PREMIER LEAGUE DATA ANALYSIS & PREDICTION

CAPSTONE PROJECT

Prepared by: Vivek Kashyap
Mentor: Varnith Chordia

CONTEXT AND BACKGROUND

Soccer(a.k.a football) is a very popular sport and is closely followed by a large number of people. There numerous events be it FIFA World cup or club events like EPL (English premier league) where there is huge interest in getting to know outcome be it goal, MVP,corners, cards etc. Most of the soccer fans would still remember, *Paul the Octopus*, whose accurate predictions during 2010 world cup got him worldwide attention

CONTENT

OVERALL PROCESS

Project was divided into four broad phases which are described below

Data extraction

- Used combination of Python site libraries and RPA process
- Data was extracted from English premier league site

Data Cleansing

- Data extracted from site was mostly clean.
- Few missing or not available values was defaulted

Data Analysis

- Data Analysis performed using Pandas data frame
- Visualization was performed using Matplotlib and Seaborn libraries

Machine Learning

- Poisson model was used to get predicted score and probability of Win , Draw or loss for a team
- Other classification like KNeighborsClassifier, RandomForestClassifier and SVM method was used

EXPLORATORY DATA ANALYSIS

DATA ANALYSIS FOR LAST 8 SEASONS OF PREMIER LEAGUE

RANKING OF DIFFERENT TEAMS ACROSS SEASONS

	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
0	Manchester City	Manchester United	Manchester City	Chelsea	Leicester City	Chelsea	Manchester City	Manchester City
1	Manchester United	Manchester City	Liverpool	Manchester City	Arsenal	Tottenham Hotspur	Manchester United	Liverpool
2	Arsenal	Chelsea	Chelsea	Arsenal	Tottenham Hotspur	Manchester City	Tottenham Hotspur	Chelsea
3	Tottenham Hotspur	Arsenal	Arsenal	Manchester United	Manchester City	Liverpool	Liverpool	Tottenham Hotspur
4	Newcastle United	Tottenham Hotspur	Everton	Tottenham Hotspur	Manchester United	Arsenal	Chelsea	Arsenal
5	Chelsea	Everton	Tottenham Hotspur	Liverpool	Southampton	Manchester United	Arsenal	Manchester United
6	Everton	Liverpool	Manchester United	Southampton	West Ham United	Everton	Burnley	Wolverhampton Wanderers
7	Liverpool	West Bromwich Albion	Southampton	Swansea City	Liverpool	Southampton	Everton	Everton
8	Fulham	Swansea City	Stoke City	Stoke City	Stoke City	AFC Bournemouth	Leicester City	Leicester City
9	West Bromwich Albion	West Ham United	Newcastle United	Crystal Palace	Chelsea	West Bromwich Albion	Newcastle United	West Ham United
10	Swansea City	Norwich City	Crystal Palace	Everton	Everton	West Ham United	Crystal Palace	Watford
11	Norwich City	Fulham	Swansea City	West Ham United	Swansea City	Stoke City	AFC Bournemouth	Crystal Palace
12	Sunderland	Stoke City	West Ham United	West Bromwich Albion	Watford	Leicester City	West Ham United	Newcastle United
13	Stoke City	Southampton	Sunderland	Leicester City	West Bromwich Albion	Crystal Palace	Watford	AFC Bournemouth
14	Wigan Athletic	Aston Villa	Aston Villa	Newcastle United	Crystal Palace	Swansea City	Brighton and Hove Albion	Burnley
15	Aston Villa	Newcastle United	Hull City	Sunderland	AFC Bournemouth	Burnley	Huddersfield Town	Southampton
16	Queens Park Rangers	Sunderland	West Bromwich Albion	Aston Villa	Sunderland	Watford	Southampton	Brighton and Hove Albion
17	Bolton Wanderers	Wigan Athletic	Norwich City	Hull City	Newcastle United	Hull City	Swansea City	Cardiff City
18	Blackburn Rovers	Reading	Fulham	Burnley	Norwich City	Middlesbrough	Stoke City	Fulham
19	Wolverhampton Wanderers	Queens Park Rangers	Cardiff City	Queens Park Rangers	Aston Villa	Sunderland	West Bromwich Albion	Huddersfield Town

- Teams like Manchester City, Chelsea, Arsenal, Liverpool, Tottenham Hotspur consistently features in Top 5 of League table from 2011/12 seasons onwards
- Swansea City, Stoke City, West Bromwich Albion featured in middle of League table from 2011/12 to 2017/18 and got relegated out of the League for 2018/19 season
- 2015/16 was an exception with Leicester City topping the table
- Everton has been mostly placed in middle of League table across all season

SEASON STATISTICS : OUTCOME, GOAL SCORED AND RANKING

Season: 2011/12

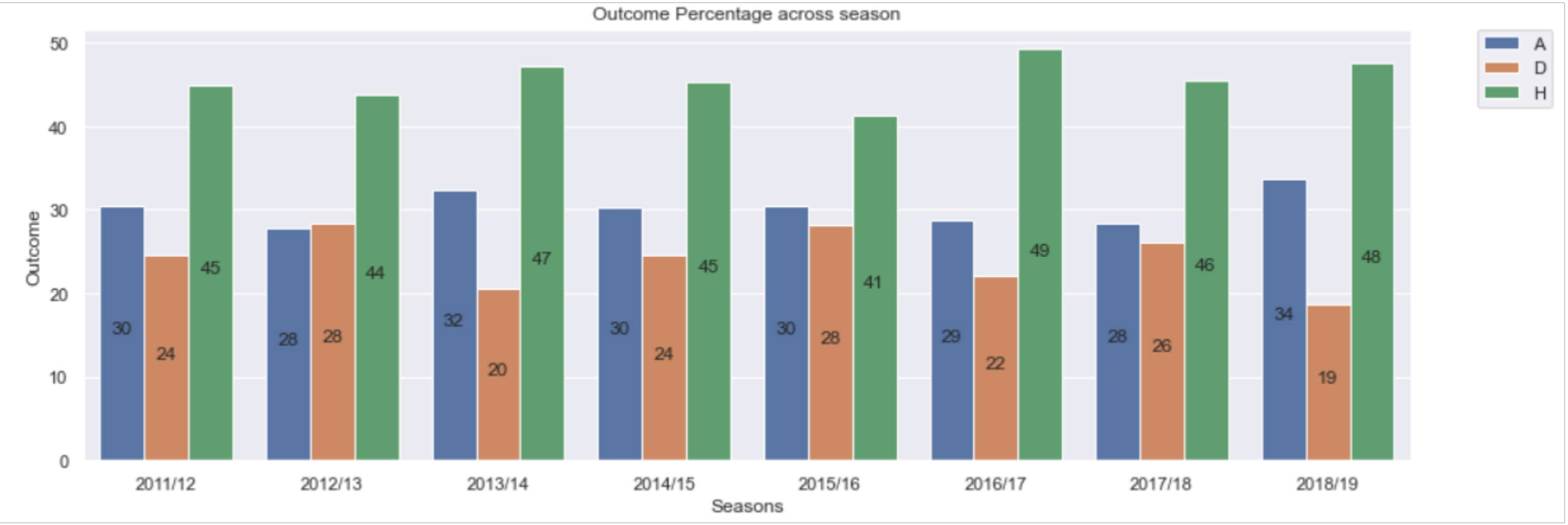
team	seasonlabel	total_game	win	win_h	win_a	draw	draw_h	draw_a	loss	loss_h	loss_a	GS	GA	GD	Points	SeasonsRank
Manchester City	2011/12	38	28	18	10	5	1	4	5	0	10	93	29	64	89	1
Manchester United	2011/12	38	28	15	13	5	2	3	5	2	13	89	33	56	89	2
Arsenal	2011/12	38	21	12	9	7	4	3	10	3	9	74	49	25	70	3
Tottenham Hotspur	2011/12	38	20	13	7	9	3	6	9	3	7	66	41	25	69	4
Newcastle United	2011/12	38	19	11	8	8	5	3	11	3	8	56	51	5	65	5
Chelsea	2011/12	38	18	12	6	10	3	7	10	4	6	65	46	19	64	6
Everton	2011/12	38	15	10	5	11	3	8	12	6	5	50	40	10	56	7
Liverpool	2011/12	38	14	6	8	10	9	1	14	4	8	47	40	7	52	8
Fulham	2011/12	38	14	10	4	10	5	5	14	4	4	48	51	-3	52	9
West Bromwich Albion	2011/12	38	13	6	7	8	3	5	17	10	7	45	52	-7	47	10
Swansea City	2011/12	38	12	8	4	11	7	4	15	4	4	44	51	-7	47	11
Norwich City	2011/12	38	12	7	5	11	6	5	15	6	5	52	66	-14	47	12
Sunderland	2011/12	38	11	7	4	12	7	5	15	5	4	45	46	-1	45	13
Stoke City	2011/12	38	11	7	4	12	8	4	15	4	4	36	53	-17	45	14
Wigan Athletic	2011/12	38	11	5	6	10	7	3	17	7	6	42	62	-20	43	15
Aston Villa	2011/12	38	7	4	3	17	7	10	14	8	3	37	53	-16	38	16
Queens Park Rangers	2011/12	38	10	7	3	7	5	2	21	7	3	43	66	-23	37	17
Bolton Wanderers	2011/12	38	10	4	6	6	4	2	22	11	6	46	77	-31	36	18
Blackburn Rovers	2011/12	38	8	6	2	7	1	6	23	12	2	48	78	-30	31	19
Wolverhampton Wanderers	2011/12	38	5	3	2	10	3	7	23	13	2	40	82	-42	25	20

Season: 2018/19

team	seasonlabel	total_game	win	win_h	win_a	draw	draw_h	draw_a	loss	loss_h	loss_a	GS	GA	GD	Points	SeasonsRank
Manchester City	2018/19	38	32	18	14	2	0	2	4	1	14	95	23	72	98	1
Liverpool	2018/19	38	30	17	13	7	2	5	1	0	13	89	22	67	97	2
Chelsea	2018/19	38	21	12	9	9	6	3	8	1	9	63	39	24	72	3
Tottenham Hotspur	2018/19	38	23	12	11	2	2	0	13	5	11	67	39	28	71	4
Arsenal	2018/19	38	21	14	7	7	3	4	10	2	7	73	51	22	70	5
Manchester United	2018/19	38	19	10	9	9	6	3	10	3	9	65	54	11	66	6
Wolverhampton Wanderers	2018/19	38	16	10	6	9	4	5	13	5	6	47	46	1	57	7
Everton	2018/19	38	15	10	5	9	4	5	14	5	5	54	46	8	54	8
Leicester City	2018/19	38	15	8	7	7	3	4	16	8	7	51	48	3	52	9
West Ham United	2018/19	38	15	9	6	7	4	3	16	6	6	52	55	-3	52	10
Watford	2018/19	38	14	8	6	8	3	5	16	8	6	52	59	-7	50	11
Crystal Palace	2018/19	38	14	5	9	7	5	2	17	9	9	51	53	-2	49	12
Newcastle United	2018/19	38	12	8	4	9	1	8	17	10	4	42	48	-6	45	13
AFC Bournemouth	2018/19	38	13	8	5	6	5	1	19	6	5	56	70	-14	45	14
Burnley	2018/19	38	11	7	4	7	2	5	20	10	4	45	68	-23	40	15
Southampton	2018/19	38	9	5	4	12	8	4	17	6	4	45	65	-20	39	16
Brighton and Hove Albion	2018/19	38	9	6	3	9	5	4	20	8	3	35	60	-25	36	17
Cardiff City	2018/19	38	10	6	4	4	2	2	24	11	4	34	69	-35	34	18
Fulham	2018/19	38	7	6	1	5	3	2	26	10	1	34	81	-47	26	19
Huddersfield Town	2018/19	38	3	2	1	7	3	4	28	14	1	22	76	-54	16	20

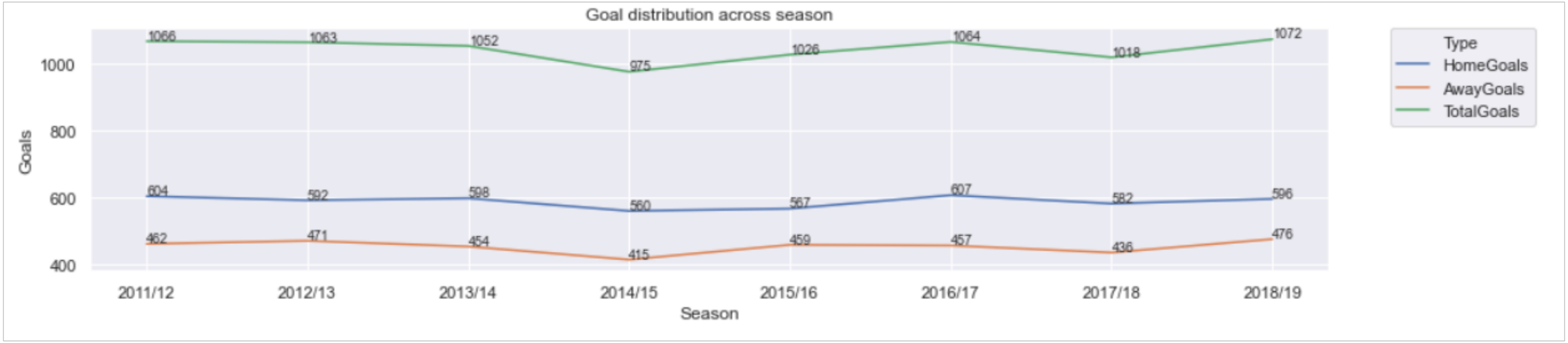
- Data was used to build ranking table for season
- For Illustration only first and last season is shown here but data can be used to build the similar table for rest of the season

MATCH OUTCOME (HOME AND AWAY)



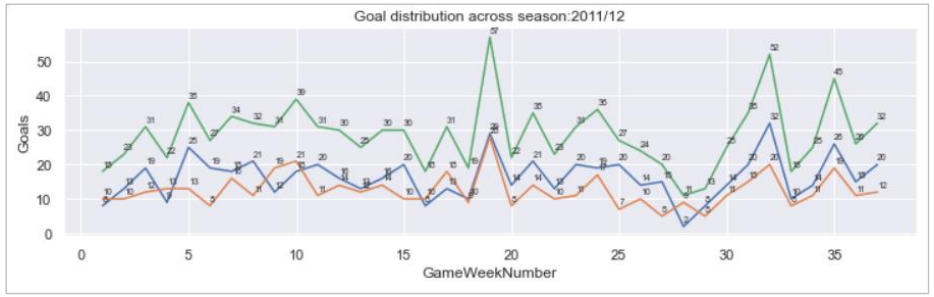
- Percentage of Match which resulted in Home Win >41% across season
- Percentage of Match which resulted in Away Team Win has been between 28%-34%
- Percentage of Match which resulted in Draw has been between 19%-28%

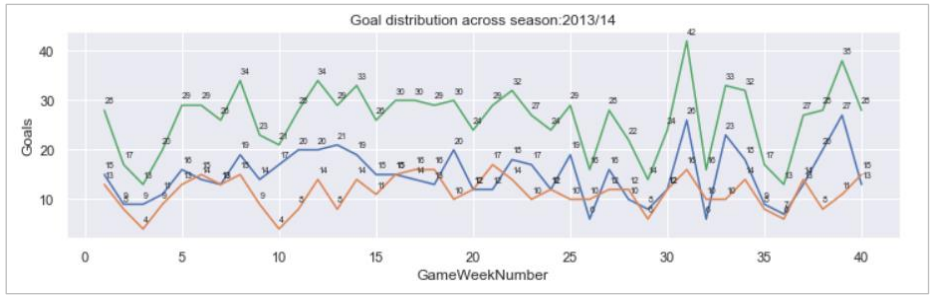
GOAL DISTRIBUTION (HOME AND AWAY)

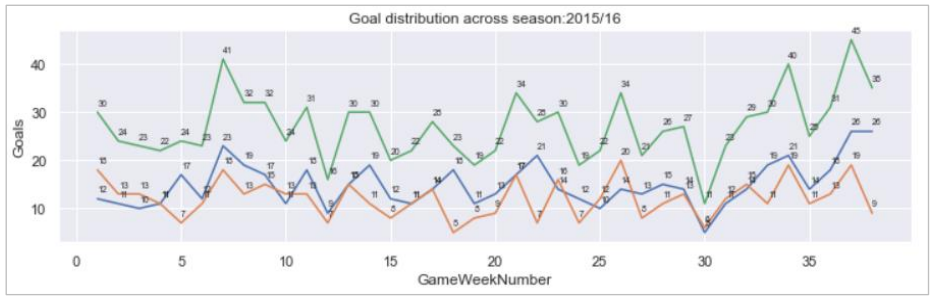


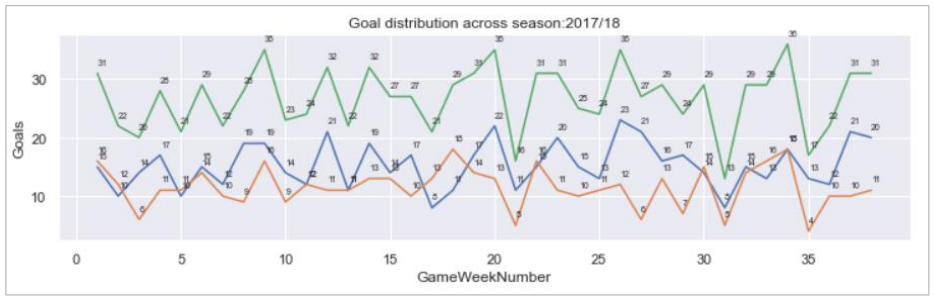
- Number of Goals scored in each season from 2011/12 to 2018/19 is in range of 1050 +/- 30 goals except in 2014/15
- Number of Goals scored by Home and away team also remains flat except 2014/2015
- Home team goals are always higher than away team goals across season

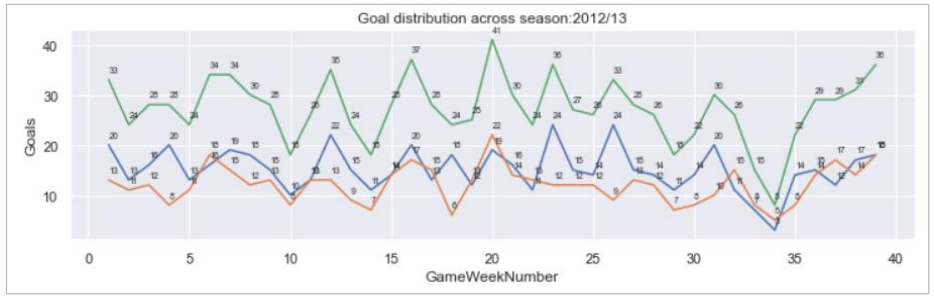
GOAL DISTRIBUTION (HOME AND AWAY)

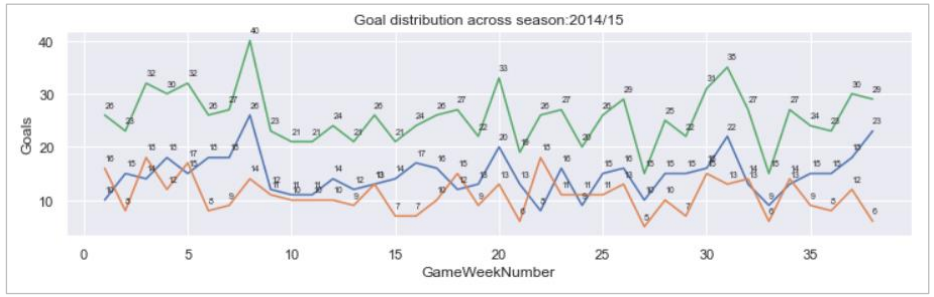


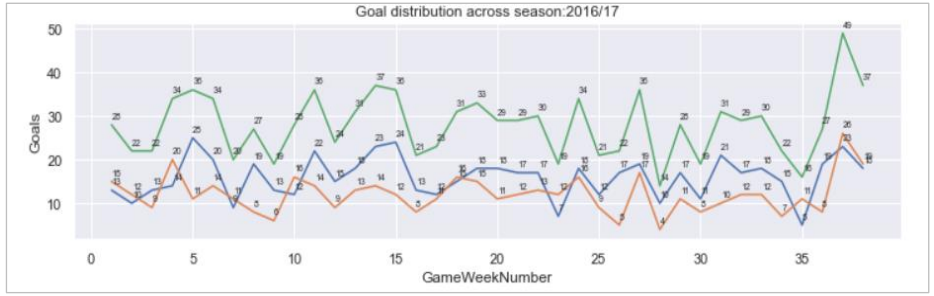


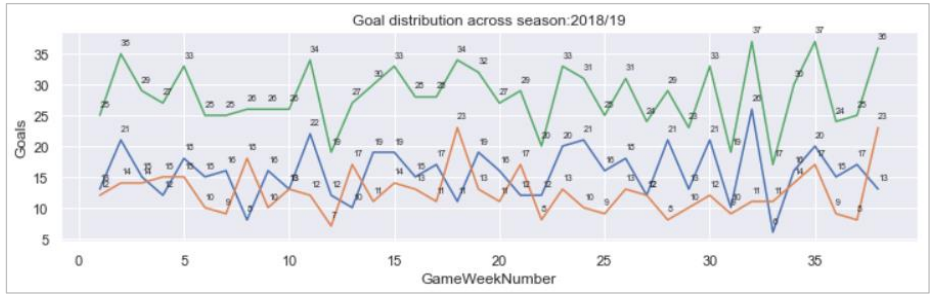






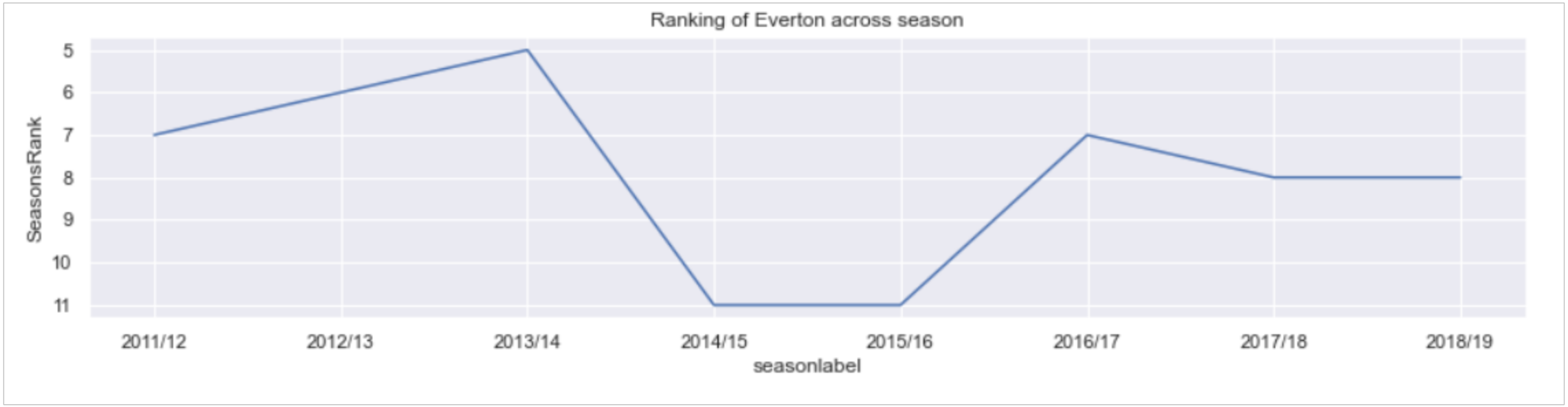






- Home goals are higher than away goals for most of the week
- Spike in total numbers of goal scored in matches are seen between Week 15-20 and Week 30-35

ANALYSIS OF TEAM (HOME ,AWAY ,OVERALL) ACROSS SEASON



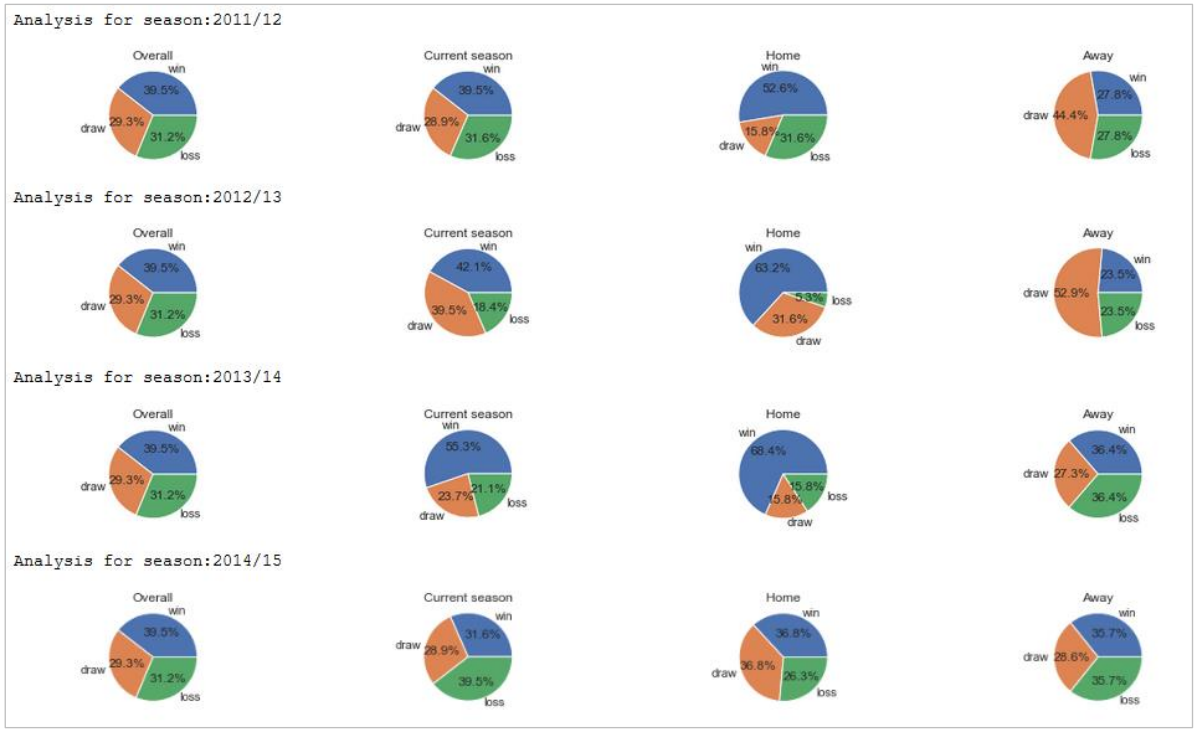
Picked ***Everton*** for Analysis

- Played almost all the season from 2011/12 – 2018/19
- Mostly placed in middle of League table across all season
- Form has been up and down with best ranking of 5 in 2013/14 and lowest ranking of 2014/15 and 2015/16.

We will look at some of statistics of what worked and what didn't for them

ANALYSIS OF TEAM (HOME ,AWAY ,OVERALL) ACROSS SEASON

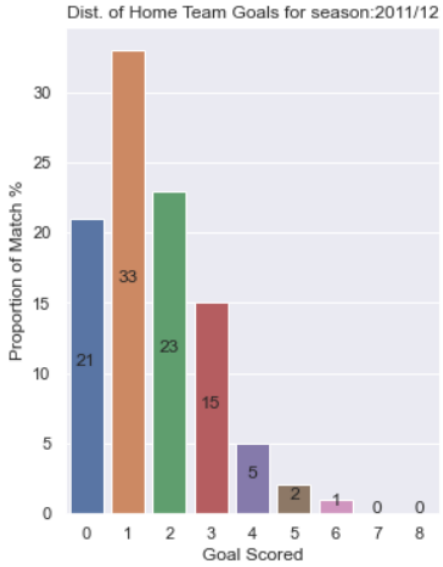
Analysis for Everton (Home & Away)

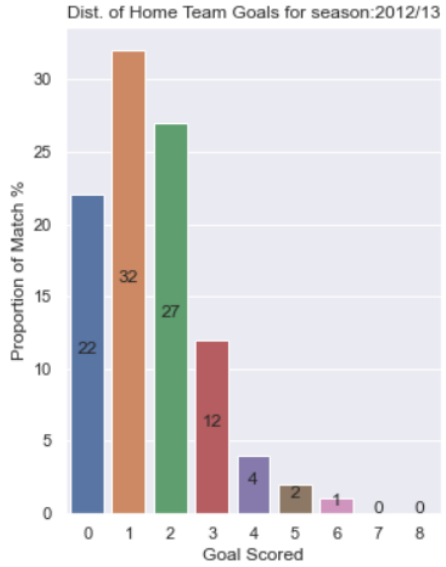


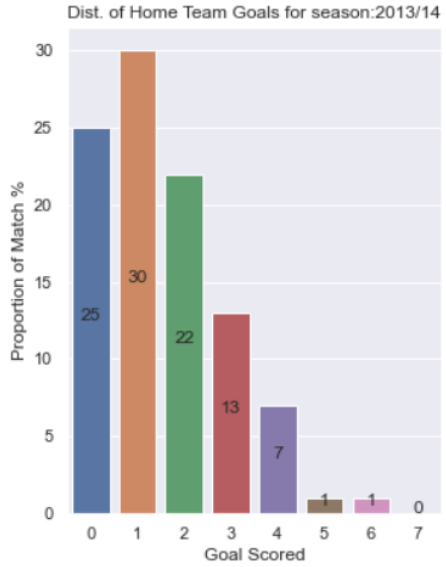
- Season 2014/15 : Home performance dipped compared to previous season - win dropped from ~68% to ~37%, draws increased from ~15% to ~36%.
- Season 2015/16: Both home and away performance dipped compared to 2013/14 season
- Season 2016/17 : With home performance improving , Everton moved back up in League standing

GOAL DISTRIBUTION ACROSS SEASON (HOME AND AWAY)

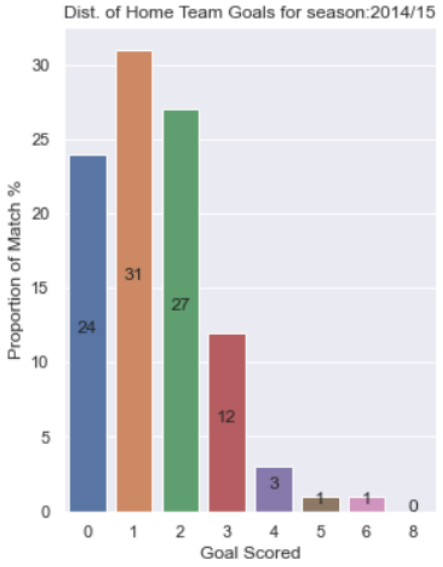
Goal distribution for Home Team

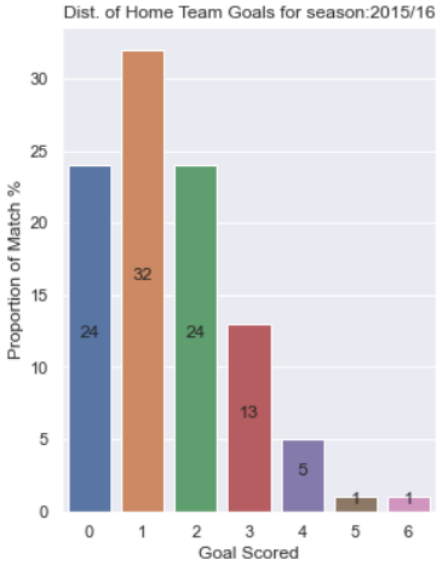


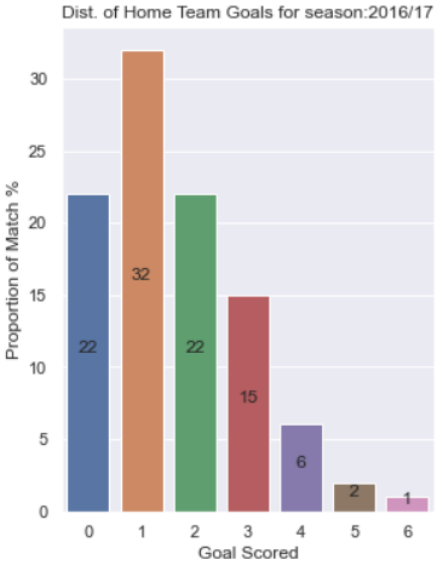


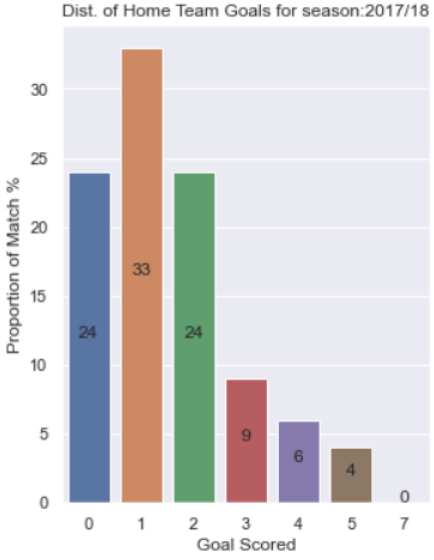


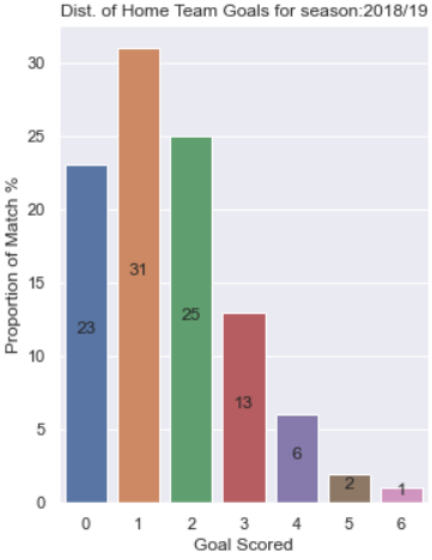
- 30%-32% of Matches Home team scores 1 Goal
- 22%-25% of Matches Home team scores 2 Goal
- Around 20% of Matches Home team scores more than 2 Goal





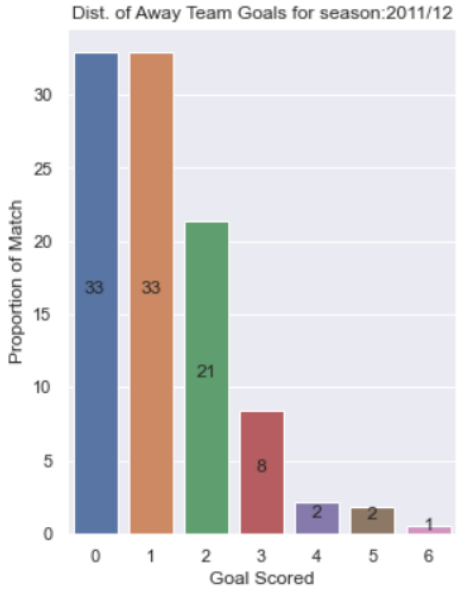


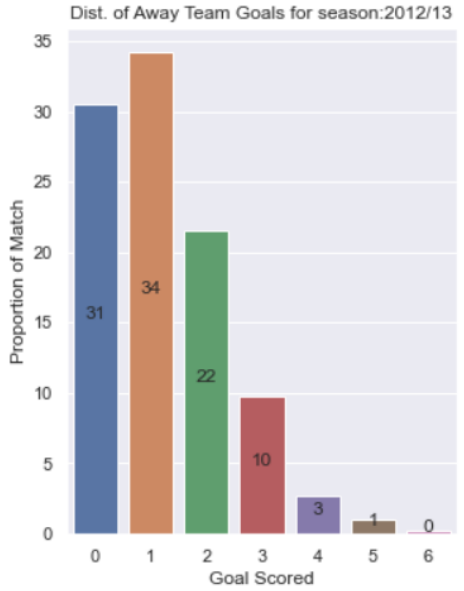


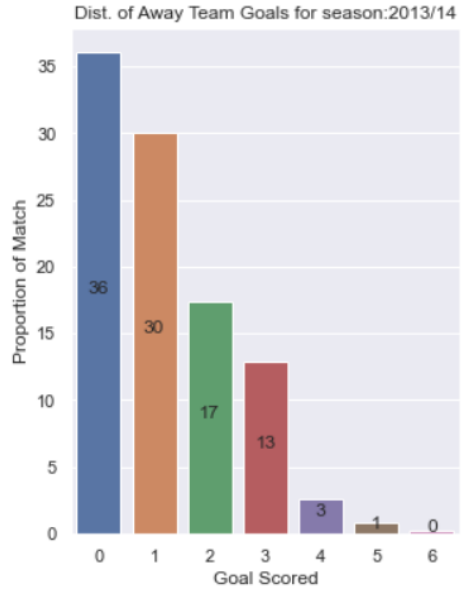


GOAL DISTRIBUTION ACROSS SEASON (HOME AND AWAY)

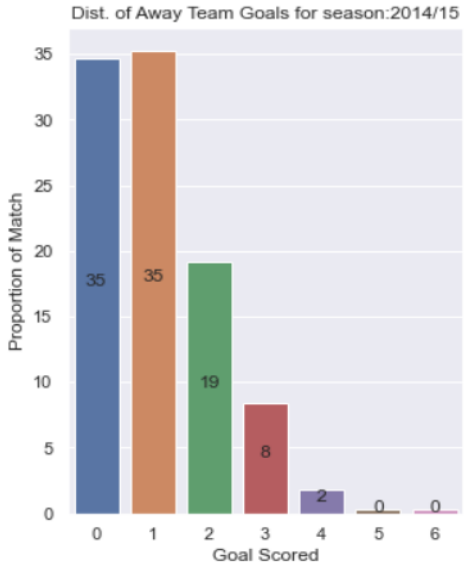
Goal distribution for Away Team

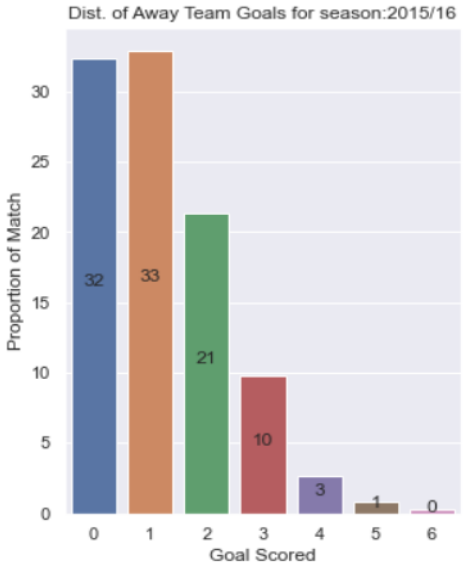


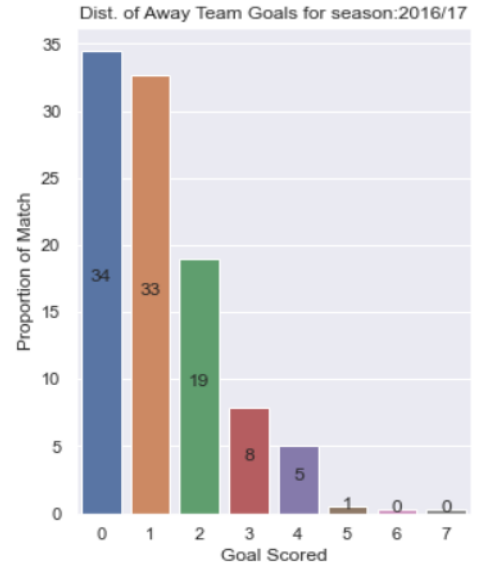


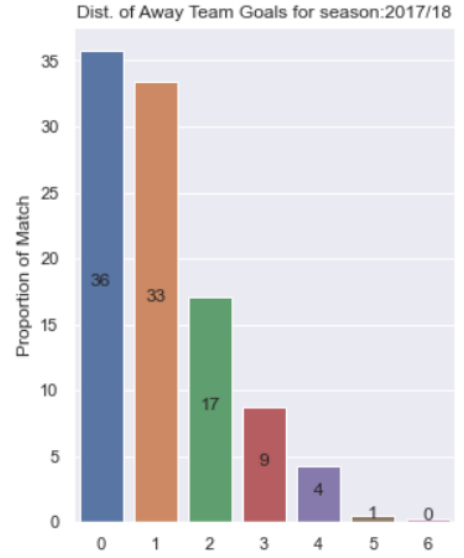


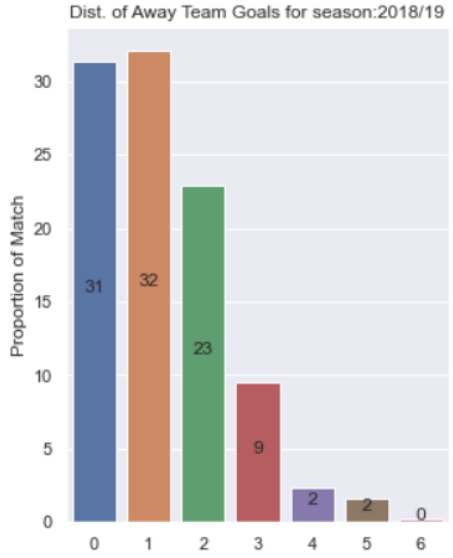
- 32%-34% of Matches Away team scores 1 Goal
- 17%-21% of Matches Away team scores 2 Goal
- Around 15% of Matches Home team scores more than 2 Goal



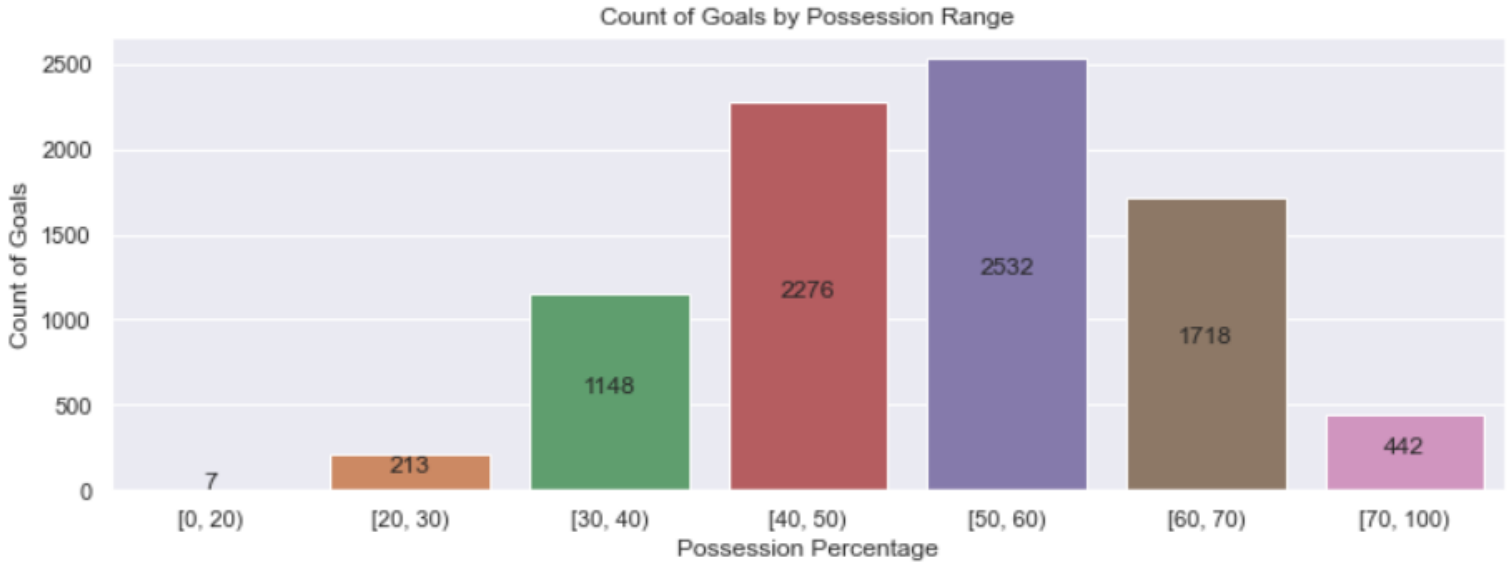


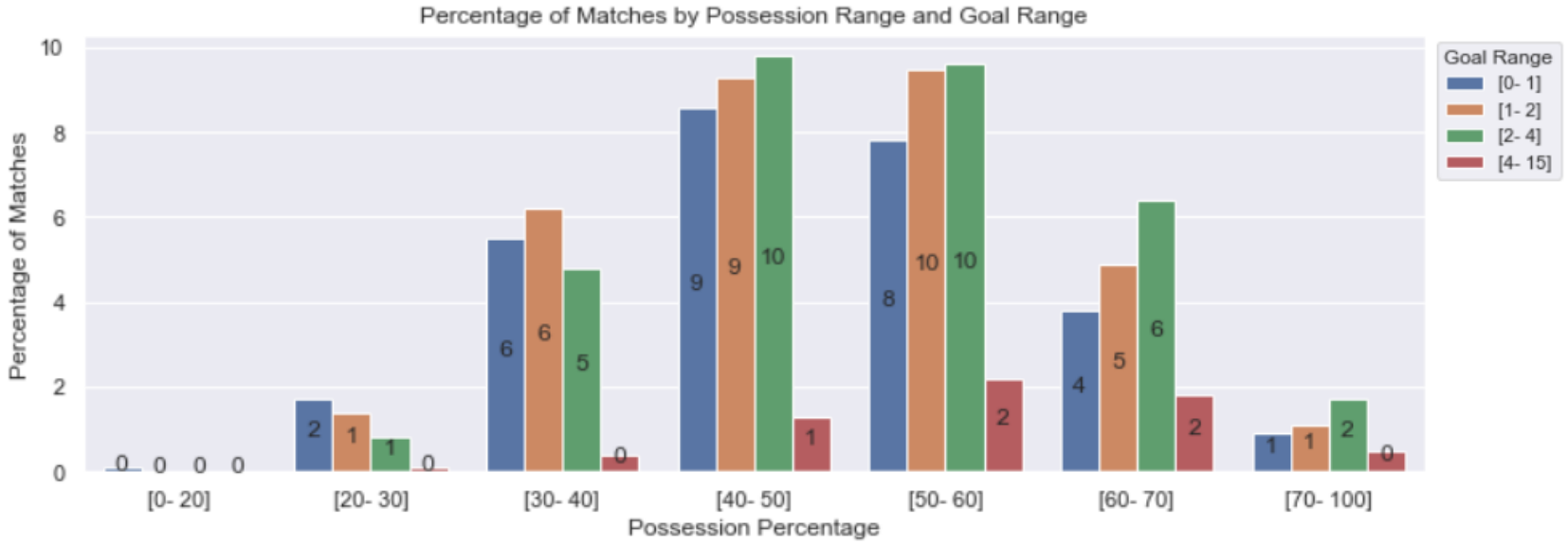






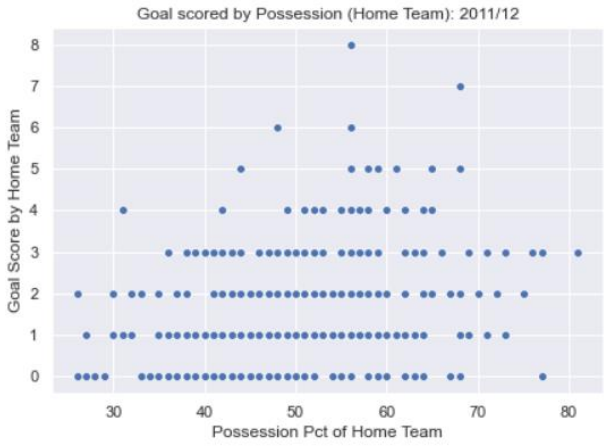
ANALYSIS ON GOAL AND POSSESSION

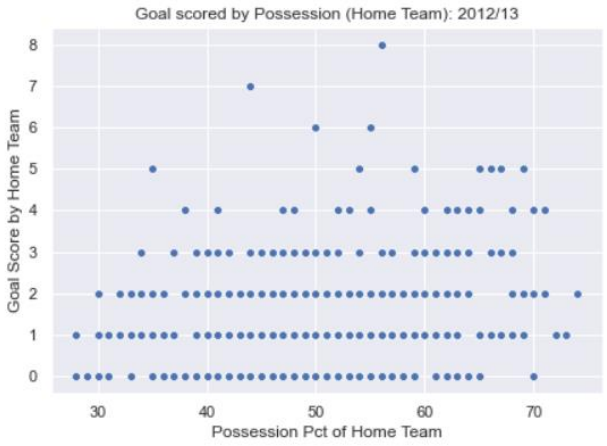


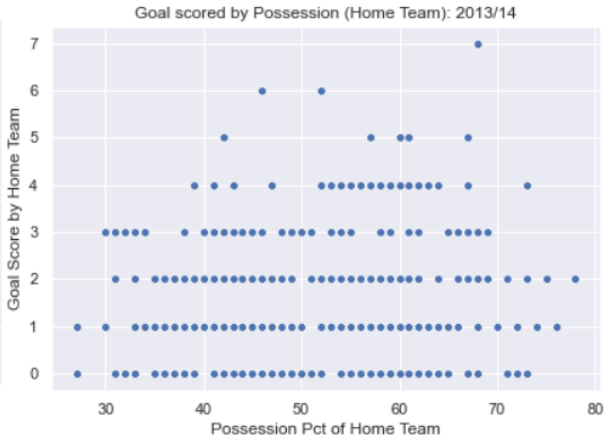


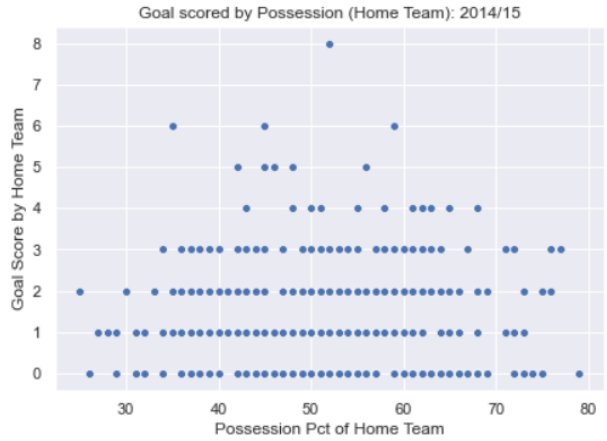
- Maximum goals were scored when the possession was between 50-60%
- As possession increases, number of goal scored also increases
  - 30% matches where team had possession of 50%-60% goal scored were 2532
  - 17% of matches where team had possession of 60%-70% goal scored were 1718
  - 4% of matches where team had possession of >70% and it resulted in 442 goals

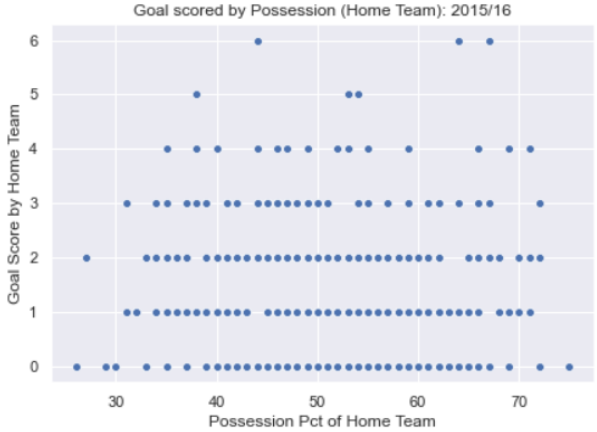
ANALYSIS OF POSSESSION VS GOAL : HOME TEAM

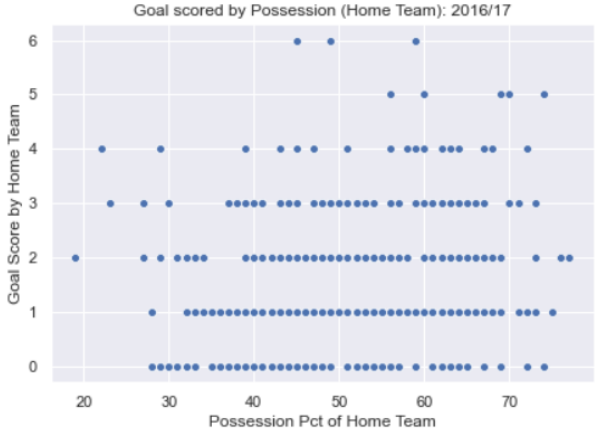


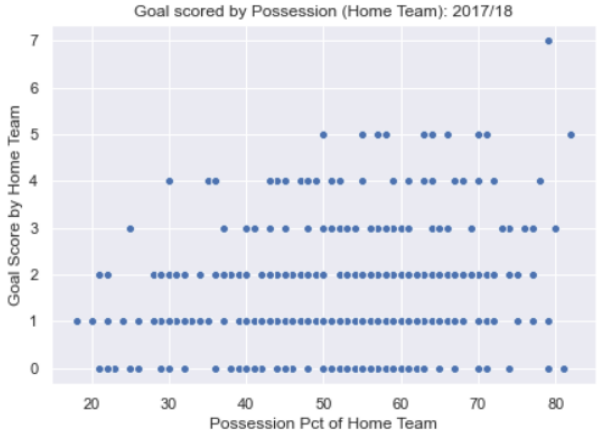


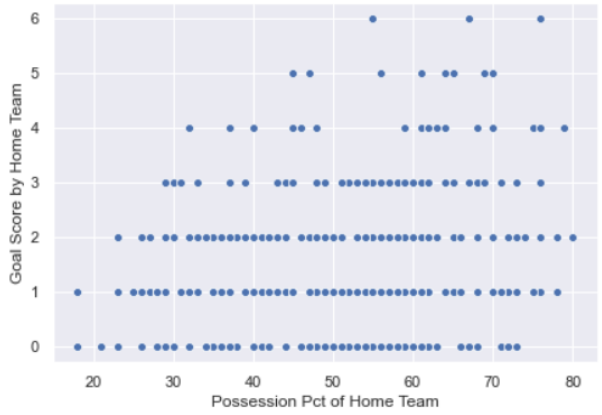






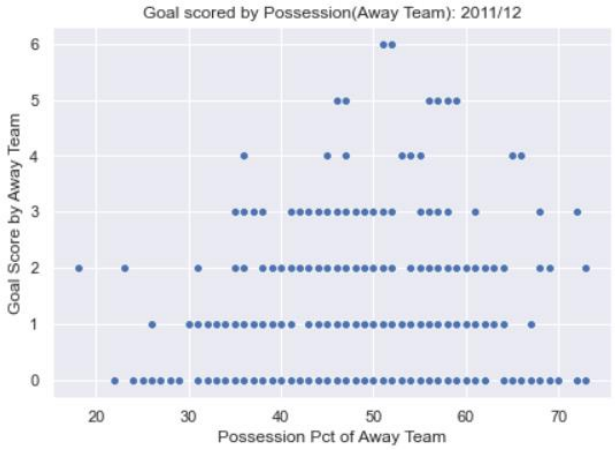


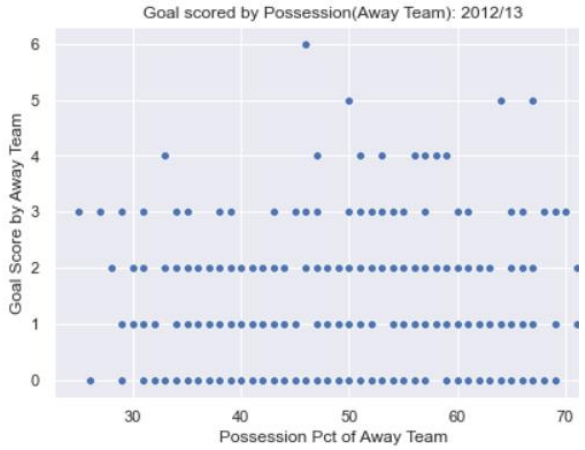


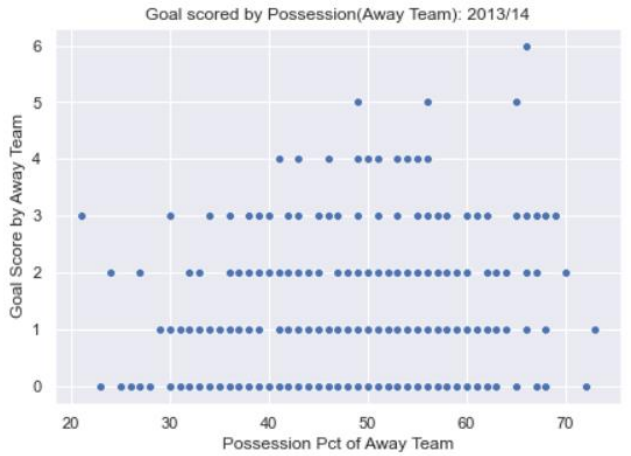


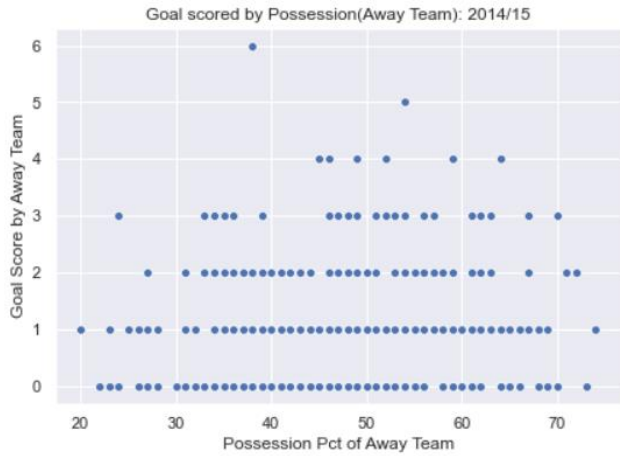
- Distribution of Possession and Goal scored is almost similar across the season from 2011/12 to 2018/19

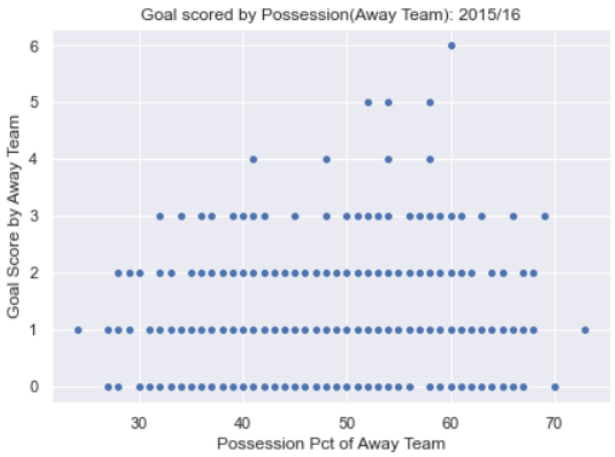
ANALYSIS OF POSSESSION VS GOAL : AWAY TEAM

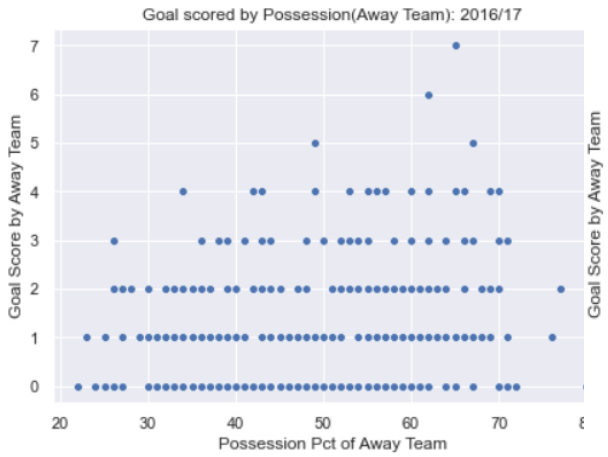


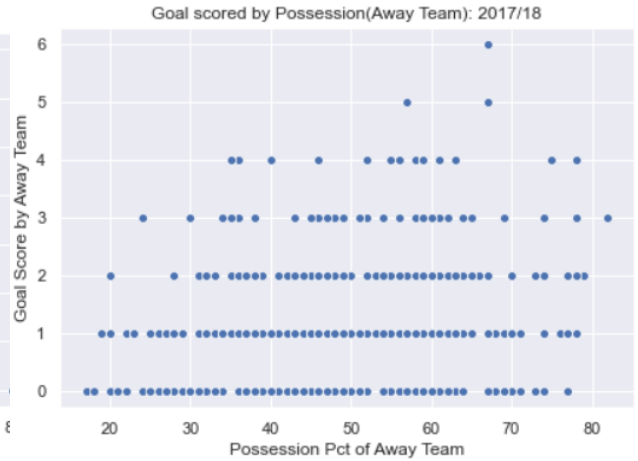


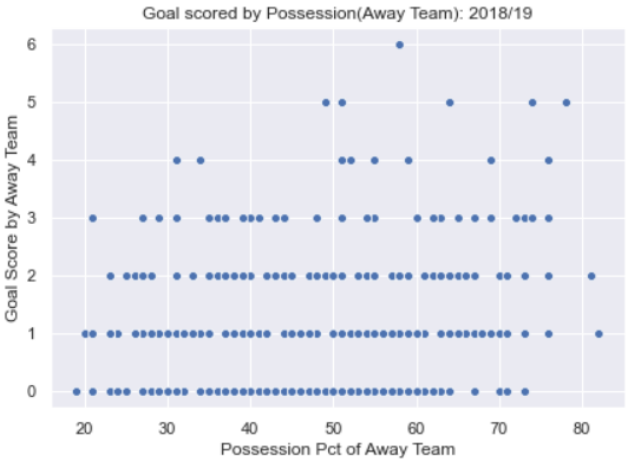










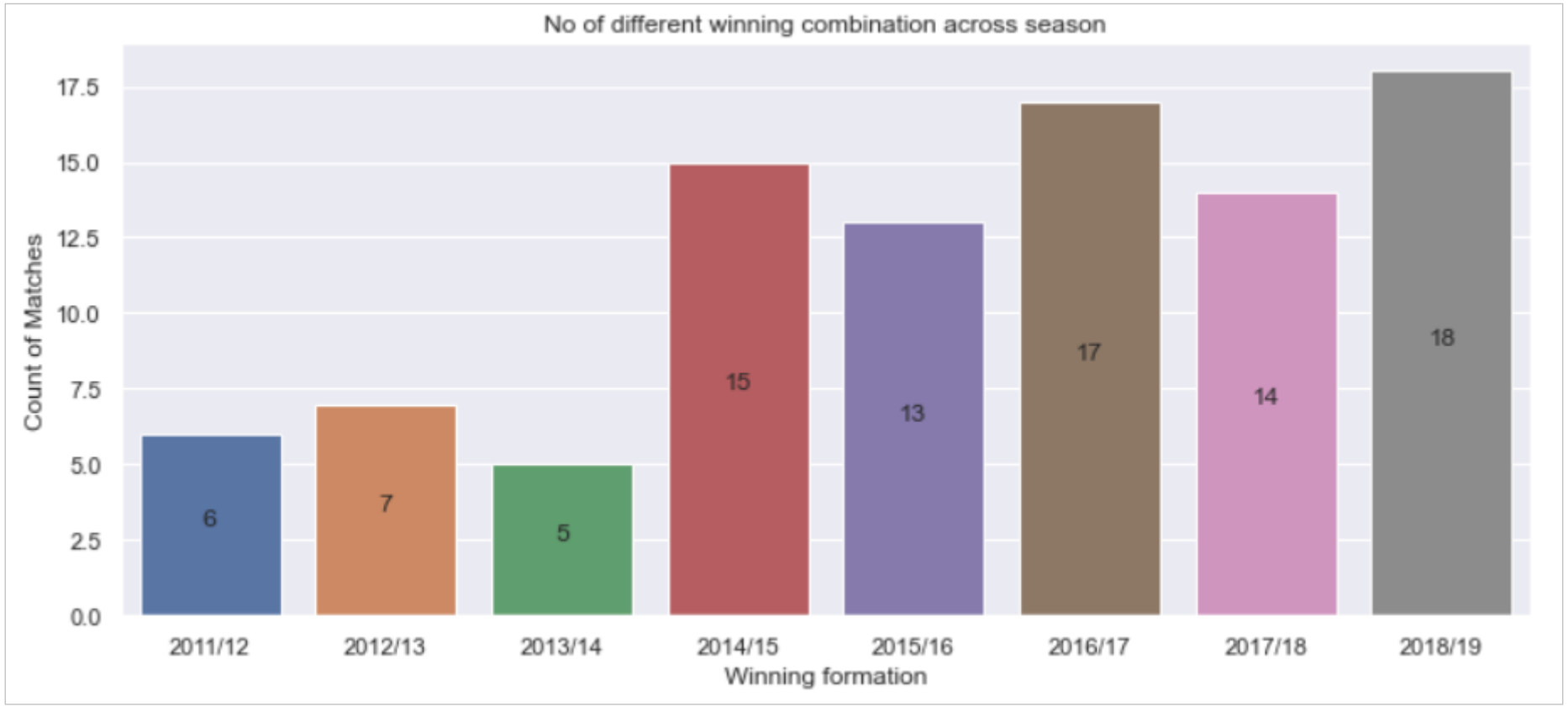


- Distribution of Possession and Goal scored is almost similar across the season from 2011/12 to 2018/19

ANALYSIS OF GOAL VS DISTANCE COVERED

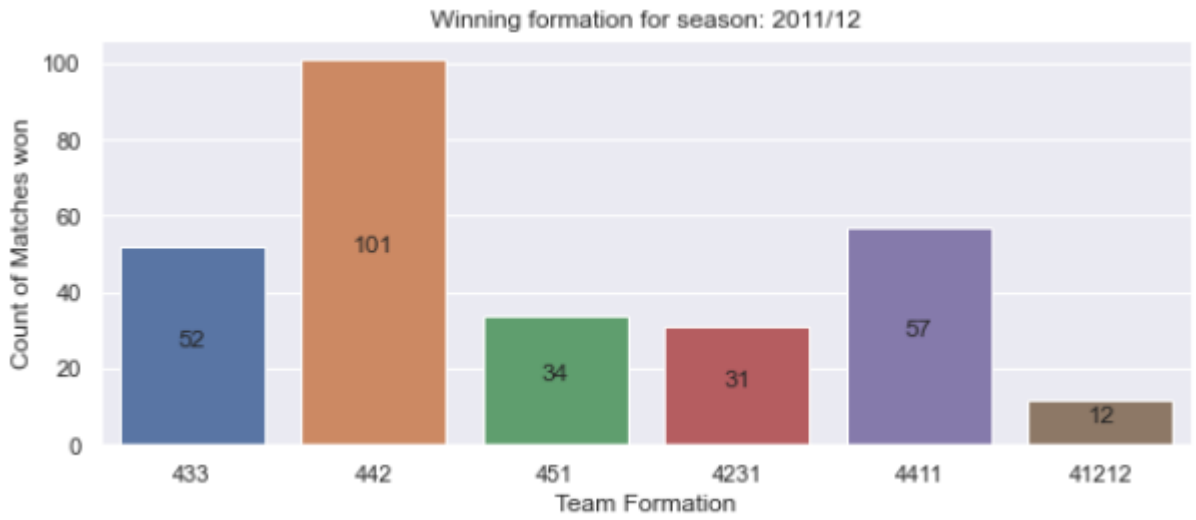
- Distance covered by players across last three season has been mostly between 210km – 230km

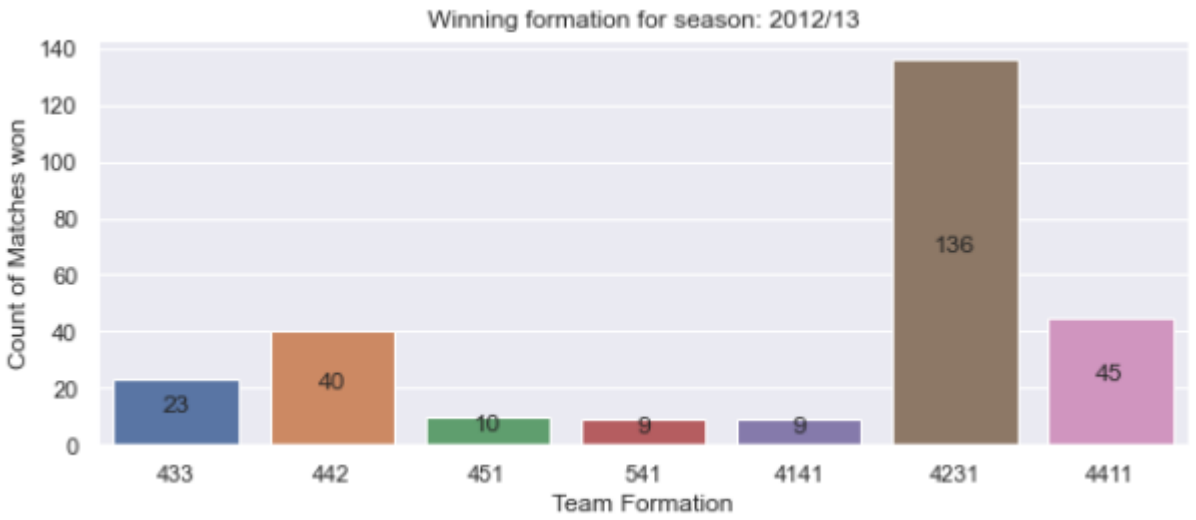
WINNING FORMATION ACROSS SEASON

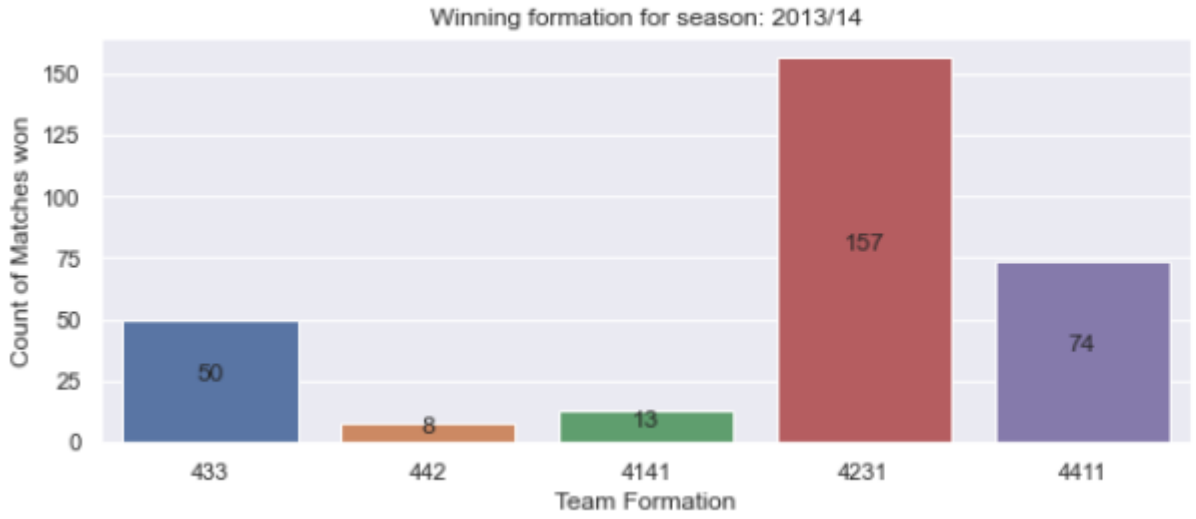


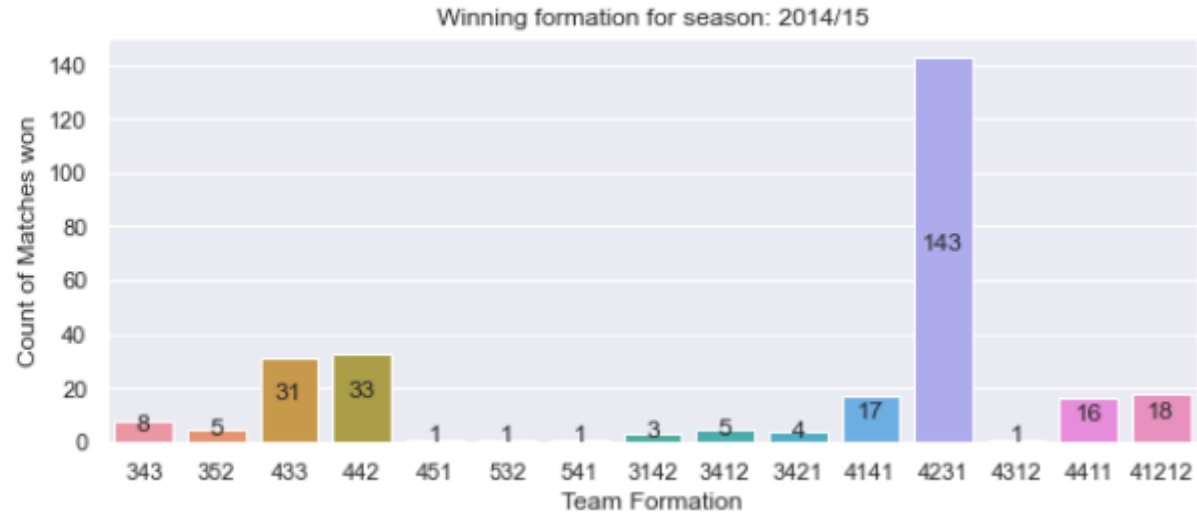
- Number of winning formation which were tried by different team has increased from 2011/12 to 2018/19

WINNING FORMATION ACROSS SEASON



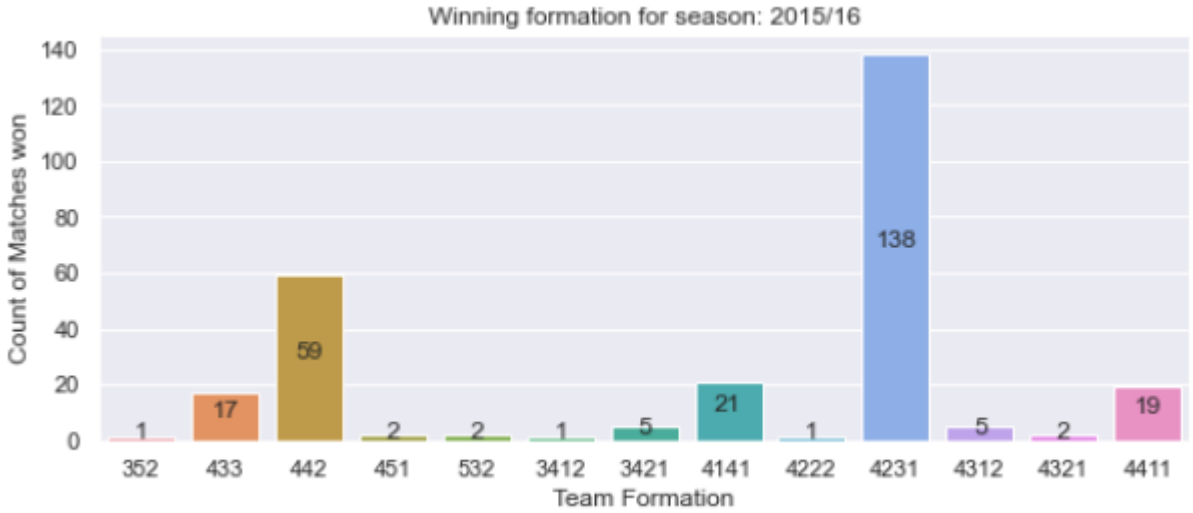


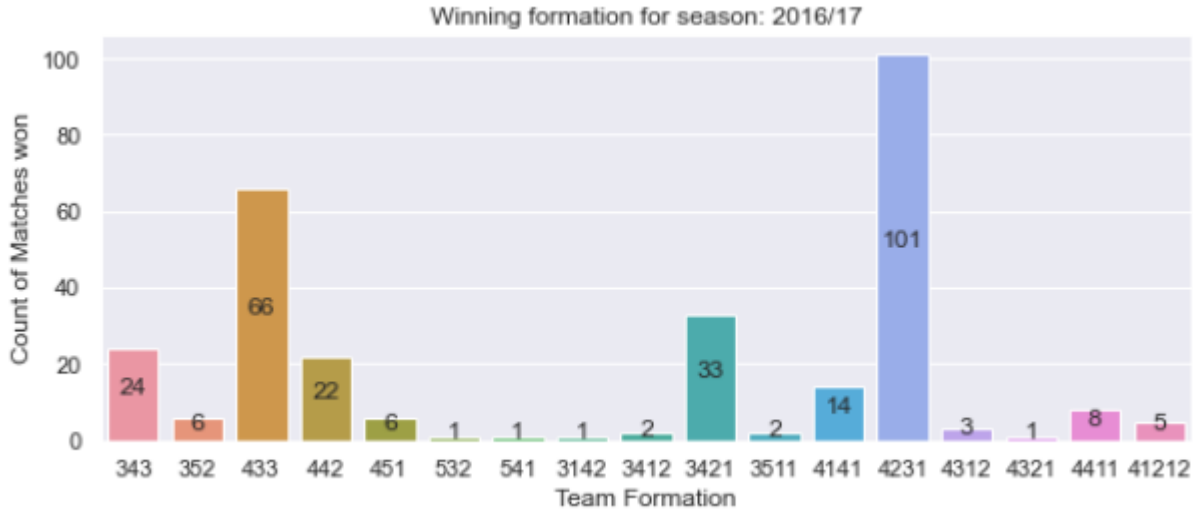


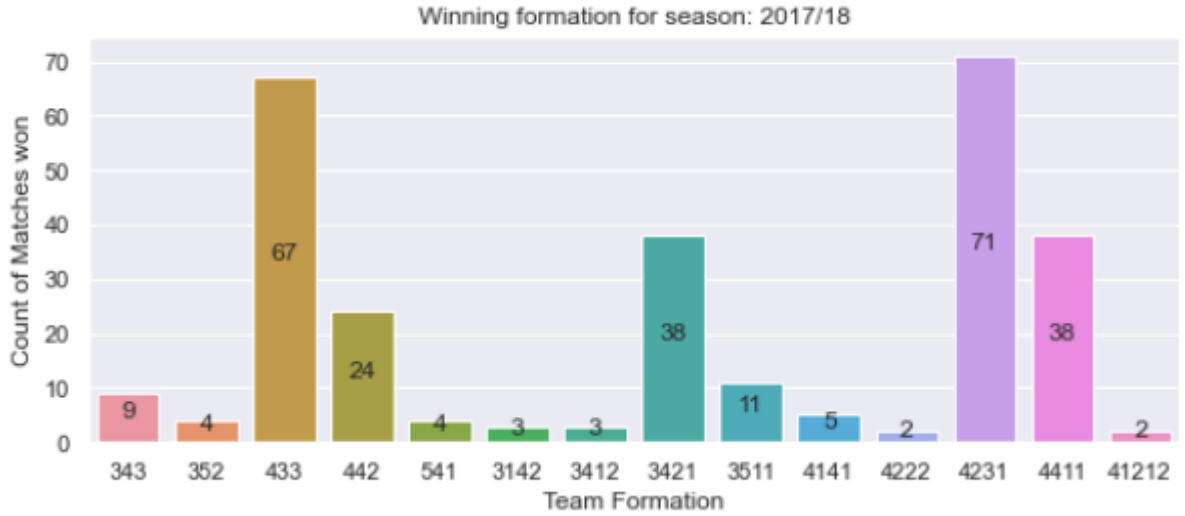


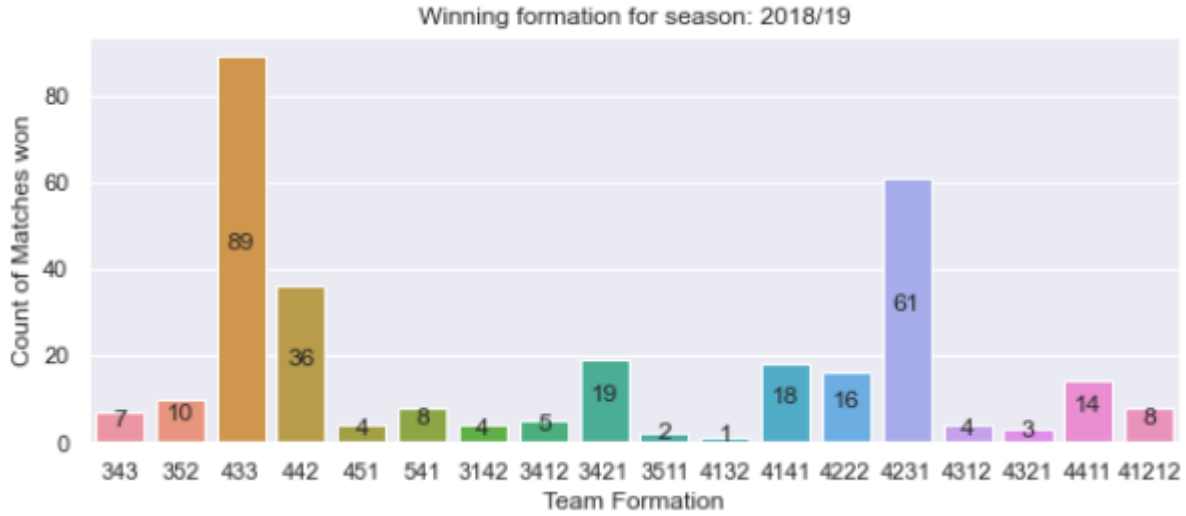
- ***442*** seems to be winning formation tried by team in 2011/12 which gradually changed to ***4231***

WINNING FORMATION ACROSS SEASON









- ***433*** winning formation has been on rise from 2015/16 season

WINNING FORMATION FOR TOP 5 TEAMS

team	seasonlabel	SeasonsRank	WinForm-Season	451	4231	4141	433	343	442	4411	41212	3421
Arsenal	2016/17	5	4231	0% (0)	53.3% (30)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	87.5% (8)
Arsenal	2017/18	6	4231	0% (0)	50.0% (8)	100.0% (1)	44.4% (9)	0% (0)	0% (0)	0% (0)	0% (0)	50.0% (20)
Arsenal	2018/19	5	433	0% (0)	52.9% (17)	0% (0)	0% (0)	66.7% (3)	33.3% (3)	0% (0)	50.0% (2)	57.1% (7)
Chelsea	2016/17	1	4231	0% (0)	0.0% (1)	60.0% (5)	0% (0)	86.4% (22)	0% (0)	0% (0)	0% (0)	88.9% (9)
Chelsea	2017/18	5	4231	0% (0)	0% (0)	0% (0)	0% (0)	75.0% (4)	0% (0)	0% (0)	0% (0)	50.0% (20)
Chelsea	2018/19	3	433	0% (0)	0% (0)	0% (0)	55.3% (38)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Everton	2016/17	7	4231	0% (0)	46.7% (15)	0% (0)	50.0% (12)	0.0% (1)	0.0% (1)	0% (0)	0% (0)	66.7% (3)
Everton	2017/18	8	4231	0% (0)	42.9% (14)	66.7% (3)	44.4% (9)	0.0% (1)	0.0% (2)	0.0% (1)	0% (0)	25.0% (4)
Everton	2018/19	8	433	0% (0)	43.3% (30)	50.0% (2)	0% (0)	0% (0)	0% (0)	0.0% (3)	0% (0)	50.0% (2)
Liverpool	2016/17	4	4231	0% (0)	0% (0)	0% (0)	54.3% (35)	0% (0)	0% (0)	0% (0)	100.0% (2)	0% (0)
Liverpool	2017/18	4	4231	0% (0)	50.0% (4)	0% (0)	53.3% (30)	0% (0)	0.0% (1)	100.0% (1)	0% (0)	0% (0)
Liverpool	2018/19	2	433	0% (0)	90.9% (11)	0% (0)	70.8% (24)	0% (0)	100.0% (2)	100.0% (1)	0% (0)	0% (0)
Manchester City	2016/17	3	4231	0% (0)	68.8% (16)	60.0% (10)	100.0% (6)	0.0% (2)	0% (0)	0% (0)	0.0% (1)	0.0% (2)
Manchester City	2017/18	1	4231	0% (0)	0% (0)	0% (0)	85.3% (34)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Manchester City	2018/19	1	433	0% (0)	66.7% (3)	0% (0)	84.4% (32)	0% (0)	0% (0)	100.0% (1)	0% (0)	0% (0)
Manchester United	2016/17	6	4231	0% (0)	36.8% (19)	0% (0)	52.9% (17)	0% (0)	100.0% (1)	0% (0)	0% (0)	0% (0)
Manchester United	2017/18	2	4231	0% (0)	63.6% (22)	0% (0)	70.0% (10)	0% (0)	0% (0)	0% (0)	100.0% (1)	100.0% (1)
Manchester United	2018/19	6	433	0% (0)	83.3% (6)	0.0% (1)	66.7% (15)	0% (0)	0.0% (1)	0.0% (1)	100.0% (1)	0% (0)
Tottenham Hotspur	2016/17	2	4231	0% (0)	57.1% (21)	50.0% (2)	100.0% (1)	100.0% (1)	0% (0)	0% (0)	0% (0)	80.0% (10)
Tottenham Hotspur	2017/18	3	4231	0% (0)	75.0% (24)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	0.0% (2)	50.0% (2)
Tottenham Hotspur	2018/19	4	433	0% (0)	66.7% (12)	0% (0)	100.0% (1)	0% (0)	100.0% (1)	0% (0)	45.5% (11)	100.0% (2)

- Arsenal
  - Played with 4231 combination in 2016/17 and 2018/19 season and had good success
- Chelsea
  - Played with 343 in 2016/17 and were extremely successful with number one in League table
- Everton
  - Played with 4231 and 433 combination mostly with ~45% success rate
- Liverpool
  - Played mostly with 433 in 2016/17 and 2017/18 but went with combination of 433 and 4231 in 2018/19. They went to number 2 in League table for 2018/19 season
- Manchester City
  - Played mostly with 433 formation with > 80% success in 2017/18 and 2018/19 season. They were number one in League standing in those season
- Manchester United
  - Played with 4231 and 433 combination in last 3 season
- Tottenham Hotspur
  - Played mostly with 4231 with 41212 tired in 2018/19 season

Legend

%age (number) : Winning Percentage (Total matches played in season)

COMEBACK GAMES ACROSS SEASONS

Season	Comeback Games
2011/12	49
2012/13	62
2013/14	51
2014/15	41
2015/16	53
2016/17	46
2017/18	50
2018/19	46

- There has been 40-50 comeback games every season since 2011/12

TOP 5 COMEBACK TEAMS ACROSS SEASONS

Season: 2011/12

Teams	Games
Stoke City	7
Wolverhampton Wanderers	5
Blackburn Rovers	4
Fulham	4

Season: 2012/13

Teams	Games
Liverpool	6
Tottenham Hotspur	6
Everton	5
Reading	5

Season: 2013/14

Teams	Games
West Bromwich Albion	8
Cardiff City	5
Arsenal	4
Manchester United	4

Season: 2014/15

Teams	Games
Burnley	5
Leicester City	4
Swansea City	4
Tottenham Hotspur	4

Season: 2015/16

Teams	Games
Crystal Palace	5
Leicester City	5
West Ham United	5
Newcastle United	4

Season: 2016/17

Teams	Games
Crystal Palace	5
Tottenham Hotspur	5
Liverpool	4
AFC Bournemouth	3

Season: 2017/18

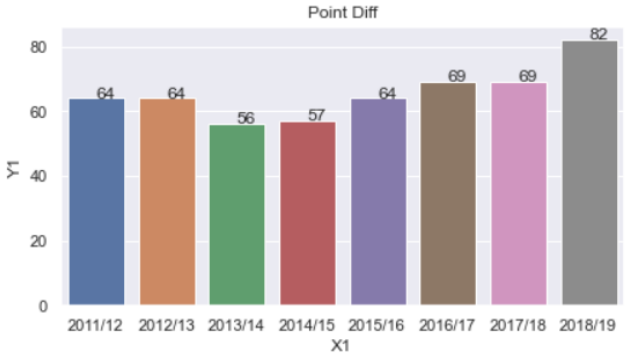
Teams	Games
Crystal Palace	5
Burnley	4
AFC Bournemouth	3
Arsenal	3

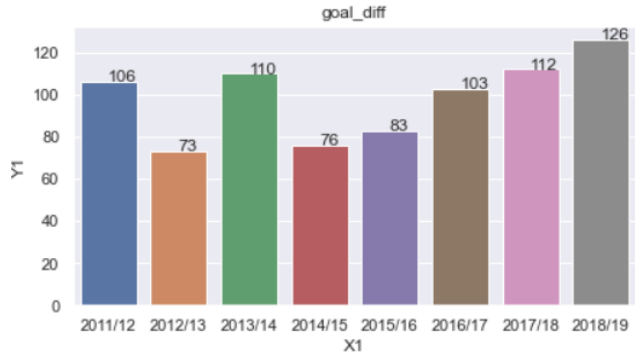
Season: 2018/19

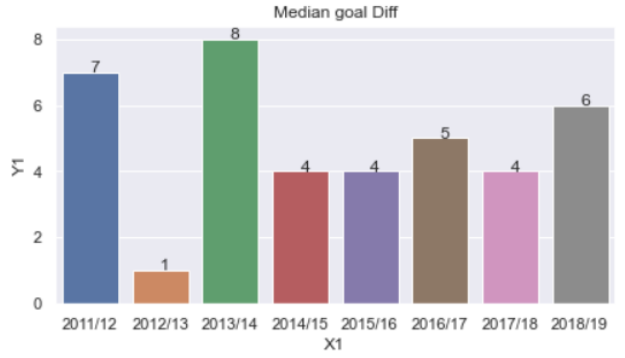
Teams	Games
Watford	5
Brighton and Hove Albion	4
Leicester City	4
Wolverhampton Wanderers	4

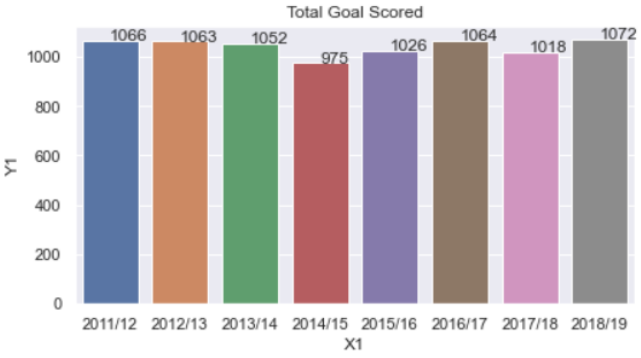
- Crystal Palace made 5 comebacks in seasons 2015/16 – 2017/18
- Leicester City also had 5 comebacks in 2015/16 and they were top of League table in that season

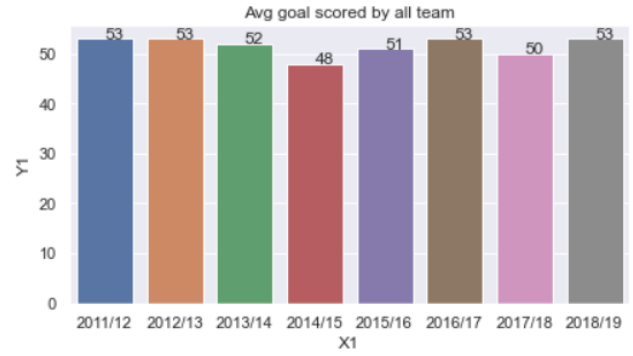
SEASON COMPETITIVENESS

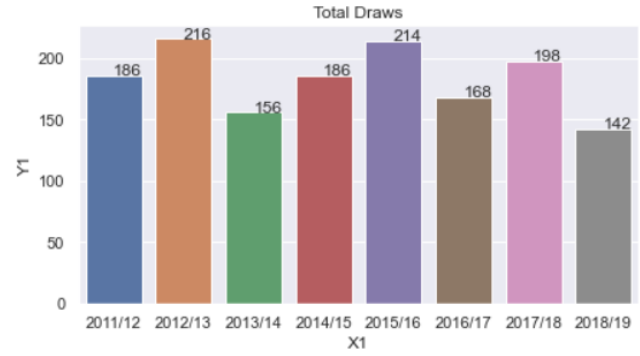


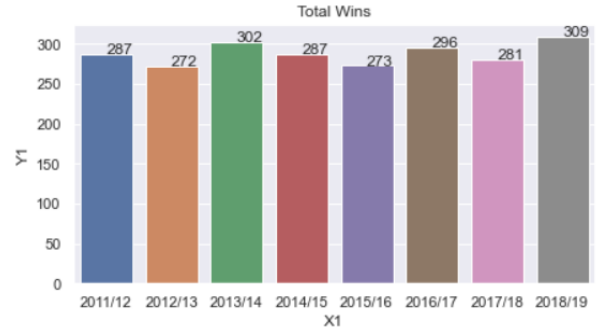


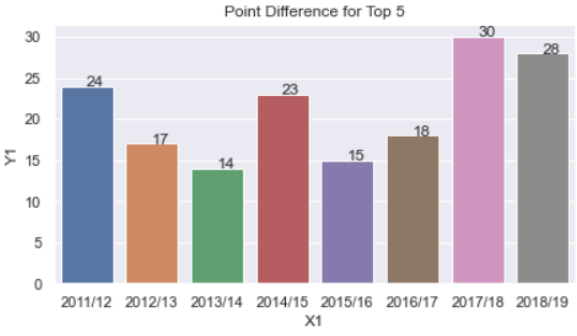












- Based on the different parameters 2016/17 season seems to be most competitive of the 8 season analyzed

PREDICTION USING MACHINE LEARNING

PREDICTING SCORES AND MATCH OUTCOME USING DIFFERENT MACHINE LEARNING MODELS

POISSON DISTRIBUTION

Feature selection for model:

To begin with following features were selected for model:

- game_type: Home or Away Game
- team : Team
- opponent : Opponent playing against team
- teamFormation : Formation of Team
- possession : Possession Pct of Team

Model Used:

- ***Statsmodel glm function***

API to predict Score and Probability :

```
    tm1sc = pmodel.predict(pd.DataFrame(data={'team': tm1, 'opponent': tm2,'possession':pos1,
                                             'game_type': 'H','teamFormation':fm1}, index=[1])).values[0]
    tm2sc = pmodel.predict(pd.DataFrame(data={'team': tm2, 'opponent': tm1,'possession':pos2,
                                             'game_type': 'A','teamFormation':fm2}, index=[1])).values[0]
    sc_mat = util.get_probability(pmodel, tm1sc, tm2sc,max_goals=8)
    wp = util.win_probability(sc_mat)
    dp = util.draw_probability(sc_mat)
    lp = util.loss_probability(sc_mat)
    lst_pred_output.append({'Date':gd,'Team':tm1,'Opponent':tm2,'Actual Score':str(row['score_h'])+':'+str(row['score_a']),'Predicted Score
df_pred_output = df_pred_output.append(lst_pred_output)
```

POISSON DISTRIBUTION : MODEL OUTPUT

Using Poisson model to predict score and outcome for last 10 games of 2018/19 season

Model output:

Date	Team (T)	Opponent Actual	Score(Actual)	Score(Predicted)	Win Prob.(T)	Draw Prob	Loss Prob.(T)	High Probability Event Happened ?
5/12/2019	Brighton and Hove Albion	Manchester City	1:4	0.77:1.85	14.88	21.95	63.16	Yes
5/12/2019	Burnley	Arsenal	1:3	0.9:1.72	19.38	23.71	56.9	Yes
5/12/2019	Crystal Palace	AFC Bournemouth	5:3	1.64:1.18	47.92	24.45	27.62	Yes
5/12/2019	Fulham	Newcastle United	0:4	1.16:1.58	27.91	24.96	47.13	Yes
5/12/2019	Leicester City	Chelsea	0:0	1.16:1.68	26.29	24.14	49.57	No
5/12/2019	Liverpool	Wolverhampton Wanderers	2:0	2.7:0.76	78.06	13.74	8	Yes
5/12/2019	Manchester United	Cardiff City	0:2	2.56:0.58	79.85	13.65	6.37	No
5/12/2019	Southampton	Huddersfield Town	1:1	1.9:0.52	70.46	20.16	9.36	No
5/12/2019	Tottenham Hotspur	Everton	2:2	1.81:0.94	57.65	22.91	19.43	No
05/12/2019	Watford	West Ham United	1:4	1.51:1.2	44.32	25.44	30.23	No

Model change & output:

What if we have half time information and run the model added ? ***Added half time score to the model*** in feature list and model comes out with following output

Date	Team (T)	Opponent Actual	Score(Actual)	Score(Predicted)	Win Prob.(T)	Draw Prob	Loss Prob.(T)	High Probability Event Happened ?
5/12/2019	Brighton and Hove Albion	Manchester City	1:4	0.96:3.25	7.53	11.3	80.54	Yes
5/12/2019	Burnley	Arsenal	1:3	0.64:1.02	22.72	33.65	43.64	Yes
5/12/2019	Crystal Palace	AFC Bournemouth	5:3	4.81:1.43	82.65	6.62	5.11	Yes
5/12/2019	Fulham	Newcastle United	0:4	0.9:2.63	10.51	15.24	74.09	Yes
5/12/2019	Leicester City	Chelsea	0:0	0.87:0.92	32.3	32.97	34.73	Close
5/12/2019	Liverpool	Wolverhampton Wanderers	2:0	2.02:0.66	69.4	19.49	11.08	Yes
5/12/2019	Manchester United	Cardiff City	0:2	1.26:0.94	43.53	28.67	27.8	No
5/12/2019	Southampton	Huddersfield Town	1:1	1.76:0.44	70.04	21.39	8.56	No
5/12/2019	Tottenham Hotspur	Everton	2:2	1.8:0.75	62.56	22.45	14.97	No
05/12/2019	Watford	West Ham United	1:4	1.02:2.38	14.58	17.90	67.44	Yes

PREDICTION OF MATCH OUTCOME WITH OTHER CLASSIFIERS

Feature selection for model:

To begin with following features were selected for model:

- team_h : Home Team
- team_a : Away Team
- formation_h : Formation for Home Team
- formation_a : Formation of Away Team
- possession_pct_h : Possession Pct of Home Team
- corner_taken_h : Corner taken by Home Team
- corner_taken_a : Corner taken by Away Team

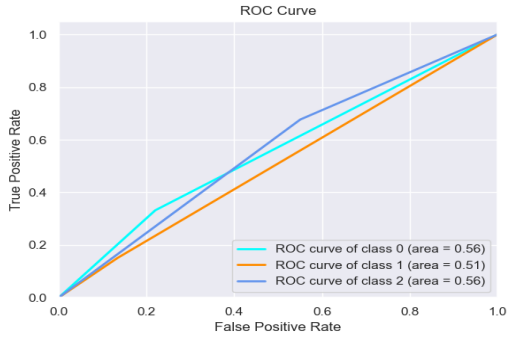
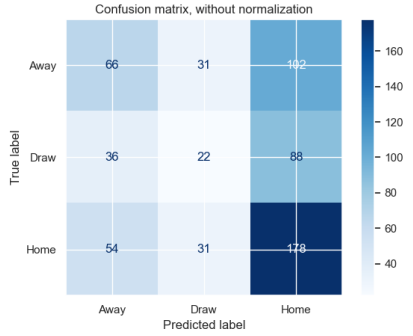
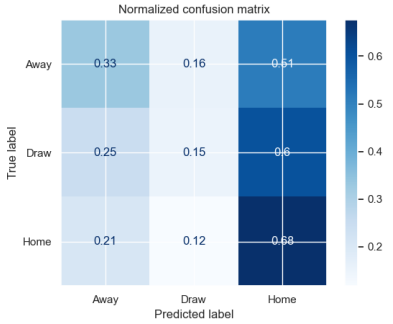
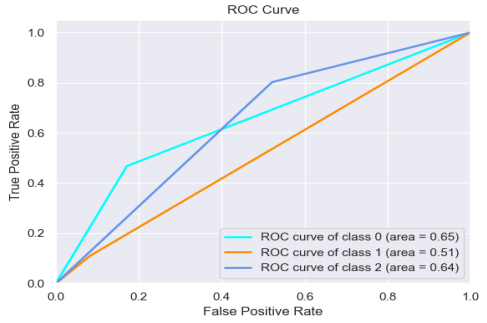
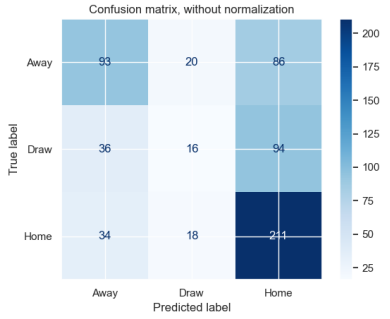
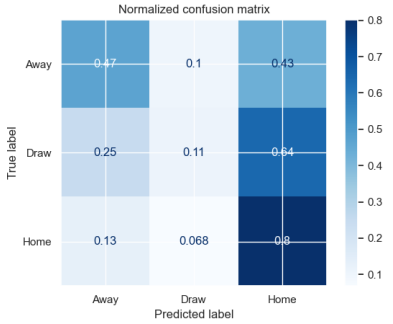
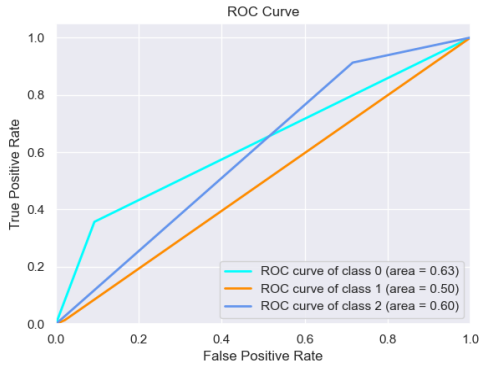
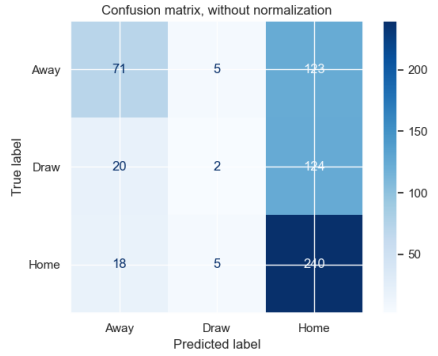
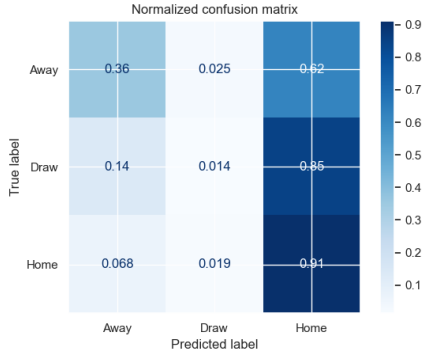
Model Used:

- KNeighborsClassifier
- RandomForestClassifier
- SVC

MODEL OUTPUT

Model	Accuracy Score	Precision	Recall	F1-Score	ROC Area	Support	Prediction
KNeighborsClassifier	0.44	0.42	0.33	0.37	0.56	199	Away Win
		0.26	0.15	0.19	0.51	146	Draw
		0.48	0.68	0.56	0.56	263	Home Win
RandomForestClassifier	0.53	0.57	0.47	0.51	0.65	199	Away Win
		0.3	0.11	0.16	0.51	146	Draw
		0.54	0.8	0.65	0.64	263	Home Win
SVM	0.51	0.65	0.36	0.46	0.63	199	Away Win
		0.17	0.01	0.03	0.50	146	Draw
		0.49	0.91	0.64	0.60	263	Home Win

MODEL OUTPUT

	ROC	Confusion Matrix	Confusion matrix (Normalized)
KNeighborsClassifier			
RandomForestClassifier			
SVM			

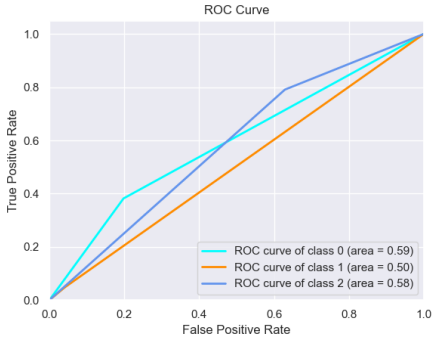
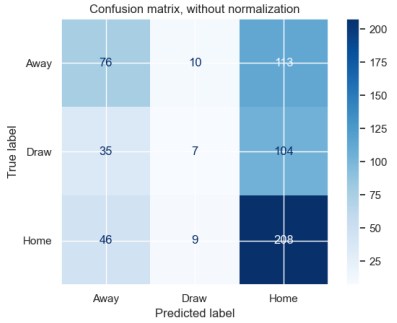
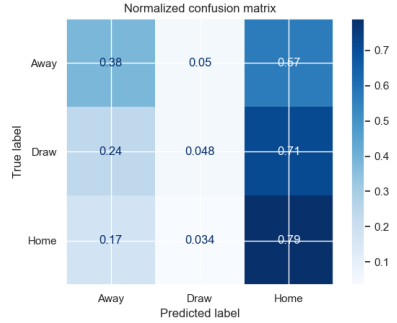
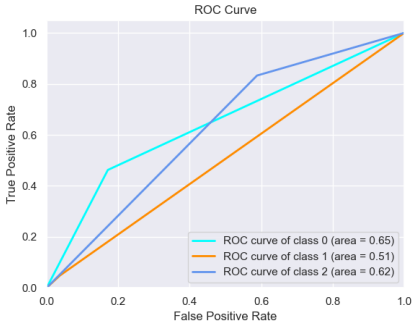
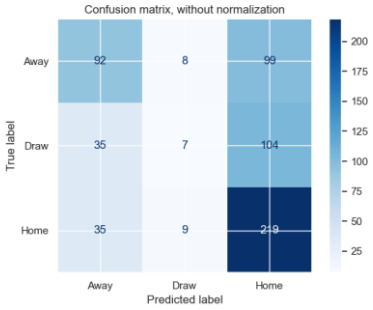
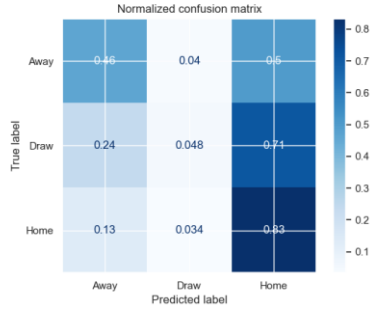
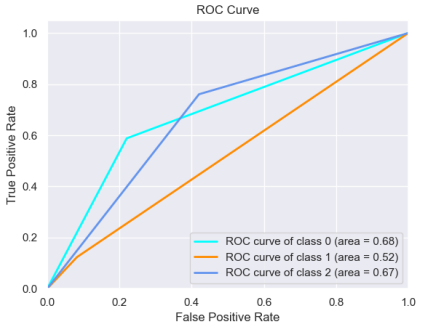
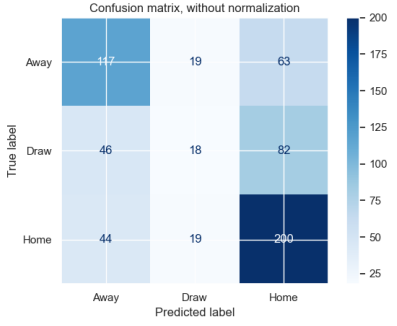
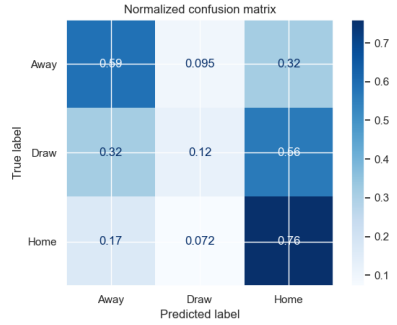
IMPROVING THE MODEL

- Applying feature scaling using StandardScaler
- Applying cross validation using RandomizedSearchCV

MODEL OUTPUT

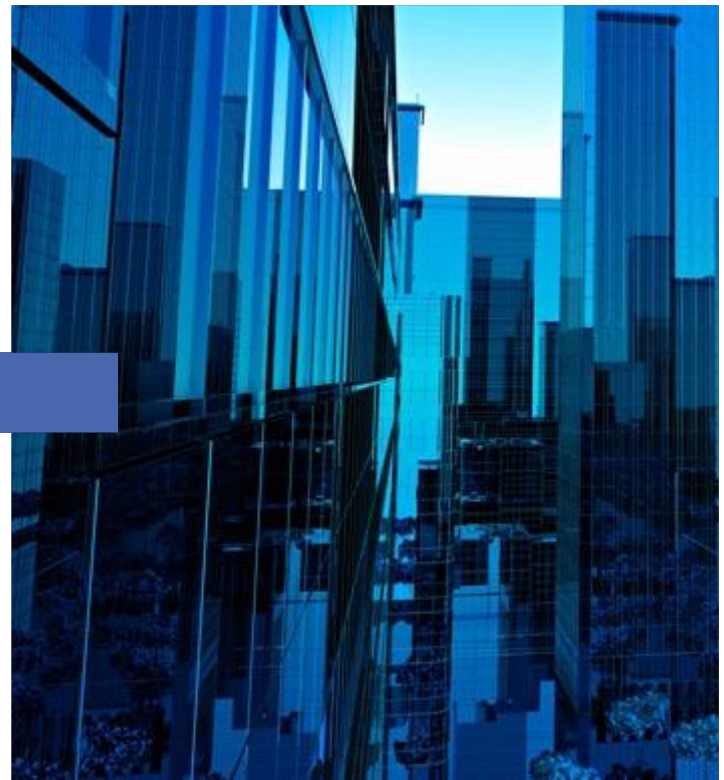
Model	Accuracy Score	Precision	Recall	F1-Score	ROC Area	Support	Prediction
KNeighborsClassifier	0.48	0.48	0.38	0.43	0.59	199	Away Win
		0.27	0.05	0.08	0.50	146	Draw
		0.49	0.79	0.6	0.58	263	Home Win
RandomForestClassifier	0.53	0.58	0.48	0.52	0.65	199	Away Win
		0.41	0.05	0.09	0.51	146	Draw
		0.52	0.84	0.64	0.62	263	Home Win
SVM	0.55	0.57	0.59	0.58	0.68	199	Away Win
		0.32	0.12	0.18	0.52	146	Draw
		0.58	0.76	0.66	0.67	263	Home Win

MODEL OUTPUT

	ROC	Confusion Matrix	Confusion matrix (Normalized)
KNeighborsClassifier			
RandomForestClassifier			
SVM			

CONCLUSION

- Using the past data we ran exploratory data analysis to understand some trends and relational between different variables or aspects of the game.
- We built different Machine learning model to predict the probability of winners and to predict the outcome of the game using the same historical data
- Classification Machine learning models were improved 4%-5% by applying techniques of feature scaling and right parameter selection. These can be improved further by applying or including additional features in our models



THANK YOU