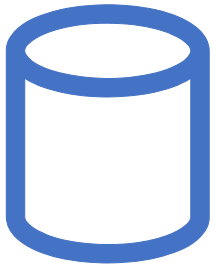
Predicting Playing Time for Women's Soccer

Nancy Binowski

Can we predict playing time for women soccer players?

Data, Tools, Algorithms

Tools:
- Python
- BeautifulSoup,requests
- Pandas
- Re
- Numpy
- Seaborn
- Matplotlib
- Statsmodels
- sklearn

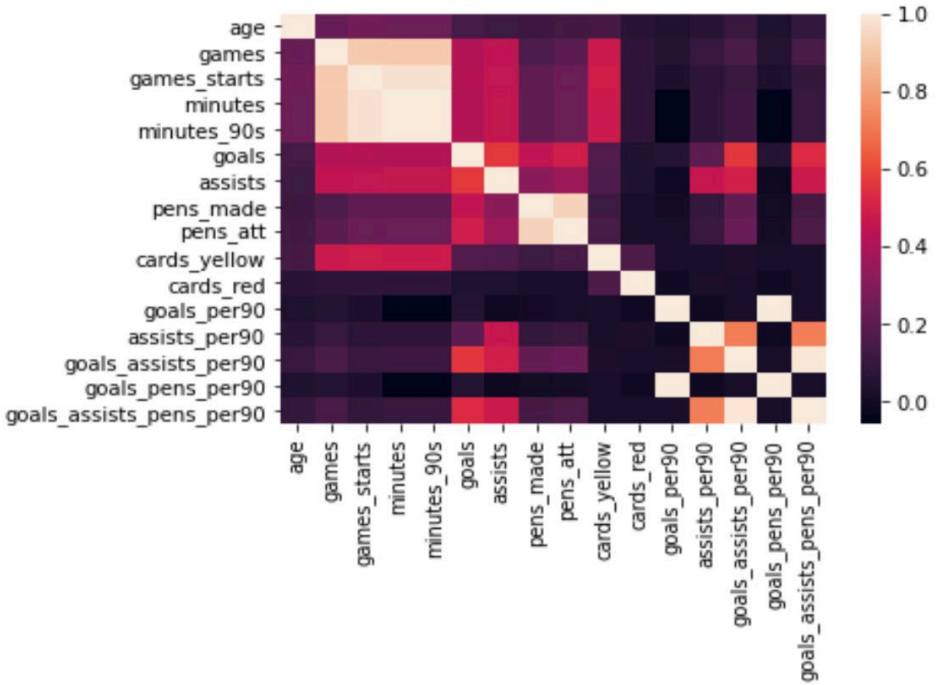


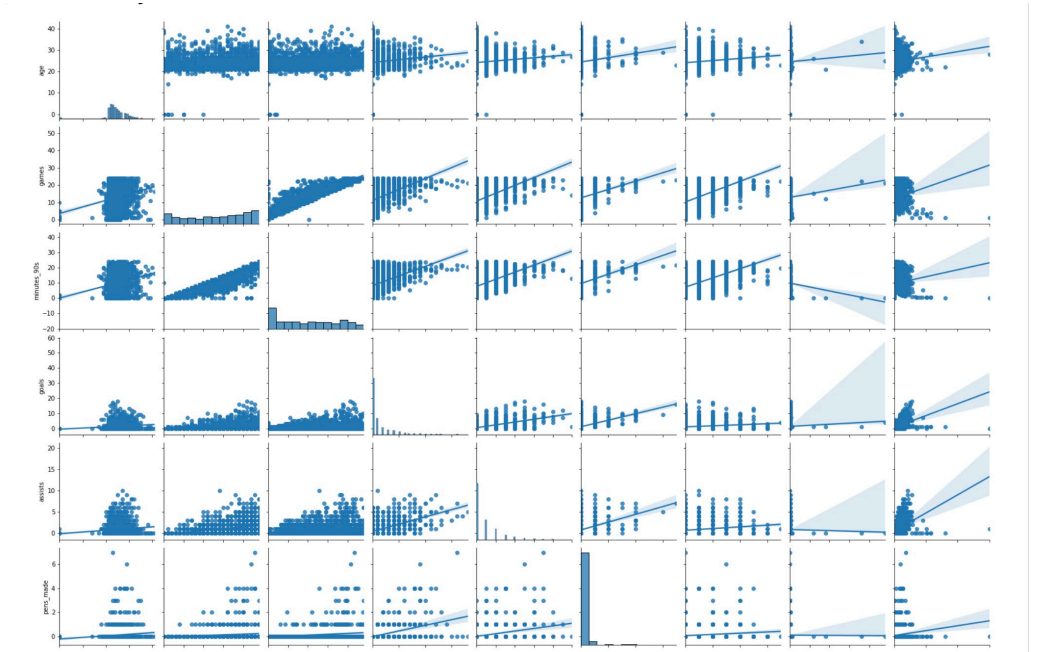
Features considered:
- Age
- Games
- Game starts
- minutes
- Goals
- Assists
- Penalty shots made
- Penalty shots attempted
- Yellow cards
- Red cards
- Various features/90 minutes

- Data webscraped from FBref.com
- Scope of Data:
  - US National Women Soccer League
  - Years 2013-2019
- Data Cleaning
  - Excluded goal keepers
  - Missing data replaced with 0 /none
  - Reformat some data: age, minutes to convert to int
  - No outlier data needed to be removed

Preliminary Analysis

- Correlation analysis showed correlation with minutes/minutes90 and several features
- Plotting showed linear relationships





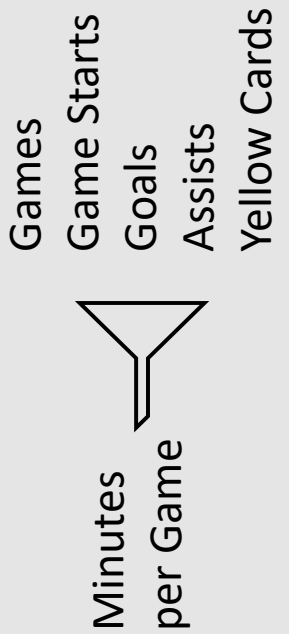
Feature Engineering

Analysis of correlation between independent variables (VIF) resulted in the following features with a target being minutes per game (minutes90)

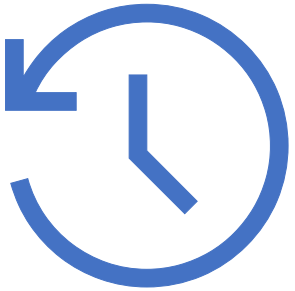
- Age
- Games
- Goals
- Assists
- Penalty shots made
- Yellow cards received
- Goals and Penalties per game
- Goals, Assists and Penalties per game

Results

- OLS Model Accuracy (R-squared): 0.857
- RidgeCV Model Accuracy: 0.955
- LassoCV Model Accuracy: 0.955
- ElasticNet Model Accuracy: 0.955



Future Study



- Test model on Olympic Player data and compare results with NWSL model