Betting on the NHL

By: Dylan M, Shawn G, and Michael G

Motivation

- Increased popularity in sports gambling
  - 13 states have legal, online sports gambling
  - Large state tax revenue (source)
    - NJ -> $61,484,646 (since June 2018)
      - amount wagered -> $6,865,133,129
- As popularity ⬆, available data ⬆
- More data -> better chance of finding trends!

DRAFT KINGS
THE GAME INSIDE THE GAME.

bet365

Overarching Question

By looking over historical NHL sports book lines, can we find trends/arbitrage opportunities?

Data Source

Historical lines were found on: SportsBook Reviews

Format: CSV

NHL Seasons (by Year): 2014-15 -> 2015-16 -> 2017-18 -> 2018-19

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date	Rot	VH	Team	1st	2nd	3rd	Final	Open	Close	Puck Line		Open OU		Close OU	
2	1003	1	V	Montreal	1	1	0	2	184	210	1.5	-125	6	-110	6	-110
3	1003	2	H	Toronto	1	1	0	3	-220	-240	-1.5	105	6	-110	6	-110
4	1003	3	V	Boston	0	0	0	0	115	105	1.5	-290	5.5	-110	5.5	-120
5	1003	4	H	Washington	2	4	1	7	-135	-115	-1.5	230	5.5	-110	5.5	100
6	1003	5	V	Calgary	0	0	2	2	-130	-130	-1.5	200	5.5	-110	5.5	-115
7	1003	6	H	Vancouver	1	0	4	5	110	120	1.5	-240	5.5	-110	5.5	-105
8	1003	7	V	Anaheim	1	1	3	5	148	160	1.5	-185	5.5	-110	5.5	100
9	1003	8	H	SanJose	1	1	0	2	-170	-180	-1.5	165	5.5	-110	5.5	-120
10	1004	51	V	Washington	3	1	2	6	140	145	1.5	-200	6	-110	6	-125
11	1004	52	H	Pittsburgh	2	3	1	7	-160	-160	-1.5	170	6	-110	6	105
12	1004	53	V	NYIslanders	0	1	0	2	120	150	1.5	-200	6	-110	6.5	105
13	1004	54	H	Carolina	0	0	1	1	-140	-165	-1.5	170	6	-110	6.5	-125
14	1004	55	V	Boston	2	1	1	4	-120	-135	-1.5	185	5.5	-110	6	100
15	1004	56	H	Buffalo	0	0	0	0	100	125	1.5	-215	5.5	-110	6	-120
16	1004	57	V	Columbus	1	1	0	3	-140	-130	-1.5	210	5.5	-110	5.5	-125
17	1004	58	H	Detroit	0	2	0	2	120	120	1.5	-260	5.5	-110	5.5	105

Data Cleaning

- 6 rows out of 13340 rows have data that was input wrong (Puck Line over 1.5 and -1.5)
- To remove them, used Df.drop(condition) to remove unwanted data that would skew our data.

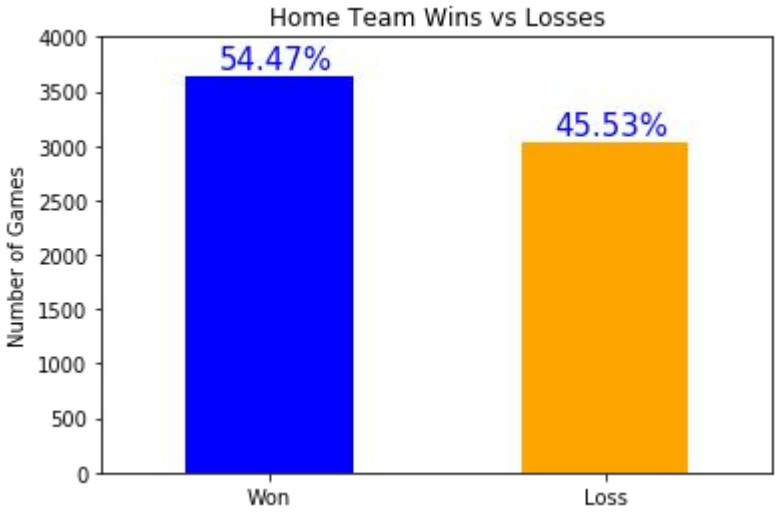
Win Rate of Home Teams

- See if there was an advantage for home teams in terms of how often they win.
- Split the data with groupby(“VH”)
- Calculate the difference between the two teams
- Set team to win or lost
- Add columns to dataframe

```
for i in range(0, home.shape[0]):
        home_dif = home['Final'].iloc[i] - away['Final'].iloc[i]
        h_dif.append(home_dif)
        if home_dif > 0:
            winloss.append("Won")
        else:
            winloss.append("Lost")
```

	VH	Team	1st	2nd	3rd	Final	Point Dif	Result
0	H	Toronto	2	0	1	3	-1	Lost
1	H	Boston	1	0	1	2	1	Won
2	H	LosAngeles	0	0	0	0	-4	Lost
3	H	Calgary	0	2	0	2	-2	Lost
4	H	Washington	1	0	0	1	-1	Lost

Results of Home Team Statistics



Total Games: 6670
Expected Value: 3335
Home Wins: 3633
Home Losses: 3037

Degrees of Freedom: 1
Critical Value: 3.84146

Chi-squared Statistic: 53.25577
P-value: 2.92826 e-13

Chi-squared test confirms statistically significant difference

Conclusion

- Teams have a statistically significant advantage when playing on their home rink
- Home team wins 54.47% of the time
- A limitation is that we are only looking at hockey and that we only have 5 years worth of data.
- This conclusion may not apply to all sports or might only be a trend of the last 5 years.
- Lot of other confounding variables can impact if a home team wins or not
    - Travel of the away team
    - Ratio of number of home fans vs number of away fans
        - How loudly each team is cheered for compared to the other (motivation)

Point Differential of Games versus Closing Odds

- How accurate are Closing Odds at predicting game outcomes?
- Benefits to predicting outcomes:
  - Smarter bets
- Find a correlation between:
  - closing odds (independent)
  - point differential (dependent)
- Data:
  - Create ‘Point Differential’ Column using by looping through Final scores

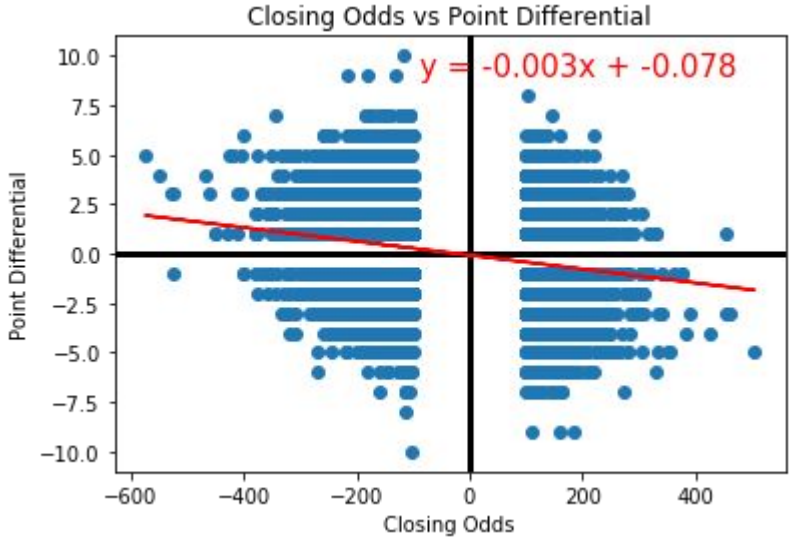
```
p_dif = []

for i in range(1, df.shape[0], 2):
        home_dif = df['Final'].iloc[i] - df['Final'].iloc[i-1]
        away_dif = df['Final'].iloc[i-1] - df['Final'].iloc[i]
        p_dif.append(away_dif)
        p_dif.append(home_dif)

df['Point_Dif'] = p_dif
```

	VH	Team	Final	Close	Point_Dif
0	V	Montreal	4	-105	1
1	H	Toronto	3	-115	-1
2	V	Philadelphia	1	165	-1
3	H	Boston	2	-185	1
4	V	SanJose	4	130	4

Results



r-value = -0.22
r-squared = 0.05

Conclusions

- There is a negative correlation between the value of the Closing Odds and the Point Differential
  - The more favored a team is to win, the more goals they win by
- A linear regression can only account for 5% of the data
  - r-squared = 0.05
- Limitations:
  - Model does not take into account style of play
    - Favored team might be defensively focused
      - Wins by less goals than offensively focused team
  - The league is competitive
    - It is not uncommon for “underdogs” to win

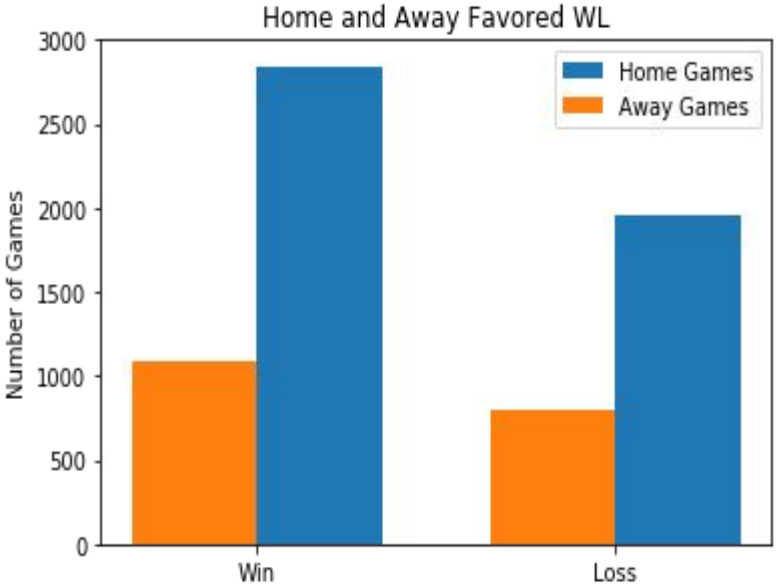
Home and Away Games for the favored team

- Find if there was a significant advantage to being the home team when favored.
- The data I used were sub tables of our original Data Frame.
  - First iterated through our original table and looked at the game (row i and i+1) and looked at the Close and whose ever was lower was declared the favorite.
  - Then from our favored table, I split it into two tables one with home favored and away favored.

```
# Create favored team table
fav = []

# Iterate through the table and see which team had the lower number in closing (Favored to win)
for i in range(1,df.shape[0], 2):
    if (df['Close'].iloc[i] < df['Close'].iloc[i-1]):
        fav.append(df.iloc[i])
    else:
        fav.append(df.iloc[i-1])
```

Results



Total Favored Home Games: 4789
Expected Value: 2394.5
Favored Home Wins: 2835
Home Losses: 1954

Degrees of Freedom: 1
Critical Value: 3.84146

Chi-squared Statistic: 162.03800
P-value: 4.05971e-37

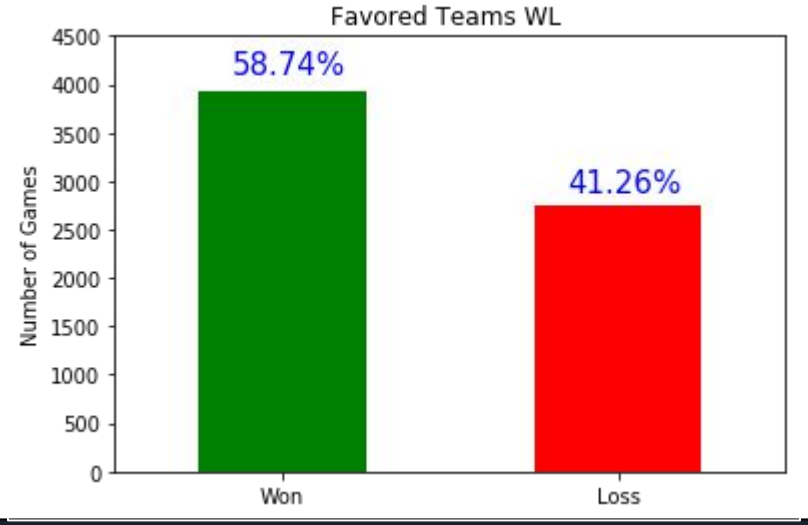
Total Favored Away Games: 1881
Expected Value: 940.5
Favored Home Wins: 1083
Home Losses: 798

Degrees of Freedom: 1
Critical Value: 3.84146

Chi-squared Statistic: 43.15904
P-value: 5.04564e-11

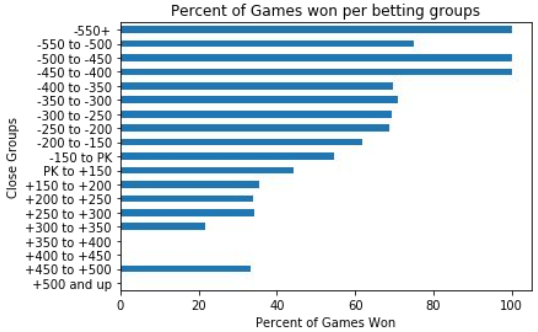
Findings / Conclusion

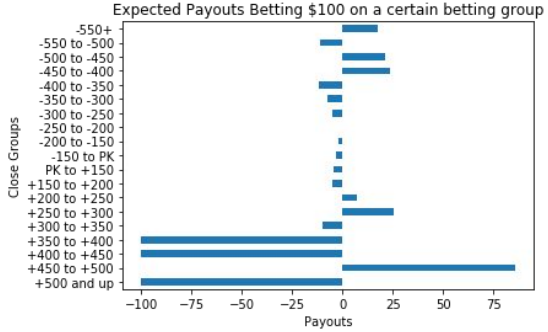
- Found that 58.74% of the time that the favored team wins.

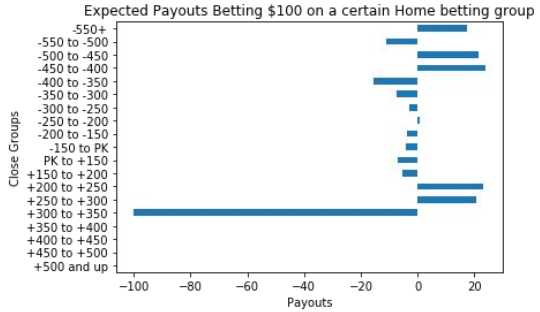


The chi test results show a number with a greater critical value and very low p value thus we can reject the null hypothesis which is that the team would win and lose the same amount while playing at home.

Win Percentage and Expected Payout by Closing Odds Groups







Expected Payout for $100 bet on the close group = (Win % * Avg Payout)+(Loss % * -$100)

Conclusion

While home teams and favored team win more, this has been taken into account in the odds

- Expected payouts are still variable/mostly negative

High Favorites and Low Underdogs are unreliable

- Not enough data to support a stance
  - Only 16 games with a team at -400 or worse odds

Best bets are:

- Moneyline: +200 to +300 (home or away)
- Puck Line: +200 to +250 (home team)

Thanks for Listening!
Any Questions?