Homework 3

Question 1: Elo Ratings

Part a: Table containing top 3 teams for each division as ranked by Elo ratings at the end of the 2017 season.

Div	Team	Elo Rating
Bundesliga	Bayern Munich	1350.621424
Bundesliga	Schalke 04	1159.177933
Bundesliga	Hoffenheim	1142.152808
EPL	Man City	1429.659400
EPL	Tottenham	1283.911192
EPL	Man United	1258.732460
La_Liga	Barcelona	1415.462495
La_Liga	Real Madrid	1306.832652
La_Liga	Ath Madrid	1220.575107
Ligue_1	Paris SG	1352.520538
Ligue_1	Monaco	1264.722861
Ligue_1	Lyon	1240.383883
Serie_A	Juventus	1414.044716
Serie_A	Napoli	1337.681320
Serie_A	Roma	1282.244900

Part b:

In some instances, it may be a good idea to temporarily use a higher value of K if we would like to place more weighting on certain games. Some games played throughout the season are of more importance than others; for example, games played at the end of the season will be more important, especially for some teams who are on the cusp of relegation as they rely on winning at the end of the season to stay within the league. Although it is not the case for European football leagues, some American sports participate in playoffs. It could be beneficial to place higher weighting on playoff games or games played at the end of the season to qualify for the playoffs. These games are higher stake than other games during the regular season.

Part c:

A new feature for the difference in Elo ratings (between home and away teams) has been included in a model from Homework 2. The model was previously built on features for Historical Average Goal Differential, Historical Average Expected Goal Differential, Historical Average Shots on Target Differential, and Winning Streak. Prior to adding the Elo feature, the model had a Brier score of 0.215536. Upon adding the difference in Elo ratings feature, the Brier score improved to 0.214157. It is evident that the Elo difference feature provided some predictive power.

Question 2: Market-Implied Probabilities

Part a: Table containing the 7 greatest upsets before the 2018 season.

Div	Y	Home Team	Away Team	pH	pA	Home Goals	Away Goals
La_Liga	16	Barcelona	Alaves	0.891147	0.028831	1	2
La_Liga	14	Barcelona	Malaga	0.875453	0.040021	0	1
La_Liga	14	Barcelona	Celta	0.861781	0.043664	0	1
Bundesliga	15	Bayern Munich	Mainz	0.856920	0.044404	1	2
La_Liga	17	Real Madrid	Betis	0.876513	0.048646	0	1
La_Liga	15	Levante	Ath Madrid	0.052018	0.798875	2	1
Bundesliga	14	Bayern Munich	M'gladbach	0.821218	0.054292	0	2

Part b:
It does not appear that the market is less accurate at the start of a season. When teams have each payed less than 5 games in a season, the Brier score of the market is 0.210581. However, including all games, the Brier score of the market is 0.210606. Therefore, it appears that the market is not any less accurate at correctly predicting the game winner. This goes to show that using market implied probabilities is highly reliable and our models can benefit from including these features.

Part c:
The market-implied probabilities were incorporated into the same baseline model from Homework 2 which used features for Historical Average Goal Differential, Historical Average Expected Goal Differential, Historical Average Shots on Target Differential, and Winning Streak. The Brier score was 0.215536 without the addition of these market-implied probabilities.

A new feature was added using the historical average of the probability of a given team winning. This was calculated by taking the cumulative seasonal average for a given team. At the start of a season, the team's cumulative average is reset to 0.5 as they had not yet played any games that season and are assumed to have a fair chance of winning against their opponent. This was done because some teams did not play in the previous season and therefore have no historical data to rely on. These features were included as 'p_cumAvg_Home' and 'p_cumAvg_Away' indicating the cumulative average for the home and away teams for the given season, respectively. Based on the model summary, the coefficients were not only in the correct direction, but the coefficients were much larger than those assigned to the other features. Additionally, the low Z values indicated high predictive power. Including these new features improved the Brier score to 0.214648.