Question 3

Amir Hossein Mansoori - 99243069

Import required modules:

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

Load the dataset:

In [3]:
```
laliga_matches = pd.read_csv('LaLiga_Matches_1995-2021.csv')
```

Store number of matches in which one of the teams were ahead by 2 goals, and number of those where the ahead team actually won the match:

In [4]:
```
total_matches_with_2_goals_ahead = 0
matches_won_with_2_goals_ahead = 0
```

Then check every match:

In [5]:
```
for index, row in laliga_matches.iterrows():
    home_first_half_goals = row['HTHG']
    away_first_half_goals = row['HTAG']
    result = row['FTR']

    if abs(home_first_half_goals - away_first_half_goals) != 2:
        continue

    total_matches_with_2_goals_ahead += 1

    if (result == 'H' and home_first_half_goals - away_first_half_goals
        matches_won_with_2_goals_ahead += 1
```

Finally, dividing these two numbers should yield the asked probability:

In [7]:
```
print(f"{(matches_won_with_2_goals_ahead / total_matches_with_2_goals_ah
```

```
91.07438016528926%
```

which is pretty high!