

Football part 1

Ahmed Ayman

October 22, 2022

Agenda

- 1 The question
- 2 The accepted answer
- 3 Improve: compare best
- 4 Improve: Non parametric
- 5 The power of business understanding
- 6 Answers evaluation
- 7 Manger reaction

The question



Ahmed Ayman • You

Data man

2mo • 🌐



لو افترضنا ان الزمالك خد الدورى

وانا عندى سؤال

- هل الزمالك خد الدورى عشان الزمالك حلو؟

- ولا خد الدورى عشان الاهلى وحش؟

ازاى اقدر اجاب ع السؤال ده بال Data science؟

ياريت محدش يدخلنى فى نقاشات كرويه او تعصب كروى

انا حابب اشوف الافكار من وجهة نظر احصائيه 😊

The accepted answer

- compare Zamalek result with old years results
- compare Ahly result with old years results



The accepted answer

- compare Zamalek result with old years results
- compare Ahly result with old years results



Old me Vs. New me



The accepted answer

- compare Zamalek result with old years results
- compare Ahly result with old years results



Old me

Vs.

New me



old me



current me



compare best

I get the data from the archive

season	p	play	p%
2022	77	34	75%
2021	80	34	78%
2020	89	34	87%
2019	80	34	78%
2018	88	34	86%
2017	84	34	82%
2016	76	34	76%
2015	87	38	76%
2011	61	30	68%
2010	65	30	72%
2009	63	30	70%
2008	70	30	78%
2007	73	30	81%
2006	72	26	92%
2005	74	26	95%

compare best

I get the data from the archive

season	p	play	p%
2022	77	34	75%
2021	80	34	78%
2020	89	34	87%
2019	80	34	78%
2018	88	34	86%
2017	84	34	82%
2016	76	34	76%
2015	87	38	76%
2011	61	30	68%
2010	65	30	72%
2009	63	30	70%
2008	70	30	78%
2007	73	30	81%
2006	72	26	92%
2005	74	26	95%

count	13
min	70%
mean	81%
median	84%
max	95%

I'll split into 3 clusters

Good	(70 - 78.3)
V.Good	(78.3 - 86.6)
Amazing	(86.6 - 94.9)

compare best

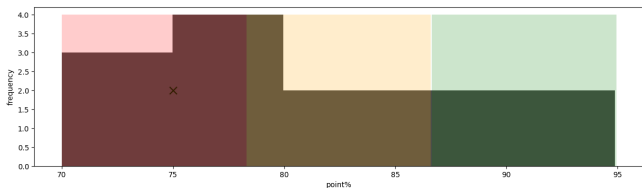
I get the data from the archive

season	p	play	p%
2022	77	34	75%
2021	80	34	78%
2020	89	34	87%
2019	80	34	78%
2018	88	34	86%
2017	84	34	82%
2016	76	34	76%
2015	87	38	76%
2011	61	30	68%
2010	65	30	72%
2009	63	30	70%
2008	70	30	78%
2007	73	30	81%
2006	72	26	92%
2005	74	26	95%

count	13
min	70%
mean	81%
median	84%
max	95%

I'll split into 3 clusters

Good	(70 - 78.3)
V.Good	(78.3 - 86.6)
Amazing	(86.6 - 94.9)



- Zamalek, not above expectation
- The main reason for wining, because Ahly was pathetic

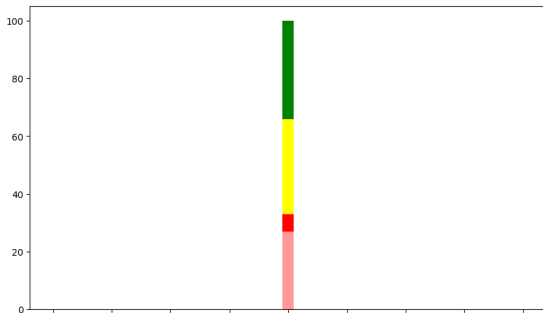
Improve: Non parametric

Someone says "the data not normally distributed, We have to use ranking"

season	p	play	p%	rank	rank%
2022	77	34	75%	4	27%
2021	80	34	78%	9	60%
2020	89	34	87%	13	86%
2019	80	34	78%	8	53%
2018	88	34	86%	12	80%
2017	84	34	82%	11	73%
2016	76	34	76%	6	40%
2015	87	38	76%	5	33%
2011	61	30	68%	1	7%
2010	65	30	72%	3	20%
2009	63	30	70%	2	13%
2008	70	30	78%	7	47%
2007	73	30	81%	10	67%
2006	72	26	92%	14	93%
2005	74	26	95%	15	100%

I'll split into 3 clusters

Good	(0 - 33%)
V.Good	(34% - 66%)
Amazing	(67% - 99%)

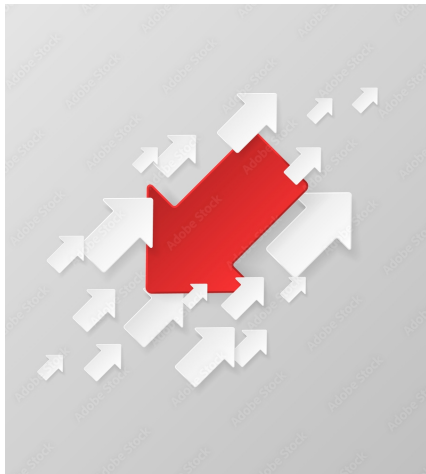


- It leads to the same conclusion

The power of business understanding

Without business understanding, you will go in the wrong direction

season	p
2022	77
2021	80
2020	89
2019	80
2018	88
2017	84
2016	76
2015	87
2011	61
2010	65
2009	63
2008	70
2007	73
2006	72
2005	74



Final thoughts

Answers evaluation

- Spend 1M for one question

Final thoughts

Answers evaluation

- Spend 1M for one question
- Each answer should be based on statistical method

Final thoughts

Answers evaluation

- Spend 1M for one question
- Each answer should be based on statistical method

Manger reaction



Final thoughts

Answers evaluation

- Spend 1M for one question
- Each answer should be based on statistical method

Manger reaction



- My club is bad?

Final thoughts

Answers evaluation

- Spend 1M for one question
- Each answer should be based on statistical method

Manger reaction



- My club is bad?
- Why It's bad?

Final thoughts

Answers evaluation

- Spend 1M for one question
- Each answer should be based on statistical method

Manger reaction



- My club is bad?
- Why It's bad?
- How could I fix It?