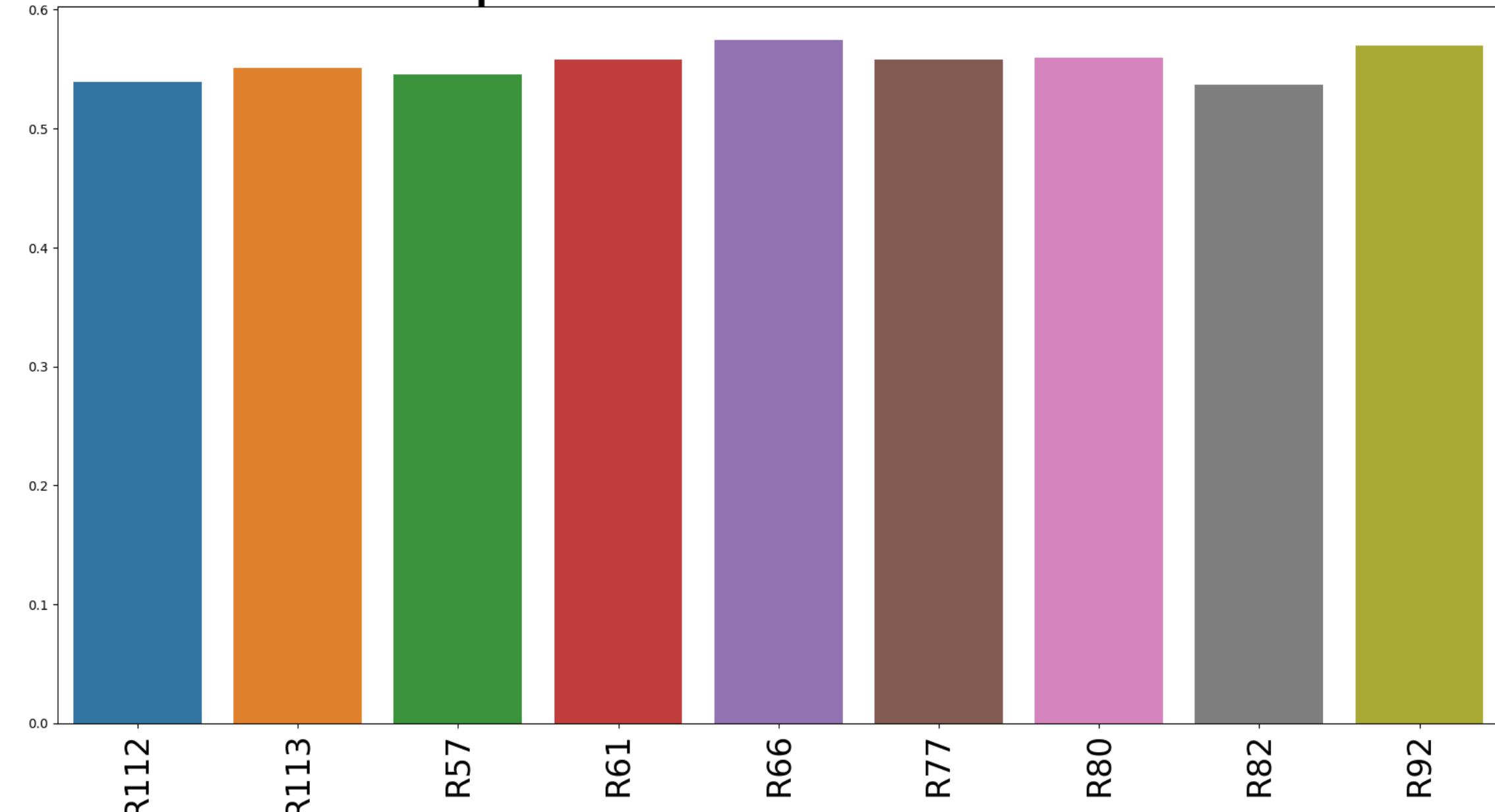


Ranked All Pick picking order

Radiant	Radiant	Radiant	Radiant	Radiant
Dire	Dire	Dire	Dire	Dire

Value	Count	Frequency (%)
22	2250651	90.8%
4	121509	4.9%
3	98666	4.0%
2	4235	0.2%
5	3656	0.1%
16	169	< 0.1%
1	14	< 0.1%

Top Winrates of Heroes

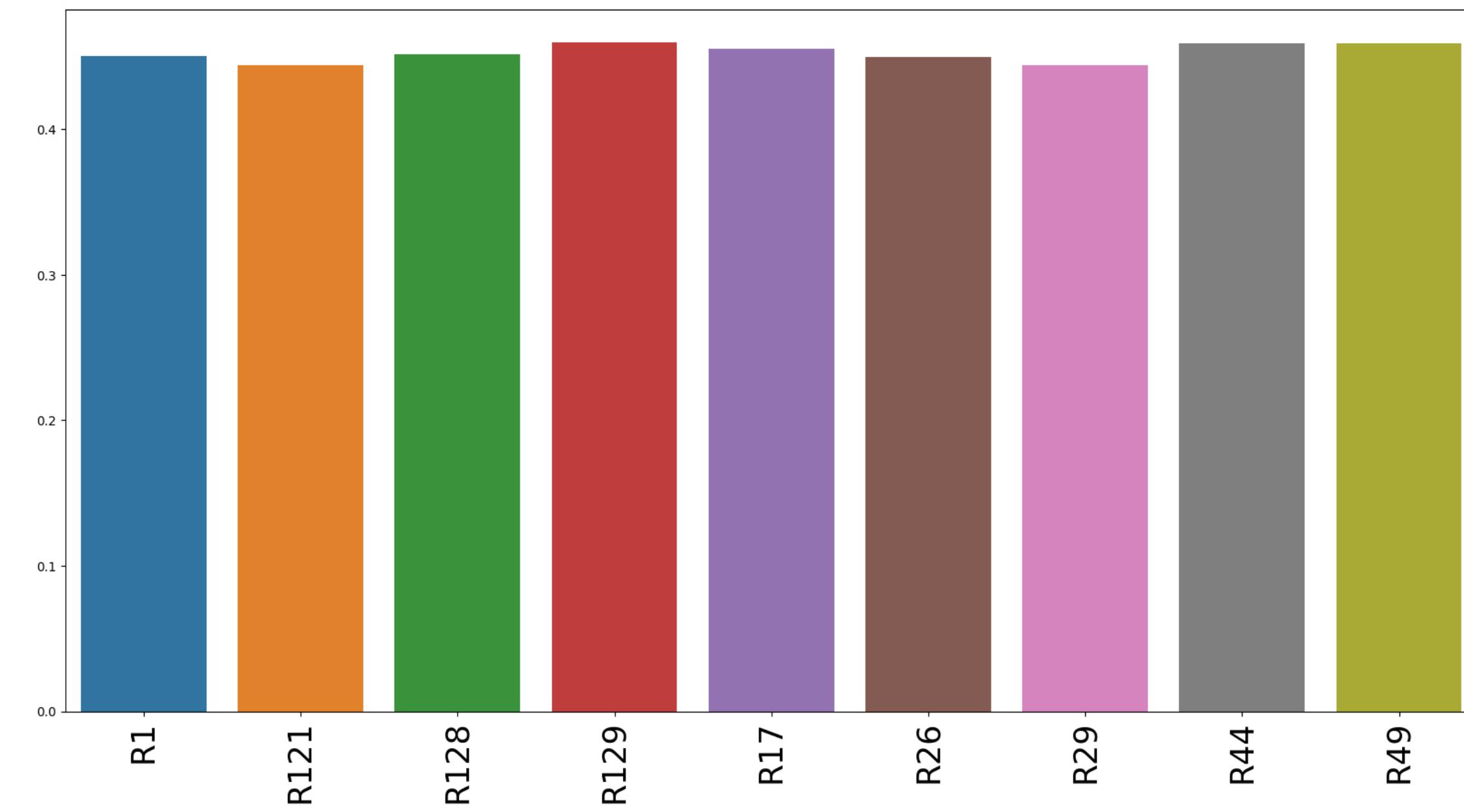


HERO INDIVIDUAL WINRATE

Top heroes' winrates
above 53.5% and
bottom winrate below
46%

As Dota 2 is a zero-sum
game, anything above 50%
is considered good, and
anything below 50% is
considered bad

Bottom Winrates of Heroes





DATA PREPROCESSING



DATA FILTERING AND TRANSFORMATION

Keep matches that fulfil the following criteria:



Ranked Matches with average rank above 70

- Assumption: Players that play high-ranked are very familiar with the game and more likely to be playing seriously and will be seeking to optimise their hero to the best of their ability



All Draft / All Pick Matches

- Ensure that all heroes are available to be picked for comparison



Match Duration above 15 minutes

- Assumption: Matches below 15 minutes probably indicates that the match ended due to external factors (e.g. people leaving early) rather than team strength

Result: 77,994 rows of data.



DATA PREPROCESSING



Dataset Split

Reverse Mapping

Pycaret / Neural
Network

Train-test-split

- Generate holdout set and validation set

Switch Teams and target labels

- To capture team compositions that are found in the opponent team but not in our team
- Doubles our number of data rows

Accuracy Metric

- Dataset is balanced due to reverse mapping
- No downside for incorrect prediction



ML MODELLING



TRADITIONAL MODELS

Model	Accuracy	
catboost	CatBoost Classifier	0.5730
ridge	Ridge Classifier	0.5609
lda	Linear Discriminant Analysis	0.5609
lr	Logistic Regression	0.5607
lightgbm	Light Gradient Boosting Machine	0.5518
svm	SVM - Linear Kernel	0.5485
gbc	Gradient Boosting Classifier	0.5462
ada	Ada Boost Classifier	0.5454
xgboost	Extreme Gradient Boosting	0.5415
rf	Random Forest Classifier	0.5352
et	Extra Trees Classifier	0.5329
nb	Naive Bayes	0.5113
knn	K Neighbors Classifier	0.5097
dt	Decision Tree Classifier	0.5096
dummy	Dummy Classifier	0.5007
qda	Quadratic Discriminant Analysis	0.4985

57.3%

TRAINING
ACCURACY

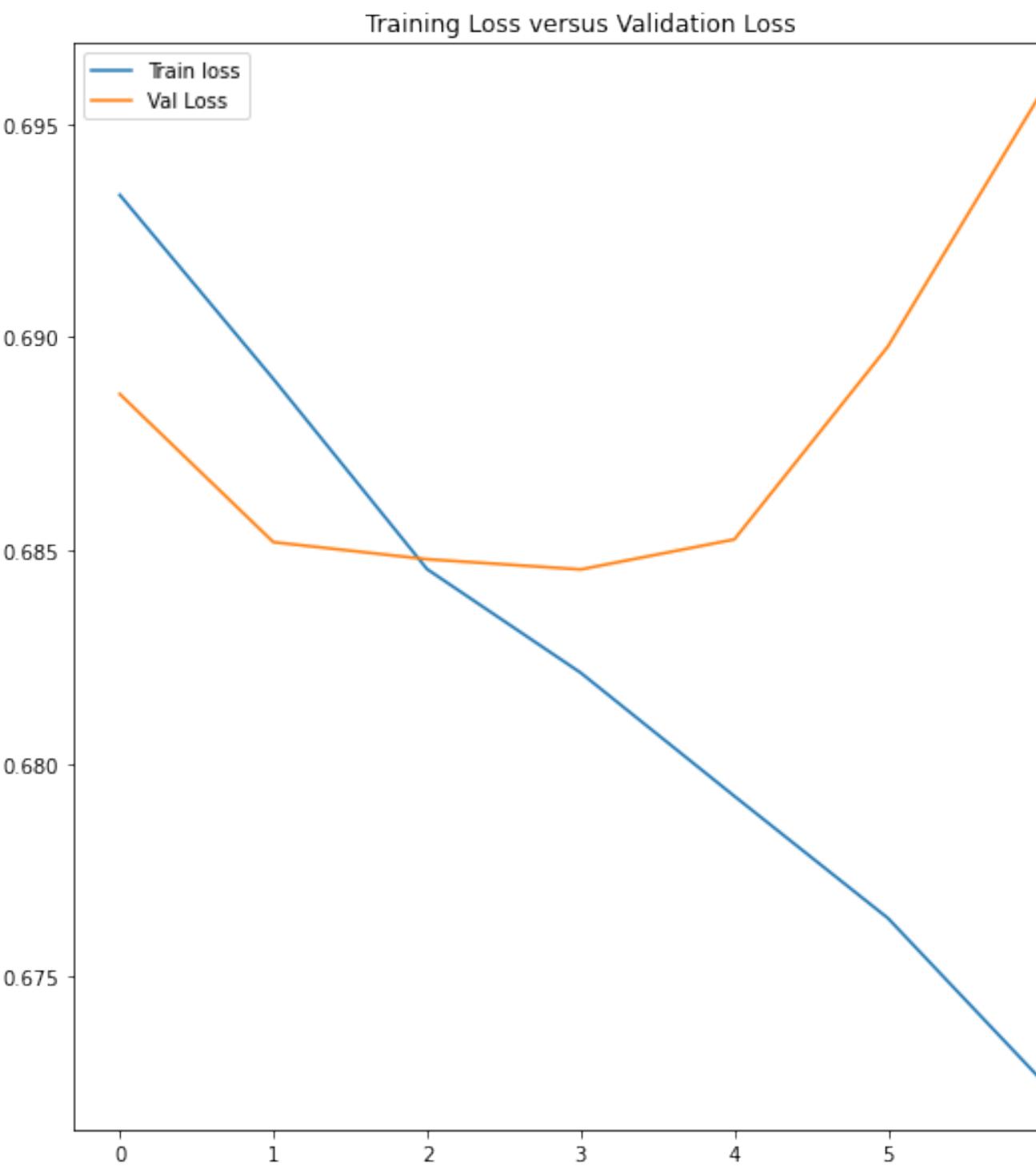
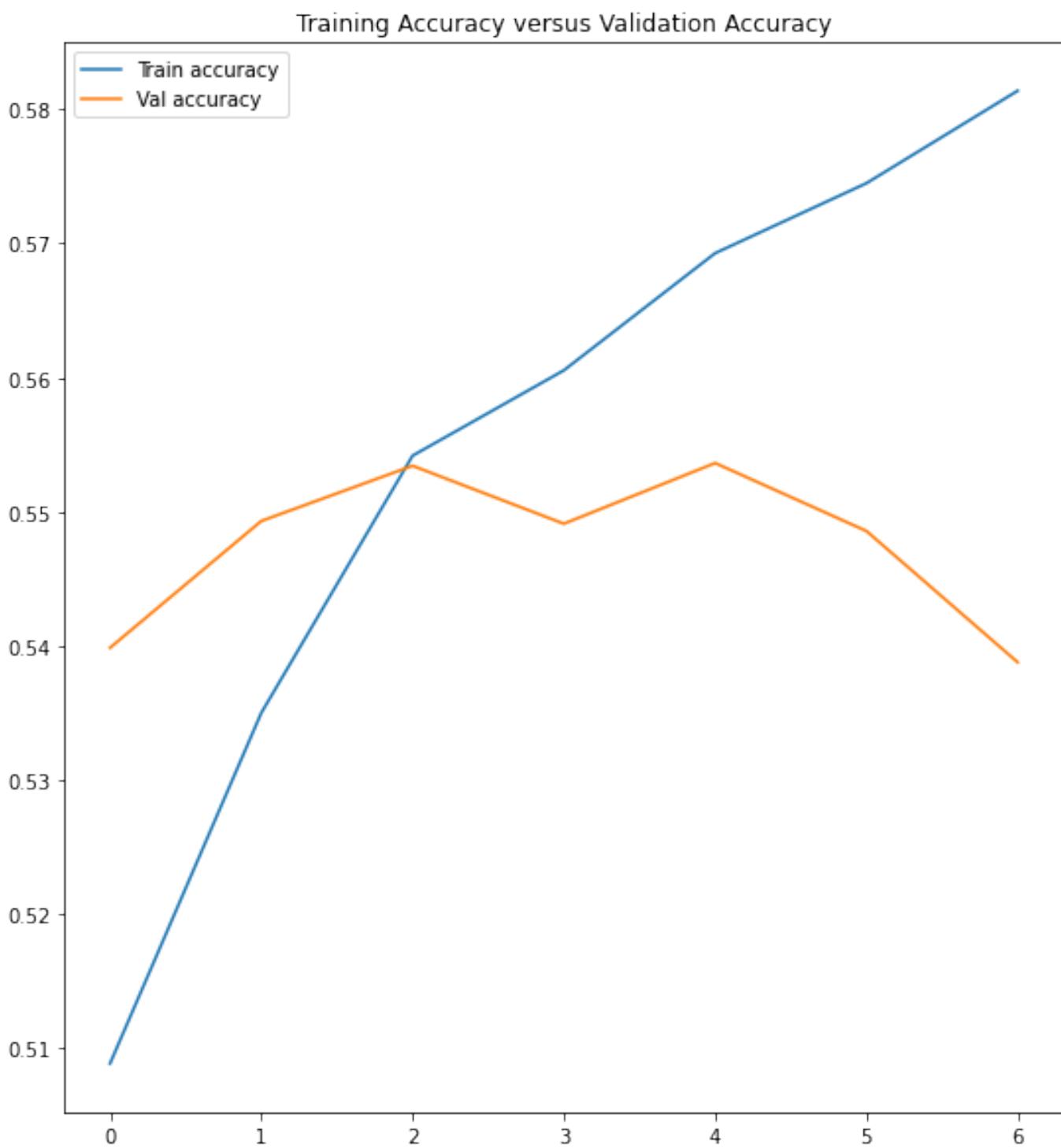
Model	Accuracy
0 Voting Classifier	0.5512

55.1%

VALIDATION
ACCURACY



NEURAL NETWORK

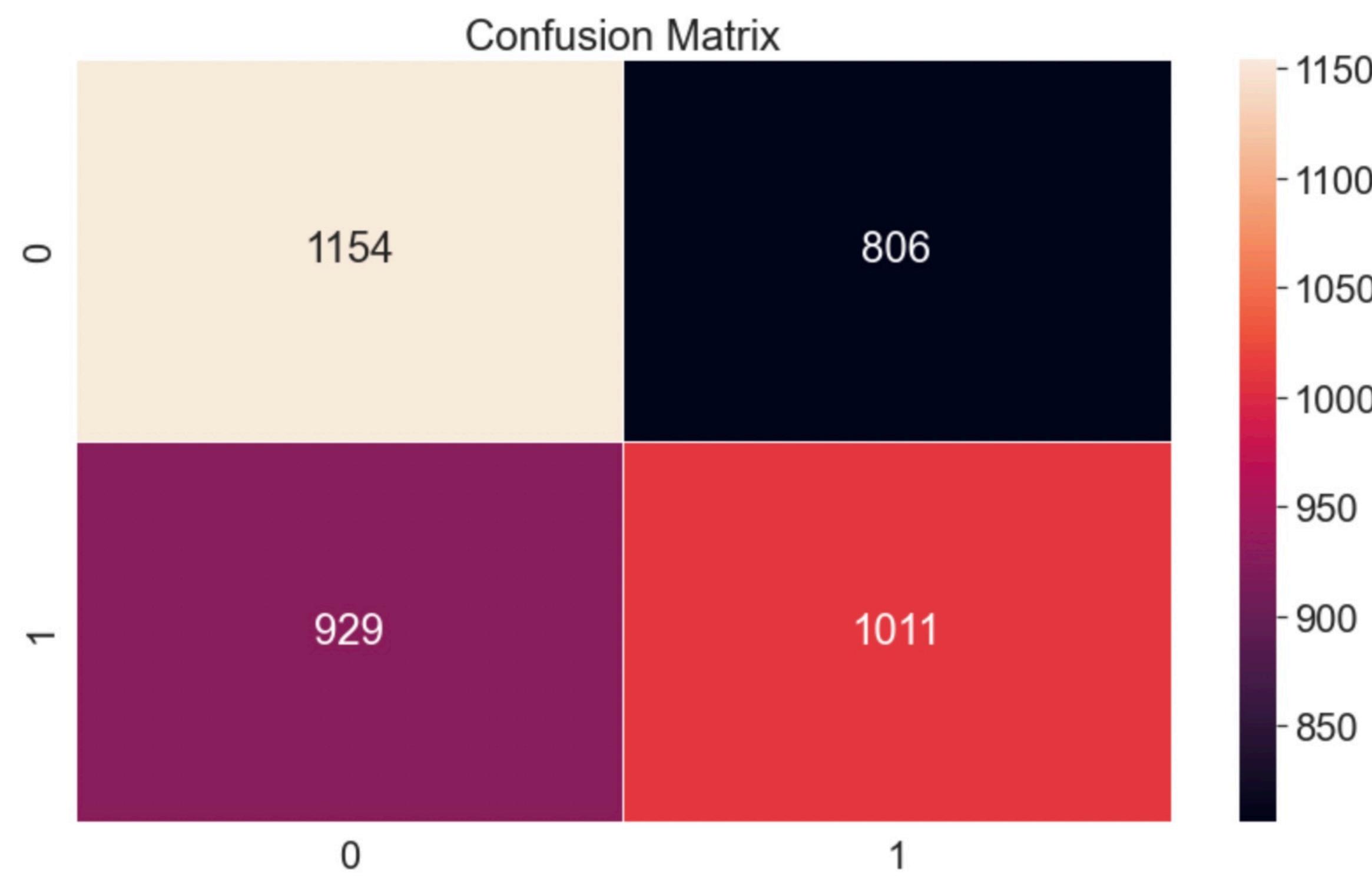


57.0%
TRAINING
ACCURACY

55.4%
VALIDATION
ACCURACY



CONFUSION MATRIX OF HOLDOUT SET



55.6%

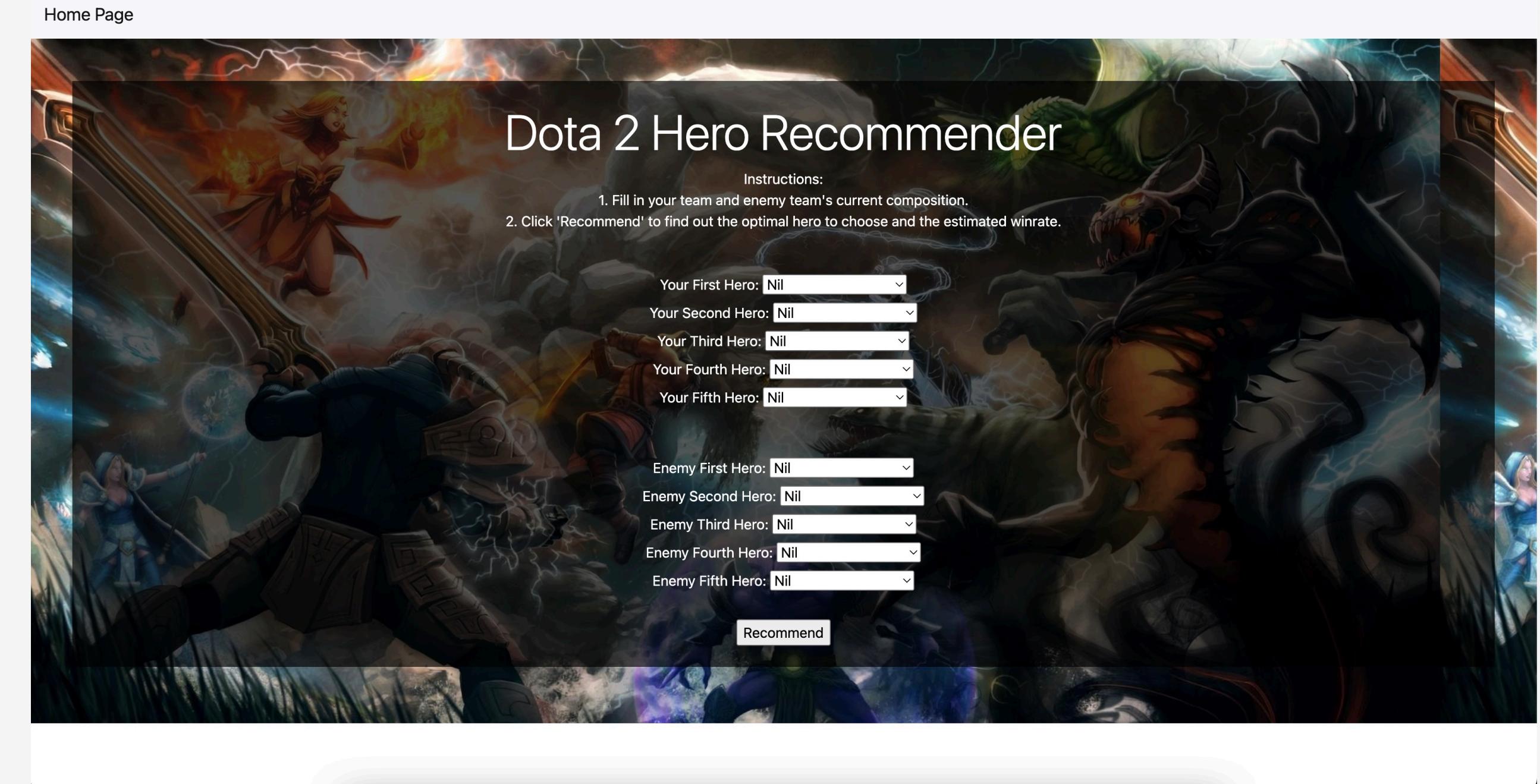
HOLDOUT
ACCURACY



RECOMMENDER SYSTEM

RECOMMENDER WEB APP

- 1 Website app deployed via Flask and Heroku where user inputs their current team and opponent's team hero selection
- 2 Outputs the recommended top 3 heroes and the estimated win rate based on the trained model



Website: <https://dota-2-hero-recommender.herokuapp.com/>



THANK YOU

Zackarias Chia

zackariaschia16@gmail.com

Linkedin: <https://www.linkedin.com/in/zackariaschia/>