GROUP 3.5 ÜSTÜ

Enes Özeren
Öykü Selen Uysal
Ogün Gürcan
Musab Emir Baş
Süheyla Şeker

TASK I

I.I

In this part of the Task I, in order to create theHomeGoal Table and AwayGoal Table, the HomeGoal and AwayGoal columns are added to the matches table and the number of goals are stated. The histograms below belong to HomeGoal, AwayGoal and HomeGoal-AwayGoal respectively.

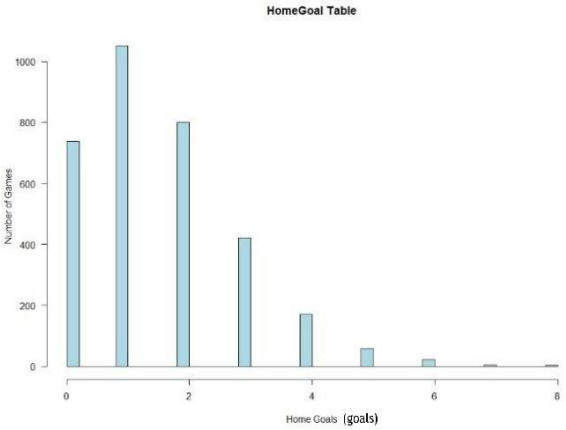


Figure 1: HomeGoal table with number of games

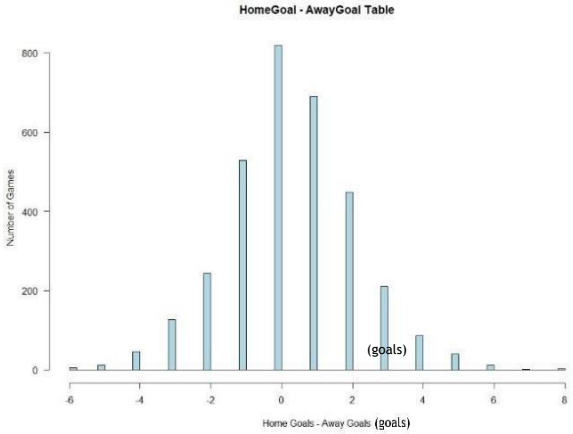


Figure 2: Difference of HomeGoal and AwayGoal table with number

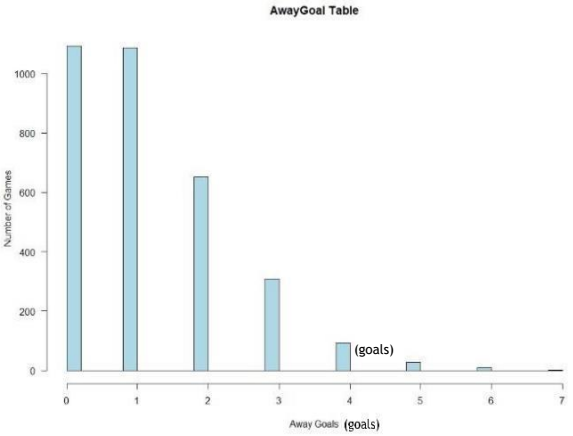


Figure 3: AwayGoal table with number of games

I.II

In this part of the Task 1, observing the distribution type, we claim that HomeGoal and AwayGoal are Poisson distributed. In order to verify our statement, calculating the mean of the HomeGoal, AwayGoal and HomeGoal-AwayGoal we found parameters for Poisson distribution to compare the sample distribution and theoretic Poisson distribution. The Poisson distribution with lambda=1.553776 in range 0:8 is plotted over to each histogram.

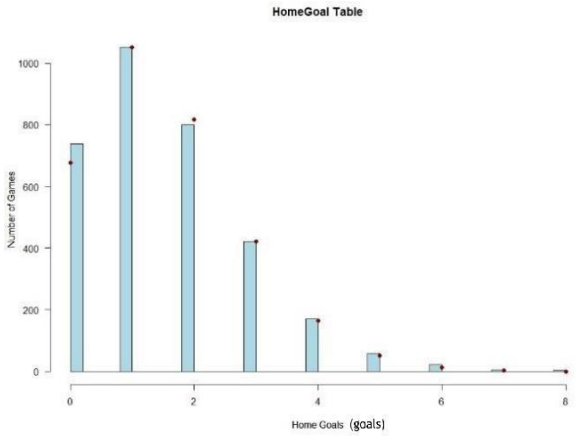


Figure 4: Home goals with poisson distribution

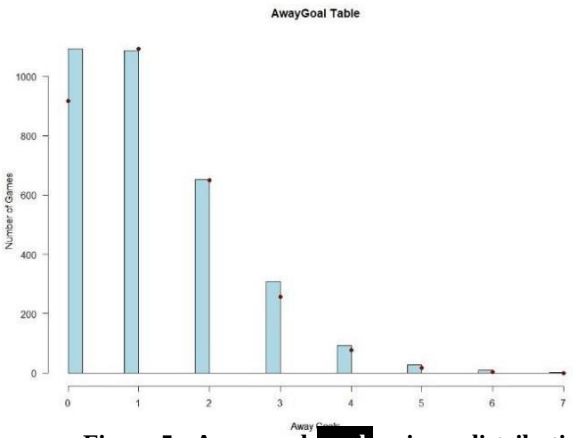


Figure 5 : Away goals [redacted] poisson distribution

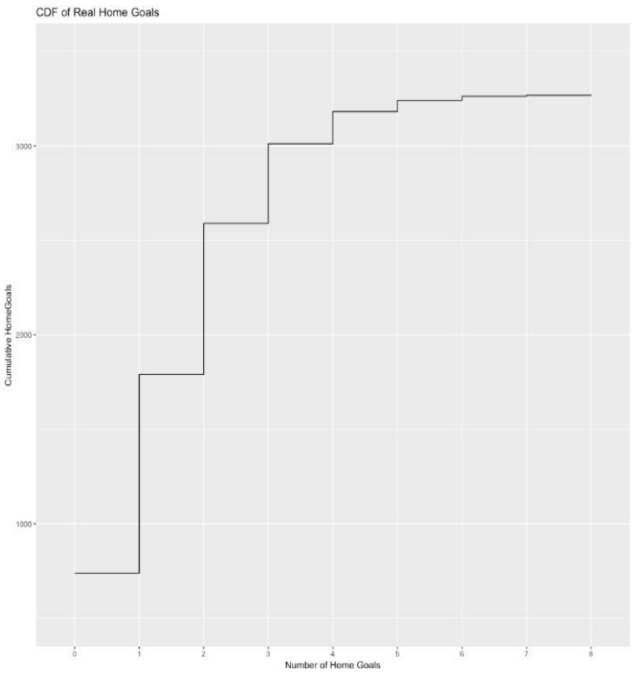
	Home Goal	Real Number of Matches	Poisson Number of Matches
1	0	739	691.6462272
2	1	1051	1074.6636092
3	2	800	834.8934958
4	3	421	432.4126132
5	4	171	167.9681321
6	5	59	52.1969851
7	6	22	13.5170742
8	7	5	3.0003588
9	8	3	0.5827358

Table 1: Number of Matches from Data and Number of Matches from Theoretical Poisson Distribution According to Home Goal

	Away Goal	Real Number of Matches	Poisson Number of Matches
1	0	1092	994.7771092
2	1	1087	1184.1153506
3	2	652	704.7453899
4	3	309	279.6270731
5	4	93	83.2122861
6	5	27	19.8100548
7	6	10	3.9300914
8	7	1	0.6683021

Table 2: Number of Matches from Data and Number of Matches from Theoretical Poisson Distribution According to Away Goal

We observed the similarity between the data and theoretical Poisson distribution from the Figure 4 and Figure 5. However, in order to support our claim, we formed Table 1 and Table 2 and obtained the numerical values for each Home Goal number and Away Goal number for data and Poisson distribution. Evaluating the result from Figure 4, Figure 5 and Table 1, Table 2, we concluded that our claim is true, and data is consistent with the Poisson distribution.



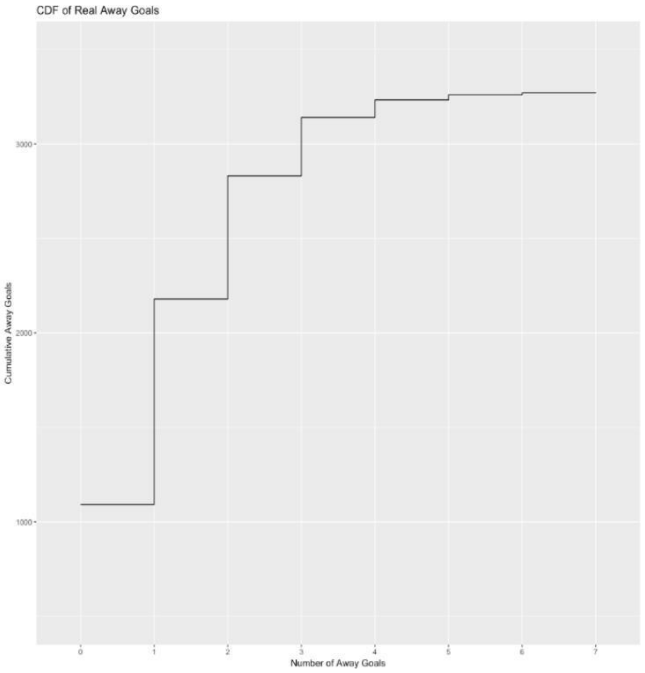
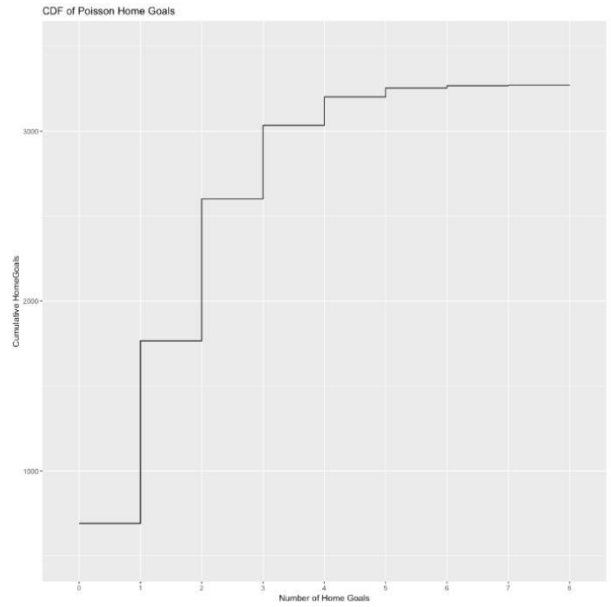


Figure 6: CDF of Real Home Goals with Number of Home **Figure 7: CDF of Real Away Goals with Number of Home**

In order to calculate the expected number of games corresponding to each quantile (number of goals) with Poisson distribution, we plotted cumulative distribution functions for Real Home Goals, Real Away Goals and Poisson Home Goals and Poisson Away Goals.

The similarity between Real and Poisson plots are observed to be high again. To conclude, our claim is turned to be true.



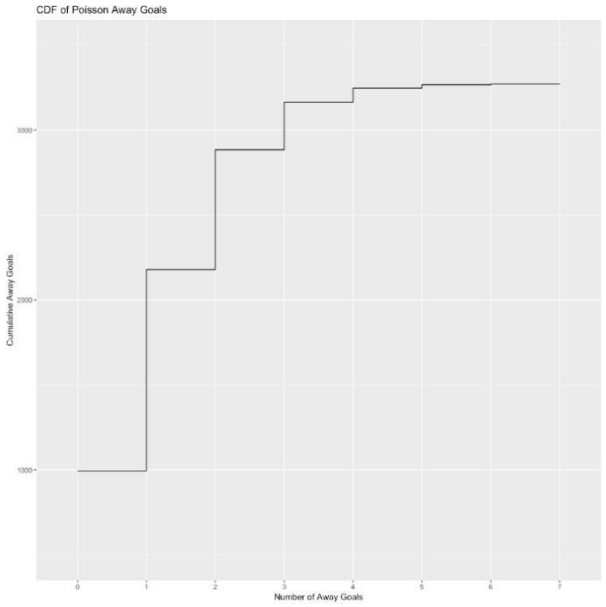


Figure 8: CDF of Poisson Home Goals with Number of Home Goals and Cumulative Home Goals

Figure 9: CDF of Poisson Away Goals with Number of Away Goals and Cumulative Away Goals

TASK II

In this part, we first calculated the P(home win),P(away win) and P(tie) by dividing 1 by odds for each result that are given by bookmarkers. Since the total probability given by bookmarkers sum up to a value bigger than 1, we normalized the probabilities. Then, we created two plots for each bookmarker that is chosen, the first plots for each bookmarker shows the non-normalized probabilities and the second ones represent the normalized probabilities. We discretized P(home win)–P(away win) values into bins like [-1,-0.95),[-0.95,-0.90).. to (0.95,1] and calculate the number of games ended as raw in the corresponding bin. In order to observe the differences between them, the two plots are put one under the other.

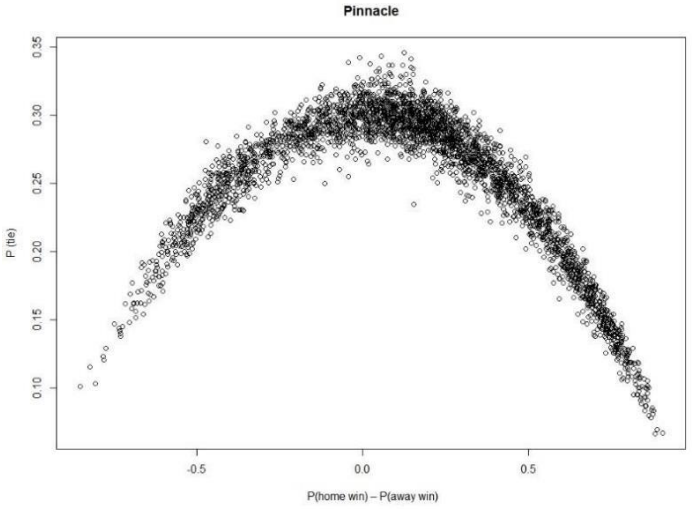


Figure 10: Pinnacle's P(home win)-P(away win) vs P(tie)

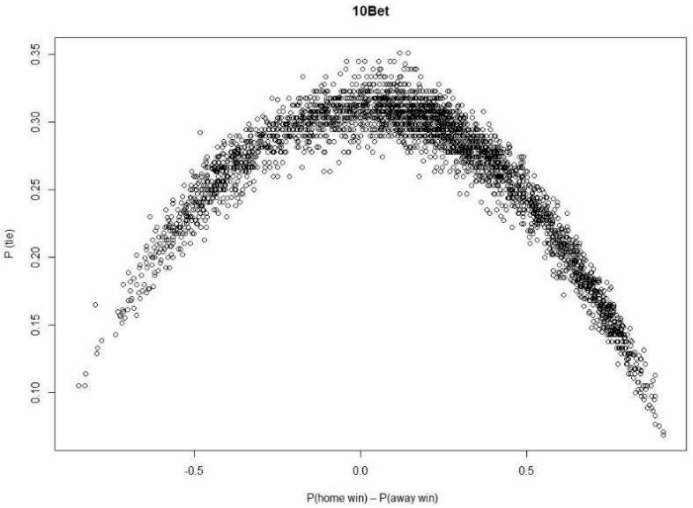


Figure 11: 10Bet's P(home win)-P(away win) vs P(tie)

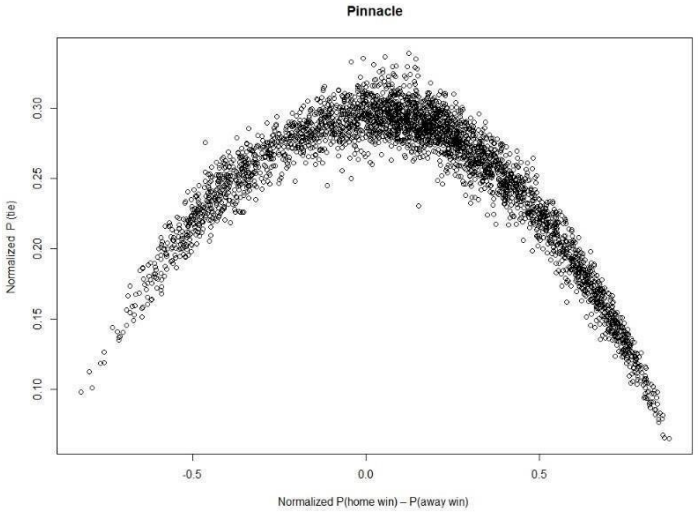


Figure 12: Pinnacle's Normalized P(home win)-P(away win) vs Normalized P(tie)

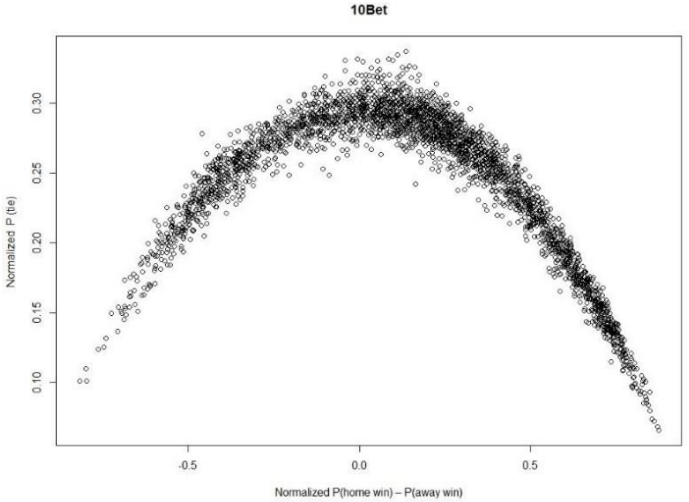


Figure 13: 10Bet's Normalized P(home win)-P(away win) vs Normalized P(tie)

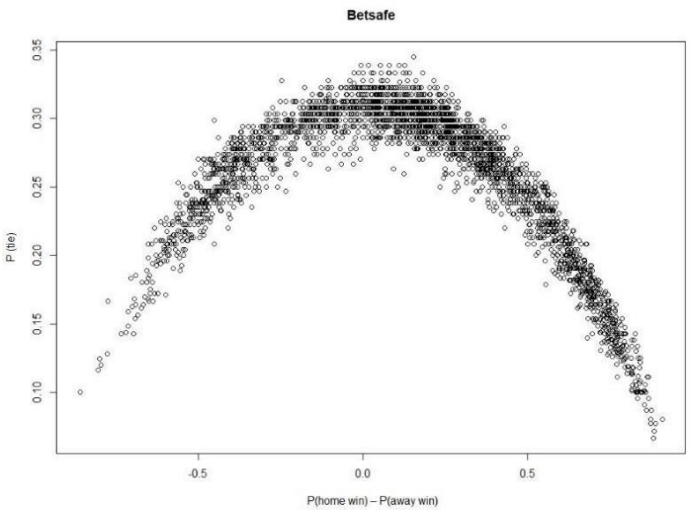


Figure 14: Betsafe's P(home win)-P(away win) vs P(tie)

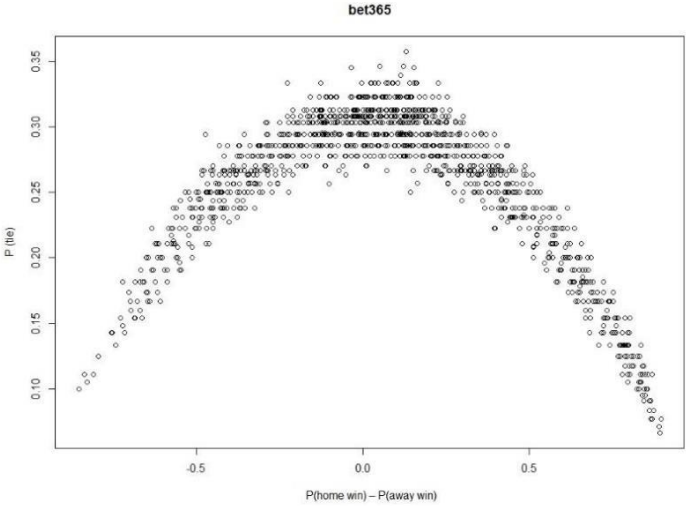


Figure 15: Bet365's P(home win)-P(away win) vs P(tie)

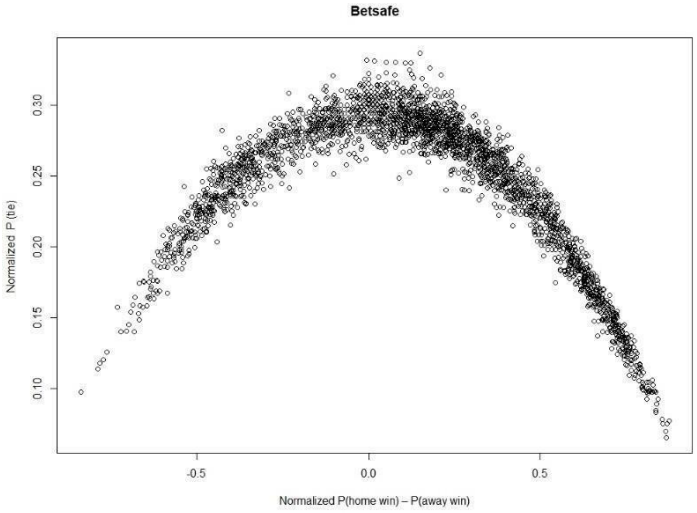


Figure 16: Betsafe's Normalized P(home win)-P(away win) vs Normalized P(tie)

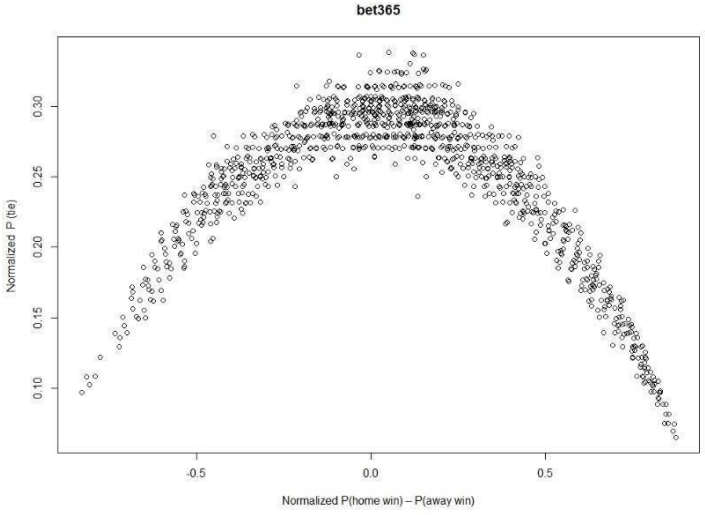


Figure 17: Bet365's Normalized P(home win)-P(away win) vs Normalized P(tie)

II.IV

When the difference between real draw ratio and bookmaker draw ratio is observed, it makes sense to bet on draw if real draw ratio is bigger than bookmaker draw ratio. Since the real probability of a match ending draw has larger probability than the probability that bookmaker estimate, there is a chance to gain money due to bias. When the difference between Real Draw Ratio and Normalized Bookmaker Draw Ratio is positive for a specific bucket that means we can make money in the long run if we play in that bucket. We can see the different buckets for different bookmaker in Figure 19, 20, 21, and 22.

	Difference Bucket	Real Draw Ratio	Normalized Bookmaker Draw Ratio
1	(0.4,0.45]	0.22137405	0.24427221
2	(0.3,0.35]	0.18978102	0.26453540
3	(0.55,0.6]	0.19387755	0.21033311
4	(0.2,0.25]	0.31862745	0.28644441
5	(0.8,0.85]	0.05714286	0.09973495
6	(0.1,0.15]	0.28333333	0.29633432
7	(-0.25,-0.2]	0.22058824	0.27827478
8	(0,0.05]	0.33812950	0.29310537
9	(0.05,0.1]	0.30653266	0.29716992
10	(-0.5,-0.45]	0.18750000	0.23280019
11	(-0.1,-0.05]	0.35869565	0.28875852
12	(0.25,0.3]	0.32748538	0.27757611
13	(0.45,0.5]	0.24468085	0.23509168
14	(-0.45,-0.4]	0.18085106	0.24512662
15	(0.7,0.75]	0.14018692	0.15171085
16	(-0.3,-0.25]	0.38095238	0.26909601
17	(0.5,0.55]	0.26000000	0.22266236
18	(0.35,0.4]	0.26984127	0.25702583
19	(-0.2,-0.15]	0.22093023	0.28477631
20	(-0.4,-0.35]	0.25000000	0.24977531
21	(0.65,0.7]	0.12605042	0.16763914
22	(-0.35,-0.3]	0.23611111	0.25840123
23	(-0.05,0]	0.32031250	0.29097735
24	(0.6,0.65]	0.14912281	0.18972266
25	(-0.65,-0.6]	0.15625000	0.17653442
26	(0.75,0.8]	0.18085106	0.13132180
27	(-0.15,-0.1]	0.25490196	0.29099750
28	(0.15,0.2]	0.33163265	0.28888540
29	(-0.55,-0.5]	0.14084507	0.21544736
30	(0.85,0.9]	0.04166667	0.06717189
31	(-0.6,-0.55]	0.15384615	0.20259313

Table3:Result Summary Ratios of Pinnacle

	Difference Bucket	Real Draw Ratio	Normalized Bookmaker Draw Ratio
1	(0.4,0.45]	0.23076923	0.25030808
2	(0.3,0.35]	0.19594595	0.26886001
3	(0.5,0.55]	0.27551020	0.22819467
4	(0.25,0.3]	0.31481481	0.27752156
5	(0.8,0.85]	0.02272727	0.10113684
6	(0.05,0.1]	0.27918782	0.29345696
7	(-0.25,-0.2]	0.23076923	0.27895784
8	(0.55,0.6]	0.20952381	0.21571813
9	(0,0.05]	0.36619718	0.29332571
10	(-0.5,-0.45]	0.16279070	0.24193434
11	(-0.05,0]	0.30656934	0.28916868
12	(0.2,0.25]	0.31794872	0.28329060
13	(0.45,0.5]	0.22222222	0.23747883
14	(-0.45,-0.4]	0.19230769	0.24317599
15	(0.75,0.8]	0.14893617	0.13214189
16	(-0.3,-0.25]	0.32352941	0.26933027
17	(-0.35,-0.3]	0.20000000	0.26451801
18	(0.35,0.4]	0.28030303	0.25792597
19	(0.1,0.15]	0.30687831	0.29385515
20	(0.65,0.7]	0.15000000	0.16889931
21	(-0.4,-0.35]	0.29702970	0.25579555
22	(0.6,0.65]	0.12121212	0.19399908
23	(-0.2,-0.15]	0.21686747	0.28360449
24	(-0.65,-0.6]	0.20000000	0.18167693
25	(0.85,0.9]	0.05882353	0.07815754
26	(0.7,0.75]	0.16363636	0.15288085
27	(-0.15,-0.1]	0.27173913	0.29147600
28	(0.15,0.2]	0.32673267	0.28758693
29	(-0.55,-0.5]	0.15873016	0.21809512
30	(-0.1,-0.05]	0.32967033	0.28788765
31	(-0.75,-0.7]	0.12500000	0.15375154
32	(-0.7,-0.65]	0.10526316	0.16134259
33	(-0.6,-0.55]	0.12500000	0.21009868

Table4:Result Summary Ratios of Betsafe

	Difference Bucket	Real Draw Ratio	Normalized Bookmaker Draw Ratio
1	(0.4,0.45]	0.25217391	0.25042292
2	(0.3,0.35]	0.20512821	0.26359855
3	(0.5,0.55]	0.24691358	0.22522813
4	(0.2,0.25]	0.31603774	0.28372233
5	(0.8,0.85]	0.05128205	0.10617939
6	(0.1,0.15]	0.31147541	0.29319351
7	(-0.25,-0.2]	0.23214286	0.28227795
8	(0.55,0.6]	0.20754717	0.21213220
9	(0,0.05]	0.37795276	0.29086113
10	(0.05,0.1]	0.30392157	0.29441225
11	(-0.5,-0.45]	0.17171717	0.23322525
12	(-0.05,0]	0.28260870	0.29163384
13	(0.45,0.5]	0.26605505	0.23608316
14	(-0.4,-0.35]	0.22988506	0.25744736
15	(0.7,0.75]	0.14285714	0.15284542
16	(-0.3,-0.25]	0.33870968	0.26904089
17	(0.35,0.4]	0.21875000	0.25504764
18	(-0.2,-0.15]	0.16129032	0.28421456
19	(0.65,0.7]	0.13043478	0.16863237
20	(0.25,0.3]	0.34965035	0.27645610
21	(-0.35,-0.3]	0.23456790	0.25923358
22	(0.6,0.65]	0.13636364	0.19256793
23	(-0.65,-0.6]	0.12903226	0.16594387
24	(0.75,0.8]	0.17094017	0.13479046
25	(-0.15,-0.1]	0.29347826	0.28985382
26	(-0.1,-0.05]	0.37894737	0.28949929
27	(0.85,0.9]	0.04545455	0.06922465
28	(0.15,0.2]	0.30541872	0.29166810
29	(-0.45,-0.4]	0.25316456	0.24719164
30	(-0.6,-0.55]	0.13725490	0.20102960
31	(-0.7,-0.65]	0.10526316	0.17279586
32	(-0.55,-0.5]	0.14285714	0.22421959
33	(-0.75,-0.7]	0.22222222	0.15651553

Table5:Result Summary Ratios of bet365

	Difference Bucket	Real Draw Ratio	Normalized Bookmaker Draw Ratio
1	(0.4,0.45]	0.22388060	0.25176551
2	(0.3,0.35]	0.21768707	0.26921618
3	(0.55,0.6]	0.17708333	0.21634941
4	(0.25,0.3]	0.30232558	0.27916130
5	(0.8,0.85]	0.02222222	0.09691772
6	(0.1,0.15]	0.28089888	0.29667044
7	(-0.2,-0.15]	0.18604651	0.28336971
8	(0.5,0.55]	0.24752475	0.22802017
9	(0,0.05]	0.35460993	0.29462447
10	(-0.5,-0.45]	0.15116279	0.23705182
11	(-0.1,-0.05]	0.40000000	0.28822675
12	(0.2,0.25]	0.31000000	0.28549651
13	(0.45,0.5]	0.26315789	0.23954403
14	(-0.45,-0.4]	0.24752475	0.24721638
15	(0.7,0.75]	0.12931034	0.16054317
16	(0.05,0.1]	0.30434783	0.29778915
17	(-0.3,-0.25]	0.28787879	0.27297022
18	(0.35,0.4]	0.25600000	0.25814526
19	(-0.4,-0.35]	0.25000000	0.25481084
20	(-0.35,-0.3]	0.24637681	0.26242757
21	(-0.05,0]	0.29457364	0.29133075
22	(0.6,0.65]	0.18867925	0.19525950
23	(-0.25,-0.2]	0.22727273	0.27761734
24	(-0.65,-0.6]	0.09677419	0.17815202
25	(0.75,0.8]	0.21359223	0.13970571
26	(-0.15,-0.1]	0.29896907	0.29260328
27	(0.65,0.7]	0.10891089	0.17004927
28	(0.85,0.9]	0.04545455	0.07397260
29	(0.15,0.2]	0.34653465	0.28967966
30	(-0.6,-0.55]	0.15909091	0.21022644
31	(-0.75,-0.7]	0.16666667	0.16379956
32	(-0.7,-0.65]	0.17647059	0.17155849
33	(-0.55,-0.5]	0.14285714	0.21773768

Table6:Result Summary Ratios of Bet10

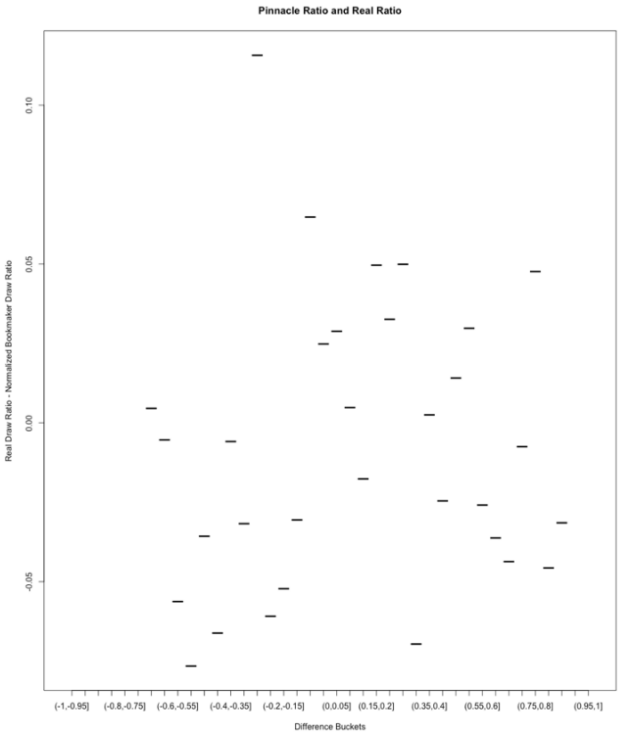


Figure 18: Pinnacle Odd Ratio Real Ratio Difference

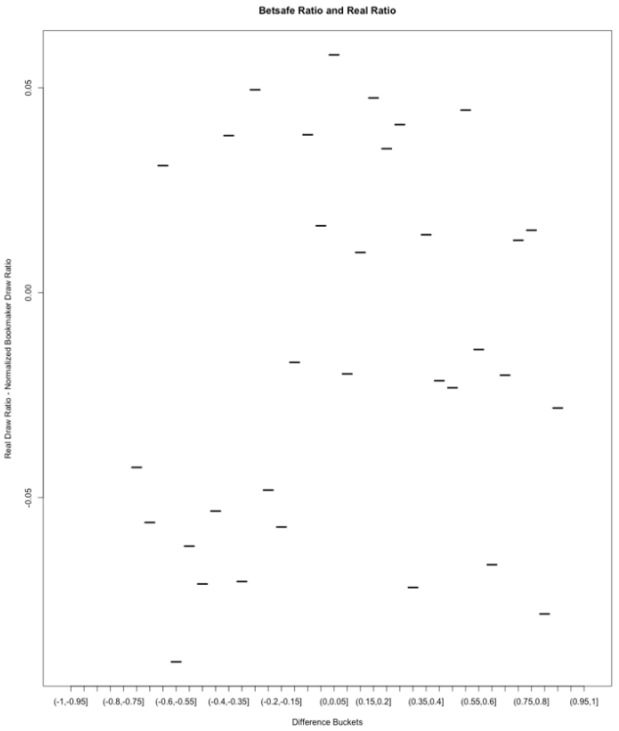


Figure 19: Betsafe Odd Ratio Real Ratio Difference

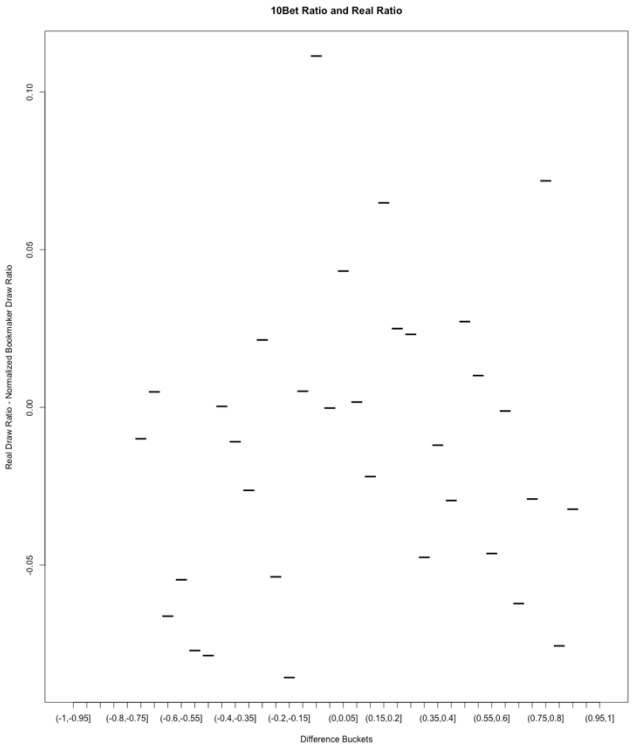


Figure 20: 10Bet Odd Ratio Real Ratio Difference

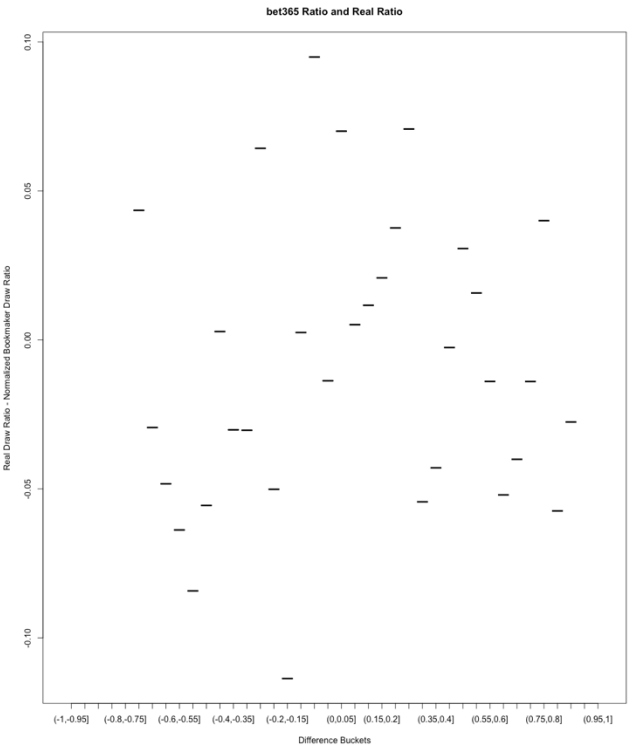


Figure 21: bet365 Odd Ratio Real Ratio Difference

III.I

In this part, first we filtered the odds table by taking the ‘ou’ betType and 2.5 totalhandicap for the bookmarker Pinnacle. We observed that there are more than 1 over odds for some matches given by the Pinnacle so by filtering initial and final odds , we created new tables called initial_odds and latest_odds. Then, similar to the previous task, we calculated the probabilities of over and under using the odds given by Pinnacle and then normalized the probabilities.

In order to detect the over ending matches, we wanted to search for total goals bigger than 2.5, thus more than or equal to 3. To do so, we calculated TotalGoal and merged it to the latest_odds table. For the over ending matches, we calculated real_over_ratio and bookmaker_over_ratio. While doing so, there occurred NA cells in TotalGoal column due to matches not played yet, we solved this problem by using na.rm=TRUE.

Result Summary Table

	diff_bucket	real_over_ratio	bookmaker_over_ratio
1	(0.45,0.5]	0.4826667	0.4782296
2	(0.5,0.55]	0.5552408	0.5256846
3	(0.4,0.45]	0.3844156	0.4305291
4	(0.65,0.7]	0.6727273	0.6674302
5	(0.6,0.65]	0.5785953	0.6228733
6	(0.55,0.6]	0.5692308	0.5749451
7	(0.35,0.4]	0.3711340	0.3877869
8	(0.3,0.35]	0.2500000	0.3402778
9	(0.7,0.75]	0.7500000	0.7058902

Table 7:Result Summary Table for Pinnacle

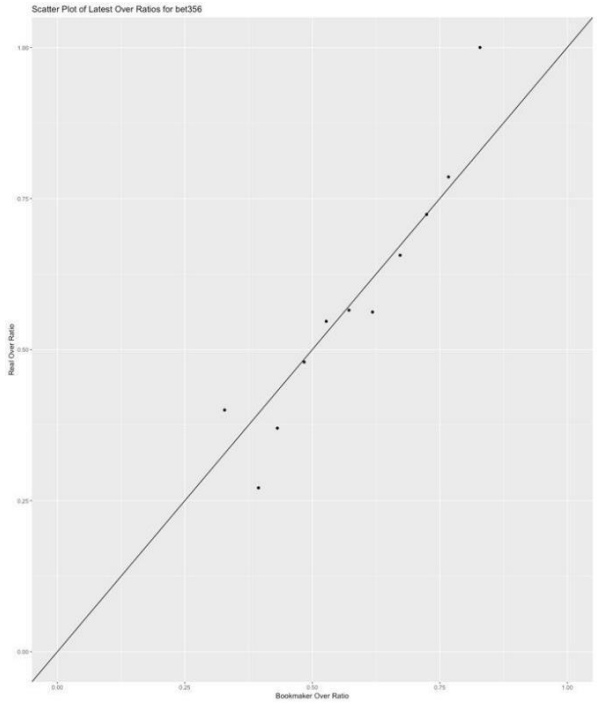


Figure 22: Scatter Plot of Latest Over Ratios for bet365

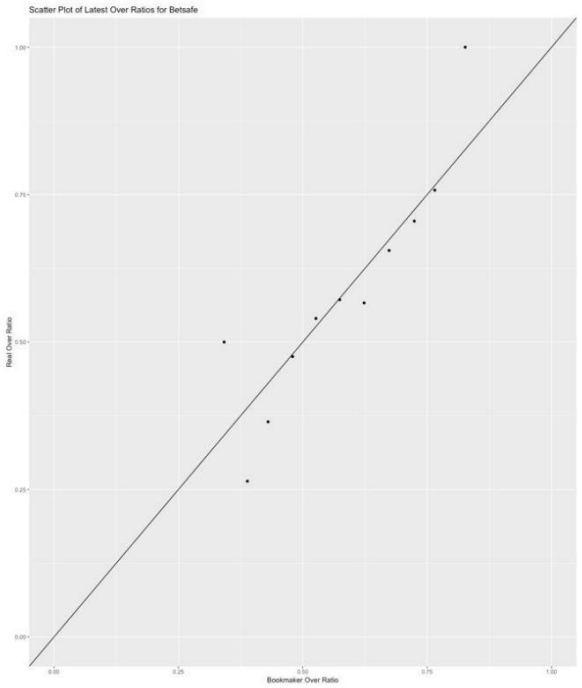


Figure 23: Scatter Plot of Latest Over Ratios for Betsafe

In order to visualize our work, we used ggplot2 library and created a scatter plot with reference line x=y. We did this for 4 different bookmakers' latest odds and observed how data is distributed along the reference line.

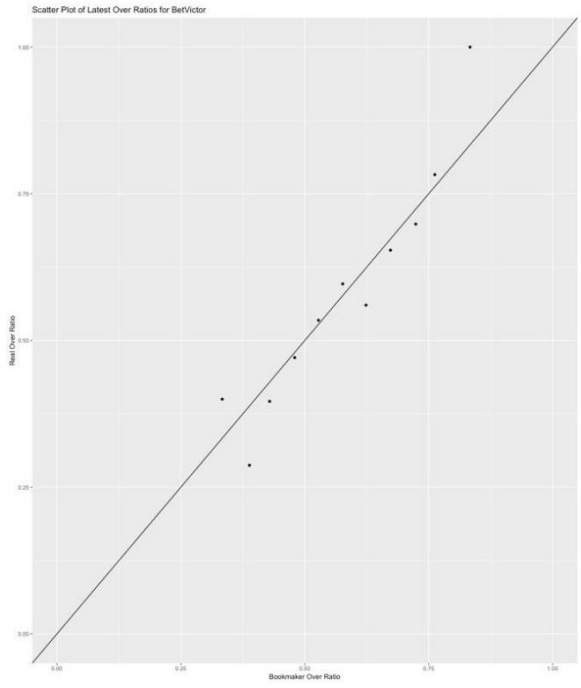


Figure 24: Scatter Plot of Latest Over Ratios for BetVictor

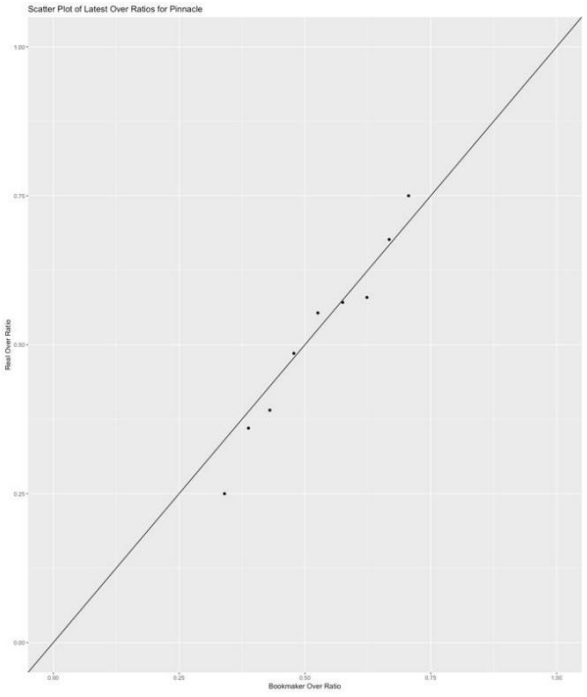


Figure 25: Scatter Plot of Latest Over Ratios for Pinnacle

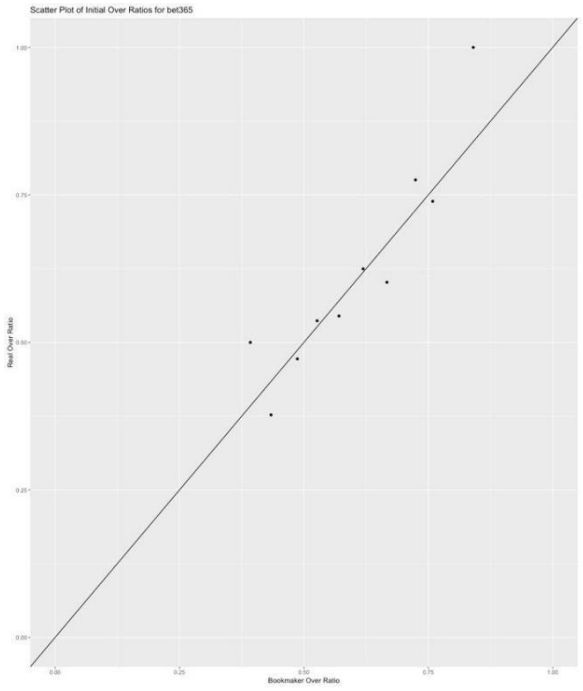


Figure 26: Scatter Plot of Initial Over Ratios for bet365

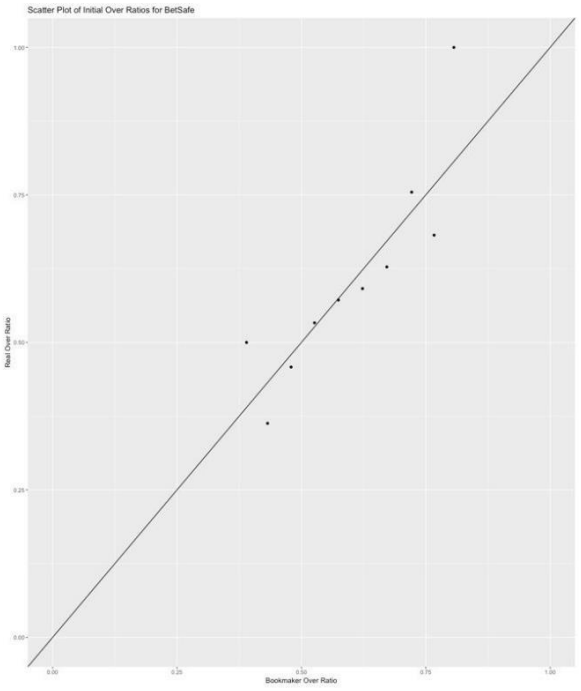


Figure 27: Scatter Plot of Initial Over Ratios for BetSafe

We did the same to visualize initial odds and observed(real) ratios.

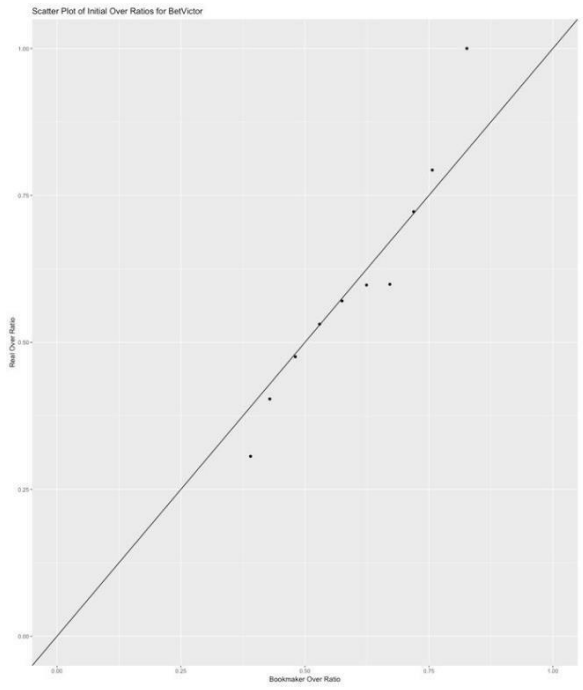


Figure 28: Scatter Plot of Initial Over Ratios for BetVictor

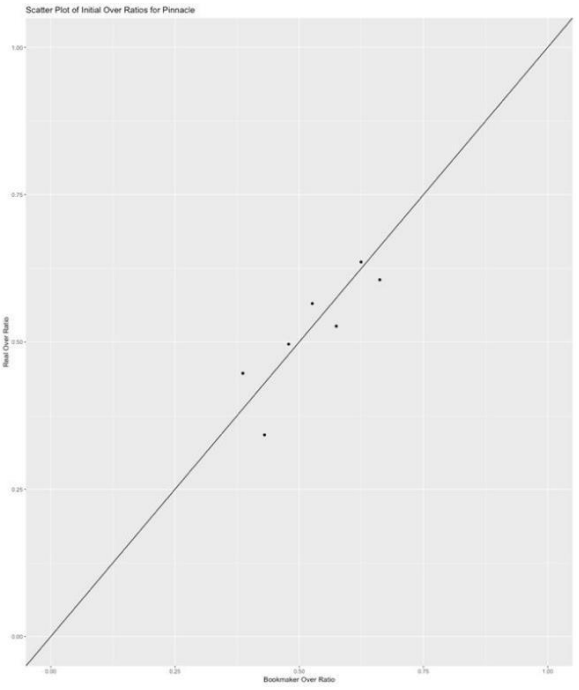


Figure 29: Scatter Plot of Initial Over Ratios for Pinnacle

III.II

In this part, we tried to observe the reliability of a bookmaker in years. We chose Pinnacle to analyze its odds for years. We converted the epoch time units to Turkey's local time and date. Next, we determined a certain bucket range in which we compared the mean of over probabilities given by bookmaker and mean of real over probabilities in each year. As we can observe in the graph, it is not possible to say that Pinnacle improved its odds over years.

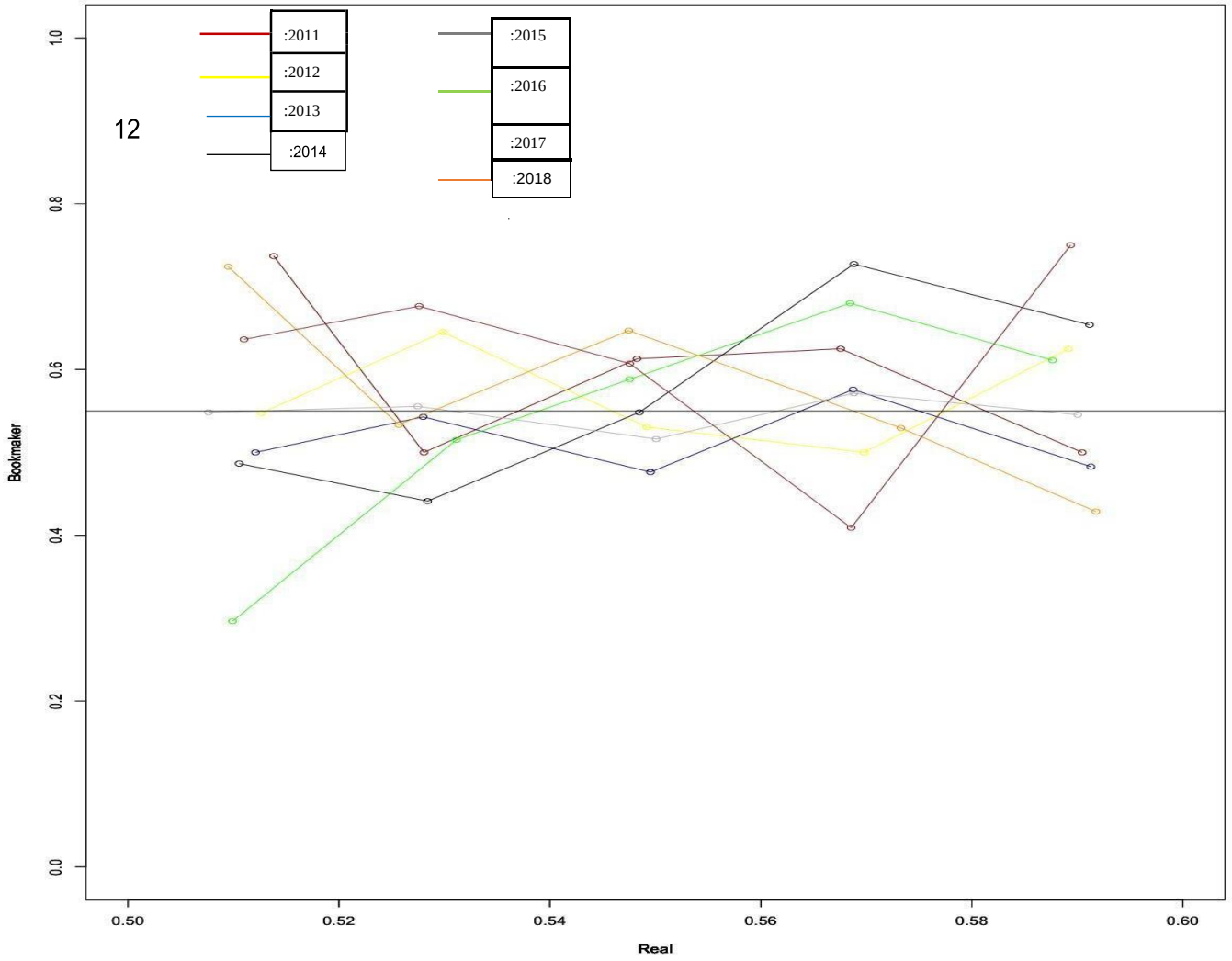


Figure 30: Representation of change of odds

R codes for Plotting Figures

i. Code for Figure 1:

```
summary_by_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary_by_homegoal$HomeGoal)
table_for_homegoal=table(summary_by_homegoal$HomeGoal)
table_for_homegoal

hist(summary_by_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab
= "Number of Games", las =1, breaks = 30,col='light blue')
```

ii. Code for Figure 2:

```
summary_by_homegoal_and_awaygoal=matches[,list(count=.N),by=list(matchId,HomeGoal,Away
Goal)]
summary_by_homegoal_and_awaygoal
homegoal_minus_awaygoal=summary_by_homegoal_and_awaygoal[,list(count=.N),by=list(matchId
,HomeGoal-AwayGoal)]
homegoal_minus_awaygoal[,c("HomeGoal-AwayGoal"):=HomeGoal]
homegoal_minus_awaygoal$HomeGoal=NULL
homegoal_minus_awaygoal$count=NULL
homegoal_minus_awaygoal
hist(homegoal_minus_awaygoal$`HomeGoal-AwayGoal`,main = "HomeGoal - AwayGoal Table",
xlab = "Home Goals - Away Goals", ylab = "Number of Games", las =1, breaks = 60,col='light blue')
```

iii. Code for Figure 3:

```
summary_by_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]
factor(summary_by_awaygoal$AwayGoal)
table_for_awaygoal=table(summary_by_awaygoal$AwayGoal)
table_for_awaygoal
hist(summary_by_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab
= "Number of Games",las=1, breaks = 30,col='light blue')
```

iv. Code for Figure 6:

```
summary_by_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary_by_homegoal$HomeGoal)
table_for_homegoal=table(summary_by_homegoal$HomeGoal)
table_for_homegoal

hist(summary_by_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab
= "Number of Games", las =1, breaks = 30,col='light blue')
mean_homegoal=mean(matches$HomeGoal,na.rm = T) mean_homegoal

par(new=TRUE)
plot(dpois(x=0:8,lambda=mean_homegoal), xlab = "Home Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)
```

```
  HomeGoal_pois=c(dpois(0,mean_homegoal)*sum(table_for_homegoal),
                  dpois(1,mean_homegoal)*sum(table_for_homegoal),
                  dpois(2,mean_homegoal)*sum(table_for_homegoal),
                  dpois(3,mean_homegoal)*sum(table_for_homegoal),
                  dpois(4,mean_homegoal)*sum(table_for_homegoal),
                  dpois(5,mean_homegoal)*sum(table_for_homegoal),
                  dpois(6,mean_homegoal)*sum(table_for_homegoal),
                  dpois(7,mean_homegoal)*sum(table_for_homegoal),
                  dpois(8,mean_homegoal)*sum(table_for_homegoal))

real_vs_poison_homegoal=data.table(Real_HomeGoal=table_for_homegoal,Poison_HomeGoal=Ho
meGoal_pois)

ggplot(real_vs_poison_homegoal, aes(real_vs_poison_homegoal$Real_HomeGoal.V1,
cumsum(real_vs_poison_homegoal$Real_HomeGoal.N))) + geom_step(aes(group=1))+

  ggtitle("CDF of Real Home Goals")+
  xlab("Number of Home Goals")+
  ylab("Cumulative HomeGoals")+
  ylim(500, 3500)
```

v. Code for Figure 7:

```
summary_by_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]
factor(summary_by_awaygoal$AwayGoal)
table_for_awaygoal=table(summary_by_awaygoal$AwayGoal)
table_for_awaygoal
hist(summary_by_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab
= "Number of Games",las=1, breaks = 30,col='light blue')

mean_awaygoal=mean(matches$AwayGoal,na.rm = T)
par(new=TRUE)
plot(dpois(x=0:7,lambda=mean_awaygoal), xlab = "Away Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)

  AwayGoal_pois=c(dpois(0,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(1,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(2,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(3,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(4,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(5,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(6,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(7,mean_awaygoal)*sum(table_for_awaygoal))

real_vs_poison_awaygoal=data.table(Real_AwayGoal=table_for_awaygoal,Poison_AwayGoal=Awa
yGoal_pois)

ggplot(real_vs_poison_awaygoal, aes(real_vs_poison_awaygoal$Real_AwayGoal.V1,
cumsum(real_vs_poison_awaygoal$Real_AwayGoal.N))) +
geom_step(aes(group=1))+
```

```
  ggtitle("CDF of Real Away Goals")+
  xlab("Number of Away Goals")+
  ylab("Cumulative Away Goals")+
  ylim(500, 3500)
```

vi. Code for Figure 8:

```
summary_by_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary_by_homegoal$HomeGoal)
table_for_homegoal=table(summary_by_homegoal$HomeGoal)
table_for_homegoal

hist(summary_by_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab
= "Number of Games", las =1, breaks = 30,col='light blue')
mean_homegoal=mean(matches$HomeGoal,na.rm = T) mean_homegoal

par(new=TRUE)
plot(dpois(x=0:8,lambda=mean_homegoal), xlab = "Home Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)

  HomeGoal_pois=c(dpois(0,mean_homegoal)*sum(table_for_homegoal),
                  dpois(1,mean_homegoal)*sum(table_for_homegoal),
                  dpois(2,mean_homegoal)*sum(table_for_homegoal),
                  dpois(3,mean_homegoal)*sum(table_for_homegoal),
                  dpois(4,mean_homegoal)*sum(table_for_homegoal),
                  dpois(5,mean_homegoal)*sum(table_for_homegoal),
                  dpois(6,mean_homegoal)*sum(table_for_homegoal),
                  dpois(7,mean_homegoal)*sum(table_for_homegoal),
                  dpois(8,mean_homegoal)*sum(table_for_homegoal))

real_vs_poison_homegoal=data.table(Real_HomeGoal=table_for_homegoal,Poison_HomeGoal=Ho
meGoal_pois)
ggplot(real_vs_poison_homegoal, aes(real_vs_poison_homegoal$Real_HomeGoal.V1,
cumsum(real_vs_poison_homegoal$Poison_HomeGoal))) + geom_step(aes(group=1))+

  ggtitle("CDF of Poisson Home Goals")+
  xlab("Number of Home Goals")+
  ylab("Cumulative HomeGoals")+
  ylim(500, 3500)
```

vii. Code for Figure 9:

```
summary_by_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]
factor(summary_by_awaygoal$AwayGoal)
table_for_awaygoal=table(summary_by_awaygoal$AwayGoal)
table_for_awaygoal
hist(summary_by_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab
= "Number of Games",las=1, breaks = 30,col='light blue')

mean_awaygoal=mean(matches$AwayGoal,na.rm = T)
par(new=TRUE)
```

```
plot(dpois(x=0:7,lambda=mean_awaygoal), xlab = "Away Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)



  AwayGoal_pois=c(dpois(0,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(1,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(2,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(3,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(4,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(5,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(6,mean_awaygoal)*sum(table_for_awaygoal),
                  dpois(7,mean_awaygoal)*sum(table_for_awaygoal))

real_vs_poison_awaygoal=data.table(Real_AwayGoal=table_for_awaygoal,Poison_AwayGoal=Awa
yGoal_pois)

ggplot(real_vs_poison_awaygoal, aes(real_vs_poison_awaygoal$Real_AwayGoal.V1,
cumsum(real_vs_poison_awaygoal$Poison_AwayGoal))) +
geom_step(aes(group=1))+
  ggtitle("CDF of Poisson Away Goals")+
  xlab("Number of Away Goals")+
  ylab("Cumulative Away Goals")+
  ylim(500, 3500)
```

viii. Code for Figure 10:

```
filtered_odds=odds[betType=='1x2' & bookmaker=='Pinnacle']
filtered_odds[,c('betType','bookmaker','totalhandicap'):=NULL]
filtered_odds=filtered_odds[order(matchId, oddtype,date)]
latest_odds=filtered_odds[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
help(dcast)
latest_odds=dcast(latest_odds,matchId~oddtype,value.var='final_odd')
temp=matches[,list(matchId,date_of_match,home,away,MatchResult)]
matches_with_odds=merge(temp,latest_odds,by='matchId')
summary_odds_by_result=matches_with_odds[,list(mean_home=mean(odd1),

mean_draw=mean(oddX),mean_away=mean(odd2),.N),by=list(MatchResult)]

matches_with_odds[,prob_home:=1/odd1]
matches_with_odds[,prob_draw:=1/oddX]
matches_with_odds[,prob_away:=1/odd2]

matches_with_odds[,total_prob:=prob_home+prob_draw+prob_away]
matches_with_odds[,home_away_diff:=prob_home-prob_away]

plot(matches_with_odds[,list(home_away_diff,prob_draw)])
cut_levels=c(-20:20)/20
```

```
matches_with_odds[,diff_bucket:=cut(home_away_diff,cut_levels)]

result_summary=matches_with_odds[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,draw_prob_bookmaker=mean(prob_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

ix. Code for Figure 11:

```
filtered_odds2=odds[betType=='1x2' & bookmaker=='10Bet']
filtered_odds2[,c('betType','bookmaker','totalhandicap'):=NULL]
filtered_odds2=filtered_odds2[order(matchId, oddtype,date)]
latest_odds2=filtered_odds2[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds2=dcast(latest_odds2,matchId~oddtype,value.var='final_odd')

temp2=matches[,list(matchId,date_of_match,home,away,MatchResult)]
matches_with_odds2=merge(temp2,latest_odds2,by='matchId')
summary_odds_by_result2=matches_with_odds2[,list(mean_home=mean(odd1),

mean_draw=mean(oddX),mean_away=mean(odd2),.N),by=list(MatchResult)]

matches_with_odds2[,prob_home:=1/odd1]
matches_with_odds2[,prob_draw:=1/oddX]
matches_with_odds2[,prob_away:=1/odd2]

matches_with_odds2[,total_prob:=prob_home+prob_draw+prob_away]
matches_with_odds2[,home_away_diff:=prob_home-prob_away]

plot(matches_with_odds2[,list(home_away_diff,prob_draw)])
matches_with_odds2[,diff_bucket:=cut(home_away_diff,cut_levels)]

result_summary2=matches_with_odds2[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,draw_prob_bookmaker=mean(prob_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xi. Code for Figure 12:

```
matches_with_odds[,P_home:=prob_home/total_prob]
matches_with_odds[,P_away:=prob_away/total_prob]
matches_with_odds[,P_draw:=prob_draw/total_prob]
matches_with_odds[,P_home_away_diff:=P_home-P_away]

P_summary_odds_by_result=matches_with_odds[,list(mean_home=mean(P_home),

mean_draw=mean(P_draw),mean_away=mean(P_away),.N),by=list(MatchResult)]


plot(matches_with_odds[,list(P_home_away_diff,P_draw)])

matches_with_odds[,P_diff_bucket:=cut(P_home_away_diff,cut_levels)]
```

```
P_result_summary=matches_with_odds[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,P_draw_prob_bookmaker=mean(P_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xii. Code for Figure 13:

```
matches_with_odds2[,P_home:=prob_home/total_prob]
matches_with_odds2[,P_away:=prob_away/total_prob]
matches_with_odds2[,P_draw:=prob_draw/total_prob]
matches_with_odds2[,P_home_away_diff:=P_home-P_away]

P_summary_odds_by_result2=matches_with_odds2[,list(mean_home=mean(P_home),

mean_draw=mean(P_draw),mean_away=mean(P_away),.N),by=list(MatchResult)]

plot(matches_with_odds2[,list(P_home_away_diff,P_draw)])

matches_with_odds2[,P_diff_bucket:=cut(P_home_away_diff,cut_levels)]
P_result_summary2=matches_with_odds2[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,P_draw_prob_bookmaker=mean(P_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xiii. Code for Figure 14:

```
filtered_odds3=odds[betType=='1x2' & bookmaker=='Betsafe']
filtered_odds3[,c('betType','bookmaker','totalhandicap'):=NULL]
filtered_odds3=filtered_odds3[order(matchId, oddtype,date)]
latest_odds3=filtered_odds3[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds3=dcast(latest_odds3,matchId~oddtype,value.var='final_odd')

temp3=matches[,list(matchId,date_of_match,home,away,MatchResult)]
matches_with_odds3=merge(temp3,latest_odds3,by='matchId')
summary_odds_by_result3=matches_with_odds3[,list(mean_home=mean(odd1),

mean_draw=mean(oddX),mean_away=mean(odd2),.N),by=list(MatchResult)]

matches_with_odds3[,prob_home:=1/odd1]
matches_with_odds3[,prob_draw:=1/oddX]
matches_with_odds3[,prob_away:=1/odd2]

matches_with_odds3[,total_prob:=prob_home+prob_draw+prob_away]
matches_with_odds3[,home_away_diff:=prob_home-prob_away]

plot(matches_with_odds3[,list(home_away_diff,prob_draw)])
matches_with_odds3[,diff_bucket:=cut(home_away_diff,cut_levels)]
```

```
result_summary3=matches_with_odds3[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,draw_prob_bookmaker=mean(prob_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xiv. Code for Figure 15:

```
filtered_odds4=odds[betType=='1x2' & bookmaker=='bet365']
filtered_odds4[,c('betType','bookmaker','totalhandicap'):=NULL]
filtered_odds4=filtered_odds4[order(matchId, oddtype,date)]
latest_odds4=filtered_odds4[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds4=dcast(latest_odds4,matchId~oddtype,value.var='final_odd')

temp4=matches[,list(matchId,date_of_match,home,away,MatchResult)]
matches_with_odds4=merge(temp4,latest_odds4,by='matchId')
summary_odds_by_result4=matches_with_odds4[,list(mean_home=mean(odd1),

mean_draw=mean(oddX),mean_away=mean(odd2),.N),by=list(MatchResult)]

matches_with_odds4[,prob_home:=1/odd1]
matches_with_odds4[,prob_draw:=1/oddX]
matches_with_odds4[,prob_away:=1/odd2]

matches_with_odds4[,total_prob:=prob_home+prob_draw+prob_away]
matches_with_odds4[,home_away_diff:=prob_home-prob_away]

plot(matches_with_odds4[,list(home_away_diff,prob_draw)])
matches_with_odds4[,diff_bucket:=cut(home_away_diff,cut_levels)]

result_summary4=matches_with_odds4[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,draw_prob_bookmaker=mean(prob_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xv. Code for Figure 16:

```
matches_with_odds3[,P_home:=prob_home/total_prob]
matches_with_odds3[,P_away:=prob_away/total_prob]
matches_with_odds3[,P_draw:=prob_draw/total_prob]
matches_with_odds3[,P_home_away_diff:=P_home-P_away]

P_summary_odds_by_result3=matches_with_odds3[,list(mean_home=mean(P_home),

mean_draw=mean(P_draw),mean_away=mean(P_away),.N),by=list(MatchResult)]
```

```
plot(matches_with_odds3[,list(P_home_away_diff,P_draw)])

matches_with_odds3[,P_diff_bucket:=cut(P_home_away_diff,cut_levels)]
P_result_summary3=matches_with_odds3[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,P_draw_prob_bookmaker=mean(P_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xvi. Code for Figure 17:

```
matches_with_odds4[,P_home:=prob_home/total_prob]
matches_with_odds4[,P_away:=prob_away/total_prob]
matches_with_odds4[,P_draw:=prob_draw/total_prob]
matches_with_odds4[,P_home_away_diff:=P_home-P_away]

P_summary_odds_by_result4=matches_with_odds4[,list(mean_home=mean(P_home),

mean_draw=mean(P_draw),mean_away=mean(P_away),.N),by=list(MatchResult)]

plot(matches_with_odds4[,list(P_home_away_diff,P_draw)])

matches_with_odds4[,P_diff_bucket:=cut(P_home_away_diff,cut_levels)]
P_result_summary4=matches_with_odds4[,list(real_draw_ratio=sum(MatchResult=='draw', na.rm
= T)/.N,P_draw_prob_bookmaker=mean(P_draw[MatchResult=='draw'], na.rm =
T)),by=list(diff_bucket)]
```

xvii. Code for Figure 18:

```
names(P_result_summary)[1]<-("Difference Bucket")
names(P_result_summary)[2]<-("Real Draw Ratio")
names(P_result_summary)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary)
plot(P_result_summary$`Difference Bucket`,
 P_result_summary$`Real Draw Ratio` - P_result_summary$`Normalized Bookmaker Draw
Ratio`,
 xlab ="Difference Buckets", ylab = "Real Draw Ratio - Normalized Bookmaker Draw Ratio",
 main = "Pinnacle Ratio and Real Ratio")
```

xix. Code for Figure 19:

```
names(P_result_summary3)[1]<-("Difference Bucket")
names(P_result_summary3)[2]<-("Real Draw Ratio")
names(P_result_summary3)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary3)
plot(P_result_summary3$`Difference Bucket`,
 P_result_summary3$`Real Draw Ratio` - P_result_summary3$`Normalized Bookmaker Draw
Ratio`,
 xlab ="Difference Buckets", ylab = "Real Draw Ratio - Normalized Bookmaker Draw Ratio",
```

```
main = "Betsafe Ratio and Real Ratio")
```

xx. Code for Figure 20:

```
names(P_result_summary2)[1]<-("Difference Bucket")
names(P_result_summary2)[2]<-("Real Draw Ratio")
names(P_result_summary2)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary2)
plot(P_result_summary2$`Difference Bucket`,
P_result_summary2$`Real Draw Ratio` - P_result_summary2$`Normalized Bookmaker
Draw Ratio`,
xlab ="Difference Buckets", ylab = "Real Draw Ratio - Normalized Bookmaker Draw Ratio",
main = "10Bet Ratio and Real Ratio")
```

xxi. Code for Figure 21:

```
names(P_result_summary4)[1]<-("Difference Bucket")
names(P_result_summary4)[2]<-("Real Draw Ratio")
names(P_result_summary4)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary4)
plot(P_result_summary4$`Difference Bucket`,
P_result_summary4$`Real Draw Ratio` - P_result_summary4$`Normalized Bookmaker
Draw Ratio`,
xlab ="Difference Buckets", ylab = "Real Draw Ratio - Normalized Bookmaker Draw Ratio",
main = "bet365 Ratio and Real Ratio")
```

xvii. Code for Figure 22:

```
filtered_odds2=odds[betType=='ou'& bookmaker=='bet365'& totalhandicap==2.5]

filtered_odds2=filtered_odds2[order(matchId,date)]
latest_odds2=filtered_odds2[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds2=dcast(latest_odds2,matchId~oddtype,value.var='final_odd')

latest_odds2[,prob_over:=1/over]
latest_odds2[,prob_under:=1/under]
Total_odds2=latest_odds2$prob_over+latest_odds2$prob_under

latest_odds2[,Total_odds2:=latest_odds2$prob_over+latest_odds2$prob_under]

latest_odds2[,P_over:=prob_over/Total_odds2]
latest_odds2[,P_under:=prob_under/Total_odds2]
```

```
cut_levels=c(0:20)/20
latest_odds2[,diff_bucket:=cut(prob_over,cut_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]
temp2=matches[,list(matchId,date,TotalGoal)]

latest_odds2=merge(temp2,latest_odds2,by='matchId')
```

```
result_summary2=latest_odds2[, list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,
bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r2=result_summary2$real_over_ratio
b2=result_summary2$bookmaker_over_ratio

ggplot(result_summary2,aes(x=b2,
 y=real_over_r2))+ geom_point()+
 geom_abline(slope = 1, intercept = 0)+
 ggtitle("Scatter Plot of Over Ratios for
 bet356")+ xlab("Bookmaker Over Ratio")+
 ylab("Real Over Ratio")+
 xlim(0,1)+
 ylim(0,1)
```

xviii. Code for Figure 23:

```
filtered_odds3=odds[betType=='ou'& bookmaker=='Betsafe'& totalhandicap==2.5]

filtered_odds3=filtered_odds3[order(matchId,date)]
latest_odds3=filtered_odds3[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds3=dcast(latest_odds3,matchId~oddtype,value.var='final_odd')

latest_odds3[,prob_over:=1/over]
latest_odds3[,prob_under:=1/under]
Total_odds3=latest_odds3$prob_over+latest_odds3$prob_under

latest_odds3[,Total_odds3:=latest_odds3$prob_over+latest_odds3$prob_under]

latest_odds3[,P_over:=prob_over/Total_odds3]
latest_odds3[,P_under:=prob_under/Total_odds3]

cut_levels=c(0:20)/20
latest_odds3[,diff_bucket:=cut(prob_over,cut_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]
temp3=matches[,list(matchId,date,TotalGoal)]

latest_odds3=merge(temp3,latest_odds3,by='matchId')

result_summary3=latest_odds3[,
                    list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r3=result_summary3$real_over_ratio
b3=result_summary3$bookmaker_over_ratio
```

```
ggplot(result_summary3,aes(x=b3, y=real_over_r3))+
  geom_point()+
```

```
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Over Ratios for
  Betsafe")+ xlab("Bookmaker Over Ratio")+
  ylab("Real Over Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xix. Code for Figure 24:

```
filtered_odds4=odds[betType=='ou'& bookmaker=='BetVictor'& totalhandicap==2.5]

filtered_odds4=filtered_odds4[order(matchId,date)]
latest_odds4=filtered_odds4[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds4=dcast(latest_odds4,matchId~oddtype,value.var='final_odd')

latest_odds4[,prob_over:=1/over]
latest_odds4[,prob_under:=1/under]
Total_odds4=latest_odds4$prob_over+latest_odds4$prob_under

latest_odds4[,Total_odds4:=latest_odds4$prob_over+latest_odds4$prob_under]

latest_odds4[,P_over:=prob_over/Total_odds4]
latest_odds4[,P_under:=prob_under/Total_odds4]

cut_levels=c(0:20)/20
latest_odds4[,diff_bucket:=cut(prob_over,cut_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]
temp4=matches[,list(matchId,date,TotalGoal)]

latest_odds4=merge(temp4,latest_odds4,by='matchId')

result_summary4=latest_odds4[,
                    list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r4=result_summary4$real_over_ratio
b4=result_summary4$bookmaker_over_ratio

ggplot(result_summary4,aes(x=b4, y=real_over_r4))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Over Ratios for BetVictor")+
  xlab("Bookmaker Over Ratio")+ ylab("Real Over
  Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xx. Code for Figure 25:

```
filtered_odds=odds[betType=='ou'& bookmaker=='Pinnacle'& totalhandicap==2.5]

filtered_odds=filtered_odds[order(matchId,date)]
latest_odds=filtered_odds[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds=dcast(latest_odds,matchId~oddtype,value.var='final_odd')

latest_odds[,prob_over:=1/over]
latest_odds[,prob_under:=1/under]
Total_odds=latest_odds$prob_over+latest_odds$prob_under

latest_odds[,Total_odds:=latest_odds$prob_over+latest_odds$prob_under]

latest_odds[,P_over:=prob_over/Total_odds]
latest_odds[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
latest_odds[,diff_bucket:=cut(prob_over,cut_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]
temp=matches[,list(matchId,date,TotalGoal)]

latest_odds=merge(temp,latest_odds,by='matchId')

result_summary=latest_odds[,list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,
bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r=result_summary$real_over_ratio
b=result_summary$bookmaker_over_ratio

ggplot(result_summary,aes(x=b, y=real_over_r))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Over Ratios for Pinnacle")+
  xlab("Bookmaker Over Ratio")+ ylab("Real Over
  Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xvi. Code for Figure 26:

```
filtered_odds2=odds[betType=='ou'& bookmaker=='bet365'&
totalhandicap==2.5] filtered_odds_for_init2=filtered_odds2[order(matchId,date,
decreasing = TRUE)] filtered_odds2=filtered_odds2[order(matchId,date)]

initial_odds2=filtered_odds_for_init2[,list(initial_odd=odd[.N]),by=list(matchId,oddtype)]
initial_odds2=dcast(initial_odds2,matchId~oddtype,value.var='initial_odd')
latest_odds2=filtered_odds2[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds2=dcast(latest_odds2,matchId~oddtype,value.var='final_odd')

initial_odds2[,prob_over:=1/over]
```

```
initial_odds2[,prob_under:=1/under]
Total_odds=initial_odds2$prob_over+initial_odds2$prob_under

initial_odds2[,Total_odds:=initial_odds2$prob_over+initial_odds2$prob_under]

initial_odds2[,P_over:=prob_over/Total_odds]
initial_odds2[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
initial_odds2[,diff_bucket:=cut(prob_over,cut_levels)]


temp=matches[,list(matchId,date,TotalGoal)]

initial_odds2=merge(temp,initial_odds2,by='matchId')

result_summary2=initial_odds2[,
                    list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r=result_summary2$real_over_ratio
b=result_summary2$bookmaker_over_ratio

##Ggpolt
ggplot(result_summary2,aes(x=b, y=real_over_r))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Initial Over Ratios for bet365")+
  xlab("Bookmaker Over Ratio")+ ylab("Real Over
  Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xvii. Code for Figure 27:

```
filtered_odds3=odds[betType=='ou'& bookmaker=='Betsafe'&
totalhandicap==2.5] filtered_odds_for_init3=filtered_odds3[order(matchId,date,
decreasing = TRUE)] filtered_odds3=filtered_odds3[order(matchId,date)]

initial_odds3=filtered_odds_for_init3[,list(initial_odd=odd[.N]),by=list(matchId,oddtype)]
initial_odds3=dcast(initial_odds3,matchId~oddtype,value.var='initial_odd')
latest_odds3=filtered_odds3[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds3=dcast(latest_odds3,matchId~oddtype,value.var='final_odd')

initial_odds3[,prob_over:=1/over]
initial_odds3[,prob_under:=1/under]
Total_odds=initial_odds3$prob_over+initial_odds3$prob_under
```

```
initial_odds3[,Total_odds:=initial_odds3$prob_over+initial_odds3$prob_under]

initial_odds3[,P_over:=prob_over/Total_odds]
initial_odds3[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
initial_odds3[,diff_bucket:=cut(prob_over,cut_levels)]


temp=matches[,list(matchId,date,TotalGoal)]

initial_odds3=merge(temp,initial_odds3,by='matchId')

result_summary3=initial_odds3[,
                        list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r=result_summary3$real_over_ratio
b=result_summary3$bookmaker_over_ratio

##Ggpolt
ggplot(result_summary3,aes(x=b, y=real_over_r))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Initial Over Ratios for
  BetSafe")+ xlab("Bookmaker Over Ratio")+ ylab("Real
  Over Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xviii. Code for Figure 28:

```
filtered_odds4=odds[betType=='ou'& bookmaker=='BetVictor'& totalhandicap==2.5]
filtered_odds_for_init4=filtered_odds4[order(matchId,date, decreasing = TRUE)]
filtered_odds4=filtered_odds4[order(matchId,date)]

initial_odds4=filtered_odds_for_init4[,list(initial_odd=odd[.N]),by=list(matchId,oddtype)]
initial_odds4=dcast(initial_odds4,matchId~oddtype,value.var='initial_odd')
latest_odds4=filtered_odds4[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds4=dcast(latest_odds4,matchId~oddtype,value.var='final_odd')

initial_odds4[,prob_over:=1/over]
initial_odds4[,prob_under:=1/under]
Total_odds=initial_odds4$prob_over+initial_odds4$prob_under

initial_odds4[,Total_odds:=initial_odds4$prob_over+initial_odds4$prob_under]
```

```
initial_odds4[,P_over:=prob_over/Total_odds]
initial_odds4[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
initial_odds4[,diff_bucket:=cut(prob_over,cut_levels)]


temp=matches[,list(matchId,date,TotalGoal)]

initial_odds4=merge(temp,initial_odds4,by='matchId')

result_summary4=initial_odds4[,
                        list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r=result_summary4$real_over_ratio
b=result_summary4$bookmaker_over_ratio

##Ggpolt
ggplot(result_summary4,aes(x=b, y=real_over_r))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Initial Over Ratios for
  BetVictor")+ xlab("Bookmaker Over Ratio")+ ylab("Real
  Over Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xix. Code for Figure 29:

```
filtered_odds=odds[betType=='ou'& bookmaker=='Pinnacle'& totalhandicap==2.5]
filtered_odds_for_init=filtered_odds[order(matchId,date, decreasing = TRUE)]
filtered_odds=filtered_odds[order(matchId,date)]

initial_odds=filtered_odds_for_init[,list(initial_odd=odd[.N]),by=list(matchId,oddtype)]
initial_odds=dcast(initial_odds,matchId~oddtype,value.var='initial_odd')
latest_odds=filtered_odds[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds=dcast(latest_odds,matchId~oddtype,value.var='final_odd')

initial_odds[,prob_over:=1/over]
initial_odds[,prob_under:=1/under]
Total_odds=initial_odds$prob_over+initial_odds$prob_under

initial_odds[,Total_odds:=initial_odds$prob_over+initial_odds$prob_under]
```

```
initial_odds[,P_over:=prob_over/Total_odds]
initial_odds[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
initial_odds[,diff_bucket:=cut(prob_over,cut_levels)]


temp=matches[,list(matchId,date,TotalGoal)]

initial_odds=merge(temp,initial_odds,by='matchId')

result_summary=initial_odds[,
                    list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

real_over_r=result_summary$real_over_ratio
b=result_summary$bookmaker_over_ratio

##Ggpolt
ggplot(result_summary,aes(x=b, y=real_over_r))+
  geom_point()+
  geom_abline(slope = 1, intercept = 0)+
  ggtitle("Scatter Plot of Initial Over Ratios for
  Pinnacle")+ xlab("Bookmaker Over Ratio")+ ylab("Real
  Over Ratio")+
  xlim(0,1)+
  ylim(0,1)
```

xxx. Code for Figure 30:

```
require(lubridate)
matches[,timestamp:=as_datetime(date,tz='Turkey')]
matches[,date_of_match:=date(timestamp)]
latest_odds[,date_of_match:=date(timestamp)]

latest_odds[,timestamp:=as_datetime(date,tz='Turkey')]

filtered_odds[,timestamp:=as_datetime(date,tz='Turkey')]
odds[,timestamp:=as_datetime(date,tz='Turkey')]
```

```
temp=matches[,list(matchId,date_of_match)]
latest_odds=merge(latest_odds,temp,by='matchId')

matches_of_2011=latest_odds[date_of_match.x>'2011-01-01' & date_of_match.x<'2012-01-01']

cut_levels=c(25:30)/50
matches_of_2011[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2011=matches_of_2011[complete.cases(matches_of_2011)]

result_summary_2011=matches_of_2011[,
                        list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2012=latest_odds[date_of_match.x>'2012-01-01' & date_of_match.x<'2013-01-01']


matches_of_2012[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2012=matches_of_2012[complete.cases(matches_of_2012)]

result_summary_2012=matches_of_2012[,
                        list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2013=latest_odds[date_of_match.x>'2013-01-01' & date_of_match.x<'2014-01-01']

matches_of_2013[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2013=matches_of_2013[complete.cases(matches_of_2013)]

result_summary_2013=matches_of_2013[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2014=latest_odds[date_of_match.x>'2014-01-01' & date_of_match.x<'2015-01-01']

matches_of_2014[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2014=matches_of_2014[complete.cases(matches_of_2014)]

result_summary_2014=matches_of_2014[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]
matches_of_2015=latest_odds[date_of_match.x>'2015-01-01' & date_of_match.x<'2016-01-01']
```

```
matches_of_2015[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2015=matches_of_2015[complete.cases(matches_of_2015)]

result_summary_2015=matches_of_2015[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2016=latest_odds[date_of_match.x>'2016-01-01' & date_of_match.x<'2017-01-01']

matches_of_2016[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2016=matches_of_2016[complete.cases(matches_of_2016)]


result_summary_2016=matches_of_2016[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2017=latest_odds[date_of_match.x>'2017-01-01' & date_of_match.x<'2018-01-01']

matches_of_2017[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2017=matches_of_2017[complete.cases(matches_of_2017)]

result_summary_2017=matches_of_2017[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]
matches_of_2018=latest_odds[date_of_match.x>'2018-01-01' & date_of_match.x<'2019-01-01']

matches_of_2018[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2018=matches_of_2018[complete.cases(matches_of_2018)]

result_summary_2018=matches_of_2018[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,

bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]

matches_of_2019=latest_odds[date_of_match.x>'2019-01-01' & date_of_match.x<'2020-01-
01'] matches_of_2019[,diff_bucket:=cut(prob_over,cut_levels)]

matches_of_2019=matches_of_2019[complete.cases(matches_of_2019)]

result_summary_2019=matches_of_2019[,list(real_over_ratio=sum(TotalGoal>=3,na.rm =
TRUE)/.N,
bookmaker_over_ratio=mean(prob_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff_bucket)]
```

```
order1<-order(result_summary_2011$diff_bucket)
result_summary_2011=result_summary_2011[order1,]
order2<-order(result_summary_2012$diff_bucket)
result_summary_2012=result_summary_2012[order2,]
order3<-order(result_summary_2013$diff_bucket)
result_summary_2013=result_summary_2013[order3,]
order4<-order(result_summary_2014$diff_bucket)
result_summary_2014=result_summary_2014[order4,]
order5<-order(result_summary_2015$diff_bucket)
result_summary_2015=result_summary_2015[order5,]
order6<-order(result_summary_2016$diff_bucket)
result_summary_2016=result_summary_2016[order6,]
order7<-order(result_summary_2017$diff_bucket)
result_summary_2017=result_summary_2017[order7,]
order8<-order(result_summary_2018$diff_bucket)
result_summary_2018=result_summary_2018[order8,]
order9<-order(result_summary_2019$diff_bucket)
result_summary_2019=result_summary_2019[order9,]


plot(result_summary_2011$bookmaker_over_ratio,result_summary_2011$real_over_ratio,axes=T,c
ol='dark red', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2011$bookmaker_over_ratio,result_summary_2011$real_over_ratio,col='dark
red')
par(new=TRUE)

plot(result_summary_2012$bookmaker_over_ratio,result_summary_2012$real_over_ratio,axes=F,co
l='yellow', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2012$bookmaker_over_ratio,result_summary_2012$real_over_ratio,col='yell
ow')
par(new=TRUE)

plot(result_summary_2013$bookmaker_over_ratio,result_summary_2013$real_over_ratio,axes=F,co
l='dark blue',xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result_summary_2013$bookmaker_over_ratio,result_summary_2013$real_over_ratio,col='dark
blue')
par(new=TRUE)

plot(result_summary_2014$bookmaker_over_ratio,result_summary_2014$real_over_ratio,axes=F,co
l='black', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2014$bookmaker_over_ratio,result_summary_2014$real_over_ratio)
par(new=TRUE)

plot(result_summary_2015$bookmaker_over_ratio,result_summary_2015$real_over_ratio,axes=F,co
l='gray', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2015$bookmaker_over_ratio,result_summary_2015$real_over_ratio,col='gray
')
par(new=TRUE)
```

```
plot(result_summary_2016$bookmaker_over_ratio,result_summary_2016$real_over_ratio,axes=F,co
l='green', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2016$bookmaker_over_ratio,result_summary_2016$real_over_ratio,col='gree
n')
par(new=TRUE)

plot(result_summary_2017$bookmaker_over_ratio,result_summary_2017$real_over_ratio,axes=F,co
l='orange', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2017$bookmaker_over_ratio,result_summary_2017$real_over_ratio,col='oran
ge')
par(new=TRUE)

plot(result_summary_2018$bookmaker_over_ratio,result_summary_2018$real_over_ratio,axis=F,co
l='brown', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")
lines(result_summary_2018$bookmaker_over_ratio,result_summary_2018$real_over_ratio,col='bro
wn')
par(new=TRUE)
abline(h=0.55)
```

R codes for Tables

i. Code for Table 1:

```
summary_by_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary_by_homegoal$HomeGoal)
table_for_homegoal=table(summary_by_homegoal$HomeGoal)
table_for_homegoal

hist(summary_by_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab
= "Number of Games", las =1, breaks = 30,col='light blue')
mean_homegoal=mean(matches$HomeGoal,na.rm = T) mean_homegoal

par(new=TRUE)
plot(dpois(x=0:8,lambda=mean_homegoal), xlab = "Home Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)

  HomeGoal_pois=c(dpois(0,mean_homegoal)*sum(table_for_homegoal),
                  dpois(1,mean_homegoal)*sum(table_for_homegoal),
                  dpois(2,mean_homegoal)*sum(table_for_homegoal),
                  dpois(3,mean_homegoal)*sum(table_for_homegoal),
                  dpois(4,mean_homegoal)*sum(table_for_homegoal),
                  dpois(5,mean_homegoal)*sum(table_for_homegoal),
                  dpois(6,mean_homegoal)*sum(table_for_homegoal),
                  dpois(7,mean_homegoal)*sum(table_for_homegoal),
                  dpois(8,mean_homegoal)*sum(table_for_homegoal))

real_vs_poison_homegoal=data.table(Real_HomeGoal=table_for_homegoal,Poison_HomeGoal=Ho
meGoal_pois)
```

ii. Code for Table 2:

```
summary_by_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]factor(summary_by_a
waygoal$AwayGoal)

table_for_awaygoal=table(summary_by_awaygoal$AwayGoal)
table_for_awaygoal
hist(summary_by_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab
= "Number of Games",las=1, breaks = 30,col='light blue')
mean_awaygoal=mean(matches$AwayGoal,na.rm = T)
par(new=TRUE)
plot(dpois(x=0:7,lambda=mean_awaygoal), xlab = "Away Goals",ylab="Number
of Games",axes=F,col='dark red',pch=19)

  AwayGoal_pois=c(dpois(0,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(1,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(2,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(3,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(4,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(5,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(6,mean_awaygoal)*sum(table_for_awaygoal),
            dpois(7,mean_awaygoal)*sum(table_for_awaygoal))


real_vs_poison_awaygoal=data.table(Real_AwayGoal=table_for_awaygoal,Poison_AwayGoal=Awa
yGoal_pois)
```

iii. Code for Table 3,4,5,6: By proceeding Pinnacle's, Betsafe's, bet365's, Bet10's Normalized P(home win)-P(away win) vs Normalized P(tie) codes in figure 12,13,16,17

```
names(P_result_summary)[1]<-("Difference Bucket")
names(P_result_summary)[2]<-("Real Draw Ratio")
names(P_result_summary)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary)

names(P_result_summary2)[1]<-("Difference Bucket")
names(P_result_summary2)[2]<-("Real Draw Ratio")
names(P_result_summary2)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary2)

names(P_result_summary3)[1]<-("Difference Bucket")
names(P_result_summary3)[2]<-("Real Draw Ratio")
names(P_result_summary3)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P_result_summary3)

names(P_result_summary4)[1]<-("Difference Bucket")
names(P_result_summary4)[2]<-("Real Draw Ratio")
names(P_result_summary4)[3]<-("Normalized Bookmaker Draw Ratio")
```

```
grid.table(P_result_summary4)
```

iv. Code for Table 7:

```
filtered_odds=odds[betType=='ou'& bookmaker=='Pinnacle'& totalhandicap==2.5]

filtered_odds=filtered_odds[order(matchId,date)]
latest_odds=filtered_odds[,list(final_odd=odd[.N]),by=list(matchId,oddtype)]
latest_odds=dcast(latest_odds,matchId~oddtype,value.var='final_odd')

latest_odds[,prob_over:=1/over]
latest_odds[,prob_under:=1/under]
Total_odds=latest_odds$prob_over+latest_odds$prob_under

latest_odds[,Total_odds:=latest_odds$prob_over+latest_odds$prob_under]

latest_odds[,P_over:=prob_over/Total_odds]
latest_odds[,P_under:=prob_under/Total_odds]

cut_levels=c(0:20)/20
latest_odds[,diff_bucket:=cut(prob_over,cut_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]
temp=matches[,list(matchId,date,TotalGoal)]

latest_odds=merge(temp,latest_odds,by='matchId')

result_summary=latest_odds[,
                    list(real_over_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,
```