IPL Auction Price Data

- The Indian Premier League (IPL) is a professional League for Twenty20(T20) cricket championship that was started in 2008 in India.
- The first IPL auction was held in 2008 for ownership of the team for 10 years with a base price of USD 50 million.
- The franchise acquires players, through an English auction that is conducted every year. However, there are several rules imposed by the IPL for example there is a maximum cap on the money a franchise can spend on buying players.
- The performance of the players could be measured through several metrics. Although the IPL follows the T20 format of the game, it is possible that the performance of the players in the other formats of the game such as Test matches, One-Day matches can influence player pricing. Few players have excellent records in the Test Matches, but their records in T20 matches are not very impressive.
- The Dataset consists of the performance of 130 Players measured through various performance metrics such as Batting strike rate, economy rate, wicket in Tets matches, Bowling average..etc in the year 2013

Load Dataset

In [ ]:
```
ipl <- read.csv("IPL IMB381IPL2013.csv")
```

In [ ]:
```

```

In [ ]:
```
# View the Strucutre of DataFrame
str(ipl)
```

In [ ]:
```
# Get the Total numbers of Columns
colnames(ipl)
```

In [ ]:
```
# Shape/Dimension of DataFrame
dim(ipl)
```

In [ ]:
```
nrow(ipl)
ncol(ipl)
```

In [ ]:
```
# Quick Descriptive Summary
summary(ipl)
```

In [ ]:
```
# Display First Few Records
head(ipl, 10)
```

In [ ]:
```
# Display last few records
tail(ipl, 10)
```

In [ ]:
```
# Transpose of DataFrame
t(ipl)
```

Slicing and Indexing of DataFrame

In [ ]:
```
ipl[0:5,] # Display 1-5 records
```

In [ ]:
```
total_row = nrow(ipl)
last_few_row = total_row - 5

ipl[last_few_row: total_row,]
```

In [ ]:
```
# Access Specific Columns
ipl$PLAYER.NAME
```

In [ ]:
```
head(ipl[,c("PLAYER.NAME", "COUNTRY")])
```

In [ ]:
```
# Access Specific Rows & Columns
ipl[4:9,1:4]
```

In [ ]:
```
# How Many players from different country ?
table(ipl$COUNTRY)
```

In [ ]:
```
# What % of players from different country?
prop.table(table(ipl$COUNTRY))
```

In [ ]:
```
# In Which Age group we have maximum players for each role
table(ipl$AGE, ipl$PLAYING.ROLE)
```

In [ ]:
```
# Sort data by Columns

# Get the players sold price in order Minimum to Maximum sold price
head(ipl[order(ipl$SOLD.PRICE),c("PLAYER.NAME","SOLD.PRICE")])
```

In [ ]:
```
# Costly Players in IPL year 2013
head(ipl[order(ipl$SOLD.PRICE, decreasing=TRUE),c("PLAYER.NAME","SOLD.PRICE")])
```

In [ ]:
```

```

In [ ]:
```
# Add new columns
# Calculate premium amount on base price
ipl$Premium <- ipl$SOLD.PRICE - ipl$BASE.PRICE
```

In [ ]:
```
ipl[order(ipl$Premium, decreasing=TRUE),c("PLAYER.NAME","BASE.PRICE","SOLD.PRICE","Pre
```

In [ ]:
```
# Get Unique Role
unique(ipl['PLAYING.ROLE'])
```

Split dataset by PLAYING ROLE

split() function perform partition on the dataset by specific columns/variables.

In [ ]:
```
result_list = split(ipl, ipl$PLAYING.ROLE)
```

In [ ]:
```
str(result_list)
```

In [ ]:

Subset()

This is most effective and commonly used function which acts as a filter for the records

In [ ]:
```
subset(ipl, TEAM=='CSK')
```

In [ ]:
```
subset(ipl, TEAM=='CSK' & PLAYING.ROLE=="Bowler")
```

In [ ]:
```
subset(ipl, TEAM=='CSK' & PLAYING.ROLE %in% c("Bowler", "Allrounder"))
```

In [ ]:
```
# Subset in Row index
ipl[ipl$TEAM=="CSK", ]
```

In [ ]:
```
# Subset in Row index
ipl[ipl$TEAM=="CSK" & ipl$PLAYING.ROLE %in% c("Bowler", "Allrounder"), ]
```

In [ ]:
```
# Check Missing Value
colSums(is.na(ipl))
```

In [ ]:
```
missing_data = ipl[is.na(ipl$AGE),]
```

In [ ]:
```
complete_ipl_data = ipl[!is.na(ipl$AGE),]
```

In [ ]:
```
dim(complete_ipl_data)
```

In [ ]:
```
ipl <- complete_ipl_data
```

Data Visualization

In [ ]:
```
## Univariate Analysis
```

In [ ]:
```
ipl$SoldPriceLacs=ipl$SOLD.PRICE/ 100000
```

In [ ]:
```
# Check how many players sold price<= 5Lacs
temp <- ipl[ipl$SoldPriceLacs<=5,]
temp[order(temp$SoldPriceLacs),]

# how many players are <= 5 Lacs
dim(temp)
```

In [ ]:
```
# Check how many players sold price > 5 Lacs
temp_above <- ipl[ipl$SoldPriceLacs>5,]
temp_above[order(temp_above$SoldPriceLacs,decreasing =TRUE),]

# How many players are > 5 Lacs
dim(temp_above)
```

In [ ]:
```
#Display the frequency distribution of Sold Price
```

```
hist(ipl$SoldPriceLacs,
     xlab="Sold Price",
     border="blue",
     col="yellow")
```

In [ ]:
```
# Check how many players sold <= 5Lacs
temp <- ipl[ipl$SoldPriceLacs<=5,]
temp[order(temp$SoldPriceLacs),]
```

In [ ]:
```
temp_above <- ipl[ipl$SoldPriceLacs>5,]
temp_above[order(temp_above$SoldPriceLacs,decreasing =TRUE),]
```

In [ ]:
```
### Correlation
# method = c("pearson", "kendall", "spearman")
cor(ipl$SOLD.PRICE,
    ipl$AVE,
    method = c("kendall"))

plot(ipl$SOLD.PRICE~ipl$T.RUNS,
     main="Sold Price ~ T.Runs",
     xlab="T.RUNS",
     ylab="Sold Price")
```