Betfair Task Interview

Conor McCool

Data Cleaning & Exploration

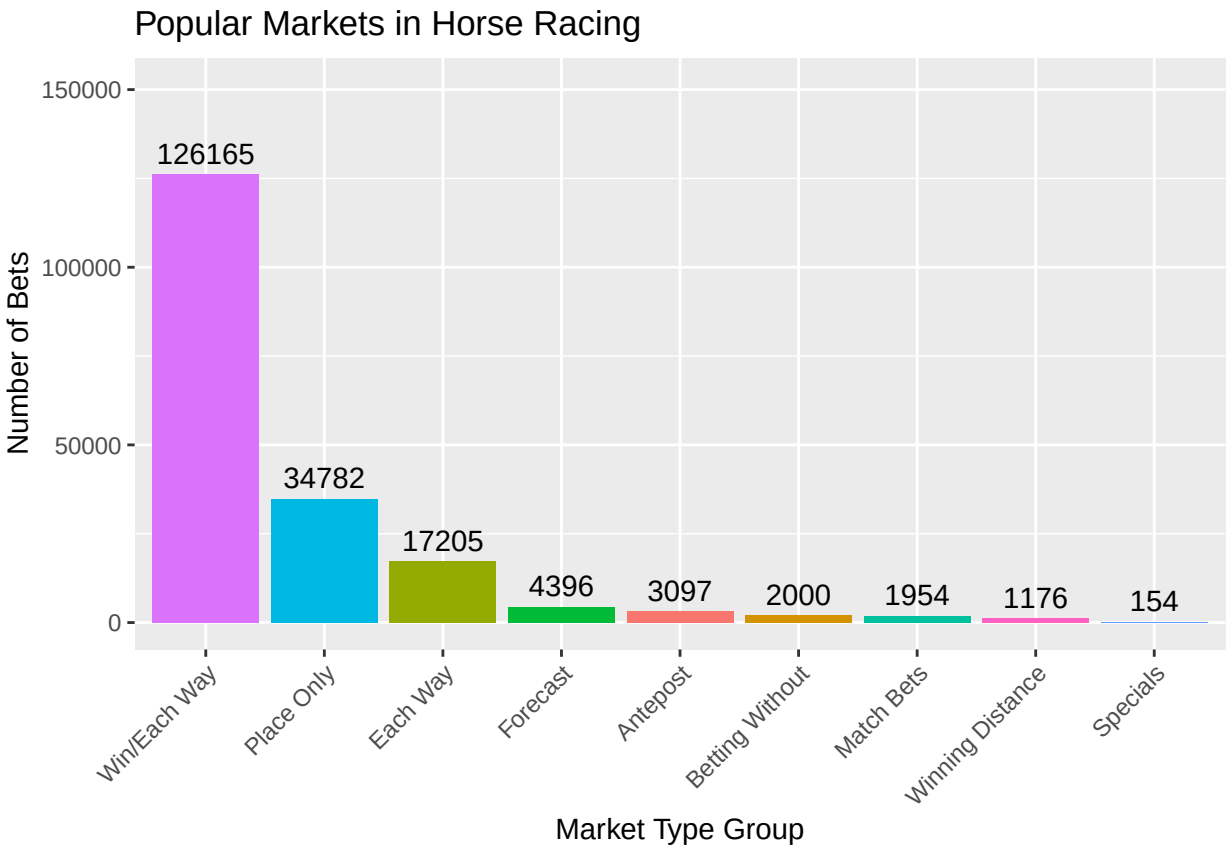
Data was collected and loaded into the Dataframe, and the three parts joined into one dataset to allow for ease of analysis.

Cricket is showing an extortionately high total compared to other sports, display mean taking into account the amount of "bets" placed. Lets see which are the most popular markets in terms of bets placed. Use this as a means to decipher the data and focus on what is most popular within the general public.

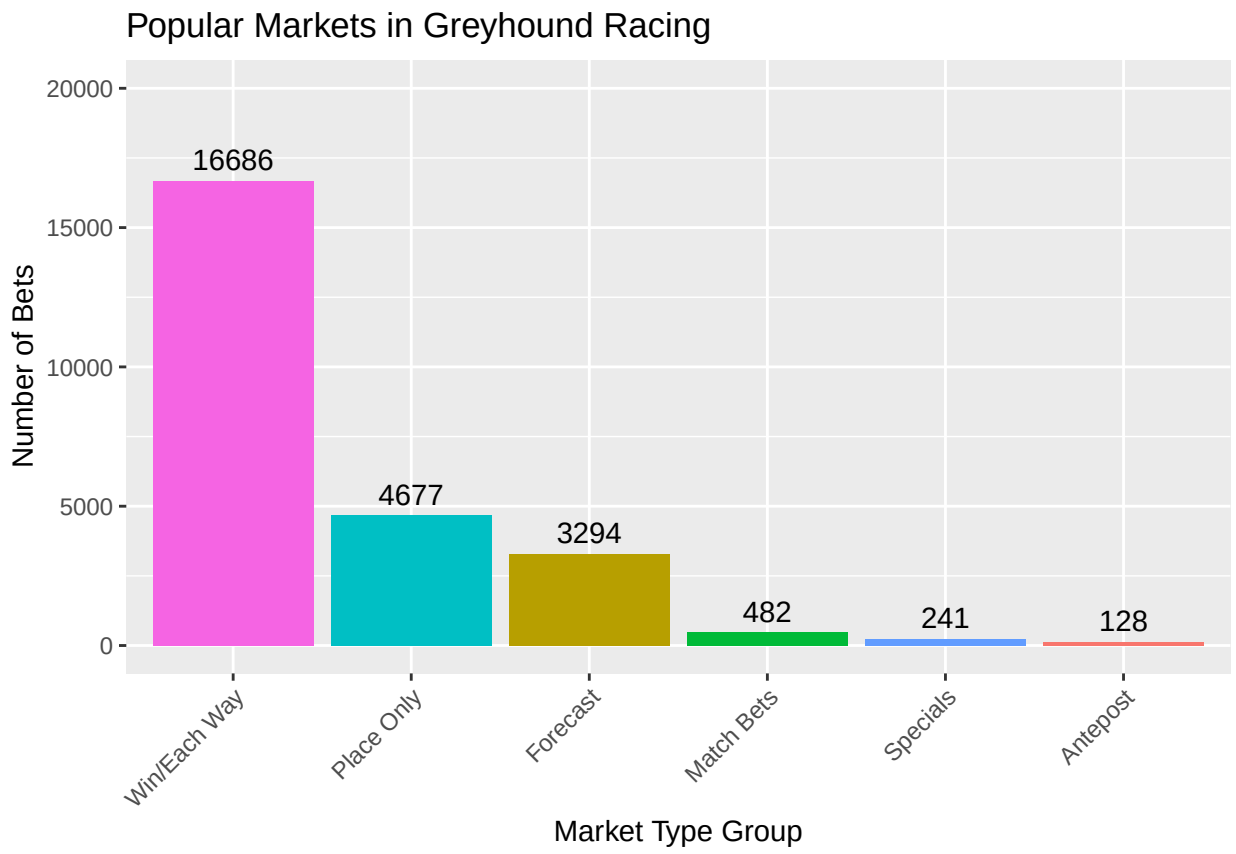
From observing within R, we can see that Greyhound Racing, Horse Racing and Soccer are the three most popular. Use this as a filter for which to focus our statistical analysis.

Lets make a plot for the three most popular markets.

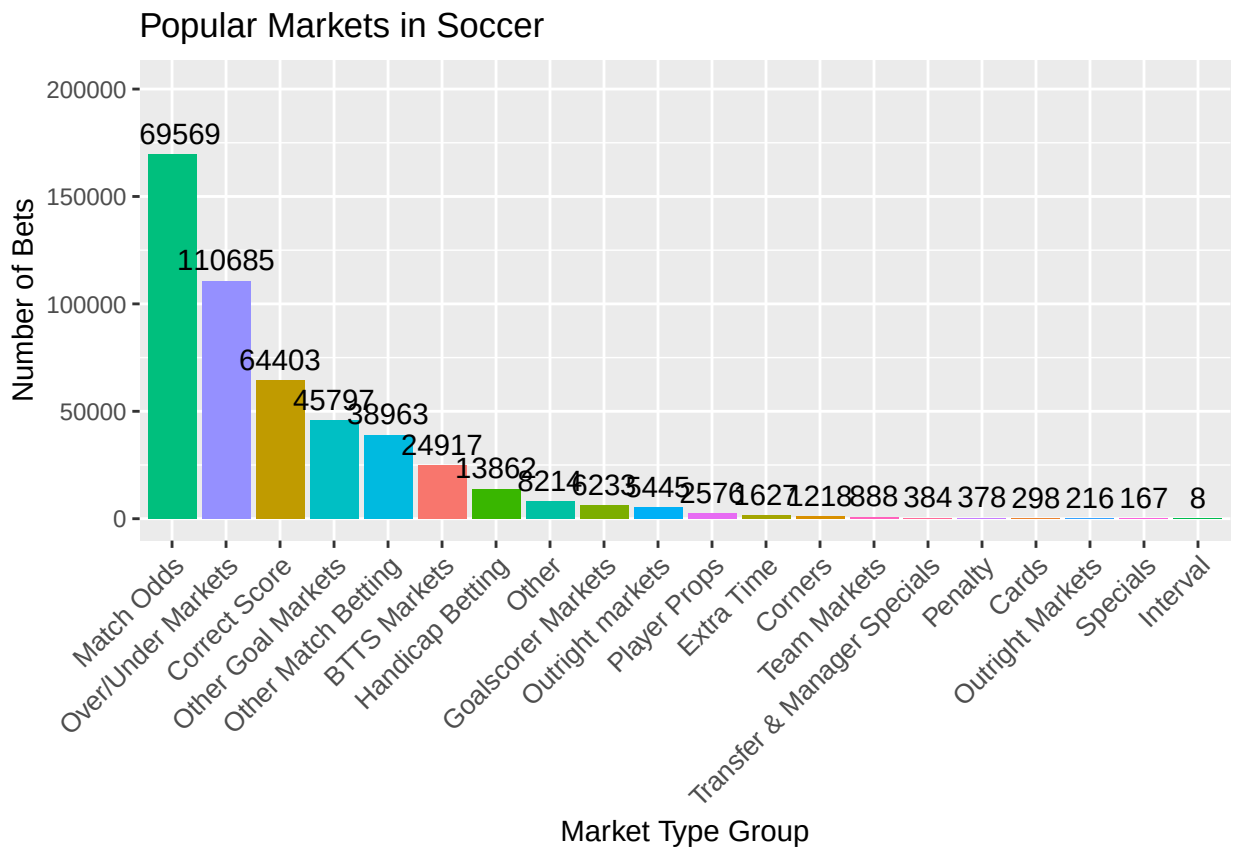
```
#print the plots
print(plot_horse_racing)
```




```
print(plot_greyhound_racing)
```

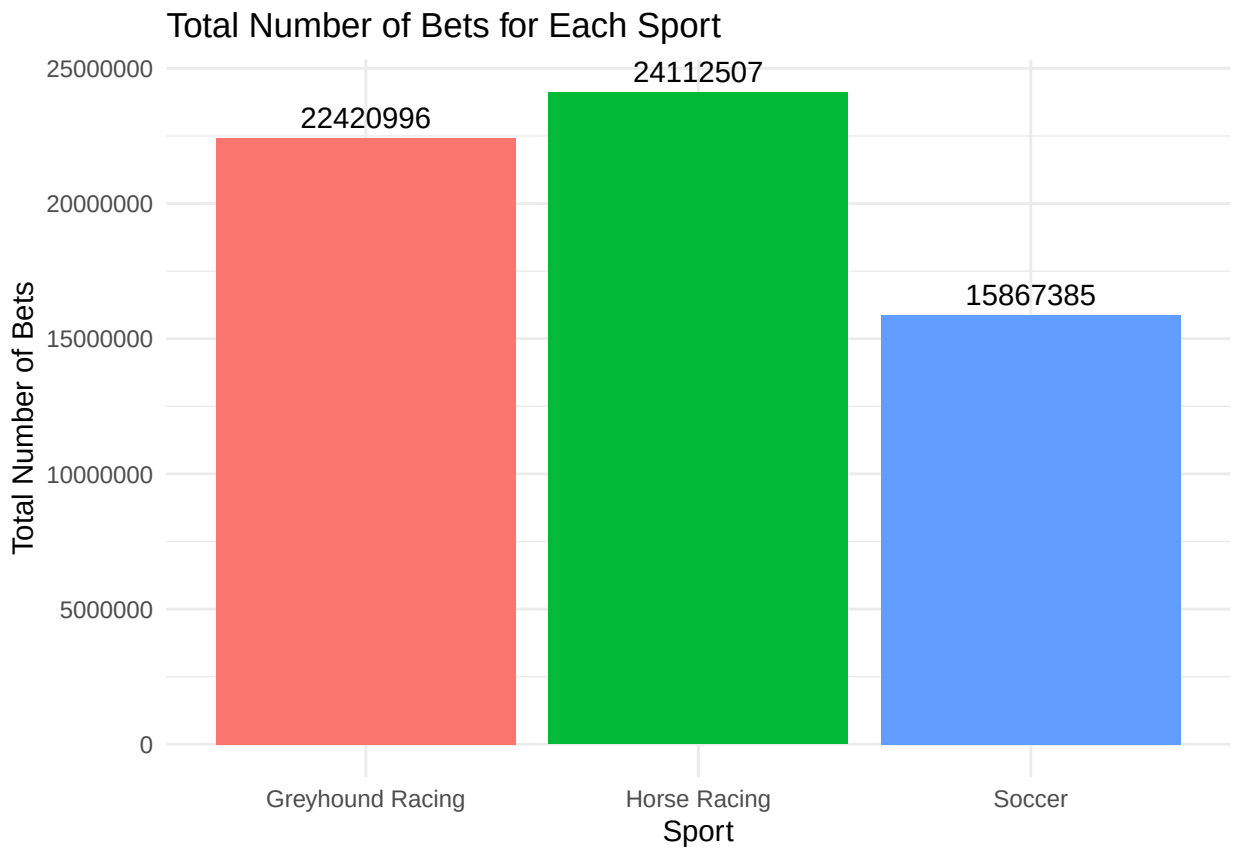



```
print(plot_soccer)
```



View these markets in terms of popularity

```
#plot a histogram
ggplot(total_bets_per_sport, aes(x = sport_name, y = total_bets, fill = sport_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Number of Bets for Each Sport", x = "Sport", y = "Total Number of
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = total_bets), vjust = -0.5)
```

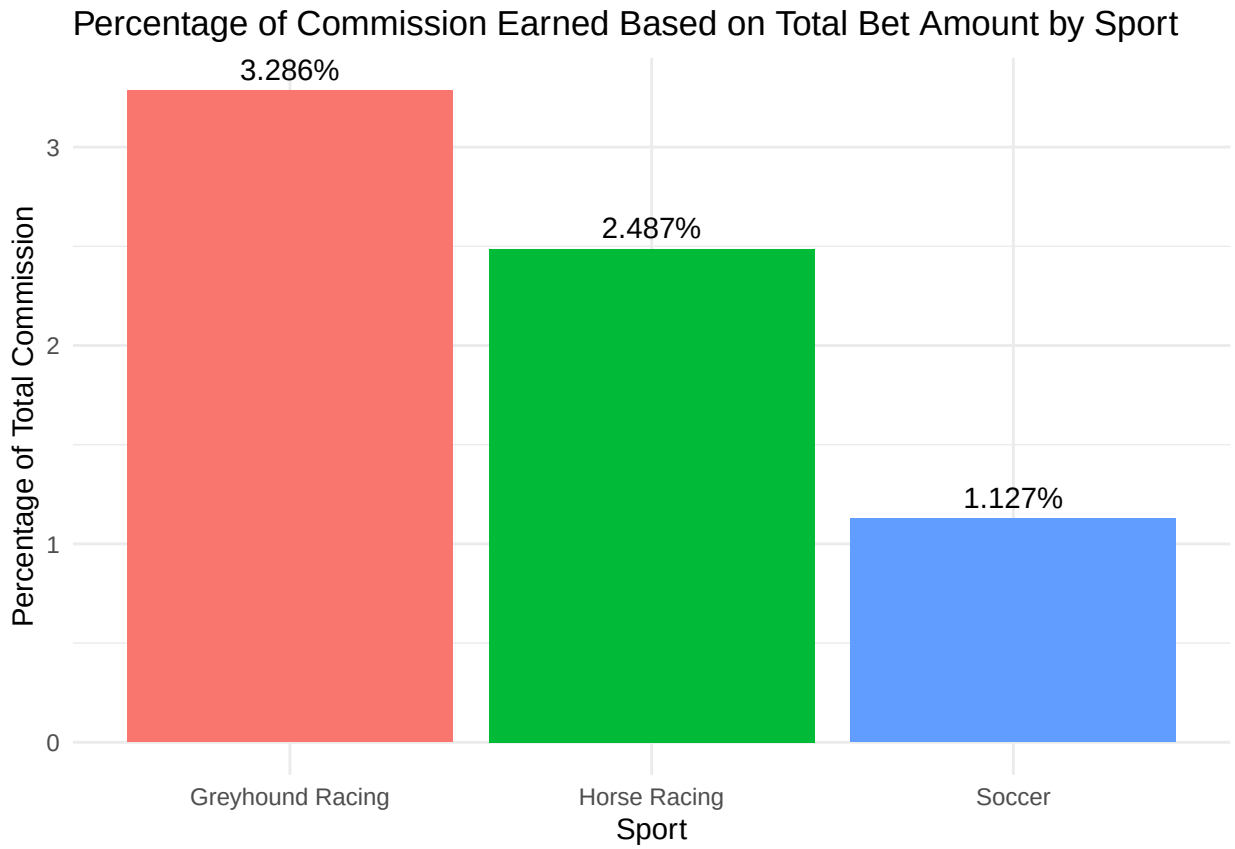


Here, we can see out of 62,400,888 bets, 35.93% were placed in Greyhound Racing, 38.64% were placed in Horse Racing, and 25.43% in Soccer, respectively.

Next, lets see which who the most profitable customers are, based on the amount of commission they earn.

It may be wise to eliminate these accounts from availing of the free bet offer, considering these are individuals already in profit, and will likely continue to use the service. Let's look at the individual markets that offered the least amount of commission.

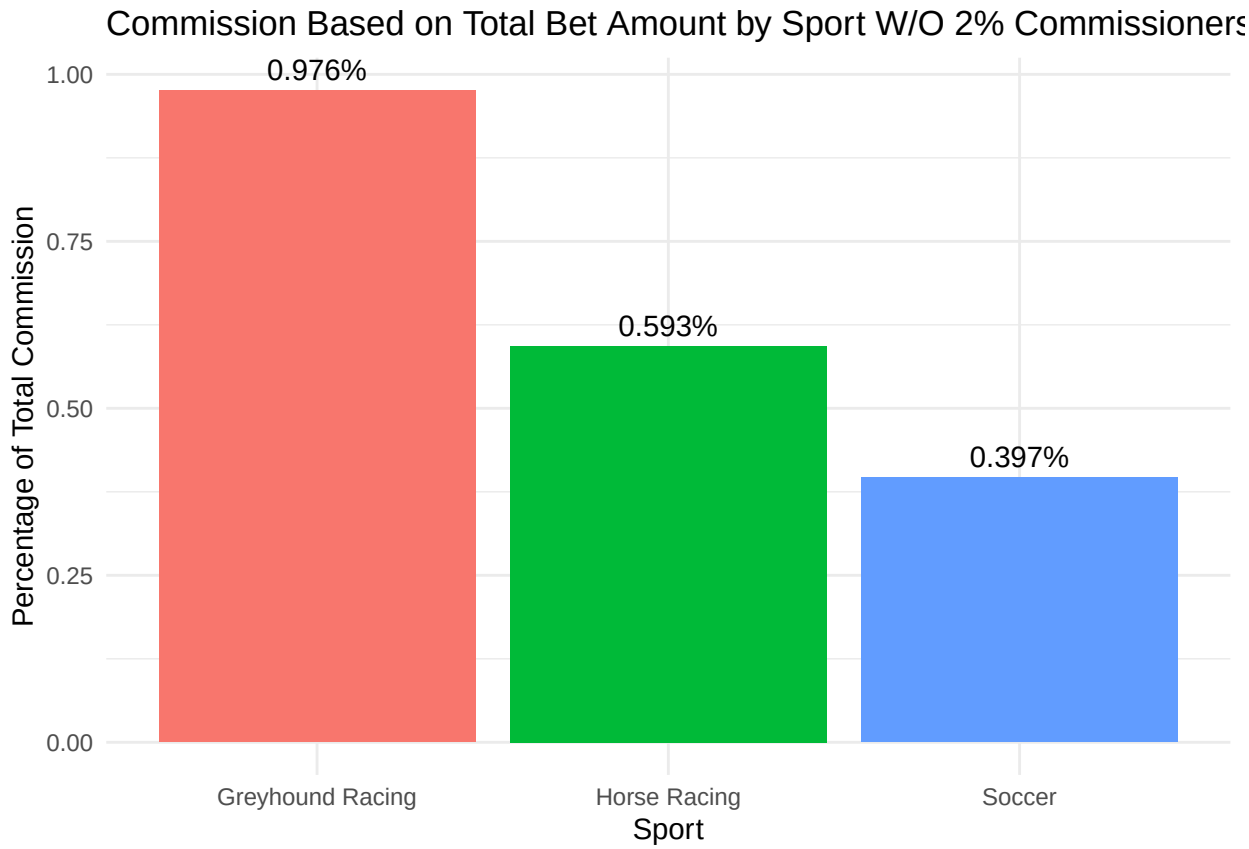
```
#plot a histogram of the percentage commission for each sport
ggplot(commission_summary, aes(x = sport_name, y = percentage_commission, fill = sport_n
  geom_bar(stat = "identity") +
  labs(title = "Percentage of Commission Earned Based on Total Bet Amount by Sport",
       x = "Sport",
       y = "Percentage of Total Commission") +
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = paste0(round(percentage_commission, 3), "%")), vjust = -0.5)
```



Here, we can see Greyhound Racing clears the Total Amount of Commission Earned in relation to the amount of money being put down, indicating that those markets have an "insider knowledge" advantage. While Horse Racing does experience something similar, it does not quite compare to the yield of Greyhound Racing for the average customer.

Lets look how it compares without the customers earning above 2% Commission.

```
#plot a histogram of the percentage commission for each sport
ggplot(commission_summary, aes(x = sport_name, y = percentage_commission, fill = sport_n
  geom_bar(stat = "identity") +
  labs(title = "Commission Based on Total Bet Amount by Sport W/O 2% Commissioners",
       x = "Sport",
       y = "Percentage of Total Commission") +
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = paste0(round(percentage_commission, 3), "%")), vjust = -0.5)
```



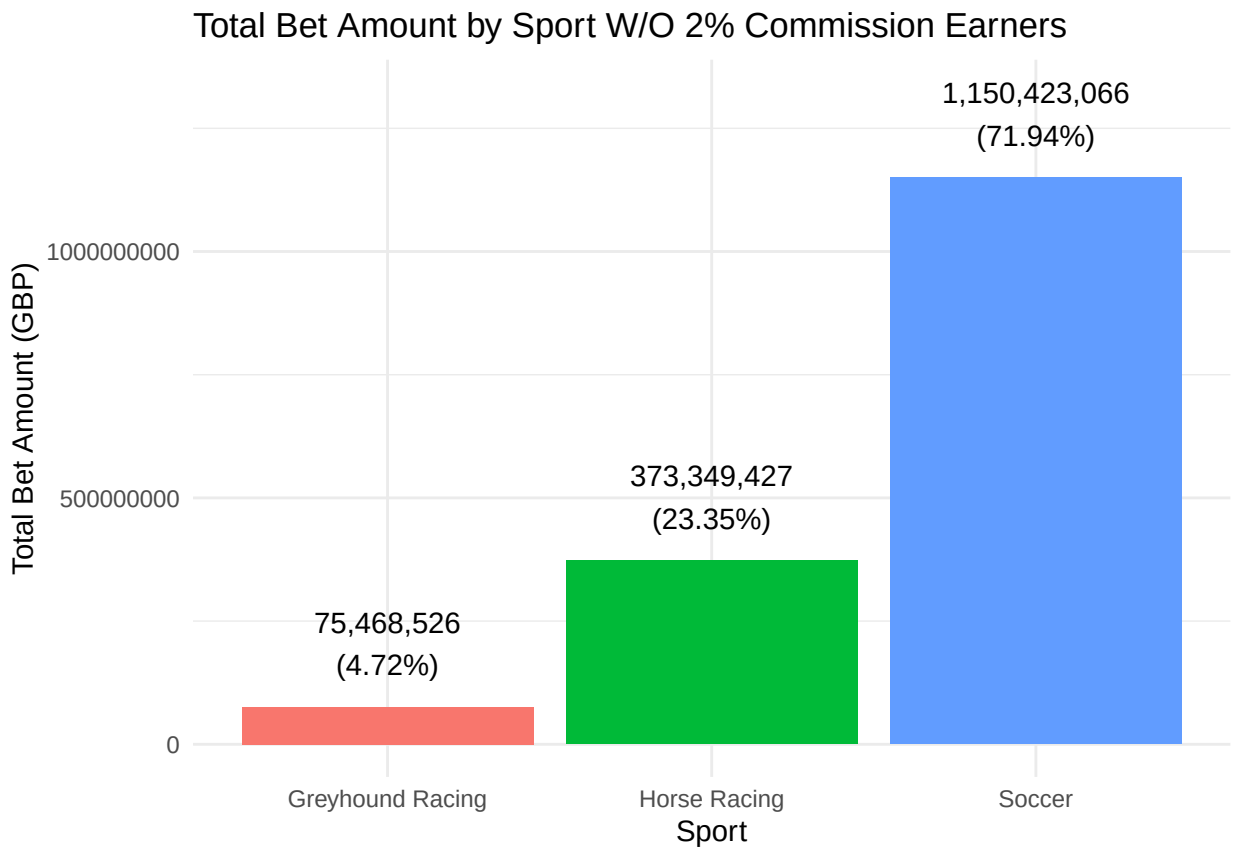
Somewhat a similar story, but much less commission earned. These would be the type of individuals that should be availing of the free bet offer to promote use of the exchange.

How will this be distributed?

While narrowing down the exchange to the three main markets, observing how much commission earned relative to the amount of bets placed, we can calculate how this can be distributed evenly among the markets.

However, when observing the commission_summary, we see that there is a much greater volume of bets for Soccer compared to the others.

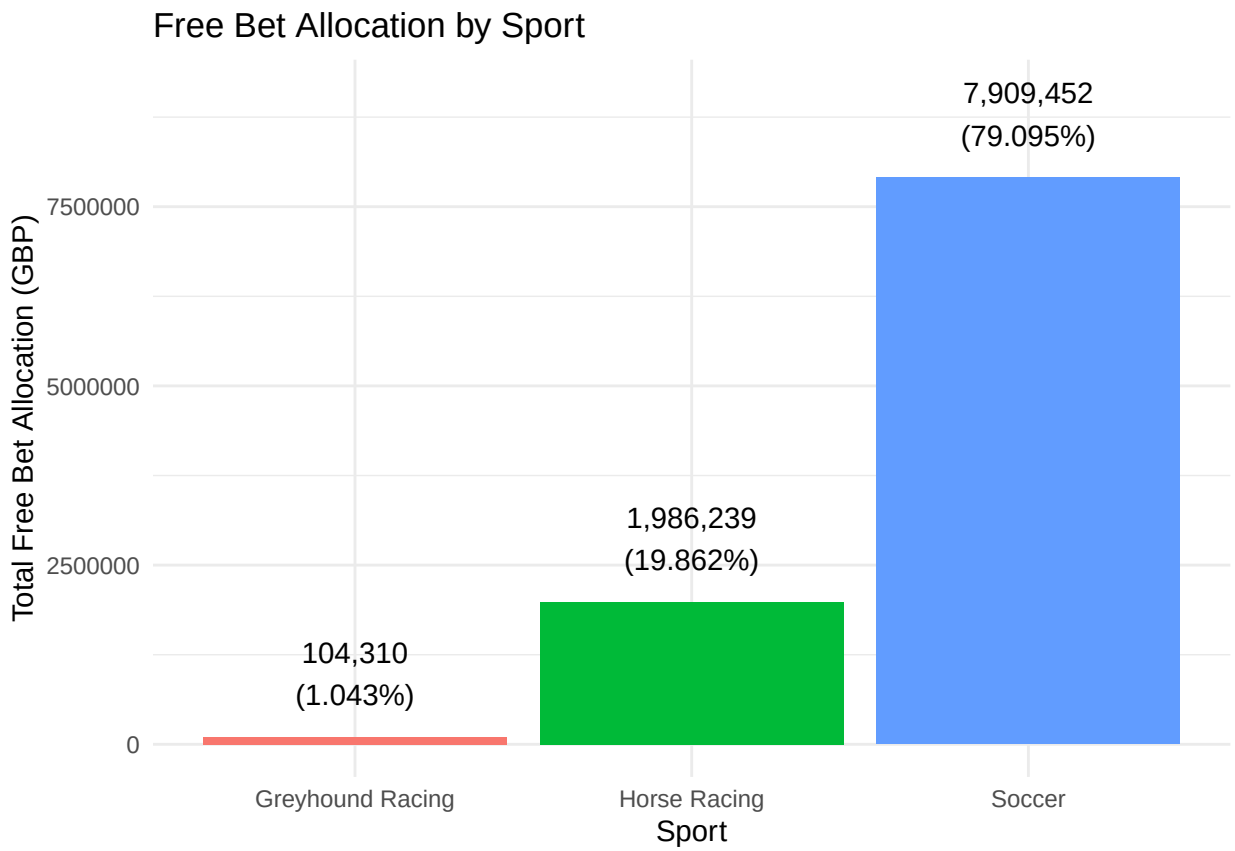
```
ggplot(commission_summary, aes(x = sport_name, y = total_bet_amount, fill = sport_name))
  geom_bar(stat = "identity") +
  labs(title = "Total Bet Amount by Sport W/O 2% Commission Earners",
       x = "Sport",
       y = "Total Bet Amount (GBP)") +
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = paste0(scales::comma(total_bet_amount), "\n(", round(percentage
            vjust = -0.5) +

    ylim(0, max_y_value)
```



Now, we can evenly distribute the allocation of the free bets to each market. However, Soccer contains a lot of arbitrage betting involving a large amount being placed by laying on Match Odds and Over/Under Markets. It would make sense to promote the most free bets to Soccer, since it has the lowest commission rate and the highest amount of total bets.

It is worth noting that the commission earned negatively affects the amount of free bet allocated to the specific market. Lets look how the graph alters as the commission is weighted and normalized.

```
#plot a bar chart for free bet allocations
max_y_value <- max(sport_allocations$total_free_bet_allocation) * 1.15 #higher y axis
ggplot(sport_allocations, aes(x = sport_name, y = total_free_bet_allocation, fill = spor
  geom_bar(stat = "identity") +
  labs(title = "Free Bet Allocation by Sport",
       x = "Sport",
       y = "Total Free Bet Allocation (GBP)") +
  theme_minimal() +
  theme(legend.position = "none") +
  geom_text(aes(label = paste0(scales::comma(total_free_bet_allocation), "\n(", round(p
            vjust = -0.5) +
  ylim(0, max_y_value)
```



In conclusion, taking into account the filtered data, excluding the individuals earning >2% commission, and adding a weighted value to the commission called "betWeight", which takes into account the commission percentage, we can determine that the Soccer market should be allocated roughly 7.91 million, Horse Racing 1.97 million, and Greyhound Racing a mere 104,310, respectively.

To note, this includes all markets within the dataset. There existed a lot of short prices that would influence the total amount of bets placed in GBP, and a wide range of markets within the Soccer dataset that would have contributed to the skewed dataset, such as Over/Under Markets, and perhaps Match Odds. Horse Racing would include a more diverse range of odds which would result in the lack of volume of bets being placed in terms of GBP, but had the highest popularity. Perhaps Horse Racing and maybe even Greyhound Racing should be allocated more than what the data suggests, but from a business standpoint, giving the free bets to soccer would be most beneficial to the company.

Some pointers to include are that there was no major international tournament during 2023, which would have yielded more Soccer bets. And, perhaps it may have been much better insight to analyze when to give out the free bets, for example in March/April to allocate for Horse Racing customers during Cheltenham/Aintree/Punchestown, or in October/November for the Shelbourne Derby in Greyhound Racing.

Some additional data that may have been useful would have been to analyze when the bets are placed, regional trends such as a large influx of bettors in a certain region betting on a horse running locally, customers net balance relative to their expenditure, or even the Age

bracket and what markets are most popular for each age group.