Stat 451 Project 28

Alex Tang, Michael Beers, Joseph Beltrand, Lucas Sickels

Project Terminology

- **Sportsbook**: an establishment that takes bets on sporting events and pays out winnings
- **Moneyline**: The odds a sportsbook sets on a team or competitor to win any given match
  - A plus sign at the beginning of the number indicates the team/player is an underdog
    - Value tells us how much you would win if you were to bet $100
      - Example: if you were to bet $100 on a team that is +230 and they win, you would win $230
  - A minus sign at the beginning of the number indicates the team/player is a favorite
    - Value tells us how much you would have to bet in order to win 100
      - Example: if you were to bet $400 on a team that is -400 and they win, you would win $100

Project Terminology

- **Decimal Odds:** conversion of odds from moneyline to a decimal format
  - If moneyline is positive:
    - Value is (moneyline/100) + 1
    - Moneyline odds of +200 are 3.0 in decimal format
  - If moneyline is negative:
    - Value is (100/ -moneyline) + 1
    - Moneyline odds of -200 are 1.5 in decimal format
- **Parlay**: two or more bets are linked together to create one bet with a greater payout.
  - All the selections must win for the bet to win
  - Odds are calculated by multiplying decimal odds together for all selections

Dataset

	game_id	book_name	book_id	team_id	a_team_id	price1	price2	is_home	wl	season_year
0	41100314	Pinnacle Sports	238	1610612759	1610612760	165.0	-183.0	f	L	2011
1	41100314	5Dimes	19	1610612759	1610612760	165.0	-175.0	f	L	2011
2	41100314	Bookmaker	93	1610612759	1610612760	160.0	-190.0	f	L	2011
3	41100314	BetOnline	1096	1610612759	1610612760	165.0	-190.0	f	L	2011
4	41100314	Bovada	999996	1610612759	1610612760	155.0	-175.0	f	L	2011
...	...	...	...	...	...	...	...	...	...	...
125281	41700215	Heritage	169	1610612755	1610612738	105.0	-125.0	f	L	2017
125282	41700215	Intertops	180	1610612755	1610612738	100.0	-120.0	f	L	2017
125283	41700215	YouWager	139	1610612755	1610612738	105.0	-125.0	f	L	2017
125284	41700215	JustBet	1275	1610612755	1610612738	100.0	-120.0	f	L	2017
125285	41700215	Sportsbetting	999991	1610612755	1610612738	105.0	-125.0	f	L	2017

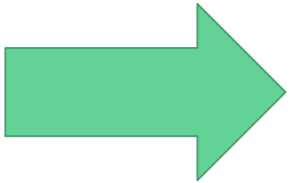
125286 rows × 10 columns

Data Cleaning

- Used only the moneyline from the most frequently used sportsbook in the dataset
- Converted “wl” column to a column where {“w” : 1, “l” : 0}
- Changed the name of “price1” to “moneyline”
- Dropped all columns other than “moneyline” and “wl”
- Filtered to only include rows where moneyline was between -500 and 500 so that our dataset was sufficiently populated

Data Cleaning

	game_id	book_name	book_id	team_id	a_team_id	price1	price2	is_home	wl	season_year
0	41100314	Pinnacle Sports	238	1610612759	1610612760	165.0	-183.0	f	L	2011
1	41100314	5Dimes	19	1610612759	1610612760	165.0	-175.0	f	L	2011
2	41100314	Bookmaker	93	1610612759	1610612760	160.0	-190.0	f	L	2011
3	41100314	BetOnline	1096	1610612759	1610612760	165.0	-190.0	f	L	2011
4	41100314	Bovada	999996	1610612759	1610612760	155.0	-175.0	f	L	2011
...	...	...	...	...	...	...	...	...	...	...
125281	41700215	Heritage	169	1610612755	1610612738	105.0	-125.0	f	L	2017
125282	41700215	Intertops	180	1610612755	1610612738	100.0	-120.0	f	L	2017
125283	41700215	YouWager	139	1610612755	1610612738	105.0	-125.0	f	L	2017
125284	41700215	JustBet	1275	1610612755	1610612738	100.0	-120.0	f	L	2017
125285	41700215	Sportsbetting	999991	1610612755	1610612738	105.0	-125.0	f	L	2017



	moneyline	wl
1	165.0	0
20	190.0	1
30	375.0	0
40	-280.0	0
50	285.0	1
...	...	...
125227	260.0	1
125237	-165.0	0
125247	-210.0	0
125257	420.0	1
125267	-245.0	1

11333 rows × 2 columns

Data Manipulation

- Shifted the moneyline by 100 towards origin to have continuity in our model since a -100 and +100 moneylines are the same
- Convert moneyline to decimal odds to make it easier to calculate parlay odds

	moneyline	wl	decimal	shifted
1	165.0	0	2.650000	65.0
20	190.0	1	2.900000	90.0
30	375.0	0	4.750000	275.0
40	-280.0	0	1.357143	-180.0
50	285.0	1	3.850000	185.0
...	...	...	...	...
125227	260.0	1	3.600000	160.0
125237	-165.0	0	1.606061	-65.0
125247	-210.0	0	1.476190	-110.0
125257	420.0	1	5.200000	320.0
125267	-245.0	1	1.408163	-145.0

Will a parlay of heavy favorites beat a straight bet of equal odds?

Method

- Logistic regression model
  - Was a bet successful or not?
- Grid search to determine optimal hyperparameter C

```
Best Model Index: 0
Best Accuracy on Validation Data: 0.6584289496910856
Best Model and Params: LogisticRegression(C=0.01, max_iter=5000), {'C': 0.01}
```

```
1 model = linear_model.LogisticRegression(C=1000)
2 X = df[["shifted"]]
3 y = df["wl"]
4 model.fit(X,y)
5 model.score(X,y)
```

```
0.6664607782581841
```

```
1 model = linear_model.LogisticRegression(C=0.01, max_iter=5000)
2 X = df[["shifted"]]
3 y = df["wl"]
4 model.fit(X,y)
5 model.score(X,y)
```

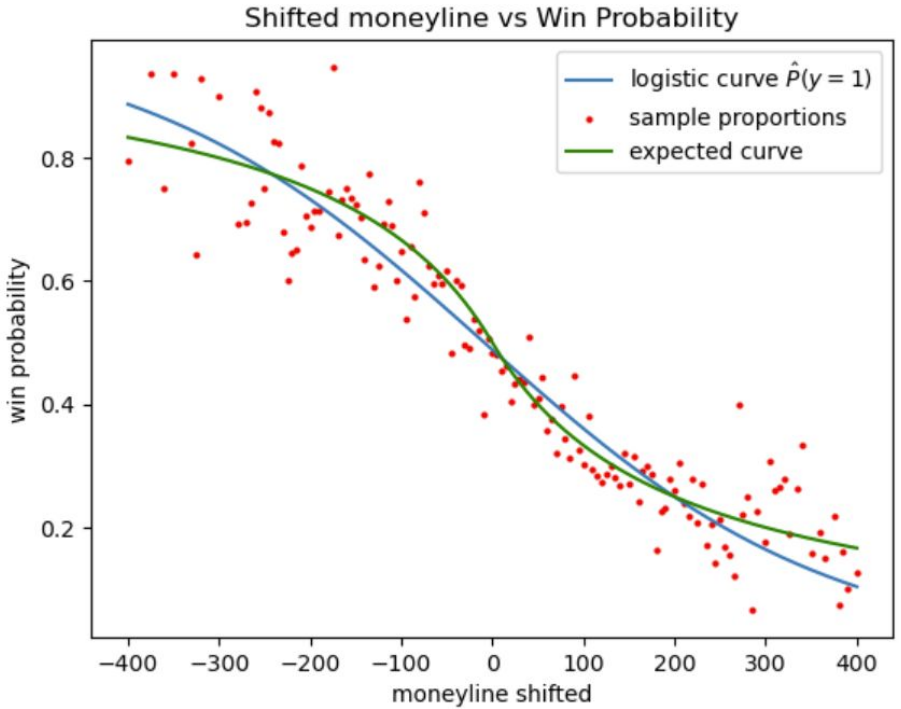
```
0.6664607782581841
```

Outcomes

- High amount of variability in sample proportions
- Logistic model curve vs. expected curve

Expected Curve

- For each moneyline, there is a formula that gives you the expected win percentage for that team
- For example, if a team's moneyline is -400, they are expected to win 80% of the time



Outcomes

Expected Value

- Used our model to get the predicted win probability for each moneyline
- Calculated the number of bets placed in a parlay needed to reach -100 odds for each moneyline
- Calculated expected value for each parlay

	moneyline	decimal	shifted	expected_w%	predicted_w_prob	bets	expected_value
0	-500.0	1.200000	-400.0	0.833333	0.887088	3.801784	0.268267
1	-499.0	1.200401	-399.0	0.833055	0.886559	3.794833	0.266455
2	-498.0	1.200803	-398.0	0.832776	0.886028	3.787882	0.264642
3	-497.0	1.201207	-397.0	0.832496	0.885495	3.780932	0.262829
4	-496.0	1.201613	-396.0	0.832215	0.884960	3.773981	0.261016
...	...	...	...	...	...	...	...
797	496.0	5.960000	396.0	0.167785	0.105976	0.388302	-0.163403
798	497.0	5.970000	397.0	0.167504	0.105478	0.387938	-0.164247
799	498.0	5.980000	398.0	0.167224	0.104982	0.387575	-0.165091
800	499.0	5.990000	399.0	0.166945	0.104488	0.387213	-0.165935
801	500.0	6.000000	400.0	0.166667	0.103996	0.386853	-0.166778

802 rows × 7 columns

Outcome

- Expected Value of parlays with odds better than -343 is greater than zero
- Graph is based on expected value from betting $1 on the parlay

- A parlay of heavy odds is more probable to hit than a parlay of straight odds based on our logistic regression model