Data

We have attached two CSV files containing football data. The file games.csv contains the fixtures played in a range of European leagues over the last 5 seasons. The file events.csv contains the times of red cards and goals in these games.

Question

We would like you to tackle the following question: **Based on the data in this dataset, do red cards seem to lead to more goals?**

We mean more goals in general here - we might expect the team with the red card to get fewer goals than we would have otherwise expected, and their opponent to get more, but the question is whether these effects combined lead to more goals overall.

Deliverables

We would like you to submit:

- A Jupyter notebook containing any code you wrote for your analysis (please put your name in the filename)
- A write-up of your analysis that explains your approach and conclusions. It's up to you whether you include your write-up in your Jupyter notebook or separate it out. Your write-up should cover:
  - A high-level explanation of your approach
  - Why you picked the approach you did
  - Your conclusions, with any data or visualizations needed to support them

Notes

- You don't need to worry about cleaning the data – we would prefer you to spend your time thinking about the analytical approach needed to answer the question. The main limitation of the data you should be aware of is that the minutes values for the first and second half are capped at 45 and 90 respectively. Any event that occurred in first half injury time will be assigned the minute 45; any event that occurred in second half injury time will be assigned the minute 90.
- Please use Python if you can (though Jupyter supports many other languages too if you have no experience with Python).