Beat the Book

Zach Norman   George Doty   Filip Tomovski   Travis DeBruyn

ABSTRACT

This project intends to, at a baseline, inform users' sports bets. Each sportsbook has different biases and odds that depend on a multitude of factors. With our visualization, we will convey these biases in a concise and readable manner to allow users to quickly find the necessary information when they want to place a bet. Our information needs to be quickly accessible and digestible as betting odds are quickly changing. The three main goals of our visualization are: compare the general performance of user bets by sportsbook, compare specific bet performance across multiple sportsbooks, and to allow users to understand which sportsbook to use (in a more broad sense) for which sports. These domain goals are important to support as sportsbooks are very good at masking their biases and intentions as seen in the related work section of this paper.

1 INTRODUCTION

As stated previously, this project is intended to inform any bettors as they bet. Each sportsbook has unique characteristics that change the way that users interact with them. With our tool, we are aiming to aid users in their interactions across multiple sportsbooks, and help them decide which to use for their current betting needs. Comparing the general performance of user bets by sportsbook is the first task that we want users to be able to complete using our tool. Users should be able to get a general idea for how well their fellow bettors tend to perform on a given sportsbook. We also want bettors to be able to compare specific bet performance across multiple sportsbooks. We intend for users to be able to find their desired bet, and see the odds of this bet across multiple sportsbooks. For instance, if I think that the Knicks will win by less than 3.5 points tonight, then our visualization tool should be able to show which sportsbook will give me the biggest return on my bet. Finally, our tool should allow users to gain a general understanding of where individual sportsbooks' biases lie. Some sportsbooks may favor users when they bet on home teams, while others may favor user bets in basketball. We want our users to be able to gain this information so that they are not always beholden to our visualization tool. This information will also serve users who are not quite as interested in spending time finding the perfect bet, and would just like to generally increase their odds. Our expected value of specific bets will often come from the Pinnacle sportsbook as it would be outside the scope of this project to try to create our own odds for hundreds of different bets. Pinnacle is widely regarded as the sportsbook with the most accurate lines, and is generally known to be the most difficult to beat.

2 RELATED WORK

There is currently a fair bit of prior research when it comes to sports books and bettor bias. Chin [1] utilizes bar charts to visualize the proportion of bets placed on random teams, teams with the advantage, and east coast teams versus west coast teams in afternoon and evening games. Since the human body typically performs best in the late afternoon, the "circadian advantage" can help explain the difference of performance in NFL games, especially when teams based in different time zones play each other. Sports books take

advantage of people not knowing about this by setting odds that bait bettors into making losing wagers on the disadvantaged team. If informed bettors place restricted wagers, the sports books encourage overbetting by offer-spreading that account to uninformed bettors, who see the wagers as being relatively cheap. When looking at bets across sports books, this method of visualization is a good way to see all options compared against each other to make the best-informed decision. Rodney and Weinbach [2] look at how bettors are biased when it comes to NHL bets, utilizing line charts to show the trend of percentage bet on favorite as favor odds increase or decrease, color-coding for bet percentage based on odds, actual bet percentage, and actual favorite win percentage. Bettors have a preference for favorites, and especially road favorites, though sportsbooks do not always price to exploit known biases. Biases for road favorites are not always priced into odds. Instead, prices are set based on game outcome forecasts. The overall idea is that this discourages entry by informed traders, therefore allowing for collecting commission on losing bets and ensuring long-run viability for sportsbooks. Visualizing findings using line charts could also be utilized when showing trends over time between sports books and/or various sports, as well as for comparison with multiple color-coded trend lines overlaid on each other. In general, by taking this research into account, considering betting biases, and displaying bets across sportsbooks, bettors could see what is the best bet to place without bias.

3 USE CASE

For this project we want to make a visualization around sports betting. Specifically, player and game odds for the NBA. Finding the right site to bet on is extremely difficult. With odds constantly changing and slight discrepancies between bookies locking in a bet with the best odds is nearly impossible. However, this does not have to be the case! By using a bet API, we can download, clean, and visualize bets across different betting sites. By creating a visualization to show what bookie to bet with sports betters can be comfortable knowing they got the best odds.

There is a clear use case for this project. To beat the book.

With sports betting moving into Mass this is the perfect time to invest in this. By web scraping betting data, we can create a seamless visualization by plotting bets in a jitter plot by expected value. Using different tools in D3 we can use color to have the best and worst bets stand out. We can also include filters on bookie, game, and even dates.

As you hang out with your friends on Saturday afternoon you decide to go watch some games at a bar later this evening. To make the games more interesting you and your friends want to place some bets. After booting up our visualization, updating the data, and filtering to games this evening you see that draft kings (marked in red) has the Nets by 3.5 points where everyone else has them winning by 3. You swap to Draft Kings and lock in Nets over 3.5. The Nets win by 3. You don't push, you win!

4 DATA

The data is collected by the odds api (https://the-odds-api.com/). It is currently set up to only pull today's games but the dataset could be expanded. It is unknown how they get it because after trying to webscape betting sites there is no way that is their solution. Many different betting sites offer paid API usage or have a minimum

balance requirement to use their API so they probably just get it from that. However, regardless of how they get the data they have comprehensive betting data across multiple sports and leagues that is up to date when I check the sites manually. It is strictly spreads and money line odds so there should be no bias as sport betting companies are trying to make money. There are often deals that different companies put out to get new bettors. One of the main reasons for this project is to find the best ones and exploit them. There are no missing values nor true outliers.

After reading in the data using a Jupyter notebook, the data was cleaned using python before exporting it to a .csv. For the cleaning in Python, first the game column was split into home and away teams to avoid confusion, and then renamed the columns in accordance with home and away. Numerous duplicate columns were also deleted and renamed so they made more sense. Lastly, the data was grouped by game ID and then Bookie so bets are sorted by game and the bookies spread and ML bets are next to each other.

5 DESIGN PROCESS
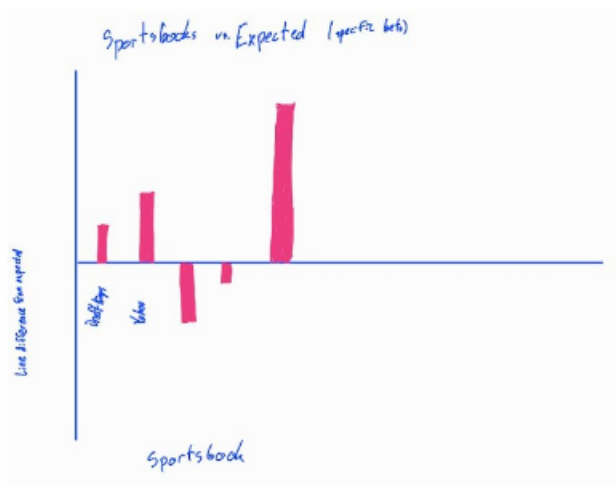
The first sketch was our initial idea to compare the performance of different sportsbooks against an expected value. While this was helpful, it was missing a great deal of information.
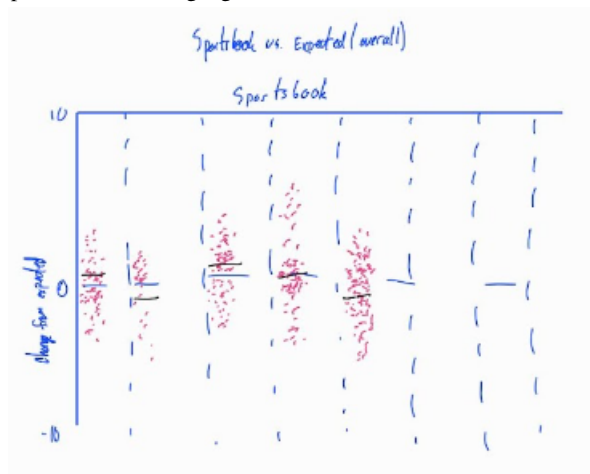
Our second sketch was our first attempt at a jitter plot. We intend to use this to show the distribution of the change in odds from a specific sportsbook to the expected odds. This proved to be the most effective for our efforts, but we still needed to improve on its clarity for our final sketch.
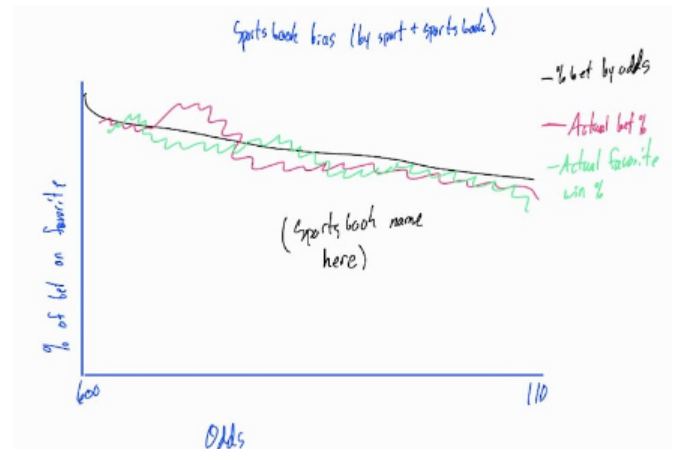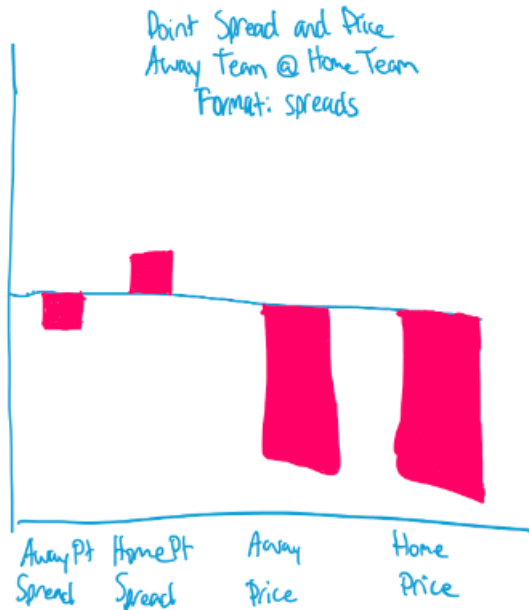

Our third sketch was an attempt to create a plot similar to one we had seen in work similar to ours. It could prove useful to show the bias of an individual sportsbook, but maybe this could be a secondary plot to our main jitter plot as is often seen in visualizations from the Viz of the Day.
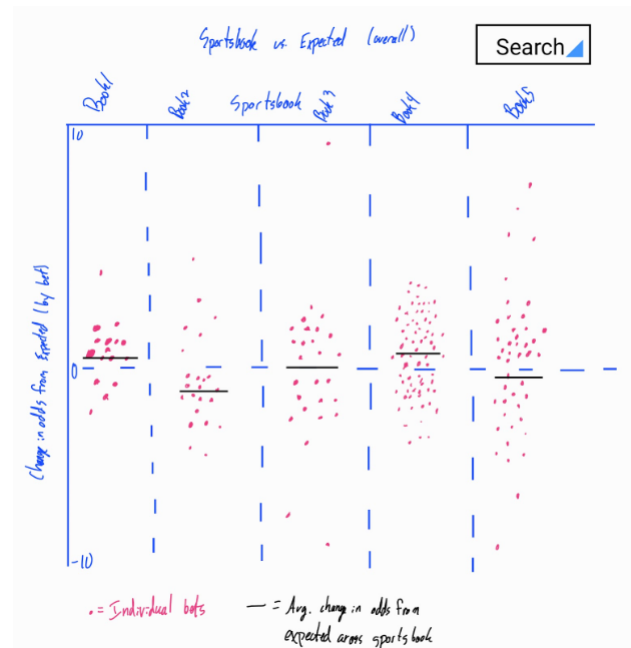
Our final sketch utilized the jitter plot idea that we originally sketched out. We decided to go with a similar color scheme as of the limitations of our drawing software, but we are still considering changing the data points in each individual sportsbook (having one color per sportsbook). One other limitation of our drawing software was that we were unable to change the opacity of the data points. We want to increase or lower opacity depending on how densely populated with points an area of the graph is. This is to help users better understand the placement of the weighted average represented by the black bar. Dots were chosen as this is a rough sketch, but utilizing a shape like a hexagon in conjunction with opacity would also help with the visual clarity of the tool. Also, we intend to have individual nodes be selectable by users so that they may see which bets are represented by each node. When a node is selected, it should highlight all related nodes in the other sportsbooks to allow quick comparison between sportsbooks. Using the search bar in the top right corner can also aid in selecting the desired bets.

Point Spread and Price
Away Team @ Home Team
Format: spreads

Away Pt Spread   Home Pt Spread   Away Price   Home Price

Visualization

Sportsbooks

DraftKings
MyBookie.ag
BetOnline.ag
LowVig.ag
WynnBET
BetUS
PointsBet (US)
FanDuel
BetMGM
Bovada
Barstool Sportsbook
BetRivers
TwinSpires

Difference Between Average Sportsbook Spread

Filter by state
Check all | Uncheck all

Point Spread and Price
Dallas Mavericks @ Sacramento Kings, BetMGM
Format: spreads

We didn't have many ideas for the second visualization at first, but as we began to develop our visualization, we realized that there was more information about each bet a user would likely be interested in, and it would be a lot of information to fit in a tooltip. Therefore, we made this bar chart to show point spreads and prices for a game a user selects, by clicking on a point. 0 is in the middle, and the bar charts go up or down depending on whether it's a positive or negative value, and the length of the bar represents the value. A tooltip when hovering over each bar would also show the exact value.

As we began to implement our visualizations, we decided to change our original idea of a search bar to state filters, as we thought this would be the most useful way to filter data for our users. As we want to encourage legal betting, users can select their state (or multiple states they're interested in), and the visualization will only show the sports books that are active in their selected states. After implementing our two visualizations with filtering, we then conducted usability testing. In our usability testing, we first asked users about visualization 1, and to identify which sportsbooks have worse bets, to make sure they understand what it is showing and can parse information from it. We then asked them to look at visualization 2 and to give values for prices and point spreads, as this is the core of what visualization 2 is showing, so we again want to make sure users understand what the visualization is showing and how to get information from it. Finally, we asked about the usability of the state filters, such as whether it's clear what the filters do and how it affects or changes the data. For all of these, we timed users completing the tasks and answering the questions, as this would tell us that the visualization isn't easy to understand at a glance if the users take a long time. What we found is that users found it unclear what the y-axis was representing in visualization 1 and how it related to good vs bad bets, users noticed the state filters working but weren't sure whether it was for states the games were in or that sportsbooks were in, and that it could be difficult parse data for each sportsbook. From this, we changed the y-axis title of visualization 1 to be clearer, added a label to the state filters to clarify what they filter, and added a legend and lines separating each sportsbook in visualization 1. In addition, while developing our visualizations, we were finding that it was tedious to uncheck every box if you only care about one state, as our visualization defaults to having all states shown, so we added buttons to easily uncheck or check every state.
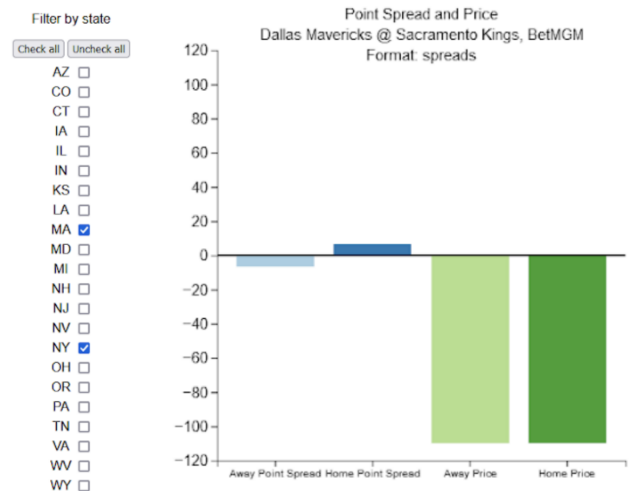
Our final visualization works in two parts. The beeswarm plot on the left separates data points out by sportsbook, and then arranges them based on their distance from the average spread of that game. The list of states in the middle allows users to choose the states in which they would bet. On the right, there is a chart that is linked to the chart on the left that shows the point spread and betting line of the users' chosen game.

Step by step use guide:

1. Select the state in which you will be placing a bet from the list in the middle. 'Check all' or 'Uncheck all' buttons are provided for added usability.

2. Click one of the bets that you'd like to investigate in the beeswarm to bring up its prices on the bar chart.

3. Evaluate the bet's performance based on its distance from the mean. A positive y-value means that a bet's spread is greater than the average spread, and the opposite is true for a negative y-value.

4. Bet responsibly!

7 DISCUSSION

Our visualization tool helps a great deal when trying to wade through the mass of sportsbooks aiming to capitalize on the new wave of laws being passed pertaining to sports betting. While we realize that our beeswarm visualization can be difficult to understand, it becomes much easier to understand when interactive features are added.

The domain problem that we set out to solve was the difficulty of understanding the nuances of each sportsbook. While we may not have completely solved this problem, we believe that our visualization tool can help mitigate it. This problem is very complex, and sportsbooks do not provide enough data to the public that would allow this problem to be totally solved, at least in the scope of this project. When we initially decided on a domain problem, we did not consider the experience or knowledge level of potential users of our tool. This led to some us overlooking the usability of our tool to those users who are new to sports betting.

Some improvements to our tool could include changing the background to a gradient that changes depending on the improvement that any individual bet makes over the average spread. This could aid newer users in understanding which bets can be more lucrative in different scenarios.

If we were to make improvements to our tool in the future, we could create a machine learning model to understand the biases of different sportsbooks. This model would allow us to create an expected spread for each bet in each sportsbook, and would allow for newer users to get a much better grasp on the topic of sports betting.

8 CONCLUSION

For this project we created a visualization to help active sports bettors find the best sportsbook to use. Because of the different biases and odds that each sportsbook can have, we created a visualization tool to help users decipher the differences in these biases. We created a beeswarm plot that links with a barchart to help users see the difference in a sportsbook's price when compared to the mean of all sportsbooks for that specific bet. These charts allow users to quickly distinguish high-value bets from low-value ones, and allow them to budget their funds in the most informed fashion. While not perfect, we hope that our tool can serve as an aid for new and experienced bettors alike when trying to navigate the confusing world of NBA sports betting.

A DATA ABSTRACTION

Each Item in the dataset represents a different bookies odds for each game split up by ML and Spread. Game ID, bet_format, and site_title make of the PK. Game ID (string) specifies what game the bet is related to. Bet format (currently a string but might make an int or bool depending on how I want to filter) is either h2h or spreads to show the different bet types. Point_spread_away (float) is always 0 for ML or how many points the away team is getting or giving for spreads. Point_spread_home (float) is the same thing but for the home team. Site_title (string) is what bookie the bet is from. Home_team (string) is the home team. Away_team (string) is the away team. Price away (Currently float, might make int because all end in .0) is what the odds you will get when betting on the away team for h2h bets, when it is spread it is defaulted to -110. Price home is the same thing but for the home team.

B TASK ABSTRACTION

Users want to compare the general performance of other users by sportsbook when choosing where to bet.

Using Munzner's taxonomy, we can aggregate bets that have similar changes in odds compared to the expected odds. Using one node to represent multiple bets can help users quickly identify the required information. Each individual bet is important so we will not be filtering.

Users want to compare individual bets across multiple sportsbooks, and see which usually offers the most favorable odds.

Completing this task abstraction will require us to optionally filter our items. We will allow users to search for specific bets via a dropdown menu, and filter out bets that they do not want to see. This allows for more immediately digestible information for our users.

Users want to understand which sportsbook to use when betting on different sports.

This is a general task that abstracts the previous two. We will again use Munzner's taxonomy to aggregate bets and change the opacity of the aggregated nodes to denote the number of bets in each node. Also, we will create a weighted average of the bets and denote this with a black bar across the visualization. Having this black bar will allow users to immediately comprehend the desired information.

REFERENCES

[1] D. M. Chin. A TEST OF UNBIASEDNESS AND SPORTS BOOK PROFITS IN THE NFL POINT SPREAD BETTING MARKET USING CIRCADIAN ADVANTAGE. *The journal of gambling business and economics*, 5(3):15–26, 2013. doi: 10.5750/jgbe.v5i3.571

[2] R. J. Paul and A. P. Weinbach. Sportsbook pricing and the behavioral biases of bettors in the NHL. *Journal of economics and finance*, 36(1):123–135, 2012. doi: 10.1007/s12197-009-9112-4.