Blackjack Data Analysis

Kasey Mallette

Main Findings

- Out of 130, 745 hands, the player won 43.7% of the hands and pushed 8.6% of the hands
- The more the dealer busts, the more the player wins, but the dealer only busts 23.7% of the time
- The player wins more hands than the dealer in a shoe only 33% of the time
- The five most common player hands are: 12, 13, 20, 14, and 15, and the player loses more hands than wins on 12, 13, 14, and 15
- Using basic strategy to predict hand outcome accounted for 69% of the variance of the test set
- Changing the rules of basic strategy for hands with a win percentage in or below the 10th quantile resulted in a statistically significantly worse overall win percentage

Player Win and Win Percentage of Shoes

Average number of hands per shoe: 45.5 hands

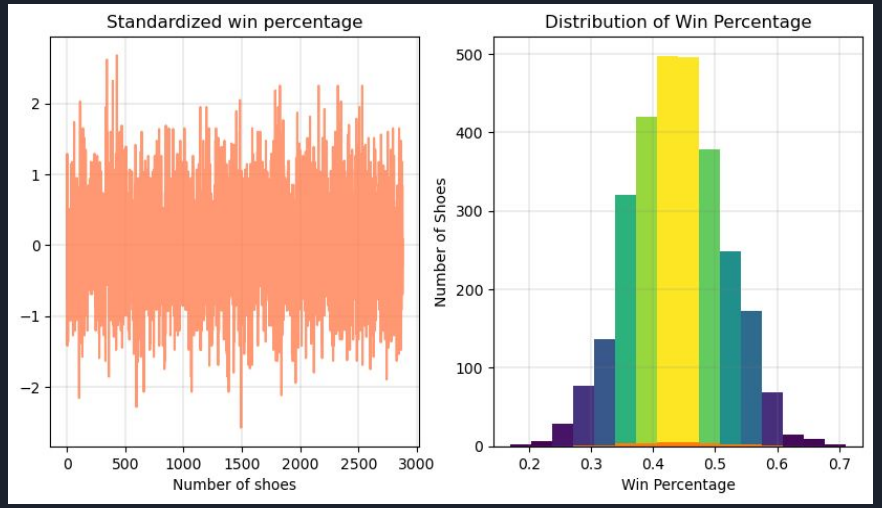
Player win	
count	2880.000000
mean	19.860417
std	3.438331
min	8.000000
25%	18.000000
50%	20.000000
75%	22.000000
max	32.000000

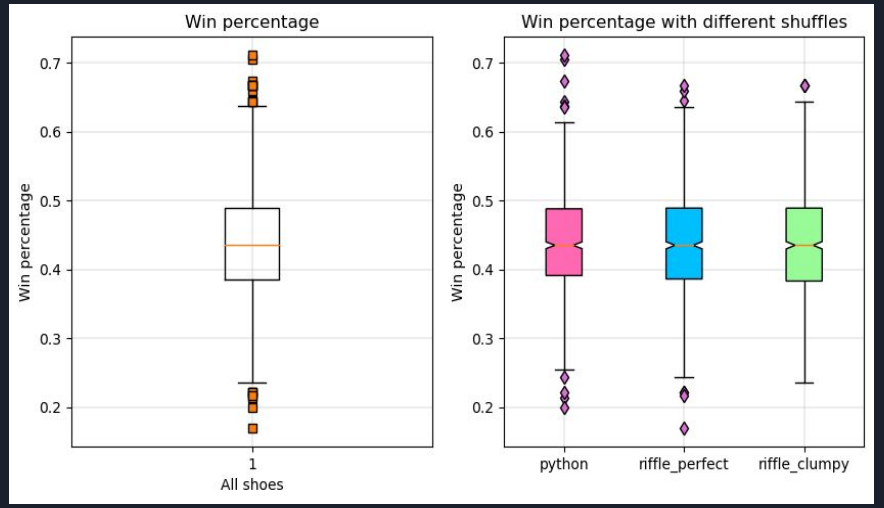
Name: player_win, dtype: float64

Win percentage	
count	2880.000000
mean	0.436624
std	0.074743
min	0.170000
25%	0.386000
50%	0.435000
75%	0.489000
max	0.711000

Name: win_pct, dtype: float64

Shoe Win Percentage Dashboard





Defining a Good Shoe

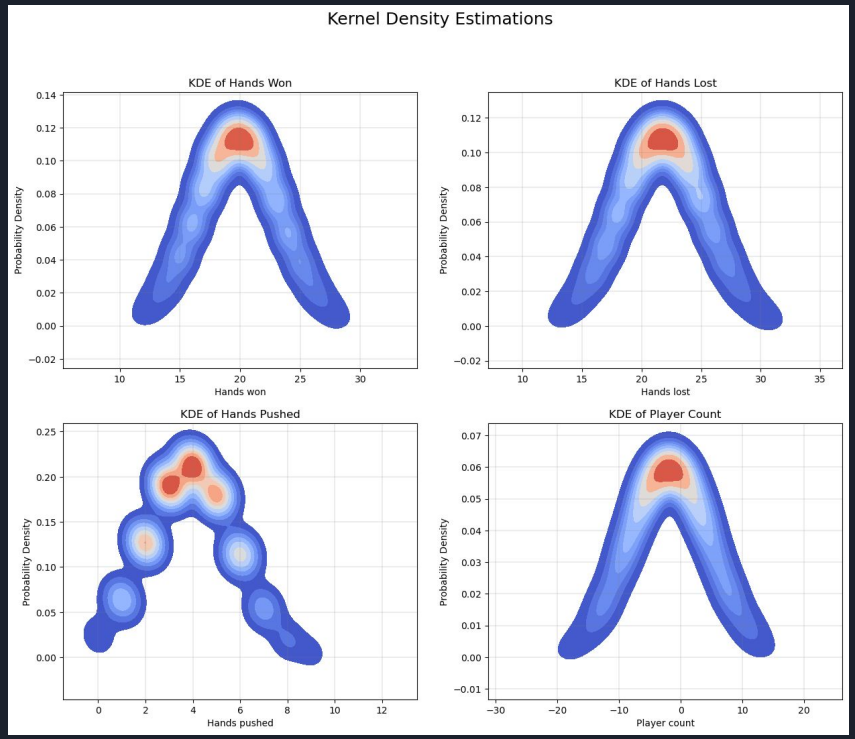
Define a good shoe as:

- A player wins more hands than the dealer and pushes less than 4 times

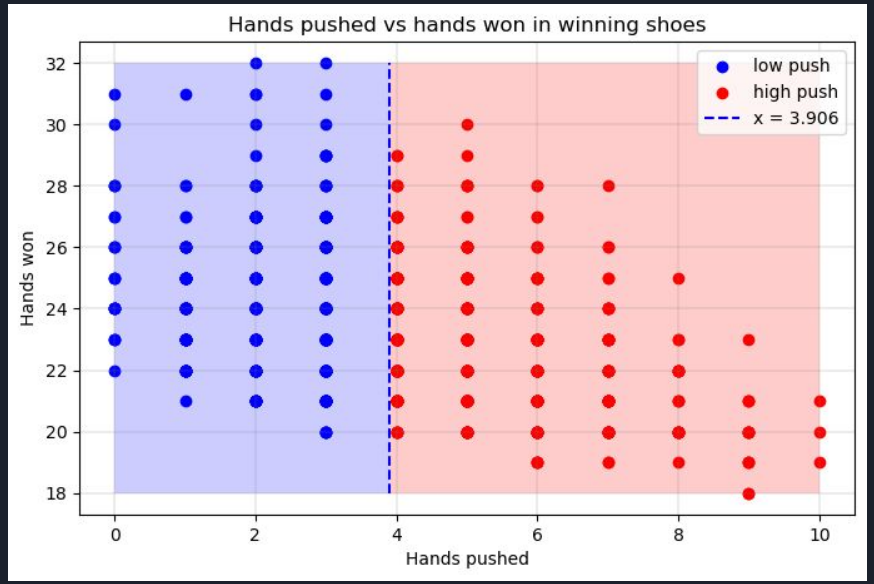
Find the probability of a good shoe:

- The probability of winning more hands than losing is: 33.1%
- The probability of pushing at most 3.906 times is: 50.0%

The probability of a good shoe is: 16.6%



Defining Hands Pushed in Good Shoes



Run a t-test comparing means of player win and hands pushed between two groups

1. Player won more hands than dealer and pushed less than 4 times
2. Player won more hands than dealer and pushed more than 4 times

Ttest_indResult(statistic=12.961842608203566, p-value=1.0794042793599065e-35)

Group 1 mean: 24.183 , count: 459
Group 2 mean: 22.522 , count: 569

Variables Correlated with Player Win

```
             player_win  dealer_bust
player_win     1.000000     0.626211
dealer_bust    0.626211     1.000000
p =  1.85842642744e-313
```

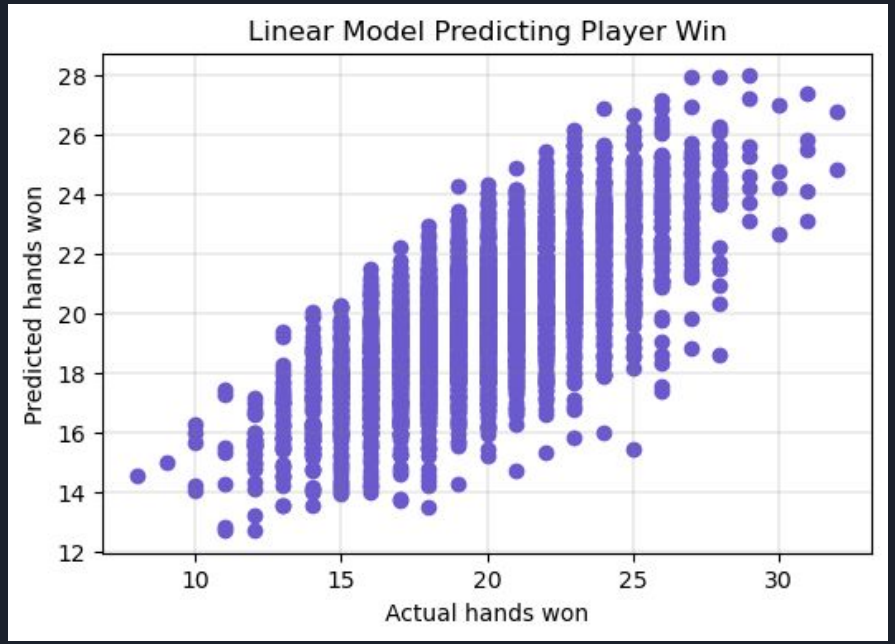
```
             player_win  player_bj
player_win     1.000000   0.278204
player_bj      0.278204   1.000000
p =  2.4105363082935726e-52
```

```
             player_win     push
player_win     1.00000 -0.24682
push          -0.24682  1.00000
p =  3.1369267063762485e-41
```

```
Check for multicollinearity using VIF
       feature       VIF
0  dealer_bust  4.740760
1         push  3.800423
2    player_bj  2.925478
```

Multiple Linear Regression

```
linear model coeff:  [ 0.74496978 -0.23840653
0.77052521]

linear model intercept:  11.261

R-squared score (training):  0.508

R-squared score (test):  0.492


Cross-validated scores: [0.47760746 0.47103567
0.4699999  0.56187995 0.52859468]

Cross-Predicted accuracy: 0.5041378832672987

Negative Mean Absolute Error:  [-1.95321744
-1.92117196 -1.98712102 -1.92348185 -1.92829369]
```



Hand Outcome Dashboard

Frequency of Hands

Frequency of hand outcomes:

	hand outcome	frequency	percentage
0	loss	62367	47.701%
1	win	57143	43.706%
2	push	11235	8.593%

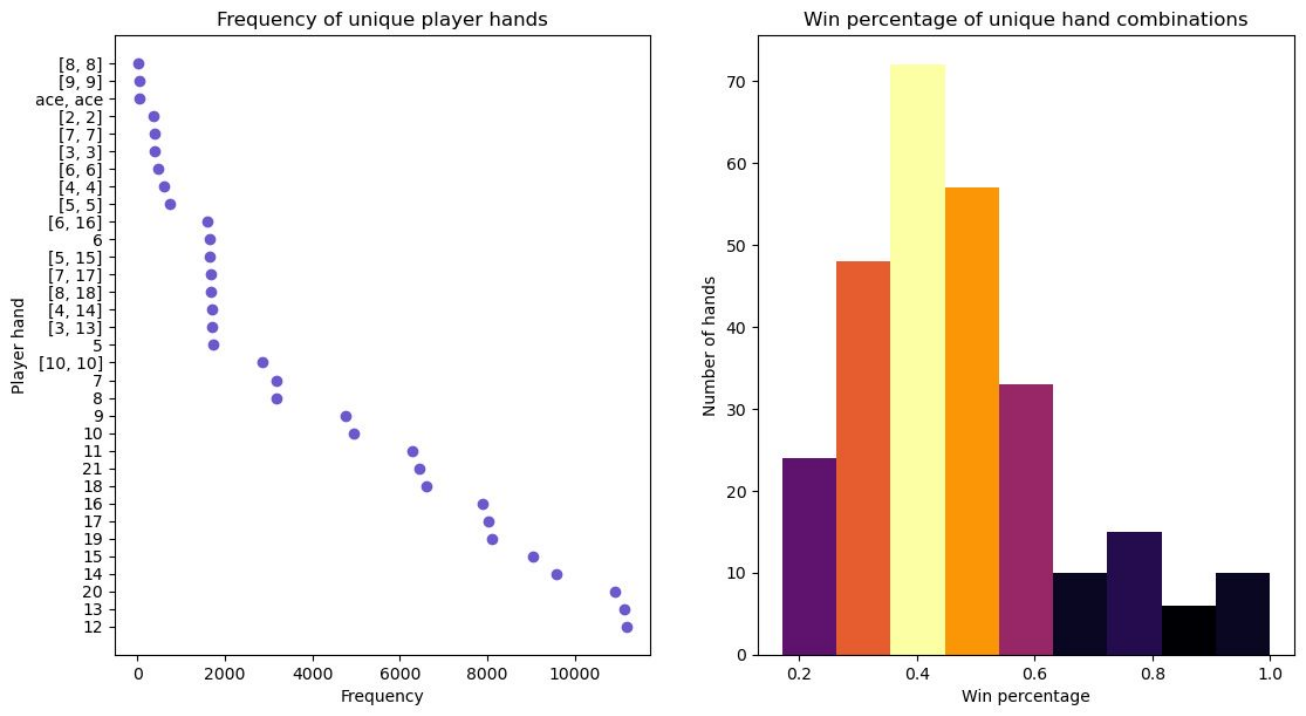
Frequency of dealer outcomes:

	dealer outcome	frequency	percentage
0	stand	58706	44.901%
1	draw	40998	31.357%
2	bust	31041	23.742%

The five most frequent player hands:

	hand	frequency	percentage
0	12	11169	8.543%
1	13	11136	8.517%
2	20	10919	8.351%
3	14	9577	7.325%
4	15	9043	6.917%

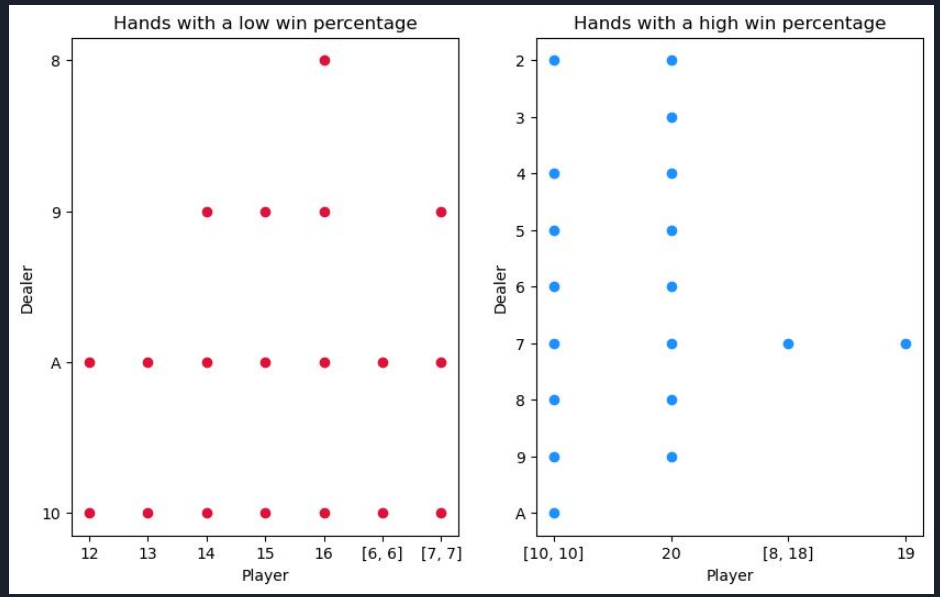
Unique Player Hands



Hands with Low and High Win Percentages

Hands with win percentage in or below 10th quantile

	player	dealer_up	move	frequency	avg_win_pct
0	12	10	hit	3312	0.260
1	12	A	hit	855	0.265
2	13	10	hit	3290	0.248
3	13	A	hit	834	0.261
4	14	A	hit	722	0.221
5	14	10	hit	2920	0.227
6	14	9	hit	709	0.266
7	15	A	hit	668	0.199
8	15	10	hit	2779	0.223
9	15	9	hit	634	0.238
10	16	10	hit	2295	0.188
11	16	9	hit	567	0.216
12	16	A	hit	558	0.227
13	16	8	hit	621	0.262
14	[6, 6]	10	hit	227	0.245
15	[6, 6]	A	hit	67	0.277
16	[7, 7]	A	hit	65	0.220
17	[7, 7]	9	hit	56	0.250
18	[7, 7]	10	hit	225	0.256



Logistic Regression vs Decision Tree Classifier

Using player hand and dealer up card to predict win/loss of hard hands

Logistic Regression

Train score: 0.634

Test score: 0.63

Decision Tree Classifier

Train score: 0.695

Test score: 0.693

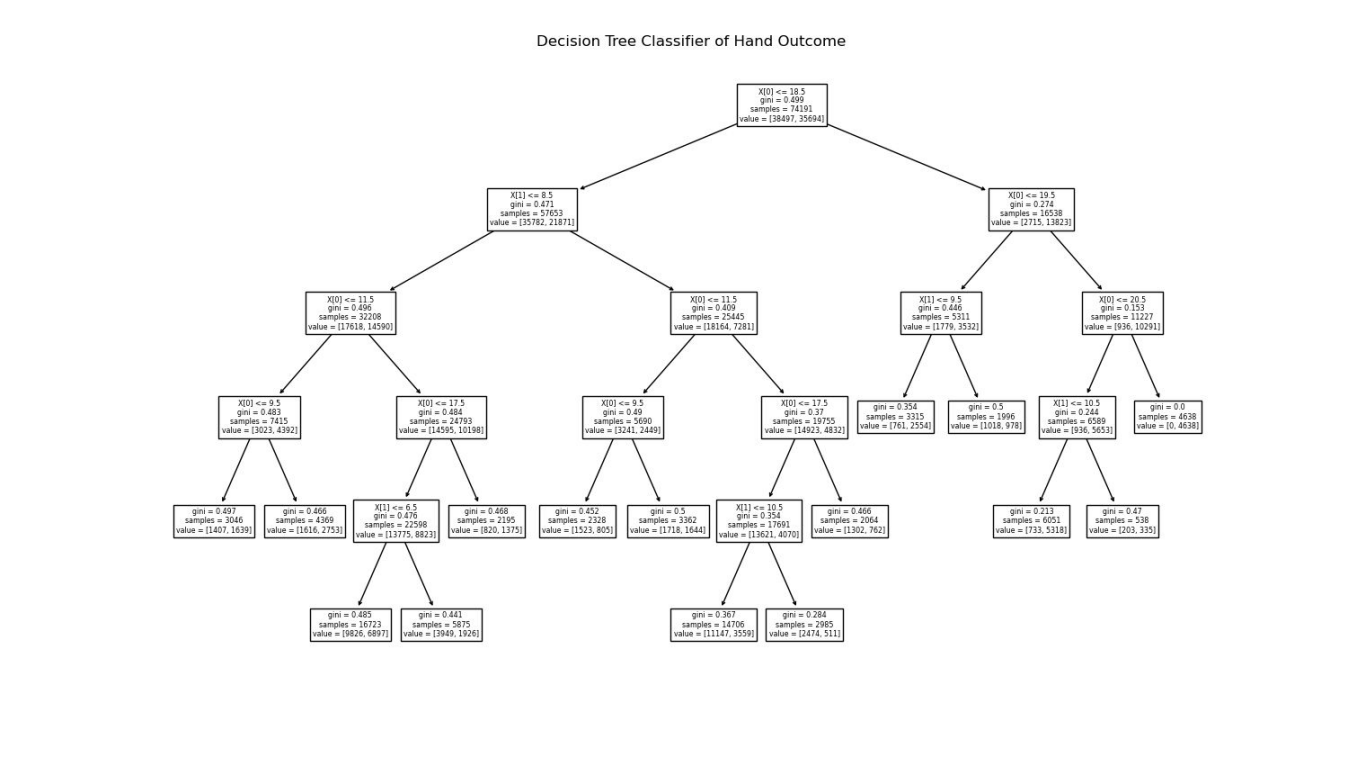
Decision tree classifier (max_depth = 2)

```
 [[10960  1911]
 [ 5684  6176]]
```

Decision Tree Classifier Report

	precision	recall	f1-score	support
not 1	0.66	0.85	0.74	12871
1	0.76	0.52	0.62	11860
accuracy			0.69	24731
macro avg	0.71	0.69	0.68	24731
weighted avg	0.71	0.69	0.68	24731

Decision Tree Classifier



Playing Against Basic Strategy

Changing basic strategy rules from hit to stand for hands with a win percentage in or below the 10th quantile results in a statistically significant worse loss percentage

Average win percentage

T-test of win percentage of low win hands and their rule changes

Ttest_indResult(statistic=3.7511988377494747, p-value=0.0006188824210168993)

Original hands: 0.239

Rule changes: 0.201

Average loss percentage

T-test of loss percentage of low win hands and their rule changes

Ttest_indResult(statistic=-9.81188841514193, p-value=1.0292760912087157e-11)

Original hands: 0.693

Rule changes: 0.799

Conclusion

- If you bet the same amount each hand and play basic strategy, you will likely break even
- Basic strategy works, but the house will always have an advantage
- There is a very low probability of getting that one great shoe, and it is most likely based on betting strategies and luck
- It’s better to be a blackjack dealer than a blackjack player, unless you win