19/10/2022

Pokemon Battle Analysis

DATA1030 Midterm Presentation

Taemin Huh

Brown University DSI

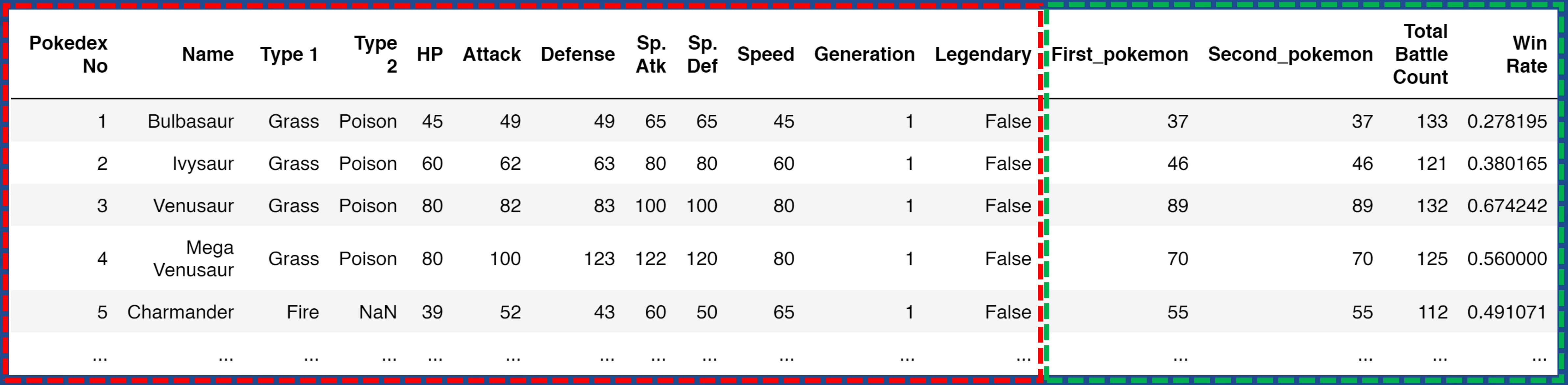
https://github.com/taemin-huh/data1030-project/

Introduction

- Target variable: Win Rate (regression problem)
- 12 features: pkmn dataset (800 Pokemon info datapoints – Kaggle)
  - ID: Pokedex Number, Pokemon Name
  - Type: Type 1, Type 2
  - 6 stats: HP, Attack, Defense, Sp. Atk, Sp. Def, Speed
  - Class: Generation, Legendary
- 3 features: battle dataset (50,000 Pokemon battle datapoints – Kaggle)
  - First Pokemon, Second Pokemon, Winner

Feature Engineering

15 features total (excl. target variable)



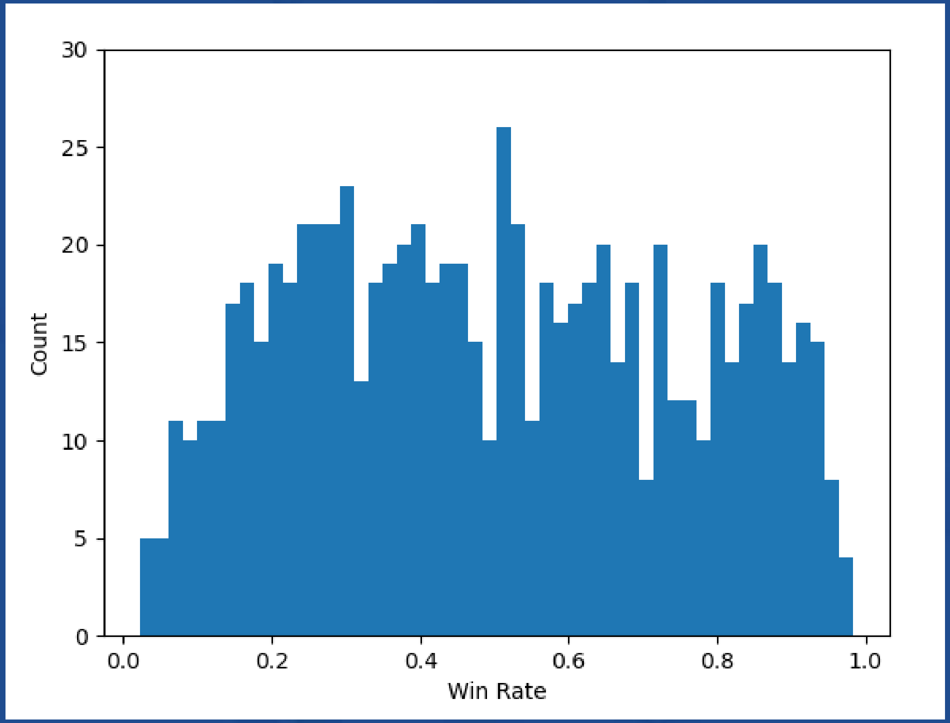
Pokedex No	Name	Type 1	Type 2	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary	First_pokemon	Second_pokemon	Total Battle Count	Win Rate
1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	1	False	37	37	133	0.278195
2	Ivysaur	Grass	Poison	60	62	63	80	80	60	1	False	46	46	121	0.380165
3	Venusaur	Grass	Poison	80	82	83	100	100	80	1	False	89	89	132	0.674242
4	Mega Venusaur	Grass	Poison	80	100	123	122	120	80	1	False	70	70	125	0.560000
5	Charmander	Fire	NaN	39	52	43	60	50	65	1	False	55	55	112	0.491071
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

🟥 Original columns from pkmn dataset 🟩 Feature engineered from battle dataset

Exploratory Data Analysis

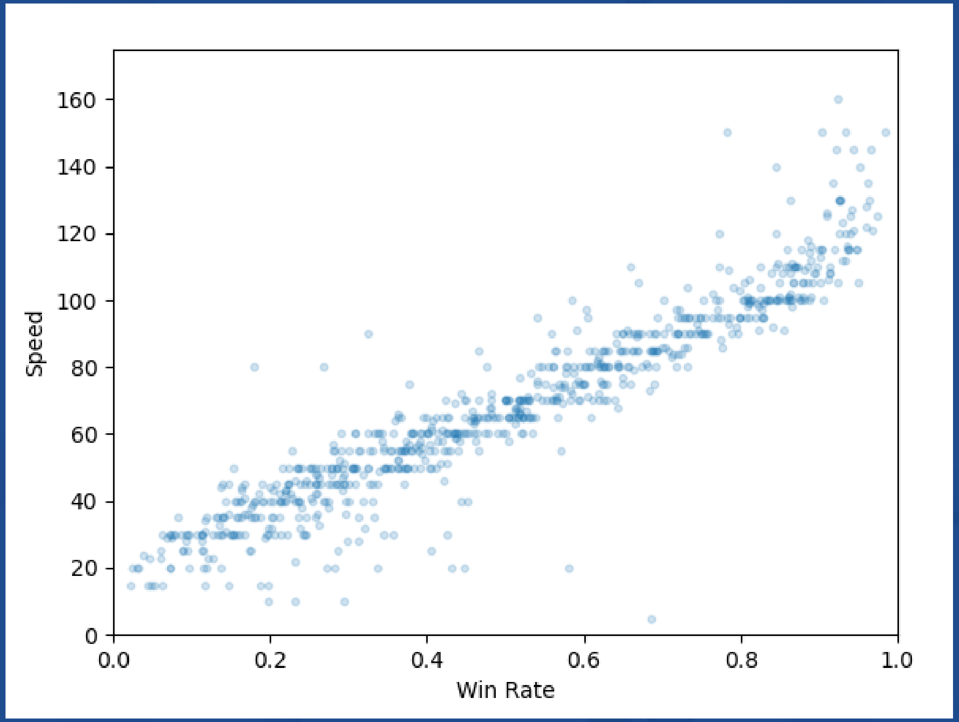
Observations

Histogram: Win Rate Distribution



Symmetrical distribution

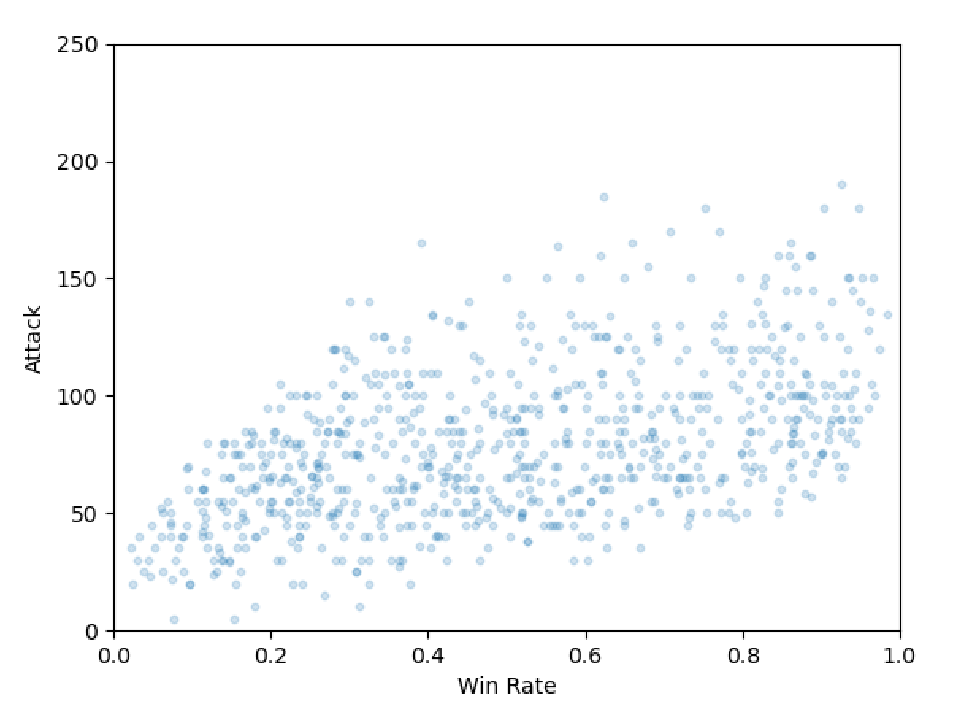
Scatter Plot: Speed vs. Win Rate



Strong correlation (~0.94)

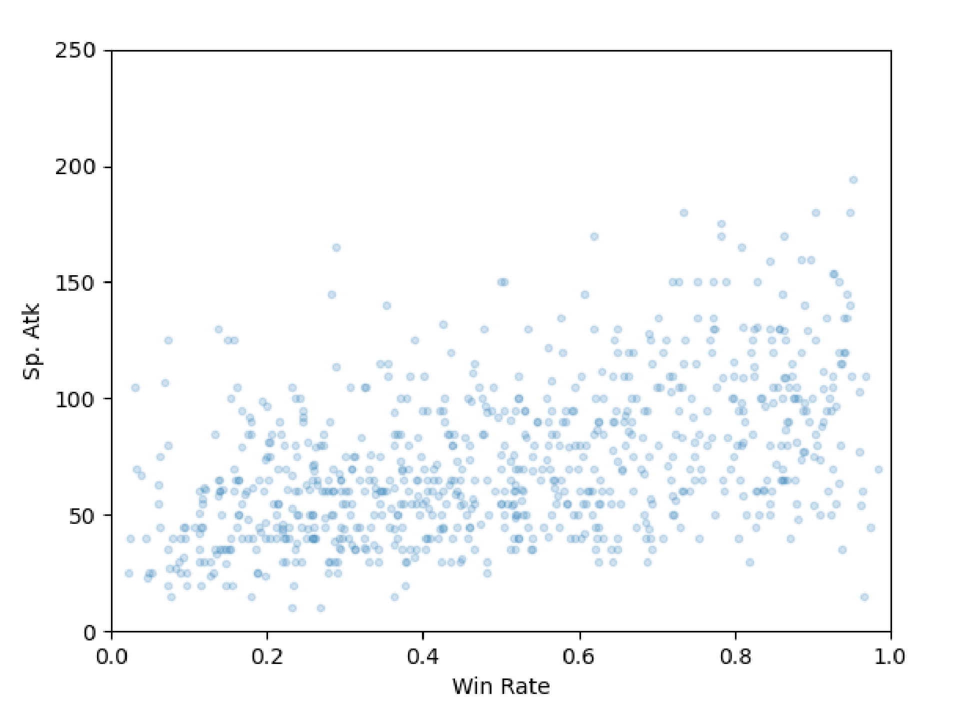
Observations (Cont.)

Scatter Plot: Attack vs. Win Rate



Some correlation (~0.50)

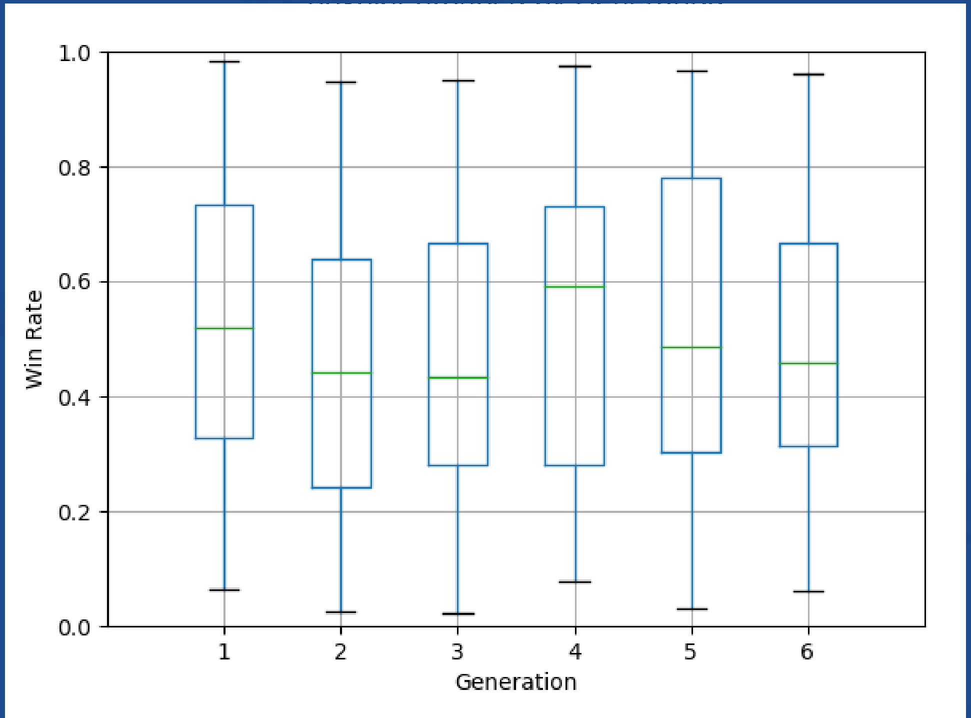
Scatter Plot: Sp. Atk vs. Win Rate



Some correlation (~0.48)

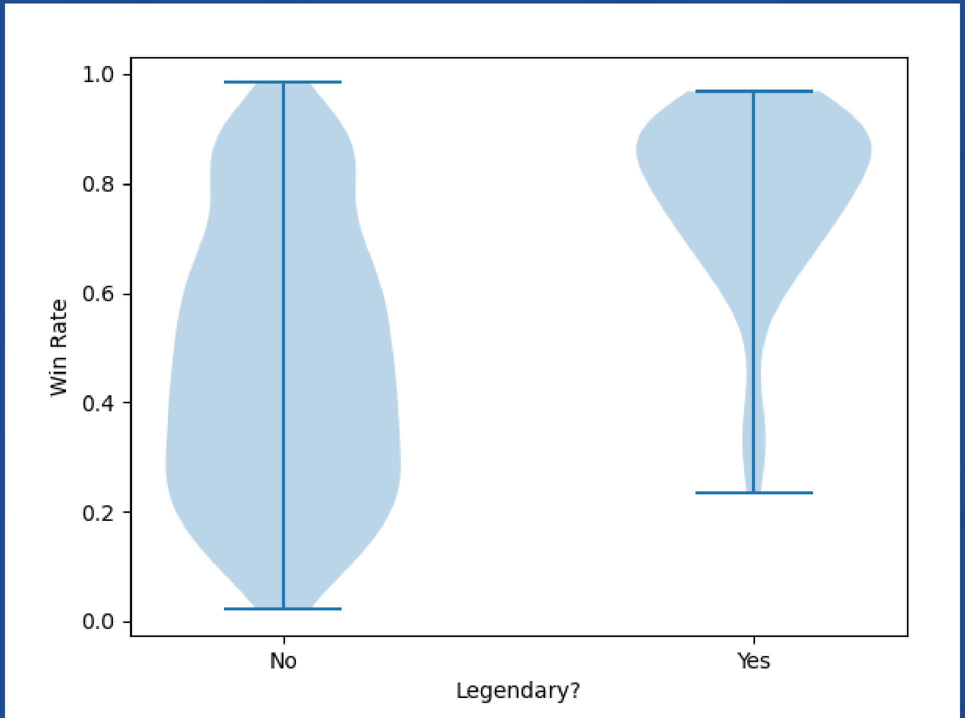
Observations (Cont.)

Box Plot: Win Rate by Generation



No power inflation over generations

Violin Plot: Win Rate of Legendary vs. Not

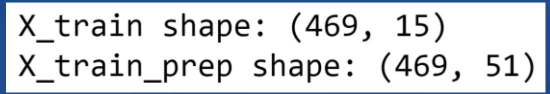


Some correlation

Pre-processing

Pre-Processing

- Basic split (IID, large # of datapoints)
  - 60%/20%/20% for train/test/split
- Pre-processors
  - OneHotEncoder: Type1 (18), Type2 (19), Generation (6), Legendary(2)
  - MinMaxScaler: HP, Attack, Defense, Sp. Atk, Sp. Def, Speed (0-255 each)
- 51 features after pre-processing (15-5+17+18+5+1)



```
X_train shape: (469, 15)
X_train_prep shape: (469, 51)
```

- Missing values: Name (1), Type 2 (386), Win Rate (17)

Thank You