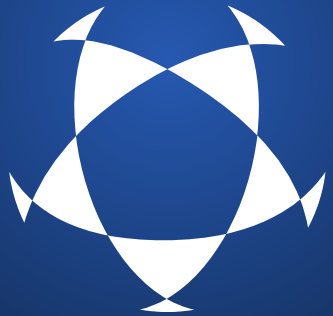
POSSESSION RATE vs SUCCESS

By Burak Basogul

TABLE OF CONTENTS



INTRODUCTION

- Client: Super Lig team in Turkey
- Is there correlation between possession rate and rankings in top European soccer leagues?
- Results will aid in narrowing down candidates to be tracked for future coaching role(s).

I

BACKGROUND

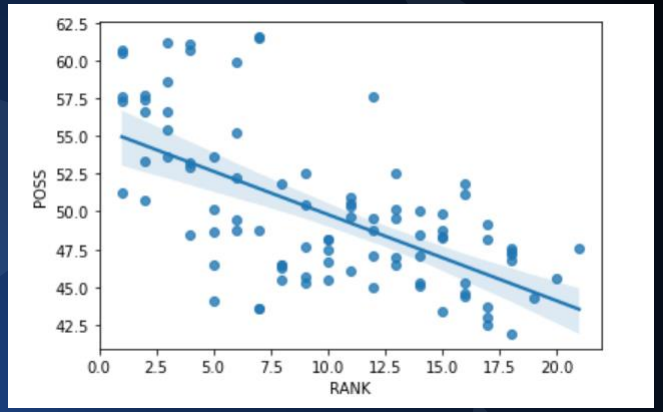
Leicester City	EPL Rank = 1	POSS = 42.6%	2016
PSV Eindhoven,	Eredivisie Rank =1	POSS = 50.4%	2018
PSV Eindhoven,	Eredivisie Rank =1	POSS = 51.7%	2015
Monaco	Ligue1 Rank = 1	POSS = 51.9%	2017
Inter	Serie A Rank = 1	POSS = 52.4%	2021
Atletico Madrid	La Liga Rank = 1	POSS = 52.4%	2021

- Are possession rates just over 50% sufficient to win the championship in a top tier European soccer league OR does a team need more often to be dominating the possession to reach the top spot?

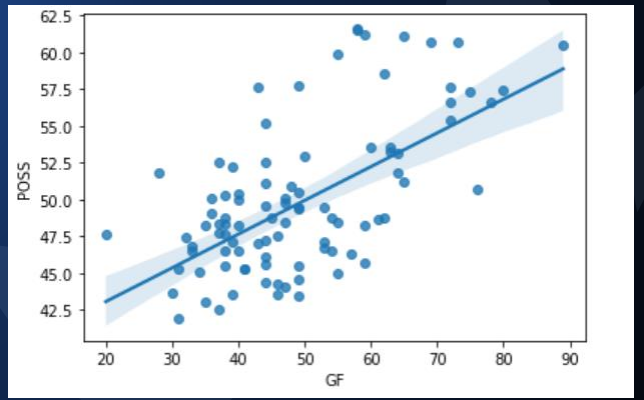
2

BASELINE

- Limited available data for possession rate in Super Lig, only for five seasons.
- Started modeling with RANK, WIN (W), LOSS (L), GOALS FOR (GF), GOALS AGAINST (GA) & POINTS (PTS) features for POSSESSION (POSS) target variable.
- Detected multicollinearity and eliminated collinear features one by one based on Variance Inflation Factor (VIF) values.
- Resulting model included features RANK, GF and GA with a R^2 value of 0.478.



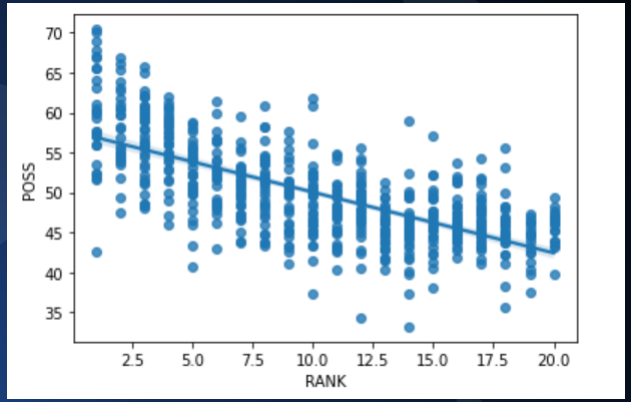
Super Lig – 2017~2021



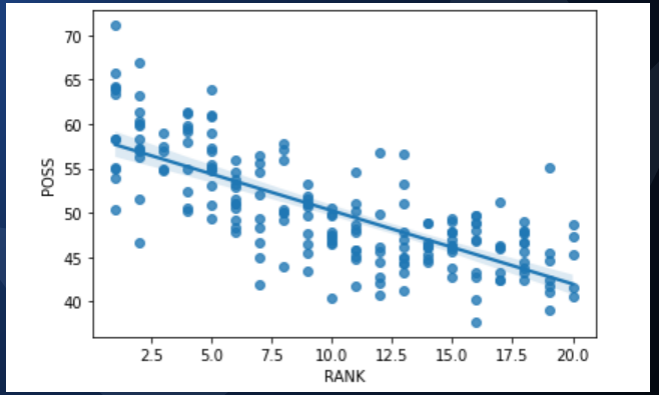
Super Lig – 2017~2021

3
ANALYSIS

- Combined tables from top 8 European leagues from 2015 thru 2021.
- Split data into Training, Validation and Test Sets (60,20,20)
- Started modeling train data with RANK, WIN (W), LOSS (L), GOALS FOR (GF), GOALS AGAINST (GA) & POINTS (PTS) features for POSSESSION (POSS) target variable.
- Detected multicollinearity and eliminated collinear features one by one based on Variance Inflation Factor (VIF) values.
- Resulting train model included features RANK, GF and GA with a R^2 value of 0.578.



UEFA Leagues Training Model – 2015~2021



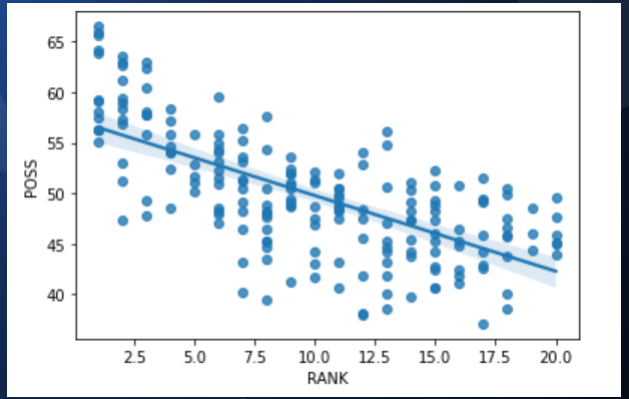
UEFA Leagues Validation Model – 2015~2021

3
ANALYSIS

- R^2 values did not improve after removal of outlier data
- Scored test data as **R^2 = 0.594**
- Validated findings via sklearn modelling

Training Data	Validation Data	Model Package
R^2 = 0.578	R^2 = 0.639	Statsmodels
R^2 = 0.586	R^2 = 0.620	Scikit-Learn

Training Data	Validation Data	Outlier Data
R^2 = 0.578	R^2 = 0.639	Included
R^2 = 0.532	R^2 = 0.603	Excluded
RMSE = 3.98	RMSE = 3.76	Included
RMSE = 3.97	RMSE = 3.87	Excluded
MAE = 3.10	MAE = 2.92	Included
MAE = 3.13	MAE = 2.93	Excluded



UEFA Leagues Test Model – 2015-2021

4
CONCLUSION

- Modelling shows that in top European leagues possession rate have significant direct correlation with ranking, goals scored and inverse correlation with goals conceded.
- Client is advised to identify candidates for coaching role(s) who have managed teams that demonstrated possession rates above 55% over a full season.