# Exploratory Data Analysis with Excel

*Using Excel to analyze my friends' NBA bets*

**Gollnick's NBA Playoffs Predictions Game** was a betting game that I developed from 2018 to 2022 for the sake of having a little bit of fun with my friends during the NBA playoffs. Betting was actually free and there was no prize money.

******In order to make it easier to read, I'm replacing the original columns letters (e.g "A:A") by what the column means in square brackets (e.g: [Year])***

## FIRST HYPOTHESIS:

Participants would make conservative bets in the first two rounds of the Playoffs and more audatious bet in the last two rounds

*-- Creating columns to find if the bet was conservative:*

```
[Conservative]
= IF([Bet_Games Won by Higher Ranked team] > [Bet_Games Won by Lower Ranked
team], 1, 0)

[Conservative T/F]
= IF([Conservative] = 1; TRUE; FALSE)
```

*-- Finding out conservative bets by year and round (e.g: first round from 2021 season):*

```
[Conservative Bets Sum]
= SUMIF('2021bets'![Round]; "1st Rd"; [Conservative])
```

*-- Counting ALL bets by year and round (e.g: first round from 2021 season)*

```
[Bets count]
= COUNTIF('2021bets'![Round]'; "1st Rd")
```

*-- Calculating conservative bets %:*

```
[Conservative Bets %]
= ([Conservative bets Sum] * 100) / [Bets count]
```

**RESULT:** First hypothesis was wrong, participants would usually be conservative in all rounds (except for for the 2019 finals).

**% of the Bets that Were Conservative**

| | 2018 | 2019 | 2020 | 2021 | TOTAL |
|---|---|---|---|---|---|
| 1R | * | 87,5 | 82,8 | 82,4 | 84,0 |
| SFC | 72,5 | 66,1 | 86,7 | 82,4 | 77,7 |
| FC | 71,4 | 68,8 | 70,0 | 97,1 | 78,2 |
| FIN | 90,0 | 33,3 | 73,3 | 87,5 | 69,6 |
| | | | | | 80,3 |

**There was no betting in the first round of the 2018 season*

## SECOND HYPOTHESIS:

Conservative bets would usually pay off

*-- Creating columns to find out it the bet was right about the winning team:*

```
[Bet_Games Won Margin]
= [Bet_Games Won by Higher Ranked team] -  [Bet_Games Won by Lower Ranked team]

[Bet_Winning Team]
= IF([Bet_Games Won Margin] > 0; "Higher-Ranked Team"; "Lower-Ranked Team")
```

*-- Bringing the results of the real games to from another table (with all the real life results) to this one:*

```
[Real_Games Won Margin]
= VLOOKUP ([Series ID]; 'Real life Results Table'; [Games Won Difference]; 0)

[Real_Winning Team]
= IF([Real_Games Won Margin] > 0; "Higher-Ranked Team"; "Lower-Ranked Team")
```

*-- Comparing colums:*

```
[Bet was right]
= IF([Bet_Winning Team] = [Real_Winning Team]); 1; 0)

[Bet was right T/F]
= IF([Bet was right] = 1); TRUE; FALSE)
```

*-- Finding out conservative bets that were RIGHT, by year and round (e.g: first round from 2021 season):*

```
[Right Conservative Bets]
= SUMIFS('2021bets'![Bet was right]; [Round]; "1st Rd";
'2021bets'![Conservative]; "TRUE")
```

*-- Finding out ALL the conservative bets by year and round (e.g: first round from 2021 season):*

```
[Conservative Bets]
= COUNTIFS('2021bets'![Round]; "1st Rd"; '2021bets'![Conservative]; "TRUE")
```

*-- Calculating % of the conservative bets that were right:*

```
[Right Conservative Bets %]
= ([Right Conservative Bets] * 100) / [Conservative Bets]
```

**RESULT:** Second hypothesis was right, 76,6% of the conservative bets were right. Also, splitting that calculation by year and round may better show in which series the underdogs had a great performance.

**% of Conservative Bets That Were Right**

| | 2018 | 2019 | 2020 | 2021 | TOTAL |
|---|---|---|---|---|---|
| 1R | * | 100,0 | 98,1 | 86,6 | 94,6 |
| SFC | 79,3 | 89,2 | 26,9 | 25,0 | 48,3 |
| FC | 0,0 | 72,7 | 66,7 | 100,0 | 73,3 |
| FIN | 100,0 | 100,0 | 100,0 | 0,0 | 64,1 |
| | | | | | 76,6 |

**There was no betting in the first round of the 2018 season*

## THIRD HYPOTHESIS:

Series results: in average, participants would bet that the series from the first two rounds would end with a 4-0 or a 4-1 result; The last two rounds would be in general 4-3 or 4-2.

*-- In this case I wanted to test nested functions. Let's find the average bet of how many games each team won in a series, by year and round (e.g: first round from 2021 season)*

```
[Higher-Ranked Team, Average Games Won]
= SUMIF('2021bets'![Round]; "1st Rd";'2021bets'![Bet_Games Won by Higher
Ranked team])
/ COUNTIF('2021bets'![Round]; "1st Rd")
```

```
[Lower-Ranked Team, Average Games Won]
= SUMIF('2021bets'![Round]; "1st Rd";'2021bets'![Bet_Games Won by Lower Ranked
team])
/ COUNTIF('2021bets'![Round]; "1st Rd")
```

*-- And, for the "Total" column that would sum data from the four years up, I decided to test even more functions in the same expression:*

```
= ( SUMIF('2021bets'![Round]; "1st Rd";'2021bets'![Bet_Games Won by Lower Ranked team]) +
SUMIF('2020bets'![Round]; "1st Rd";'2020bets'![Bet_Games Won by Lower Ranked team]) +
SUMIF('2019bets'![Round]; "1st Rd";'2019bets'![Bet_Games Won by Lower Ranked team]) +
SUMIF('2018bets'![Round]; "1st Rd";'2018bets'![Bet_Games Won by Lower Ranked team]) )
/ ( COUNTIF('2021bets'![Round]; "1st Rd") +
COUNTIF('2020bets'![Round]; "1st Rd") +
COUNTIF('2019bets'![Round]; "1st Rd") +
COUNTIF('2018bets'![Round]; "1st Rd") )
```

**RESULT:** Inconclusive, as I could not see any pattern on the resulting data apart from some outliers.

**Average Participants' Bets on Games Won by Team**

| | 2018 | | 2019 | | 2020 | | 2021 | | TOTAL | |
|---|---|---|---|---|---|---|---|---|---|---|
| | High-Rank | Low-Rank | High-Rank | Low-Rank | High-Rank | Low-Rank | High-Rank | Low-Rank | High-Rank | Low-Rank |
| 1R | * | * | 3,82 | 1,85 | 3,74 | 2,01 | 3,73 | 2,25 | 3,76 | 2,05 |
| SFC | 3,45 | 2,25 | 3,38 | 2,79 | 3,85 | 2,28 | 3,75 | 2,66 | 3,63 | 2,52 |
| FC | 3,43 | 2,64 | 2,11 | 2,53 | 3,57 | 2,67 | 3,97 | 1,71 | 3,71 | 2,33 |
| FIN | 3,80 | 1,90 | 0,71 | 3,53 | 3,53 | 2,53 | 3,81 | 2,19 | 3,43 | 2,59 |

**There was no betting in the first round of the 2018 season*

*-- In the [Participant Surname] column, I used this string manipulation to keep participants discrete and reveal only their surname:*

```
[Participant Surname]
= RIGHT([Participant Full Name]; LEN([Participant Full Name]) - SEARCH(" ";
[Participant Full Name]; 1))
```