**Predicting Bruin’s Performance**

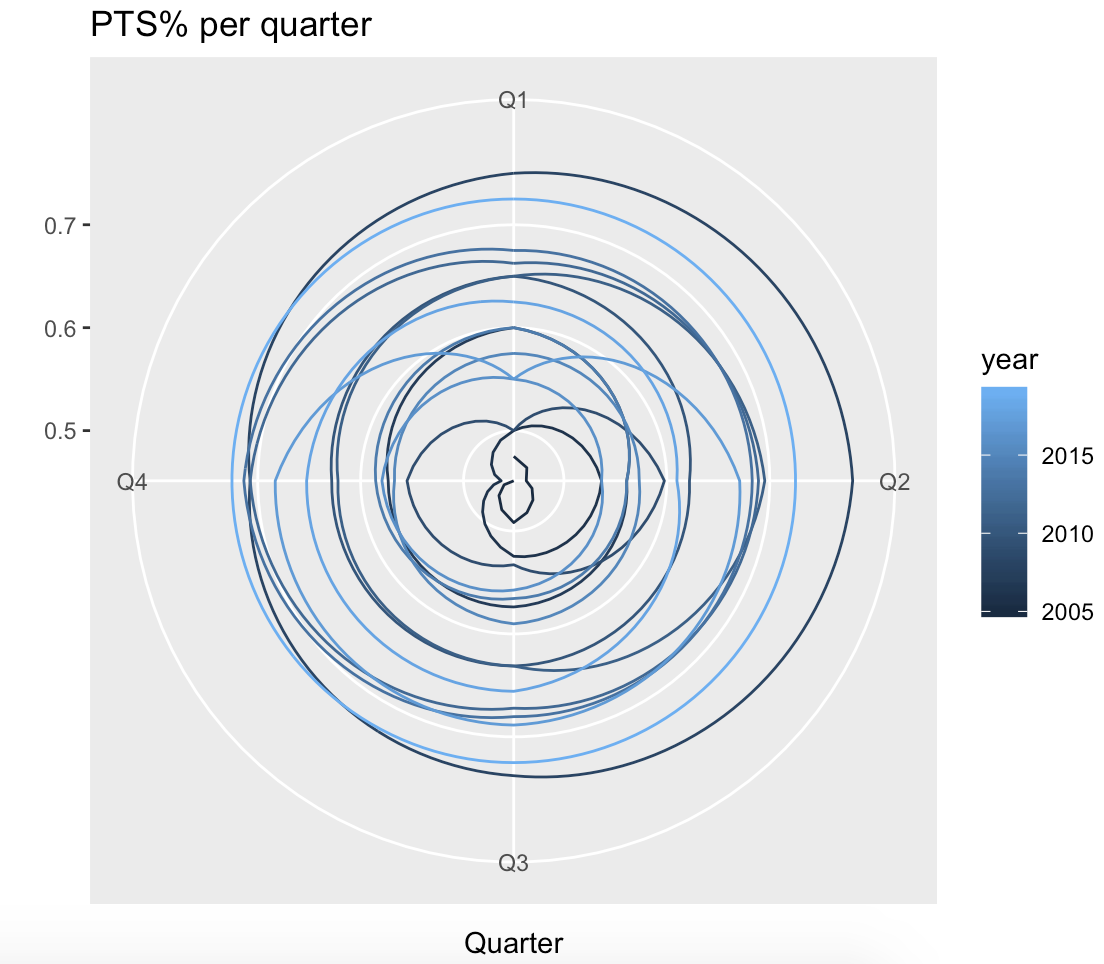
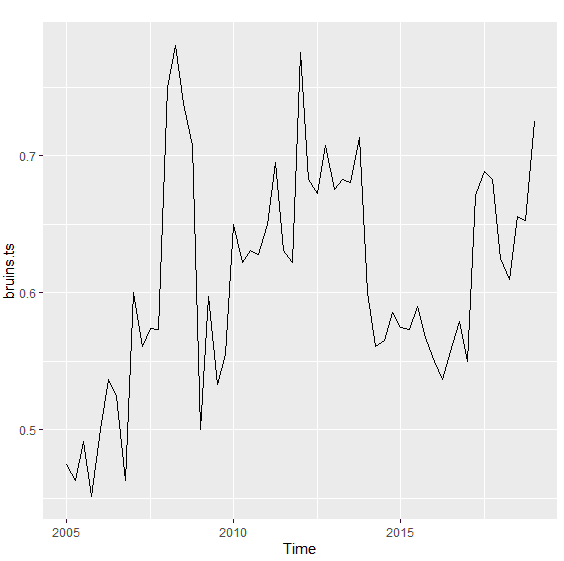
**Introduction**

In the NHL, a team with more points over the course of a season is ranked higher in the standings. Predicting the number of points a team accrues in a season is hard. With 14 years of quarterly point percentage (pts%) data for the Boston Bruins, it will be interesting to see how accurately we can predict the point percentage for the next quarter. In this project, the fundamental model of predicting a team’s percentage points over the next quarter season.

**Data Preparation & Exploratory Data Analysis**

At the preliminary stage of building a model, several steps are carried out to better understand the data. First, we visualized our data to observe potential trends (if any) through the time series plot. Then we have tried to assess underlying seasonalities, and special patterns using both polar map.

The time - series plot indicates that there is a lot of variation around the mean; however there seems to be an absence of any strong pattern or trend.

From the polar plot, we cannot conclude the team performed extremely well in the same number of games played in different years as different color would overlap in the same year. We also cannot detect a clear trend in which the Bruins are playing better or worse over the course of time. Overall, polar plot did not provde meaningful information about the data.   
 

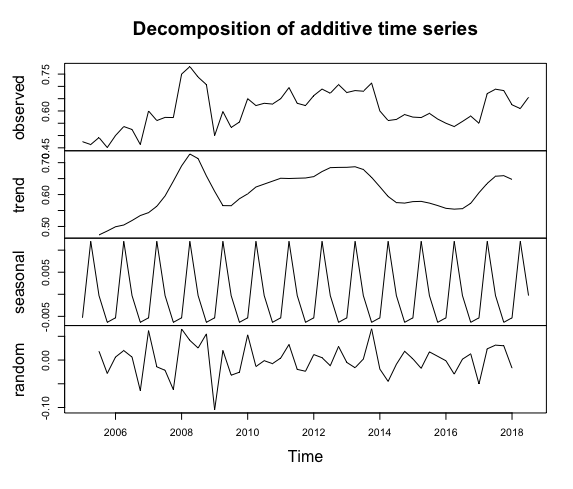
At the beginning of the 2012-2013 season, there was a dispute over the collective bargaining agreement between the league and the player’s association. Typically, the two parties work over the summer recess and come to an agreement on the rules of the game on and off the ice. However, during this particular year, no agreement could be made. Without an agreement, no games could be played. As the standoff pushed into the beginning of the season, games were cancelled. A total of 38 games were cancelled before the agreement was signed. There were still time to schedule 48 games for the season, so they still had enough of a sample size to salvage the season. In a time series dataset, the order of observations are crucial as correlation plays an important role in analyzing time series.

**Theory and Method**

* Naive Model:

A naive model is first created as a benchmark for comparison purposes. At the end, all models will be compared to the naive model.

* Decomposition Model

To continue from the polar plot a decomposition is done. From the plot, there is no clear upward or downward trend. From the seasonal graph, we do see a yearly seasonality. The accuracy measures are saved and compared afterwards. (Decomposition)

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MSE | MAE | MAPE |
| Naive | 0.1013 | 0.06998 | 119.03 |
| MA Decomposition | 0.05681 | 0.037603 | 6.22944 |

**Conclusion**

After evaluating 2 different models (Naive vs. MA Deomposition) using different testing statistics such as small MSE/ MAPE/MAE value, the MA Decomposition model is considered the best model to predict Bruin’s point percentage for the next game quarter. This is precisely because of the significantly lower MAPE and lower MSE/MAE.

**Future Work**

Although, the MA decomposition model seems to be a better model in predicting the win percentage for Bruins in the next season; there is always an opportunity to use advanced forecasting models like ETS Models or ARIMA Models. Also, there was no regressor taken into consideration when making any prediction models. In real life, numerous factors can dominate the outcome of a game. One of the most important variable’s is the opponent’s team. Each game, Bruin faces a different team. If an overall score for the opponent can be evaluated and used as a regressor, the prediction can be more accurate. The overall score can combine statistics from a variety of places such as win rate from previous games, combination of each player’s overall score, age of each player and qualitative variables such as whether a player has injured, whether a coach has changed, whether they are playing in their home field, etc. It is easier said than done, this process requires extremely detailed data gathering and processing. Since this analysis is only the tip of an iceberg, much more can be explored and added to make a comprehensive and accurate prediction model.