Hockey Bets

BUS 462 D100
Group Lambda

TABLE OF CONTENTS

Introduction

Key Question: Can we use the historical game statistics from the NHL to predict a win or loss?

Ultimate Goal: Create a betting model based on NHL data that is profitable in the long run.

Secondary Goal: Create a reusable method for transforming and analyzing data across multiple sports.

Analysis Method: LOGIT Regression

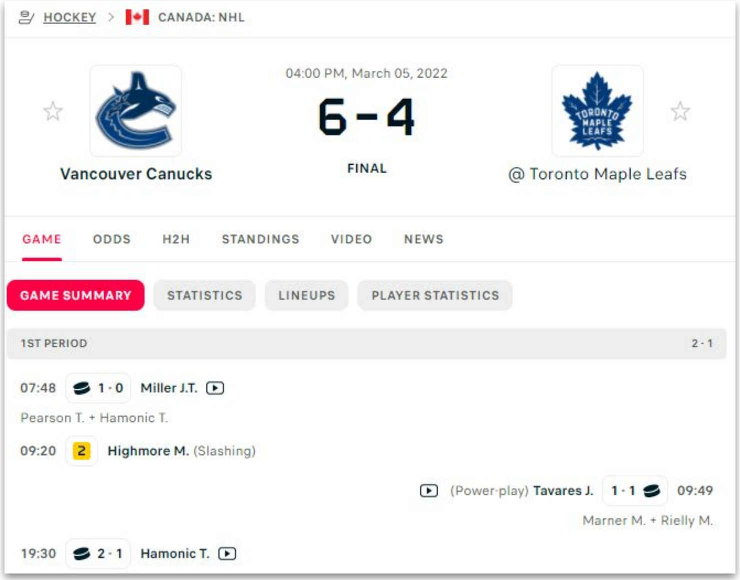
Findings Preview: You should bet if the probability of winning is 0.3 or greater

Data Description

Format: csv

Data Collection:
NHL season data from 2010 to 2021

Source: Flashscore.com



Scraped Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
13067	2019-11	NHL	Florida I	Buffalo :	2	5	Konec	2.12	4	2.93	https://w	45	26	4.44% (	19.23%	21	43	80.77%	95.56%	2	4	4	8	1	0	0	0
13068	2019-11	NHL	Detroit F	Carolina	0	2	Konec	3.57	4.04	1.88	https://w	19	36	0% (0/1	5.56% (	34	19	97.14%	100% (1	2	1	4	2	0	0	0	1
13069	2019-11	NHL	Arizona	Edmont	3	4	Po náje:	2.29	3.88	2.71	https://w	30	37	10% (3/	8.11% (	34	27	91.89%	90% (27	1	2	2	4	1	1	0	0
13070	2019-11	NHL	Columb	Ottawa	1	0	Konec	1.97	4.06	3.27	https://w	19	25	5.26% (	0% (0/2	25	18	100% (2	94.74%	2	3	4	6	0	0	0	0
13071	2019-11	NHL	New Yo	Minneso	3	2	Po prod	2.68	3.92	2.3	https://w	31	28	9.68% (	7.14% (	26	28	92.86%	90.32%	2	4	4	8	1	0	0	0
13072	2019-11	NHL	Tampa E	Buffalo :	5	2	Konec	1.7	4.35	4.14	https://w	29	30	17.24%	6.67% (	28	24	93.33%	82.76%	4	4	8	8	1	0	2	0
13073	2019-11	NHL	Pittsbur	Calgary	3	2	Po prod	1.9	4.02	3.46	https://w	38	34	7.89% (	5.88% (	32	35	94.12%	92.11%	2	3	4	6	1	1	0	0
13074	2019-11	NHL	Philadel	Vancouv	2	1	Konec	2.17	3.95	2.87	https://w	34	17	5.88% (	5.88% (	16	32	94.12%	94.12%	2	1	4	2	0	0	0	0
13075	2019-11	NHL	Nashvill	St. Louis	3	2	Po náje:	2.05	3.99	3.1	https://w	39	25	5.13% (	8% (2/2	23	37	92% (23	94.87%	4	4	11	11	0	0	0	0
13076	2019-11	NHL	Dallas S	Vegas G	4	2	Konec	2.13	3.93	2.97	https://w	28	28	14.29%	7.14% (	26	24	92.86%	85.71%	5	4	10	8	2	0	0	0
13077	2019-11	NHL	Anahein	New Yo	3	0	Konec	2.74	3.85	2.29	https://w	23	26	13.04%	0% (0/2	26	20	100% (2	86.96%	3	3	9	9	0	0	0	0
13078	2019-11	NHL	Los Ang	San Jos	3	4	Po prod	2.58	3.89	2.38	https://w	36	25	8.33% (	16% (4/	21	33	84% (21	91.67%	2	5	4	10	0	0	0	0
13079	2019-11	NHL	New Jer	Minneso	2	3	Konec	2.14	3.96	2.92	https://w	34	29	5.88% (	10.34%	26	32	89.66%	94.12%	4	2	8	4	1	1	0	0
13080	2019-11	NHL	Montrea	Boston I	1	8	Konec	2.57	3.9	2.39	https://w	37	24	2.7% (1/	33.33%	16	36	66.67%	97.3% (	3	3	6	6	0	2	0	0
13081	2019-11	NHL	Chicago	Dallas S	3	0	Konec	2.51	3.88	2.45	https://w	38	32	7.89% (	0% (0/3	32	35	100% (3	94.59%	2	1	4	2	1	0	0	0
13082	2019-11	NHL	Washing	Florida I	4	3	Konec	1.91	4.15	3.39	https://w	20	40	20% (4/	7.5% (3/	37	16	92.5% (	80% (16	4	3	8	6	0	0	0	0
13083	2019-11	NHL	Ottawa	Boston I	1	2	Konec	3.15	3.92	2.05	https://w	34	21	2.94% (	9.52% (	19	33	90.48%	97.06%	2	2	4	4	0	0	0	0
13084	2019-11	NHL	Pittsbur	Vancouv	8	6	Konec	1.99	4.03	3.23	https://w	40	22	20% (8/	27.27%	16	32	72.73%	82.05%	3	5	6	10	2	1	0	0
13085	2019-11	NHL	New Yo	Carolina	3	2	Konec	3.12	4	2.04	https://w	25	43	12% (3/	4.65% (	41	22	95.35%	88% (22	5	5	10	10	1	1	0	0
13086	2019-11	NHL	Buffalo :	Calgary	2	3	Po prod	2.69	3.86	2.32	https://w	36	29	5.56% (	10.34%	26	34	89.66%	94.44%	2	2	4	4	0	0	1	0
13087	2019-11	NHL	Columb	Philadel	2	3	Konec	2.43	3.87	2.54	https://w	30	22	6.67% (	13.64%	19	28	86.36%	93.33%	6	1	20	2	0	1	0	0
13088	2019-11	NHL	Detroit F	Toronto	0	6	Konec	4.29	4.21	1.71	https://w	25	54	0% (0/2	11.11%	48	25	88.89%	100% (2	1	0	2	0	0	1	0	0
13089	2019-11	NHL	Tampa E	St. Louis	3	4	Konec	2.05	3.99	3.1	https://w	35	33	8.57% (	12.12%	29	32	87.88%	91.43%	2	4	4	8	1	1	0	1
13090	2019-11	NHL	Nashvill	Vegas G	3	4	Po prod	2.12	3.99	2.94	https://w	34	30	8.82% (	13.33%	26	31	86.67%	91.18%	2	2	4	4	0	0	0	0
13091	2019-11	NHL	Arizona	Anahein	4	3	Po náje:	2.13	3.9	2.99	https://w	28	34	10.71%	8.82% (	31	25	91.18%	89.29%	4	3	8	6	1	0	0	0
13092	2019-11	NHL	Colorad	Edmont	4	1	Konec	2.21	3.97	2.79	https://w	50	20	8% (4/5	5% (1/2	19	46	95% (19	92% (46	4	9	8	29	0	0	0	0
13093	2019-11	NHL	San Jos	Winnipe	1	5	Konec	2.23	3.96	2.77	https://w	33	26	3.03% (	19.23%	21	32	84% (21	96.97%	5	9	18	34	0	1	0	0
13094	2019-11	NHL	Los Ang	New Yo	4	1	Konec	2.99	3.88	2.13	https://w	30	25	13.33%	4% (1/2	24	26	96% (24	89.66%	4	5	8	10	1	0	0	0
13095	2019-11	NHL	Montrea	New Jer	4	6	Konec	1.99	4.06	3.21	https://w	48	35	8.33% (	17.14%	29	44	85.29%	91.67%	3	2	6	4	0	0	0	0
13096	2019-11	NHL	Boston I	New Yo	3	2	Po prod	1.77	4.25	3.81	https://w	27	28	11.11%	7.14% (	26	24	92.86%	88.89%	6	2	17	7	0	0	0	0
13097	2019-11	NHL	Buffalo :	Toronto	6	4	Konec	3.06	3.98	2.07	https://w	36	29	16.67%	13.79%	25	30	86.21%	85.71%	5	3	10	6	0	0	0	0
13098	2019-11	NHL	Minneso	Ottawa	7	2	Konec	1.85	4.3	3.44	https://w	35	35	20% (7/	5.71% (	33	28	94.29%	80% (28	2	4	4	8	1	0	0	0
13099	2019-11	NHL	Chicago	Colorad	2	5	Konec	2.32	3.94	2.65	https://w	36	23	5.56% (	21.74%	18	34	78.26%	94.44%	5	3	13	9	0	1	0	0
13100	2019-11	NHL	San Jos	Los Ang	4	1	Konec	1.95	4.07	3.32	https://w	22	34	18.18%	2.94% (	33	18	97.06%	81.82%	1	3	2	6	0	0	0	0
13101	2019-11	NHL	Anahein	Winnipe	0	3	Konec	2.28	3.9	2.71	https://w	24	20	0% (0/2	15% (3/	17	24	85% (17	100% (2	8	4	35	11	0	1	0	0
13102	2019-11	NHL	Philadel	Detroit F	6	1	Konec	1.59	4.52	4.85	https://w	35	33	17.14%	3.03% (	32	29	96.97%	82.86%	4	3	6	6	1	0	0	0
13103	2019-11	NHL	Washing	Tampa E	4	3	Po prod	2.32	3.97	2.63	https://w	35	30	11.43%	10% (3/	27	31	90% (27	88.57%	5	3	10	6	2	1	0	0
13104	2019-11	NHL	Vegas G	Arizona	2	1	Po náje:	2.04	4	3.12	https://w	38	36	2.63% (	2.78% (	35	37	97.22%	97.37%	2	5	4	10	1	0	0	0

Data Description: Transformation

1 match is being represented as 2 rows → from the home team POV and the away team POV

Transforming all data to average over last n games for every team.

Created dummy variables for each team as a home team and as an away team

Draws are being interpreted as Lose because we are looking only at the bet_outcome (Win=1 or lose=0)

0.228773	Arizona Coyotes	Boston Bruins	1	6	4.13	8/12/2017 1:00	L
0.543007	Boston Bruins	Arizona Coyotes	6	1	1.74	8/12/2017 1:00	W

In the next step, we created a dataframe with averaged values from the last 5 games.

Key Variables

Variable	Description	Unit
shots_on_goals	Shots on goal for team	Number of shots on goal
goalie_saves	Saves by goalie	Number of saves
penalties	Penalties for team	Number of penalties
PIM	Penalty minutes for team	Number of penalty minutes
PP_goals	Powerplay goals for team	Number of powerplay goals
SH_goals	Shorthanded goals for team	Number of shorthanded goals
bodychecks	Body checks for team	Number of bodychecks
faceoffs_won	Faceoffs won for team	Number of faceoffs won

Key Variables continued...

Variable	Description	Unit
faceoffs_%	Faceoff win percentage for team	Percentage
empty_net_goals	Empty net goals for team	Number of empty net goals
blocked_shots	Shots blocked for the team	Number of blocked shots
win_prob	Win probability for the team	Probability
team_goals	Goals for the team	Number of goals
opponent_goals	Goals for the away team	Number of goals
bet	Win multiplier	Number
bet_outcome	Win or lose for the bet	Win=1 or lose =0

Analysis: Model 1

Logit Model using the last 5 games
DV: Bet Outcome
IV: All the variables we have, except (“giveaways”, “takeaways”, “shots_off_goal”)

Confusion Matrix (test_set):

	Predicted Loss	Predicted Win
Actual Loss	**3184**	**251**
Actual Win	**1812**	**271**

Test accuracy = 0.630

```
Optimization terminated successfully.
         Current function value: 0.646304
         Iterations 5
                         Results: Logit
==============================================================================
Model:                 Logit              Pseudo R-squared:    0.030
Dependent Variable:    Bet_outcome        AIC:                 28683.1535
Date:                  2022-04-04 22:24   BIC:                 29299.3090
No. Observations:      22071              Log-Likelihood:      -14265.
Df Model:              76                 LL-Null:             -14703.
Df Residuals:          21994              LLR p-value:         2.3539e-136
Converged:             1.0000             Scale:               1.0000
No. Iterations:        5.0000
------------------------------------------------------------------------------
                         Coef.  Std.Err.    z     P>|z|   [0.025   0.975]
------------------------------------------------------------------------------
shots_on_goal           0.0046   0.0040   1.1439 0.2527 -0.0033  0.0125
goalie_saves           -0.0094   0.0040  -2.3448 0.0190 -0.0172 -0.0015
penalties               0.0155   0.0258   0.5998 0.5487 -0.0351  0.0661
PIM                    -0.0003   0.0075  -0.0465 0.9629 -0.0150  0.0143
PP_goals               -0.0222   0.0405  -0.5486 0.5833 -0.1015  0.0571
SH_goals               -0.0837   0.1032  -0.8110 0.4174 -0.2859  0.1185
bodychecks             -0.0003   0.0028  -0.1080 0.9140 -0.0057  0.0051
faceoffs_won            0.0093   0.0071   1.3252 0.1851 -0.0045  0.0232
faceoffs_%             -0.0064   0.0050  -1.2633 0.2065 -0.0162  0.0035
empty_net_goals         0.0966   0.0880   1.0971 0.2726 -0.0760  0.2691
blocked_shots           0.0059   0.0022   2.6327 0.0085  0.0015  0.0102
win_prob                0.1836   0.3256   0.5639 0.5728 -0.4546  0.8218
```

Analysis Method

If we bet on Tampa Bay, where our Model 1 predicted a win probability of **0.5**, then we would have won on our test set **41 times** our betting amount we set. Which means, that the return on this rule is roughly **14%**.

	index	team	prob	won	profit
85	85	Tampa Bay Lightning	0.5	346.06	41.06
84	84	Tampa Bay Lightning	0.4	670.6	29.6
300	300	Vegas Golden Knights	0	328.68	20.68
301	301	Vegas Golden Knights	0.1	328.68	20.68
302	302	Vegas Golden Knights	0.2	328.68	20.68
95	95	Winnipeg Jets	0.5	71.95	18.95
272	272	Anaheim Ducks	0.2	715.46	18.46
304	304	Vegas Golden Knights	0.4	281.98	16.98
273	273	Anaheim Ducks	0.3	650.2	15.2
215	215	Colorado Avalanche	0.5	89.75	13.75
34	34	Minnesota Wild	0.4	345.62	12.62
303	303	Vegas Golden Knights	0.3	317.04	12.04
150	150	Columbus Blue Jackets	0	732.06	11.06
151	151	Columbus Blue Jackets	0.1	732.06	11.06
271	271	Anaheim Ducks	0.1	724.25	10.25

Analysis: Model 2

Logit Model without dummy variables ("team", "opponent")

Confusion Matrix (test_set):

	Predicted Loss	Predicted Win
Actual Loss	**3271**	**164**
Actual Win	**1875**	**208**

Test accuracy = 0.630

```
Optimization terminated successfully.
         Current function value: 0.648289
         Iterations 5
                         Results: Logit
==================================================================
Model:              Logit            Pseudo R-squared: 0.027
Dependent Variable: Bet_outcome      AIC:              28646.7693
Date:               2022-04-04 22:39 BIC:              28766.7996
No. Observations:   22071            Log-Likelihood:   -14308.
Df Model:           14               LL-Null:          -14703.
Df Residuals:       22056            LLR p-value:      3.0459e-159
Converged:          1.0000           Scale:            1.0000
No. Iterations:     5.0000
------------------------------------------------------------------
                  Coef.  Std.Err.    z     P>|z|   [0.025  0.975]
------------------------------------------------------------------
shots_on_goal     0.0057   0.0039   1.4570 0.1451 -0.0020  0.0133
goalie_saves     -0.0038   0.0035  -1.0642 0.2872 -0.0107  0.0032
penalties         0.0298   0.0253   1.1769 0.2392 -0.0198  0.0793
PIM              -0.0020   0.0074  -0.2733 0.7846 -0.0165  0.0124
PP_goals         -0.0377   0.0401  -0.9409 0.3468 -0.1162  0.0408
SH_goals         -0.0906   0.1020  -0.8882 0.3745 -0.2906  0.1094
bodychecks        0.0016   0.0024   0.6625 0.5077 -0.0032  0.0064
faceoffs_won      0.0036   0.0067   0.5339 0.5934 -0.0096  0.0167
faceoffs_%        0.0006   0.0045   0.1370 0.8910 -0.0083  0.0095
empty_net_goals   0.1045   0.0874   1.1952 0.2320 -0.0669  0.2759
blocked_shots     0.0064   0.0022   2.9323 0.0034  0.0021  0.0107
win_prob          0.8740   0.2603   3.3572 0.0008  0.3637  1.3842
Team_Goals        0.0464   0.0216   2.1452 0.0319  0.0040  0.0888
Opponent_Goals    0.0185   0.0183   1.0100 0.3125 -0.0174  0.0545
Bet              -0.5398   0.0234 -23.1012 0.0000 -0.5857 -0.4940
==================================================================
```

Analysis Method

If we bet on Tampa Bay, where our Model 2 predicted a win probability of winning **0.4**, then we would have won on our test set **26.71 times** our betting amount we set. Which means, that the return on this rule is roughly **4.6%**.

index	team	prob	won	profit
84	Tampa Bay Lightning	0.4	590.71	26.71
272	Anaheim Ducks	0.2	711.43	23.43
300	Vegas Golden Knights	0	328.68	20.68
301	Vegas Golden Knights	0.1	328.68	20.68
302	Vegas Golden Knights	0.2	328.68	20.68
150	Columbus Blue Jackets	0	732.06	11.06
151	Columbus Blue Jackets	0.1	732.06	11.06
83	Tampa Bay Lightning	0.3	774.57	10.57
304	Vegas Golden Knights	0.4	249.56	10.56
271	Anaheim Ducks	0.1	724.25	10.25
274	Anaheim Ducks	0.4	364.64	9.64
270	Anaheim Ducks	0	724.25	9.25
95	Winnipeg Jets	0.5	35.49	8.49
215	Colorado Avalanche	0.5	77.21	8.21
255	New York Rangers	0.5	49.84	7.84

Analysis: Model 3

Logit Model ONLY using “bet”, “win_prob”

Confusion Matrix (test_set):

	Predicted Loss	Predicted Win
Actual Loss	**3298**	**137**
Actual Win	**1906**	**177**

Test accuracy = 0.630

```
                         Results: Logit
=================================================================
Model:              Logit            Pseudo R-squared: 0.025
Dependent Variable: Bet_outcome      AIC:              28660.1040
Date:               2022-04-04 23:36 BIC:              28676.1081
No. Observations:   22071            Log-Likelihood:   -14328.
Df Model:           1                LL-Null:          -14703.
Df Residuals:       22069            LLR p-value:      5.8205e-165
Converged:          1.0000           Scale:            1.0000
No. Iterations:     5.0000
-----------------------------------------------------------------
               Coef.   Std.Err.     z      P>|z|   [0.025  0.975]
-----------------------------------------------------------------
Bet           -0.4745    0.0164 -28.9084  0.0000 -0.5067 -0.4423
win_prob       1.9347    0.1098  17.6214  0.0000  1.7195  2.1499
=================================================================
```

Analysis Method

If we bet on Anaheim Ducks, where our Model 3 predicted a win probability of **0.2**, then we would have won on our test set **24.46 times** our betting amount we set. Which means, that the return on this rule is roughly **3.5%**.

index	team	prob	won	profit
272	Anaheim Ducks	0.2	715.46	24.46
300	Vegas Golden Knights	0	328.68	20.68
301	Vegas Golden Knights	0.1	328.68	20.68
302	Vegas Golden Knights	0.2	328.68	20.68
84	Tampa Bay Lightning	0.4	598.75	19.75
83	Tampa Bay Lightning	0.3	777.69	11.69
215	Colorado Avalanche	0.5	69.54	11.54
154	Columbus Blue Jackets	0.4	271.4	11.4
304	Vegas Golden Knights	0.4	242.15	11.15
150	Columbus Blue Jackets	0	732.06	11.06
151	Columbus Blue Jackets	0.1	732.06	11.06
271	Anaheim Ducks	0.1	724.25	10.25
303	Vegas Golden Knights	0.3	299.74	9.74
270	Anaheim Ducks	0	724.25	9.25
135	Pittsburgh Penguins	0.5	140.49	7.49

Comparison of Models

	Model 1	Model 2	Model 3
Pseudo R-squared	0.030	0.027	0.025
AIC	28683.154	28646.769	28660.104
Test accuracy	0.63	0.63	0.63

Interpreting Coefficients

	Coef.	Std.Err.
shots_on_goal	0.0057	0.0039
goalie_saves	-0.0038	0.0035
penalties	0.0298	0.0253
PIM	-0.0020	0.0074
PP_goals	-0.0377	0.0401
SH_goals	-0.0906	0.1020
bodychecks	0.0016	0.0024
faceoffs_won	0.0036	0.0067
faceoffs_%	0.0006	0.0045
empty_net_goals	0.1045	0.0874
blocked_shots	0.0064	0.0022
win_prob	0.8740	0.2603
Team_Goals	0.0464	0.0216
Opponent_Goals	0.0185	0.0183
Bet	-0.5398	0.0234

For every 1 shots_on_goal, log odds of bet outcome increases by 0.0057

Intuitively makes sense. The more shots on goal probably means you are in control of the game

For every 1 penalties, log odds of bet outcome increases by 0.0298

Not as intuitive. Could be indicative of aggressive play resulting in positive outcome with symptom of penalties.

For every 1 empty net goal, log odds of bet outcome increases by 0.1045

Intuitively makes sense. Empty net goal means opposing goalie was pulled.

Conclusion

Key Findings

There are possible patterns in historical data of NHL, where we can be profitable in long-run. (Tampa Bay)

We will need more sophisticated modelling methods to make better use of this data.

Average of stats over last 5 games may not best represent a team’s momentum, further analysis is needed.

Next Steps

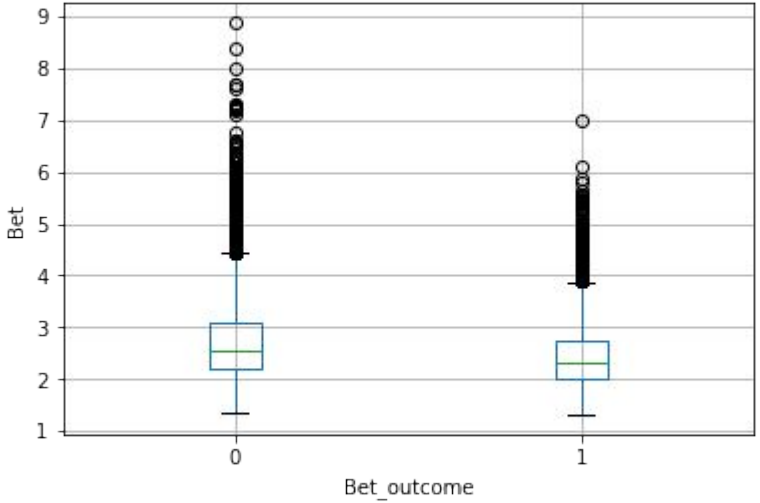
1. **Acquire more data.**
   - Individual players and their statistics are probably more significant than the team's average as a whole.
2. **Create rules to better contextualize the data.**
   - Specific data is more significant in some situations than others.
3. **Use ML to create these rules.**
   - There are probably too many situations to manually model.
   - Regression is maybe too simple for the complex model like this.

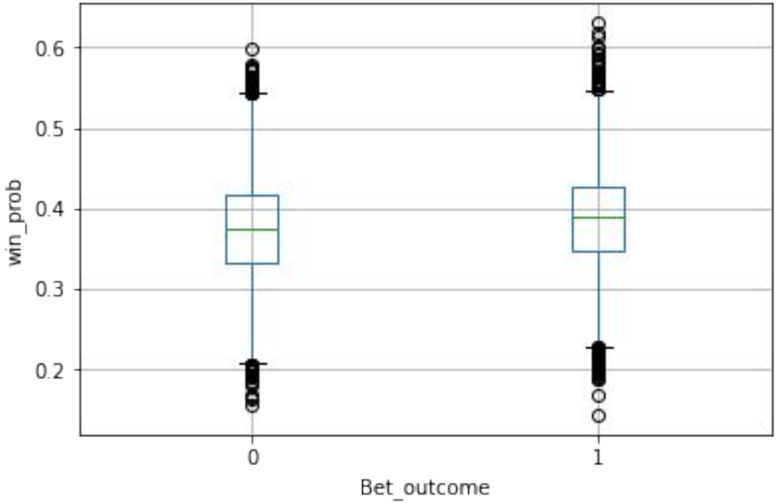
BUS 462 D100
Group Lambda

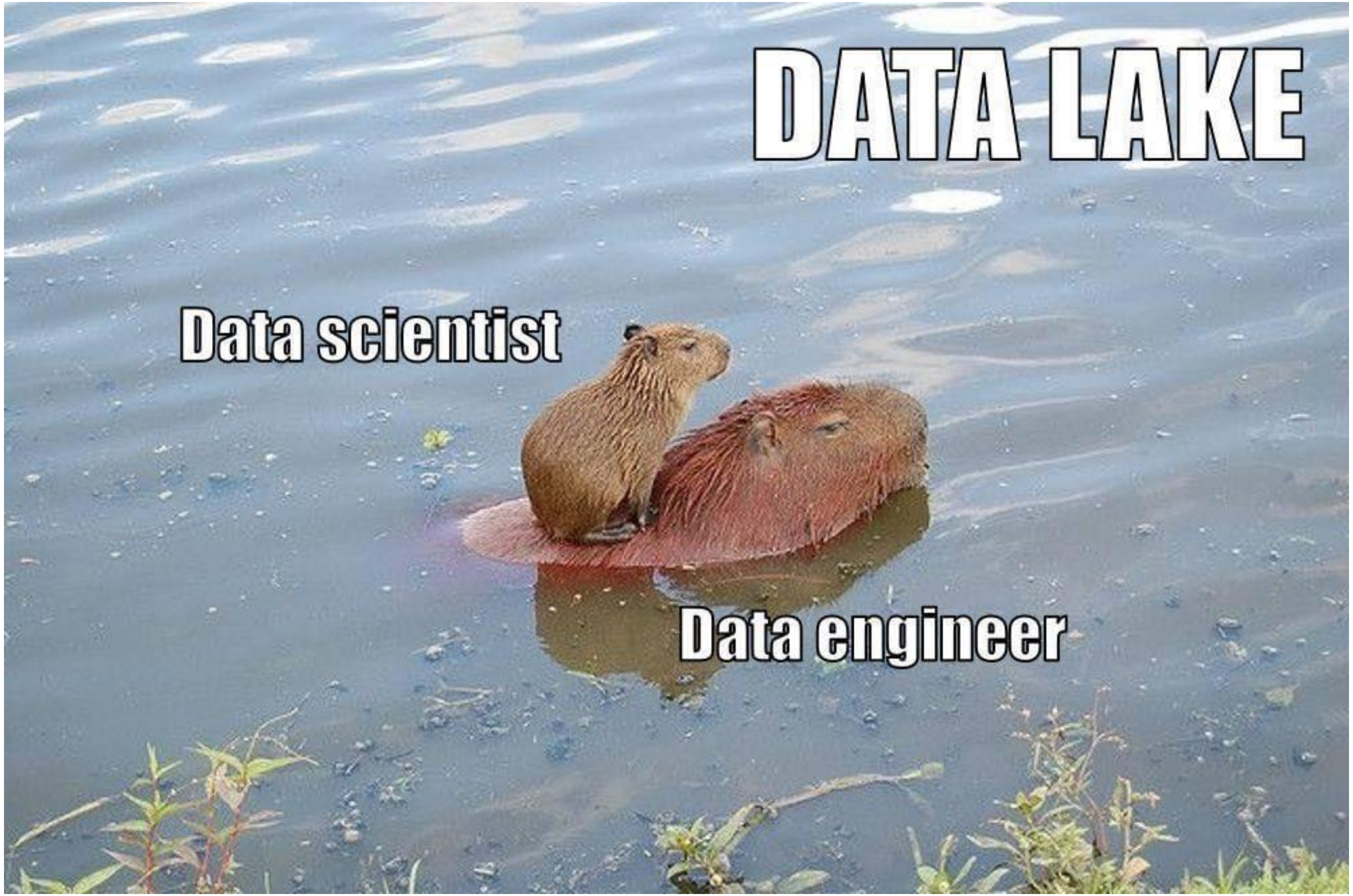
Thank You

Appendix

Box Plots


Bet
9
8
7
6
5
4
3
2
1
0
1
Bet_outcome


win_prob
0.6
0.5
0.4
0.3
0.2
0
1
Bet_outcome


DATA LAKE
Data scientist
Data engineer