

# BI Analyst

Technical Assignment

# Dataset

4 dataset were given to analyze the data

1. **Bet Data** : Contains information of the bets placed
2. **Betting Event**: Contains mapping of which betting event corresponds to which event type
3. **User Profile**: Contains user demographic information
4. **Sports Name**: - Contains mapping of which event type corresponds to which sport/game

# Tool

Tool I could use are R programming in R studio, SQL, Tableau

I can use these all tools but I choose to do this in Rstudio, I like R programming and you can get the idea of my r Programming skills

R studio was used to read the data and for data cleaning, data processing and data transformation.

R studio was also used for data visualization. R studio is considered to be best language for data analysis because of its packages which include best libraries and functions for statistical analyzing and visualizing.

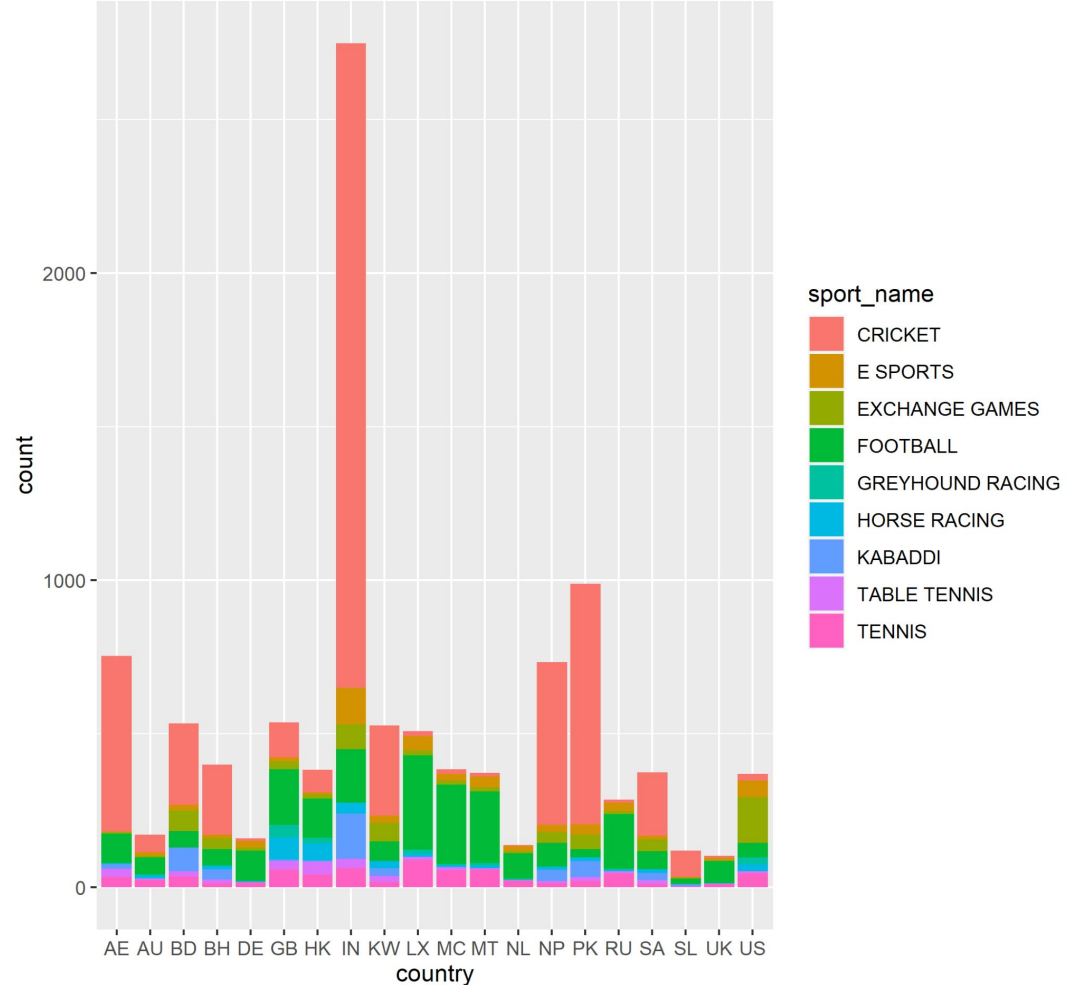
Few important Packages include

- **Tidyverse**: For data cleaning and processing
- **ggplot2**: For graphical representation of data

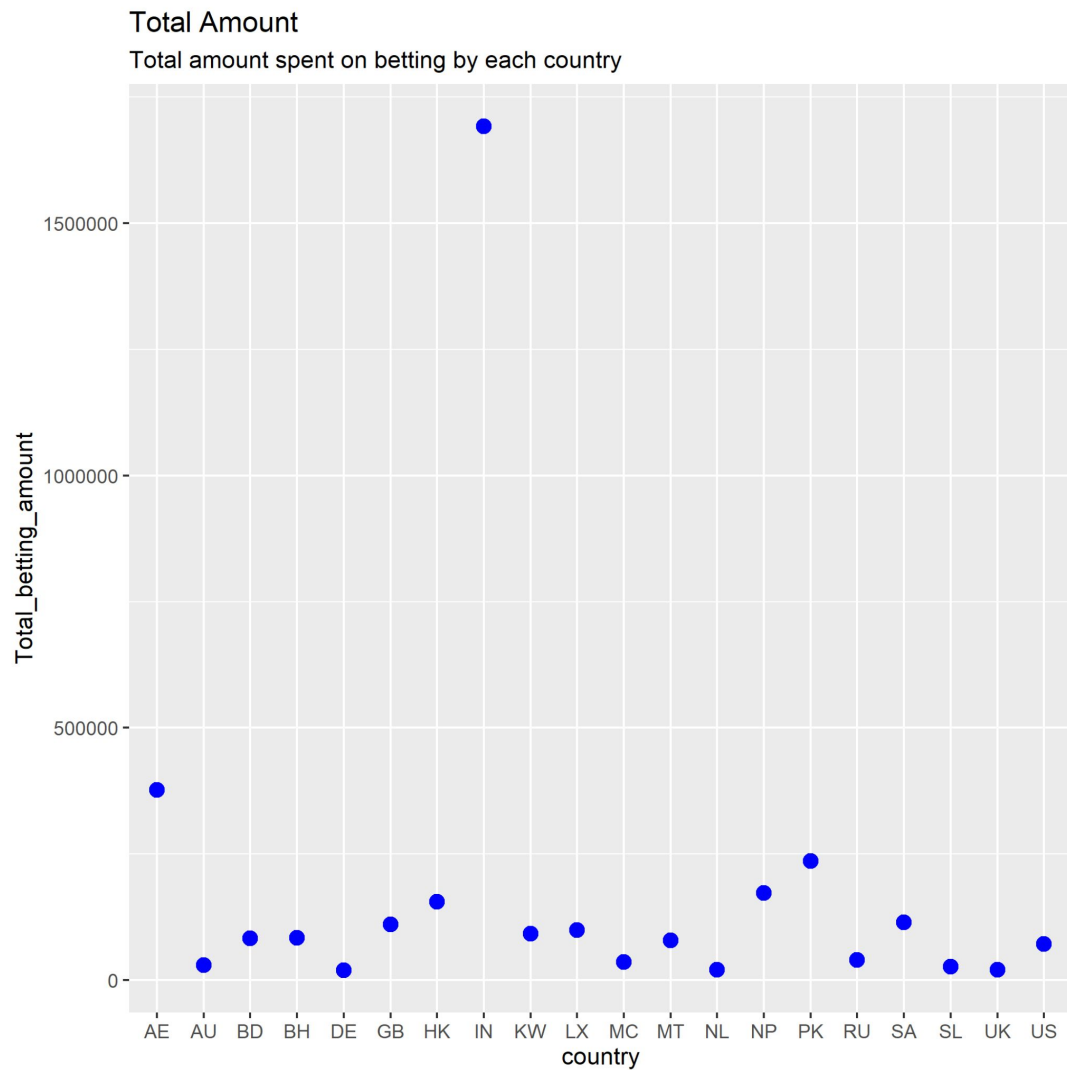
Data was format and clean for analysis and we merge all 4 dataset to one for better comparison

	Filter													
	bet_id	event_id	user_id	bet_size	outcome	odds	bet_time	bet_date	event_type_id	sport_name	gender	age	country	
1	2102181920249260	1NHSL	79240	17042.26	WON	1.61	19:37:48	2021-02-18		1	FOOTBALL	M	49	AE
2	2102172009450390	1OZAC	566741	40.73	WON	1.65	21:54:03	2021-02-17		1	FOOTBALL	M	64	AE
3	2101122023431390	1HATX	633820	2.04	WON	1.66	21:10:26	2021-01-12		1	FOOTBALL	M	19	AE
4	2102232005562290	1HATX	501753	188.23	LOST	1.66	20:43:41	2021-02-23		1	FOOTBALL	M	51	AE
5	2101140837040130	1YNKJ	500055	72.10	LOST	1.67	05:04:23	2021-01-18		1	FOOTBALL	M	30	IN
6	2102181806242890	1YNKJ	460365	9.00	LOST	1.67	19:49:30	2021-02-18		1	FOOTBALL	M	31	IN
7	2102162030539060	1ETEB	544967	1.47	WON	1.72	21:27:31	2021-02-16		1	FOOTBALL	M	39	IN
8	2102181827259480	1KZDK	626779	1.07	WON	1.75	19:02:21	2021-02-18		1	FOOTBALL	M	45	IN
9	2102181923309770	1KZDK	217230	54.65	LOST	1.75	19:39:34	2021-02-18		1	FOOTBALL	M	48	IN
10	2101132005068360	1UXQG	440188	32.59	LOST	1.76	21:38:54	2021-01-13		1	FOOTBALL	M	62	AE
11	2101131955587310	1NRGD	453649	2.01	WON	1.77	21:38:54	2021-01-13		1	FOOTBALL	M	47	IN
12	2102172127536680	1FGZU	494686	3.34	WON	1.80	21:42:33	2021-02-17		1	FOOTBALL	M	41	IN
13	2102181759319830	1BWAU	465414	137.36	LOST	1.83	19:43:38	2021-02-18		1	FOOTBALL	M	30	AE
14	2102181802361060	1KNOP	526111	5.86	WON	1.84	18:29:49	2021-02-18		1	FOOTBALL	M	37	AE
15	2102162017496890	1DPFH	633820	2.04	LOST	1.86	21:52:25	2021-02-16		1	FOOTBALL	M	19	AE
16	2102181859325890	1DPFH	79240	8521.13	WON	1.86	19:18:42	2021-02-18		1	FOOTBALL	M	49	AE
17	2101140827574640	1FCXK	619920	1687.29	LOST	1.87	05:04:23	2021-01-18		1	FOOTBALL	M	27	IN
18	2102231934027530	1YTIS	478465	3.09	WON	1.90	21:53:05	2021-02-23		1	FOOTBALL	M	41	AE
19	2102181902166190	1KMVS	79240	17042.26	WON	1.92	19:39:34	2021-02-18		1	FOOTBALL	M	49	AE

It clearly show from india most of the bet are made.



This Point graph clearly tell us that most amount on betting is from india and other countries are much below.



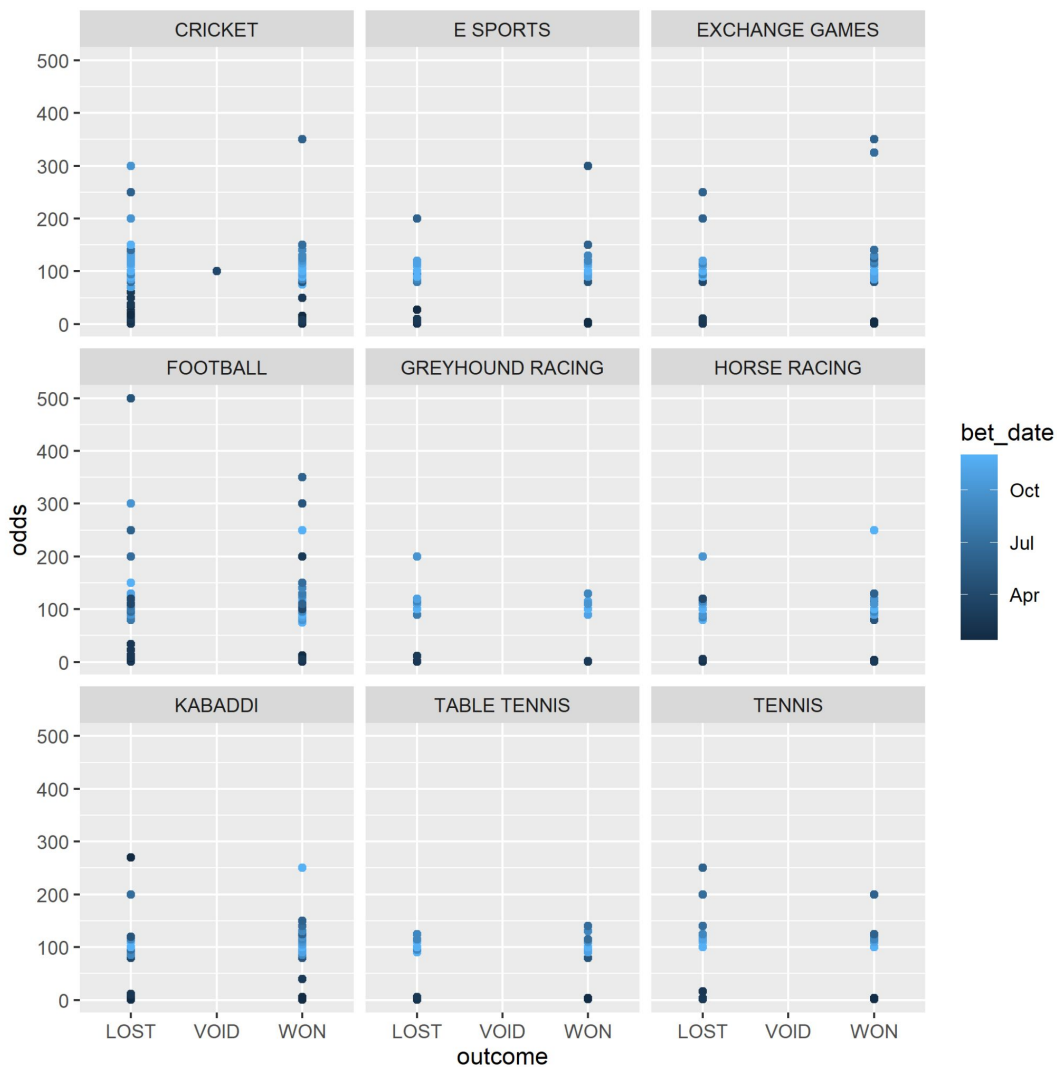
To understand more properly.

This Graph represent the data of Odds vs outcome by considering bet date also.

In simple words-

“When the bet was won or lost by what odds on each sports”

This help a lot in understanding bets on each sports

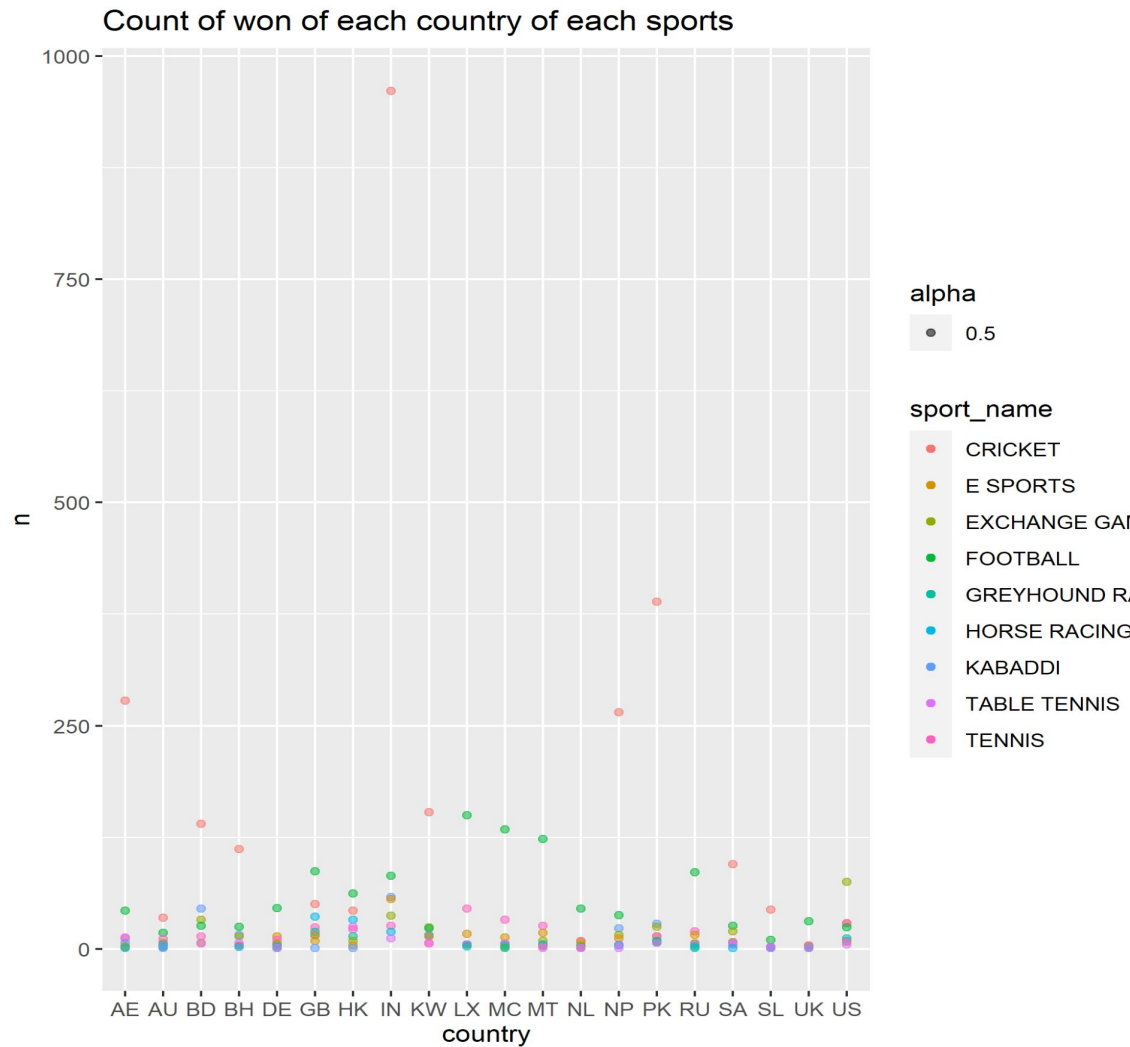


n = count of won of each sports from each country.

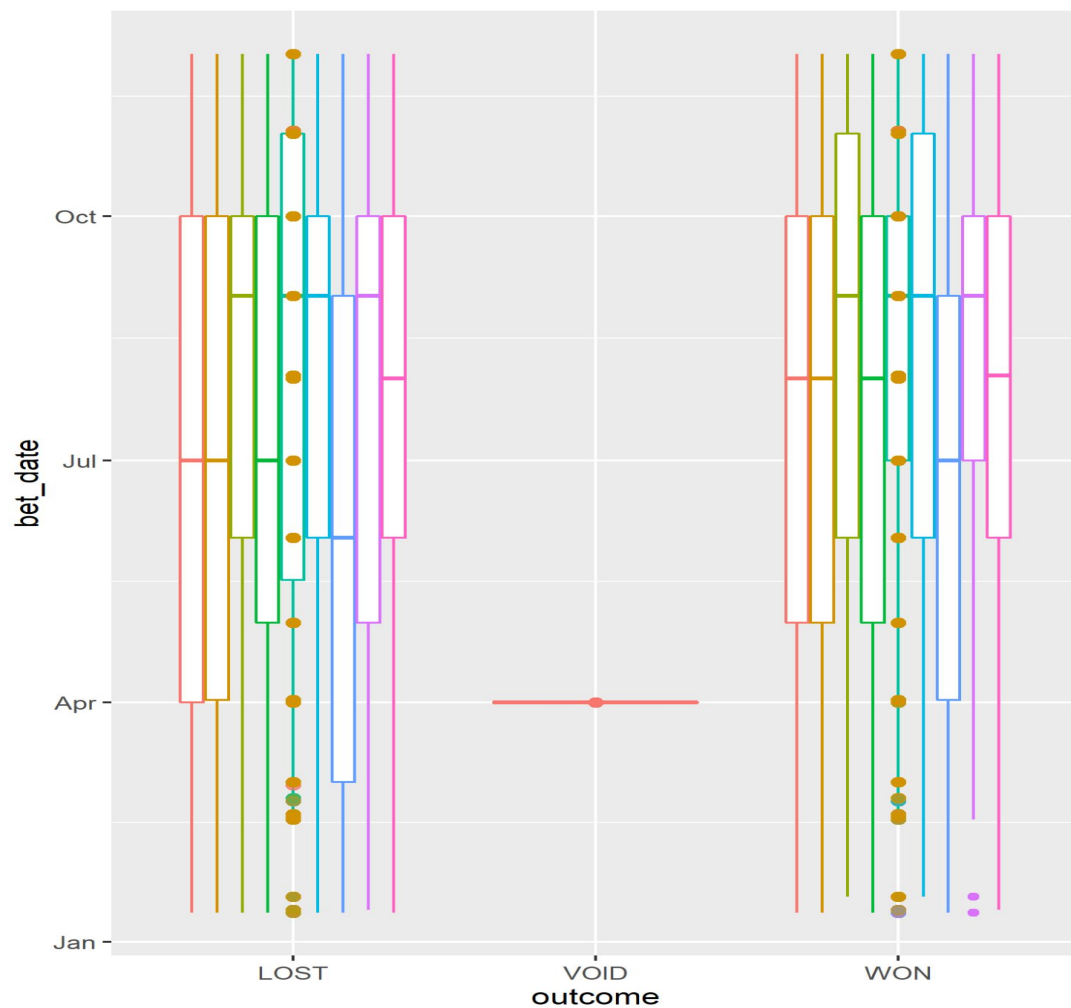
I create this plot for verification. This plot clearly clarify that India do most of the betting and they bet most on cricket and also won most of the bet on cricket.

Pakistan do the second most and also on cricket.

Least bet is on Tennis and table tennis.







This Plot may seem confusing but it not.

This represent the time period/range that when the bet on each sports is placed divided into lost and won section.

Purpose of this plot is to also understand and extract useful information.