

Classification of Online Sports Gambling Patterns

Dan MacLean

11/16/2021

Objective

To create classification among online sports gambling behavior to detect common patterns.

Background

In 2021 the sports gambling industry took a leap in the US, legalizing online sportsbooks allowing millions of new gamblers to place wagers instantly without being near a physical sportsbook. This is a new experience for both the gamblers and the book keepers. In this Analysis we are looking to find patterns to classify bettors into common wagering behaviors.

This data is sourced from a study done in europe from bwin Interactive Entertainment. A majority (58%) of this data is from Germany. This data is sourced from 46,339 bettors and their habits over an 18 month period from Feb 1, 2005 to August 31, 2006. This data contains wagers, deposits, profits and whether bets were fixed (placed before the beginning of an event) or live (placed during an event). We will be utilizing an unsupervised machine learning method to discover patterns in the behaviors of these bwin patrons.

Data importing and cleaning

With the data available in this set we see that we have an entry of all bets per bettor in the collection period with their aggregated totals for amount wagered, wins and total bets. All these variables have filters based on fixed and live wagers. In order to find patterns in behavior we will be using the average amount per wager and average NET of each wager. These will tell us how much bettors are putting on each game and if they are on average taking in a profit or not and to what magnitude. Later we will separate these 3 measures of behavior and split by those who solely bet on Fixed events, Live events and those who do both.

Data Exploration

Table 1: Countries in data by proportion

CountryID	n	Percent_of_sample
276	26649	57.508794
792	2795	6.031636
616	2770	5.977686

Here we see using the country codes (ISO 3166-1) in the data set that a majority of the respondents are from Germany (58%) followed by Turkey and Poland (6%).

Table 2: Proportions of Bettor types

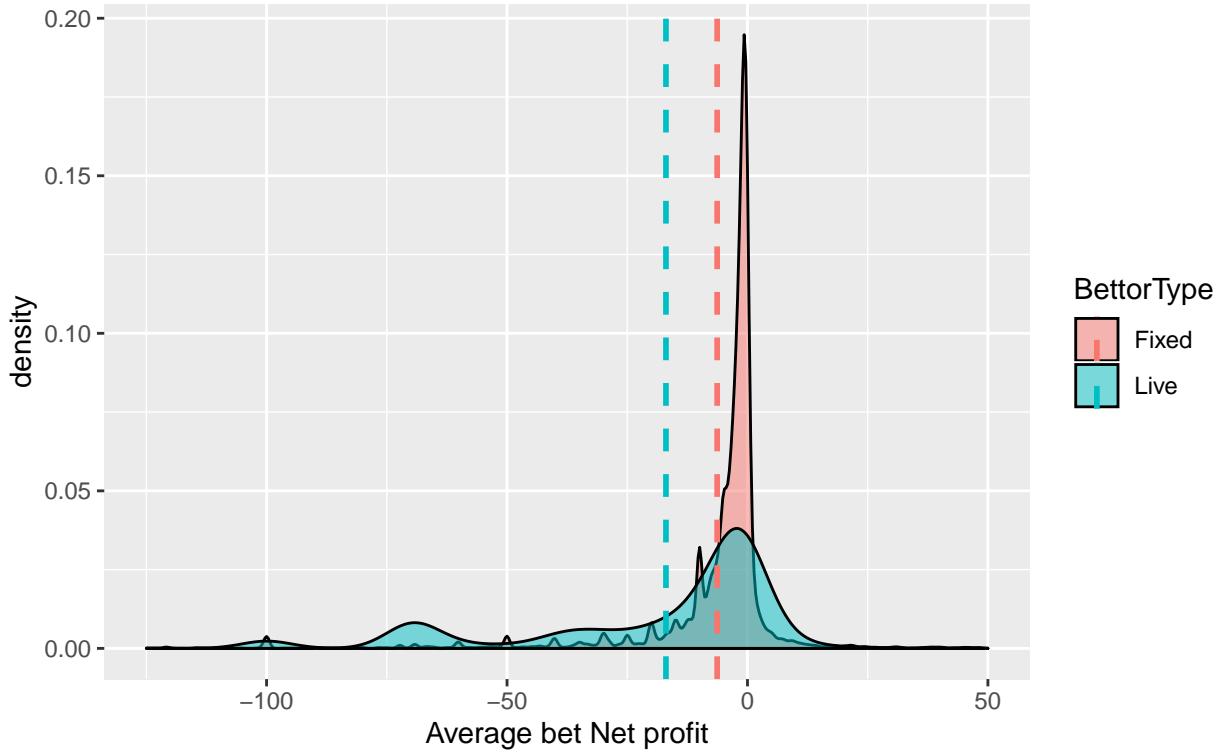
Live	Live_Who_Bet_Fixed	Fixed	Fixed_Who_Bet_Live	Both
0.6771402	0.9765441	0.9841171	0.6719295	0.6612573

From the main data set we can see that 68% of our sample has placed a live bet during a sporting event and 98% have bet on an event before it has started. Looking into the cross usage we see that out of those who place live bet, 97% of these people also place fixed bets. This is much higher compared to those who bet fixed wagers which only 67% of people also bet live. We will look into overall trends as well as those strategies among fixed vs live betting approaches.

To see the overall performance and behavior of these two types of wagers we look to plot the density of each type and their average return per wager. As expected, on average each type of wager end up negative but the live betting style seems like it is less kind to those who take part averaging -17 Euros per wager. Fixed bettors have less of a spread on their net being more centered towards 0 still averaging -6.3 Euros per wager.

Density Plot of those who Live bet vs Fixed

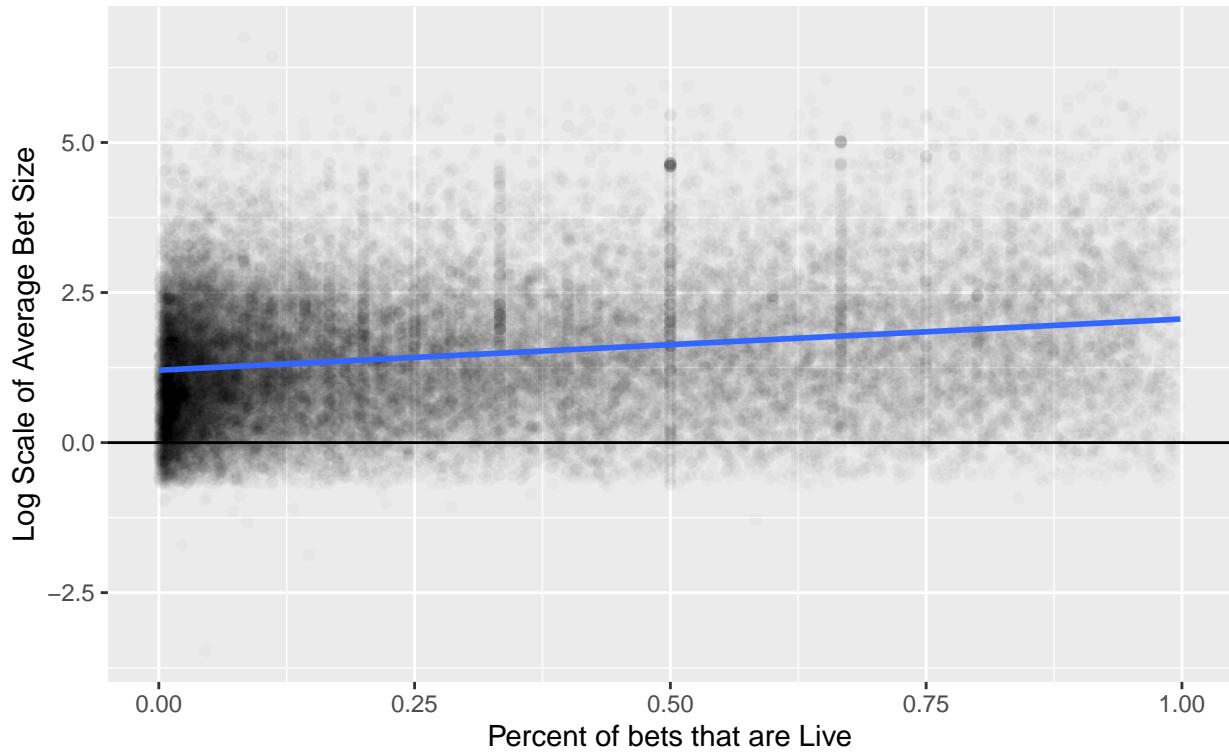
(Those who wager both types are excluded here)



From this chart we can conclude that those who partake in live wagers are willing to take on extra risk to profit, we see that the live betting style has a longer tail on the negative end of the chart but also has more on the positive side. This flatter curve summarizes the variance of profit/loss a live wager can take versus a fixed bet which appears to be more stable as decisions are made in advance of a match starting opposed to during.

To confirm this is true in the instance of those who place both types of wagers, the plot of average stake of a bet vs the percentage of live bets placed was created below. From the results of this plot we can see from the blue line of best fit has a positive slope and we can conclude that there is a positive relationship with wager size and live betting among those who partake in both types of wagers.

Plot of Average bet size vs percentage of live bets placed
 Among those who place both types of bets



Now that we have a general idea of what our data contains and a couple trends to work off of we will take the data and conduct a cluster analysis using the elbow method to determine the most definitive number of distinct behaviors (clusters) to use. This number will be used to classify bettors when we run our kmeans clustering algorithm to assign these general behaviors to bettors in our data. This will be done among those who exclusively bet on live events, those who exclusively bet on fixed events and those who bet on both.

Cluster determination

To run the cluster analysis 3 sets of data filtered by Live, Fixed and Both gambler types have been created. This data has been scaled so all variables being used to find patterns within these groups now have a mean of 0 and a standard deviation of 1. In the “Both” data we have included the additional variable that contains the percentage of live bets this person takes.



After running the elbow method to determine the number of definitive behaviors, we can visually see on the charts where the number of clusters reduces in the distance between points in the clusters assigned. We look to pick the point where there is a break or “elbow” in the line, this is where increasing the number of clusters has less of an impact and overfits to the data. Marked by the red lines for our “both” category we choose 4 clusters. For the fixed and live exclusive sets we will also use a 4 cluster approach. Now we will assign these clusters back to our data to find characteristic of each group.

Cluster characteristics

After assigning our cluster labels back to our original data set we can now pull summary statistics of the 4 groups we created for Live, Fixed and Both types of wagers.

Both - group descriptions:

Table 3: Both - Cluster Group Summary Stats

Cluster	Percent	n	Bets_Made	Wager_Size	Wager_Net	Percent_LiveBets
1	0.6484237	19869	298.5739	6.198387	-1.0237763	0.1390557
2	0.0150773	462	7053.6970	8.021747	-0.6947153	0.4625088
3	0.0179166	549	253.9253	130.740677	-22.3428779	0.4930990
4	0.3185823	9762	387.9536	10.421587	-1.0965369	0.6584227

In our group of those who have wagered on both Live and Fixed events we can see initially that group 1 is the largest and 4 is large compared to groups 2 and 3. Based on this table we can now further describe the trends within this group. 1. *Majority Fixed Bettor* - This group favors the fixed wagoer more only allotting

14% of bets to live action. This group makes up most of the people who do partake in both types of bets which checkes out as a majority of our sample bets on fixed wagers. This group Is pretty averse across the board. With volume, size and profit. It is on the lower end of wager size but nothing too far from the other groups. 2. *The cautious volume shooter* - This group is the smallest but does the best on average of returns from wagers, with this being said their wager size is comparable to groups 1 and 4 but their volume of bets is by far the highest. This group wagers the most, loses the lease on average and has about an even split when it comes to live/fixed wagers. We can conlue this group wagers an average amount on events that are likely to happen. 3. *The occassional big spender* - This group is about the same size as group 2 and makes up a small portion of our sample. This group makes the least amount of wagers, not by far in comparison to group 1 but also wagers by far the most per bet and at the same time loses the most on average per bet made. This group alike group 2 splits live/fixed wagers evenly. These gamblers appear to take some big swings and on average take a bigger hit with a loss in exchange at the opportunity of a larger payout. 4. *Majority Live bettor* - This group makes up 31% of those who bet both wager types and has the second most wagers placed, still far behind group 2. This group however, has the highest portion of live bets made with 65%, this makes sense as we found Live bettors tend to wager more.

Live - group descriptions:

Table 4: Live - Cluster Group Summary Stats

Cluster	Percent	n	Bets_Made	Wager_Size	Wager_Net
1	0.0095109	7	524.285714	16.96273	-0.6608290
2	0.5774457	425	18.590588	9.86841	-0.7175247
3	0.2269022	167	2.652695	89.18015	-56.5600622
4	0.1861413	137	6.795620	43.17586	-19.8473591

1. *Live bet enthusiast* - This group is very small but distict, this small group wagers a significant amount more but has an average to smaller bet size and has thr best returns on average.
2. *Common Live bettor* - This group contains the majority of this type ot bettor, on averse placing 18 bets with the smaller wager size out of the groups and haing the second best returns.
3. *Big game bettor* - This group has the least wagers placed on average with the largest wager size and also the worst return. This group may just place wagers on large events when they are wathing them to make it more interesting. Unfortunately this group doesnt tend to win very much.
4. *Occasional Live bettor* - This group, alike the big game bettors has a larger bet size with not so goo returns, they place more wagers but have similar results. This group is a toned down version of the previous with more bet volume.

Fixed - group descriptions:

Table 5: Fixed - Cluster Group Summary Stats

Cluster	Percent	n	Bets_Made	Wager_Size	Wager_Net
1	0.0072188	108	2298.657407	3.235759	-0.5610316
2	0.0441147	660	11.166667	115.249781	-63.6624747
3	0.0012700	19	1.473684	657.168876	-657.1688763
4	0.9473966	14174	63.320375	9.696625	-2.8020924

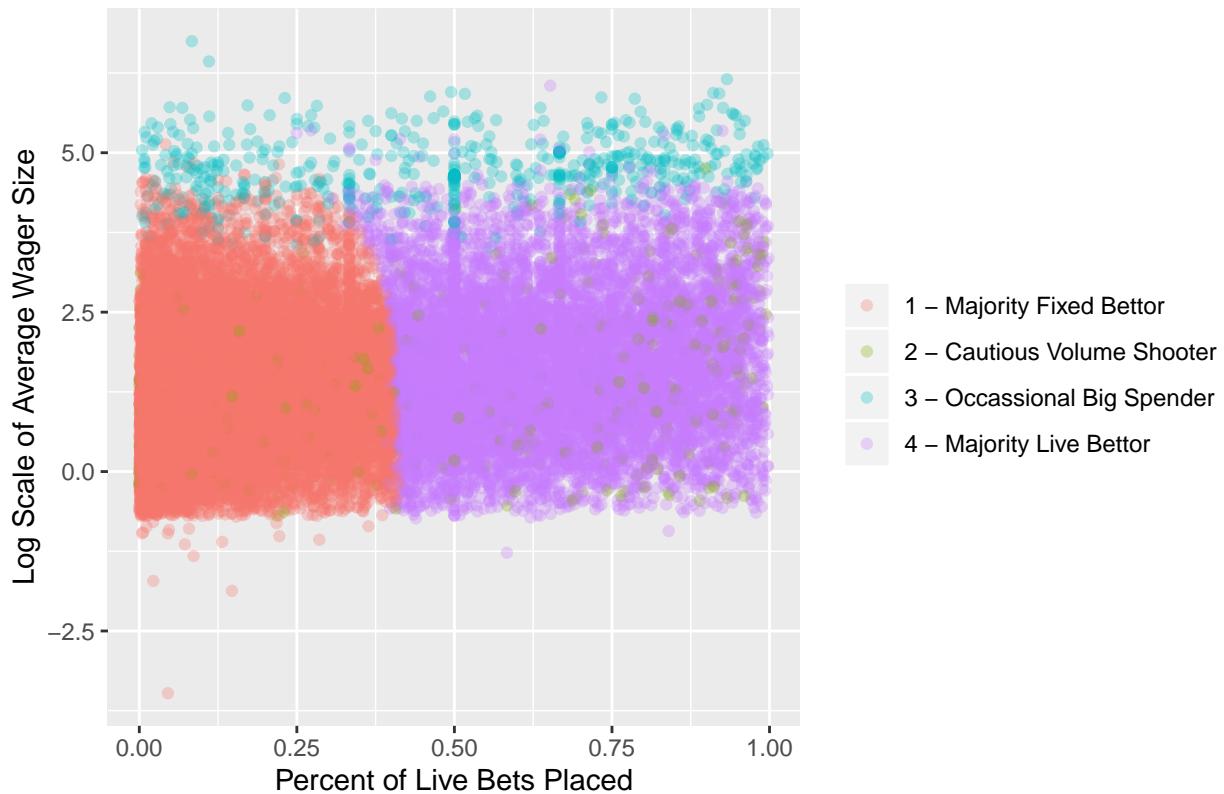
1. *Bets Everything* - This group appears to bet on a significant amount of events compared to others in the group. With the smallest wager size and best returns this group appears to wager smaller amounts to have skin in whatever it is they are watching. They are a very small portion of the sample.
2. *Causal risk takers* - This group has the second largest wager size and the second worst returns. They do not bet often but when they do they have large stakes.

3. *Rare risk takers* - A far less common group that makes a very small portion of the sample. These rare gamblers barely ever place a bet but have an average stake size that makes them very distinct from the rest of bettors. With this risk comes the worst return of all fixed bettors.
4. *Common Fixed bettors* - This group contains the large majority of fixed bettors and is average across the board in wager volume, size and returns. Making up for 95% of fixed bettors, most bettors in this category share similar attributes, and there are the extremes in the previous 3 groups.

Cluster visualization

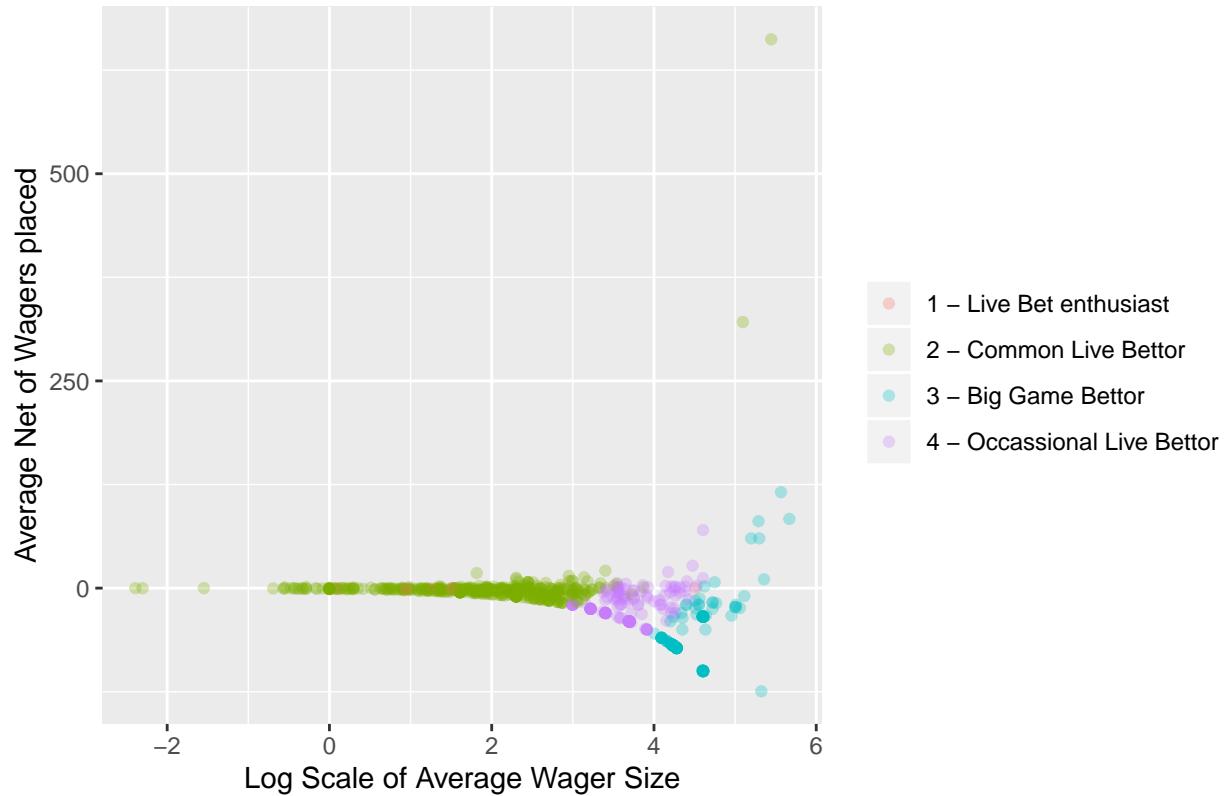
From the plots we can visualize the defining aspects of each subgrouping and different behaviors of gamblers for a visual understanding of the relationships between variables for these groups.

Clusters of Both Bet types



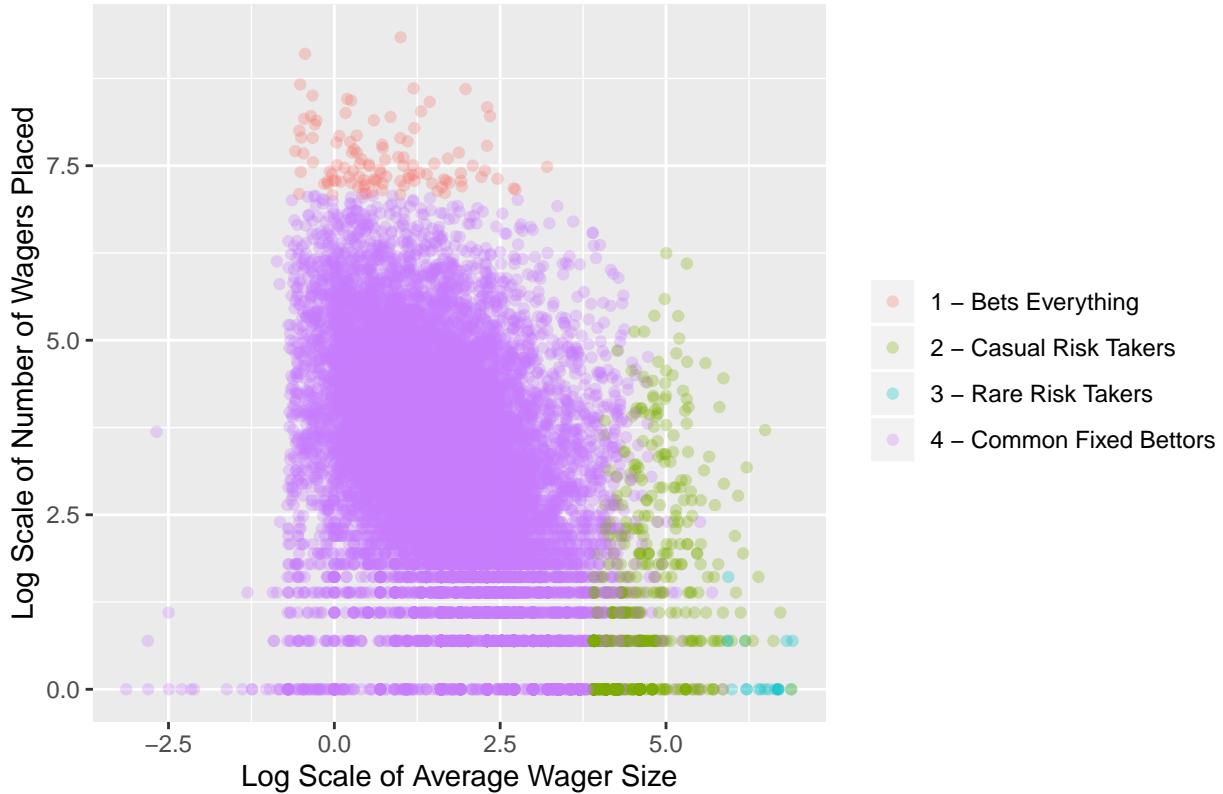
Those who use both types of bets we can see the difference in how groups 1 and 4 differ in % of live bets placed. By crossing this with average wager size we can see that group 3 is separated from 1 and 4 by their wager size. Group 2 looks to get lost in this chart as it is the smallest and shares similar traits with groups 1 and 4, the difference which is not captured in this chart is that group 2 has a much larger volume of wagers placed.

Clusters of Live Bettors



In the Live cluster chart we see something similar as group 1 is the smallest and appear to blend into the other groups. Similar to the above chart, this group has a much higher volume of bets which is not shown. What is shown in chart 2 for Live bettors is the relationship between the wager size and the profitability of each group. We see the more common group 2 has less variance and does not haе a large wager size, compared to groups 4 and 3 we see as the wager size increases the profitability falls. This could be a result of placing large wagers on long shot outcomes that are less likely to occur.

Clusters of Fixed Bettors



The fixed Cluster chart shows the volume of wagers vs the size of wagers placed. This helps us visualize all 4 clusters as we can see group 1 has the highest volume of bets spread across wager sizes, group 3 is the opposite betting large amounts with a low volume. Group 4 takes up most of this chart as it is the majority of this category having a range of wagers placed and wager size. We can see off of group 4 to the right is group 2 who place slightly less bets but are willing to take on more risk and wager higher amounts.

Cluster analysis and conclusions

By performing this analysis we were able to classify types of gamblers, how many people fit into these groups and how they behave and place wagers. By taking our initial data set we were able to classify those who take part in both live and fixed betting events (66%), those who exclusively bet fixed events (32%) and those who exclusively bet live events (2%). In taking a look at the exclusive groups we can get a sense of how people in general feel towards each type of wager and use this to draw conclusions on the majority who uses both. Through the analysis we found live bettors tend to have a riskier profile than their fixed bet counterparts. Both exclusive groups have extremes where there are those who rarely bet in very large amounts, those who bet slightly more often with relatively large amounts, conservative bettors who place many bets and then the majority who are an average of wagers placed and risk. Overall we found the same breakdown with the added information of how often they place a live wager compared to fixed. Those who had a larger live percentage bet slightly more often with a higher wager price. We also saw, regardless of live/fixed betting preference the small groups who have higher stakes and then those who place significantly more wagers.

Overall this analysis was helpful to identify underlying behaviors of gambling, we were able to find 4 typical behaviors and how many people fall into each category. Some behaviors were clearly more risk averse than others, those who fell into the majority. From the data supplied we were able to make generalizations about these groups, having information on someone's running balance would have been helpful information as we could truly analyze how risky certain bettors were being instead of deducing this from average profit

and average wager sizes. This could have enabled us to look at % of account wagered at a time. Another piece of information that would be helpful are since a majority are fixed bettors would be to see if wagers made were in a parlay, betting multiple fixed bets together for better odds. These would be helpful pieces to further assess levels of risk taken into this behavioral analysis.

This paper utilized data from the Transparency Project (www.thetransparencyproject.org), Division on Addiction, the Cambridge Health Alliance, a teaching affiliate of Harvard Medical School